The Impact of Player Turnover on Football Teams Performance Pedro Miguel Branco Pires<br>Department of Engineering and Management, Instituto Superior Técnico pedro.branco.pires@tecnico.ulisboa.pt


#### Abstract

In the past decades, people have witnessed a massive transformation of football clubs into authentic public companies, where good financial and asset management are crucial. Players are becoming the most crucial club asset. Moreover, across the past decade, it has been seen increasing club capability of performing player transfers. On the other hand, one problem arises related to team performance when a club has many changes from one season to the next. This dissertation addresses performance and club stability by applying Data Envelopment Analysis and Linear Regression Model. The main goal is to update the literature regarding the topic and eventually help football clubs making decisions. Both models have as the primary input variable the percentage of new players minutes and as the output variable the variation of points regarding the previous season. Regarding the results, important to highlight: the team with the lowest \%NPMin and the team with the highest variation of points regarding the previous season are permanently assigned maximum efficiency. Also, $58 \%$ of the cases (average from all leagues) with less than 20\%NPMin had efficiency equal to 1 . From the results of DEA models, no direct conclusions were possible to make. It was only possible to see an evident tendency line reflecting the relation between \%NPMin and VarPts. Finally, the final linear regression model allowed to prove that stability affects the team's performance. Clubs that have fewer new players fielded are more likely to have a better performance than before.


Keywords: Player turnover; Team Stability; Football; Performance; DEA; Regression Model

## 1. Introduction and Contextualisation

According to all studies, it is unanimous that football is the most popular sport in the world. The number of followers differs slightly from study to study. However, it is believed to be around 4 billion followers worldwide. Many factors make football the most popular and viewed sport globally. Based on SportsShow (2020) and TotalSportek (2017), some will be shown further. As mentioned before, football has the most significant fan base and audience across the world.

Similarly, it is always possible to watch football on TV, where the biggest broadcast deals and sponsorship deals regarding sports happen. The popularity and presence on the Internet and in the media are also quite evident. Furthermore, football is the most popular and practised sport in many countries since it has the most significant professional and amateur leagues. This number of leagues can be explained due to the accessibility to play the sport since it does not require huge investments. In contrast, the salary of the football players makes them within the highest salaries among athletes. Finally, the most followed athlete is a football player - Cristiano Ronaldo, with more than 300 million followers on Instagram.

The information regarding players contracts and their specifications, incentivising players to sign, is somehow difficult to find. However, it is important to highlight what a football player contract includes and explain all the processes of transferring players from one club to another. Additionally, the modern football world is facing a rise of complex decision-making mechanisms in the transfer market. On the one hand, clubs have to compete amongst themselves to recruit the most promising players strongly. On the other hand, they also need to consider the rise in professional football players' bargaining power.

Moreover, the stakeholders involved in these negotiations adopt fewer formal ways to negotiate and sign contracts, e.g., via electronic devices, such as emails, WhatsApp, and others. At the same time, this can speed up the terms of negotiation. In contrast, it can also create some disorder, mainly due to the number of bureaucratic documents sent and resented that can get lost. It is acknowledged that clubs discuss multiple player options and negotiate many possible signings simultaneously. However, players and agents do so as well. Thus, this previously described disorder will dramatically increase.

As mentioned, the action of moving a player who is under contract with a club to another club is called a transfer. This term is mainly used because the registration details are transferred from one association football club to another. The new club usually pays for the player's services. Thus, as compensation for the previous club losing the player, the new club pays them a previously determined amount of money - "transfer fee". This compensation can vary between straight monetary value or player exchange, with or without a monetary add-on. The player that is "offered" in exchange is in a difficult position since he does not have to approve the shift of clubs. Also, the club can reject to receive this player. Since 2002, UEFA established two fixed periods where it is possible to transfer players - called transfer window. The first one beginning before the start of the season, therefore from $1^{\text {st }}$ July until $31^{\text {st }}$ August, and the second taking place in the entire month of January. However, players without a contract with any team - called free agents - can be assigned at any moment of the season (Football-Stadiums.co.uk, 2020).

There can be different results in the procedure surrounding the signing of a player agreement. Such examples of results that have already occurred are:

1. Clubs that had a tentative agreement, but one of the parties dismisses from the final contract signing.
2. Clubs may consent to the move, but there is no agreement on the employment contract.
3.Finally, a dramatic situation in which two clubs complete the entire transfer process and the player dies unexpectedly.
According to Lukomski (2020), the events to procedure entirely with a player transfer are the following:
3. Agreement to move the player from one club to another
a. An arrangement and a written deal to transfer the sporting rights of the player should be concluded between the former club and the current club
b. This arrangement can be a financial amount or an exchange of players
c. The move can be permanent or can go from halfseason to two seasons via loan (there is no official maximum of seasons, however usually loans do not go over two seasons)
2.The approval of medical tests exercised by the new player's club
a. This procedure is usually performed before the signature of the employment contract to make sure everything is fine with the player.
4. Signature of a contract of employment between the player and the new club. This contract includes: Wages; Duration of the contract; Signing-on bonus; Bonus payments; Agent fee; Release/Buyout clause
4.Player's registration with the new football association
a. This procedure is only needed when the transfer is made between clubs from different countries.
b. The registration is a multistage procedure requiring the previous club to give the information and upload the transfer agreement and employment contract in the FIFA Transfer Matching System. After, the new club obtains the International Transfer Certificate (ITC) of the player. Finally, the player can be registered on the new football association.
c. A working permit is also needed in some countries, such as in the English Premier League, where it gets difficult to obtain it in some cases.
The most singular and unique topic about a football player's contract is the possibility of including a buyout clause. This clause allows the player to cancel his contract (without having a just cause) with the present club by paying, typically, a tremendous amount of money. When the player completes this process, he becomes a free agent and can sign with another club. The buyout clause is not mandatory to include in the contract. It is only a possibility of agreement when signing the contract. However, in Spain, this clause is indeed mandatory to include in the contract. Consequently, FIFA directives regarding this topic are not international law, and each country can decide independently (Madina Tatraeva, 2020). The commentary on FIFA regulations takes a more concise and direct explanation of how the buyout clause operates: "The parties may, however, stipulate in the
contract the amount that the player shall pay to the club as compensation in order to unilaterally terminate the contract (a so-called buyout clause). The advantage of this clause is that the parties mutually agree on the amount at the very beginning and fix this in the contract. By paying this amount to the club, the player is entitled to unilaterally terminate the employment contract. With this buyout clause, the parties agree to allow the player to cancel the contract at any moment and without a valid reason, i.e. also during the protected period, and as such, no sporting sanctions may be imposed on the player as a result of the premature termination." (FIFA, 2006).

