Production Efficiency of Teams in the Premier League and Serie A

An Analysis Based on Teams' On-the-field Performance during 2015/16 Season

Mingchuan WANG
#355668

Erasmus University Rotterdam
August 2016

Abstract

This paper estimates the production functions of professional football teams in the EPL and Serie A using on-the-field performance statistics as input. The identification of the production functions facilitates efficiency rankings within a league as well as cross-league comparisons. Result suggests there exist distinctions in the process of *chance creation* between the two leagues while *goal conversion* is universal. With the additional information on player valuation, the implications of the efficiency rankings are thoroughly discussed. In general, for a team to improve, they should preserve their efficient departments and restructure their inefficient departments.

Key Words: association football (soccer), production function, efficiency, player valuation, OLS regression.

Table of Contents

1. Introduction

Association football[1] has been existing for more than a century. Its popularity and influence has boomed during the recent decades. For some people, football represents 'the most important of the less important things in the world'; while for others, it can be 'much more serious than the matter of life and death'[2]. Believer or not, there is no denying the fact that *The Beautiful Game* is the most popular sport, or perhaps even the most enjoyed leisure activity globally (Bridgewater, 2014).

The recent development of the professional football industry benefited largely from the popularisation of live television broadcasting started in the late eighties. Television broadcasts enable enthusiasts around the globe to enjoy live actions of their favourite teams without leaving their own living-rooms. In the meantime, astronomical amount of revenue is generated for the clubs through television rights deals, allowing them to grow into multinational giants. According to the latest annual reports published by Deloitte, the Annual Review of Football Finance (2015) and Football Money League (2016), the total European football market is estimated to be beyond €20 billion (2013/2014 season), while the top ten European football clubs generate a combined revenue of €4.65 billion (2014/15 season). For professional football clubs, or professional competitive sports in general, the economic performances are always related to their athletic performances and commercialisation made that link tighter. For instance, it is estimated that failing to qualify for the Champions League costed Manchester United a £17.7m (17%) decrease in matchday revenue in the 14/15 season (Deloitte, 2016). Evidence from stock market is even more staggering: after suffering a dismal spell in the middle of the recent season (15/16), the Red Devil's market capitalisation in the New York Stock Exchange fell by £412m comparing to the start of the season (Kearns, 2016). With enormous economic benefits at stake, the incentive to win is higher than ever. While it is safe to say that consistent on-the-field success will eventually lead to financial success, the reverse is not always true. The results of football matches are by nature highly random, even though big clubs can use high salary and transfer fee to attract better players to increase their chances of winning, there always exists a probability that the underdog would prevail and that is part of the attraction of any competitive sport.

With its ever-growing influence, the competitive sports industries have drawn the attention of economists; one of the field of interest is to study the efficiency of the clubs by using production functions. Under such settings, clubs are regarded as production firms, and depending on the focus of

[1] Or soccer, which will be referred to as 'football' in the remainder of this paper.
[2] Quotes by Carlo Ancelotti, an Italian football manager and former footballer (Kuper, 2014); and Bill Shankly, former Liverpool manager (Corbett, 2009).

the study, the relationship between a variety of input and output variables can be investigated. There are usually two types of variables, performance or financial based, either can be used as input or output. For instance, the output of a team can be sporting results such as scores, points, wins (rates), etc.; they can also be financial statistics such as revenues, turnovers, attendance, etc., or even both. Similarly, the input can be in-match performance stats such as shots on goal (football), rebounds (basketball), base runs (baseball), etc.; or financial stats such as player wage, transfer fee, etc.

The idea of using teams' output and input variables to measure their efficiency was first introduced by Scully (1974) on baseball. During the following decades, similar techniques were widely applied to other US based interactive competitive sports such as basketball and American football. It was not until the new millennium that such research was conducted on football. The lack of related research on football is mainly due to the continuous and interactive nature of the game. During a discussion on the nature of the production functions of sports, Scully (1995) explains that the level of complementarity and interactive effects of players (commonly known as 'team plays') affects the applicability and performance of production function models. In general, he concludes that it would be much easier to model sports with low level of complementarity and interactivity. On the lower end of Scully's spectrum is baseball, where player input are largely independent of each other, and thus their contribution to team outputs, whether in terms of wins or revenue, could be accurately measured. Scully's example on the other end is American football (while basketball and hockey falls in between). The high level of team plays means the production technology is not additive, but exhibits complementary features. For instance, good players' performance could be dragged down by poor teammates, *vice versa*; also, a collective production environment allows management, morale, chemistry, form and other factors to compensate the lack of quality amongst individual players. As a result, the relationship between input and output can no longer be accurately distinguished and measured. Scully's discussion was focused on the US based major professional team sports, however we can safely assume that the same theory applies to football. In terms of complementarity and interactivity, football is on an even higher level comparing to American football or any other US based competitive team sports (Carmichael et al., 2000), which creates a natural barrier to the application of production function analyses.

The high level of complementarity and interactivity of football is also reflected on the fact that football matches are highly continuous with little amount of set plays. Carmichael et al. (2001) argue that football's continuously interactive nature and its lack of set plays does not facilitate decomposition, recording, and measurement; resulting in a scarcity of player performance statistics comparing to football's American counterparts, another difficulty in modelling the production function of football teams.

Despite the above mentioned obstacles, attempts were made in the field of football. As it will be discussed in the next section, the majority of the research on productive efficiency of football teams tend to choose financial statistics over performance stats (probably due to the scarcity of the latter). Another point worth mentioning is that these research are also more in favour of using the non-parametric data envelopment analysis (DEA) models over econometric parametric models. DEA models allow the use of multiple inputs and outputs, and do not impose any functional form on the variables, nor do they make distribution assumptions on the error term. The advantage of such models is that they are flexible, especially helpful in the cases where the relationship of input and output are unknown or cannot be derived. The disadvantage is obvious: since a DEA model works like a black box, it provides little value when the interest is in the relationship between the input and output variables and its underlying economic interpretations.

Thanks to the advancements in the data collecting processes, the lack of match-play statistics has been largely alleviated during the recent years. The enrichment of performance data gave rise to new possibilities in football analytics, however, the 'old' problem of complementarity and interactivity persists: it is much more difficult to interpret football performance stats comparing to the likes of baseball or basketball. The high level of complementarity and interactivity of football means the stats are highly correlated and they provide little information standing alone. Wang (2015) discovers that defensive players are usually undervalued while attackers are usually overvalued in terms of their transfer market valuation. A plausible explanation is that it is rather difficult to interpret the defensive stats. The interpretations of these stats (including tackles, interceptions, and duals won, etc.) are heavily reliant on other team-level stats: the lack of tackles and interceptions doesn't necessarily mean a team/defender is bad at defending, it can be the result of the lack of attacking plays from the opposite side.

In order to understand and make use of these newly discovered data, this paper studies the performance of the top-level football clubs in England (the English Premier League, EPL) and Italy (Serie A) – two distinctive yet equally influential football culture. Parametric models are used to investigate the relationship between output (goals, attempts) and various combinations of performance based input variable. By choosing the best fitted model, I can estimate how a game could have ended given how well each side actually performed. The residuals of the regression analyses can then be used to rank the relative efficiencies of teams within their respective leagues. Parameters recovered from regression analyses could provide a clear indication on the relative importance of the different aspects of game plays in a certain league, and thus makes it possible to spot differences in the style of play across leagues.

While previous research on production efficiency of football teams are limited to use seasonal statistics, it is now possible to use per match data. There are several advantages of doing so. Firstly, seasonal statistics have too few data points, under such settings, the number of data points equals to the number of teams (20 for a typical EPL/Serie A season); using per match data yields significantly more data points ($10 * 38 = 380$ games played per season). Secondly, due to the interests on match forecasts, betting odds, and etc., there exists vast literatures on modelling game results, some of them provide valuable insights (e.g. it would be much easier to model the effect of home advantage on a per match basis). Additionally, using statistics on a per match basis evokes the concept of relative performance. As an interactive and competitive sport, the output of a football game is always determined by the performance of both sides. In other words, it is the relative performance of each side that really matters. For instance, for a given football match, the amount of goals scored by the home team is determined not only by the quality of attacks from the home side but also the defensive displays of the visitors, *vice versa*. It seems more reasonable to associate one team's performance to their opponent's, as it can be argued that there is no such thing as an 'absolute performance'. To analyse the sport on a per match basis captures the nature of football matches being zero-sum games: one team's win is inevitably another team's loss, and one team's over-performance (positive residual) is equivalent to the opponent's under-performance (negative residual).

In practice, the efficiencies of two aspects of football game are assessed: the creation of goal-scoring opportunities via passing, and the conversion of chances into goals via shooting; each aspect can be measured from either an offensive or defensive perspective. Results from statistical analyses suggest that the process of goal conversion is quite similar between the leagues, but the factors of creating chances are different. The results strongly imply there exists certain dissimilarities across leagues, however, due to the presence of multicollinearity, the details cannot be identified at the current level.

With the efficiency rankings, we can identify the relative strength and weakness of the teams. During the course of a season, it can be helpful for managers to arrange training sessions and to make match preparations against specific opponents. They are also helpful for clubs in making transfer decisions and the firing of managers. Since an additional analysis suggests player valuation is not influenced by the team's efficiency in the short run, a team would be better-off to preserve the players in their efficient departments, sell players from the inefficient departments, and to buy players from other teams' inefficient departments.

The remainder of this paper organises as follows. **Section 2** surveys the past literatures on production functions in team sports and modelling football game results. The potential contributions of this thesis are also discussed. **Section 3** proposes a theoretical framework on the econometric models of production function. The choices of independent variables are extensively discussed in this section. **Section 4** describes the data and sample. **Section 5** presents the results from regression analyses, a comparison between two leagues, and efficiency rankings based on residuals. **Section 6** discusses the interpretations and applications of the ranking lists. The last section concludes the paper, addresses the limitations, and discusses the direction of future research.

2. Literature Review

The infamous reserve clause and other economic issues regarding player compensations led to the first ever players' strike in baseball history in 1972. The opposing parties clashed on whether players are underpaid or overpaid. In the wake of the event, economists' interests were drawn to the labour market of sport industries. Scully (1974) argues that player salary should in theory be determined by his marginal revenue product (MRP). He uses a two-step approach to formulate this idea: firstly, the quality of the games, measured by percent wins, is related to player performance; and secondly, the revenue is related to the quality of games. Scully (1974) is one of the first to model a competitive sport under a production function, he focuses more on the analyses on marginal revenue rather than the percent wins, even so, the crude (in his own words) estimation on the percent win function using key match-play stats as input is able to produce surprisingly good results (an R-squared as high as 0.88), indicating a good penitential of the application of such methodology.

The relative low level of complementarity and interactivity and an abundance in match-play statistics made baseball an ideal field for performance based production function studies. Zech (1981) investigates the relationship between team victory and player skills using Cobb-Douglas production functions. His research is more focused on the technique aspect of the game rather than financial; the majority of the input variables are player performance data. In order to have a complete coverage on player skills, Zech made a comprehensive discussion where skills are categorised into four main categories, hitting, running, defence, and pitching, each with detailed subskills; he then matches the skills with the most representative match-play data. By doing so, Zech's models are able to explain the majority of the variation in team victories (R^2 around 0.78); moreover, the results provide interesting insights about the game which are hidden beneath the numbers. For instance, Zech debunks the conventional wisdom that "pitching is 75% of the (baseball) game", where his results indicate that hitting for average contributes almost 6 times as much as pitching to a team's success. Zech (1981) sets a good example on how to translate complicated competitive sport into mathematical expressions. Even though football has a complete deferent skillset comparing to baseball, the fundamental idea should be more than similar: game-play is breakdown into categories of skills; match the available match-play data to these skills; and carry out econometric analysis using production functions.

Before long, economists started to use production functions to assess the performance of teams in other major US based competitive sports. Based on the theoretical framework on the frontier production function by Timmer (1971), Afriat (1972), and Richmond (1974), Zak et al. (1979) investigate the production efficiency of teams in the National Basketball Association (NBA). By introducing the concept of frontier production, Zak et al. (1979) are able to estimate the maximum number of wins attainable (output) for a given combination of performance inputs, and thus answer the question whether a team is playing up to their potential. Since basketball involves more team-play comparing to baseball, it has a higher level of complementarity and interactivity. Zak et al. (1979) spend a lot of effort in justifying their choices of parameters and model formulation. “The essence of sports competition” they point out, “is to gauge one's performance relative to others.” Based on this principal, they employ the Cobb-Douglas[3] functional form using ratios rather than absolute values as output and input. The advantage of a frontier production function is that in addition to identifying the relationship between input and output, quality index and efficiency index of the teams can be constructed. Hofler and Payne (1997) conduct a similar research on the same sport, they utilize the production frontier methodology advanced by Aigner et al. (1977), where the residual of the OLS is decomposed into two parts, a random error term and a team specific nonpositive error term, the latter one functions as a measurement of inefficiency. A research on the offensive potential of National Football League (NFL) teams is also carried out by the duo (Hofler and Payne, 1996) where they use the exact same theoretical framework.

Due to the lack of recorded match-play data elsewhere, such research was largely bounded to the US for a long period of time. One of the few exceptions was Schofield (1988) where he studies the production functions of professional cricket in England, a sport akin to baseball; it has a high level of set plays and individual’s contribution to team success can be easily evaluated. It was not until the second half of the 90’s when a ‘true’ interactive sport was studied outside of the US. Following the footsteps of Scully (1974) and Schofield (1988), Carmichael and Thomas (1995) investigate the production and efficiency of British rugby league football. The authors use a multiple equation system where they first constructed a production function using performance data as input; and then further related the performance inputs to performance influencing variables. Carmichael and Thomas (1995) are able to accomplish three objectives: they identify how player’s on-the-field attacking and defending performances are affected by various team/player factors (player fitness, experience, coaching skills…); they identify the relative importance of the performance inputs on team winning; and they construct an efficiency ranking of the 35 teams involved based on the residuals of the regression analyses.

