Working Paper in Economics

202226

Causal and Consequences of Multiple Dismissals: Evidence from Italian Football League

Kaori Narita
J.D. Tena
Babatunde Buraimo

Causes and Consequences of Multiple Dismissals: Evidence from Italian Football League

Kaori Narita* J .D. Tena† Babatunde Buraimo‡

November 12, 2022

Abstract

Previous research in leadership succession focuses on establishing whether such an event has a positive impact on the subsequent performance of an organisation. However, factors that can affect the effectiveness of leadership change are not well understood. The aim of this study is to identify the causes of first and second within-season head coach dismissals and estimate the impact of the two types of dismissal on field performance using data from the Italian professional football league (*Serie A*). We employ inverse propensity score weighting together with machine learning techniques in order to mitigate selection bias. Our analysis shows that the determinants of the two decisions are not identical in that the second replacement is likely to be taken with greater caution. Despite this, we find some positive effects of first dismissals on subsequent performance, whilst the second dismissals do not appear to make any difference. These findings suggest that frequent changes in leadership are not favourable options even when a recent replacement has not improved the situation. This could be because the potential benefit of leadership replacement may be counteracted by disruptive effects, or a source of underperformance may lie elsewhere rather than a manager.

Keywords: leadership succession, machine learning, inverse propensity score weighting, football managers

JEL Codes: J63, M51, Z22

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

†Email: jtena@liverpool.ac.uk. Address: University of Liverpool Management School, Liverpool, L69 7ZH, United Kingdom.

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

1 Introduction

Leadership succession is a critical event in any organisation, be it a political party, a corporation, or a sports club. As such, this issue has been intensively studied in many contexts. For instance, as is evident from the reviews by Giambatista et al. (2005) and Berns and Klarner (2017), CEO successions in publicly traded companies have been an active research area in management. Farah et al. (2020) also provides a review of studies related to leadership succession in different contexts, such as privately-owned businesses and political organisations. In addition, professional sports clubs have been a popular field to explore the causes and consequences of leadership succession, where the researchers have examined the impact of replacing a manager or a head coach on a club's performance (Rowe et al., 2005).

This study falls into the last category in that we will provide further evidence of the causes and consequences of head coach dismissals using data from the Italian football league, although our findings can be generalised to broader contexts of leadership succession. The previous studies in the domain of professional football focus on establishing whether replacing a manager improves team performance, however, there is little guidance on how to make such a decision. A recent study by Flepp and Franck (2021) fills this gap by providing evidence that the consequence of dismissal depends on whether the decision was subject to misperceptions of performance. In particular, the study indicates that to implement managerial changes successfully, one needs to identify whether the poor performance is due to low manager ability or bad luck.

To better understand the conditions under which a managerial dismissal can bring about a favourable outcome, we analyse the decision of replacing a manager more than once within a season. In particular, we separately identify the determinants and estimate the impacts of the two types of managerial dismissals: (1) first dismissals that occur in a given season (*single dismissal*), and (2) second dismissals that happen following the first one within the same season (*multiple dismissal*). Such an analysis is economically relevant in at least two ways. First, it helps us to understand how a principal (club) may adjust their expectation with respect to the agent (manager)'s contribution to the productivity of working teams, given an (unfavourable) outcome of the first replacement. Second, it allows us to understand how first and second dismissals are operationalised in a causal analysis. Given that these two decisions could be motivated by different factors, they could also affect team performance differently.

Our empirical study is based on observational data from the top-tier Italian football league (*Serie A*) for the seasons from 2004/2005 to 2017/2018, where we observe 114 and 36 cases of single and multiple dismissals, respectively. Firstly, we identify the predictors of each type of dismissal by forming appropriate control and treatment groups for single and multiple dismissals and employing classification models. In addition to logistic regression, which is normally employed in this context, we apply a machine learning algorithm, gradient boosting machine (GBM), to identify the most influential variables.

Whilst one can assume that a manager is likely to be dismissed when a club is performing poorly, whether it is the first or second time in a given season, the set of specific predictors for each event could be different. For instance, if a club has already replaced a manager yet the situation has not improved, they may revise their belief on a manager's role in the adverse outcomes; hence the decision to replace yet another manager may be made with greater caution.

Allowing for the possible differences in the set of predictors between the single/multiple dismissals is also important in order to take into account the pre-treatment differences between the treated and control groups for the respective type of dismissal. A decision to dismiss a manager, like many other managerial decisions, does not occur randomly. For instance, such decisions are more likely to be made when a club is performing poorly. This means that one cannot simply attribute post-treatment differences to the treatment effects of dismissal decisions, without taking into account the differences in the pre-treatment status that could also affect outcomes. Therefore, to estimate the average treatment effects (ATE) of single and multiple dismissals, we employ the inverse propensity score weighting (PSW) method, where such differences in pre-treatment characteristics are taken into account through the weighting based on the propensity score, i.e. the predicted probability of being treated.

Our main findings are the following. First, we find that the predictors of dismissals can depend on the situation, specifically, whether a club has dismissed a manager in the specific season or not. For instance, a rather crude measure of performance such as the average points obtained in recent matches can influence a decision to dismiss a manager for the first time in a given season, as much as relative performance against expectations can. On the other hand, our results suggest that the second within-season dismissal is mostly influenced by relative performance. In addition, whilst the threat of relegation can increase the probability of the first dismissals, it does not appear to affect the second dismissal decision. It is also shown that different sets of managerial characteristics can influence the two types of dismissal. Our study also finds that the consequences of single and multiple managerial changes can be different. In particular, whilst we find a limited boost in club performance after the first managerial change, the second change seems to have no impact on club performance. Therefore, if a club has not dismissed a manager in a given season, making changes could be beneficial, although such positive effects could be short-lived. On the other hand, should a club have changed a manager in a given season, going through yet another change does not make any difference.

The remainder of the paper is structured as follows. Section 2 reviews related studies and offers our hypotheses regarding the causes and consequences of single and multiple dismissals. In Section 3, we describe our data set and variables used in the analysis. Section 4 then describes relevant empirical challenges and our methodology. We present our results in Section 5 and conclude our study in Section 6.

