Working Paper in Economics

202224

Managerial Contribution to Firm Success: Evidence from Professional Football Leagues

Kaori Narita
Benjamin Holmes
Ian McHale

Managerial Contribution to Firm Success: Evidence from Professional Football Leagues

Kaori Narita* Benjamin Holmes† Ian G. McHale‡

November 12, 2022

Abstract

Previous research in economics and management finds significant heterogeneity across senior managers in their contribution to organisational success. It remains, however, challenging to disentangle the impact of managers from that of other inputs, such as labour, due to limited data availability in many general organisations. In contrast, individual workers (players)' performance and their characteristics are publicly available in professional football leagues. Therefore, this study employs data from the industry to estimate individual managers' contributions to field performance, given the resources at hand. To measure a club's output, we adopt expected goals, which are less influenced by randomness than conventional measures. Controlling for players' quality based on their historical performance as well as a club's financial strengths, we yet find significant impacts of managerial inputs. In addition, we compare our estimated manager coefficients with a more naive measure of managerial performance, such as winning percentage. This highlights the importance of taking into account the differences in resources that a manager has at his disposal as well as the randomness of the outcome when evaluating a manager's performance.

Keywords: firm performance, managers, football, performance evaluation, leadership

JEL Codes: M54, Z22, L25

*Corresponding author. Email: K.Narita@liverpool.ac.uk. Address: University of Liverpool Management School, Liverpool, L69 7ZH, United Kingdom.

†Email: sgbholme@liverpool.ac.uk. Address: The University of, Mathematical Sciences Bldg, Liverpool L69 7ZL, University of Liverpool.

‡Email: ian.mchale@liverpool.ac.uk. Address: University of Liverpool Management School, Liverpool, L69 7ZH, United Kingdom.

1 Introduction

The relationship between a firm owner and managers is often described in the principal-agent model, where a firm owner delegates day-to-day management tasks to managers, who are typically more specialised in such tasks and have superior knowledge in the respective industry. Based on the evidence in previous management and economics research (Hendricks et al., 2015; Siebert and Zubanov, 2010; Bloom et al., 2014, 2013), there is probably not much room for debate on the importance of managerial inputs in firm productivity. However, the nature of managerial tasks, such as monitoring and motivating workers, makes the evaluation of managers rather challenging since these are not easily quantified. A hidden information problem is also present in the sense that it is not easy to disentangle the contributions of different inputs. Some managers may seem to be more competent than others, however, this may as well be the case that they are fortunate to have high-quality workers.

This study aims to quantify the managerial contribution to a firm's success in the context of professional football. A sports club is not any different from firms operating in other industries in that it employs managers, labour, and capital to produce output. Frick and Simmons (2008) draw an analogy between professional football managers and firm executives:

> Soccer head coaches have roles that resemble that of CEOs. They propose hiring and firing decisions to the board of directors (most often through a Director of Football), and they impose team playing strategies and make tactical adjustments within games. Head coaches have important motivational roles in trying to raise individual players and team performance. (Frick and Simmons, 2008, p.594)

In addition, top managers in a corporate setting and football managers share similar characteristics such as age, accountability, experience, and resilience to pressure (Pieper et al., 2014).

However, a great deal of publicity at an individual employee level, be it player or manager, is perhaps unique to the professional sports industry. The huge public interest in the industry means the characteristics of individual managers and related events, for instance, managerial turnover, are highly publicised through media. A club's output or performance is clearly and regularly measured. This makes the industry an attractive place to study issues in management and economics. As noted by Kahn (2000), it offers a (natural) laboratory whereby researchers can explore the implications of certain policies or decisions on the outcome of interest. Furthermore, unlike laboratory settings, sports events occur in professional environments with high stakes. Therefore, the sports industry produces valuable data with a sophisticated level of quality and quantity, contributing to the field of management (Day et al., 2012; Lefgren et al., 2015).

At the same time, the conclusions drawn based on football managers are valuable in their own right.

In professional football, not only is it an employer who is interested in knowing the abilities of potential employees, but also the stakeholders, including the huge fan base worldwide. In fact, much effort has been devoted to evaluating players by the communities and researchers. For instance, the EA Sports FIFA website, known as *SoFIFA.com*[1] and *WhoScored.com*[2] offer ratings for the comprehensive sets of players created by scout communities. In academia, Decroos et al. (2020) and Kharrat et al. (2020), among others, contribute to the field by providing objective evaluation systems of players within team sports.

Contrarily, we have not seen much development in the evaluation of individual managers to date. Managers arguably attract as much attention as players; the media covers every event related to managerial change, particularly in the top-tier domestic football leagues. Therefore, advancement in the understanding of managerial contributions and the evaluation system of individual managers would be beneficial to making better-informed decisions regarding the employment of a manager.

This study builds on the previous research on managerial contributions to firm outcomes, including Bertrand and Schoar (2003), Peeters et al. (2020), Muehlheusser et al. (2018), and Buzzacchi et al. (2021), whilst overcoming some of the limitations in these studies. Firstly, the previous studies rely on the separate estimations of firm and leader fixed effects in order to disentangle the contributions of managers from firm-specific effects. However, this limits the scope of estimation of individual leaders' effects since this is only feasible for those observed in multiple firms. Therefore, instead of estimating the two sets of fixed effects, we explicitly measure other inputs (labour and capital) to control for resources available for a manager to produce output, whilst capturing individual managers' contributions with their fixed effects.

To do so, we collected individual players' historical performance and clubs' transfer budgets as proxies for labour and capital inputs available for each club. The former is based on player ratings provided by *WhoScored.com* mentioned above, which consists of more than a million data points. We propose some adjustments to the observed ratings of individual players in order to take into account the different levels of difficulty across the leagues. This makes ratings observed in different leagues comparable to each other.

These variables are also allowed to change over time; therefore, we can quantify an individual manager's contribution to the match outcome, given the quality of a squad and financial resources in a particular period of time. This allows us to control for the most up-to-date fitness of the squad and financial strengths of a club, which may not be well-captured by individual club-fixed effects as in the previous studies. We also argue that this is a valid way to evaluate a football manager since the main role of a manager is to optimise a match day outcome given these resources, especially due to the emergence of the role of Director of Football at many clubs, which has taken some control away from a manager in terms of recruitment of players.

[1] Available at https://sofifa.com/.
[2] Available at https://www.whoscored.com/.

Secondly, we extend the studies by Muehlheusser et al. (2018) and Buzzacchi et al. (2021), who provide the rankings of managers within the German and Italian football leagues, respectively, by considering multiple leagues in Europe, the "Big Five" in particular. This is a relevant extension due to the fact that the labour market for football managers is mobile, and many managers are observed across different leagues. The proposed measures of the inputs are also comparable across the leagues, making this extension feasible.

Finally, in order to take into account the randomness of match outcome, we measure a club's performance with expected goals, which capture the quality of chances created, using information related to individual shot data. As shown by recent studies such as Flepp and Franck (2021), this measure of performance is more informative than the conventional crude measure of outcome based on average points as in Muehlheusser et al. (2018) and Buzzacchi et al. (2021).

Employing the advanced measures of output and inputs, our study finds the overall significance of managerial input and significant heterogeneity among managers. Furthermore, our findings confirm that taking into account the strength of players and finance as well as uncertainties associated with the randomness of match outcome matters in the evaluation of managerial contribution.

The remainder of this paper is organised as follows. The next section provides an overview of related studies that give us some direction to achieve our objectives set out above. Section 3 describes data used in this analysis. Section 4 explains constructions of the important variables and models to be estimated. Our empirical results are presented in Section 5, followed by concluding remarks in Section 6.

2 Related literature

There are a few strands of literature that can give us some directions to achieve our objectives set out in the introduction; quantifying the managerial contribution to the firm success and the heterogeneity among managers. First, previous studies in economics and management have estimated the contributions of managerial inputs by empirically disentangling the individual "firm" and "manager" effects on production. One of the most cited studies by Bertrand and Schoar (2003) employs a CEO-firm matched data to estimate managers' fixed effects on the performance and behaviour of firms, separating them from the firms' fixed effects on the variables of interest. This approach is followed by many others, including Graham et al. (2012) and Lazear et al. (2015), to establish a relationship between the heterogeneity in managerial ability and various measures of corporate outcomes.

In the sports domain, Peeters et al. (2020), Muehlheusser et al. (2018), and Buzzacchi et al. (2021), among others, have applied a similar method to investigate the significance of managers in professional sports clubs. In particular, Muehlheusser et al. (2018) and Buzzacchi et al. (2021) employ data from the top-tier professional football leagues, German Bundesliga and Italian Serie A, respectively. Both studies show that individual managers' effects are economically and statistically meaningful for explaining the variations in field performance measured by the average points obtained per match per half-season. Buzzacchi et al. (2021) in addition find that financial performance measured by the growth in players' market values is partly explained by individual manager's effects.

The studies mentioned above conclude that managerial contributions are statistically significant, that is, managers do matter to determine the firm's success both in corporate and sports settings. This view is supported by another strand of literature on technical efficiency, which suggests that the extent to which a firm can achieve technical efficiency, i.e. the optimal output given production factors available, is dependent on the quality of managers. For instance, Frick and Simmons (2008) and Dawson et al. (2000) provide empirical evidence that technical inefficiencies are reduced by employing more competent managers in German Bundesliga and English Premier League, respectively. Such relationship is also established in other sports (Hofler and Payne, 2006; Brian, 2013) and beyond the sports industry (Papadopoulos, 2021).

Therefore, it is not surprising that leaders are rewarded or punished according to the field performance. The most common causes of dismissals of football managers are indeed poor performance in recent matches (van Ours and van Tuijl, 2016; Tena and Forrest, 2007). Nevertheless, the prospect of improvement in post-succession performance is not well-supported by empirical evidence. Employing the data from the Dutch football league, van Ours and van Tuijl (2016) show that seemingly positive effects of managerial change are not due to the succession itself, but merely a regression to the mean performance level. Tena and Forrest (2007) find positive effects of a new manager in Spanish football, however, this is limited to performance in home matches, concluding that such improvement in performance at home may simply reflect a boost in support by the fans. Evidence from National Hockey League (Rowe et al., 2005) suggests that within-season managerial change can deteriorate performance, followed by positive long-term effects.