According to Deloitte's annual study of the European football market size, the total revenue of European football continues to increase - the value of the season 17/18 was $28.4 €$ billion. (Barnard et al., 2019) With a contribution of $14.7 €$ billion only in the top 5 leagues (England, Germany, Spain, Italy and France). Portugal was in the $9^{\text {th }}$ position in this study, with overall revenue of $431 €$ million. It is possible to see this revenue growing substantially in the past years due to the broadcast deals. Clubs are receiving enormous quantities of these deals, and their leading receivables are from the broadcast. For instance, Real Madrid sold its sponsorship rights to a company for $\$ 224$ million in a four-year deal. Also, Barcelona signed a shirt sponsorship for \$246 million.

For those reasons and many others, Real Madrid and Barcelona are at the top of ranking the most valuable clubs in the world in 2019, made by Forbes. The top 5 of this ranking are: Real Madrid evaluated in $\$ 4.239$ billion, followed by Barcelona with $\$ 4.021$ billion. The rest of the top 5 are Manchester United ( $\$ 3.02$ billion), Bayern Munich ( $\$ 3.024$ billion) and Manchester City ( $\$ 2.688$ billion). On the rest of the top, essential to highlight that only clubs from the Big-five leagues are ranked, with most clubs from the English Premier League. (Forbes, 2019a) Another example of the increased amount of money in the football deals is the tournaments' prize money. For instance, the prize money for the complete edition of the 2018-2019 champions league achieved a record of $\$ 2.28$ billion, more than $30 \%$ than the previous edition. (Forbes, 2019b)

Another aspect verified in the past years is that the overall player's transfer fees increase yearly. Mainly because of this growth in football investments. From the top ten highest transfer fees, the seven highest are from 2017 until 2019 summer. On the top are the Neymar's transfer (2017-2018) evaluated in $222 €$ million, followed by Philippe Coutinho's (2017-2018) with $145 €$ million. The last of these top ten is the transfer of Eden Hazard to Real Madrid for $100 €$ million in the 2019 summer. (Statista, 2019) It is also notable the tendency of higher transfer fees over the last two decades. Furthermore, according to the 2020 ranking of the top100 highest-paid athletes globally, there are fourteen football players in the ranking. (Forbes, 2020). These high values are one more proof that players are one of the best assets of the clubs, and they must be well managed to provide the best club performance possible.

## 2. Literature Review

## Job turnover in general companies

There are some critical studies regarding collective turnover which resulted in contributing with essential insights of the influence on companies' performance outputs such as customer service (Koys, 2001), financial performance (Batt, 2002; Huselid, 1995; Michele Kacmar et al., 2006) or labour productivity (Guthrie, 2001; Siebert \& Zubanov, 2009). Another approach can also be held, which is studying what influences the collective turnover. Hancock et al. (2017) took that approach. They analysed the impact of recent prior firm performance on the turnover and the effect of contagious turnover, suggesting that contagion effects can happen at an aggregate level. Hausknecht \& Trevor (2011) also studied collective turnover through a framework that allowed to analyse the antecedents and the consequences of the collective turnover. Hancock et al. (2013) and Park \& Shaw (2013) focused on the performance impact. Hancock et al. (2017), to prevent the turnover, focuses on High Commitment HR Systems allied with trials on increasing satisfaction, commitment and perception for fairness levels to limit the collective turnover. Mohsin et al. (2015) explained the staff turnover in luxury hotels by some other antecedents. In this study, the feeling of job security, earnings and loyalty to the company were taken as variables. O'Fallon \& Rutherford (2010) appointed as significant causes of staff turnover the treatment by superiors, excess of working hours, job pressure, scheduling, training, better opportunities in a different place, and physical demands of the job. Ogbonna \& Harris (2002) concluded that low remuneration is the crucial point for turnover. Allen et al. (1979) concluded that an outsider substitute has a worse impact on the team. Also, Grusky (1963) found that insiders' successions are less troublemaking than an outsider succession.

## Manager turnover in sports

Sports teams also offer a starting point for study and performance comparisons through the skills of managers (Audas et al., 2002). The study of turnover started with Grusky's study in 1963. The author studied the managerial succession and organisational effectiveness with data of the MLB - Major League Baseball, the professional baseball American league. The most popular and accurate conclusion regarding manager turnover admits that the changes in managers are a scapegoating ritual Gamson \& Scotch (1963). ter Weel, in 2011, also tried to conclude whether the manager turnover improves firm performance with data from the Dutch Soccer. The author ter Weel (2011) concluded, if managers still have an extended period left on the contract, they are less likely to be fired because of the financial implications it would take for the company. The authors van Ours \& van Tuijl (2016) put together data from 14 seasons about the number of coach dismissals on the top European leagues. The highest rate of sacked coaches happened in Serie A (Italy), with a mean of 8.4 per season. Secondly, La Liga (Spain) with 6.7 but closely followed by Bundesliga (Germany) with 6.6 dismissals per season. Premier League (England) with 5.6 and Ligue 1 (France) with 4.7 complete the ranking.