2 Related literature and hypotheses

2.1 Causes of dismissal

The relationship between a football club and a head coach can be characterised as a principal-agent (PA) model, just as that of a firm owner and a CEO (Desai et al., 2018). Club owners invest in capital and pay workers, including a head coach and players, to maximise performance. As a sports club, output that will generate monetary revenue is (favourable) performance on the field. A club, therefore, hires a head coach and delegates responsibilities for field management to him/her, who is specialised in training a squad, creating a tactic, and game management. As is commonly the case in any PA relationship, the delegation of tasks entails asymmetric information. In our example, from the point of view of a club owner, who has inferior information in terms of sporting aspects, the effort level of a head coach is unknown since such input is neither directly measurable nor observable. Furthermore, hidden information arises since the contribution of a head coach towards production (match outcome) cannot be easily disentangled from that of players.[1]

Incentive theory suggests that the threat of dismissal improves effort by workers under an environment characterised with asymmetric information, as documented in Kwon (2005) and Sparks (1986). In particular, the use of dismissal threat in compensation of top management such as CEOs is investigated by Wang et al. (2017), Jensen and Murphy (1990a) and Jensen and Murphy (1990b). Indeed, involuntary dismissal of a head coach is a common practice in the world of professional sports. As a club owner, his/her interest is to maximise a manager's effort given the cost that entails such incentive design (firing and hiring a manager). The decision is based on a signal or proxy of managerial contribution, such as realised match outcomes. A head coach, in turn, exerts effort to maximise the probability of winning (hence minimising the probability of being dismissed) with a given set of sub-ordinates.

As shown in the previous studies on the causes of managerial turnover, a football manager's tenure is, in fact, heavily dependent on field performance. The most common factors of dismissals are: falling short of expected performance (van Ours and van Tuijl, 2016; Bryson et al., 2021b), a streak of unfavourable match outcomes (D'Addona and Kind, 2014; van Ours and van Tuijl, 2016), and a threat of relegation (Tena and Forrest, 2007). Although it is suggested to be of less importance, some studies also find managerial

[1] There is a strand of literature which looks into the contribution of a manager to organisational success. A study of the largest U.S. firms by Bertrand and Schoar (2003) suggests that a wide range of managerial practices is affected by individual manager effects. In the sports domain, Muehlheusser et al. (2018) and Peeters et al. (2020) provide further evidence that a firm's productivity depends on the quality of an individual manager after controlling for individual firm characteristics. In addition, Frick and Simmons (2008) show that relative coach salaries have a significant impact on team efficiency. Therefore, it is evident that a manager can affect a firm's success. However, imperfect information remains in terms of the extent to which a manager is responsible for a firm's performance.

characteristics can affect the likelihood of dismissal. For instance, Bryson et al. (2021b) find that a less experienced manager is more likely to be dismissed, after controlling for the recent performance. As is set out in the introduction, one of our objectives in this paper is to identify the causes and consequences of a first within-season replacement (single dismissal) and dismissal of a new manager following the first replacement (multiple dismissal). To the best of our knowledge, this issue has not previously been investigated. We conjecture, however, that the causes of these two types of dismissals may well be different, as we elaborate below.

Traditional theories around managerial dismissal (Grusky, 1963; Gamson and Scotch, 1964) suggest that performance should improve following dismissal *if* an incumbent is to blame for poor performance. When a club dismisses a manager, therefore, the consequence of such an event reveals some information regarding the factors contributing to the underachievement. If a post-dismissal performance bounces back, this may indicate that a dismissed manager was indeed responsible for poor performance. On the other hand, if the firm continues to stagnate, the problem may well lie out of a head coach's control, hence they may seek an alternative solution.

Yet, theoretically speaking, a new manager resulting from recent turnover is now under the watchful eyes of a board, and he/she can be fired anytime based on realised match outcomes. Particularly, in the context of professional sport, the majority of variables (e.g. playing talent and club's finance) are fixed or at most adjustable with limitations within a season. Therefore, a head coach may often be replaced with the hope of turning a situation around, and this sometimes happens multiple times within a season. Given the discussion above, however, a club may revise their belief about a manager's role in adverse outcomes based on an outcome of the recent turnover. If a recent dismissal turned out to be unsuccessful, the decision to replace yet another manager might be made with greater caution. Accordingly, we test the following hypothesis regarding the determinants of the single and multiple within-season dismissals:

Hypothesis 1 (H1): *Causes of single and multiple dismissals are not identical, in that multiple dismissal is considered with greater caution.*

Studying this issue is relevant since factors that lead to a certain decision could affect the consequences of such a decision, and therefore they need to be taken into account when evaluating their effectiveness. The relevance is not limited to the methodological point of view; this would add to the literature by providing some empirical evidence on how the outcome of the recent dismissal can influence the motivating factors to dismiss yet another manager within a short period of time.

2.2 Consequences of dismissal

Needless to say, one hopes leadership change (as a result of dismissal) will improve club performance. However, most of the previous studies suggest that there are no causal effects of managerial replacement on performance. For instance, Scelles et al. (2020) and van Ours and van Tuijl (2016) estimate the consequences of a managerial change in the French *Ligue 1* and the Dutch *Eredivisie*, respectively. In these studies, the control group is formed with counterfactual observations that followed a similar path of performance against expectation, yet did not replace their manager. Considering all the remaining matches following an actual and counterfactual turnover as treated, they identified a statistically significant improvement in both actual and counterfactual cases. Therefore, these studies suggest that the seemingly positive effect of replacement is, in fact, due to regression to the mean. Similarly, ter Weel (2011) and Bruinshoofd and Ter Weel (2003) compare the subsequent performance of clubs that replaced their head coach with a control group formed by untreated clubs with similar previous performance in the four post-turnover matches. Results in these papers do not support the hypothesis that a managerial change improves performance.

However, the studies cited above, among others, primarily focus on establishing whether managerial replacements are effective in improving club performance. Knowing specific circumstances or moderating factors that may affect the consequences of dismissals would be useful for decision-makers. Muehlheusser et al. (2016) and Flepp and Franck (2021) fill this gap by identifying some conditions under which a dismissal can bring about favourable outcomes. Muehlheusser et al. (2016) find that a dismissal can positively influence club performance, given that a club is homogeneous, i.e. the ability of the top and bottom players within the squad is rather similar.[2] Flepp and Franck (2021) also offer further insights into the consequences of managerial dismissals by distinguishing "wise" and "unwise" cases, where the former is led by actual poor performance on the pitch and the latter follows seemingly poor performance due to bad luck. They find that "wise" dismissals can improve performance, whilst "unwise" dismissals have no impact.

This study aims to contribute to understanding factors that may affect the consequence of managerial dismissal by distinguishing single and multiple dismissals as described above. Whilst no previous study compares the consequences of the first and second within-season managerial replacements, there are some studies in other professional sports and organisations that indicate that the frequency of leadership succession can affect the effectiveness of such events. Boyne and Dahya (2002) and Gordon and Rosen (1981)'s hypothesis predicts a negative relationship between the frequency of executive succession and its consequence. The latter states that "too many managerial replacements in too brief a period can be disruptive to a unit" (Gordon and Rosen, 1981, p.238).