This unfortunate phenomenon, particularly in the short term, prevails beyond the sports industry. A meta-analysis conducted on the relationship between CEO successions and firm outcomes by Schepker et al. (2017) concludes that leadership succession is on average costly to organisations since it causes disruptions in the short term, and any strategic change could take a while to manifest its effect. In this respect, a firm seeking a new leader to replace the incumbent would need to identify one that can at least partly offset these possible negative effects of succession. According to the studies cited above, however, this proves challenging, since we would otherwise see more successful cases of leadership changes.

One of the aspects that could be hindering an effective evaluation of managers is the randomness of outcomes. In professional football, in particular, any measure of performance that relies on match outcome is particularly susceptible to this concern due to the low-scoring nature of the sport. For instance, findings in a recent study of managerial dismissals in European football leagues by Flepp and Franck (2021) imply that other metrics, such as expected goals that capture the qualities of each shot within a match, can better reflect a manager's performance than conventional measure purely based on match outcome. In particular, they find that a club's performance improves after a streak of under-performance measured by match outcome, regardless of replacing a manager, whereas improvement is only expected with clubs who replace a manager when a club is experiencing under-performance based on expected goals. In the former case, it can be argued that performance simply reverts back to a mean level, since adverse performance leading up to dismissal was merely due to bad luck. On the other hand, dismissing a manager with poor performance based on expected goals improves the situation since this measure better captures a manager's actual ability, rather than luck. This suggests that it is important to account for the randomness of match outcome when evaluating a manager's performance.

The current study extends the previous studies estimating the managerial contribution to firm performance, such as Muehlheusser et al. (2018) and Buzzacchi et al. (2021) in the following ways. Firstly, the past studies disentangle the manager's effects from other contributors to firm performance by measuring the latter by individual firm effects. In the context of football clubs, these are club-specific fixed effects, which capture heterogeneity among individual clubs, or "how big a club is." However, these effects are effectively held constant over time. This can be a rather strong assumption since the resources available in a specific club is likely to be time-variant. For instance, the quality of players and a club's financial strengths can vary over time. In addition, separately estimating the two sets of fixed effects (managers and clubs) could significantly limit the scope of evaluation of individual managers. This is because such estimation requires managers to be observed in more than one club.

Therefore, we will explicitly control for the resources that are available for a manager to work with in order to achieve on-field success. As mentioned in the introduction, there are sophisticated rating systems for individual players in professional football, which can be used as a proxy for the quality of labour input. This measure of labour quality can provide more accurate information on the player's fitness at a particular point in time since this is updated every time a player appears on a pitch. Furthermore, a club's financial resources will be controlled with transfer budgets available, which also vary over time.

Second, Muehlheusser et al. (2018) and Buzzacchi et al. (2021) measure field performance with the average points obtained in a given half season. Following the discussion above, however, this measure is less capable of capturing a club's actual performance. Therefore, we measure a club's performance using expected goals. Finally, whilst the previous studies focus on a single league, we estimate the individual manager's contribution using multiple leagues. This is a relevant matter since the labour market for football managers is characterised by great mobility. Our measures of labour and capital inputs are comparable across the leagues, making the analysis with multiple leagues feasible.

These extensions allow us to compare managers who are operating in the different leagues in a way robust to random variation, without limiting the pool of managers to those hired by multiple clubs during the sample period. Therefore, this study contributes to understanding of managerial inputs and the development of the evaluation system of a manager by estimating heterogeneity amongst managers based on their contributions to match-day performance, given the resources at hand in the particular period of time.

3 Data

3.1 Match and manager data

We collected data from the top-tier football leagues in Spain, France, England, Italy, and Germany, known as the "Big Five", for seasons from 2014/2015 to 2020/2021. A very similar league format is followed by the five leagues included in our sample. Each league features the 18 or 20 most competitive clubs in the respective countries. A club competes with all others twice during a season, once at its home stadium and once away. With 20 participating clubs, this amounts to 38 matches per club per season or 380 matches per league per season. A club is rewarded with points after each match based on a match outcome; 3, 1, and 0 points for a win, draw, and loss, respectively. At the end of a season, the championship title is awarded to the club with the highest number of cumulative points within the respective league, whilst the weakest clubs are relegated to the second-tier league.[3] Table 1 provides the number of matches used in the main analysis by leagues.

Table 1: Number of matches used in the analysis by leagues

League	Matches
England Premier League	2,658
France Ligue 1	2,554
Germany Bundesliga	2,140
Italy Serie A	2,658
Spain LaLiga	2,658
Total	**12,668**

Notes: Seasons 2014/2015-2020/2021.

During the sample period, we observe 405 managers across 146 clubs, who together managed 12,668 matches. As mentioned in Section 2, the previous studies relied on the "mover condition," where a manager has to be observed or matched with more than one club, to estimate the individual effects of the managers.

[3] In general, the bottom three clubs are relegated, whilst the number could vary slightly since some leagues feature playoffs.

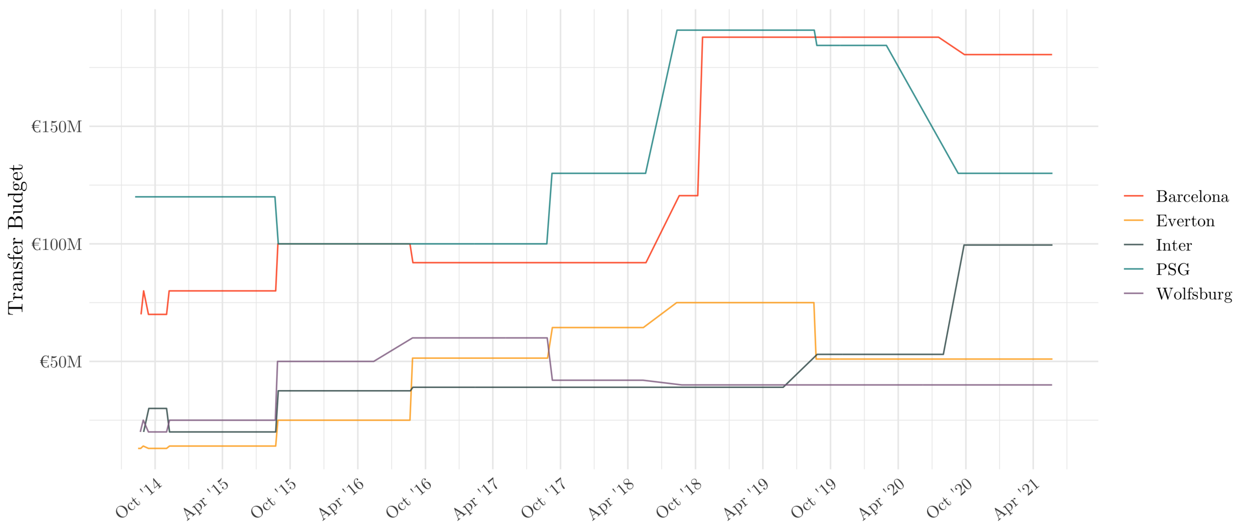
We argue that this condition could be quite restrictive since the number of managers whose effects can be estimated would be reduced significantly, which hinders the comparison among different managers. One way to overcome this issue is to expand the number of individual clubs or seasons, across which a manager may be observed.

We indeed build on the previous studies by covering more than a single league. This is certainly relevant in the context of European football since the labour market for football managers is mobile, particularly at the top tier. Nevertheless, there is a limit on expanding the time horizon. This is because some data are only available for more recent seasons. Whilst match-level statistics go back to as far as the establishment of the leagues, more advanced data, such as event-level data, have only recently been introduced. As explained later in this section, part of our analysis relies on this type of data. Regardless, the value added by including older seasons can be limited, due to the attrition of managers and the fact that the more recent data carry more useful information to predict future performance.

Imposing the mover condition on our sample would shrink the pool of managers to 142 individuals. To avoid this, we employ alternative measures to control for club heterogeneity, in terms of resources available for a manager to work with. In particular, we collected information on a club's transfer budgets, which change over time. This is obtained from the EA Sports FIFA website, known as *SoFIFA.com*[4]. The website provides individual club profiles that are updated regularly. Another advantage of this measure is that it is directly comparable among clubs that are operating within different leagues. Figure 1 plots the transfer budgets over the sample period for the five selected clubs: Barcelona (Spain), Everton (England), Inter (Italy), PSG (France), and Wolfsburg (Germany). The Figure suggests that there is significant heterogeneity in budgets available among clubs and that it varies across time within clubs.

[4] Available at https://sofifa.com/

Figure 1: Transfer budget for selected clubs (2014/2015-2020/2021)



We do, however, impose a restriction on managers' appearances. In particular, we estimate the individual fixed effects of managers who managed more than 10 matches in the sample period. The threshold is used since with our sample this threshold fairly separates permanent managers from caretaker managers. This leaves us with 322 managers. The number of matches managed by the 322 individuals ranges from 11 to 266, with the mean and median values of 77.15 and 53, respectively.

3.2 Event data

In addition to the match-level data, shot information is extracted from the Opta F24 feed, so-called "event data." Within the sample period, 321,074 shots are identified. For each shot, we gathered the outcome (whether it led to a goal), and the following characteristics: location ((x, y) coordinate), body part involved, and situation. The (x, y) coordinate is used to obtain the distance from the centre of the goal (*Distance*) and the angle between the coordinates of the shot origin and the two goal posts (*Angle degrees*). The categorical variable *Body part* indicates whether a shot involved a header or other body parts (right/left foot or other body parts). *Situation* variable is categorised into free-kick, open play, penalty, or corner kick. The descriptive statistics of shot data are presented in Table 2.