## Squad stability, player turnover and team performance

The best performing teams had a percentage of new signings of $31.3 \%$. On the other hand, the least competitive had around $42 \%$. There were some league champions with a high percentage of new signings, although they were in leagues of countries where squads are traditionally unstable. The record percentage among champions occurred in 2011 with PFC Ludogorets Razgrad of Bulgaria $-92 \%$ of new signings on the squad. On the top ten in the table of champions with the lowest percentage of new signings are many clubs of the five major leagues, except for Italy (Serie A):
FC Bayern Munich (Germany - 2016): 9.1\%
FC Barcelona (Spain - 2012): 11.5\%
Manchester United (England - 2010): 12.9\%
LOSC Lille (France - 2010): 13.6\%
Chelsea FC (England - 2009): 14.8\%
These numbers are especially remarkable when compared with the average percentage among champions of the major leagues from 2009 until 2017: Liga NOS $(38,4 \%)$; Serie A $(33,4 \%)$; EPL $(30,7 \%)$; Ligue 1 ( $28,9 \%$ ); Bundesliga (22,4\%) and La Liga (20,8\%). The average on all champions of all leagues was 34\%. According to the authors, this number allowed to confirm that clubs considered stable are in the majority (CIES, 2018).

In the same way as the previous authors, Kounetas (2014) studied efficiency related to Greek clubs' performance before and after the Euro 2004. The study is divided into two analyses: first, a bootstrapped DEA to determine the efficiency and, secondly, an investigation to determine possible factors affecting the efficiency.

Cohen \& Bailey (1997) defined a team as a group of individuals working as independent entities inside an organisation fighting for a common pursued outcome. However, organisational researchers use the term team for groups that have included high interdependence, which is a term related to the coordination of team members to achieve the desired outcome. Interdependence is as high as the necessity of coordination between team members to accomplish a task (Gully et al., 1995). When this coordination is optimised, the performance is efficiently maximised. Team stability also becomes a factor with such importance when talking about interdependence.

## 3. Methodology

The data collection started with deciding which leagues would take part in the study. The big-5 - Spain, Germany, England, Italy, France - were predictable since they are the leagues with the most impact worldwide. This impact is economically speaking and in terms of popularity, as it was possible to observe in the second chapter of this dissertation. Moreover, these leagues are the ones who contribute the most with winners of the UEFA Champions League and the FIFA Club World Cup, which are the most important clubs' competitions in the world. The insertion of the Portuguese football league Liga NOS - is also easily explained by the assiduous presence of 2 or 3 clubs in the Champions League. Besides this, Portugal is getting close to France on the UEFA ranking, and the overall club comparison is similar - except for Paris Saint-Germain. (UEFA, 2020)

Secondly, the step was to decide in which seasons the study would focus. The intention was to consider the most recent seasons and the most recent data. However, one detail impacted the decision: Ligue 1 2019/2020 (France) was not concluded due to the pandemic situation in 2020. Consequently, the data from the rest of the leagues would be disproportional in comparison to France. Thus, the decision was to collect data from 4 seasons to have three variations from one year to another to include in the study: 2015/2016; 2016/2017; 2017/2018; 2018/2019.

The dataset collection started with aggregating all the clubs that participated in these leagues, divided by league and year. These lists allowed us to understand which clubs were qualified and disqualified from year to year and will not have enough data to be included in the model. Then, the minutes of all players participating in these leagues by the club were collected from the website and data store ZeroZero - each data set collected was of all players fielded from each club on each season. (ZeroZero, 2020) This data would then be used to calculate some variables such as the percentage of new players besides the same percentage although from the players that already were in the club the previous season. Afterwards, the data related to the clubs were collected. This data was collected from SoccerStats and TransferMarkt, two significant websites and data stores related to football worldwide (SoccerStats, 2020; TransferMarkt, 2020). This data compilation resulted in many variables: Games Played; Points; Victories; Draws; Losses; Goals Scored; Goals Conceded; Goal Average (Difference between scored and conceded); \% Clean Sheets (games percentage without conceding a goal); \% Failed to Score (games percentage without a goal scored); Players Fielded and Squad Value.

Additionally, on 2016/2017, on 2017/2018 and 2018/2019, since all had a previous season included in the study, it was also calculated four additional variables regarding variations:

- Sum of New Players Minutes
- Sum of Previous Season Players Minutes
- \% of New Players Minutes
- \% of Previous Season Players Minutes

Finally, all variables were calculated to establish the variation from season to season. After the aggregation of both collected and calculated variables, the matrix of all variables included 464 observations on the totality of the clubs.

## Variables

The literature review made it possible to analyse multiple studies regarding performance and turnover to collect some insights into interesting variables to include in this dissertation. The studies included variables like the salary turnover, previous year's attendance, current and previous season winning percentage, number of championships, city population, recently built arena, expenses on transfers and contract renewals, operational costs, total expenses, profit margin, assets to debt ratio, club's age, location or presence on international competitions. These variables are from different areas of investigation such as finances, demographics or performance; however, all can be important for explaining player turnover. In terms of output variables, the studies were more similar and less divergent using different
variables. The variables used were points earned, goal average and total attendance. Of course, these variables represented all variables used in studies compared with this dissertation and were not all included.

First, it was done a covariance table to help choose which variables to use. This table aims to analyse which variables can be related to each other and minimise the impact of one variable hiding the impact of the other. Thus, this table confirmed that some variable sets were not possible together in a model. It was assumed that any correlation value (in absolute value) above 0.6 was too high and could interfere with the model results. The essential variable and the first chosen was one related to turnover directly. The options were: Sum of New Players Minutes, Sum of Previous Season Players Minutes, percentage of New Players Minutes and percentage of Previous Season Players Minutes. In order to be different from the previous studies and to have a more consistent and more accessible way to analyse, the primary variable chosen was the percentage of New Players Minutes. This variable is critical since just the variable itself can give essential information regarding how important the goal of this dissertation is and highlight the balance of new players and "old" players on every squad from 2016 to 2019 in the six major leagues.

On the other hand, the usage of this variable requires information from two seasons, which results in the significant limitation of this dissertation. Clubs that were present in the league in the previous season are not considered. Consequently, the observations are cut from 464 to 302 in total.

The process of choosing the rest of the input variables was complicated. However, some were highlighted and tested after to conclude regarding the insertion in the model. The list of possibilities was:

- Squad Value: essential to distinguish a big club from a small club
- Goals Scored: essential to analyse the offensive performance of the team
- Goals Conceded: essential to analyse the defensive performance of the team
- \% Failed to Score: also related with the offensive performance, however on a different approach
- Players Fielded: number of players the club used for the entire season

The choice was squad value, and the number of players fielded. The first is because, as mentioned above, it is essential to distinguish between big and small clubs. The covariance values with the variables involved are all above 0.2 and below 0.6 (in absolute value), which is a variable related to the remaining variables. The second variable, because the first has significant correlation values and has a negative correlation, is vital to increase the understatement of results. Both Goals Scored and Conceded have either values out of the range or values close to zero on the covariance, so it was decided to take them out of the DEA model; however, it is included in the regression control variable.