Despite its global influence, the lack of match-play performance data and the high interactive nature made football a difficult subject for production function analyses. A new mathematical technique was developed based on the pioneer work of Charnes et al. (1978) and follow-ups like Coelli et al. (1998)

[3] Their argument is that the marginal products of each input in the Cobb-Douglas form depend on the levels of other inputs.

and Cooper et al. (1999), the Data Envelopment Analysis. It provides a way to bypass the problem of misspecification of the input variables and thus promoted related research on football. Instead of performance stats, financial data such as player/manager salary are favoured as proxies of team/player quality on the input side of the equation. Examples can be found in the works of Haas (2003a and 2003b) on English Premier League (EPL) and Major League Soccer (MLS); Barros and Leach (2006) on the EPL; Guzmán and Morrow (2007) on the EPL; Espitia-Escuer and Garcia-Cebrian (2010) on Champions League; and several others. The key feature of the non-parametric DEA approach is that it is does not possess a functional form, thus poses few restrictions on the choices of input and output variables, and therefore is able to include financial data such as revenue, attendance (output); turnover, salary (input) in the analyses. However, the cost of such an approach is that the relationship between the output and input cannot be discovered.

Research using parametric production functions with match-play performance data as input are rare. Carmichael et al. (2000) are among the first to implement performance variables as input to investigate the production and efficiency among football teams. Unlike most of the production efficiency research which use seasonal summary stats of teams as data points, their paper (the only one in the long list of literatures reviewed) uses the full season fixture list of 380 matches as data points. Soon afterwards, the same group of authors (Carmichael et al., 2001) conduct an alternative research on the same subject (EPL) using seasonal data. Similar to the earlier works on rugby league (Carmichael and Thomas, 1995), by using a multi-equation model, the authors are able to measure four aspects of relative efficiency based on seasonal team performance: the conversion of goals into league points; defensive ability in terms of preventing opposition attempts/conceding goals; the conversion of shots into goals; and converting possession into shot attempts. The highlight of this research is that they made a fine attempt in rationalising the production process of football by quantifying the relationship between attacking/defending and scoring/conceding. Other influential literatures on the topic of production functions of football teams include the works of Bosca et al. (2006), Barros and Leach (2007), Barros and Garcia-del-Barrio (2008), and many others. The key features of the literatures are listed in **Table 1** in **Appendix I**.

Data collecting on top-tier football league matches has been drastically improved since the times of Carmichael et al. (2000 and 2001). The scale, variety, and accuracy of match-play statistics are higher than ever. The development on this front makes it possible to carry out more detailed analyses on per match basis. Choosing per game data as data points means each football match is treated like a production run where the production output (dependent variable) is set to be game results (whether in win/draw/loss, points, or goals scored/conceded). Due to the interests in forecasting game results, especially from betting markets, modelling football results is by no means a new topic. Even though most of the research on this field are focused on the accuracy of predictions rather than investigating the relationship between results and performance, some can still provide valuable insights. For instance, Peeters (2014) and Wang (2015) attempt to model game results using player (transfer market) valuations. Their models are similar to a production-function-setting in the sense that the market value of players (proxies of player quality) can be regarded as a type production input. The double counting strategy used by the authors, where a game is recorded twice: one from the home team's perspective and the other from the away team's perspective, will be essential to the model construction and residual analyses of this paper.

This thesis may contribute in the following ways. By following the footsteps of Scully (1974), Carmichael et al. (2000), and many others, this paper is intended to fill the gap of the lack of performance based production efficiency analyses on football, and to make sense of the dazzling amount of match-play data that has recently been made available. In the meantime, by investigating the relationships between game results and performances, this paper is one of the few attempts to carry out cross-league comparison between major football cultures. Lastly, most of the past literatures on modelling football game results use (past) game results as the sole predictor, other more informative variables such as player valuation or performance data are largely underused. I hope this paper can inspire other researchers to adopt such variables in modelling (or even forecasting) football results.

3. Theoretical Framework

3.1. Model

A production function studies the relationship between a firm's output level and the level of its inputs or factors of production. It enables us to distinguish the allocative efficiencies of the firms. For a professional competitive team sport such as football, each single contest can be viewed as an analogue to a factory production run (Zak et al., 1979; Carmichael et al., 2000; and others). Under a league[4] football context, assuming the objective of any team is to be as high as possible on the table/standings at the end of the season, then during each contest, either side will try their best to win by attempting to score as many goals as possible while to concede as few as possible. We can thus assign the number of goals scored (attempts made) by one of the sides as a convenient measure of output of the production function. A goal-oriented model has an advantage over a point/margin-oriented model since it captures the competitive interactions of attacking and defending, and therefore avoid misinterpretation on the independent variables, this point will be elaborated in detail in the latter part of the section.

[4] As opposed to friendlies or cup football, in which the incentives of the teams might be slightly different.

On the other side of the equation, the factors of this goal-producing process (input) include the quality of the on-field players and substitutes, the quality of manager and other coaching staffs (Scully, 1974), and other factors such as facilities (stadia and training equipment), weather conditions (Schofield, 1988), etc. Note that it is inappropriate to use factors such as club capital *alongside* the above-mentioned factors since the wealth of the clubs are already reflected on the quality of players and staffs they hire.

According to Scully (1995), it is relatively easier to measure and model sports with low levels of complementarity and interactivity. With low levels of inter-play between players (e.g. baseball and basketball), the quality of a team can be regarded as a (linear/log linear) aggregation of the quality of its individual players. To construct a production function, we just need to plug in the proxies for individual player qualities such as player game stats, height/weight (physicality), age (experience), etc. For football however, such an approach is inadequate due the lack of *ex ante* player talent indicators. The continuous nature of football games makes it difficult to breakdown matches into set plays and thus does not facilitate measurement and recording (Carmichael et al., 2001). As a result, *ex post* team performance data have to be implemented as substitutes for *ex ante* player talent.

Similar to Scully (1974), Zak et al. (1979), and many others, if we regard each football match as the unit of observation, then the production function can be defined as:

$$y_i = f(x_1, x_2, x_3, \dots, x_j) \tag{1}$$

where y, the output of either the home or away team in game i, is a function of $\boldsymbol{x}$, a vector of j numbers of *ex post* team performance parameters. Equation 1 assumes that *all* the information of factors of production are reflected on the in-game performance if $\boldsymbol{x}$ is exhaustive enough. In reality, it can be argued that the actual/realized performance should capture not only the talent of the players, but also the wisdom of the managers, the effect of morale, form, weather conditions, and other latent input variables of the football production process.[5]

In football, the number of goals scored by a team is directly related to the intensity and accuracy of the attempts they make as well as the capacity of the opponent to block and save the attempts. In the meantime, the intensity of attempts, or the creation of goal-scoring opportunities, are directly related to the team's capability of passing and the opponent's capability to interrupt the build-ups. Therefore, similar to Carmichael et al. (2001), two types of efficiency are measured in this thesis: the efficiently to convert pass into goal-scoring opportunities, and the efficiency to convert these opportunities into goals. To account for these two relationship, two equations are used, the first one describes the production of goals:

[5] See footnote 14 of Carmichael et al. (2001) for more discussion on *ex ante* talent vs. *ex post* performance.

$$\boldsymbol{GF} = f(\textit{shooting related variables}) \qquad (\boldsymbol{GF} \geq 0) \tag{2}$$

where **GF** (goals for) is the total number of goals scored by a specific team in a specific game, and a second equation to describe the production of goal-scoring opportunities:

$$\boldsymbol{Shot} = f(\textit{passing related variables}) \qquad (\boldsymbol{Shot} \geq 0) \tag{3}$$

where **Shot** is the total number of attempts made by the team.

There exist considerable scope/scale effect in any competitive team sports, especially football; they can be positive or negative. For instance, a team with good attacking section can put the opponent on defensive and indirectly reduce the pressure of its defence; while the weak link of a team can be focused and exploited by the opponent and becomes a breakpoint. The economies of scope/scale are the results of the cross-effects and complementarity between player input and externalities (negative input) from the other team (Scully, 1995). The multiplicative Cobb-Douglas specification of the production function seems to be natural fit to accommodate the effects of economies of scope/scale, however this only applies when players' individual input is separable. When the input variables are team oriented, the effects of economies of scope/scale might already be encoded in the data and a linear production function becomes the preferred option. Another disadvantage of the multiplicative form is that it narrows the selection of independent variables, the production function cease to work when variables on the input side take the value zero, where in reality zero value can be quite common in any performance data (Carmichael et al., 2000).

Therefore, this thesis adopts the linear specification of the production function:

$$y_i = c + \alpha h + \sum_{j=1}^{n} \beta_j x_j + e_i \tag{4}$$

where $\boldsymbol{h}$ is a dummy indicator of **Home Advantage**, it takes the value 1 when the team under investigation is playing at home and 0 otherwise; α and $\boldsymbol{\beta}$ are the corresponding coefficient(s) of the home advantage dummy and the performance parameters; c is a constant term; and e is a game specific error term represented by the residual of regression analyses.

Before conducting any statistical analyses, the assumption of homogeneity has to be made. It is not a far-fetched concept since the league system is designed to let teams on the similar level to compete against each other; and thus it is fair to assume that they have access to similar levels of production technology. Furthermore, even though football evolves over time just like any other industry, the changes only happen on a long horizon, it is safe to assume that within a short period of time, the 'way of playing', in terms of team organisation, style of play, and football culture are rather constant within regions.

By assuming the matches from a specific league during a specific season are homogenous, the residual term e can be used to evaluate the efficiency of teams. If each game is measured twice, one from the home team's and the other from the away team's perspective, then for this specific football match, four residuals can be recovered: for **Equation 2**, from the home team's perspective, the residual is the number of goals scored less that would be expected given the performance, it can be used as a measure of *the offensive efficiency of goal conversion*, a zero residual represents the team is of average efficiency, a positive (negative) residual indicates the team scored more (less) than expected, thus is more (less) efficient; from the away team's perspective, the (reverse) residual can be used as an indicator of *the defensive efficiency of goal conversion*, it measures the number of goals conceded less that would be expected, a positive (negative) value means less (more) efficient. Similarly, for **Equation 3**, the residual for *the offensive (defensive) efficiency of chance creation* can be obtained for that specific match. It represents the number of chances created (conceded) less that would be expected given the performance.

We can assess the seasonal performance of teams by aggregating the residuals, from which four efficiency rankings can be constructed: *the offensive efficiency of chance creation* measures the capability of a team to create goal-scoring opportunities, *the defensive efficiency of chance creation* measures the capability of a team to prevent the creation of goal-scoring chances, *the offensive efficiency of goal conversion* measures the capability of a team to convert opportunities into goals, and *the defensive efficiency of goal conversion* measures the capability of a team to stop opponent's attempts to become goals.

Figure 1

Model Structure

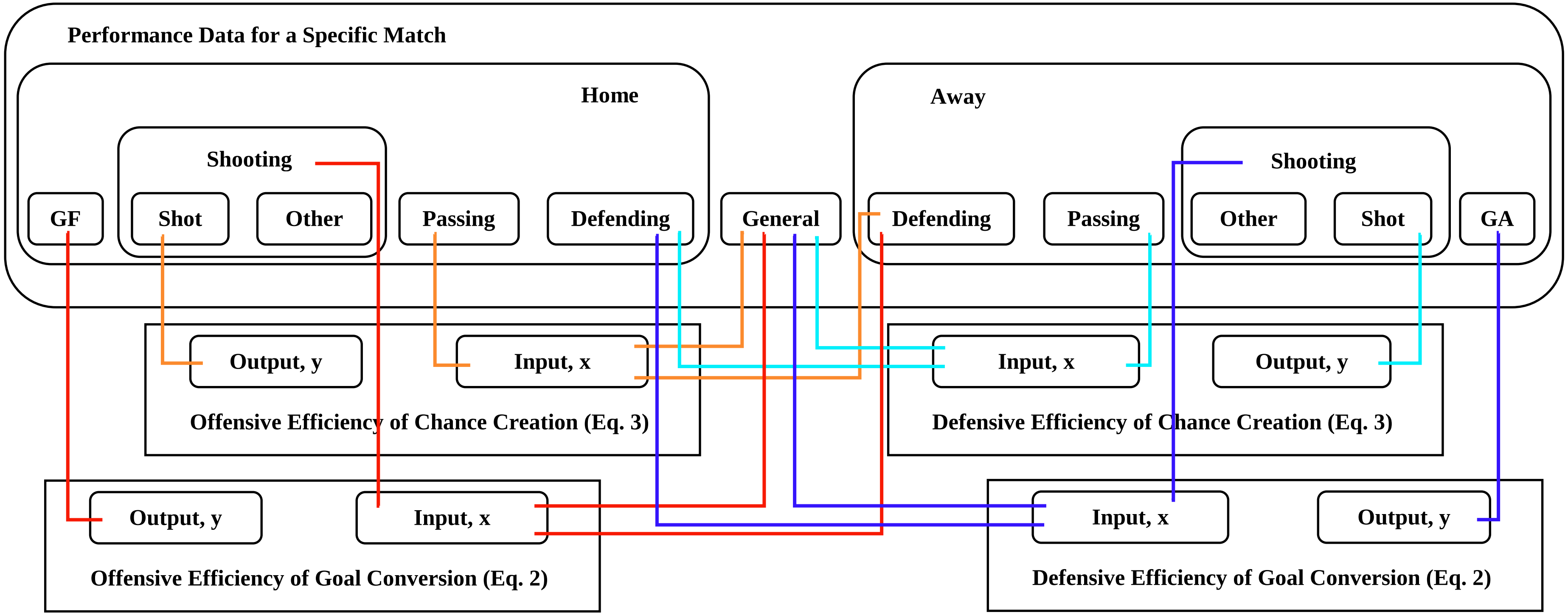

Figure 1 presents the detailed structure of the models. The **defending** and **general** variables will not be used repetitively. In other words, the ones used in **Equation 3** will not be used again in **Equation 2** (They will be insignificant anyway if included, e.g. **Home Advantage**. For more discussion, see **Section 5.1**.)

The concept of "offensive" and "defensive" are all from the (nominal) home team's perspective. The definition and description of the variables are discussed in the next subsection.

