A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

Misbehavior Vs Market Efficiency in the context of International Football

A Special view on Academy Players

João Maria Ferraz Leal Trovão

26185

A Project carried out on the Masters in finance Program, under the supervision of:

Prof Francisco Queiró

January 3rd 2020

Abstract

This paper tackles the reality of academy players in the context of the international football world. A set of behavioral and non-behavior phenomena are studied to understand if clubs are biased towards buying players in the market rather than betting in their own players. A mathematical analysis is conducted to analyze both the non-academy and academy player's variation of the market value in a window of 1 year and conclude if indeed there is an underestimation towards academy players. This paper allows for a financial study on football players valuation in a short-term window of time and the consequent return to clubs.

Keywords

Sport, Behavioral bias, Market Value

Introduction

Rational expectations and market efficiency are two essential pillars in neoclassical economics. First, by the theory of rational behavior people are expected to act rationally in the process of their economic decisions. Second, consumer perceptions of a certain product are expected to affect its price and demand, which along with competition lead, ultimately, to an efficient allocation of resources within the economy, creating the market equilibrium. In short, people are expected to make unbiased predictions and the market incorporates them into unbiased estimates with useful value.

In this paper we tackle these two building blocks of modern economics in a different but stimulating way: the European football context, particularly the choices of European clubs between their academy players and buying new players. Implicitly, teams that choose to invest in a new player rather than just promoting an academy player to his place are predicting that the first one will bring greater benefits than the latter.

What drives a team to buy a player instead of betting on a youngster from its own academy? Do teams, nowadays, have a significant bias towards buying players, throwing their own products to second plan? In this paper, we investigate deeply what drives teams to choose one way over the other and what are the results of this choice. The paper's initial conjecture is that teams do not have rational expectations when predicting future performance between academy and non-academy players.

It seems that teams incur in a combination of behavioral and non-behavioral phenomena that work towards a systematic bias: teams overestimate non-academy player performance in a way that is inconsistent with rational expectations. A combination of known behavioral episodes appear to open roads where teams underestimate the return academy players can provide to the club both in terms of performance and financials. It was considered that this phenomenon would not be eliminated by market forces because, even if there are some smart teams that do not follow this pattern, they are just outliers in a trend with hundreds of clubs. Hence, these few teams do not hold the power to change the status-quo.

In this paper it was concluded that academy players have a greater percentual change, in one year, in their market value compared to non-academy players. From this follows that teams wanting short-term financial results are better off by betting in their academy players rather than searching the market for new players with the same expected quality. Meaning that, clubs have a higher financial return from their academy players due to the increased variation of their market value in the same window of time as other players. As the horizon of time is equal between these two classes of players (1-year window) one can only conclude that, in the short-term, teams ought to bet in their academy to achieve increased financial results.

Literature Review

The transfer market implicates two processes which have constantly been under spotlights in the psychology field – predicting the future and bidding competitively. Psychological researchers have provided a span of fundamental studies on these two tasks which will be the roots of this paper. Even though we cannot be certain that any of these biases are to be blamed for the decision making, surely, they invigorate the central prediction: teams overvalue players in the transfer market over academy players.

In the psychological field, overconfidence is a concept that goes hand in hand with the process of predicting the future. This effect is a well-known bias where a person's confidence in his judgements is greater than the objective certainty of those same judgements. Meaning that, in the football world, coaches and scouts believe their sporting knowledge is more precise than it is in fact. Following Don Moore & Paul Healy's paper in 2008 on overconfidence, this idea can be separated into 3 pillars: overestimation, overplacement and overprecision. The first variety can be thought as an overestimation of a person's actual ability, performance, or level of control. For example, a student incurs in overestimation when, in a 10-question quiz, he believes he has 5 correct when in fact he only answered correctly in 3 of them. The second definition occurs when people think of themselves as better than the median. If a student thinks his score is the best in class when, truly, half of the class scored above him, he is overplacing his score relative to others. Lastly, overprecision can be described as "excessive certainty regarding the accuracy of one's beliefs". Researchers have been thoroughly studying this concept using numerical questions (e.g. "How long is the Nile River?") and asking individuals to project 90% confidence intervals. Results demonstrate that these intervals are often too narrow, meaning that participants are too convinced they know the precise answer. These 90% confidence intervals embody the correct answer less than 50% of the time (Alpert & Raiffa, 1982; Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999; Soll & Klayman, 2004). This shows that people are not recognizing the limits of their cognitive capabilities nor applying sufficient weight to the world's uncertainty.

For the purpose of this paper it is also essential to understand how individual's confidence varies with the amount of information available. Rationally one can deduce that confidence in the decision-making process will increase when people are provided with more information, however, frequently, this confidence increases more than the actual capacity of predicting the future. Stuart Oskamp (1965) studied this phenomenon through an experience whereas participants analyze information about a case "(a) their confidence about the case increases markedly and steadily but (b) the accuracy of their conclusions about the case quickly reaches a ceiling". The trial was divided into 4 sections and after judges read each one of the sections, they had to answer a set of 25 questions about the case. Throughout the 4 stages of the study, the accuracy of the participant's answers maintained rather constant, but confidence rose steeply with the increase of information, supporting the initial hypothesis of the paper.

Accordingly, individuals shifted from being reasonable in the beginning to being highly overconfident in the presence of more information.

In the context of international football, clubs experience similar situations to Stuart Oskamp's experiment. Consistently teams must evaluate players while increasing the amount of information about them. Clubs often follow players from a young age, increasing drastically the available information when they reach the professional or international level. When coaches and scouts actively target a player, additional observations are conducted to present a thorough report on the strengths and weaknesses of the athlete along with the possible benefits for the club if they sign the player. Although one may think that such reports improve club's decisions about players, the previous research points the other way. On the contrary, in the face of complete information on players, a team's confidence on their capacity to differentiate between athletes may surpass any true upgrade in their judgements.

Competitive bidding brings another set of topics into the picture. As Thaler explains in his 1988 paper when many players compete for a product with a common but uncertain value, the winner of the auction often overpays. This phenomenon is the well-known winner's curse, "first discussed in the literature by three Atlantic Richfield engineers, Capen, Clapp, and Campbell (1971)" on oil-lease bids (Richard H. Thaler, 1988). As the author points, avoiding this phenomenon is not easy, even in the case where bidders have unbiased expectations of the product's value, the winner of the auction is very probable to be a player who has overestimated the true worth of the item. If teams were to be rational during the bidding process, they would reduce their bids, especially when the number of clubs in the auction increases, eliminating the winner's curse anomaly. However, as reported by Kagel & Levin in 1986, "Auctions with large numbers of bidders produce more aggressive bidding than with small numbers, resulting in negative profits, the winner's curse". Thus, clubs often tend to facilitate this phenomenon when a large number of teams are in the race for a player, causing an overvaluation of the athlete.

Consequently, the winner team overpays for the athlete and gains a smaller profit than expected or, in sometimes, it even incurs in a negative profit. Following this line of thought it is easy to understand that, systematically, teams overpay for players in the transfer market when, if they were to be rational, it would be more profitable to bet on an academy player with the same expected quality. Betting on the academy does not involve any kind of bidding process with other clubs and, furthermore, with the lower wages and familiarity to the club of academy players, teams end up saving valuable money.

Richard Thaler discusses in his 2015 book “Misbehaving” the bias known as *the weight of the present* and how this concept affects the task of teams competitive bidding. This idea suggests that owners, managers and coaches of a club feel the urge to win now. Therefore, when bidding against other teams in the transfer market there is the illusion that the desired player will turn a losing team into a winning team. It is of the utmost importance to note that this phenomenon leads clubs searching the market for players. As academy players are still young athletes, the structure of the club feels that these athletes do not have the capacity to change the present situation right now. In their eyes the best solution is to find desired players in the market and guarantee their signings hoping they can turn the situation around. Ultimately, the weight of the present undermines a club’s confidence in their academy products.

Together, all these biases lead teams towards undervaluing academy players. These phenomena are the foundations to a story where clubs are blindly used to shop in the market to solve their problems instead of betting in their own products. Rationally, there are powerful incentives for clubs to overcome these biases and football has been around for long enough to have valuable market information that clubs have had time to study and learn the consequent lessons. It is fundamental to understand that sports are one of the few industries where employers can easily monitor the performance of candidates who were hired but also the ones that could not get the job. For example, every club in the world observed Renato Sanches’s failure in Munich after the 2016 Euro, not only Bayern. This is just one between thousands of available examples for clubs which should, in theory, facilitate learning.