Regarding the output variables, there were some options analysed, such as:

- Points: points earned from victories and draws
- Variation of points: difference of points from the current season to the previous one
- Goals Scored
- Goals conceded
- Goal Average: the difference between scored and conceded goals

Both goals scored and conceded were easily eliminated since they did not have robust results that could justify what was expected. Also, goals scored when performed alone, and with \%NPMin showed results with a high number of clubs with maximum efficiency and no evidence was concluded from those results. The variable of goal average proved to have results very similar to the variable points. Also, this can be justified since both have numbers of covariance very similar with all the input variables. Note that this variable was one of the most used in studies related to football. In this dissertation, the variable "goal average" is no more than confirming what is observable with the output "points". The output variable points will be used better to control the results regarding the variation of points. Thus, the two output variables used were points and variation of points. The second one is important to analyse the impact of the new players on the difference in points from the previous season, which is the main focus of the study.

In conclusion, on the input side, the variables are the percentage of minutes of new players (\%NPMin), squad value (SqValue), and the number of players fielded (PFielded). On the output side, points (Pts) and variation of points (VarPts) were the choices as mentioned before. Next, it will show some statistics regarding the variables to understand better which values the variables take.

## - \% New Player Minutes:

1. The average from all clubs for the six leagues was 34,18\%
2.The minimum value found was on Premier League with 2.95\% (Tottenham Hotspur FC - season 2018/2019)
3.The maximum value found was on Liga NOS which has the highest values on every season and with values almost $10 \%$ higher - with 79,87\% (Moreirense FC - season 2017/2018)
As expected, this is a variable with a significant deviation

- Squad Value
1.The average from all clubs for the six leagues was 201,94 Million Euros
2.The lowest value was observed on Liga NOS justifiable by the smaller economic size in comparison with the rest - with 11,48 Million Euros (CD Aves - season 2018/2019)
3.The highest value was observed on La Liga with 1160 Million Euros (FC Barcelona - season 2018/2019)
Also, as expected and justified by the growth of big clubs in modern football, the deviation from big clubs to small clubs is quite noticeable.

[^0]Leverkusen (Bundesliga), Brighton \& Hove Albion FC and Manchester City FC (Premier League)
3. The highest value was observed on Ligue 1 with 41 players fielded (AS Monaco FC - season 2017/2018)

## - Points

1.The average points from all clubs for the six leagues were 52,52
2.The lowest score in points observed occurred on Premier League with 16 (Huddersfield Town AFC season 2018/2019)
3.The highest score in points observed also occurred on Premier League with 100 (Manchester City FC - season 2017/2018)
4. Note that Bundesliga and Liga NOS have fewer points since both have only 18 teams instead of 20 as the other four leagues. Consequently, each team play less than four games. Even though this is not a variable with a significant deviation

- Variation of Points
1.The average of the difference in points of one season from the previous one was $-1,50$ points
2.The lower variation was recorded on Ligue 1 with -44 points (AS Monaco FC - season 2018/2019)
3.The highest variation occurred in Premier League with 43 points (Chelsea FC - season 2016/2017)
4.As expected, this variable has a more significant deviation than the variable "Points" and interestingly has more observations with negative than with positive variation
The main variables of interest are the input side, the percentage of minutes of new players, and the variation of points on the output side. Both variables are related to the change from one season to the next, and the goal of the dissertation is indeed analysing the impact of that change.


## Model

The methodology applied was two different approaches to have more robust results and analyse the data differently. First, was performed a DEA and then a multiple linear regression.

DEA-VRS (also called DEA-BCC) was used with an input orientation. The software allowed to have the results was MATLAB, and the toolbox was retrieved from Álvarez et al. (2016). The choice of this model came after analysing the potential of both DEA-CRS and DEA-VRS. The first was the first DEA model performed by Charnes, Cooper and Rhodes. Generally, DEA-CRS gives more conservative results since it does not divide the efficiency into Technical Efficiency and Scale Efficiency. Also, this was proved by testing the model on the MATLAB DEA toolbox that was used. The results were difficult to analyse when the model performed was DEA-CRS since the variety of efficiency scores was lower. Also, many DMUs were scored with 0 on the efficiency ranking. From this evidence, it could be concluded that DEA-CRS would be more effective with a model with a higher number of independent variables. However, this was refuted since the number of zeros was extremely high on all the observations. Consequently, the model chosen to be used was DEA-VRS, even with the
downside of sometimes having an excess of maximum efficiency DMUs.

First, the DEA model was performed on each league and each year. The main reason was to avoid club repetitions and comparisons between teams that do not compete against each other. Thus, each output of the model is related to a specific league season. The goal was to conclude regarding the impact of the new players on the performance of the team. For that reason, the primary analysis was done with \%NPMin as the only independent variable and VarPts as the dependent variable. Besides this, it was also performed DEA analysis with a combination of the other two variables - SqValue and PFielded - with the primary independent variable \%NPMin. These combinations were done both with VarPts and Pts as the dependent variable.

Secondly, to confirm the results of the DEA model and analyse whether the input variables can predict the behaviour of the dependent variable, a linear regression was performed in STATA software. Each model was performed with fixed effects on league and year, i.e. the model is performed without affecting the league and the year, which allows avoiding analysing the same clubs and comparing clubs from different leagues (which do not compete against each other). The linear regression also allowed us to understand whether exist any variable interaction. This interaction happens when a variable can only take different values if another variable also changes to achieve the output.

## 4. Results and Discussion

Across the results, it is easily noticeable that there is no direct and extreme relation between a low percentage of new player minutes and efficiency (when the output variable is VarPts). In DEA models, instead of that, what is noticeable is that generally, a lower percentage is related to higher efficiency, although with some outliers that can be explained with the behaviour regarding VarPts. On the other hand, on Linear Regression Models, the interaction effect between \%NPMin and PFielded affects team performance.