As it is displayed in **Figure 1**, the offensive efficiency are the mirror images of defensive efficiency. If each game is measured twice, where in the reverse observation the away team are regarded as the nominal home team and *vice versa*, then I just need to run the regression once (per equation) to obtain two residuals: one describes the nominal home teams' offensive efficiency, the other for their defensive efficiency. In this process, the effect of home advantage is modelled (**Home Advantage** = 0 for the reverse observations).

Unfortunately, the models are unable to produce an overall efficiency for each team. There are two main reasons. First of all, since the efficiency of chance creation and goal conversion have different units (number of attempts vs. number of goals), I cannot simply sum up the two values. In order to obtain the overall efficiency, I have to use the fitted value from **Equation 3** as an input of **Equation 2**. Such a strategy is invalid since the total number of chance created (**Shot**) will directly influence other shooting-related input variables (e.g. shots on target, long shots, etc.) and it is impossible to modify them correctly[6]. On the other hand, it is mathematically possible to sum up the offensive and defensive efficiency from the same category. However, since one of the important features of these models is to separate the offensive and defensive departments, taking an overall efficiency would be a step backward. As we will see in **Section 5.2** and **6**, most of the implication of the efficiency rankings depend on the identification of the (in)efficient departments. Thus an overall efficiency would have little real-life relevance on these fronts. More details on the construction of efficiency rankings will be discussed in **Section 5.2**.

In addition to the rankings within a league, the results from regressions can facilitate cross-league comparisons. The coefficient parameters β_j recovered from the regressions indicate the relative importance of different technical aspects of the game (x_j), by comparing the β of different leagues, we could identify the difference in the playing styles and football cultures.

3.2 Independent Variables

The various types of performance data provide limited information standing alone, some of them are highly correlated and thus must be analysed together on the team level. The identification and construction of the performance parameters x_j should be treated with great care. A football match is always a head-to-head contest between two teams, both sides are trying to score as many goals as possible and to concede as few as possible in the same time, therefore, it makes more sense to apply relative performance – the ratios and differences of the stats between the two teams rather than the

[6] As we can see, the independent variables are not exactly independent. The consequences of high correlation and multicollinearity will be discussed in **Appendix II**.

absolute data. This strategy is widely used in the past literatures. For instance, all the 20 independent variables used by Carmichael et al. (2000) are differences: difference in shots on/off target, difference in number of tackles made, etc. The ratio/difference of the same stats between the two sides appears to be a good representation of the relative strength of the sides on that specific area (e.g. the ratio of procession can be interpreted as the difference in the capability of controlling the ball), however, for some variables, ratio/difference does *not* capture the essence of the real production process. The reason is that in a football contest, a team's capability to attack is related to the other team's capability to defend (rather than to attack). It would make more sense to associate a team's attacking performance with the opponent's defending performance, and therefore one could argue that the ratio/difference of a 'pure' attacking/defending parameter such as 'the difference of tackles' doesn't have any practical meaning.

My solution to this problem is to separate a team's attack and defence by using one-sided outputs. Under this setting, a single football match produces two observations, the production process of the two teams are measured separately. In this way I can avoid misinterpretation by leaving out the variables that are obviously irrelevant. For instance, a team's tackling capability would have little influence on the team's capability to create chances or to score goals (but would be highly influential on their opponent's). The input variables of this one-sided model should include: pure attacking parameters of the team; pure defensive parameters of the opponent; and relative parameters of general play such as the ratio/difference of possession, total passes, fouls, etc.

A wide range of variables are selected as the ingredients for constructing the performance parameters x_j. These variables are selected under the principle of being exhaustive while avoiding redundancy, in the meantime, the choices are also influenced by data availability and results from past research. There are three types of variables under consideration: attacking, defending, and general, as it is presented in **Table 2**, the variables are further classified into six categories according to their affiliation to either passing or shooting.

Table 2

Candidates of Input (Independent) Variables

Variable	Description	Home		Away	
		G	S	G	S
General					
Home Advantage	Dummy, 1 if team playing at home	+	+	-	-
Possession	The percent possession of a team		+		-
Possession Ratio	**Possession** divided by **Possession** of the opponent		+		-
Pass Ratio	**Pass** divided by **Pass** of the opponent		+		-
Dual	The number of areal duals won	+	+	-	-
Dual Ratio	**Dual** divided by the total number of areal duals between to two sides	+	+	-	-
Foul	The number of fouls committed		-		+
Foul Ratio	The number of fouls committed divided by the number of fouls suffered		-		+
Shooting					
Shot	Total number of shots	+			
Shot On Target	Number of shots on target	+			
Shot Accuracy	**Shot On Target** divided by **Shot**	+			
Shot Off Target	Number of shots off target	-			
Shot Block	Number of on-target shots that are blocked	-			
Shot In-box	Number of shots inside the penalty area	+			
Goal In-box	Number of goals scored inside the penalty area	+			
Shot Out-Box	Number of shots outside the penalty area	+			
Goal Out-Box	Number of goals scored outside the penalty area	+			
Shot Set Play	Number of attempts made from set plays	+			
Goal Set Play	Number of goals scored from set plays	+			
Penalty	Number of penalties scored	+			
Penalty Attempt	Number of penalties rewarded	+			
Penalty Convert Rate	**Penalty** divided by **Penalty Attempt**	+			

Table 2 (*continued*)

Variable	Description	Home G	Home S	Away G	Away S
Passing					
Pass	Total number of successful passes		+		
Pass Attempt	Total number of attempted passes		+		
Pass Accuracy	**Pass** divided by **Pass Attempt**		+		
Forward Pass	Number of successful forward passes		+		
Forward Pass Attempt	Number of attempted forward passes		+		
Forward Pass Accuracy	**Forward Pass** divided by **Forward Pass Attempt**		+		
Forward Pass Ratio	**Forward Pass** divided by **Pass**		+		
Backward Pass	Number of successful backward passes		+/-		
Backward Pass Attempt	Number of attempted backward passes		+/-		
Square Pass	Number of successful square passes		+/-		
Square Pass Attempt	Number of attempted square passes		+/-		
Long Pass	Number of successful long passes		+		
Long Pass Attempt	Number of attempted long passes		+/-		
Short Pass	Number of successful short passes		+		
Short Pass Attempt	Number of attempted short passes		+/-		
Attacking Pass	Number of successful passes in the final third		+		
Attacking Pass Attempt	Number of attempted passes in the final third		+		
Middle Pass	Number of successful passes in the middle third		+/-		
Middle Pass Attempt	Number of attempted passes in the middle third		+/-		
Defensive Pass	Number of successful passes in the defensive third		+/-		
Defensive Pass Attempt	Number of attempted passes in the defensive third		+/-		
Other Attacking					
Cross	Number of successful crosses		+		
Cross Attempt	Number of attempted crosses		+		
Cross Accuracy	**Cross** divided by **Cross Attempt**		+		
Corner	Number of successful corners (short corners excluded)		+		

Table 2 (*continued*)

Variable	Description	Home G	Home S	Away G	Away S
Corner Attempt	Total number of corners rewarded		+		
Corner Accuracy	**Corner** divided by **Corner Attempt**		+		
Take-on	Number of successful take-ons		+		
Take-on Attempt	Number of attempted take-ons		+		
Take-on Accuracy	Success rate of take-ons, **Take-on** divided by **Take-on Attempt**		+		
Defending					
Tackle	Number of successful tackles				-
Tackle Attempt	Number of attempted tackles				-
Tackle Accuracy	Success rate of tackles				-
Clearance	Number of successful clearances				-
Clearance Attempt	Number of attempted clearances				-
Clearance Accuracy	Success rate of clearances				-
Interception	Number of interceptions				-
Break	Total number of attack break-ups, the sum of **Tackle**, **Clearance**, and **Interception**				-
Own Goal	Number of own goals			+	
Goalkeeping					
Save	Number of shots on target that are saved by the goalkeeper			-	
Save Rate	**Save** divided by **Shot On Target** of the opponent			-	
Goalkeeper Distribution	Number of successful goalkeeper passes		+		
Goalkeeper Distribution Attempt	Number of attempted goalkeeper passes		+		
Goalkeeper Distribution Accuracy	Pass accuracy of the goalkeeper		+		

Table 2 presents an incomplete list of the candidates of the input variables. The accuracy and ratio of various types of passing variables are not listed due to the limitation of space, they are constructed similar to **Forward Accuracy** and **Forward Ratio**.
The last 4 columns present the expected effect of the variables on the output of home or away side (from the home team's perspective), G represents the effect on goals scored, S represents the effect on chances created (**Shot**); + (-) indicates a positive (negative) effect, +/- indicates the effect is ambiguous.

The first category consists of the indicators of overall performance, they are relevant to both attack and defence. **Possession**, **Possession Ratio**, and **Pass Ratio** measure a team's overall control of the game, they are expected to have positive (negative) effects on the team's (opponent's) capacity to create goal-scoring opportunities. **Dual** and **Dual Ratio** measure the physicality (height, jumping, strength) of the team, they too are expected to have positive (negative) effects on the team's (opponent's) capacity to create (deter) goals or chances. **Foul** and **Foul Ratio** are discipline indicators, they are expected to have negative (positive) effects on the team's (opponent's) output.

Shooting is probably the most important attacking skillset. It is obvious that the number of goals scored are directly related to the intensity and accuracy of shooting. Therefore, we expect the effect of most of the parameters in this category to have a positive effect on the number of goals scored. The exceptions are **Shot Off Target** and **Shot Block,** they either measure the lack of shooting skills or strong defensive display from the opponent, thus they are expected to have a negative effect on **GF**.

When we consider a team's capacity to attack, passing skills are as much, if not more important than shooting. Passing is essential for the creation of goal-scoring opportunities. The passing parameters measure not only players' technical capability to move the ball from one area of the pitch to another, but also the players' vision, reaction, and reading of the game. In general, it is safe to expect that more passes and better passing accuracy would lead to more chances created. However, since fast paced counter attacks are deeply embedded in modern football, it is sometimes dangerous for a team to 'pass without a purpose'. Such situations include passing backward, passing between defenders, booming the ball forward with no apparent target, etc. Therefore, the expectation of the effect of backward passes, long passes, passes in the defensive third of the pitch can be ambiguous.

Other attacking parameters include the capability to complete crosses, corners, and take-ons. Crosses and corners are directly related to the creation of goal scoring opportunities; they are expected to have a positive effect on the number of chance created. A successful take-on describes the situation when a player (attacker) beats another player (defender) one-on-one via either skill or pace, the **Take-on** parameters are also expected to have a positive effect on the **Shot**.

Defensive parameters have little influence on a team's capability to create chances or to score goals, but are highly influential on its opponent's. The capability to execute tackles, clearances, and interceptions interrupt the build-ups of the opponent, they are expected to have a negative effect on the number of goal-scoring chances created. Own goals are rare events when a team put the ball into their own net, they can be interpreted as defensive mistakes or 'bad luck', either way, they are expected to have a one-to-one relationship with the number of goals scored by the opponent.

Parameters in the last category measures the quality of the goalkeeper. Two aspects of goalkeeper quality are taken into consideration. Goalkeeper distribution parameters proxy the passing ability of the goalie, they are expected to have a positive effect on **Shot**. However, the effect should be very small since goalkeepers are far away from the scoring area and thus their involvement in the build-up of an attack is minimum. On the other hand, as the last line of defence, goalkeepers are heavily involved in defending activities and therefore the saving parameters are expected to have a significant negative effect on the number of goals conceded.

4. Sample and Data

The subjects of this study are professional football clubs competing in the Premier League and Serie A. The EPL and Serie A are respectively at the top of the English and Italian football league system. The two leagues are widely recognised as two of the best of men's professional association football leagues alongside the Spanish La Liga, German Bundesliga, and French Ligue 1 (commonly known as *the Big Five*). The competition of the EPL started in 1992 after the First Division clubs resigned from the Football League and formed a new league. Since then, it has enjoyed tremendous commercial success through sponsorships and television rights deals. It is now the richest and most-watched football league in the world (Deloitte, 2015 & 2016) and ranks third in the UEFA coefficients of leagues based on performances in European competitions over the past five seasons. Serie A hosts some of the world's most famous clubs in Juventus, Roma, Napoli, Milan, and Internazionale, it has enjoyed a meritorious history of over eighty years and is often depicted as the most tactical football league. Despite their recent dip in influence, Italian football and Serie A clubs are still regarded as an important force of European football. Serie A currently ranks fourth in the UEFA coefficients, just behind La Liga, Bundesliga and the EPL.

The Premier League and Serie A shares a similar structure in terms of number of participants and competition organisation. Under their current format, both league is contested by 20 clubs with similar relegation rules. In a typical season, each team will play an opponent twice, home and away, which yield 38 games per team and 380 games in total. The time allocation of the fixtures is also similar, for both leagues, a season starts in August and ends in May of the following year. The playing style and football culture of the two countries however, are rather distinct. The English football is renowned for its affiliation to physicality, a typical English club bears the traits of being tall, strong, and fast. The Premier League endorses a fast-paced playing style and is not shy of physical contact. The Italians however, are famous for their resilience and their organised way of defending. The Italian style of play, *Catenaccio*, has an almost bigoted passion in defending and tactics. With such a mentality, the Italian teams are usually regarded as unfavourable opponents by any other side since they are notoriously hard to beat. The resemblance in structure and distinction in playing style of the two leagues make them interesting subjects of comparative analyses. One of the objectives of this paper is to identify these differences.

The econometric analyses are based on the detailed match-play performance stats of 760 individual games that took place during the 2015/16 season in the EPL and Serie A. For each game, 50 distinctive performance variables (100 entries if differentiate between home and away), and a further 31 derived variables (58 entries if differentiate between home and away) are manually collected. The data is obtained from the online database of the famous football magazine *FourFourTwo* (StatsZone, powered by Opta), all of which are publicly accessible. The descriptive stats of the most relevant variables are presented in **Table 3**.