[2]The authors suggest this is due to more intense competition among players to impress a new manager.

Consistent with their theories, Kim et al. (2021) find that high frequencies of CEO turnover have a negative impact on future performance in Korean firms. Khaliq et al. (2006) study CEO turnover in hospitals and conclude that frequent managerial turnover can significantly reduce the morale of the workforce. In sports literature, Hill (2009) finds a non-linear relationship between the number of occurrences of managerial turnover and its impact on field performance in professional baseball teams. The latter study suggests that a moderate incidence of managerial change could be successful, whilst a high frequency of such events entails detrimental effects. It can be argued that constant managerial change can send an adverse signal to workers, undermining their confidence in the firm and their leader, which in turn hinders a new manager from implementing their strategies effectively.

Given the discussion above, one can argue that the first and second within-season dismissals can have quite different impacts on subsequent club performance. In particular, our hypothesis is as follows:

Hypothesis 2 (H2): *Consequences of single and multiple dismissals are not identical, in that single dismissal can have a positive impact, whilst multiple dismissal is detrimental to firm performance, or at most ineffective.*

3 Data

In order to test our hypotheses, we collected club-match level data from the top-tier Italian football league (*Serie A*) that took place over the seasons from 2004/2005 to 2017/2018. During this period, the 20 most competitive clubs in the country participated in the tournament each season. Akin to most of the domestic professional football leagues in the world, the competition operates in a format of a round-robin tournament. Hence, each competitor meets each other club twice during a campaign, once at its home stadium and once at the opponent's (away) stadium. Therefore, for any given season s, club i appears in each game week t, where $s \in \{1, ..., 14\}$, $i \in \{1, ..., 20\}$, and $t \in \{1, ..., 38\}$. This gives a total number of observations $N = 10,640$.

For each match, three points are awarded to a winner, none to a loser, and one to each competitor in case of a draw. At the end of a campaign, a club with the highest accumulated points receives the championship title, while the bottom three clubs are demoted to the second tier league (*Serie B*), and replaced by the three most competitive clubs promoted from *Serie B*.

Our data set is suitable for the analysis of managerial turnover since the role of a head coach, whose tenure can be terminated at any point during a season by a club owner, is akin to that of a top manager in a corporation (Pieper et al., 2014). In particular, the fact that some clubs replaced their manager multiple times in a given season facilitates testing for the possible differences in causes and consequences of single and multiple dismissals. Moreover, such events are publicly observable with precise timing. This allows us to fairly specifically associate a match outcome with a particular manager's contribution.

3.1 The treatment variable and treatment groups

The data set indicates a club i's manager who was in charge at any specific time, i.e. a game week t in a particular season s, and recorded as the variable $\textit{Head coach}_{its}$. Among other variables that are related to match statistics, this information was obtained from the official website of *Serie A TIM*, which provides a match report for the individual matches. Based on this, we define our treatment variable $\textit{New coach}_{its}$ as follows.

$$\textit{New coach}_{its} = \begin{cases} 0 & \text{if } \textit{Head coach}_{i,t,s} = \textit{Head coach}_{i,(t-1),s}, \\ 1 & \text{otherwise.} \end{cases} \tag{1}$$

Therefore, a club-match observation at time t is deemed to be treated ($\textit{New coach}_{its} = 1$), when club i dismisses its manager after a match $t-1$ in a season s, followed by a new coach in charge from a match t onwards. In order to make sure any managerial replacement considered in the analysis is due to an involuntary departure of a manager dismissed by a club rather than a voluntary decision or mutual consent between the two parties, we consulted the archives from the official website of the league and individual clubs, as well as the two most read national sports newspapers in Italy, *Corriere dello Sport-Stadio* and *La Gazetta dello Sport*.[3] Furthermore, in case a caretaker manager is introduced between the departure of the outgoing manager and the appointment of a replacement, matches overseen by the interim head coach are excluded from our analysis.[4]

Before presenting descriptive statistics, we form different treatment groups in the following fashion. First, we split the sample into two sub-samples (Sample I and Sample II), for the two treatments, Treatment I and Treatment II, defined as follows. Sample I consists of club-match observations where a match $t-1$ does not lead to managerial dismissal ($\textit{New coach}_{its} = 0$) or lead to dismissal ($\textit{New coach}_{its} = 1$), given that a club i has not replaced their manager in Season s. Thus Treatment I refers to a dismissal that happens for the first time during Season s. Sample II, on the other hand, is represented by clubs which have already replaced their manager in Season s, and either retain a new manager who replaced a first manager ($\textit{New coach}_{its} = 0$), or replace them with yet another new manager after match $t-1$ ($\textit{New coach}_{its} = 1$). Therefore, Treatment II is a second dismissal that happens during season s.[5] Thus, the average treatment

[3] We identified 15 cases of voluntary departures in the 14 seasons analysed in this study. As Bryson et al. (2021b) and Bryson et al. (2021a) show, the motivations and outcomes of quits and dismissals are distinct from one another.

[4] During the relevant seasons, eight individuals served as caretakers amid a transition from an outgoing to an incoming manager.

[5] There are very few cases where a club dismisses a manager more than twice in a given season. We do not consider these cases since it is neither common enough to be of interest nor easy to form a counterfactual control group to evaluate the treatment effect.

effects (ATE) of each treatment are obtained by comparing the post-treatment performance between the treated and control groups within the respective sub-samples. Finally, we refer to treated and control groups in Sample I as Treated I and Control I, respectively, and those in Sample II as Treated II and Control II.[6]

Table 1 summarises the number of club-match observations that led to single and multiple dismissals, and the number of corresponding control observations that did not lead to the respective type of dismissal. During the 14 seasons, we observe a total of 114 single turnovers and 36 multiple turnovers. That is, on average, more than 8 out of 20 clubs dismissed their managers at least once during a season, and almost three clubs go through managerial dismissals twice within a season on average. Both single and multiple treatments are observed in every season in our sample. For instance, AFC Fiorentina underwent managerial change twice during their 2004/2005 campaign. After seven matches in the season, Emiliano Mondonico was dismissed and replaced by Sergio Buso. He was then dismissed after 13 matches, and a new manager Dino Zoff was appointed and was in charge for the rest of the season. Another example is Inter Milan in Season 2011/2012, where we observe the first turnover when Gian Piero Gasperini was dismissed and replaced by Claudio Ranieri after just three matches of the season, and the second time in the dismissal of Ranieri, whose role was taken over by Andrea Stramaccioni, the then vice coach.