However, there are also non-behavior explanations to justify the fact that teams prefer to buy new players instead of promoting their own young players. Corruption seems to have a significant role in today's football world, and it shows no signs of slowing down. Today's transfer market consists in deals between player's agents and club executives which are ultimately crooked and made under a "culture of corruption" that regulatory bodies seem to look away from. The football's transfer market allows for corruption by placing agents in a position which provides them with too much power (Michael J. Weir, 2007). Agents take advantage of this positioning and often bribe club executives to buy or sell players damaging the entire transfer system. Hence, clubs across the world are indeed biased towards buying players because it is in their executive's best interest to do so. Managers that run clubs act against their own organization and put their personal interests in front of the team's sustainability. What is alarming about this reality is that it seems to be reinforced every year and it is becoming part of our sports culture. For example, one can just look at the empire created by Jorge Mendes (Gestifute) and see the dimension of this actuality. Clubs like Wolverhampton, Benfica or Atletico Madrid too often negotiate players with a hand of Mendes present, and these are dangerous paths for football to go through.

The Football Transfer Market

Carmichael & Thomas (1993) divide a formal transfer market into two pillars: " (1) to facilitate and organize the acquisition and exchange of players by clubs to enable the reconstitution of teams with the aim of increasing player strengths and improving team performance; (2) to facilitate the movement of players between clubs in their search for better opportunities, higher earnings or increased job satisfaction". Of course, all sporting transfer markets have a set of restrictions and regulations through which clubs and players guide their actions. Normally, these controls prevent players from moving to new clubs without the approval of their current team and the obligation for all transfers to be inside the respective regulatory body's guidelines. Withal, the controls imposed in these markets differ between sports.

For the purpose of this paper it is necessary to explain how the football labor market, meaning the transfer market, works. The most notable difference to "usual" labor markets is that in the football market, players cannot resign their jobs with the same ease as a regular worker in a company. The transfer market has made fundamental changes towards a world where players experience greater freedom and power. Peter J. Sloane (1969) compares the pre-1961 market to a slave market because players had a maximum wage and unless a club allowed to sell an athlete, they were "trapped" in the team as long as the owners wanted (Retain and Transfer System). Afterwards, in 1961 the maximum wage was eradicated and the retain and transfer system was altered in 1963, however a transfer fee was still necessary if the club so decided. This system prevented significant differences between clubs, meaning that, it did not allow a considerable concentration of high-quality players in few teams. If bigger teams desired star players, a transfer fee (monetary compensation) was demanded by smaller clubs. Accordingly, this incentivized teams to develop in-house players due to the compensations for their investments.

However, the football transfer market would suffer key changes with the Bosman ruling, ending the controversial retain and transfer system. This decision turned, formally, football into an economic activity, hence, being subject to the provisions of the treaty of Rome regarding the freedom of movement of the players. Meaning that, this ruling now gives players whose contracts are expired the right to seek employment in any club desired, without the former club receiving any monetary compensation for the departure. As previously explained, this did not happen before the Bosman case as clubs would still receive financial compensation for out of contract players. The Bosman ruling also had another key implication to the world of football. Before this verdict clubs could only have three foreigners in their roster. After the decision clubs were allowed to employ as many UE-citizens and citizens from countries with agreements with the EU as they pleased. However, it is important to note that citizens from nations unprotected by the Bosman ruling are still subject to national regulations, differing across countries, nonetheless, being usually more restrictive. For example, in France players from outside the "Bosman Area" must have at least 1 international cap to receive a work permit.

"The challenge for economic theory is to find a dynamic balance between love and money necessary to analytically grasp the passionate and pragmatic complexities of the beautiful game" (Vrooman, 2007a). As this quote suggests, football is different from the "average" industries we are used to study in the business world. Workers in this market, meaning players, are earning millions of Euros per year and have the objective of maintaining club reputation. Satisfying fans is one of the top priorities of cub owners, and, in the point of view of these fans, winner clubs in football are the ones that win trophies not the ones that generate more revenue. Owners understand that, in order to win and satisfy fans, it is essential to secure the best players in the market, nonetheless, competition is intense. For example, global spending on transfer fees during the 3 months of summer of 2018 was approximately USD 5.44 billion, constituting a new record for the industry, with the big 5 leagues having a fundamental role spending USD 4.21 billion. In total, until September 2018 transfer fees amounted to USD 7.10 billion, being 11.5% more than the amount registered during the entire 2017 year (FIFA big 5 summer, 2018). Accordingly, it will be important to spend some time explaining how football clubs behave in such powerful industry so that the assumptions/expectations made are realistic.

Football club behavior

Rationality and Profit Maximization

As previously explained, the first thorough study on the economics of sports was performed by Simon Rottenberg in 1956 where he studied the American baseball labor market. Rottenberg formulated the uncertainty of outcome hypothesis where the welfare of the league depends on the "equal distribution of playing talent among opposing talents" (Rottenberg, 1956). No team can become too dominant leading to a retention of player talent which, ultimately, drives consumers away. The other important piece for this paper is that Rottenberg considers that club owners are rational profit maximizers: "A rational team will seek to maximize the rent it derives from each player" (Rottenberg, 1956). These two pillars combined lead us to a critical finding: "the relationship between revenue and number of star players turns negative at some point" (Anders & Christian Gulbrandsen, 2011). Accordingly, at some time, a small team will value a star athlete higher than a big team, concluding that, the profit maximizing solution would be a state where clubs are nearly equal and, therefore, the league is sustainable. However, a scenario in which clubs are equal across leagues is most definitely not the case today.

Win at any cost vs. sustainability

Another important view on the behavior of football clubs is the Sloane (1969) approach where teams are viewed as utility maximizers subject to solvency constraints. For Sloane it is the structure of clubs, meaning owners and managers, that play with the weights given to performance and profits. The problem associated with this power are factors causing dangerous non-profit maximizing behavior. For example, wealthy owners or demanding supporters can lead to a risky willingness to win at any cost. In these scenarios, the organization of a club is forced to shift their principal focus to the sporting performance and fans, maximizing the combination of these two elements. The key takeaway from these situations is that this type of reckless behavior often leads to the breakdown of the NPV condition.

Research Hypothesis

The international football market seems the perfect landscape to overcome the previous behavioral phenomena, due to its particular conditions, however, as the author of Moneyball, Michael Lewis, considers: "If professional baseball players, whose achievements are endlessly watched, discussed and analyzed by tens of millions of people, can be radically mis-valued, who can't be? If such a putatively meritocratic culture as professional baseball can be so sloppy and inefficient, what can't be?". Thus, it would be no surprise if the football market was to be labelled as "sloppy and inefficient" as Michael Lewis found the major league baseball to be.

Academy players are valuable. As the European Club Association affirms "It makes sense to invest in youth development (because with an efficient youth academy the clubs save money on transfers and inflated salaries") vis à vis to the process of bringing a new player with the same expected quality into the organization. Also, the player's loyalty, identification with the club and the supporter's base will grow with this bet. This suggests that if teams are profit maximizing on their choices when signing players, academy players ought to be a more rational choice than non-academy players, having the same expected quality. Teams seem to be overvaluing both return and capacity of non-academy players over in-house players and, constantly, they end up paying substantial quantities of money to secure those players instead of betting in their own talents.

The main hypothesis of this paper is that it is more advantageous, in financial terms, for teams that want short-term results to bet on academy players rather than spending money in the labor market searching for athletes with the same expected quality. Consistently with the behavioral phenomena above presented, teams are biased towards buying players in the market, regularly overpaying for them. In this paper it is evaluated the change in the market value, in one season, of both non-academy and academy players throughout a set of different teams and countries. The study investigates whether academy players have a greater percentual change in their market value compared with the other players, evaluating if, in fact, financially, it is more valuable for clubs to wager in their own players instead of buying players with the same projected quality. As previously mentioned, it is important to remember that when betting in academy players, no transfer fee is payed which immediately brings down the costs associated with this type of players.

Data Explanation

For the purpose of this paper data was collected from transfermarkt regarding player's personal information and market information. Transfermarkt is a German-based website where valuable information on the football world is available to everyone. All kinds of data are present in this website, from results and transfer news to players and club's values. Concerning personal information, it was gathered information on their (1) nationality, (2) age, (3) height and (4) dominant foot. On the market side, Transfermarkt provided information on the (5) previous club, (6) the actual club, (7) the country of the actual club, (8) the length of the contract, (9) the date of entry in the team, (10) the position, (11) whether players are from the academy or not, in a window of 1 (Formation1y), 2 (Formation2y) or 3 years (Formation3y), and most important, the difference in one season of the player's market value, both in (12) absolute value and (13) percentage. The data was set up in an Excel workbook divided into leagues. Each league (sheet) had all this information about players, teams and countries. After that, a panel data was constructed in another sheet, with all league's information in one sheet, to analyze the results in Stata.