Table 3
Summary Statistics

	EPL				*Serie A*			
Variable	Mean	SD	Max	Min	Mean	SD	Max	Min
GF	1.49	1.26	6	0	1.47	1.22	6	0
GA	1.21	1.15	6	0	1.11	1.08	5	0
Dual (H)	15.89	6.23	44	3	13.31	5.34	33	2
Dual (A)	15.63	6.06	34	3	12.70	5.24	29	0
Foul (H)	10.07	3.41	22	2	15.13	4.35	28	6
Foul (A)	11.52	3.41	22	4	15.13	4.35	28	6
Possession (H)	51.75	10.34	75.9	25.7	50.81	11.21	76.9	19.7
Possession (A)	48.25	10.34	74.3	24.1	49.19	11.21	80.3	23.1
Shot (H)	14.21	5.45	37	1	13.94	5.46	36	2
Shot (A)	11.29	4.47	27	2	11.45	4.66	30	2
Shot On Target (H)	4.62	2.57	15	0	4.64	2.52	14	0
Shot On Target (A)	3.87	2.15	12	0	3.70	2.16	10	0
Shot Out-box (H)	5.91	2.99	19	0	6.25	3.22	20	0
Shot Out-box (A)	4.54	2.50	16	0	5.34	2.71	14	0
Pass Accuracy (H)	0.78	0.07	0.92	0.54	0.79	0.07	0.92	0.48
Pass Accuracy (A)	0.77	0.07	0.92	0.57	0.78	0.07	0.92	0.47
Short Pass (H)	332.66	107.36	711	124	331.39	118.15	726	107
Short Pass (A)	305.92	102.65	691	100	318.28	112.99	759	98
Long Pass (H)	20.26	6.18	42	8	20.28	6.73	49	6
Long Pass (A)	18.57	5.52	36	7	18.86	6.34	46	5
Forward Pass (H)	172.07	51.87	350	71	175.57	55.96	341	65
Forward Pass (A)	154.85	48.44	335	60	167.58	54.42	397	57
Attacking Pass (H)	103.02	43.81	251	35	90.39	38.39	309	19
Attacking Pass (A)	85.19	35.29	241	30	77.85	33.90	234	24
Cross (H)	5.22	3.06	20	0	5.19	3.08	17	0
Cross (A)	3.99	2.34	12	0	3.99	2.57	15	0
Tackle (H)	19.06	5.67	41	7	17.92	5.50	38	5

Table 3 (*continued*)

Variable	*EPL* Mean	SD	Max	Min	*Serie A* Mean	SD	Max	Min
Tackle (A)	19.68	5.71	37	6	17.52	5.26	34	6
Save (H)	2.61	1.75	9	0	2.54	1.78	9	0
Save (A)	3.08	2.06	13	0	3.18	2.05	10	0
			n=380				n=380	

Table 3 presents an incomplete list of the summary statistics of the output and input variables. Due to the limitation on space, only the most important variables are listed.
For each game, **GF** (**GA**), goals for (against), is the total number of goals scored by the home (away) team. See **Table 2** for definitions and descriptions of other variables. The suffixes indicate whether the variable belong to either the home (**H**) or away (**A**) team.

5. Results

5.1. Model Specification and Cross-league Comparison

In general, performance data prove to be quite successful in explaining the variations in the input variables, especially for goal conversions. The results further suggest that the process of converting opportunities into goals (**Table 5**) are quite similar between the Premier League and Serie A while the process of creating these chances (**Table 4**) are rather distinct. However, due to the presence of multicollinearity[7], the interpretations of individual parameters are limited, thus the details of their differences cannot be thoroughly identified.

Unlike the majority of the past research on production efficiency of football that suffer from the lack of predictors to choose from, this thesis is confronted with the opposite challenge. There is a long list of potential input variables and these variables can be highly correlated (**Table 6**). Since the performance variables are ultimately related to the quality of the players, a confounding factor that is unobservable, it is inevitable that some levels of spurious relationship exist in the pool of input variables; thus makes it difficult to identify the best specifications of the models.

The first relationship I try to identify is the creation of goal-scoring opportunities through passing (**Equation 3**). Since the main objective is to produce the most reliable efficiency rankings, my top priory is to improve the explanatory power of the models as much as possible while maintaining the

[7] The presence of multicollinearity is presented in **Table 6** (correlation matrix) and **Table 7** (VIF test) in **Appendix II**.

significancy of individual parameters; reducing multicollinearity and misspecification is of lessor concern. In the process of filtering the masses amount of possible combinations of input variables, I found it is almost impossible to fit the two samples in the same model without losing a significant portion of explanatory power[8], a clear indication that the mechanics of producing goal-scoring opportunities are rather distinct between the two leagues. **Table 4** reports the descriptive statistics of best fitted model for either leagues and a comparative model.

Table 4
Descriptive Statistics of Chance Creation Models

	Model 1	Model 2	Model 3	
	EPL	*Serie A*	*EPL*	*Serie A*
Constant	-7.8853	-12.5074	-1.4623	-13.9935
	(-2.8599)***	(-5.3385)***	(-0.3183)	(-3.0289)***
Home Advantage	1.0182	0.705	0.9706	0.6652
	(3.8807)***	(2.8321)***	(3.6868)***	(2.6353)***
Possession (H)	0.1772	0.0753	0.1671	0.0809
	(5.8652)***	(2.7592)***	(5.2767)***	(2.7409)***
Dual Ratio (H)	2.4144	2.1066	2.5571	2.0078
	(2.0602)**	(2.0125)**	(2.1792)**	(1.8959)*
Foul (H)	-0.0729	-0.0374	-0.0671	-0.0383
	(-1.6313)	(-1.2498)	(-1.4970)	(-1.2748)
Foul Ratio (H)	0.6402		0.596	0
	(2.6204)***		(2.4301)**	-
Long Pass (H)		-0.113	0.0767	-0.1477
		(-2.5260)**	(-1.1208)	(-2.5272)**
Long Pass Accuracy (H)	9.0044		8.8187	2.9859
	(4.6401)***		(3.1559)***	(-1.288)
Long Pass Ratio (H)	-48.1702	31.3387	-68.5533	31.5389
	(-4.3250)***	(2.0886)**	(-3.4368)***	(1.9038)*
Short Pass (H)	0.0742	0.027	0.0794	0.0337
	(6.0841)***	(5.2498)***	(5.5608)***	(2.1633)**
Short Pass Attempt (H)	-0.0984		-0.1011	-0.0091
	(-6.8641)***		(-4.9156)***	(-0.4598)
Forward Pass Attempt (H)	0.0476		0.0588	0.0019
	(4.5963)***		(2.9785)***	(-0.1066)
Forward Pass Ratio (H)		10.9066	-7.2875	9.0013
		(3.0959)***	(-1.0282)	(-1.2411)
Backward Pass Attempt (H)	-0.0208		-0.0312	-0.0055
	(-1.8730)*		(-2.5364)**	(-0.4395)
Attacking Pass Accuracy (H)		6.3059	-0.3216	3.8331
		(2.5675)***	(-0.0937)	(-1.2197)

[8] There are several variations which could fit both samples with all independent variables significant at 5%. They are reduced from Model 3 and the coefficients of individual variables are similar to Model 3. E.g. a combination of **Home Advantage**, **Possession**, **Short Pass**, **Attacking Pass Ratio**, **Cross Accuracy**, **Shot Set Play**, **Tackle**, **Tackle Attempt**. Due to the limitation of space, these results (amongst many others) will not be reported. In such cases, the adjusted R^2 are no higher than 0.55 (EPL) and 0.59 (Serie A).

Table 4 (*continued*)

	Model 1	Model 2	Model 3	
	EPL	*Serie A*	*EPL*	*Serie A*
Attacking Pass Ratio (H)	21.2547	15.6756	19.4752	18.3119
	(9.4493)***	(5.0117)***	(5.1048)***	(4.9639)***
Middle Pass Attempt (H)		-0.0191	-0.0119	-0.0152
		(-3.0198)***	(-1.4141)	(-2.1941)**
Defensive Pass Accuracy (H)	4.7801		3.3146	3.6362
	(1.8435)*		(-1.2414)	(-1.259)
Cross Attempt (H)	-0.0845		-0.0841	-0.0042
	(-3.7284)**		(-3.7101)***	(-0.1894)
Cross Accuracy (H)	4.9734	5.4182	5.1039	5.3521
	(4.0853)***	(4.5100)***	(4.1583)***	(4.4420)***
Corner Attempt (H)	0.1783	0.158	0.171	0.1601
	(2.9113)***	(2.6735)***	(2.7820)***	(2.3969)**
Shot Set Play (H)	0.9416	0.7157	0.9477	0.699
	(12.1384)***	(8.6295)***	(12.2206)***	(8.3243)***
Take-on Accuracy (H)	2.265	-1.0276	2.4844	-1.2721
	(1.8795)*	(-0.9512)	(2.0433)**	(-1.1680)
Tackle (A)	-0.1112	-0.216	-0.1034	-0.228
	(-2.0129)**	(-4.0878)***	(-1.8684)*	(-4.2725)***
Tackle Attempt (A)	0.079	0.1374	0.0686	0.1441
	(2.0480)**	(3.7314)***	(1.7578)*	(3.8641)***
Residual diagnostics				
R2	0.6159	0.6205	0.619	0.6232
Adj R2	0.6055	0.6118	0.6065	0.6114
N	760	760	760	760

*Significant at the 10% level. **Significant at the 5% level. ***Significant at the 1% level.

Table 4 reports the results of the creation of goal-scoring chances (**Equation 3**). **Model 1** and **3** are applied to the EPL sample, while **Model 2** and **3** are applied to Serie A sample.
Since each game is measured twice (**N** = 760), in the reverse observation (**Home Advantage** = 0), the actual home team is regarded as the nominal away team and *vice versa*. The independent variable is **Shot (H)**, the total number of attempts made by the nominal home team. The t-statistics are reported in parentheses and the number of stars indicates their significancy.

In general, a large proportion of the variation in the creation of goal-scoring opportunities can be explained by the variation of general, offensive (passing, crossing, set-pieces, etc.), and defensive (mainly tackling) variables. The explanatory power is very similar between the two samples. Apart from a few **Attempt** variables, the signs of the parameters accord with the predictions.

In terms of general performance, the home advantage, possession, and ratio of aerial duals won have a positive effect on the number of chance created, while the number of fouls committed has a negative effect. Across the two leagues, the reward of having more possession is higher in the EPL and the punishment for committing more fouls are also higher in the EPL. The marginal contribution of winning more aerial duals are larger in the EPL, a result that confirms the common belief that physicality plays an important role in English football. The reward of home advantage is also higher for the EPL, an interesting result considering the fact that the EPL enjoys a much higher attendance rate comparing to Serie A and therefore their home grounds could be more intimidating.

Due to the fact that different sets of passing variables are all fundamentally related to the players' capability of passing, there exist a considerable level of multicollinearity in this category and I will restrain myself on interpretations of the individual variables. The exceptions are the long passing related variables (**Long Pass Attempt** excluded), since the ability to perform accurate long passes is a rare trait shared by only the best midfielders on the planet, their correlation with other passing variables are rather low. Oddly, the effect of long pass ratio (the successful long passes over total successful passes) is significantly positive for Serie A and significantly negative for the EPL. It seems the pay-off of playing long balls are much lower for the EPL comparing to Serie A (even if the passes are accurate). Since the success rate of long passes are on average much lower than short passes, the number of long passes is directly related to the risk appetite of the team. The distinction in long pass ratio could be a result of the differences in playing style: defending in Serie A are in general more organised, and therefore teams are forced to make risker long passes in attempting to break the defence of the opponent; where in the EPL there is no need to take such risks, and thus long passes are excessive and not favoured.

For other attacking parameters including crosses, corners, and set-play related variables, the effects are positive and similar across samples. Interestingly, the reward of beating opponent one-on-one (**Take-on Accuracy**) is only significant in the EPL sample, a result consistent with the fact that Italian football are considered to be defensively well-organised and thus beating only one defensive player has little effect on creating goal threats.

Tackling appears to be the only subset of significance amongst the category of defensive stats. It seems that the effect of interceptions and clearances are already accounted for in the effects of passing or general variables. It is also worth noticing that the marginal effect of performing a successful tackle in Serie A is roughly twice as much as a successful tackle in the EPL – another indicating that Serie A emphasis more on defending.

The goalkeeper's capability to distribute the ball has little influence on the creation of goal-scoring chances in all the models for either league. None of the effect of successful passes, attempted passes, nor the goalie's passing accuracy appears to be significant in any model.

While the process of creating goal-scoring chances are rather distinctive between the EPL and Serie A, the process of converting these chances into goals are surprisingly similar between the two. **Table 5** reports the descriptive stats of the three models based on **Equation 2**.

Table 5
Descriptive Stats of Goal Conversion Models

	Model 4	Model 5	Model 6	
	EPL	*Serie A*	*EPL*	*Serie A*
Constant	-0.1301	-0.0988	-0.1547	-0.1703
	(-4.5217)***	(-2.3336)**	(-2.6376)***	(-2.1644)**
Home Advantage		0.0578	-0.0059	0.0589
		(2.7559)***	(-0.3185)	(2.8055)***
Shot (H)	0.3852	0.1088	0.3865	0.116
	(20.9447)***	(7.0160)***	(20.6450)***	(6.8712)***
Shot On Target (H)	0.4645	0.6961	0.4648	0.6934
	(23.9752)***	(33.6409)***	(23.9437)***	(33.2903)***
Shot Off Target (H)	-0.3702	-0.0965	-0.3694	-0.0965
	(-19.4681)***	(-5.9400)***	(-19.3631)***	(-5.9405)***
Shot Block (H)	-0.3847	-0.0996	-0.3841	-0.0999
	(-20.6015)***	(-6.2860)***	(-20.5168)***	(-6.3016)***
Shot In-Box Accuracy (H)	1.0003	0.8896	0.9951	0.9
	(9.6154)***	(9.2321)***	(9.3167)***	(9.2946)***
Shot Out-Box (H)	-0.0066		-0.0109	-0.0147
	(-1.3534)		(-1.0671)	(-1.0777)
Shot Out-Box Accuracy (H)	0.6435	0.8282	0.6419	0.8352
	(6.1018)***	(7.8995)***	(6.0611)***	(7.9521)***
Shot Out-Box Ratio (H)		-0.1757	0.06	-0.0223
		(-2.6512)***	-0.4899	(-0.1422)
Shot Set Play Accuracy (H)		0.1235	0.0205	0.1252
		(2.4423)**	-0.4261	(2.4743)**
Penalty Attempt (H)	0.0459	0.0582	0.0475	0.0571
	(1.7306)*	(2.1138)**	(1.7724)*	(2.0714)**
Own Goal (A)	1.0275	1.0556	1.0292	1.0517
	(25.3924)***	(19.5398)***	(25.3207)***	(19.4249)***
Save (A)	-0.8306	-0.7826	-0.8304	-0.781
	(-50.764)***	(-47.265)***	(-50.653)***	(-46.978)***
Residual diagnostics				
R2	0.9607	0.9438	0.9608	0.9439
Adj R2	0.9602	0.9429	0.9601	0.9429
N	760	760	760	760

*Significant at the 10% level. **Significant at the 5% level. ***Significant at the 1% level.