However, there are many cases, 16 out of 36 cases to be precise, where a dismissed manager in the first dismissal is re-hired following the second within-season dismissal. For instance, Massimo Ficcadeni was in charge of Cagliari Calcio before the club fired him after the first ten matches of Season 2011/2012. His replacement, Davide Ballardini, was eventually dismissed after 17 matches, which resulted in the re-hiring of Massimo Ficcadeni. The reasons behind the re-hiring of a dismissed manager are likely to be related to unique rules regarding head coach dismissals in *Serie A*. A head coach who is dismissed by a club is not allowed to be hired by another club in the division for the remaining of the season. Meanwhile, a club is obliged to pay the dismissed manager until the end of the season. This implies that the dismissed coach is likely to be available, and the club may have the incentive to bring him back since he continues to be paid regardless. The existence of the unusual rule in Italy could potentially limit the generalisability of our results in terms of the likelihood of re-hiring the dismissed coach following the second dismissal.[7]

[6] In this context, treatment decisions are sequential in the sense that the second managerial change can only happen given the first managerial change has occurred. Therefore, these events cannot be simultaneously captured in a multi-level treatment variable.

[7] However, it is impractical to distinguish the two categories of second dismissal because of low numbers in each category. Therefore, we recommend that future research examines the issue in other countries.

Table 1: Number of treated and control units by seasons

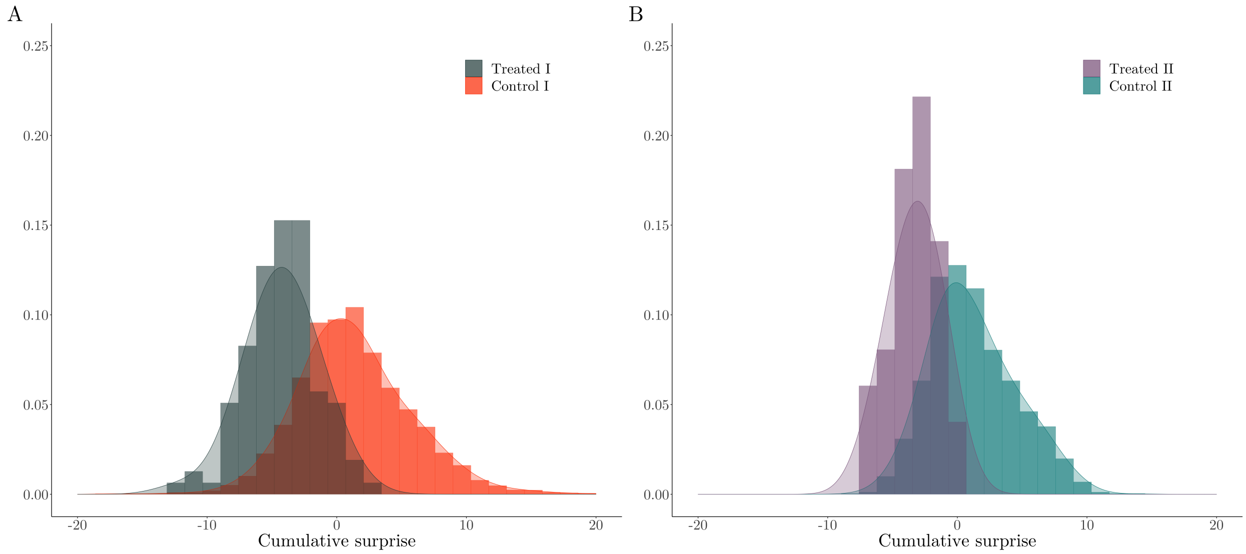
	Treated I, N = 114	**Control I**, N = 7,890	**Treated II**, N = 36	**Control II**, N = 1,903
Season				
2004/2005	7	604	1	111
2005/2006	7	588	4	81
2006/2007	9	578	3	128
2007/2008	8	562	5	118
2008/2009	8	578	3	120
2009/2010	11	480	3	203
2010/2011	9	579	2	128
2011/2012	10	478	4	210
2012/2013	8	543	2	157
2013/2014	10	524	3	170
2014/2015	5	623	1	100
2015/2016	9	563	2	122
2016/2017	5	631	2	86
2017/2018	8	559	1	169

3.2 Determinants of dismissals

An identification issue that typically arises in analyses of observational data is the endogeneity of treatment assignment. That is, the treatment is not randomly assigned to individuals, hence there are fundamental differences between those sorted to a treated group and those sorted to a control group. Following the previous subsection, we have two types of treatment: the first dismissal that happens within a season (single treatment) and the new manager dismissal after the first dismissal that occurs within the same season (multiple dismissal). Therefore, identifying the determinants of the two treatments is important from the methodological point of view, as further described in Section 4. At the same time, our aim is to test the possible differences in the factors leading to these two types of dismissals.

Whilst there is no previous study which has differentiated the causes of the two types of treatments discussed above, there is ample literature that examines the causes of dismissals in general. For instance, using the data from the top-tier football league in the Netherlands (*Eredivisie*) covering the seasons 2000/2001-2013/2014, van Ours and van Tuijl (2016) show that within-season dismissals are most likely to occur at clubs which accumulated negative "match surprise." Match surprise refers to unexpected outcomes and is captured by the difference between actual and *ex-ante* expected outcomes, where the latter can be measured with, for example, betting odds. To illustrate the pre-treatment differences with respect to this variable,

Figure 1: Kernel density of *Cumulative surprise* in treated and control groups



Notes: Each figure plots cumulative surprise on the x-axis and density on the y-axis in Sample I (Panel A) and Sample II (Panel B). Each contains the densities for treated and control groups in the respective sample.

Cumulative surprise, the Kernel densities of this variable for treated and control groups for respective treatment types are presented in Figure 1. Panel A compares the distributions of *Cumulative surprise* for the treated and control groups for Treatment I (single dismissal), and Panel B compares those for the treated and control groups for Treatment II (multiple dismissal). For both types of treatment, the underlying differences between treated and control groups are noticeable. For both cases, cumulative surprise for the treated units is mostly distributed below zero, whilst that for the control counterpart is distributed more evenly about zero.

There are many other factors that are identified as a predictor of dismissal in the literature. The two main categories are: (1) recent performance variables and (2) characteristics of a manager.

Predictors capturing recent team performance are most commonly found in prior research. The short-term performance history measured with the most recent four matches (van Ours and van Tuijl, 2016) and within a couple of weeks prior to the current match (D'Addona and Kind, 2014) are shown to be significant predictors of involuntary departure of a manager. Unlike the cumulative surprise variable described above, however, these performance indicators are considered a rather crude measure of performance since they do not take into account expectations. Among others, Bryson et al. (2021b) provide evidence that league position can also determine the likelihood of dismissal. Furthermore, Tena and Forrest (2007) find that the threat of demotion is the most common factor in dismissal decisions. Whilst performance indicators employed tend to vary across the studies, all the evidence points to a negative relationship between recent performance and the probability of dismissal.