Table 2: Descriptive statistics for shot data

Variable	N = 320,050[1]
Goal	
0	285,941 (89%)
1	34,109 (11%)
Distance	17 (12, 25)
Angle degrees	20 (15, 31)
Body part	
Foot	265,069 (83%)
Header	54,981 (17%)
Situation	
Corner	46,894 (15%)
Free kick	35,866 (11%)
Open play	233,255 (73%)
Penalty	4,035 (1.3%)

Notes: [1]n (%); Median (IQR).

3.3 WhoScored rating

The data for individual players were collected from *WhoScored.com*.[5] They provide players' ratings (Who Scored (WS) ratings), which reflect the contribution of individual players to the outcome in every match. This means that the rating for a particular individual is updated every time they appear on the pitch. The initial value of the rating is set at 6.0 and the values can increase (up to the maximum value of 10.0) or decrease based on in-match and post-match statistics. In-match statistics are related to their action (dribble, shots, etc.). This is then weighted with its impact measured by the area on the pitch and outcome of the event, and reflected in their rating. At the end of the match, the overall outcome of the match is added to the rating, where such an outcome is weighted based on the individual's influence, determined by playing position, minutes played, etc. The advantages of using WS ratings are that the player's performance is evaluated regularly and that the ratings are obtained for virtually all professional leagues around the world, using the same method. Previous studies, Frick and Simmons (2008), for instance, used players' wage bills to capture labour input. However, this measure is generally only observed when a new contract is agreed,[6] and is available for limited countries.

The website covers 18 major professional football leagues, and there are 1,048,877 WS ratings available that resulted from 37,442 matches within these leagues over the seasons 2013/2014-2020/2021. The number of WS ratings collected by leagues is summarised in Table 3.

[5]Available at https://www.whoscored.com/.
[6]Hence it could even be less often than annually.

Table 3: Number of WS ratings by leagues

League	N(Ratings)
Argentina Superliga	46, 463
Belgium Jupiler Pro League	8, 703
Brazil Brasileirão	75, 063
England Championship	122, 343
England League One	15, 610
England League Two	27, 556
England Premier League	83, 761
France Ligue 1	82, 229
Germany Bundesliga	68, 963
Germany Bundesliga II	47, 442
Italy Serie A	86, 121
Netherlands Eredivisie	65, 720
Portugal Liga NOS	43, 782
Russia Premier League	53, 107
Scotland Premiership	5, 659
Spain LaLiga	85, 878
Turkey Super Lig	63, 873
USA Major League Soccer	66, 604
Total	**1,048,877**

Notes: Seasons 2013/2014-2020/2021.

For each WS rating collected, our data set indicates the date on which a rating is observed, i.e. the date of the match on which a rating is based, the name of the player associated with the rating, their date of birth, and the league in which they played on the date. In light of the adjustments we are going to propose in the following section, we identify the players who were observed in more than one league during the sample period and played at least 10 matches within a given league. We identify 3,478 players who satisfy these conditions, and the number of ratings observed for these players amounts to 406,239.

4 Methodology

Our ultimate goal is to estimate a club's production function relative to its opponent at a match level and disentangle an individual manager's contributions from those of other inputs, labour and capital, measured at the time when a match takes place. Therefore, using the data presented in Section 3, we construct our output and input variables at a match level. Our output measure is the so-called expected goals (xG). This is based on the probability of a shot leading to a goal, given observed characteristics of the shot. To obtain the output measure at a match level, therefore, these probabilities are aggregated over all the observed shots in a specific match.

As for input measures, we use transfer budgets and individual players' historical performance (*WhoScored* ratings) as proxies for financial strengths and labour quality. The latter is observed at an individual player level, hence we construct a variable to capture the strength of the squad on a particular match day, using the historical performance of the players within the squad. The player ratings used in this study are based on an impact that he makes in a particular match and created at a match-level across most of the major leagues across the world.[7] However, these ratings are not directly comparable across the leagues in that it is easier for a player to make an impact in lower-level leagues than in higher-level ones. Therefore, the first adjustment we make is to take into account the different levels of difficulty across the leagues. Once we obtain league-adjusted ratings for individual players, we compute the weighted average of ratings observed within two years prior to the match for which we estimate the production function. The weights applied to each rating available in the time window are based on how far it is in the time horizon so that a more recent rating is given a higher weight than an old rating. In addition, the initial value of rating (the rating of a particular player exactly two years prior to the match day) is set at the average value of league-adjusted ratings in the league where the match under consideration is observed. This is so that the weighted average of ratings for players with very few observations is shrunk towards this initial value, rather than skewed towards these few observations. We then obtain the mean value of the weighted average of league-adjusted historical ratings over players in the squad available for the match under consideration to capture the overall strength of the squad on the day.

Finally, we estimate the match-level production function using the output and input metrics computed as above together with the individual manager dummies, where all output and input are measured for a home club relative to an away club. The remainder of this section describes each step in detail.

4.1 xG model

The conventional way to measure output is based on the tertiary outcome (win, loss, and draw) or points earned in each match. However, this measure is susceptible to random variation in match outcome. This is particularly relevant due to the low-scoring nature of association football, where relying on match outcomes that are heavily affected by random forces can lead to misjudgment (Brechot and Flepp, 2020). Therefore, this study measures performance with scoring chance and its quality, which is observed much more frequently than goals. In particular, we adopt the concept of expected goals (xG), which assigns the probability of scoring to each shot that occurred in a match. Whilst the number of studies that employ xG as a measure of performance in professional football is yet sparse (Kharrat et al., 2020; Flepp and Franck, 2021; Brechot and Flepp, 2020), the measure is shown to add to the predictive power for future performance (Brechot and

[7] See Table 3 for the list of the leagues in which the individual players' historical ratings are observed.

Flepp, 2020).

Using the shot data described in Section 3.2, we first estimate the log odds of a shot leading to a goal, given the characteristics of the shot as follows:

$$\begin{aligned}\text{Ln}\left(\frac{\text{P}[\text{Goal}_{ikts}=1]}{\text{P}[\text{Goal}_{ikts}=0]}\right) = f(&\text{ Angle degrees}_{ikts},\ \text{Distance}_{ikts},\\ &\text{Body part}_{ikts},\ \text{Situation}_{ikts}\ ),\end{aligned} \tag{1}$$

where i, k, t, and s represent a club, shot, match, and season, respectively. We estimate the model (1) by means of logistic regression, using observations from the first season (2013/2014). Once the model is estimated, we obtain the fitted $\hat{\text{P}}[\text{Goal}_{ikts}=1]$ for each shot and aggregate these values for home and away clubs in a particular match in the rest of the seasons (2014/2015 - 2020/2021). This gives us the expected goals for home and away clubs in the match.

4.2 Adjustments of WS ratings

We propose some adjustments to WS ratings presented in Section 3.3 in order to (1) make the ratings comparable across the leagues, (2) mitigate the possible bias towards very few observations for some players, and (3) take into account the fact that newer observations carry more useful information than older ones. Given that WS ratings essentially reflect an individual player's contributions to within and overall match performance, the rating of an individual player is influenced by how difficult it is to make an impact in the particular league in which he plays, because different leagues have different standards of play. For instance, a player in the top-tier English league may have a low rating, however, his rating may well be higher if measured within the second-tier league. To measure the quality of a player on a particular match day, historical ratings of the player prior to the match day will be taken into account. However, a player typically moves across leagues, as his career progresses. Therefore, it is important to adjust the ratings in order to make them comparable across the leagues in which they are observed.

To achieve this, we first identify the degree of heterogeneity among leagues by running the following regression:

$$\text{WS} \ = \ \beta_0 \ + \ X'\gamma \ + \ L'\rho \ + \ \beta_1\text{Age} \ + \ \beta_2\text{Age}^2 + \varepsilon, \tag{2}$$

where X and L are vectors of dummy variables for individual players and leagues, respectively. A variable *Age* and its quadratic term are included to control for the general progression of a player's quality over their career. Therefore, ρ is the vector of league coefficients of interest. To estimate ρ, we use the observations for raw WS ratings that belong to players who have been observed in at least two leagues where they appeared in a minimum of 10 matches. As described in Section 3.3, we identify 3,478 players who meet the criteria, for whom we have 406,239 observed ratings.

Once we obtain the league coefficients, we add the negative value of the respective league coefficient to a realised (raw) WS rating to obtain league adjusted WhoScored (LAWS) rating defined as follows:

$$\mathrm{LAWS} \;=\; \mathrm{WS} \;+\; (-\hat{\rho}_l)\,, \tag{3}$$

where ρ_l is the estimated league coefficient for a league l. The term $(-\hat{\rho}_l)$, therefore, can be interpreted as league strength, or the level of difficulty for a player to make a positive effect on the game.

Using LAWS ratings, we measure a player's ability at time t using all the available LAWS ratings in the preceding periods of two years, (LAWR_1, LAWR_2, ..., LAWR_n). Previous literature on the prediction of outcome in football (Boshnakov et al., 2017; Kharrat et al., 2020; Dixon and Coles, 1997) suggests that recent performance carries more predictive power for future performance. Accordingly, we apply the following weighting function for each LAWS rating i:

$$w_i = \exp\left(\xi(\mathrm{Date}_i - \mathrm{Date}_t)/3.5\right), \tag{4}$$

where ξ is the time-weighting parameter, Date_t is the date when the player's ability is measured, and Date_i is the date when LAWS_i is observed. Following the studies mentioned above, time distances are scaled in half-week units (= 3.5 days), and set $\xi = 0.002$. Therefore, LAWS_i that are recently observed (closer to time t) are given exponentially higher importance to measure a player's ability at time t.