Table 5 reports the results of the creation of goal-scoring chances (**Equation 2**). **Model 4** and **6** are applied to the EPL sample, while **Model 5** and **6** are applied to Serie A sample.
Similar to **Table 4**, each game is measured twice. The independent variable is **GF**, the number of goals scored by the nominal home team. The t-statistics are reported in parentheses and the number of stars indicates the significancy.

The three models of goal conversion share the majority of their explanatory variables, and there are no big distinctions in terms of the size of the coefficients between the two leagues. The overall performance of these models are superb in the sense that almost all variations in the number of goals scored (**GF**) can be explained by the variations of the input variables specified in the models.

The signs of the coefficients are all in line with prediction. In detail, the effect of total number of shots (**Shot (H)**, the dependent variable of the previous estimation) are positive in both leagues, however, the marginal benefit of an extra shot is on average higher in Serie A. The effect of the number of shots on target, close-range accuracy, and long-range accuracy are all positive and similar across the models. Long shots (**Shot Out-box** and **Shot Out-box Attempt**) in general negatively affects the goal production since they are rarely effective. Committing penalties and own goals are displays of defensive weakness, they all contribute positively to the output of the opponent; on the contrary, goalkeeper saves have a negative effect on opponent's output. Inaccurate shots and blocked shots both negatively affect the number of goals scored, such effects are larger in the EPL. Lastly, home advantage appears to be significant only in Serie A sample, however the effect is rather small.

Combining the results for **Table 4** and **Table 5**, I can conclude that **1**) the home advantage affects the teams mainly through the build-ups and it doesn't affect a team's goal conversion (or the effect of home advantage on shooting is already incorporated in other parameters), such effects are on average higher for the English team, **2**) the process of creating goal-scoring opportunities are distinctive between the two leagues; Serie A seems to be more defending-oriented, however due to multicollinearity, the details of the differences cannot be thoroughly identified at the current level, and **3**) the process of converting chances into goals are universal (between the EPL and Serie A).

Finding 1 confirms the presence of the effect of home advantage, it further implies that the home advantage improves the output of a team even after their performance is controlled (home advantage is significantly positive in the presence of performance variables). It seems that the atmosphere of the stadiums is indeed important, and better attendance could actually improve a team's production.

In an average football game, the whole team are usually involved in the build-up of chances but most of the final executions (shooting) are performed by the attackers. **Finding 2** and **3** imply that the criterion of good attackers is universal while the virtues of other positions (midfielders and defenders) could be regional. The findings provide a plausible explanation on why attackers are in general overvalued in transfer markets: quality forwards (and maybe goalkeepers) are widely appreciated and they can easily attract the interest of many managers across different leagues, on the contrary, managers could be hesitant in pursuing midfielders and defenders out of their own leagues since they are more likely to flop in a new environment.

The results are comparable to those from Carmichael et al. (2000 and 2001). The signs and significancy of most of the variables used in their research are confirmed by this paper. Additionally, new features of this paper, such as the inclusion of extra independent variables, and being able to analyse on per game basis, largely improves the performance of the models. In their 2001 paper, the explanatory variables only account for a small portion of the variation in goal production (seasonal data), while in this paper, almost all the variations in goals can be explained.

5.2. Efficiency Rankings

After specifying the models, the predicted values and residuals can be recovered. For each team, four types of efficiency can be evaluated by aggregating the residuals (two per model). Take Roma for example, their seasonal efficiency of chance creation is computed by aggregating the residuals from **Equation 3** (**Model 2**): the offensive efficiency of chance creation is the sum of the residuals of the 38 observations where the nominal home team is Roma, 19 observations where Roma is the real home team (**Home Advantage** = 1) plus the 19 reverse observations where Roma is the away team but regarded as the nominal home team (**Home Advantage** = 0). The defensive efficiency of chance creation is the sum of the residuals of the 38 observations where the nominal away team is Roma, 19 observations where Roma is the real away team (**Home Advantage** of Roma's opponent = 1) plus the 19 reverse observations where Roma is the home team but regarded as the nominal away team (**Home Advantage** of Roma's opponent = 0). Similarly, Roma's seasonal offensive/defensive efficiency of goal conversion can be obtained by manipulating the residuals from **Equation 2** (**Model 5**).

Table 8 presents eight ranking lists based on their sums of residuals during the 15/16 season, four for each league. The residuals of the EPL is obtained through **Model 1** (**Equation 3**) and **4** (**Equation 2**), while for Serie A I apply **Model 2** and **5**. Since different models are used on different leagues, the rankings are only comparable within their respective sample.

Table 8
Efficiency Rankings

8.1.1.1 Chance Creation, Offensive, EPL

Rank	Team	Residual	Standing
1	Tottenham	126.68	3
2	West Ham	44.68	7
3	Liverpool	38.56	8
4	Leicester City	38.08	1
5	Watford	34.58	13
6	Newcastle	14.88	18
7	Sunderland	11.57	17
8	Southampton	5.23	6
9	Swansea City	0.15	12
10	Stoke City	-0.48	9
11	Man. City	-0.78	4
12	Bournemouth	-14.46	16
13	Everton	-16.62	11
14	Crystal Palace	-19.07	15
15	Chelsea	-23.44	10
16	West Brom	-25.49	14
17	Arsenal	-28.10	2
18	Aston Villa	-48.74	20
19	Norwich	-55.75	19
20	Man United	-81.47	5

8.1.2.1 Goal Conversion, Offensive, EPL

Rank	Team	Residual	Standing
1	Leicester City	2.71	1
2	Man. City	2.67	4
3	Newcastle	0.90	18
4	Liverpool	0.75	8
5	Southampton	0.11	6
6	Swansea City	0.05	12
7	Bournemouth	-0.27	16
8	Everton	-0.28	11
9	West Ham	-0.38	7
10	Crystal Palace	-0.40	15
11	Tottenham	-0.43	3
12	Stoke City	-0.73	9
13	Aston Villa	-1.21	20
14	Watford	-1.26	13
15	Man United	-1.29	5
16	West Brom	-1.48	14
17	Chelsea	-1.76	10
18	Arsenal	-1.85	2
19	Sunderland	-2.94	17
20	Norwich	-3.36	19

8.1.1.2 Chance Creation, Defensive, EPL

Rank	Team	Residual	Standing
1	Man. City	-66.99	4
2	Norwich City	-20.85	19
3	Aston Villa	-17.80	20
4	Liverpool	-14.59	8
5	Leicester City	-12.22	1
6	Everton	-10.60	11
7	Bournemouth	-5.83	16
8	Man United	-4.86	5
9	Stoke City	-1.62	9
10	Watford	-1.12	13
11	Sunderland	0.82	17
12	Southampton	1.50	6
13	Newcastle	5.44	18
14	West Brom	6.93	14
15	West Ham	7.70	7
16	Crystal Palace	11.79	15
17	Tottenham	18.09	3
18	Arsenal	30.60	2
19	Swansea City	36.16	12
20	Chelsea	37.44	10

8.1.2.2 Goal Conversion, Defensive, EPL

Rank	Team	Residual	Standing
1	West Ham	-3.28	7
2	Man United	-2.92	5
3	Leicester City	-2.05	1
4	Liverpool	-1.69	8
5	Southampton	-1.44	6
6	West Brom	-1.43	14
7	Sunderland	-1.27	17
8	Crystal Palace	-1.27	15
9	Norwich City	-0.68	19
10	Watford	-0.53	13
11	Arsenal	-0.46	2
12	Everton	-0.20	11
13	Newcastle	0.11	18
14	Chelsea	0.44	10
15	Swansea City	0.59	12
16	Aston Villa	0.64	20
17	Tottenham	0.74	3
18	Man. City	0.88	4
19	Bournemouth	1.08	16
20	Stoke City	2.29	9

8.2.1.1 Chance Creation, Offensive, Serie A

Rank	Team	Residual	Standing
1	Napoli	60.20	2
2	Frosinone	47.57	19
3	Torino	46.07	12
4	Chievo	29.92	9
5	Empoli	28.05	10
6	Carpi	14.84	18
7	Juventus	13.71	1
8	Milan	12.41	7
9	Fiorentina	5.07	5
10	Roma	4.46	3
11	Verona	0.26	20
12	Atalanta	-0.21	13
13	Lazio	-7.27	8
14	Udinese	-19.47	17
15	Sassuolo	-20.70	6
16	Palermo	-20.86	16
17	Genoa	-36.86	11
18	Bologna	-43.84	14
19	Sampdoria	-46.19	15
20	Internazionale	-67.17	4

8.2.2.1 Goal Conversion, Offensive, Serie A

Rank	Team	Residual	Standing
1	Napoli	3.82	2
2	Roma	1.83	3
3	Chievo	1.51	9
4	Genoa	1.21	11
5	Carpi	0.40	18
6	Sassuolo	0.34	6
7	Atalanta	-0.42	13
8	Fiorentina	-0.42	5
9	Juventus	-0.62	1
10	Torino	-0.86	12
11	Udinese	-0.90	17
12	Verona	-0.93	20
13	Sampdoria	-1.04	15
14	Bologna	-1.13	14
15	Milan	-1.28	7
16	Frosinone	-1.38	19
17	Internazionale	-1.72	4
18	Lazio	-1.86	8
19	Palermo	-2.07	16
20	Empoli	-3.99	10

8.2.1.2 Chance Creation, Defensive, Serie A

Rank	Team	Residual	Standing
1	Carpi	-79.17	18
2	Fiorentina	-49.83	5
3	Juventus	-39.99	1
4	Bologna	-39.90	14
5	Atalanta	-32.10	13
6	Internazionale	-24.53	4
7	Napoli	-16.66	2
8	Torino	-12.15	12
9	Udinese	-5.14	17
10	Verona	-4.89	20
11	Sassuolo	-3.38	6
12	Milan	15.71	7
13	Sampdoria	19.52	15
14	Roma	22.74	3
15	Genoa	22.75	11
16	Empoli	31.89	10
17	Palermo	33.11	16
18	Lazio	34.33	8
19	Chievo	41.67	9
20	Frosinone	86.00	19

8.2.2.2 Goal Conversion, Defensive, Serie A

Rank	Team	Residual	Standing
1	Bologna	-2.62	14
2	Fiorentina	-2.39	5
3	Frosinone	-1.77	19
4	Internazionale	-1.76	4
5	Genoa	-1.49	11
6	Atalanta	-1.33	13
7	Juventus	-1.22	1
8	Chievo	-0.67	9
9	Carpi	-0.51	18
10	Roma	-0.31	3
11	Palermo	-0.30	16
12	Verona	-0.10	20
13	Torino	-0.06	12
14	Sampdoria	0.17	15
15	Sassuolo	0.35	6
16	Empoli	0.46	10
17	Lazio	0.46	8
18	Milan	0.50	7
19	Napoli	1.03	2
20	Udinese	2.07	17

The rankings indicate how well each team have performed given their respective on-the-field performance. For instance, in the EPL, Tottenham Hotspur on average produced 126.68 / 38 = 3.33 more chances than what we would expect from their in-game performances, which makes them the most efficient team in converting passes into chances (most creative); during the same time, Manchester United produced on average 81.47 / 38 = 2.14 less chances than their performance indicates, making them the least creative side in the league.

For most teams, the first step towards improvement is to recognise their relative strength and weakness. Efficiency rankings not only provides this information, the different types of ranking can also help managers and director to identify the source of their (in)efficiency. The four aspect of efficiency can be directly related to the performance of different departments of a football team. The offensive efficiency of goal conversion is related to strikers; the offensive efficiency of chance creation is related to midfielders, strikers and other attackers; the defensive efficiency of goal conversion is related to goalkeepers, defenders, and defensive midfielders; and the defensive efficiency of chance creation is related to midfielders and defenders.

In the short run, this information helps the match preparations of the managers. Knowing the relative strength and weakness of the opponent enables a manager to undertake precautions and counter-measures. If we further assume that it is more difficult for a team to improve on the areas they are already good at, then the most efficient way for a team to increase their long-run production is to work on the departments with the lowest relative efficiency. This can be achieved by manger changing tactics and organisation of these departments, player transfer, and managerial change. Managerial changes could have significant negative spillovers on the efficient departments and thus are usually considered to be the last resort when every department of the team fails (e.g. Chelsea). Player investment are related to the valuation of players; details on this topic will be discussed in the next section.

It is worthy to note that the offensive and defensive efficiency are not entirely symmetric with respect to zero. From a defending perspective, the effect of defence can only go as far as reducing the opponent's output to zero but not lower, therefore there exists a cap on the defensive efficiency which is equal to the output of the opponent. However, there are no such restrictions on the positive side of the axis. Consequently, the aggregate efficiency of the whole league will always be slightly smaller or equal to zero[9]. This provides yet another explanation to the overpricing of attackers. The misevaluation of different positions is justified in the sense that the contribution of a defensive player is limited while there is no upper boundary on the potential of an attacker.