The database consists in 803 different observations from 31 different European clubs. This data was collected from 7 different European leagues, the Premier league (England), La Liga (Spain), Bundesliga (Germany), Serie A (Italy), Ligue 1 (France), Liga NOS (Portugal) and Eredvisie (Netherlands). It is important to state that the teams chosen in each of the national leagues were teams in which the average market value of players was equal or above the €10

Million threshold. Meaning that, clubs present in the database are top-tier European teams. To make this decision the assumption made was that top-tier clubs have more conditions and, therefore, are more likely to have and bet on academy players to their senior team. It was important to have rosters with academy players available so that the analysis could be significant and the consequent findings meaningful.

For the development of the work a set of variables were created in order to run significant regressions and analyze the results. These predictors can be grouped into 2 categories: Independent and dependent variables. For the independent variables there is only the **Difference** variable: The percentual difference in a player's market value in the last year. Regarding the dependent variables a couple of variables were built. **CodeClub**: A code from 1 to 31 to list all the different clubs. **CodePosition**: A code from q to 4 to list the 4 football positions. 1) Goalkeeper, 2) Defender, 3) Midfielder and 4) Striker. **Formation1y, Formation2y and Formation3y** refer to the window of time a player has been promoted from the academy. Either in the previous year, 2 or 3 years ago respectively. **CodeCC**: (Code Club Country) A code from 1 to 7 to list the countries from the teams evaluated. 1) England, 2) Spain, 3) Germany, 4) Italy, 5) France, 6) Portugal and 7) Netherlands. **AgeUseful**: The age of each player. Lastly, **GroupA, GroupB, GroupC, GroupD** and **GroupE** are variables which group clubs into classes with similar power (market value).

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Difference	772	.3032746	1.738421	-.5	36.5
CodeClub	802	16.29426	8.901777	1	31
Codeposition	802	2.684539	.9888191	1	4
Formation1y	802	.0773067	.2672442	0	1
Formation2y	802	.1122195	.3158332	0	1
Formation3y	802	.1359102	.3429071	0	1
CodeCC	802	3.220698	1.876305	1	7
Ageuseful	802	25.40399	4.522534	16	41
GroupA	126	3.079365	1.434451	1	5
GroupB	172	9.023256	2.026011	6	12
GroupC	191	16.02094	1.941129	13	19
GroupD	214	23.6028	2.320796	20	27
GroupE	99	29.52525	1.13698	28	31

Regarding the independent variable, **Difference**, the paper contains 772 observations with a maximum variation of 3650% and a minimum (Absolut) variation of -50%. It is important to refer that the most common variation is of 0% (Median). Moving on to the dependent variables, the **CodeClub** variable comprises 802 observations and the club which appears more often is the club 17 (Leipzig), meaning that, the German team has the most extensive roster. It is important to refer that the value 3, midfielders, is the most frequent position in the **Codeposition** variable (median). Regarding the **Formation1y**, **Formation2y** and **Formation3y** variables, they all have 802 observations and a median value of 0, meaning that, the large majority of the paper's data are non-academy players. The median value of the **CodeCC** variable is 3, meaning that, Germany is the country with more players in the data. Concerning the **Ageuseful** variable, the paper has 802 players with a minimum age of 16 years and a maximum age of 41 years.

Regression Analysis

In order to analyze the main hypothesis of this paper a set of procedures was followed to ensure the significance of the findings. Interactions between variables, using the # command, were constructed to build indicators for each combination of the categories of the variables. Also, the **i.** command was used to specify indicators for each category of the variable. Lastly, fixed effects regressions were built to remove biased effects. The model in this paper comes out of the following regressions:

Regression 1:

$$Difference = \alpha + \beta_1 Xi + \varepsilon_i$$

Where Xi = Formation1y, Formation2y and Formation3y

Regression 4:

$$Difference = \alpha + \beta_1 Formation1y + \beta_2 CodeClub + \varepsilon_i$$

Regression 5:

$$Difference_A = \alpha + \beta_1 Formation1y_A + \varepsilon_i$$

Regression 6:

$$Difference_{CP} = \alpha + \beta_1 Formation1y_{CP} + \varepsilon_i$$

Regression 7:

$$Difference_C = \alpha + \beta_1 Formation1y_C + \varepsilon_i$$

Regression 8:

$$Difference_{CC} = \alpha + \beta_1 Formation1y_{CC} + \varepsilon_i$$

Regression 9:

$$Difference = \alpha + \beta_1 Formation1yZi + \beta_2 Zi \quad \varepsilon_i$$

Where Zi = CodeClub, CodeCC, Codeposition, GroupA, GroupB, GroupC, GroupD, GroupE

Discussion

```
. reg Difference Formation1y

      Source |       SS       df       MS              Number of obs =     375
-------------+------------------------------           F(  1,   373) =   44.07
       Model |  233.296279     1  233.296279           Prob > F      =  0.0000
    Residual |  1974.48845   373  5.29353472           R-squared     =  0.1057
-------------+------------------------------           Adj R-squared =  0.1033
       Total |  2207.78473   374  5.90316773           Root MSE      =  2.3008

------------------------------------------------------------------------------
  Difference |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Formation1y |   2.644907   .3984092     6.64   0.000     1.861497    3.428316
       _cons |   .3151746   .1251453     2.52   0.012     .0690958    .5612533
------------------------------------------------------------------------------
```

Before continuing it is important to refer that players promoted from the academy 4 or more years ago were excluded from the regressions to only compare academy players, in a window of 3 years, with bought players in the same span of time. As previously mentioned, for the purpose of this paper, academy players were divided into 3 windows so that we can see players that were promoted from the club academy 1, 2 or 3 years ago. Accordingly, first, to test the influence on the market value variation of coming from the club academy, linear regressions with the 3 windows were conducted. In these 3 regressions it was only compared players with equal windows of time in the club, meaning that, players promoted from the academy 1 year ago were only compared to bought players with 1 year in the team, and so on, in order to have results with higher level of significance. Looking at the 3 regressions, one can conclude that coming from the academy is statistically highly significant to a positive market value variation due to the p-values = 0,000 **(see Annex 2 and 3).** Furthermore, as expected, the coefficient of each independent variable (2.64 for 1 year, 1.67 for 2 years and 1.37 for 3 years) is positive but decreasing with the growth of the window. Meaning that, players promoted in a recent window of time tend to have higher variation in their market value than players promoted a few years ago. All results analyzed, it can be concluded that players coming from the academy have, in fact, a higher market value variation in one season comparing to non-academy players, regardless of the window of time. It is important to explain that, from now on, all regressions are made with promoted academy players and bought players in a window of 1 year due to the short period of time and the more convincing results it produces, nevertheless, regressions with a window of time of 2 and 3 years were also built to test the significance of the work.

Before moving on, it would be interesting to understand how the different clubs in the data valuate their players, meaning that, a ranking of clubs was constructed to analyze how players, both from academy and outside, fluctuate their market value across the various clubs. Looking at **Annex 4**, the ranking of the clubs, it is easy to see that big clubs like Real Madrid (-0,14) or

Liverpool (0,092) have small positive variation or even negative variation in player's market value. This is indicative that top-tier clubs, to pursue important trophies, prefer star players that are either experienced/mature players (have little positive or negative fluctuations in their market value) or young "stars" (who have reached their maximum valuation and tend to stabilize at that valuation). On the contrary, small teams which, usually, tend to be "exporting teams", meaning that, they raise talents and send them to big international clubs tend to have higher positive fluctuations in their player's valuation. Looking at Real Sociedad (1,79), Lille (1,41) or Valencia (0,60) these clubs do not have a club valuation so significant as top-tier clubs, but their players tend to have a higher value variation than star players from big European teams. Ultimately, this is indicative that either players here are younger athletes (with a huge margin of progression ahead) or are average players (who tend to have higher absolute variations across seasons than star players).