Accordingly, in addition to the variable *Cumulative surprise* illustrated above, we examine the following performance indicators: average points obtained in the most recent four matches (*Point last four*) and a club's current position in the league (*Standing*). Also included in this category are three dummy variables to indicate whether the most recent match outcome was loss (*Lost last*), the most recent match outcome was loss and took place at the club's home stadium (*Lost last home*), and whether a club is under threat of relegation (*Relegation*).

The second set of variables that could affect the likelihood of dismissals is the characteristics of a manager. Whilst existing evidence is not as consistent as the one for recent performance, some studies suggest specific managerial characteristics can also affect the chances of dismissal. For instance, Audas et al. (1999) show that a manager's age can influence the survival rate of football managers, whilst others (Frick et al., 2010; van Ours and van Tuijl, 2016) find significant effects of their previous experience. Other managerial characteristics such as their nationality and background have also been tested. Gilfix et al. (2020), for instance, provides some evidence that domestic managers are less likely to be dismissed relative to those who are not nationals of the country.

In our analysis, therefore, we examine a set of managerial characteristics that could potentially affect the probability of dismissal. The information was obtained from various sources, with Tranfermarkt (*trarnsfermarket.com*) being our primary source. Firstly, a manager's general and industry-specific human capital are measured with the number of years of experience (*Experience years*) and a dummy variable to indicate whether a manager has previously worked in *Serie A* (*Experience Serie A*), respectively. We also add the two dummy variables associated with managerial experience, one to indicate a new entrant to the labour market (*No experience*),[8] and the other to indicate a manager's experience in professional leagues abroad (*Experience abroad*). To take into account a manager's recent participation in the professional football manager labour market, we examine indicators of activity status in the most recent season (*Active*), as well as that in *Serie A* (*Active Serie A*) in our models.

Secondly, a set of dichotomous variables regarding a manager's background as a player are considered. Specifically, they indicate whether a head coach had a playing career in professional football (*Player*) and whether that included playing in *Serie A* (*Player Serie A*). Additionally, information regarding the position played is indicated with the variable *Player GK/DF*, where *Player GK/DF* $= 1$ if a manager is a former

[8] Peeters et al. (2017) provide interesting findings regarding market failure in the football managers' labour market, where they show that an experienced manager has an excess advantage over a novice manager. By the time the firing decision is made, however, the novice manager's ability would have become public. Therefore, the mere fact that they have no previous experience as a head coach should not affect the likelihood of dismissal, after controlling for the realised performance. Nevertheless, there may be another type of hidden information, in the sense that a new manager's particular ability to cope with the adverse situation is yet to be realised. If a club were unwilling to take the risk of not being able to turn a situation around, or worse, aggravating the situation, it may be more prone to fire a novice coach than the experienced.

defender or goalkeeper and *Player GK/DF* $= 0$ otherwise.

Thirdly, association with the club is captured by dummy variables to indicate whether a manager has previously served the club as a vice coach (*Former vice*), whether he is a former player of the club (*Player club*), and whether the club is the last club with which he was a player (*Player club last*).

Finally, we examine two personal traits, nationality and age, where *Italian* takes 1 if a manager is Italian and 0 otherwise, and *Age* measures their age in years.

In addition to the two groups of predictors described above, we include a variable *Week*[9], which indicates how far it is into the season, and a variable *Days*, the number of days between match $t-1$ and match t. Table 2 provides descriptive statistics of the pre-treatment status measured by the variables described above, for the respective treated and control groups for the two types of dismissal.

Table 2: Descriptive statistics of treatment predictors

Variable	**Treated I**, N = 114	**Control I**, N = 7,890	**Treated II**, N = 36	**Control II**, N = 1,903
Cumulative surprise	-4.27 (2.75)	1.32 (4.36)	-3.27 (1.81)	1.20 (3.29)
Point last four	0.59 (0.41)	1.45 (0.76)	0.47 (0.41)	1.24 (0.73)
Lost last				
0	16 (14%)	5,196 (66%)	5 (14%)	1,121 (59%)
1	98 (86%)	2,694 (34%)	31 (86%)	782 (41%)
Lost last home				
0	71 (62%)	6,906 (88%)	20 (56%)	1,599 (84%)
1	43 (38%)	984 (12%)	16 (44%)	304 (16%)
Standing	15.51 (4.20)	9.07 (5.41)	17.33 (3.51)	14.32 (4.52)
Relegation				
0	63 (55%)	7,226 (92%)	13 (36%)	1,324 (70%)
1	51 (45%)	664 (8.4%)	23 (64%)	579 (30%)
Experience years	11.30 (7.65)	12.23 (7.59)	14.75 (8.56)	13.42 (8.90)
Experience Serie A				
0	25 (22%)	1,228 (16%)	5 (14%)	218 (11%)
1	89 (78%)	6,662 (84%)	31 (86%)	1,685 (89%)
No experience				
0	108 (95%)	7,635 (97%)	34 (94%)	1,789 (94%)
1	6 (5.3%)	255 (3.2%)	2 (5.6%)	114 (6.0%)
Experience abroad				
0	87 (76%)	5,625 (71%)	23 (64%)	1,191 (63%)
1	27 (24%)	2,265 (29%)	13 (36%)	712 (37%)
Active				
0	16 (14%)	800 (10%)	10 (28%)	412 (22%)
1	98 (86%)	7,090 (90%)	26 (72%)	1,491 (78%)
Active Serie A				
0	51 (45%)	2,385 (30%)	17 (47%)	792 (42%)
1	63 (55%)	5,505 (70%)	19 (53%)	1,111 (58%)

[9] We also include a squared term of *Week* (*Week squared*) in logistic regression specification, following Tena and Forrest (2007).