## DEA Model results

Table 1 - DEA-VRS results of EPL 2019 with \%NPMin and VarPts

| DMU | \%NPMin <br> $(\mathrm{X})$ | VarPts <br> $(\mathrm{Y})$ | Efficienc <br> y |
| :---: | :---: | :---: | :---: |
| Liverpool FC | $21,0 \%$ | 22 | 1 |
| Manchester City FC | $3,6 \%$ | -2 | 1 |
| Tottenham Hotspur FC | $3,0 \%$ | -6 | 1 |
| Watford FC | $23,2 \%$ | 9 | 0,4967 |
| Crystal Palace FC | $19,3 \%$ | 5 | 0,4484 |
| Southampton FC | $23,8 \%$ | 3 | 0,3016 |
| Arsenal FC | $33,8 \%$ | 7 | 0,2982 |
| Everton FC | $30,1 \%$ | 5 | 0,2872 |
| West Ham United FC | $45,9 \%$ | 10 | 0,267 |
| Leicester City FC | $32,7 \%$ | 5 | 0,2639 |
| Newcastle United FC | $24,9 \%$ | 1 | 0,23 |
| Burnley FC | $13,3 \%$ | -14 | 0,2222 |
| AFC Bournemouth | $26,8 \%$ | 1 | 0,2139 |
| Manchester United FC | $14,0 \%$ | -15 | 0,2112 |
| Chelsea FC | $39,0 \%$ | 2 | 0,1658 |
| Brighton \& Hove Albion | $20,0 \%$ | -4 | 0,1624 |
| FC | $21,1 \%$ | -21 | 0,14 |

The two lowest results of our input variable were given the maximum result of efficiency, 1 - Manchester City FC and Tottenham Hotspur FC - even with negative values of the output variable. On the other hand, Liverpool FC, with approximately one new player in every five in the squad a seven times higher proportion than the previous two clubs - also had the maximum efficiency value. However, note that the variation of points on the Liverpool case is notable since they made the most significant recovery from the previous season. Consequently, the efficiency rate is easily justified by that result. Another interesting case is West Ham United FC. They had almost half of the squad composed of new players, and even though they had a rise of 10 points compared with the previous season - the second-highest score on this EPL season. However, West Ham is placed on half of the table regarding efficiency. A similar \%NPMin of Liverpool is Brighton \& Hove Albion FC, although their performance compared with the previous season declined. So those new players did not correspond to a reasonable efficiency rate.

Regarding the cases with higher \%NPMin, they are not directly connected to a lower value of efficiency, above it was already exposed the case of West Ham United FC. Moreover, Chelsea FC, the club with the second-highest \%NPMin, is the third-lowest score inefficiency. Also, if Manchester City, Tottenham (the two lowest), West Ham and Chelsea are not considered, the values of \%NPMin are relatively uniform, around $20-30 \%$. Consequently, the efficiency values are more influenced by the performance regarding the variation of points than with the independent variable. Huddersfield Town AFC, Manchester United FC and Burnley FC, which are the three
lowest in the variation of points, are all below the middle of the table when sorted by efficiency.

Assuming that Liverpool was an outlier in this dataset, the model was performed again without that DMU. Manchester City and Tottenham had the same result, which was expected accordingly with their lower percentage of new player minutes. On the other hand, Watford FC and West Ham United, even with a higher \%NPMin, have the two best VarPts and were scored with the maximum value of efficiency as Liverpool. Thus, this model proved what was said above when Liverpool was included, and the conclusion is that Liverpool was not so of an outlier. Teams with the best recovery in points in comparison to the previous have the maximum efficiency. Except for Manchester City and Tottenham, the clubs are almost ordered from the highest to the lowest variation of points in this case. Note that the exceptions of this ordering are because of a higher \%NPMin, which also penalises the team - cases of Brighton \& Hove Albion FC and Chelsea FC. This ordination can be justified from what was mentioned above in the case with all the teams. When the independent variable results become more homogenous, the dependent variable influences the results (VarPts).

The rest of the leagues have similar results regarding what has been explained above in English Premier League. The lowest value on \%NPMin has, in all leagues, the maximum value of efficiency. Even if the performance has declined compared with the previous season, look at Ligue 1, La Liga and Liga NOS - which was also observable in Manchester City and Tottenham on the above analysis. There are also cases compared with the Liverpool case, and even with more noticeable percentages - Ligue 1 and Liga NOS cases, with 48,9\% on \%NPMin and +37 on VarPts and $64,4 \%,+20$, respectively.

Moreover, in all leagues on every season, there are similar cases with the related above. Both 2018 and 2017 have similar results as 2019, important to highlight:

- The team with the lowest \%NPMin is permanently assigned with maximum efficiency
- The team with the highest variation of points regarding the previous season is permanently assigned with maximum efficiency

It is observable that every year on every league proves that the tendency is to relate significant percentages of new players with lower efficiency rates. In other words, the more stable is the squad from one year to another the better is the efficiency. Also, important to note that the league where this is more noticeable over the three years is the Bundesliga, only surpassed by the Premier League in 2017. In conclusion, all seasons proved that when the team's stability decreases - corresponding to a higher \%NPMin - the value of efficiency also decreases.

One of the study's limitations occurs when the model is performed with more than one input variable. Some datasets with two variables could give us some meaningful information, where it was possible to extract some conclusions. Ten out of sixteen clubs were assigned with maximum efficiency. Although, the majority presented several DMUs with maximum efficiency that made it impossible to conclude anything out of them.

There is not an evident pattern of variable set behaviour that leads to a maximum efficiency score. From
the graphics of tendency line, it was concluded regarding the influence of a low \%NPMin on efficiency and a negative correlation between an increase in \%NPMin and efficiency. Besides \%NPMin, only the dependent variable VarPts shows a noticeable tendency on influencing the efficiency score - generally, a higher variation of points results in higher efficiency. The remaining variables show minimal tendency on influencing efficiency. However, generally, the behaviour is:

- SqValue: lower SqValue reflects on a higher efficiency
- PFielded: lower PFielded reflects on a higher efficiency

The DMUs assigned with maximum efficiency show at least one variable influencing the most the efficiency score. On the other hand, there is not an ideal combination of variables that reflects maximum efficiency. Thus, it is acceptable that the result, when combining all variables, shows a high number of DMUs with a maximum efficiency score. Later with the results of linear regression, it will be possible to conclude the weight of each variable contributing to the variation of points. Next, it will be analysed the results with \%NPMin together with SqValue and together with PFielded.