The number of LAWS ratings available for a certain player at time t, i.e. the number of LAWS_i between time t and $t-$ (2 years), may vary significantly among players. If a player has very few observations, this could result in a bias towards such observations. To circumvent this, we set the initial values (a rating that would have been observed exactly two years prior to t) to be the average value of LAWS for a respective league. By doing so, in an extreme case where a player has no previous observation at time t, his ability at time t will shrink towards the mean value in the respective league, weighted according to the time difference of two years. On the other hand, if a player is regularly observed, this shrinkage hardly affects their rating, particularly because the time discount applied to the initial value is large. Therefore, we measure the ability of a player at time t with the following weighted value of LAWS:

$$\overline{\mathrm{LAWS}}_t \;=\; \left(\tilde{w_0} + \sum_{i=1}^{n} w_i\right)^{-1} \left(\tilde{w_0} \times \mathrm{LA\tilde{W}S}_l \;+\; \sum_{i=1}^{n} w_i \times \mathrm{LAWS}_i\right), \tag{5}$$

where $\tilde{w_0}$ is the weight with the time distance of two years; n is the number of the player's LAWS ratings available between time t and $t-$ (2 years); $\tilde{\text{LAWS}}_l$ is the average LAWS ratings of the league l to which a player belongs at time t.

We therefore calculate $\overline{\text{LAWS}}_t$ for each available player on match day t using WS ratings observed between t - (2 years) and the last appearance of a respective player prior to t. We then obtain average values separately for the starting eleven and substitutes, which gives overall strength of the squad available for a manager to work with on a particular match day t. The advantage of measuring the quality of players this way is that it can vary throughout the season, hence allowing it to capture possible drops in the strengths, for instance, due to injuries. The fact that individual players' ability is updated every time they are observed on the pitch means the fluctuation in their performance is also reflected. These are the elements that are not captured in other measures, for instance, their wage bills or even market values.

4.3 Production function

We model an individual club i's output within a match that takes place at time t in season s, (Y_{its}), as a function of inputs, i.e. labour (L_{its}) and capital (K_{its}). In addition, we include home advantage (γ_{it}) and managers' fixed effects (μ_i). Specifically, a club i's production function is defined as the following Cobb-Douglas production function:

$$Y_{its} = {L_{its}}^{\beta_l} {K_{its}}^{\beta_k} \exp(\gamma_{it}\mu_i), \tag{6}$$

where the return parameters β_l and β_k measure the impact of labour and capital. Ultimately, a match outcome is dependent on a club i's output relative to that of its opponent club j, therefore, we rewrite the model (6) in logs and take the differences between clubs i and j[8]:

$$y_{its} - y_{jts} = \gamma_{it} - \gamma_{jt} + \beta_l(l_{its} - l_{jts}) + \beta_k(k_{its} - k_{jts}) + \mu_i - \mu_j + \varepsilon_{ijts}, \tag{7}$$

where ε_{ijts} is a match-specific error term.

We now look at model (7) from a home club's perspective, and consider a match outcome as a home club's log xG net of an away club's log xG ($d\_log\_xG_{ts}$). Similarly, labour and capital inputs are included as a log-difference form of the measure for the strengths of the squad and transfer budgets, respectively. Furthermore, to obtain an individual manager's coefficients, we include a 322×1 vector M_{gt} of manager

[8]The definition and transformation of the production function for a sports club are similar to those in (Peeters et al., 2014).

dummies m_{ts}, where:

$$m_t = \begin{cases} -1 & \text{if a manager } m \text{ manages the away club in match } t, \\ 0 & \text{if a manager } m \text{ manages neither the home nor away club in match } t, \\ 1 & \text{if a manager } m \text{ manages the home club in match } t. \end{cases} \tag{8}$$

The resulting production function is, therefore:

$$\text{d\_log\_xG}_{gt} = (\gamma_h - \gamma_a) + \beta_l(l_{hgt} - l_{agt}) + \beta_k(k_{hgt} - k_{agt}) + M_{gt}{}' \mu + \varepsilon_{gt}, \tag{9}$$

where μ is a 322×1 vector of the coefficients for individual managers, and the first term $(\gamma_h - \gamma_a)$ is interpreted as a home advantage and estimated as an intercept of the model.

5 Results

5.1 xG and naive rankings

The estimation results of the expected goal model (1) using the first season (2013/2014) is summarised in Table 4. All the variables are statistically significant, and the estimated parameters have expected signs.

Table 4: Estimated xG model parameters

	Dependent variable:
	Goal
Angle degrees	0.032^{***} (0.001)
Distance	-0.084^{***} (0.004)
Body part	
Header	-1.070^{***} (0.047)
Situation	
Free kick	0.642^{***} (0.092)
Open play	0.669^{***} (0.077)
Penalty	3.275^{***} (0.129)
Constant	-2.186^{***} (0.125)
Observations	47,766
Log Likelihood	−12,991.710
Akaike Inf. Crit.	25,997.420
Note:	*p<0.1; **p<0.05; ***p<0.01

We then obtain the fitted values ($\hat{\text{P}}[\text{Goal}_{ikts} = 1]$, where i = club, k = shot, t = match, and s = season) for the observations within seasons 2014/2015-2020/2021. Finally, to obtain expected goals per club per match (xG_{its}), we aggregate these fitted values for home and away clubs in a particular match. Figure 2 plots the values obtained for xG for home and away clubs in the main sample.

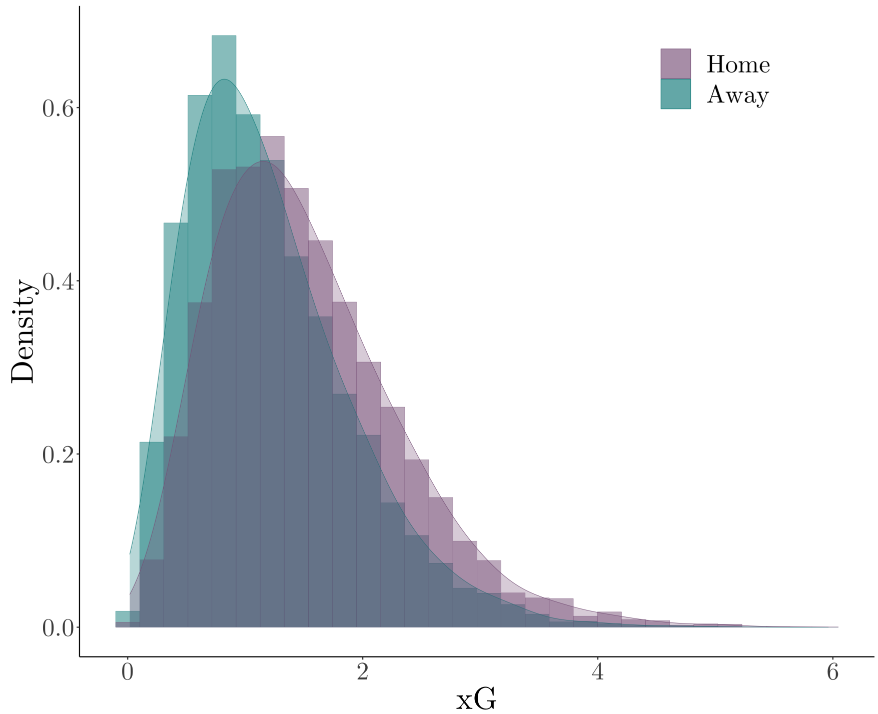


Figure 2: Estimated xG values for home and away clubs

One of the conventional ways to measure a football manager's success is to look at a winning percentage (WP) over their career.[9] As a comparison, Table 5 provides the rankings of managers in terms of WP and the average net xG (d_xG), i.e. xG differentials between the managing club and their opponents, for the top 50 and selected managers. We call these "naive" rankings since they do not take into account the strengths of a squad and the financial resources that a manager has at his disposal. The only difference between the two rankings is the performance (output) metric.

Nevertheless, the Table provides quite different pictures, implying that not taking into account the randomness of the outcome can provide misperception with respect to an individual manager's performance. For instance, Diego Simeone and Santiago Solari are ranked 10th and 11th, respectively, in terms of the WP. However, the two Argentinean managers are lower-ranked (35th and 39th, respectively) in terms of xG differentials. This may indicate that the WPs of the two reflect good luck to some extent. On the other hand, Stefano Pioli and, more notably, Graham Potter are examples of managers whose abilities may not be reflected in their WPs. Stefano Pioli, a manager of AC Milan (Italy), is ranked 47th and 27th in terms of xG differentials and WP, respectively. Graham Potter, a manager of Brighton & Hove Albion F.C. (England), is

[9] This is obtained by WP $= ((\text{N of wins}) + 0.5\,(\text{N of draws}))/(\text{N of matches})$.

ranked much higher in terms of xG differentials (62nd) than WP (173th). Note, however, that both rankings are again naive in the sense that other inputs that would have contributed to either measure of performance are not taken into account.