Player				
0	12 (11%)	510 (6.5%)	5 (14%)	188 (9.9%)
1	102 (89%)	7,380 (94%)	31 (86%)	1,715 (90%)
Player Serie A				
0	37 (32%)	2,707 (34%)	16 (44%)	661 (35%)
1	77 (68%)	5,183 (66%)	20 (56%)	1,242 (65%)
Player club				
0	97 (85%)	6,630 (84%)	30 (83%)	1,533 (81%)
1	17 (15%)	1,260 (16%)	6 (17%)	370 (19%)
Player club last				
0	108 (95%)	7,237 (92%)	35 (97%)	1,846 (97%)
1	6 (5.3%)	653 (8.3%)	1 (2.8%)	57 (3.0%)
Player GK/DF				
0	74 (65%)	6,067 (77%)	26 (72%)	1,447 (76%)
1	40 (35%)	1,823 (23%)	10 (28%)	456 (24%)
Former vice				
0	109 (96%)	7,659 (97%)	33 (92%)	1,794 (94%)
1	5 (4.4%)	231 (2.9%)	3 (8.3%)	109 (5.7%)
Italian				
0	12 (11%)	778 (9.9%)	2 (5.6%)	186 (9.8%)
1	102 (89%)	7,112 (90%)	34 (94%)	1,717 (90%)
Age	49.33 (6.82)	50.33 (6.75)	52.42 (6.77)	51.10 (7.43)
Days	8.46 (4.56)	7.20 (3.53)	7.39 (3.17)	7.08 (3.14)
Week	16.34 (8.49)	17.12 (10.66)	25.69 (6.54)	24.44 (8.05)

Notes: Mean (SD); Frequency (%)

3.3 Outcome and control variables

As discussed in the earlier section, a club owner delegates the tasks that particularly involve sporting aspects of the firm to a head coach, who normally possesses superior knowledge and experience in football. Sporting success is, of course, one of the primary aims of a football club since the number of wins is directly and indirectly translated into a club's revenue. Hence, evaluating the treatment effect of managerial turnover focuses on post-succession field performance. In particular, we measure the performance with average points ($Point$) and goal differences($Goal\ dif$) in the subsequent matches.

Whilst the most common practice is to use all the remaining matches in the relevant season to measure the post-treatment performance, it is the authors' discretion to decide how long a manager is to be considered as a "new" manager after their appointment. Since in principle, a club may be taking a decision over whether to retain or replace the manager at any point in the season, a "new" manager is also subject to dismissal from the day they arrive. Indeed, the focus of this study is on cases where a club replaces a manager more than once in a given season. Therefore, we allow flexibility on the time window within which matches are deemed to be treated.

More precisely, we examine the impact of managerial change on performance in the very first match after the change, as well as the average performance in the subsequent 5 matches, the subsequent 10, and for the rest of the season, or until the next managerial turnover, if any, whichever occurs first. Should there be a more limited number of matches than those exact time windows, we consider the average performance of the available matches. Hence, we denote our outcome variables with a suffix which indicates the maximum number of post-treatment matches considered to measure the performance. For instance, a variable *Point_10* is the average value of points within the subsequent 10 matches in the post-treatment period. However, should a manager be dismissed after 7 post-treatment matches, the variable takes the average points of these matches.

To control for factors that could affect match outcomes other than the treatment assignment, we construct the following control variables. First, a variable *Home* controls for home advantage measured by the proportion of the matches that took place at the home stadium out of the matches with which we measure the outcome variable. With the example of *Point_10* demonstrated above, therefore, we control for the proportion of matches held at a club's home stadium out of the 10 post-treatment matches (*Home_10*).

In addition, the ability level of the club (*Ability*) and that of opponents (*Opponent ability*) are controlled by the ability indicator constructed in the following manner. First, we take a club's final position in the league table in the preceding season, reversing the order so that, for example, the top club is assigned the value 20 (and the bottom club would be assigned the value 1). The order is reversed to ensure that the variable increases with club ability as captured by its performance in the preceding season. In cases where a club had not played in the top division in the preceding season, it is assigned the value 1 (i.e. treated as having been equivalent to the bottom club in the top tier). We obtain these values for the final positions over the past four seasons, then take the weighted average with higher weights given to the more recent seasons for each club.[10] The variable *Opponent ability* is the average value of the ability indicator for the opponents in the subsequent matches with which the outcome is measured. Again, when considering the preceding 10 matches, we take the average of strengths over the 10 corresponding opponents (*Opponent ability_10*).

4 Methodology

An important identification issue that arises in observational studies is that assignment of treatment is not random, hence a direct comparison of outcomes in treated and control groups is not valid to establish a treatment effect. This applies to this study where we aim to estimate the effects of single and multiple dismissals since, as we will see, these treatments are more likely to be assigned to, for instance, clubs that

[10] More precisely, the weights given to the seasons $s-1$, $s-2$, $s-3$, $s-4$ are 0.5, 0.3, 0.15, and 0.05, respectively, where s represents the current season. As reported in Dixon and Coles (1997), a club's ability is better measured by recent performance with increasing weights on the more recent information.

are performing poorly. As discussed in Section 2, the main approach to this issue has been to apply various matching techniques to form comparable counterfactual control groups. Our approach is different from those used in these prior studies as we employ the inverse propensity score weighting (PSW) method (Imbens, 2000). Intuitively, this method virtually "randomises" the treatment assignment by giving higher weights to "unusual" cases in treated and control groups. That is, if a treated unit has a lower predicted probability of being treated, it will be given a higher weight since it resembles a control unit. Similarly, if a control unit has a higher predicted probability of being treated, it will be given a higher weight since, although it is a control unit, it resembles a treated unit. By doing so, the distribution of the predicted probability of being treated, i.e. how likely a unit is being treated *ex-ante*, becomes similar between treated and control groups, as if the treatment was assigned randomly. The *ex-ante* probability of being treated is the so-called "propensity score," and we obtain these for the two types of treatment using the respective subsamples. In the following subsections, we provide further descriptions of the method and how to estimate outcome and treatment assignment models.

Another common characteristic of observational data is imbalance class. That is, the number of treated and control cases is often not approximately equal. This could be problematic particularly when implementing machine learning techniques to obtain propensity scores. Therefore, in the last part of this section, we explain how to deal with this issue using an over-sampling technique when estimating the treatment assignment models.

4.1 Estimation of outcome model

The outcome model includes post-treatment performance as a response variable (y_{its}), which is regressed with a treatment variable ($New\ coach_{its}$) as well as other control variables that can independently influence the outcome variables ($X_{its} = (Home,\ Ability,\ Opponent\ ability)$) as follows:

$$y_{its} = \delta\ New\ coach_{its} + \beta' X_{its} + \varepsilon_{its}, \tag{2}$$

Since our aim is to estimate the treatment effect that could possibly be different for single and multiple dismissals, we estimate the model (2) for the two types of dismissal separately, using the respective treatment and control groups defined in Section 3. Then, δ is the ATE of a respective managerial dismissal, β is a vector of parameters also to be estimated, and ε_{its} is a stochastic error component.