Table 2- Sum of Variation of Points with \%NPMin below 20\%

| Sum of VarPts with \%NPMin < 20\% |  |  |  |
| :--- | :--- | :--- | :--- |
|  | 2019 | 2018 | 2017 |
| English Premier League | -32 | -46 | -13 |
| Bundesliga | 9 | 20 | -17 |
| Ligue 1 | -30 | - | -7 |
| La Liga | -14 | -36 | -7 |
| Liga NOS | -3 | - | -6 |
| Serie A | 9 | 2 | -3 |

It is noticeable that only Bundesliga and Serie A have positive values of variation of points. Thus, from these results, it is noticeable that a stable squad is not a direct sign of having better performance than the previous season, as in most of the leagues shown above. Even in terms of efficiency:

- Premier League: from 13 observations with less than $20 \%$ of \%NPMin, 4 (31\%) were assigned with maximum efficiency.
- Bundesliga: from 10 observations with less than 20\% of \%NPMin, $4(40 \%)$ were assigned with maximum efficiency.
- Ligue 1: from 4 observations with less than $20 \%$ of \%NPMin, 3 (75\%) were assigned with maximum efficiency.
- La Liga: from 10 observations with less than 20\% of \%NPMin, 4 (40\%) were assigned with maximum efficiency.
- Liga NOS: from 2 observations with less than $20 \%$ of \%NPMin, 2 (100\%) were assigned with maximum efficiency.
- Serie A: from 5 observations with less than $20 \%$ of \%NPMin, 3 (60\%) were assigned with maximum efficiency.

These percentages give that an average of $58 \%$ of the cases with less than $20 \%$ of new players minutes have the maximum efficiency score. Consequently, these results confirm what was concluded earlier regarding the relation
between \%NPMin and VarPts - generally, low values of \%NPMin results in higher efficiency scores.

On the other hand, these results also showed that the clubs considered a stable squad (since they did not change more than $20 \%$ of the team) had a decline in performance compared to the previous season. However, essential to notice that this decline in performance (this case related in terms of points) can be misjudged since a team can have fewer points compared with the previous season even though it can win the league. This is one of the study's limitations since the team accomplished the season goal even with a negative variation of points.

## Linear Regression Results

Now, the results of the linear regression model done on STATA software will be shown. The main variables used were the same as before on the DEA model, although it was also included Goals Scored (GScored) as a control variable and an extra one. This extra is the result of the interaction between \%NPMin and PFielded. All the models were performed without the effect of the league and the year, using fixed effects. In other words, the DMUs were entirely considered. However, only DMUs from the same year and the same league are put together. This avoids performing models with DMUs that do not play against each other at any point of the season since they are from different leagues. It also avoids including the same team from different seasons on the same model.

One important topic to notice is that the \%NPMin positively influences the output variable, VarPts, opposing what was expected and supported by DEA models. The second model was essential to establish the squad value as a control value since it does not strongly influence VarPts.

On the other hand, SqValue is critical to be in the model to distinguish big and small clubs. Also, this model and the first are not significant and have not explanatory power at any level of significance. Finally, the third model made us conclude that by increasing one player in the total number of players fielded across the season, VarPts was affected negatively. So, the increase in the number of players fielded makes the VarPts decrease gradually. Also, this model is the only one from these single regression models that showed significance at any level.

Model 5 showed excellent results and allowed recognising that \%NPMin and PFielded may have an interaction effect that was not considered. If the model is interpreted with the two variables separately, the conclusion is that one unit of \%NPMin affects the VarPts positively. Consequently, it would be good if a team had a more significant percentage of new player minutes in comparison with minutes of long last players. However, the variable PFielded shows that each player fielded (either new or long last) negatively affect the points. In fact, there is evidence that these two variables have an interaction effect that is not being considered. Also, another important topic is that model 5 is significant in all variables and parameters at the highest significance level, which concludes a better explanatory power.

Model 6 was performed with the three variables simultaneously - NPMin, SqValue and PFielded. The results were robust and continued with the same line of thought as the previous ones - \%NPMin affecting the
variation of points positively, SqValue as a control variable and PFielded negatively affecting the output variable. Note that either \%NPMin, PFielded and $\beta_{0}$ have significant power on the lowest $p$-value - resulting in a robust explanatory model.

Table 3 - Linear regression models with three or more input variables

|  | Three or more variable models | Output variable |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  |  | 6. VarPts | 7. VarPts | $\begin{gathered} 8 . \\ \text { VarPts } \\ \hline \end{gathered}$ |
| $\frac{y}{0}$ | \%NPMin | $\begin{aligned} & 23,534 \\ & (-3,24) \\ & \hline \end{aligned}$ | $\begin{array}{r} 27,14 \\ (-3,98) \\ \hline \end{array}$ | $\begin{aligned} & 208,395 \\ & (-3,91) \end{aligned}$ |
|  | SqValue | $\begin{aligned} & 0,005 \\ & (-0,63) \end{aligned}$ | $\begin{aligned} & -0,012 \\ & (-1,03) \end{aligned}$ | $\begin{aligned} & -0,011 \\ & (-0,93) \\ & \hline \end{aligned}$ |
|  | PFielded | $\begin{aligned} & -1,721 \\ & (-5,22) \\ & \hline \end{aligned}$ | $\begin{aligned} & -1,463 \\ & (-4,35) \\ & \hline \end{aligned}$ | $\begin{aligned} & 0,807 \\ & (-1,08) \end{aligned}$ |
|  | GScored |  | $0,334$ $(-3,75)$ | $\begin{aligned} & 0,332 \\ & \\ & (-3,88) \\ & \hline \end{aligned}$ |
|  | \%NPMin. <br> PFielded |  |  | $\begin{aligned} & -6,488 \\ & (-3,5) \\ & \hline \end{aligned}$ |
|  | $\beta 0$ | $\begin{aligned} & 40,371 \\ & (-3,76) \end{aligned}$ | $\begin{aligned} & 16,116 \\ & (-1,26) \\ & \hline \end{aligned}$ | $\begin{aligned} & -45,868 \\ & (-2,07) \end{aligned}$ |
|  | Adjusted Rsquared <br> F-statistic | 0,18 | 0,25 | 0,28 |
|  |  | 7,15 | 9,12 | 9,08 |


| ${ }^{*} p<0,1$ |
| :---: |
| ${ }^{* *}$ |
| $p<0,05$ |
| $* * *$ |
| $p<0,01$ |
| Number <br> s in <br> brackets <br> are the <br> standar <br> $d$ <br> deviatio <br> $n$ |
| $n$ |