Table 5: Naive rankings (winning percentage and xG differential)

Ranking	Manager (Club)	WP	Ranking	Manager (Club)	d_xG
1	Josef Heynckes (Bayern)	0.870	1	Luis Enrique (Barcelona)	1.483
2	Hans-Dieter Flick (Bayern)	0.836	2	Josep Guardiola (Man City)	1.396
3	Luis Enrique (Barcelona)	0.825	3	Josef Heynckes (Bayern)	1.263
4	Laurent Blanc (PSG)	0.822	4	Hans-Dieter Flick (Bayern)	1.227
5	Massimiliano Allegri (Juventus)	0.821	5	Thomas Tuchel (PSG)	1.146
6	Josep Guardiola (Man City)	0.807	6	Laurent Blanc (PSG)	1.072
7	Zinedine Zidane (Real Madrid)	0.781	7	Zinedine Zidane (Real Madrid)	0.901
8	Thomas Tuchel (PSG)	0.766	8	Frank Lampard (Chelsea)	0.849
9	Antonio Conte (Inter)	0.765	9	Jurgen Klopp (Liverpool)	0.812
10	Diego Simeone (Atletico)	0.736	10	Massimiliano Allegri (Juventus)	0.811
11	Santiago Solari (Real Madrid)	0.735	11	Andrea Pirlo (Juventus)	0.802
12	Andrea Pirlo (Juventus)	0.724	12	Maurizio Sarri (Juventus)	0.731
13	Luciano Spalletti (Inter)	0.720	13	Antonio Conte (Inter)	0.730
14	Jurgen Klopp (Liverpool)	0.707	14	Ralf Rangnick (RBL)	0.728
15	Maurizio Sarri (Juventus)	0.706	15	Luciano Spalletti (Inter)	0.694
16	Carlo Ancelotti (Everton)	0.696	16	Arsene Wenger (Arsenal)	0.680
17	Ralf Rangnick (RBL)	0.691	17	Carlo Ancelotti (Everton)	0.638
18	Mauricio Pochettino (PSG)	0.675	18	Bruno Genesio (Rennes)	0.619
19	Ernesto Valverde (Barcelona)	0.675	19	Edin Terzic (Borussia Dortmund)	0.607
20	Edin Terzic (Borussia Dortmund)	0.674	20	Gian Piero Gasperini (Atalanta)	0.599
21	Jorge Sampaoli (Marseille)	0.673	21	Niko Kovac (Monaco)	0.579
22	Adi Hutter (Eintracht Frankfurt)	0.671	22	Paulo Fonseca (Roma)	0.577
23	Leonardo Jardim (Monaco)	0.669	23	Julen Lopetegui (Sevilla)	0.570
24	Bruno Genesio (Rennes)	0.668	24	Julian Nagelsmann (RBL)	0.563
25	Unai Emery (Villarreal)	0.665	25	Peter Bosz (Leverkusen)	0.545
26	Ole Gunnar Solskjaer (Man Utd)	0.665	26	Unai Emery (Villarreal)	0.535
27	Lucien Favre (Borussia Dortmund)	0.663	27	Stefano Pioli (AC Milan)	0.523
28	Julen Lopetegui (Sevilla)	0.663	28	Rudi Garcia (Lyon)	0.504
29	Gennaro Gattuso (Napoli)	0.659	29	Ole Gunnar Solskjaer (Man Utd)	0.469
30	Arsene Wenger (Arsenal)	0.658	30	Gennaro Gattuso (Napoli)	0.467
31	Andre Villas-Boas (Marseille)	0.656	31	Guus Hiddink (Chelsea)	0.453
32	Rudi Garcia (Lyon)	0.654	32	Ronald Koeman (Barcelona)	0.420
33	Jose Mourinho (Tottenham)	0.649	33	Brendan Rodgers (Leicester)	0.416
34	Louis Van Gaal (Man Utd)	0.638	34	Ernesto Valverde (Barcelona)	0.399
35	Simone Inzaghi (Lazio)	0.637	35	Diego Simeone (Atletico)	0.398
36	Julian Nagelsmann (RBL)	0.636	36	Mauricio Pochettino (PSG)	0.395
37	Hubert Fournier (Lyon)	0.632	37	Simone Inzaghi (Lazio)	0.393
38	Gian Piero Gasperini (Atalanta)	0.620	38	Jorge Sampaoli (Marseille)	0.387
39	Sergio Conceicao (Nantes)	0.614	39	Santiago Solari (Real Madrid)	0.380
40	Niko Kovac (Monaco)	0.613	40	Leonardo Jardim (Monaco)	0.370
41	Paulo Fonseca (Roma)	0.612	41	Hubert Fournier (Lyon)	0.370
42	Christophe Galtier (Lille)	0.608	42	Frank de Boer (Crystal Palace)	0.365
43	Peter Bosz (Leverkusen)	0.604	43	Adi Hutter (Eintracht Frankfurt)	0.353
44	Bo Svensson (Mainz)	0.600	44	Marco Rose (Borussia M.Gladbach)	0.351
45	Marcelino Garcia Toral (Athletic Bilbao)	0.597	45	Roger Schmidt (Leverkusen)	0.311
46	Ronald Koeman (Barcelona)	0.596	46	Franck Haise (Lens)	0.309
47	Stefano Pioli (AC Milan)	0.596	47	Alexander Zorniger (Stuttgart)	0.286
48	Marco Rose (Borussia M.Gladbach)	0.596	48	Jose Mourinho (Tottenham)	0.273
49	Guus Hiddink (Chelsea)	0.595	49	Roberto Mancini (Inter)	0.272
50	Oliver Glasner (Wolfsburg)	0.588	50	Imanol Alguacil (Real Sociedad)	0.267
66	Marcelo Bielsa (Leeds)	0.543	62	Graham Potter (Brighton)	0.147
173	Graham Potter (Brighton)	0.421	92	Marcelo Bielsa (Leeds)	0.016
189	Eddie Howe (Bournemouth)	0.410	216	Eddie Howe (Bournemouth)	−0.347
246	Gary Neville (Valencia)	0.344	258	Gary Neville (Valencia)	−0.539

Notes: Club with which a manager appeared most recently in the sample is in parentheses.

5.2 League coefficients and league adjusted player ranking

Following Section 4.2, we first estimate model (2) to obtain league coefficients. Table 6 presents the resulting ranking of the league strengths, i.e. the negative value of the estimated league-specific effects.

Table 6: Estimated league coefficients

	League	$-\hat{\rho}_l$
1	Premier League (England)	0.378
2	LaLiga (Spain)	0.348
3	Serie A (Italy)	0.292
4	Bundesliga (Germany)	0.224
5	Ligue 1 (France)	0.212
6	Brasileirão (Brazil)	0.183
7	Championship (England)	0.159
8	Super Lig (Turkey)	0.123
9	Premier League (Russia)	0.109
10	Liga NOS (Portugal)	0.089
11	League One (England)	−0.007
12	Major League Soccer (USA)	−0.010
13	Bundesliga II (Germany)	−0.024
14	Jupiler Pro League (Belgium)	−0.024
15	Eredivisie (Netherlands)	−0.060
16	Premiership (Scotland)	−0.091
17	League Two (England)	−0.147

Notes: Reference league is Superliga (Argentina).

The Table suggests that prominent differences exist across leagues. Other things being equal, being in the English Premier Leagues lowers an individual player's ratings by 0.378 points relative to the reference category of the Argentinian Superliga. It is also worth noting that the Big Five are indeed ranked in the top five.

Table 7 shows the highest values of the weighted average of ratings ($\overline{\text{WS}}_t$) and League Adjusted WhoScored ratings ($\overline{\text{LAWS}}_t$) achieved by a unique individual. Applied to both cases are time discounting based on the weighting function (4) and shrinkage towards the league average as the initial value described in Section 4.2. The only difference between the left-hand side and the right-hand side of the table, therefore, is whether the ratings are adjusted for the league strengths. Lionel Messi and Neymar attained the highest ratings in both cases. As expected, however, without the league adjustment ($\overline{\text{WS}}_t$) some players in the lower tier leagues achieved higher ratings, for instance, James Tavernier in Premiership (Scotland), who is ranked higher than Kevin De Bruyne in Premier League (England). The left-hand side of the table produces the ranking, which is probably more in line with what one would expect; the well-known players from the Big Five, for instance, Kevin De Bruyne, Harry Kane, and Sergio Agüero are included in the top ten.

To measure the strength of a squad at time t, therefore, we take the average value of $\overline{\text{LAWS}}_t$ for the players available on a match day t.

Table 7: Weighted average of WS and LAWS ratings

	Player	Date	League	$\overline{\mathrm{WS}}_t$		Player	Date	League	$\overline{\mathrm{LAWS}}_t$
1	Lionel Messi	2021/04/25	LaLiga (Spain)	8.025	1	Lionel Messi	2021/04/25	LaLiga (Spain)	8.370
2	Neymar	2021/02/07	Ligue 1 (France)	7.767	2	Neymar	2020/11/20	Ligue 1 (France)	8.053
3	Hakim Ziyech	2020/02/16	Eredivisie (Netherlands)	7.548	3	Cristiano Ronaldo	2021/01/17	Serie A (Italy)	7.822
4	Cristiano Ronaldo	2021/04/21	Serie A (Italy)	7.504	4	Kevin De Bruyne	2021/01/20	Premier League (England)	7.726
5	Robert Lewandowski	2021/05/15	Bundesliga (Germany)	7.474	5	Robert Lewandowski	2021/05/15	Bundesliga (Germany)	7.713
6	Zlatan Ibrahimovic	2021/02/13	Serie A (Italy)	7.439	6	Eden Hazard	2020/06/21	LaLiga (Spain)	7.712
7	James Tavernier	2020/12/30	Premiership (Scotland)	7.429	7	Luis Suárez	2020/06/13	LaLiga (Spain)	7.698
8	Hulk	2016/05/15	Premier League (Russia)	7.379	8	Harry Kane	2021/05/19	Premier League (England)	7.672
9	Steven Berghuis	2021/04/25	Eredivisie (Netherlands)	7.366	9	Zlatan Ibrahimovic	2021/02/13	Serie A (Italy)	7.627
10	Kevin De Bruyne	2021/01/20	Premier League (England)	7.355	10	Sergio Agüero	2019/09/28	Premier League (England)	7.626

Notes: Table shows the top 10 weighted average of WhoScored ratings ($\overline{\mathrm{WS}}_t$) and League Adjusted WhoScored ratings ($\overline{\mathrm{LAWS}}_t$) achieved by a unique player during the period from 2013/07/14 to 2021/05/23. Date and League are the date when the latest rating used to obtain the weighted average is realised, and the league the player belonged to at that time.