Following the discussion above, non-randomness of the treatment assignment means the *ex ante* probability of being treated ($\Pr[New\ coach_{its} = 1]$) varies across individuals.[11] Therefore, δ would be biased if the factors that affect such probabilities are not taken into account. Accordingly, we employ inverse propensity score weighting (PSW) to deal with this endogeneity issue. The method has been used traditionally in medical research (Austin and Stuart, 2015) and more recently in social sciences (Morgan and Todd, 2008). The PSW adjusts the pre-treatment differences between treated and control units by weighting the treated units with the inverse of the predicted probability of receiving the treatment and the control units with the inverse of the predicted probability of *not* receiving the treatment. The predicted probability of receiving the treatment is the so-called propensity score, and the weights assigned to individual units are formally defined as follows:

$$w_{its} = \begin{cases} (P[New\ coach_{its} = 1])^{-1}, & \text{if } New\ coach_{its} = 1, \\ (1 - P[New\ coach_{its} = 1])^{-1}, & \text{if } New\ coach_{its} = 0. \end{cases} \tag{3}$$

Again, since we estimate two types of treatment, we obtain the weights defined by the equation (3) separately using the two subsamples. Once we have obtained each weight (w_{its}) associated with the respective treatment type, we can separately estimate the model (2) within the weighted regression framework for each treatment. In addition, we estimate equation (2) with several post-treatment time windows as well as different performance measures discussed in the previous sections.

4.2 Estimation of treatment assignment models

Following the discussion above, the estimation of the outcome model using PSW requires obtaining the probabilities of receiving a respective treatment. Therefore, we first estimate the treatment assignment models defined as follows:

$$P[New\ coach = 1] = f(Z_{its}, \gamma), \tag{4}$$

where Z_{its} is a set of covariates that can affect the probability of receiving treatment, and γ is a vector that reveals the importance of each covariate in Z_{its}.

Whilst the estimation of the model (4) is a necessary step in the PSW procedure, we use this estimation to test our hypotheses regarding the causes of single and multiple dismissals. To fit the model (4), we employ two classes of classification algorithms: (1) logistic regression (LGT) and (2) Gradient Boosting Machine (GBM) with decision trees. While the former is the most commonly used estimation method for propensity scores in non-experimental causal analysis (Dehejia and Wahba, 2002), the use of Machine Learning (ML), including the latter model, has recently been shown to be a competitive option.

[11] In contrast, under a randomised control trial, where a treatment is randomly allocated to participants with equal probabilities.

In their reviews, Westreich et al. (2011) discuss the advantages and disadvantages of ML models against a logistic regression in the context of propensity score estimations. Perhaps the most notable advantage of tree-based models is the fact that functional forms are not required to be specified by a researcher. This is particularly attractive since underlying covariates are most likely to interact with each other or have a non-linear effect on the probability of treatment, and misspecification of functional forms may lead to a failure in mitigating the bias. On the other hand, a possible disadvantage of decision tree based models is the fact that the specific interaction of variables or polynomials cannot be identified, often referred to as the "black box" nature of the model class. As noted in Olmos and Govindasamy (2015), however, this drawback is not most concerning in this context, given that the aim of propensity score weighting is to mitigate the underlying covariate imbalancedness, not to identify such functional forms.

GBM employed in this analysis is a machine learning algorithm to assemble a series of weak learners, i.e. a classification model that alone does not have strong predictive power, in order to improve prediction. Here a weak leaner used is typically a regression tree, as is the case in this study. Intuitively, decision trees are sequentially built one after another to reduce prediction errors made by a previous learner until no improvement can be made or the specified maximum number of tree M is reached. More specifically, a new learner tries to model errors made by a previous learner and update this to a set of predicted values. By sequentially assembling simple weak learners in this way, the bias can be reduced without compensating for increased variance. As such, GBM models often outperform a single strong learner such as regression, which is subject to trade-offs between estimation errors and biases.

In each iteration, a set of such predicted values, $F_m(z)$ conditioning on covariates z, are obtained such as to minimise a loss function given the previous prediction $F_{m-1}(z)$, where $m = 1, \ldots, M$. Since our target value is binary, the loss function can be characterised as a negative value of log-likelihood.[12]

The model complexity, however, increases in the number of decision trees. Therefore, to avoid over-fitting, iterations are stopped where predictive performance is optimised on an independent data set rather than the sample used to fit the models. We implement this early stopping using cross-validation (CV) errors.[13]

4.3 Imbalancedness of classes

Another common characteristic of observational data is that the proportion of each class, for instance, treated and control, is often not even. In this study, the number of treated units in Sample I for single dismissal is 114, as opposed to that of control units being 7,890. Similarly, the number of treated units in Sample II for multiple dismissal is 36, as opposed to that of control units being 1,903. This is problematic, particularly

[12] This corresponds to the parameter estimation by maximum likelihood, where log-likelihood is to be maximised. For the formal presentation of the algorithm, see Friedman (2001).

[13] See, James et al. (2013) for more on early stopping.

for ML models, since they can bias toward the majority class.

To cope with the class imbalancedness, therefore, we use a synthesised sample to fit the treatment assignment models. Particularly, we employ the synthetic minority oversampling technique (SMOTE) proposed by Chawla et al. (2002). This approach to imbalanced data involves under-sampling the majority class (control groups) and over-sampling the minority class (treated groups) by synthesising data. More precisely, the latter is done in the following manner. First, k-nearest neighbours with respect to a vector of covariates (possible determinants of dismissals) are selected for a randomly selected member within the minority class. Second, we randomly select one of the k-nearest neighbours and obtain the distance between this and the member of the minority class under consideration. Finally, synthetic data is obtained where the difference is multiplied by a random value within the range of $(0, 1)$. In other words, synthetic data is a convex combination of the two minority observations. This is repeated until the specified number of synthesised minority cases is achieved. Then, randomly selected majority cases are also removed to achieve better class balance.[14]

The advantage of synthesising data points rather than purely duplicating members in the under-represented class is that the former can not only enlarge the minority sample size but also introduce variations. As explained in Chawla et al. (2002), the subsampling method can improve predictions in classification models. Table 4 summarises the compositions of each group and type of the treatments for all samples, and the resulting synthesised samples. Within both Sample I and Sample II, the class imbalance is significantly reduced in SMOTE samples relative to the respective original samples.