Before performing the final model, and with the justification of increasing the robustness of the final model, one more control variable was included - GScored - which helped to understand the model was well built since GScored also affects VarPts positively as expected.

$$
\begin{aligned}
\text { VarPts }=208,4 \text { NPMin } & -0,011 \text { SqValue }+0,81 \text { PFielded } \\
& +0,33 \text { GScored }-6,49 \text { NPMin. PFielded } \\
& -45,87
\end{aligned}
$$

This model has an additional variable related to that interaction effect - \%NPMin.Pfielded. Therefore, there are three variables of interest on this final model and two control variables, as explained earlier. In order to better interpret interaction effect results between the two main variables, next will be shown a different way to analyse the model. These equations represent the impact of \%NPMin on VarPts and the impact of PFielded on the same variable, respectively.

$$
\frac{\Delta \text { Pts }}{\% \text { NPMin }}=208,4-6,49 \text { PFielded }
$$

It is possible to understand that a minor variation of points happens when there is an increase in \%NPMin and a higher number of players fielded, which we could conclude earlier but now is proved with this model. This equation has a robust explanatory power since both parameters involved are significant on the highest interval possible. The same does not happen with PFielded.

$$
\frac{\Delta \text { Pts }}{\text { PFielded }}=0,81-6,49 \mathrm{NPMin}
$$

With a similar purpose, the objective this time was to understand the impact of PFielded, including the interaction effect on the output variable. However, the parameter of PFielded on the primary equation of the model has no significant power, which makes this
equation having less explanatory power than the previous. In this case, an increase in \%NPMin would affect on a large scale the variation of points negatively.

## 5. Conclusions

The world of modern football is related a lot to the enterprise environment. In the last decades, football clubs became and continue to evolve into huge companies to manage and need to optimise their more valuable assets, which are the players.

Football as we know it nowadays did not exist some decades ago. Today, the technology is a lot incorporated in studies regarding teams, players, performance in general, and many statistical methods used in the football world. In this dissertation, the impact of changes in the roster on team performance is studied.

Regarding DEA models, the main difficulty was to have strict conclusions with the results. However, it was possible to analyse a less deep look and see a tendency line in all leagues every season. All observations showed the same tendency line, only differing on the line slope. On the \%NPMin vs Efficiency analysis, it was clear that a higher \%NPMin relates to a lower efficiency rating. Bundesliga was the league where this relationship was more notable over the three years of analysis, only surpassed by the Premier League in 2017. In other words, all seasons proved that when the team's stability decreases, the value of efficiency also decreases. From DEA models, another strong tendency was analysed even though not so important as the primary variable of the study (\%NPMin), which was the behaviour of PFielded vs Efficiency. The variable PFielded had similar results as the primary variable in that when a team has a higher number of PFielded (less stability), the efficiency rating is low.

The next step was to perform the linear regression model with different variables and some new ones to have a more robust model. The first expectations were that the results would prove and support the previous DEA ones. Both models supported the same conclusions; however, these regression models allowed us to analyse and understand the interaction effect between \%NPMin and PFielded. This interaction effect was proved to hurt the overall team performance, which supported the results obtained before by the DEA models. Consequently, the main goal of this dissertation was accomplished and was aligned with the existent literature.

## Future Considerations

Every study and dissertation can help giving insights for further investigations regarding the same subject or investigations that will use the same models applied here. For further investigations, here are the recommendations:

1. New variables More leagues
Expand to other sports
2. Presidents and Directors turnover instead of player turnover
3. Apply optimisation tools and models

## 6. References

Allen, M. P., Panian, S. K., \& Lotz, R. E. (1979). Managerial Succession and Organisational Performance : A Recalcitrant Problem Revisited. 24(2), 167-180. http://www.jstor.org/stable/2392492
Álvarez, I. C., Barbero, J., Zofío, J. L., \& Barbero, J. (2016). A Data Envelopment Analysis Toolbox for MATLAB.
Audas, R., Dobson, S., \& Goddard, J. (2002). The impact of managerial change on team performance in professional sports. Journal of Economics and Business, 54(6), 633-650. https://doi.org/10.1016/S0148-6195(02)00120-0
Barnard, M., Boor, S., Winn, C., Wood, C., \& Wray, I. (2019). Annual Review of Football Finance 2019. Deloitte Annual Review of Football Finance 2019, May, 40.
Batt, R. (2002). Managing customer services: Human resource practices, quit rates, and sales growth. Academy of Management Journal, 45(3), 587-597. https://doi.org/10.2307/3069383
CIES. (2018). Monthly report 34 - The importance of squad stability: evidence from European football. CIES Football Observatory. https://footballobservatory.com/IMG/pdf/mr34en.pdf
Cohen, S. G., \& Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. In Journal of Management (Vol. 23, Issue 3). https://doi.org/10.1177/014920639702300303
FIFA. (2006). Commentary on the Regulations for the Status and Transfer of Players. 76. http://resources.fifa.com/mm/document/affedera tion/administration/51/56/07/transfer_commenta ry_06_en_1843.pdf
Football-Stadiums.co.uk. (2020). Football Player Transfers Explained. https://www.football-stadiums.co.uk/articles/football-player-transfersexplained/
Forbes. (2019a). The Business Of Soccer - 2019 Most Valuable Clubs Ranking. https://www.forbes.com/soccervaluations/list/\#tab:overall
Forbes. (2019b). The World's Most Valuable Soccer Teams 2019.
https://www.forbes.com/sites/mikeozanian/2019/ 05/29/the-worlds-most-valuable-soccer-teams2019/\#1c61300540d6
Forbes. (2020). The World's Highest-Paid Athletes List Forbes. Forbes.Com. http://www.forbes.com/athletes/list/
Gamson, W. a, \& Scotch, N. a. (1963). Scapegoating in Baseball AND DEBATES Scapegoating in Baseball. American Journal of Sociology, 70(1), 69-72.
Grusky, O. (1963). Managerial Succession and Organizational Effectiveness. American Journal of Sociology, 69(1), 21-31. https://doi.org/10.1086/223507
Gully, S. M., Devine, D. J., \& Whitney, D. J. (1995). A Meta-Analysis of Cohesion and Performance: Effects of Level of Analysis and Task Interdependence. 26(4), 497-520.