5.3 Estimated production function

Using expected goals and adjusted ratings presented above, together with transfer budgets to control for a club's financial strength, we estimate the model (9). The estimation results are reported in Table 8. The dependent variable is the log-difference of expected goals from a home club's perspective (*d_log_xG*). Other inputs (*Player ratings* and *Transfer budgets*) are also in the log-difference form according to the model (9). In columns (1) and (2), *Player ratings* is based on the mean value of average adjusted ratings of a squad, whilst (3) and (4) include separately that of the starting lineup and substitutions, *Player ratings (starting)* and *Player ratings (substitutions)*, respectively. Manager fixed effects are included in columns (2) and (4) but not in columns (1) and (3). The coefficient estimates on the manager dummies, i.e. the estimates of manager fixed effects, are not reported in Table 8 since they will be presented and discussed further in the next subsection.

All the coefficients on the relative input variables have expected signs; the relative quality of players and financial strengths positively affect the relative performance measured by expected goals. In addition, home advantage (Constant) is positive and significant in all specifications, as expected. The comparisons between columns (1) and (2) as well as (3) and (4) imply that managerial input also influences field performance. The inclusions of manager fixed effects reduce the size of coefficients for labour and capital inputs, whilst that of home advantage is not affected. This implies that players' effects in columns (1) and (3) partly capture the manager's input. Considering that the production function models the match-day contribution of a manager given the resources available on the day, it can be argued that managerial input adds to the players' impacts on the field performance through the decisions on tactics and substitution, as well as psychological supports.

The estimated parameters show statistically significant effects of each input, except for the strength of substitutions in column (4), where such effects are not significant at the 5% significance level. The possible explanation for this is that the seemingly positive effects of substitution quality in column (3) are due to a manager's decision on substitutions during the match, rather than the strengths of substitutions themselves. In other words, whilst the robust set of substitutions can marginally add to the performance, it is possibly the quality of decisions on substitutions made by a manager that matters more.

The joint significance of manager fixed effects is tested by means of the F-test in Table 9. The model numbers in Table 9 correspond to the column numbers in Table 8. The comparisons between models (1) and (2) as well as (3) and (4) confirm that manager fixed effects are jointly significant. This provides evidence that managers overall affect output level, after explicitly controlling for players' strength on a particular match day and financial resources.

Table 8: Estimated parameters for production functions

	Dependent variable:			
	d_log_xG			
	(1)	(2)	(3)	(4)
Player ratings	27.441***	18.236***		
	(0.949)	(1.377)		
Player ratings (starting)			23.662***	16.244***
			(0.906)	(1.197)
Player ratings (substitutions)			2.807***	1.543*
			(0.806)	(0.855)
Transfer budgets	0.092***	0.065***	0.085***	0.059***
	(0.009)	(0.013)	(0.009)	(0.013)
Constant	0.266***	0.266***	0.266***	0.266***
	(0.007)	(0.007)	(0.007)	(0.007)
Manager fixed effects	No	Yes (N > 10)	No	Yes (N > 10)
Observations	12,668	12,668	12,668	12,668
R^2	0.227	0.317	0.229	0.318
Adjusted R^2	0.227	0.299	0.229	0.300

Notes: *p<0.1; **p<0.05; ***p<0.01. All variables are in the log-difference form from a home club's perspective. In columns (1) and (2), *Player ratings* is based on the mean value of average adjusted ratings of a squad, whilst (3) and (4) include separately that of the starting lineup and substitutions. Manager fixed effects are included in columns (2) and (4), but not in columns (1) and (3). The individual coefficients on manager fixed effects are not reported.

Table 9: F-test for significance of manager effects

A. Model (1) v. (2)

Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
(1)	12,665	8,657.173				
(2)	12,343	7,651.564	322	1,005.609	5.038	0

B. Model (3) v. (4)

Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
(3)	12,664	8,628.030				
(4)	12,342	7,638.022	322	990.008	4.968	0

5.4 Estimated manager coefficients

In Table 10, we report the estimated coefficients for individual managers with the 50 largest coefficients and selected managers. The estimation is based on the specification in column (4) in Table 8. As explained in Section 3, we include dummy variables for the 322 managers who appeared in more than 10 matches in our sample. Therefore, our reference group is effectively the pool of managers who do not satisfy this criterion.

That is, the estimated coefficients presented in Table 10 are relative to the average performance of the 83 managers with less than or equal to 10 appearances. As described in Section 3, the latter group is mainly represented by caretaker managers.

Table 10 provides evidence of heterogeneity among managers in European football, controlling for home advantage, players' strengths, and financial resources. Our estimation suggests that Josep Guardiola, who managed FC Bayern Munich (German *Bundesliga*) and Manchester City F.C. (English *Premier League*) during our sample period, made the largest contribution to clubs' production. Specifically, the relative size of his influence to home advantage is 3.443 ($\approx 0.916/0.266$). With similar comparisons, the effects of the top 50 managers presented in Table 10 are all larger than the size of home advantage. The impact of an average manager (one with an average coefficient of 0.1163) is less than half the size of home advantage. For instance, Peter Stöger, the former manager of the German *Bundesliga* clubs, FC Cologne and Borussia Dortmund, has an estimated coefficient of 0.1189, which is equal to 0.447 relative to the size of home advantage.

Table 10: Estimated manager coefficients

ranking	manager	Estimate	Clubs (Seasons)
1	Josep Guardiola	0.916	Bayern ('14-'16), Man City ('16-'21)
2	Frank Lampard	0.839	Chelsea ('19-'21)
3	Luis Enrique	0.782	Barcelona ('14-'17)
4	Laurent Blanc	0.716	PSG ('14-'16)
5	Javier Aguirre	0.701	Leganes ('19-'20)
6	Jose Rojo Martin	0.692	SD Huesca ('20-'21)
7	Jurgen Klopp	0.640	Borussia Dortmund ('14-'15), Liverpool ('15-'21)
8	Josef Heynckes	0.619	Bayern ('17-'18)
9	Gian Piero Gasperini	0.605	Genoa ('14-'16), Atalanta ('16-'21)
10	Maurizio Sarri	0.577	Empoli ('14-'15), Napoli ('15-'18), Chelsea ('18-'19), Juventus ('19-'20)
11	Franck Haise	0.570	Lorient ('16-'17), Lens ('20-'21)
12	Graham Potter	0.561	Brighton ('19-'21)
13	Thomas Tuchel	0.561	Borussia Dortmund ('15-'17), PSG ('18-'20), Chelsea ('20-'21)
14	Frank Kramer	0.555	Arminia Bielefeld ('20-'21)
15	Massimiliano Allegri	0.547	Juventus ('14-'19)
16	Ralf Rangnick	0.536	RBL ('18-'19)
17	Stefano Pioli	0.520	Lazio ('14-'16), Inter ('16-'17), Fiorentina ('17-'19), AC Milan ('19-'21)
18	Sabri Lamouchi	0.504	Rennes ('17-'19)
19	Julen Lopetegui	0.496	Real Madrid ('18-'19), Sevilla ('19-'21)
20	Pascal Plancque	0.484	Nimes ('20-'21)
21	Jose Bordalas	0.478	Getafe ('17-'21)
22	Nuno Espirito Santo	0.466	Valencia ('14-'16), Wolves ('18-'21)
23	Vahid Halilhodzic	0.459	Nantes ('18-'19)
24	Antonio Conte	0.458	Chelsea ('16-'18), Inter ('19-'21)
25	Julian Nagelsmann	0.455	Hoffenheim ('15-'19), RBL ('19-'21)
26	Clarence Seedorf	0.451	Deportivo ('17-'18)
27	Brendan Rodgers	0.447	Liverpool ('14-'16), Leicester ('18-'21)
28	Bo Svensson	0.447	Mainz ('20-'21)
29	Imanol Alguacil	0.445	Real Sociedad ('17-'21)
30	Sergio Conceicao	0.443	Nantes ('16-'17)
31	Ole Gunnar Solskjaer	0.428	Man Utd ('18-'21)
32	Bruno Genesio	0.415	Lyon ('15-'19), Rennes ('20-'21)
33	Niko Kovac	0.414	Eintracht Frankfurt ('15-'18), Bayern ('18-'20), Monaco ('20-'21)
34	Jose Luis Mendilibar	0.414	Levante ('14-'15), Eibar ('15-'21)
35	Pablo Machin	0.408	Girona ('17-'18), Sevilla ('18-'19), Espanyol ('19-'20), Deportivo Alaves ('20-'21)
36	Giampiero Ventura	0.402	Torino ('14-'16), Chievo ('18-'19)
37	Rudi Garcia	0.400	Roma ('14-'16), Marseille ('16-'19), Lyon ('19-'21)
38	Carlo Ancelotti	0.398	Real Madrid ('14-'15), Bayern ('16-'18), Napoli ('18-'20), Everton ('19-'21)
39	Stephane Moulin	0.394	Angers ('15-'21)
40	Urs Fischer	0.386	Union Berlin ('19-'21)
41	Paulo Fonseca	0.385	Roma ('19-'21)
42	Luciano Spalletti	0.384	Roma ('15-'17), Inter ('17-'19)
43	Marcelo Bielsa	0.383	Marseille ('14-'16), Lille ('17-'18), Leeds ('20-'21)
44	Mikel Arteta	0.382	Arsenal ('19-'21)
45	Scott Parker	0.380	Fulham ('18-'21)
46	Hubert Fournier	0.376	Lyon ('14-'16)
47	Julien Stephan	0.375	Rennes ('18-'21)
48	Cristobal Parralo	0.373	Deportivo ('17-'18)
49	Bernard Blaquart	0.371	Nimes ('18-'20)
50	Edin Terzic	0.367	Borussia Dortmund ('20-'21)
150	Jose Mourinho	0.144	Chelsea ('14-'16), Man Utd ('16-'19), Tottenham ('19-'21)
154	Eddie Howe	0.136	Bournemouth ('15-'20)
308	Gary Neville	−0.333	Valencia ('15-'16)

Notes: Table reports the estimated manager coefficients for those with the 50 largest values and selected managers. The reference group is a set of managers who appeared in less than 10 matches during the sample period. Also presented are clubs and seasons where a manager was observed during the sample period.