To better evaluate the results of the regressions, using academy players, the use of fixed effects was of the utmost importance. It was assumed that certain variables could impact or bias the outcome variable, hence, by removing those effects it could be assessed the net effect of significant predictors on this outcome variable. For example, being a male or female could influence the opinion regarding certain matters. Therefore, by removing these effects, results are free of weighty bias which creates greater significance in consequent findings.

```
Fixed-effects (within) regression               Number of obs      =       375
Group variable: Ageuseful                       Number of groups   =        22

R-sq:  within  = 0.0599                         Obs per group: min =         1
       between = 0.8421                                        avg =      17.0
       overall = 0.1057                                        max =        42

                                                F(1,352)           =     22.41
corr(u_i, Xb)  = 0.3174                         Prob > F           =    0.0000

------------------------------------------------------------------------------
  Difference |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Formation1y |   2.175907   .4596466     4.73   0.000     1.271908    3.079906
       _cons |   .3614492   .1285268     2.81   0.005     .1086721    .6142262
-------------+----------------------------------------------------------------
     sigma_u |  .41584055
     sigma_e |  2.3288157
         rho |  .03089955   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Using a fixed effects model, the first regression made removed the age effect, meaning that, with this regression the model is only comparing players with the same age. This is important because, naturally, the variation of a promoted academy player with 22 years old is not the same as the variation of a promoted player with 17 years. By looking to the regression one can see that coming from the academy in a 1-year window (Formation1y) is highly significant to a positive market value variation due to a p-value = 0,000. Also, the coefficient of the 1-year window promotion (2,18) lead us to conclude that his variable is positively related to the variation of a player's market value, implying that, when our dummy takes the value 1, the market value increases in average 218%. Removing the age effect, the coefficient of the independent variable decreases from 2,64 to 2,18 leading to significant unbiased results.

```
Fixed-effects (within) regression               Number of obs      =       375
Group variable: Codeposition                    Number of groups   =         4

R-sq:  within  = 0.1134                         Obs per group: min =        49
       between = 0.2085                                        avg =      93.8
       overall = 0.1057                                        max =       120

                                                F(1,370)           =     47.32
corr(u_i, Xb)  = -0.0606                        Prob > F           =    0.0000
```

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y	2.724998	.3961458	6.88	0.000	1.946019	3.503978
_cons	.3072722	.1238573	2.48	0.014	.0637196	.5508248
sigma_u	.44785203					
sigma_e	2.275925					
rho	.03727812	(fraction of variance due to u_i)				

After, in order to achieve impartial results, a regression fixing the position effect was constructed to compare players with the same position. It is critical to understand that players from different positions have different fluctuations in their market value (e.g. Goalkeepers tend to be cheaper than strikers), therefore, the results reached in this case, would be highly biased. Looking at the table above one can assert that leaving the academy 1 year ago or less continues highly significant, statistically speaking, due to the p-value of 0,000. Regarding the coefficient of the regressor, a value of 2,72 suggests that academy players have an increase of 272% in their market value compared to non-academy players with the same position. In this case, by fixing the position effect, the coefficient of the regressor actually increases from 2,64 to 2,72 which represents a more accurate result.

```
Fixed-effects (within) regression               Number of obs      =       375
Group variable: CodeClub                        Number of groups   =        31

R-sq:  within  = 0.0812                         Obs per group: min =         4
       between = 0.3081                                        avg =      12.1
       overall = 0.1057                                        max =        18