[9] It only happens to the efficiency of goal conversion (**Equation 2**), since all of the fitted values of chances created (**Shot**) from **Equation 3** are above zero in both samples.

6. Player Valuation and Efficiency

6.1. Squad Value and Efficiency

Production function measures the allocative efficiency of resources of a firm; in the particular case of this thesis, it answers the question on whether a team are performing up to their standard. Of course, each team have their own standard depending on the quality of their players. A highly efficient team could produce bad results due to the lack of quality in the players. Therefore, in order to make full use of the efficiency rankings, the quality of the players should also be taken into consideration. The performance statistics used in the preceding sections have proven to be useful approximations of player quality; however, they do not provide a summarised single-number indicator for the overall quality of one player/team. An alternative can be found in prior research of Peeters (2014) and Wang (2015), where their results suggest the aggregate transfer market valuation of players is a good proxy for team quality[10]. **Table 9** presents the total market values of each club at the start and end of the 2015/16 season.

[10] The term "market value of players" specifically refers to those values obtained from *transfermarkt.de*, a public internet forum where the valuations of individual football players are assessed via aggregating the opinions of thousands of registered users. These valuations can be regarded as products of 'wisdom of the crowd' and they are proved to be quite accurate. Models using the squad valuation as the sole predictor for match result can comfortably beat the bookmakers' odds. For more discussions on player valuation and wisdom of the crowd see Peeters (2014) and Wang (2015).

Table 9
Total Market Value of Players

Premier League					*Serie A*				
R	Team	End	Start	S	R	Team	End	Start	S
1	Man. City	501.75	480.85	4	1	Juventus	397.3	388.1	1
2	Chelsea	495.75	531.75	10	2	Napoli	303	260.75	2
3	Arsenal	440	402	2	3	Roma	264.3	258.6	3
4	Man. United	418.25	377.25	5	4	Internazionale	248.3	238.98	4
5	Liverpool	365.35	330.25	8	5	Milan	187.1	203.75	7
6	Tottenham	312.5	283	3	6	Lazio	175.9	181.05	8
7	Everton	234	194.3	11	7	Fiorentina	164.6	131.7	5
8	Southampton	206.75	185.75	6	8	Sampdoria	99.3	100.93	15
9	Newcastle	179	157.5	18	9	Sassuolo	88.1	75.9	6
10	West Ham	172	165	7	10	Genoa	86.35	88.78	11
11	Stoke City	136.25	116.25	9	11	Torino	83	70.35	12
12	Leicester City	127	89.8	1	12	Udinese	77.6	76.75	17
13	Crystal Palace	125.75	117	15	13	Bologna	70.9	63.85	14
14	Swansea City	115.08	129.25	12	14	Empoli	65.48	43.75	10
15	Watford FC	106.5	82.5	13	15	Atalanta	62.2	57.45	13
16	Aston Villa	105.25	105.55	20	16	Palermo	47.4	54	16
17	West Brom	101	93.5	14	17	Verona	39.85	41.5	20
18	Norwich City	92	72.65	19	18	Chievo	36.65	34.4	9
19	Bournemouth	88.8	52.5	16	19	Carpi	36.3	42.4	18
20	Sunderland	85.75	113.25	17	20	Frosinone	31.8	28.1	19

Source: TransferMarkt.de

Table 9 presents the total squad value of the teams at the start and end of the season. In order to avoid the interference of personnel changes, the starting value is not measured at the very start of the season (Match Day 1) but at the end of the summer transfer window (Match Day 5 for the EPL, 3 for Serie A).
The unit is in Mill. €. The teams are ranked (R) according to their end-season values. The final league standings (S) are listed in the last column.

With the additional information from **Table 9**, it is now possible to extend the discussion on the implications of the efficiency rankings. In general, the final standing of a team is largely bounded to its squad value. For both leagues, none of the teams in the most valuable quartiles ended up in the bottom half of the table; the closest team to achieve that is Chelsea, the second most valuable team in EPL (1st if we rank according to the value at the start of the season) was only able to finish at tenth place. As the champions of 14/15 season, Chelsea endured the worst-ever title defence in Premier League history[11]. Their catastrophic track record is reflected on the efficiency rankings: Chelsea ended up on the lower half of all the ranking lists, two of which in the bottom quartile, and on one occasion at the very bottom. From a production point of view, it seems that each and every department of the Blues (shooting/passing

[11] The second worst is shared by Blackburn (1996) and Manchester United (2014), in both cases the title holders finished seventh.

or attacking/defending) is functioning on a suboptimal level, a clear indication that there is something wrong with the club management.

In the meantime, almost all the relegated teams (bottom three in the final standings) came from the least valuable quartile (two out of three for the EPL and three out of three for Serie A). For these teams, it is almost impossible to break into the top half of the final standings. The exception is Chievo, the third least valuable team in Serie A (second least at the start) was able to finish ninth. Accordingly, Chievo enjoys a fairly high ranking in three of the four lists.

It is impossible to comment on the 15/16 season without mentioning one of the greatest footballing miracles in the history of the sport. At the start of the season, not even their most optimistic supporters would have expected that Leicester City, which just pulled off one of the greatest escape from relegation of EPL history in the previous season, followed up by a scandalous off-season, managerial changes, and losing their best player, would eventually win the league title in such a dominate fashion. By the end of an uneventful summer transfer window, the overall squad value of the Midland club was 17th, just marginally higher than the three newly promoted sides; and not even a quarter of those title contenders' in absolute value. Not surprisingly, results from **Table 8** suggest they are highly efficient on all fronts. The Foxes are in the top quartile of all four efficiency rankings, and is the most clinical team in terms of converting chances into goals. It seems that the team possesses something extra that is unmeasurable by the performance data. In production terms, Leicester City appears to be on a higher level of production technology comparing to the rest of the league.

Among other major sides of the EPL, it is widely acknowledged that Tottenham have exceeded expectations this season by being a title challenger until the very end of the season. Unlike Leicester which relied on good defence and fast counters, Spurs adopted an attacking, possession based style and is a favourite of the critics. Their achievement is reflected on the statistics: the young and energetic Tottenham are the most creative side, their comparative advantage on this front is so much superior even comparing to the second highest ranked team (West Ham). Meanwhile, despite finish as runners-up, Arsène Wenger and his Arsenal were heavily criticised during the course of the season. The criticisms are justified by my results: Arsenal are on the bottom half of all four efficiency rankings and are 17th or lower on three occasions; implying the team were not playing even close to their potential either offensive or defensive wise. Manchester United were another team which endured a difficult season. Louis Van Gaal's side were constantly criticised for being unadventurous and boring. Despite having a lot of the ball, United struggled to create real threats and we can spot this weakness easily from efficiency ranking: they are by far the worst team in terms of converting passes into chances. The last team worth addressing from the EPL sample are Newcastle United. They had an above average team in terms of player valuation and they exhibited a reasonable level of efficiency, yet they were relegated in the end. If we assume the efficiency rankings are valid, then the only explanation would be that the perceived quality of their players are heavily overestimated.

For Serie A, the influence of squad value is much larger. The standings are very similar to the ranking of squad value[12]. At the end of the season, the top four most valuable teams occupied their respective places on the league table. Despite being overall more efficient on almost every department, the second most valuable team Napoli were not able to prevent Juventus from defending their fifth consecutive title. At the same time, at the lower end of the table, strong efficiency rankings couldn't save Carpi and Frosinone from the fate of relegation (note that the squad value of these two teams are less than 1/10 of the champions Juventus).

6.2. Value Change and Efficiency

There is little doubt that the players are the most valuable assets for professional football clubs. In modern football industry, player transfer is an important source of club revenue; and thus the valuation of players is crucial for club finance and asset investment. It would be interesting to know whether efficiency affects the perceived value of players. Since my production functions are built on team data, performance of individual players cannot be separated. Currently, the hypothesis can only be tested on the team level.

For the alternative hypothesis (H_a), I suggest the aggregate relative value change of a team's players is affected by their efficiency during that period. Mathematically, I assume:

$$d_r V = c + \sum_{k=1}^{4} \beta_k CR_k + e \qquad (5)$$

where $d_r V$ is the relative value change of a squad during a certain period, $d_r V = (V_{End} - V_{Start})/V_{Start}$; CR are the cumulated residuals of that period calculated from **Model 1**, **2**, **4**, and **5**, they can be regarded as measures of efficiency (see **Section 5.2**); β are the coefficients of different types of efficiency; c is a constant term; and e is the error term.

Since the perceived value of players varies little in the short run, the time period should be long enough to capture these changes. Meanwhile, personnel changes could also affect the squad value; such interference can be avoided by leaving out the periods when transfer window is opened. To account for the issues mentioned above, two changes will be recorded for each team in the league (20*2=40 observations per league): from the end of the summer transfer window to the opening of winter transfer

[12] For Serie A, the correlation between league points and squad value is as high as 0.93 while for the EPL the correlation is 0.54. There could be several reasons behind this difference, for instance the Serie A teams might be more homogenous, or there exist a higher level of misevaluation among the EPL players. Since the topic is beyond the scope of efficiency, I will refrain the discussion on this front.

window (EPL: Match Day 5 – 16; Serie A: 3 – 16), and from the end of the winter transfer window until the end of the season (EPL: Match Day 24 – 38; Serie A: 23 – 38). The results of the OLS regression is presented in **Table 10**.

Table 10
Descriptive Stats, Value Change - Efficiency Model

	EPL	*Serie A*
Constant	0.0271	0.0303
	(3.8753)***	(4.4301)***
Offensive Efficiency of Chance Creation	-0.0004	0.0004
	(-0.9800)	-0.9919
Defensive Efficiency of Chance Creation	-0.0001	-0.0002
	(-0.2399)	(-0.5594)
Offensive Efficiency of Goal Conversion	0.0148	-0.0045
	(2.3483)**	(-0.6793)
Defensive Efficiency of Goal Conversion	0.0177	-0.0118
	(2.2119)**	(-1.9209)*
R2	0.2045	0.1159
Adj R2	0.1136	0.0149
F-value	2.2493	1.1476
Prob(F)	0.0836	0.3505
N	40	40

*Significant at the 10% level. **Significant at the 5% level. ***Significant at the 1% level.

Table 10 presents the results of the regressions of value changes against efficiency. The independent variable is the value changes during the non-transfer-window periods. The dependent variables are the cumulative residuals of the according periods. The t-statistics are reported in parentheses and the number of stars indicates their significancy. For each model, the f-value and its t-value (Prob(F)) are also reported.

Results indicate that none of the individual parameters proves to be highly significant in any of the models; the overall significancy and explanatory power of the models are also very poor[13]. In conclusion, the null hypothesis (H_0) that efficiency do not affect squad valuations cannot be rejected, at least not within the span of half season.

The result is counterintuitive since on-the-filed performance should be the only factor affecting player valuation. A plausible explanation is that the evaluation of player quality is assessed based on players' long-period historical performance, much beyond the span of half season. In economic terms, the beliefs of player quality are updated slowly, they are largely independent from short-term efficiency. Interestingly, this explanation echoes an old cliché in football, "form is temporary; class is permanent".

[13] Other combinations of efficiency (cumulative residuals) are also tested, none of them prove to have any improvements on the individual/overall significancy of the models. Due to the limitation of space, these results will not be reported.

The main implication of the result is on the player transfer market. Regarding the sales of players, it is generally unwise to sell players from the efficient departments since efficiency are not accounted for in the player valuation. If we regard efficiency as a technological advantage of a team, then the responsible departments are the source of the team's comparative advantage. The cost of losing a key player from an efficient department could be much larger than his transfer market valuation. The team not only lose an in-form individual, it could further damage the chemistry of the department and the team could lose their competitive edge. This implication is most important for teams with high level of efficiency and low squad value (e.g. Leicester). For these teams, the key is it to avoid managerial changes and the loss of important players. This is also important to teams with a huge comparative advantage on one single aspect of efficiency (e.g. Tottenham, Manchester City, Napoli, Carpi, etc.). These teams should build around their efficient areas in order to preserve team chemistry and their latent advantage.

Instead, teams should sell players from their less efficient departments. Since players in these departments are performing on a suboptimum level, personnel changes in these areas might improve chemistry and the cost of selling a player would be actually smaller than his price tag indicates. This implication could be helpful for teams struggling on one particular aspect of the game (e.g. Manchester United, Frosinone, etc.). For them, changing the composition of their inefficient departments could reduce the comparative disadvantage in the according areas. For teams with low rankings on multiple efficiencies, the cost of a massive change in the dressing room could be too high especially when the overall squad valuation is high (e.g. Chelsea, Arsenal), and thus managerial changes seems to be the best way to realise the potential of all the underperforming players.

Regarding player investment, if we assume the efficiency rankings are known (to some degree) to most managers and directors, then following earlier discussion, clubs would be unwilling to sell their player from the efficient departments unless the transfer fee are much higher than the market valuation. Since efficiency is a team-level concept depending on the chemistry between several players as well as manager talent, there is no guarantee that the individual player could duplicate his form from his current team. Therefore, for most teams (excluding the wealthiest ones which interest in only the best players in the world) it would be better pursue players from the inefficient departments from other teams. Lastly, for teams with expensive squad, good efficiency, but poor results (e.g. Newcastle, Sampdoria), the best strategy would be to sell their overvalued squad and rebuild (with or without managerial change).

Up until this point, this thesis has commentated on the seasonal performance of various teams in the Premier League and Serie A based on their efficiency rankings and squad valuation. A variety of policy implications are proposed along the way. **Table 11** summarises these implications. Note that since a team can have several characteristics, the suggestions in the third column are not mutually exclusive.

Table 11
Policy Implication

Characteristics	Team	Action
High efficiency in some areas.	Tottenham, Manchester City, West Ham; Napoli, Carpi, Fiorentina	Avoid personnel changes in the according departments.
High efficiency in all/most areas.	Leicester, Liverpool; Napoli	Avoid managerial changes and the loss of key players.
Low efficiency in some areas.	Manchester United, Stoke City; Inter, Empoli, Frosinone	Training. Tactics. Player transfers in the according departments.
Low efficiency with high squad value.	Manchester United, Arsenal, Chelsea; Lazio, Milan	Managerial change.
Low efficiency in all/most areas.	Chelsea, Arsenal, Aston Villa; Lazio	Managerial change. Rebuild.
Good efficiency, good squad, poor results.	Newcastle; Sampdoria	Rebuild.