Depicted in Figure 3 are the coefficients of the top 10 and selected managers with 95% confidence intervals. These effects of the top 10 managers are individually different from zero, and even among this very top group, the heterogeneity is evident.

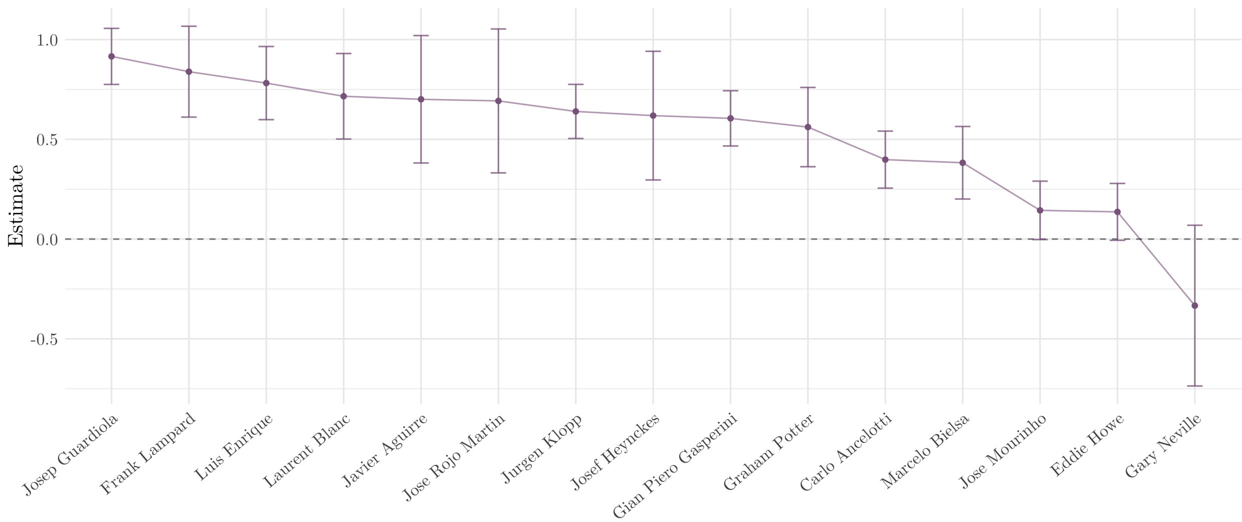


Figure 3: Manager coefficients with 95% confidence intervals

Recalling the naive rankings in Table 5, we compare our ranking based on the manager coefficients (MFE) with the two naive rankings in Figure 4. In Panel (I), the 322 managers are ordered with respect to the naive ranking based on the win percentage (WP) on the x-axis in descending order, which is plotted against the ranking based on the MFE, together with a 45-degree line. Panel (II) is similar, except that the naive ranking is based on the average net expected goals (d_xG).

It is evident from Figure 4 (I) that the two rankings produce quite different pictures. The correlation between the two rankings is 0.589. The sources of discrepancy between the two rankings are that the naive ranking is based on a more crude measure of output (WP) and that it also does not take into account the contributions of other inputs. In Figure 4 (II), the naive ranking is based on the average net expected goals, the same output measure used to obtain MFEs. Nevertheless, there is still some discrepancy between the two rankings (the correlation between the two is 0.8320), suggesting that taking into account the input levels is also important, as well as measuring the output that is more robust to randomness.[10]

In both cases, managers above (below) the 45-degree lines are potentially under (over) rated relative to our ranking, which takes into account the other inputs and is more robust to randomness. The further the distance from the 45-degree line, the greater the disparity between the respective naive ranking and the ranking based on MFEs. One of the potentially most under-rated managers is Clarence Seedorf, a former Deportivo (Spanish *LaLiga*) manager. He is ranked 270th and 79th out of the 322 managers in the naive rankings in terms of WP and xG differentials but is ranked 26th in the MFE ranking. The disparity between his WP and MFE rankings is more prominent than that between xG differential and MFE rankings, therefore,

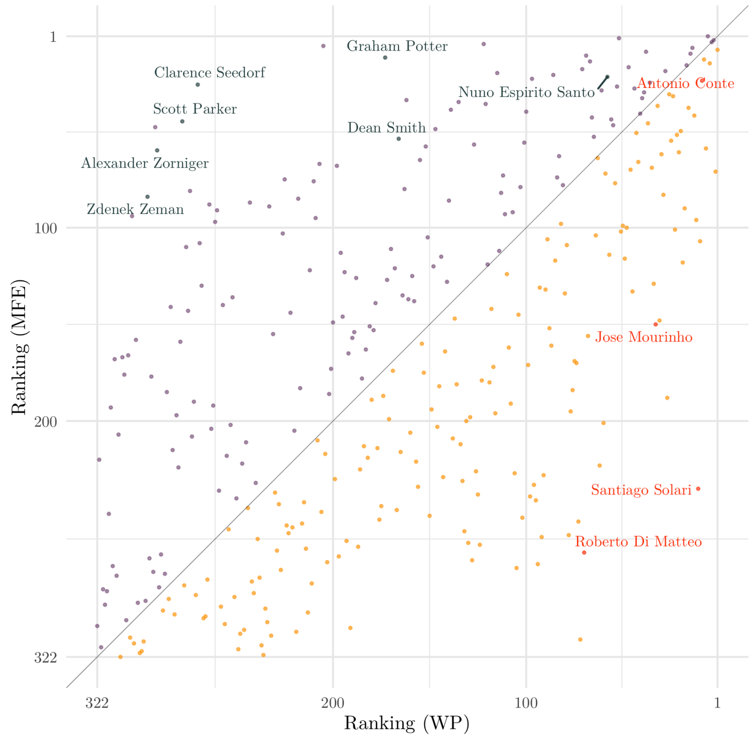
[10]The correlation between WP and MFE and that between d_xG and MFE in actual values (as opposed to rankings) are 0.6130 and 0.8164, respectively.

he can be considered as one of the "unlucky" managers, whose performance was not necessarily reflected in match outcomes. Similarly, Graham Potter (Brighton & Hove Albion F.C., England) is ranked a lot higher in the MFE ranking (12th), than the WP and xG differential rankings (173rd and 62nd, respectively). Given the resources available, his performance is notable, being just below the top ten managers.

On the other hand, our estimation suggests that some managers are over-rated when relying on naive measures of a manager's performance. For instance, Santiago Solari (Real Madrid, Spain) and Roberto Di Matteo (Schalke, Germany) are ranked in the top quartile in terms of the winning percentage but in the 3rd and the bottom quartile, respectively, in the MFE ranking.

As seen in the examples above, when there is an apparent disparity in rankings between one naive measure and MFE, it is usually the case also between the other naive measure and MFE. For instance, if a manager is ranked much higher in the WP table than they are in the MFE table, they also tend to be ranked higher in the net xG table than they are in the MFE table. One exception is Alexander Zorniger (VfB Stuttgart, '15-'16), who is ranked 291st and 27th in the WP and net xG tables, whilst his ranking is 60th in terms of MFE. In this case, therefore, he may be under-rated in the crude measure of match outcome, but the chances created with the given squad and financial source are rather over-rated.

These examples show that consideration of performance that is more robust to randomness and disentangling a manager's contribution from that of other inputs are both relevant in comparing managers.



(I) WP v. manager FE

(II) d_xG v. manager FE

Figure 4: Naive rankings and manager fixed effects

5.5 Case study

To put the results into perspective, we compare expected points (xP) obtained by a club under different managers. In doing so, we first obtain fitted values for d_log_xG with an actual manager and alternative managers based on our estimation results presented in Table 8 (column (4)) and Table 10. Then, we convert these values into expected points and observe how they are accumulated throughout a certain season.

To obtain xP, we first establish the relationship between the probabilities of each outcome (home win, draw, and away win) and the log-difference of xGs (d_log_xG). Therefore, we estimate the ordered logit model with outcome variable y where $y = 0$ for away win, $y = 1$ for draw, and $y = 2$ for home win and a predictor d_log_xG. Specifically, we estimate the parameters (μ_1, μ_2, β) in the following model:

$$y = \begin{cases} 0 \text{ (away win)} & \text{if } y^* \leq \mu_1, \\ 1 \text{ (draw)} & \text{if } \mu_1 < y^* \leq \mu_2, \\ 2 \text{ (home win)} & \text{if } \mu_2 < y^*, \end{cases} \quad \text{where } y^* = \beta \, \text{d\_log\_xG} + \varepsilon. \tag{10}$$

Our estimation results are presented in Table 11.

Table 11: Estimated parameters for ordered logit model

	Dependent variable:
	y
d_log_xG	0.981***
	(0.022)
0 | 1	-0.700***
	(0.021)
1 | 2	0.528***
	(0.020)
Observations	12,668
Note:	*p<0.1; **p<0.05; ***p<0.01

Given this, we discuss a case study of Everton F.C. (English *Premier League*) in Season 2020/2021. In the preceding season, former manager Marco Silva was replaced by Carlo Ancelotti in the first half of the season, and the club finished the season with accumulated points of 49 and is ranked 12th in the league. Carlo Ancelotti was in charge for the following season, where the club obtained 59 points overall and went up the table by two places. At the end of the season, however, he left the club, resulting in another managerial change for the club. A new manager Rafael Benítez was introduced to the club prior to Season 2021/2022.

To see the impacts of different managers on expected points as well as season outcome, we compare the predicted performance of Everton F.C. for Season 2020/2021 fixtures for the following managers: Carlo Ancelotti, Gary Neville, Graham Potter, Josep Guardiola, Marco Silva, and Rafael Benitez. For each of these managers we obtain the fitted values for the log-difference of xGs (d_log_xG), given the quality of players and financial strengths, those of opponents, and home advantage, if any. Then, convert this into expected points using the estimated ordered logit model parameters in Table 11. The accumulated expected points predicted to be obtained throughout the season under different managers are depicted in Figure 5.