                                                F(1,343)           =     30.32
corr(u_i, Xb)  = 0.0827                         Prob > F           =    0.0000
```

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y	2.44628	.4442882	5.51	0.000	1.572407	3.320152
_cons	.3347724	.1256499	2.66	0.008	.0876311	.5819137
sigma_u	.6769634					
sigma_e	2.2803174					
rho	.0809949	(fraction of variance due to u_i)				

Intuitively, clubs are not all at the same level, neither are their academies. It follows that in different clubs the market value of academy players fluctuates at different degrees, therefore, comparing players from different clubs produces bias results. To eliminate this phenomenon, a regression fixing the club effect was built to compare players in the same club and reach unbiased results **(table above)**. Looking at the p-value = 0,000, one can state that coming from the academy with a 1-year window is, statistically, highly noteworthy. Considering the coefficient of the regressor, inside the same club, academy players increase their value 244% higher than non-academy players (or players that left the academy 2 or more years ago).

Following this line of thought, clubs from different countries follow the same pattern. It would be an illusion to state that teams in the Netherlands are of the same quality as teams in England. Hence, the variation of player's market value clearly differs across nations. To remove this bias, it was constructed a regression removing the effect of the countries of the different teams **(Annex 5)**. The p-value = 0,000 proposes a statistical high significance of the 1-year window of promotion to a positive market value variation. This regressor presents a coefficient of 2,76 showing that academy players, promoted in the window of 1 year, alter their value 276% higher compared to the bought players in the previous year.

After thoroughly evaluating the available data and consequent regressions, some important questions started forming. **(I)** What club takes the biggest advantage from academy players? **(II)** In what country do academy players have a greater market value variation? **(III)** In what position (GK, Defender, Midfielder or Striker) do academy players have an increased variation of their market value? **(IV)** Grouping the clubs into fair classes (Top, medium and low-tier), which of them takes higher advantage (intra-group) from academy players?

In order to answer the first question, it was prefixed the club variable to specify indicators for each team of this variable, i.e., an interaction between the **Formation1y** and the **CodeClub** variables was created. Meaning that, a ranking of the clubs, in terms of academy players valuation, was constructed to conclude which clubs took the biggest advantage from their own youth players **(Annex 6)**. Looking at this regression, one can state that only 14 from the total 31 clubs have promoted academy players in previous year (with market fluctuations available, this is, some players recently promoted do not have a difference in their value because they did not have a market value before). This just comes to show that our hypothesis, the clubs bias towards buying players, is a reality in the context of international football. Nevertheless, as we forecasted, academy players have higher variations in their value compared to other players. Looking at the regression, one sees that academy players, in their vast majority, have positive coefficients, most of them with high values. For example, Real Sociedad's academy players (1st place in the ranking) have an 910% additional increase in the variation of their market value when comparing to players bought 1 year ago. The same happens in Ajax (2,32) or in Bayern Munich (2,97) leading one to firmly state that academy players are, in fact, more valuable to clubs financially speaking. And, it is curious to notice that, this happens in all type of clubs. Top, medium or low-tier international clubs take bigger financial advantage from their academy players than from bought players as hypothesized in the beginning of the paper.

Regarding the difference across countries, a regression pre-fixing the country variable was constructed to specify values for each nation inside this variable. A ranking of countries, regarding the variation of academy players market value was made to achieve significant conclusions about academies throughout different nations and their addition to the valuation of players **(Annex 7)**. Analyzing the regression, both French and Italian clubs did not promote academy players with a market fluctuation available, meaning that, teams inside these countries do not bet in in-house players as often as other clubs. Concerning the rest of the nations, Spain is clearly ahead with a coefficient of (5,22) followed by Germany (2,91), Netherlands (2,18), England (1,22) and Portugal (0,67). It is curious to see that the Netherlands has a coefficient so high, being illustrative of the recent reality we have been witnessing. It is no coincidence that Ajax reached the Semi-Finals of the Champions league last season with a team of young talents, with some of them being "exported" to top international clubs this summer.

It is also interesting to study how the player's position impacts its market valuation. In order to analyze this impact a regression was constructed to see the indicators for each position inside this same variable (Goalkeeper, defender, midfielder and striker). First, looking at the regression involving every type of player **(Annex 8)** one can see that the ranking of positions in terms of variation of the market value is Defenders (0,55), Strikers (0,22), Midfielders (0,17) and, lastly, Goalkeepers (-0,13) which can be explained by the average age of Goalkeepers who tend to play more time than a typical player. Therefore, in this position players are often older and experience negative fluctuations with more frequency. **Annex 9** shows us the regression using only academy players and it follows the same trend, meaning that, the ranking is exactly the same, but coefficients are different. Academy defenders are in front with a 619% addition to the variation of the market value compared to non-academy players, followed by academy strikers (1,90), academy midfielders (1,76) and, finally, goalkeepers (0,63) contrarily to the negative variation of all players. Showing that academy goalkeepers, due to their young age, do not experience those negative fluctuations, previously referred.

Lastly, in order to group clubs into fair groups, teams were divided into 5 different groups. This groups were built having the market value of clubs into consideration. Group A consists in clubs which have a market value above €1 Billion (Manchester City, Liverpool, Real Madrid, Barcelona and PSG). Group B market values range from around €700 Million to €1 Billion. Group C market values varies from around €500 Million to €650 Million. Group D fluctuates between market values of €300M to €450M and, lastly, group E are the teams below the €300M threshold. The reason behind constructing these groups is to have a better insight on how clubs of a same level take advantage of their academies. For example, it would be unfair to compare

Barcelona's academy with Lille's academy and the results produced by each of them, meaning that, Barcelona produces star players more often and, when promoted, the "leap" in terms of valuation is significantly higher compared to young Lille players. Thus, this grouping allows to compare the results produced by academies and clubs of the same level and reach valuable findings regarding the financial advantage taken by similar clubs.

Concerning group A, Manchester City, Liverpool and PSG did not promote academy players this last year (with market fluctuation available). Looking at **Annex 10**, one can see that Real Madrid takes the higher advantage from its academy with academy players adding a 232% variation to the non-academy players market valuation. Barcelona takes 2nd place with a coefficient of 0,41 meaning that academy players are also valuable to this club. It is also important to notice that both p-values < 5% producing significant results.

Regarding group B **(Annex 11)**, Tottenham, Juventus and Atletico Madrid do not have academy players promoted in this previous year in the database. Bayern Munich leads this group with a 2,97 coefficient, followed by Chelsea with a coefficient of 2,16, both having p-values below 0,05 implying statistically significant results. After, Manchester United occupies the 3rd position with the factor "academy" adding 17,2% to the non-academy players market value. In the last place is Arsenal, with the academy players accruing 5,1% to the valuation of non-academy players.

In group C **(Annex 12)** teams seem to have a significant bias against academy players. From 7 teams only one promotes academy players. Valencia presents a -0,056 coefficient, meaning that, the academy factor here decreases by 5,6% the variation of non-academy players market valuation. Everton, Dortmund, Leipzig, Napoli, Inter and AC Milan did not promote any academy player with a percentual market variation in the previous year.

Concerning group D **(Annex 13)**, Leicester, AS Roma, Lyon and AS Monaco are the clubs which did not bet in the academy this last year (did not have players with a percentual variation in the database). This group is led by Real Sociedad with a considerable distance to the rest. The factor academy accrues a 910% difference in the market value of non-academy players, meaning that, this club takes a massive advantage of academy players financially speaking (also has p-value = 0,000). In 2nd place Leverkusen presents a coefficient of 2,93 and, after, appears Ajax with the academy factor adding to non-academy players valuation a fluctuation of 233%. Benfica is the last team in this group, with a 0,74 coefficient.

Concluding the analysis, in group E **(Annex 14)** only Lille did not promote academy players to the senior team in this previous year. Seville takes the 1st place with academy players having a 293% increase in its valuation compared to bought players. After, PSV occupies the 2nd position of the group with the factor "academy" adding a 151% variation to the market value of the remaining players of the club. Finally, in the last position of the group is Porto with a coefficient of 0,55.

Conclusion

This paper attempted to study the context of the international football world with a special view on academy players and its reality. In a world where teams are biased towards buying players, this paper tried to discover the reason behind this status-quo. A set of behavioral and non-behavioral phenomena were studied and, as expected, all of them worked together towards a reality where academy players are undervalued. Thus, taking this into consideration, the main hypothesis of the paper consisted in the notion that it is more advantageous for teams to bet on academy players rather than spending money in the labor market searching for athletes with the same expected quality. It was concluded that coming from the academy is highly significant to a positive market value variation. This result is mostly significant to teams wanting short-term results due to the 1-year window variation. Using fixed-effects regressions similar results were obtained, however, with a higher degree of significance due to the removal of weighty bias. After validating the main hypothesis of the paper, a set of valuable experiments were conducted to reach important findings regarding the world of academy players. First, it was studied what club took the biggest advantage from its academy, using a rank of clubs in terms of academy usefulness with the 1st place going to Real Sociedad. Second, it was discovered that academy players in Spain have the highest variation in their market value, resorting to a ranking by country of the additional value of belonging to the academy. Lastly, it was revealed that the defender position increased most the market value of an academy player through the construction of a position ranking. Concluding, in order to reach superior results in terms of club's academy usefulness, teams were categorized into homogeneous groups of quality and size and each group was studied through a club classification.

Future Research and Limitations

Regarding future work it would be interesting to have this paper as a starting point and study the academy players reality deeper. First, it would be stimulating to understand how much do clubs spend when forming young players. This would enable future researchers to comprehend if, in reality, the profit of betting in the club academy is higher than buying players with the same expected quality.

In this paper it was not analyzed the costs involved with club academies mainly because academies are not exclusively payed with player's transfers and clubs look at them more like a source of income instead of a cost factor. After thorough research it was discovered that sponsors cover some academy costs and, in 75% of the cases, clubs receive financial aid from the National Associations softening this burden (European Club Association). Also, on the other side it was not count the commissions of agents of bought players, which tend to be high amounts.

It would also be valuable to understand with what frequency are academy players promoted and what is the proportion of academy players that reach the senior team. For future work researchers should extend the window of time in the observations to evaluate the results in a medium to long-run perspective. This work focused on short-term findings, using a window of 1 year, because most of the teams want results now. Lastly, it would also be interesting to scale up this project to other continents and understand if these leagues follow a similar trend.

References

1. Charmichael, F.; Thomas, D. 1993. *"Bargaining in the transfer market: theory and evidence".* University of Salford, University of Wales
2. Chelsea transfer ban: Fifa punishment 'after 150 rule breaches involving 69 academy players'. 2019. BBC Sport
3. European Club Association. *"Report on Youth Academies in Europe".*
4. Fédération Internationale de Football Association. 2018. *"Regulations on the Status and Transfer of Players"*
5. Frick, Bernd. 2007. *"The Football Player's Labor Market: Empirical Evidence From the major European Leagues".* Scottish Journal of Political Economy, Vol. 54, No. 3
6. Gulbrandsen, Anders Munkhaugen; Gulbrandsen, Christian Munkhaugen. 2011. *"Valuation of Football Players: A Complete Pricing Framework".* NORGES HANDELSHØYSKOLE
7. He, Miao; Cachucho, Ricardo; Knobbe, Arno. *"Football Player's Performance and Market Value".* Leiden University, the Netherlands
8. He, Yuan. *"Predicting Market Value of Soccer Players Using Linear Modeling Techniques".*
9. Kagel, John H.; Levin, Dan. 1986. *"The Winner's curse and Public Information in Common Value Auctions".* The American Economic Review, Vol. 76, No. 5, pp. 894-920
10. Kahneman, Daniel; Tversky, Amos. 1973. *"On the Psychology of Prediction".* Psychological Review, Vol. 80, No. 4. Hebrew University of Jerusalem, Israel, and Oregon Research Institute
11. Massey, Cade; Thaler, Richard H. 2005. *"Overconfidence VS Market Efficiency in the National Football League".* National Bureau of Economic Research
12. Massey, Cade; Thaler, Richard H. 2012. *"The loser's curse: Decision-Making & Market Efficiency in the National Football League Draft".*
13. Moore, Don A.; Healy, Paul J. 2008. *"The Trouble with Overconfidence".* Psychological Review. Carnegie Mellon University, Ohio State University
14. Oprean, Victor-Bogdan; Oprisor, Tudor. 2014. *"Accounting for soccer players: capitalization paradigm vs expenditure"*. Science Direct. Faculty of Economics and Business Administration, Babes-Bolyai Universit
15. Oskamp, Stuart. 1965. *"Overconfidence in Case-Study Judgements".* Journal of Consulting Psychology, Vol. 29, No. 3, 261-265

16. Rottenberg, Simon. 1956. *"The Baseball Players' Labor Market".* Journal of Political Economy, Vol. 64, No. 3, pp. 242-258
17. Sanderson, Allen R.; Siegfried, John J. 2006. *"Simon Rottenberg and Baseball, then and now: A 50th Anniversary Retrospective".* Department of Economics, Vanderbilt University, Nashville.
18. Simmons, Robert. 1997. *"Implications of the Bosman Ruling for football transfer markets".* Institute of Economic Affairs
19. Sloane, Peter J. 2006. *"Rottenberg and the Economics of Sport after 50 years: An Evaluation".* University of Wales Swansea and IZA Bonn
20. Thaler, Richard H. 1988. *"Anomalies: The Winner's curse".* Journal of Economic Perspectives, Vol. 2, No. 1, pp. 191-202
21. Torres-Reyna, Oscar. 2007. *"Panel Data Analysis Fixed and Random Effects using Stata".* Princeton University

Annex

Background Information

The field of the economics of sport began to be studied with interest in 1956 with Simon Rottenberg's paper on the baseball labor market which, ultimately, paved the path to the subsequent studies in every sports market. The key turning point of this study was Rottenberg's belief that the economics of professional sports markets could actually be studied using the same framework as for the other existing industries. Nevertheless, he acknowledged the uniqueness of a sports market by the presence of two unusual aspects – the monopsony power (monopoly) here takes an intense form and that competitors must have similar size if any of them are to be successful. Recently, the availability of detailed valuable information on transfer fees, contract lengths and players wages along with important changes in the regulatory system of the football labor market have persuaded a growing number of economists to shift their attention to this particular industry (Bernd Frick 2007).

Although it is not necessary to be a football specialist to follow the analysis of this thesis, for example, to know the difference between a central and advanced midfielder functions, it is important to have some basic insights on this industry.

First players are traded for cash settlements based on their market value contrarily to the common draft system present in most of the American sports markets. This market value which goes hand in hand with the player's salary are both in great part determined by the age and experience of the player, the number of international caps, number of goals scored and position (Bernd Frick 2007). It is also important to note that when contracts are signed between teams and players, no contract can exceed five years and by the terminating date players either become free agents or renew their contract with the team.

Second, the Union of European Football Associations (UEFA) agreed in principle in 2009 to establish a Financial Fair Play Regulation (FFP) to all European clubs preventing them to spend more than what they earn in the pursuit of success and, consequently, getting into financial distress that may threaten their long-term survival. UEFA engaged in this plan when in 2009 more than half of the 655 European clubs incurred a loss in the previous year and, even if a small percentage managed to sustain heavy losses, at least, 20% of those clubs were in financial hazard. This regulation was finally implemented at the outset of the 2011-12 season providing UEFA with sanctions to implement against European teams that exceed spending over several seasons within a budgetary framework. These penalties go from fines or withholding of prize money to player transfer bans, being the severe of them all disqualification from European competitions. However, it is important to state that, contrarily to many sports in the US, UEFA does not impose any transfer or salary cap to the teams. Meaning that, the pursuit for competitive balance across national leagues and European competitions becomes a dream hard to become true with big teams stepping up and acquiring a big percentage of the world top talents.

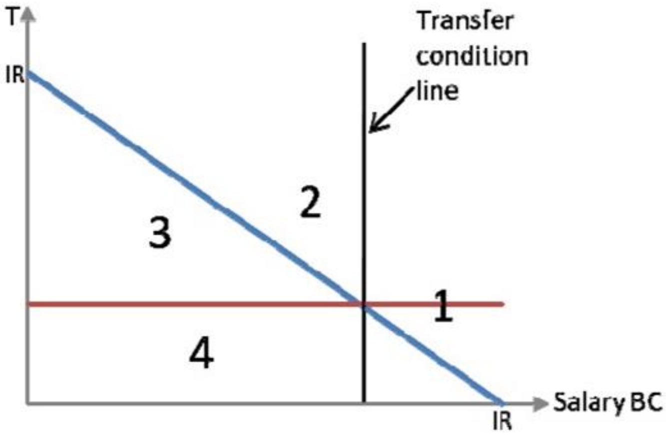
Third, according to FIFA regulations “Minors are deemed to be players under the age of 18, or between 16 and 18 for transfers within the European Union or European Economic Area” with the new club meeting the required minimum obligations to acquire players in this span of age with a non-professional contract. It is of the utmost importance to refer that only at the age of 17 is a player eligible to sign a professional contract with a football team also subject to the appropriate regulation. Meaning that, an academy player will only sign a professional contract with the club when he turns 17 also becoming possible for him to transfer to other European teams with an appropriate professional contract.

To illustrate this idea in a pragmatic way let us consider the example of Ángel Di María’s move to Manchester United in the summer of 2014. On 26 of August 2014 Di María signed a 5-year contract with Manchester United for a transfer fee of about £60 million becoming one of the most expensive players at that time and the highest transfer fee paid by a British club at the time. When he arrived at Manchester, Di María inherited the famous Manchester United No.7 previously worn by superstars like George Best, Cantona, Beckham and Cristiano Ronaldo. Needless was to say that expectations were high for the upcoming season of the Argentinian football star. However, expectations were not met, as the winger did not manage to settle in England. With only 3 goals and 10 assists in a short 27 game season, Di María was crowned by the British press the worst signing of the season and one of the greatest flops in the history of the premier league. The Argentinian player ended up leaving in the summer of 2015 to Paris Saint Germain for a transfer fee of around £44 million, meaning that, after just one season, Manchester United actually managed to lose money with the winger (accounting only for the transfer fees).

Curiously, in the 2015/16 season, when Di María left for PSG, Manchester United incorporated from its academy into the senior team the 18 years old winger Marcus Rashford. The British became an over-night sensation by scoring two goals in both his debuts on the first team. First against Midtjylland in the UEFA Europa League, making him the youngest player ever to score for Man Utd in European competitions, and against Arsenal for the premier league only three days later. After just one remarkable season for this young player, Manchester United proposed a new contract to the winger keeping him at the club until 2020. After 5 great seasons with the British club, having 57 goals and 29 assists in 199 matches at the age of 22 Rashford continues to be one of the best players in the team and in the premier league justifying a market value of €80 million.

Furthermore, consider the unique scenario of Chelsea's transfer ban. FIFA punished the British club with a regulation forbidding them to engage in any transfer until February 2020 due to the breach of 150 FIFA rules involving 69 academy players. Chelsea had been claiming that several academy players were only trialists and did not attend organized matches which was later discovered to be false. This case was detected after coming to light that the forward Bertrand Traore, who moved from Burkina Faso to London, played for the team for several years before his registration in 2014. As previously mentioned, FIFA only allows international transfers under-18 to happen within a certain criterion, for example, the parents move to the country in question for non-football reasons, to prevent children exploitation or trafficking. As the international organization did not tolerate these breaches a heavy transfer ban was imposed to the blues. Surprisingly, this punishment led to a strong bet on Chelsea academy players. Consider the example of the 22-year old forward Tammy Abraham, this season's worldwide sensation. For the last two seasons Tammy was loaned to Swansea and Aston Villa respectively. This season, the blues decided to take a chance on the academy forward and, for the surprise of many, he has been the principal star of the team with 9 goals in the premier league and 1 in the Champions league. The same logical can be applied to the midfielder Mason Mount or the defender Fikayo Tomori. All 3 players rose to the senior team and have been under the world's eyes with the transfer ban still being present in the blue's reality. We do not know if without the ban these players would have been part of Chelsea's plan or not, but one thing is certain, this punishment forced the British team to bet on their academy and the results were highly satisfying.

Graph 1



With the help of **Graph 1**, in the Annexes, let us consider the example of a very wealthy owner (e.g. a sheik) buying a small club with the desire of creating a super team capable of fighting for both national and international trophies. In order to do so, the sheik will see the club transfers falling into area 1, meaning that, the owner's club is buying star players from big clubs. In this scenario, top-tier clubs, not willing to lose their position feel the need to improve their roster quality and compete for the best players in the market. Accordingly, transfer fees and wages shift upwards making transfers in the market to occur in area 2. However, low-tier clubs are also affected by this change pushing some of them into financial hazard. Consequently, to avoid financial distress the small teams start to sell players, which corresponds to the transfers occurring in area 4. Nevertheless, the industry still sees owners with the desire of running their clubs in a sustainable and self-financing way, transferring players around area 3. However, the strong competition forces this kind of clubs to act according the established trend, meaning that, sustainable clubs do not have the power to change this financially "insane" status-quo. To some extent, this example can be seen as representative of what the football world has been witnessing in the last couple of years. Just think about Manchester City, Chelsea or PSG which have become top-tier clubs recently due to ownership changes. The result of this ecosystem has been an inflated market devastating a large number of clubs which are incurring in losses year after year.

Annex 2