By the time this thesis is completed, the summer transfer window of the 16/17 season is well under way and the new season is about to start. It appears that most of the wealthy clubs in the EPL have recognised their weaknesses and have strengthened their squads accordingly, and the current champions Leicester City ware able to maintain most of their title-winning line-up. It is safe to expect that the title fight of the new EPL season will be as fierce as ever. Meanwhile in Italy, the second and third largest team Napoli and Roma have both lost their most valuable player to the title-holder Juventus, it seems that the Old Lady will continue their dominance in Serie A in the foreseeable future.

7. Concluding Remarks

To maximize production by allocating the limited amount of resources in an efficient way is crucial for the success for any organisation, the football industry is no exception. This paper estimates the production functions of professional football teams in the Premier League and Serie A. In practice, four types of efficiency are measured based on their relative performance during the recent 2015/16 season. During the process, I was able to discover the similarity (goal conversion) and distinctions (chance creation) between the English and Italian football culture. The identification of the best-fitted models enables the construction of four efficiency rankings. The rankings have implication on almost every aspect of football including training arrangement, tactic preparation, team management, player investment, etc. With the additional information on player transfer market valuation (a proxy for player quality), I am able to interpret the efficiency rankings thoroughly and comment on each team's seasonal performance.

This research benefited largely from the development on data collection in the recent years. The large variety of detailed in-game data made it possible to carry out analyses on a per game basis and to separate the production of the two opposing sides. My results are comparable to those of Carmichael et al. (2000 & 2001), the inclusion of additional variables and more detailed per game analyses significantly improve the performance of the models.

There are two major limitations. First of all, despite the enrichment in available in-game data, the defensive aspect of the game is still overlooked. The lack of defensive data is difficult to overcome since the very centre of any football game is the ball and therefore the data collection is inevitably focused on the on-the-ball movements. In the meantime, the majority of defending activities are off-the-ball movements, it is difficult to measure these movements not to mention recording them. In this sense, the importance of defending may never be quantified and we might never truly understand the secret recipe of Italian football. Secondly, an important purpose of the production function is to estimate the marginal production of labour and therefore determine labour compensation. Such purpose cannot be fulfilled by my models due to the use of team-oriented data. Once again this problem is rooted to the complementarity and interactivity of football. A feasible solution to this problem is to come up with some types of (arbitrary) player quality index using their performance stats and to build the production functions around these indices. However, how to include the effect of manager talent, chemistry, form, and other team-level factors could still be problematic.

The prediction market (game forecasting, betting) could also benefit from the construction of *ex ante* player quality indices. The performance data are proven to be quite powerful in explaining the variations in game results. As far as I'm concerned, such variables are largely underused in forecasting. Most of the research on predictions still use nothing but past results as the only predictor (Wang, 2015). Since game results contains only a fraction of the information on what is actually happening on the field, their predictive power is limited no matter how advance their mathematical model is. A direction of future research is to exploit the use of the performance statistics on game-result forecasting. No matter where the future leads, one thing is clear: football analytics has passed the stage where data availability is causing the main problems, the challenge of the new era is to make sense of the monumental amount of new data which are increasing day by day.

References

Afriat, S. N. (1972). Efficiency estimation of production functions. International Economic Reviewed 13 (October), 568-98.

Aigner, D., Lovell, C.A.K., and Schmidt, P. (1977). Foundation and estimation of statistical production function models. Journal of Econometrics 5, 21–37.

Barros, C.P., and Garcia-del-Barrio, P. (2008). Efficiency measurement of the English football Premier League with a random frontier model. Economic Modelling 25, 994–1002.

Barros, C.P., and Leach, S. (2006). Performance evaluation of the English Premier League with data envelopment analysis. Applied Economics 38 (12), 1449–1458.

Barros, C.P., and Leach, S. (2007). Technical efficiency in the English Football Association Premier League. Applied Economics Letters 14 (10), 731–741.

Boscá, J. E., Liern, V., Martínez, A., and Sala, R. (2006) Increasing offensive or defensive efficiency? An analysis of Italian and Spanish football. Omega 37 (2009), 63 – 78.

Bridgewater, S. (2014) Will football always be the biggest sport in the world? The Conversation. Retrieved from: http://theconversation.com/will-football-always-be-the-biggest-sport-in-the-world-27761

Carmichael, F., and Thomas, D. (1995) Production and efficiency in team sports: an investigation of rugby league football. Applied Economics, 1995, 27, 859-869.

Carmichael, F., Thomas, D., and Ward, R. (2000). Team Performance: The case of English premiership football. Managerial and Decision Economics, Vol. 21, No. 1, 31-45.

Carmichael, F., Thomas, D., and Ward, R. (2001). Production and efficiency in association football. Journal of Sports Economics, Vol.2 No.3, 228–243.

Charnes, A., Cooper, W.W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429-444.

Coelli, T., Rao, D.S., and Battese, G.E. (1998). An introduction to efficiency and productivity analysis. Boston, Kluwer.

Cooper, W.W., Seiford, L.M., and Tone K. (1999). Data envelopment analysis. Boston, Kluwer.

Corbett, J. (2009, October 18). Bill Shankly: Life, death and football. Retrieved from http://www.theguardian.com/football/2009/oct/18/bill-shankly-liverpool-manager.

Deloitte Sports Business Group. (2015). Revolution, Annual Review of Football Finance 2015.

Deloitte Sports Business Group. (2016). Top of the table, Football Money League 2016.

Guzmán, I., and Morrow, S. (2007) Measuring efficiency and productivity in professional football teams: evidence from the English Premier League. Central European Journal of Operations Research, Volume 15, Issue 4, 309-328.

Haas, D. J. (2003a). Productive efficiency of English football teams — a data envelopment approach. Managerial and Decision Economics 24, 403–410.

Haas, D. J. (2003b). Technical efficiency in the Major League Soccer. Journal of Sport Economics 4 (3), 203–215.

Hofler, R.A., and Payne, J.E. (1996). How close to their offensive potential do national football league teams play? Applied Economics Letters, 1996, 3, 743–747.

Hofler, R.A., and Payne, J.E. (1997). Measuring efficiency in the National Basketball Association. Economics Letters 55, 293–299.

Kearns, S. (2016, April 12). Manchester United's value falls by more than £400m during Louis van Gaal's disastrous second season. Retrieved from http://metro.co.uk/2016/04/12/manchester-uniteds-value-falls-by-more-than-400m-during-louis-van-gaals-disastrous-second-season-5811845.

Kuper, S. (2014, January 17). Interview: Carlo Ancelotti. Retrieved from http://www.ft.com/cms/s/2/7e89c58e-7e45-11e3-b409-00144feabdc0.html.

Peeters, T. (2014). Testing the wisdom of crowds: Transfermarkt valuations and international football results (preliminary version). Erasmus University Rotterdam.

Richmond, J. (1974). Estimating the efficiency of production. International Economic Reviews 15 (June), 515-21.

Schofield, J.A. (1988). Production functions in the sports industry: An empirical analysis of professional cricket. Applied Economics 20(2), 177–193.

Scully, G.W. (1974). Pay and performance in Major League Baseball. American Economic Review, 64, 915-930.

Scully, G. W. (1995). The market structure of sports. Chicago: University of Chicago Press.

Timmer, C. P. (1971). Using a probabilistic frontier production function to measure technical efficiency. Journal of Political Economy 79, no. 4 (July/August), 776-94.

Wang, M. (2015). Modelling association football (soccer) results using the wisdom of crowds - a research on the impact of starting line-ups and player position. Erasmus University Rotterdam.

Zak, T.A., Huang, C.J., and Siegfried, J.J. (1979). Production efficiency: the case of professional basketball. Journal of Business 52, 379–392.

Zech, C. E. (1981). An empirical estimation of a production function: the case of major league baseball. The American Economist, Vol. 25, No. 2, 19-23.

Appendix

I. Literature Survey

Table 1

Literature Survey: Production Efficiency of Sports

Year	Author(s)	Sport (League)	Model	Outputs	Inputs
1974	Scully	Baseball (MLB)	Linear, Multi-equation	Percent wins	Performance variables on hitting/pitching, Managerial proxies
				Revenue	Percent wins, Geographic factors
1979	Zak, Huang, and Siegfried	Basketball (NBA)	Cobb-Douglas Frontier	Ratio of final scores	10 performance variables
1981	Zech	Baseball (MLB)	Cobb-Douglas	Wins	6 performance variables, 2 managerial variables
1988	Schofield	Cricket (County Championship and Sunday League)	Recursive multi-equation	Wins	Performance and weather variables
1995	Carmichael and Thomas	Rugby league football (RFL 1^{st} and 2^{nd} Div)	Linear and Cobb-Douglas	Percent wins	Performance, organization, managerial and other variables
1996	Hofler and Payne	American Football (NFL)	Linear frontier	Scoring	13 performance variables
1997	Hofler and Payne	Basketball (NBA)	Linear frontier	Wins	6 performance variables
2000	Carmichael, Thomas, and Ward	Football (EPL)	Linear	Match score	20 performance variables
2001	Carmichael, Thomas, and Ward	Football (EPL)	Linear, Multi-equation	League Points, goals scored/conceded, shots	15 performance variables
2003	Haas	Football (EPL)	DEA-CCR and DEA-BCC	Points, attendance, revenue	Player/manager wage

Table 1 (*continued*)

Year	Author(s)	Sport (League)	Model	Outputs	Inputs
2003	Haas	Football (MLS)	DEA-CCR and DEA-BCC	Points, attendance, revenue	Player/manager wage, city population
2006	Boscá, Liern, Martínez, and Sala	Football (La Liga and Serie A)	DEA	Goals	5 performance variables
2007	Guzmán and Morrow	Football (EPL)	DEA-CCA	Points	Turnover, expenses
2007	Barros and Leach	Football (EPL)	Stochastic frontier	Points, attendance	Operational cost
2008	Barros and Garcia-del-Barrio	Football (EPL)	Random frontier	Operational cost	Wages, capital-premises, investment, sales, points, attendance
2010	Espitia-Escuer and Garcia-Cebrian	Football (CL)	DEA	No. of games before elimination	Attacking plays, possession, goal attempts, no. of players

II. Multicollinearity

Table 6
Correlation Matrices

6.1.1 EPL, Passing

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**Shot**	1.00	0.47	0.41	0.19	0.38	0.15	-0.15	0.35	-0.22	0.40	0.46	0.42	0.43	0.19	0.57	0.55	0.47	0.44
2	**Possession**	0.47	1.00	0.87	0.17	0.68	0.33	-0.09	0.51	-0.48	0.86	0.87	0.89	0.69	-0.06	0.77	0.76	0.66	0.17
3	**Pass**	0.41	0.87	1.00	0.18	0.87	0.27	-0.26	0.58	-0.61	1.00	0.97	0.92	0.87	-0.20	0.82	0.77	0.78	0.08
4	**Pass Ratio**	0.19	0.17	0.18	1.00	0.10	0.07	-0.05	0.14	-0.03	0.18	0.19	0.17	0.11	0.02	0.24	0.23	0.10	0.11
5	**Pass Accuracy**	0.38	0.68	0.87	0.10	1.00	0.23	-0.37	0.62	-0.59	0.87	0.83	0.68	0.96	-0.26	0.66	0.54	0.84	-0.04
6	**Long Pass**	0.15	0.33	0.27	0.07	0.23	1.00	0.60	0.68	0.53	0.22	0.27	0.29	0.23	-0.03	0.21	0.22	0.18	0.02
7	**Long Pass Attempt**	-0.15	-0.09	-0.26	-0.05	-0.37	0.60	1.00	-0.15	0.65	-0.29	-0.27	-0.13	-0.42	-0.02	-0.26	-0.15	-0.41	-0.08
8	**Long Pass Accuracy**	0.35	0.51	0.58	0.14	0.62	0.68	-0.15	1.00	0.06	0.55	0.59	0.49	0.66	-0.02	0.53	0.45	0.60	0.12
9	**Long Pass Ratio**	-0.22	-0.48	-0.61	-0.03	-0.59	0.53	0.65	0.06	1.00	-0.65	-0.59	-0.54	-0.57	0.17	-0.47	-0.41	-0.52	0.03
10	**Short Pass**	0.40	0.86	1.00	0.18	0.87	0.22	-0.29	0.55	-0.65	1.00	0.97	0.92	0.87	-0.20	0.82	0.76	0.78	0.08
11	**Forward Pass**	0.46	0.87	0.97	0.19	0.83	0.27	-0.27	0.59	-0.59	0.97	1.00	0.96	0.87	0.03	0.84	0.79	0.76	0.14
12	**Forward Pass Attempt**	0.42	0.89	0.92	0.17	0.68	0.29	-0.13	0.49	-0.54	0.92	0.96	1.00	0.72	0.07	0.79	0.80	0.64	0.14
13	**Forward Pass Accuracy**	0.43	0.69	0.87	0.11	0.96	0.23	-0.42	0.66	-0.57	0.87	0.87	0.72	1.00	-0.05	0.71	0.60	0.85	0.06
14	**Forward Pass Ratio**	0.19	-0.06	-0.20	0.02	-0.26	-0.03	-0.02	-0.02	0.17	-0.20	0.03	0.07	-0.05	1.00	-0.01	0.05	-0.13	0.29
15	**Attacking Pass**	0.57	0.77	0.82	0.24	0.66	0.21	-0.26	0.53	-0.47	0.82	0.84	0.79	0.71	-0.01	1.00	0.97	0.80	0.60
16	**Attacking Pass Attempt**	0.55	0.76	0.77	0.23	0.54	0.22	-0.15	0.45	-0.41	0.76	0.79	0.80	0.60	0.05	0.97	1.00	0.67	0.64
17	**Attacking Pass Accuracy**	0.47	0.66	0.78	0.10	0.84	0.18	-0.41	0.60	-0.52	0.78	0.76	0.64	0.85	-0.13	0.80	0.67	1.00	0.37
18	**Attacking Pass Ratio**	0.44	0.17	0.08	0.11	-0.04	0.02	-0.08	0.12	0.03	0.08	0.14	0.14	0.06	0.29	0.60	0.64	0.37	1.00