Guthrie, J. P. (2001). High-Involvement Work Practices, Turnover, and Productivity: Evidence From New Zealand. Academy of Management Journal, 44(1), 180-190. http://www.jstor.org/stable/3069345
Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., \& Pierce, C. A. (2013). Meta-Analytic Review of Employee Turnover as a Predictor of Firm Performance. Journal of Management, 39(3), 573603. https://doi.org/10.1177/0149206311424943

Hancock, J. I., Allen, D. G., \& Soelberg, C. (2017). Collective turnover: An expanded meta-analytic exploration and comparison. Human Resource Management Review, 27(1), 61-86. https://doi.org/10.1016/j.hrmr.2016.06.003
Hausknecht, J. P., \& Trevor, C. O. (2011). Collective turnover at the group, unit and organisational levels: Evidence, issues, and implications. Journal of Management, 37(1), 352-388.
Huselid, M. A. (1995). The Impact Of Human Resource Management Practices On Turnover, Productivity, And Corporate Financial Performance. Human Resource Management, 3(3).
Kounetas, K. (2014). Greek foot ball clubs' efficiency before and after Euro 2004 Victory: a bootstrap approach. Central European Journal of Operations Research, 22(4), 623-645. https://doi.org/10.1007/s10100-013-0288-5
Koys, D. J. (2001). The effects of employee satisfaction, organisational citizenship behavior, and turnover on organisational effectiveness: A unit-level, longitudinal study. Personnel Psychology, 54(1), 101-114. https://doi.org/10.1111/j.17446570.2001.tb00087.x

Lukomski, J. (2020). On the finalisation of international football transfers and professional football players' contracts. The International Sports Law Journal, 0123456789. https://doi.org/10.1007/s40318-020-00170-y
Madina Tatraeva. (2020). Buy-out clauses in football contracts. ILaw.Co.Uk. https://www.ilaw.co.uk/news/buy-out-clauses-in-football-contracts
Michele Kacmar, K., Andrews, M. C., Van Rooy, D. L., Chris Steilberg, R., \& Cerrone, S. (2006). Sure everyone can be replaced . . . But at what cost? Turnover as a predictor of unit-level performance. Academy of Management Journal, 49(1), 133-144. https://doi.org/10.5465/AMJ.2006.20785670
Mohsin, A., Lengler, J., \& Aguzzoli, R. (2015). Staff turnover in hotels: Exploring the quadratic and linear relationships. Tourism Management, 51, 3548. https://doi.org/10.1016/j.tourman.2015.04.005
O'Fallon, M. J., \& Rutherford, D. G. (2010). Hotel Management and Operations (5th ed.). John Wiley \& Sons, Inc. https://books.google.pt/books?hl=ptPT\&|r=\&id=GOAnmDecKpQC\&oi=fnd\&pg=PR15\&ot $s=s p 3 E q F 9 G K A \& s i g=0 w z 0 A o K 7 K m v I D 8 J P k R 2 i B g d D$ eJ4\&redir_esc=y\#v=onepage\&q\&f=false
Ogbonna, E., \& Harris, L. C. (2002). Managing organisational culture: Insights from the hospitality industry. Human Resource Management Journal, 12(1), 33-53. https://doi.org/10.1111/j.1748-
8583.2002.tb00056.x

Park, T. Y., \& Shaw, J. D. (2013). Turnover rates and organisational performance: A meta-analysis. Journal of Applied Psychology, 98(2), 268-309. https://doi.org/10.1037/a0030723
Siebert, W. S., \& Zubanov, N. (2009). Searching for The Optimal Level of Employee Turnover: A Study of a Large U.K. Retail Organization. Academy of Management Journal, 52(2).
SoccerStats. (2020). Clubs Statistics. https://www.soccerstats.com/
SportsShow. (2020). Top 10 Most Popular Sports in the World. https://sportsshow.net/top-10-most-popular-sports-in-the-world/
Statista. (2019). Soccer players highest transfer fees alltime 2019.
https://www.statista.com/statistics/263304/transf er-fees-the-10-most-expensive-transfers-in-soccerever/
ter Weel, B. (2011). Does Manager Turnover Improve Firm Performance? Evidence from Dutch Soccer, 1986-2004. Economist, 159(3), 279-303. https://doi.org/10.1007/s10645-010-9157-y
TotalSportek. (2017). 25 World's Most Popular Sports (Ranked by 13 factors). https://www.totalsportek.com/most-popularsports/
TransferMarkt. (2020). Clubs Statistics. www.transfermarkt.com
UEFA. (2020). Country coefficients / UEFA Coefficients. https://www.uefa.com/memberassociations/uefar ankings/country/\#/yr/2020
van Ours, J. C., \& van Tuijl, M. A. (2016). In-season headcoach dismissals and the performance of professional football teams. Economic Inquiry, 54(1), 591-604. https://doi.org/10.1111/ecin. 12280
ZeroZero. (2020). Minutes played by all Players fielded. https://www.zerozero.pt/home.php


[^0]:    - Players Fielded

    1. The average from all clubs for the six leagues was 27,71 players
    2. There were three clubs with the same value, which was the minimum observed, all on 2018-2019 season, with 21 players fielded - Bayer 04