```
. reg Difference Formation2y

      Source |       SS       df       MS              Number of obs =     502
-------------+------------------------------           F(  1,   500) =   36.59
       Model |  153.344503     1  153.344503           Prob > F      =  0.0000
    Residual |  2095.35919   500  4.19071838           R-squared     =  0.0682
-------------+------------------------------           Adj R-squared =  0.0663
       Total |  2248.70369   501  4.48843053           Root MSE      =  2.0471

------------------------------------------------------------------------------
  Difference |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Formation2y |   1.668336   .2757996     6.05   0.000     1.126467    2.210205
       _cons |   .2374897   .0977039     2.43   0.015     .0455289    .4294506
------------------------------------------------------------------------------
```

Annex 3

```
. reg Difference Formation3y

      Source |       SS       df       MS              Number of obs =     572
-------------+------------------------------           F(  1,   570) =   34.26
       Model |  128.318478     1  128.318478           Prob > F      =  0.0000
    Residual |  2134.61625   570  3.74494079           R-squared     =  0.0567
-------------+------------------------------           Adj R-squared =  0.0550
       Total |  2262.93473   571   3.9631081           Root MSE      =  1.9352

------------------------------------------------------------------------------
  Difference |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Formation3y |   1.372795   .2345219     5.85   0.000     .9121621    1.833427
       _cons |   .2114584   .0871563     2.43   0.016     .0402716    .3826452
------------------------------------------------------------------------------
```

Annex 4