6.1.2 EPL, Shooting

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	**GF**	1.00	0.47	0.41	0.19	0.38	0.15	-0.15	0.35	-0.22	0.40	0.46	0.42	0.43	0.19
2	**Shot**	0.47	1.00	0.87	0.17	0.68	0.33	-0.09	0.51	-0.48	0.86	0.87	0.89	0.69	-0.06
3	**Shot On Target**	0.41	0.87	1.00	0.18	0.87	0.27	-0.26	0.58	-0.61	1.00	0.97	0.92	0.87	-0.20
4	**Shot Accuracy**	0.19	0.17	0.18	1.00	0.10	0.07	-0.05	0.14	-0.03	0.18	0.19	0.17	0.11	0.02
5	**Shot Off Target**	0.38	0.68	0.87	0.10	1.00	0.23	-0.37	0.62	-0.59	0.87	0.83	0.68	0.96	-0.26
6	**Shot Block**	0.15	0.33	0.27	0.07	0.23	1.00	0.60	0.68	0.53	0.22	0.27	0.29	0.23	-0.03
7	**Shot In-box**	-0.15	-0.09	-0.26	-0.05	-0.37	0.60	1.00	-0.15	0.65	-0.29	-0.27	-0.13	-0.42	-0.02
8	**Shot Out-box**	0.35	0.51	0.58	0.14	0.62	0.68	-0.15	1.00	0.06	0.55	0.59	0.49	0.66	-0.02
9	**Penalty**	-0.22	-0.48	-0.61	-0.03	-0.59	0.53	0.65	0.06	1.00	-0.65	-0.59	-0.54	-0.57	0.17
10	**Penalty Attempt**	0.40	0.86	1.00	0.18	0.87	0.22	-0.29	0.55	-0.65	1.00	0.97	0.92	0.87	-0.20
11	**Shot Set Play**	0.46	0.87	0.97	0.19	0.83	0.27	-0.27	0.59	-0.59	0.97	1.00	0.96	0.87	0.03
12	**Save (A)**	0.42	0.89	0.92	0.17	0.68	0.29	-0.13	0.49	-0.54	0.92	0.96	1.00	0.72	0.07
13	**Save Rate (A)**	0.43	0.69	0.87	0.11	0.96	0.23	-0.42	0.66	-0.57	0.87	0.87	0.72	1.00	-0.05
14	**Own Goal (A)**	0.19	-0.06	-0.20	0.02	-0.26	-0.03	-0.02	-0.02	0.17	-0.20	0.03	0.07	-0.05	1.00

6.2.1 Serie A, Passing

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**Shot**	1.00	0.54	0.44	0.17	0.39	0.17	-0.18	0.40	-0.22	0.44	0.49	0.44	0.46	0.08	0.66	0.62	0.57	0.45
2	**Possession**	0.54	1.00	0.86	0.16	0.70	0.31	-0.11	0.51	-0.50	0.85	0.88	0.87	0.73	-0.12	0.76	0.77	0.63	0.11
3	**Pass**	0.44	0.86	1.00	0.19	0.86	0.30	-0.16	0.54	-0.58	1.00	0.97	0.92	0.86	-0.33	0.75	0.71	0.67	-0.10
4	**Pass Ratio**	0.17	0.16	0.19	1.00	0.10	0.07	-0.11	0.20	-0.07	0.18	0.19	0.16	0.13	-0.04	0.23	0.20	0.12	0.05
5	**Pass Accuracy**	0.39	0.70	0.86	0.10	1.00	0.29	-0.26	0.59	-0.56	0.85	0.82	0.69	0.97	-0.37	0.62	0.52	0.76	-0.16
6	**Long Pass**	0.17	0.31	0.30	0.07	0.29	1.00	0.65	0.72	0.52	0.25	0.32	0.32	0.30	0.02	0.29	0.27	0.28	0.06
7	**Long Pass Attempt**	-0.18	-0.11	-0.16	-0.11	-0.26	0.65	1.00	-0.03	0.63	-0.20	-0.16	-0.06	-0.29	0.04	-0.16	-0.10	-0.24	-0.03
8	**Long Pass Accuracy**	0.40	0.51	0.54	0.20	0.59	0.72	-0.03	1.00	0.12	0.51	0.58	0.48	0.65	0.00	0.54	0.47	0.58	0.13
9	**Long Pass Ratio**	-0.22	-0.50	-0.58	-0.07	-0.56	0.52	0.63	0.12	1.00	-0.62	-0.54	-0.51	-0.52	0.31	-0.38	-0.35	-0.36	0.17
10	**Short Pass**	0.44	0.85	1.00	0.18	0.85	0.25	-0.20	0.51	-0.62	1.00	0.97	0.92	0.85	-0.33	0.74	0.71	0.67	-0.11
11	**Forward Pass**	0.49	0.88	0.97	0.19	0.82	0.32	-0.16	0.58	-0.54	0.97	1.00	0.97	0.86	-0.10	0.78	0.76	0.68	-0.01
12	**Forward Pass Attempt**	0.44	0.87	0.92	0.16	0.69	0.32	-0.06	0.48	-0.51	0.92	0.97	1.00	0.72	-0.04	0.75	0.76	0.56	-0.01
13	**Forward Pass Accuracy**	0.46	0.73	0.86	0.13	0.97	0.30	-0.29	0.65	-0.52	0.85	0.86	0.72	1.00	-0.19	0.68	0.58	0.79	-0.06
14	**Forward Pass Ratio**	0.08	-0.12	-0.33	-0.04	-0.37	0.02	0.04	0.00	0.31	-0.33	-0.10	-0.04	-0.19	1.00	-0.03	0.03	-0.12	0.39
15	**Attacking Pass**	0.66	0.76	0.75	0.23	0.62	0.29	-0.16	0.54	-0.38	0.74	0.78	0.75	0.68	-0.03	1.00	0.97	0.79	0.54
16	**Attacking Pass Attempt**	0.62	0.77	0.71	0.20	0.52	0.27	-0.10	0.47	-0.35	0.71	0.76	0.76	0.58	0.03	0.97	1.00	0.66	0.56
17	**Attacking Pass Accuracy**	0.57	0.63	0.67	0.12	0.76	0.28	-0.24	0.58	-0.36	0.67	0.68	0.56	0.79	-0.12	0.79	0.66	1.00	0.39
18	**Attacking Pass Ratio**	0.45	0.11	-0.10	0.05	-0.16	0.06	-0.03	0.13	0.17	-0.11	-0.01	-0.01	-0.06	0.39	0.54	0.56	0.39	1.00

6.2.2 Serie A, Shooting

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	**GF**	1.00	0.31	0.57	0.43	0.06	0.03	0.38	0.10	0.28	0.25	0.10	0.12	-0.49	0.13
2	**Shot**	0.31	1.00	0.67	-0.04	0.78	0.65	0.84	0.78	0.13	0.16	0.51	0.64	0.19	0.00
3	**Shot On Target**	0.57	0.67	1.00	0.64	0.26	0.20	0.66	0.41	0.19	0.24	0.25	0.87	0.17	-0.04
4	**Shot Accuracy**	0.43	-0.04	0.64	1.00	-0.37	-0.31	0.07	-0.15	0.12	0.16	-0.13	0.50	0.11	-0.04
5	**Shot Off Target**	0.06	0.78	0.26	-0.37	1.00	0.31	0.59	0.67	0.05	0.06	0.44	0.28	0.13	0.03
6	**Shot Block**	0.03	0.65	0.20	-0.31	0.31	1.00	0.51	0.55	0.04	0.04	0.39	0.23	0.11	0.03
7	**Shot In-box**	0.38	0.84	0.66	0.07	0.59	0.51	1.00	0.30	0.14	0.20	0.52	0.58	0.10	0.02
8	**Shot Out-box**	0.10	0.78	0.41	-0.15	0.67	0.55	0.30	1.00	0.06	0.05	0.30	0.45	0.21	-0.01
9	**Penalty**	0.28	0.13	0.19	0.12	0.05	0.04	0.14	0.06	1.00	0.87	0.00	0.07	-0.13	-0.01
10	**Penalty Attempt**	0.25	0.16	0.24	0.16	0.06	0.04	0.20	0.05	0.87	1.00	0.04	0.16	-0.07	-0.03
11	**Shot Set Play**	0.10	0.51	0.25	-0.13	0.44	0.39	0.52	0.30	0.00	0.04	1.00	0.24	0.04	-0.04
12	**Save (A)**	0.12	0.64	0.87	0.50	0.28	0.23	0.58	0.45	0.07	0.16	0.24	1.00	0.53	-0.02
13	**Save Rate (A)**	-0.49	0.19	0.17	0.11	0.13	0.11	0.10	0.21	-0.13	-0.07	0.04	0.53	1.00	0.04
14	**Own Goal (A)**	0.13	0.00	-0.04	-0.04	0.03	0.03	0.02	-0.01	-0.01	-0.03	-0.04	-0.02	0.04	1.00

Table 7
VIF Tests

7.1 Chance Creation Models (Equation 3)

	Model 1 (EPL)		Model 2 (Serie A)		Model 3 (EPL)		Model 3 (Serie A)	
	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
Home Advantage	1.23	0.8130	1.11	0.9009	1.24	0.8065	1.14	0.8772
Possession (H)	7.16	0.1397	6.74	0.1484	7.90	0.1266	7.87	0.1271
Dual Ratio (H)	1.10	0.9091	1.13	0.8850	1.11	0.9009	1.15	0.8696
Foul (H)	1.73	0.5780	1.21	0.8264	1.75	0.5714	1.22	0.8197
Foul Ratio (H)	1.75	0.5714			1.77	0.5650		
Long Pass (H)			6.19	0.1616	11.74	0.0852	10.55	0.0948
Long Pass Accuracy (H)	3.25	0.3077			6.77	0.1477	5.65	0.1770
Long Pass Ratio (H)	4.42	0.2262	9.26	0.1080	14.23	0.0703	11.28	0.0887
Short Pass (H)	118.90	0.0084	25.34	0.0395	163.43	0.0061	233.38	0.0043
Short Pass Attempt (H)	171.46	0.0058			353.53	0.0028	393.65	0.0025
Forward Pass Attempt (H)	18.36	0.0545			67.06	0.0149	69.63	0.0144
Forward Pass Ratio (H)			1.54	0.6494	4.99	0.2004	6.52	0.1534
Backward Pass Attempt (H)	10.44	0.0958			12.89	0.0776	18.95	0.0528
Attacking Pass Accuracy (H)			3.61	0.2770	7.26	0.1377	5.90	0.1695
Attacking Pass Ratio (H)	1.52	0.6579	3.07	0.3257	4.37	0.2288	4.27	0.2342
Middle Pass Attempt (H)			15.72	0.0636	20.18	0.0496	18.86	0.0530
Defensive Pass Accuracy (H)	1.50	0.6667			1.60	0.6250	1.94	0.5155
Cross Attempt (H)	2.34	0.4274			2.60	0.3846	2.90	0.3448
Cross Accuracy (H)	1.19	0.8403	1.13	0.8850	1.21	0.8264	1.14	0.8772
Corner Attempt (H)	2.34	0.4274	2.02	0.4950	2.37	0.4219	2.58	0.3876
Shot Set Play (H)	1.65	0.6061	1.68	0.5952	1.65	0.6061	1.72	0.5814
Take-on Accuracy (H)	1.76	0.5682	1.87	0.5348	1.79	0.5587	1.90	0.5263
Tackle (A)	7.08	0.1412	5.80	0.1724	7.13	0.1403	5.91	0.1692
Tackle Attempt (A)	6.83	0.1464	5.83	0.1715	7.01	0.1427	5.97	0.1675
Mean VIF	18.30		5.49		29.40		35.39	

7.2 Goal Conversion Models (Equation 2)

	Model 4 (EPL)		Model 5 (Serie A)		Model 6 (EPL)		Model 6 (Serie A)	
	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
Home Advantage			1.08	0.9259	1.11	0.9009	1.08	0.9259
Shot (H)	118.57	0.0084	64.31	0.0155	122.47	0.0082	76.22	0.0131
Shot On Target (H)	28.14	0.0355	24.04	0.0416	28.15	0.0355	24.37	0.0410
Shot Off Target (H)	32.43	0.0308	19.94	0.0502	32.55	0.0307	19.94	0.0502
Shot Block (H)	26.03	0.0384	11.31	0.0884	26.08	0.0383	11.31	0.0884
Shot In-Box Accuracy (H)	3.48	0.2874	2.54	0.3937	3.66	0.2732	2.57	0.3891
Shot Out-Box (H)	2.49	0.4016	1.14	0.8772	10.91	0.0917	16.58	0.0603
Shot Out-Box Accuracy (H)	1.46	0.6849	1.35	0.7407	4.93	0.2028	6.42	0.1558
Shot Out-Box Ratio (H)					1.47	0.6803	1.36	0.7353
Shot Set Play Accuracy (H)			1.13	0.8850	1.19	0.8403	1.13	0.8850
Penalty Attempt (H)			1.10	0.9091	1.12	0.8929	1.10	0.9091
Own Goal (A)	1.01	0.9901	1.01	0.9901	1.02	0.9804	1.02	0.9804
Save (A)	12.92	0.0774	10.16	0.0984	12.92	0.0774	10.24	0.0977
Mean VIF	25.17		11.59		19.04		13.33	

The correlation matrices and VIF tests show considerable level of multicollinearity amongst the (candidate) input variables, especially for variables from the passing and shooting categories. Since the main objective of this paper is to construct the most reliable rankings and multicollinearity does not affect the overall performance of the models, its presence is not a big concern. As a consequence, there might exist misinterpretations in the individual variables. Therefore, the discussions on the cross-league comparisons (the passing/shooting related variables) in **Section 5.1** could be inaccurate and are subject to debate.