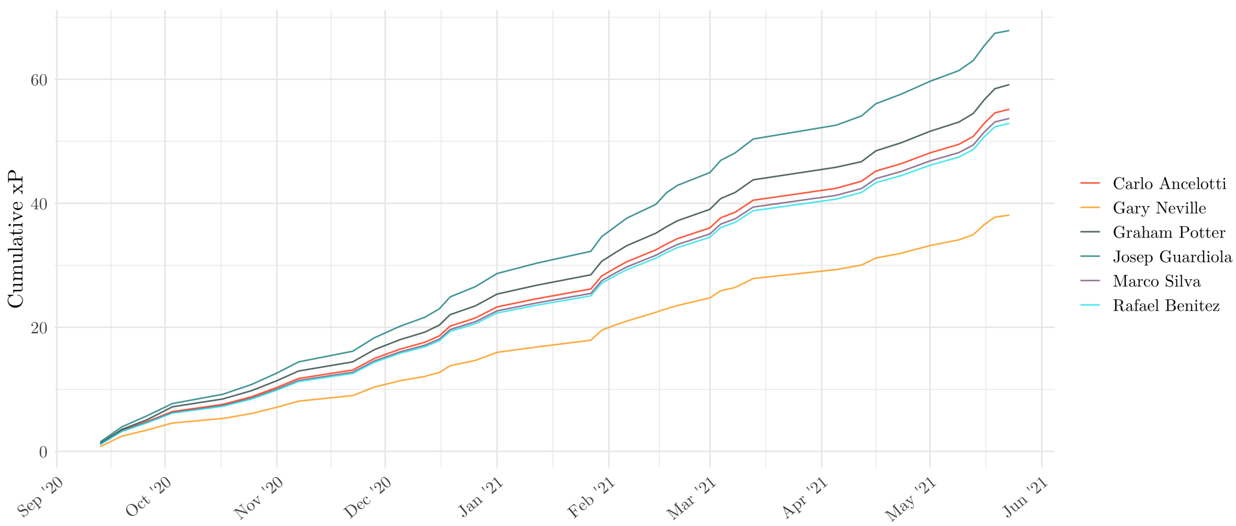


Figure 5: Cumulative expected points (xP) for Everton F.C. (2020/2021)

The Figure provides further insight into the implications of different managers, who are ranked quite differently as in Table 10, on the seasonal outcome. In terms of expected points, as opposed to actual points, Carlo Ancelotti, who was ranked 38th in our manager ranking, gained 55.16 points, which places the club the 6th in the league. Recall that the club actually finished with 12th place in Season 2020/2021, implying that the club might have had a relatively more unlucky season than the other clubs might have.

How well would the club have performed if Josep Guardiola, who is identified as the best manager in terms of contribution, had been in charge instead of Carlo Ancelotti, all other things being equal? The club would have accumulated the expected points of 67.85, and it would have ended up in the top three in the league, resulting in qualification for the UEFA Champions League. With Graham Potter, the club would have qualified for the UEFA Europa League and performed slightly better than what would have been expected under Carlo Ancelotti. Marco Silva and Rafael Benitez, who were ranked 58th and 74th in the manager ranking, respectively, would have led to the club just the top half of the league table (8th and 9th, respectively). The club could have been relegated with managers found towards the bottom of our ranking. For instance, under Gary Neville, its expected points would have been 38.09, which would have resulted in finishing the season second to the bottom of the league table. Overall, our analysis shows that the seasonal outcome could have been drastically different under a different manager, all other things being equal.

6 Conclusion

Consistent with previous studies, our analysis has shown that managers do matter in explaining firm production. The evidence we present here is arguably even more compelling since the overall significance of managers is still present even after explicitly controlling for other inputs by utilising data from professional football leagues. In addition, our analysis highlights the importance of taking into account the randomness of outcome and resources available at the hand of a manager to evaluate managerial contributions to a club's performance. In particular, we demonstrate how the ranking of football managers can differ with and without these considerations. Therefore, relying on a "naive" measure of managerial performance, such as a winning percentage, can be costly to a club since it may overestimate or underestimate a manager's contributions.

An individual manager's coefficient estimated in our model can be interpreted as the manager's match day contribution to a club's performance, given the quality of a squad and financial strengths on the day. This is useful information since the level of players and financial strength can vary over time and reflect on a manager's performance, and we, therefore, believe that this is a well-founded way of comparing different managers. In addition, this paper expands the previous studies by comparing managers observed in different leagues, which can be particularly useful for clubs seeking a manager.

Nevertheless, there are some limitations to note. Firstly, individual manager effects are effectively held constant at the average level over the past seven seasons (2014/2015-2020/2021). Therefore, this does not provide a full picture of how a manager's ability may have changed over time. It is also likely that the more recent observations carry higher predictive power, hence weighting these observations equally may not be optimal.

Secondly, our measure of managerial contribution reflects their ability to produce outcomes, given the quality of players on the day. Therefore, this captures their ability related to tactics, substitution decisions, and motivational role on the day. Although we argue that this is a valid way to evaluate a manager, it does not capture other important roles as a manager. For instance, good managers would contribute to players' growth over their career, not necessarily just bringing the best out of the player on a particular match day. Therefore, one could evaluate managers in terms of how they influenced individual players' performance over time. This could be done, for instance, using the historical performance measure employed in this study, which is allowed to vary over time.

References

Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, 118(4):1169–1208.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? Evidence from India. *Quarterly Journal of Economics*, 128(1):1–51.

Bloom, N., Lemos, R., Sadun, R., Scur, D., and Reenen, J. V. (2014). The new empirical economics of management. *Journal of the European Economic Association*, 12(4):835–876.

Boshnakov, G., Kharrat, T., and McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2):458–466.

Brechot, M. and Flepp, R. (2020). Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *Journal of Sports Economics*, 21(4):335–362.

Brian, G. (2013). Contributions of Managerial Levels: Comparing MLB and NFL. *Managerial and Decision Economics*, 34(6):428–436.

Buzzacchi, L., Caviggioli, F., Milone, F. L., and Scotti, D. (2021). Impact and Efficiency Ranking of Football Managers in the Italian Serie A: Sport and Financial Performance. *Journal of Sports Economics*, 22(7):744–776.

Dawson, P., Dobson, S., and Gerrard, B. (2000). Stochastic Frontiers and the Temporal Structure of Managerial Efficiency in English Soccer. *Journal of Sports Economics*, 1(4):341–362.

Day, D. V., Gordon, S., and Fink, C. (2012). The sporting life: Exploring organizations through the lens of sport. *Academy of Management Annals*, 6(1):397–433.

Decroos, T., Bransen, L., Haaren, J. V., and Davis, J. (2020). VAEP: An Objective Approach to Valuing On-the-Ball Actions in Soccer. *In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 4696–4700.

Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 46(2):265–280.

Flepp, R. and Franck, E. (2021). The performance effects of wise and unwise managerial dismissals. *Economic Inquiry*, 59(1):186–198.

Frick, B. and Simmons, R. (2008). The impact of managerial quality on organizational performance: Evidence from German soccer. *Managerial and Decision Economics*, 29(7):593–600.

Graham, J. R., Li, S., and Qiu, J. (2012). Managerial attributes and executive compensation. *Review of Financial Studies*, 25(1):144–186.

Hendricks, K. B., Hora, M., and Singhal, V. R. (2015). An empirical investigation on the appointments of supply chain and operations management executives. *Management Science*, 61(7):1562–1583.

Hofler, R. A. and Payne, J. E. (2006). Efficiency in the National Basketball Association: A Stochastic Frontier Approach with Panel Data. *Managerial and Decision Economics*, 27(4):279–285.

Kahn, L. M. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives*, 14(3):75–94.

Kharrat, T., McHale, I. G., and Peña, J. L. (2020). Plus–minus player ratings for soccer. *European Journal of Operational Research*, 283(2):726–736.

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics*, 33(4).

Lefgren, L., Platt, B., and Price, J. (2015). Sticking with what (Barely) worked: A test of outcome bias. *Management Science*, 61(5):1121–1136.

Muehlheusser, G., Schneemann, S., Sliwka, D., and Wallmeier, N. (2018). The Contribution of Managers to Organizational Success: Evidence from German Soccer. *Journal of Sports Economics*, 19(6):786–819.

Papadopoulos, A. (2021). The Effects of Management on Production: A Survey of Empirical Studies. In *Handbook of Production Economics*, pages 1–47. Springer, Singapore.

Peeters, T., Szymanski, S., Fumagalli, C., and Thomas, C. (2014). Financial fair play in European football. *Economic policy*, 29(78):343–390.

Peeters, T. L., Salaga, S., and Juravich, M. (2020). Matching and Winning? The Impact of Upper and Middle Managers on Firm Performance in Major League Baseball. *Management Science*, 66(6):2735–2751.

Pieper, J., Nüesch, S., and Franck, E. (2014). How performance expectations affect managerial replacement decisions. *Schmalenbach Business Review*, 66:5–26.

Rowe, W. G., Cannella, A. A., Rankin, D., and Gorman, D. (2005). Leader succession and organizational performance: Integrating the common-sense, ritual scapegoating, and vicious-circle succession theories. *The Leadership Quarterly*, 16(2):197–219.

Schepker, D. J., Kim, Y., Patel, P. C., Thatcher, S. M., and Campion, M. C. (2017). CEO succession, strategic change, and post-succession performance: A meta-analysis. *The Leadership Quarterly*, 28(6):701–720.

Siebert, W. S. and Zubanov, N. (2010). Management economics in a large retail company. *Management Science*, 56(8):1398–1414.

Tena, J. D. and Forrest, D. (2007). Within-season dismissal of football coaches: Statistical analysis of causes and consequences. *European Journal of Operational Research*, 181(1):362–373.

van Ours, J. C. and van Tuijl, M. A. (2016). In-season head-coach dismissals and the performance of professional football teams. *Economic Inquiry*, 54(1):591–604.