```
. reg Difference Formation1y i.CodeClub
```

Source	SS	df	MS
Model	439.16314	31	14.1665529
Residual	1884.81113	706	2.66970415
Total	2323.97427	737	3.15328937

Number of obs = 738
F(31, 706) = 5.31
Prob > F = 0.0000
R-squared = 0.1890
Adj R-squared = 0.1534
Root MSE = 1.6339

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y	2.691604	.2864194	9.40	0.000	2.129268	3.25394
CodeClub						
2	.0917391	.4818176	0.19	0.849	-.8542277	1.037706
3	.1278261	.4818176	0.27	0.791	-.8181407	1.073793
4	-.1046268	.4843856	-0.22	0.829	-1.055636	.846382
5	-.0917219	.4819785	-0.19	0.849	-1.038005	.8545608
6	-.1228098	.4879572	-0.25	0.801	-1.080831	.8352111
7	.0005702	.4677125	0.00	0.999	-.9177037	.9188442
8	.0985287	.4720829	0.21	0.835	-.8283257	1.025383
9	-.2321121	.5028327	-0.46	0.645	-1.219339	.7551144
10	-.1390955	.4722219	-0.29	0.768	-1.066223	.7880319
11	.2410824	.506542	0.48	0.634	-.7534267	1.235591
12	.6489903	.4874359	1.33	0.183	-.3080071	1.605988
13	1.991853	.4934837	4.04	0.000	1.022982	2.960724
14	.0038917	.4773693	0.01	0.993	-.9333417	.9411251
15	.0551911	.4819785	0.11	0.909	-.8910916	1.001474
16	.0362609	.4818176	0.08	0.940	-.909706	.9822277
17	.128801	.4677125	0.28	0.783	-.789473	1.047075
18	-.059461	.4819785	-0.12	0.902	-1.005744	.8868217
19	-.0537617	.4636292	-0.12	0.908	-.9640187	.8564953
20	.025192	.4767722	0.05	0.958	-.9108691	.9612532
21	-.005558	.4767722	-0.01	0.991	-.9416191	.9305032
22	-.0197391	.4818176	-0.04	0.967	-.965706	.9262277
23	.1636472	.4677125	0.35	0.727	-.7546268	1.081921
24	-.0851739	.4818176	-0.18	0.860	-1.031141	.8607929
25	.3757754	.4767722	0.79	0.431	-.5602858	1.311837
26	-.1343199	.459805	-0.29	0.770	-1.037069	.7684291
27	1.473654	.487262	3.02	0.003	.5169981	2.43031
28	-.2217994	.4666533	-0.48	0.635	-1.137994	.694395
29	-.0672168	.4781146	-0.14	0.888	-1.005913	.8714799
30	.0724078	.4638832	0.16	0.876	-.8383479	.9831636
31	-.0759221	.4824609	-0.16	0.875	-1.023152	.8713077
_cons	.0503913	.3406965	0.15	0.882	-.6185083	.7192909

Annex 5

```
. xtreg Difference Formation1y, fe
```

Fixed-effects (within) regression   Number of obs = 375
Group variable: CodeCC   Number of groups = 7

R-sq: within = 0.1067   Obs per group: min = 31
between = 0.1139   avg = 53.6
overall = 0.1057   max = 81

F(1,367) = 43.81
corr(u_i, Xb) = -0.1022   Prob > F = 0.0000

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y	2.757028	.4165164	6.62	0.000	1.93797	3.576086
_cons	.3041119	.1254872	2.42	0.016	.0573477	.550876
sigma_u	.3489797					
sigma_e	2.2960398					
rho	.02257994	(fraction of variance due to u_i)				

F test that all u_i=0: F(6, 367) = 1.26 Prob > F = 0.2769

Annex 6

Source	SS	df	MS
Model	623.382742	44	14.1677896
Residual	1584.40199	330	4.80121815
Total	2207.78473	374	5.90316773

Number of obs =	375
F(44, 330) =	2.95
Prob > F =	0.0000
R-squared =	0.2824
Adj R-squared =	0.1867
Root MSE =	2.1912

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#CodeClub						
1 1	0	(empty)				
1 2	0	(empty)				
1 3	0	(empty)				
1 4	2.16375	1.54939	1.40	0.163	-.8841768	5.211677
1 5	.172	2.366732	0.07	0.942	-4.483785	4.827785
1 6	.0512778	1.712915	0.03	0.976	-3.318332	3.420888
1 7	0	(empty)				
1 8	0	(empty)				
1 9	.4085	1.732271	0.24	0.814	-2.999186	3.816186
1 10	2.323636	2.288601	1.02	0.311	-2.17845	6.825723
1 11	0	(empty)				
1 12	-.0559286	2.268075	-0.02	0.980	-4.517637	4.405779
1 13	9.098667	1.183366	7.69	0.000	6.770774	11.42656
1 14	2.925125	1.643376	1.78	0.076	-.3076895	6.15794
1 15	2.970667	2.309694	1.29	0.199	-1.572914	7.514248
1 16	0	(empty)				
1 17	0	(empty)				
1 18	2.934143	2.342457	1.25	0.211	-1.673889	7.542175
1 19	0	(empty)				
1 20	0	(empty)				
1 21	0	(empty)				
1 22	0	(empty)				
1 23	0	(empty)				
1 24	0	(empty)				
1 25	0	(empty)				
1 26	0	(empty)				
1 27	0	(empty)				
1 28	.739	1.200152	0.62	0.538	-1.621914	3.099914
1 29	.5497436	1.403471	0.39	0.696	-2.211134	3.310622
1 30	2.328682	1.11206	2.09	0.037	.1410611	4.516302
1 31	1.513433	1.649452	0.92	0.360	-1.731333	4.758199
CodeClub						
2	.0196857	1.283017	0.02	0.988	-2.504237	2.543609
3	.1959	1.46988	0.13	0.894	-2.695617	3.087417
4	-.04385	1.46988	-0.03	0.976	-2.935367	2.847667
5	.1904	1.326818	0.14	0.886	-2.419689	2.800489
6	.2156222	1.222175	0.18	0.860	-2.188615	2.619859
7	.1154	1.181828	0.10	0.922	-2.209467	2.440267
8	.1336222	1.222175	0.11	0.913	-2.270615	2.537859
9	-.0416	1.249158	-0.03	0.973	-2.498917	2.415717
10	-.0572364	1.181828	-0.05	0.961	-2.382104	2.267631
11	.4240667	1.166338	0.36	0.716	-1.870329	2.718463
12	.9893286	1.141573	0.87	0.387	-1.256349	3.235006
13	.2959	1.249158	0.24	0.813	-2.161417	2.753217
14	.008275	1.12264	0.01	0.994	-2.200158	2.216708
15	.1297333	1.222175	0.11	0.916	-2.274504	2.53397
16	.1275667	1.166338	0.11	0.913	-2.166829	2.421963
17	.1504	1.181828	0.13	0.899	-2.174467	2.475267
18	-.0007429	1.283017	-0.00	1.000	-2.524666	2.52318
19	.0433091	1.181828	0.04	0.971	-2.281558	2.368176
20	.1089	1.166338	0.09	0.926	-2.185496	2.403296
21	.0373167	1.166338	0.03	0.974	-2.257079	2.331713
22	.0804769	1.153069	0.07	0.944	-2.187816	2.34877
23	.3275667	1.10769	0.30	0.768	-1.851458	2.506592
24	-.0686714	1.141573	-0.06	0.952	-2.314349	2.177006
25	.72115	1.166338	0.62	0.537	-1.573246	3.015546
26	-.0601556	1.10769	-0.05	0.957	-2.23918	2.118869
27	1.824713	1.12264	1.63	0.105	-.3837203	4.033145
28	.2586	1.200152	0.22	0.830	-2.102314	2.619514
29	.2723231	1.153069	0.24	0.813	-1.99597	2.540616
30	.1742182	1.181828	0.15	0.883	-2.150649	2.499086
31	-.0085333	1.131514	-0.01	0.994	-2.234424	2.217358
_cons	.0666	.9799202	0.07	0.946	-1.861078	1.994278

Annex 7

Source	SS	df	MS
Model	377.496101	11	34.3178273
Residual	1830.28863	363	5.04211744
Total	2207.78473	374	5.90316773

Number of obs	=	375
F(11, 363)	=	6.81
Prob > F	=	0.0000
R-squared	=	0.1710
Adj R-squared	=	0.1459
Root MSE	=	2.2455

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#CodeCC						
1 1	1.224836	.9010979	1.36	0.175	-.5471912	2.996864
1 2	5.217221	.702318	7.43	0.000	3.836098	6.598344
1 3	2.905423	1.627988	1.78	0.075	-.2960482	6.106894
1 4	0	(empty)				
1 5	0	(empty)				
1 6	.6654185	.9216763	0.72	0.471	-1.147077	2.477914
1 7	2.184615	.9078504	2.41	0.017	.3993089	3.969922
CodeCC						
2	.1821987	.4058929	0.45	0.654	-.615998	.9803954
3	-.0030867	.4700637	-0.01	0.995	-.9274768	.9213033
4	.0244273	.4099641	0.06	0.953	-.7817756	.8306302
5	.4821864	.4191779	1.15	0.251	-.3421357	1.306508
6	.1517929	.5575817	0.27	0.786	-.9447031	1.248289
7	-.045779	.5344179	-0.09	0.932	-1.096723	1.005165
_cons	.1811636	.3027786	0.60	0.550	-.4142567	.7765839

Annex 8

Source	SS	df	MS
Model	323.656922	4	80.9142305
Residual	2000.31735	733	2.7289459
Total	2323.97427	737	3.15328937

Number of obs	=	738
F(4, 733)	=	29.65
Prob > F	=	0.0000
R-squared	=	0.1393
Adj R-squared	=	0.1346
Root MSE	=	1.652

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y	2.882977	.2726785	10.57	0.000	2.347653	3.418301
Codeposition						
2	.5581866	.2070918	2.70	0.007	.1516228	.9647504
3	.1782844	.2115048	0.84	0.400	-.2369431	.5935119
4	.2180868	.2142862	1.02	0.309	-.202601	.6387747
_cons	-.1330446	.1795119	-0.74	0.459	-.4854635	.2193742

Annex 9

Source	SS	df	MS
Model	430.627867	7	61.5182668
Residual	1777.15686	367	4.84238927
Total	2207.78473	374	5.90316773

Number of obs	= 375
F(7, 367)	= 12.70
Prob > F	= 0.0000
R-squared	= 0.1950
Adj R-squared	= 0.1797
Root MSE	= 2.2005

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#Codeposition						
1 1	.6344225	.9589987	0.66	0.509	-1.2514	2.520245
1 2	6.189129	.7626719	8.12	0.000	4.689374	7.688884
1 3	1.757609	.6122864	2.87	0.004	.5535786	2.961639
1 4	1.904088	.863124	2.21	0.028	.2067987	3.601377
Codeposition						
2	.5195711	.3952706	1.31	0.190	-.2577083	1.296851
3	.1175139	.4058108	0.29	0.772	-.6804923	.9155201
4	.0908822	.4072183	0.22	0.824	-.7098919	.8916563
_cons	.0877442	.3355797	0.26	0.794	-.5721561	.7476444

Annex 10

Source	SS	df	MS
Model	5.49369988	6	.915616646
Residual	1.3876486	41	.033845088
Total	6.88134848	47	.14641167

Number of obs	= 48
F(6, 41)	= 27.05
Prob > F	= 0.0000
R-squared	= 0.7983
Adj R-squared	= 0.7688
Root MSE	= .18397

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#GroupA						
1 1	0	(empty)				
1 2	0	(empty)				
1 3	.4085	.1454413	2.81	0.008	.1147753	.7022247
1 4	2.323636	.1921508	12.09	0.000	1.93558	2.711693
1 5	0	(empty)				
GroupA						
2	.0196857	.107722	0.18	0.856	-.1978633	.2372347
3	-.0416	.1048792	-0.40	0.694	-.2534079	.1702079
4	-.0572364	.0992262	-0.58	0.567	-.2576278	.1431551
5	-.0686714	.0958464	-0.72	0.478	-.2622371	.1248942
_cons	.0666	.082274	0.81	0.423	-.0995558	.2327558

Annex 11

Source	SS	df	MS
Model	22.665323	10	2.2665323
Residual	40.2761324	52	.774541007
Total	62.9414554	62	1.01518476

Number of obs	63
F(10, 52)	2.93
Prob > F	0.0055
R-squared	0.3601
Adj R-squared	0.2370
Root MSE	.88008

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#GroupB						
1 6	0	(empty)				
1 7	2.16375	.6223106	3.48	0.001	.9149924	3.412508
1 8	.172	.9505952	0.18	0.857	-1.735509	2.079509
1 9	.0512778	.6879903	0.07	0.941	-1.329276	1.431831
1 10	0	(empty)				
1 11	2.970667	.9276859	3.20	0.002	1.109129	4.832205
1 12	0	(empty)				
GroupB						
7	-.23975	.6223106	-0.39	0.702	-1.488508	1.009008
8	-.0055	.5680893	-0.01	0.992	-1.145455	1.134455
9	.0197222	.5288623	0.04	0.970	-1.041518	1.080962
10	.2281667	.5081145	0.45	0.655	-.7914397	1.247773
11	-.0661667	.5288623	-0.13	0.901	-1.127407	.9950732
12	-.1525909	.5138561	-0.30	0.768	-1.183719	.8785368
_cons	.2625	.4400401	0.60	0.553	-.620505	1.145505

.

Annex 12

Source	SS	df	MS
Model	9.75975216	7	1.39425031
Residual	196.449135	78	2.51857866
Total	206.208888	85	2.42598691

Number of obs	86
F(7, 78)	0.55
Prob > F	0.7910
R-squared	0.0473
Adj R-squared	-0.0382
Root MSE	1.587

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#GroupC						
1 13	0	(empty)				
1 14	-.0559286	1.642704	-0.03	0.973	-3.326301	3.214444
1 15	0	(empty)				
1 16	0	(empty)				
1 17	0	(empty)				
1 18	0	(empty)				
1 19	0	(empty)				
GroupC						
14	.8739286	.6394217	1.37	0.176	-.3990623	2.146919
15	.0121667	.6624525	0.02	0.985	-1.306675	1.331008
16	.035	.6767004	0.05	0.959	-1.312207	1.382207
17	-.0065	.6624525	-0.01	0.992	-1.325342	1.312342
18	-.0780833	.6624525	-0.12	0.906	-1.396925	1.240758
19	-.0349231	.6501527	-0.05	0.957	-1.329278	1.259431
_cons	.182	.4784994	0.38	0.705	-.7706192	1.134619

.

Annex 13

Source	SS	df	MS
Model	499.205511	11	45.3823192
Residual	1081.65393	99	10.9257973
Total	1580.85944	110	14.3714494

Number of obs	=	111
F(11, 99)	=	4.15
Prob > F	=	0.0000
R-squared	=	0.3158
Adj R-squared	=	0.2398
Root MSE	=	3.3054

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#GroupD						
1 20	0	(empty)				
1 21	9.098667	1.78513	5.10	0.000	5.55658	12.64075
1 22	2.934143	3.533642	0.83	0.408	-4.077369	9.945655
1 23	0	(empty)				
1 24	0	(empty)				
1 25	0	(empty)				
1 26	.739	1.810453	0.41	0.684	-2.853331	4.331331
1 27	2.328682	1.677564	1.39	0.168	-.9999688	5.657332
GroupD						
21	.1622778	1.606145	0.10	0.920	-3.024663	3.349218
22	-.1343651	1.665775	-0.08	0.936	-3.439624	3.170893
23	.1939444	1.349432	0.14	0.886	-2.483621	2.87151
24	.5875278	1.457553	0.40	0.688	-2.304573	3.479629
25	-.1937778	1.349432	-0.14	0.886	-2.871343	2.483788
26	.1249778	1.518735	0.08	0.935	-2.888523	3.138478
27	.040596	1.485676	0.03	0.978	-2.907307	2.988499
_cons	.2002222	1.101806	0.18	0.856	-1.986001	2.386445

.

Annex 14

Source	SS	df	MS
Model	49.0016004	6	8.16693339
Residual	264.635144	60	4.41058574
Total	313.636745	66	4.75207189

Number of obs	=	67
F(6, 60)	=	1.85
Prob > F	=	0.1042
R-squared	=	0.1562
Adj R-squared	=	0.0719
Root MSE	=	2.1001

Difference	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Formation1y#GroupE						
1 28	2.925125	1.575105	1.86	0.068	-.2255533	6.075803
1 29	0	(empty)				
1 30	.5497436	1.345166	0.41	0.684	-2.140989	3.240476
1 31	1.513433	1.580928	0.96	0.342	-1.648893	4.675759
GroupE						
29	1.816437	.7425114	2.45	0.017	.3311935	3.301681
30	.2640481	.7841795	0.34	0.738	-1.304544	1.832641
31	-.0168083	.7547852	-0.02	0.982	-1.526603	1.492987
_cons	.074875	.5250349	0.14	0.887	-.9753511	1.125101