Table 4: The number of units in full and SMOTE samples

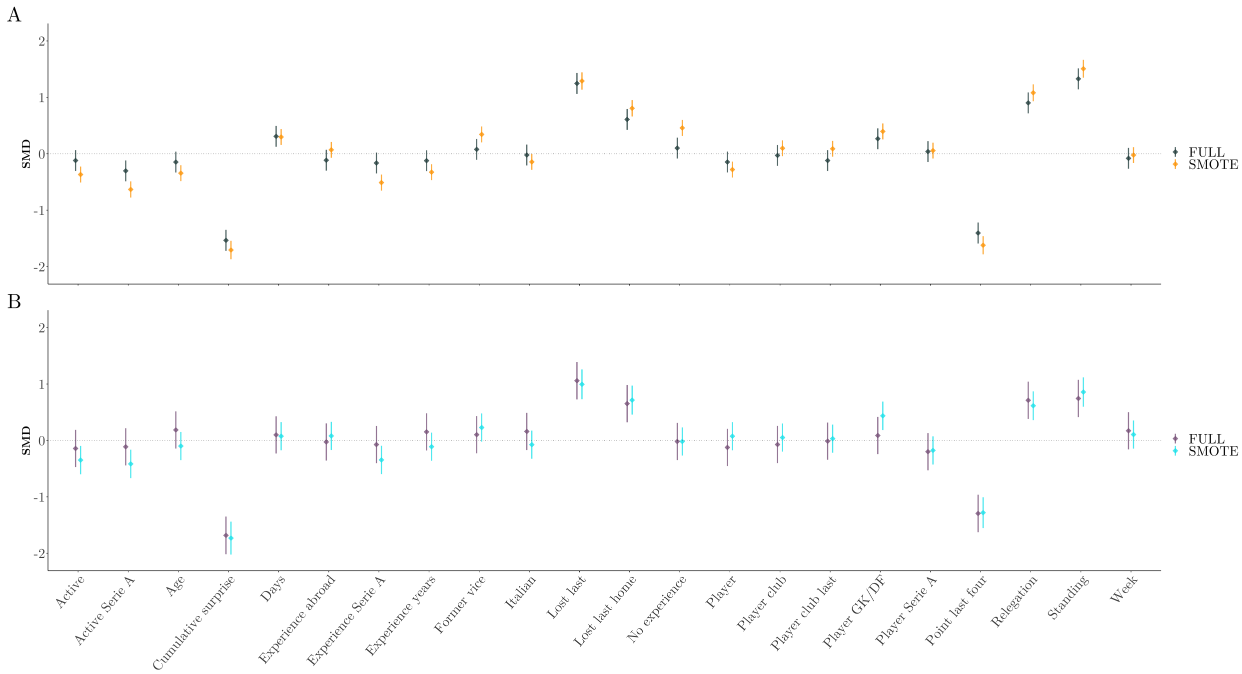
	Sample I		Sample II	
	Treated	Control	Treated	Control
All	114 (1.424%)	7890 (98.576%)	36 (1.857%)	1903 (98.143%)
SMOTE	342 (42.857%)	456 (57.143%)	108 (42.857%)	144 (57.143%)

Notes: Table shows the number of units in the respective treated and control groups in Sample I and Sample II. The numbers are presented for full and synthesised (SMOTE) samples.

Figure 2 visualises the standardised mean differences (SMD) between treated and untreated observations for each treatment (Treatment I and Treatment II) and sampling type (full sample and SMOTE sample). As the figures show, the synthesised sample preserves the underlying differences between treated and control groups for each treatment type. Therefore, our treatment assignment models are fitted using the SMOTE sample.

[14] More specifically, the SMOTE is implemented using the R function SMOTE, which uses $k = 5$ nearest neighbours and synthesises minority cases until the number of extra minority cases generated is equal to 2*(the number of minority cases in the original sample). In addition, the majority cases are randomly removed so that the number of majority cases is equal to 2*(the number of synthesised minority cases).

Figure 2: Standardised mean differences (SMD) in predictors in full and SMOTE samples



Notes: Figure shows the SMDs in all the considered predictors for Treatment I (Panel A) and Treatment II (Panel B). Each panel shows the SMDs between treated and control groups with full and synthesised (SMOTE) samples.

5 Results

In this section, we present the results from each step of our analysis. That is, we first present and discuss the results related to the causes of the two types of dismissals, i.e. single and multiple dismissals, where the logistic regression and GBM are fitted for treatment assignment models of the two. Then, we use the GBM models[15] to obtain the probabilities of each dismissal, i.e. propensity scores, and shows how well the weighting based on these propensity scores can reduce the pre-treatment unbalancedness between the respective control and treated groups. Finally, we present our results for the PSW estimators of the average treatment effects of the single and multiple dismissals.

5.1 Predictors of single and multiple dismissals

Following the procedure explained in the previous section, we proceed with the analysis by fitting classification models (4). As discussed there, the models for single dismissal are trained with SMOTE sample synthesised from Sample I, and those for multiple dismissal with SMOTE sample synthesised from Sample

[15] We select GBM for obtaining the propensity scores since it outperforms logistic regression in terms of balancing covariates, as explained further below.

II. The objective here is two-fold: to test the hypothesis regarding the motivations behind the two types of dismissals (H1) and to obtain the predicted values of being sorted into a realised treatment group, i.e. a propensity score. The classification models are estimated using (1) logistic regression and (2) Gradient Boosting Machines (GBM), in which possible determinants of treatment assignments discussed in Section 3.1 are considered.

For logistic regression models, the final set of covariates to be included is selected by means of stepwise regression with a sequential replacement algorithm, which combines forward and backward selection. Hence, the predictors under consideration are iteratively added and removed until the lowest predictive error is achieved. Following Bruce and Bruce (2017), errors are measured with the Akaike information criterion (AIC).

Table 5 shows the logit estimates with selected predictors. Included in Column (1) are the parameter estimates for the selected covariates in the treatment assignment model for Treatment I. The results suggest that most of the indicators associated with recent field performance show a strong correlation with the likelihood of Treatment I, the first dismissal that happens during a season. In particular, performance exceeding the expectation (*Cumulative surprise*) and a streak of favourable outcomes in recent matches (*Point last four*) are negatively related to the probability of dismissal. On the other hand, a loss in the most recent match (*Lost last*) is associated with a higher probability of dismissal, with an additional effect when the defeat was at the home stadium (*Last lost home*). Whilst a club's interim league position (*Standing*) itself does not appear in the selected model, there is a positive and significant effect of relegation threat (*Relegation*). These results are fairly consistent with the previous findings.[16]

Furthermore, our results suggest that some of the variables related to managerial characteristics can also influence dismissal decisions. For instance, having previous experience as a head coach within *Serie A* (*Experience Serie A*) and being a former player (*Player*) are negatively associated with a manager being dismissed. On the other hand, the probability of dismissal is higher when a manager has a previous experience as a head coach in a foreign professional league (*Experience abroad*) and is a former defender/goalkeeper (*Player GK/DF*). Other managerial characteristics, such as experience in years, are not selected in the model, suggesting that they are not an important predictor of the first dismissal.

The results further indicate that when there are more days between two matches, a club is more likely to dismiss a manager (*Days*), and the coefficients on the variable *Week* and its quadratic term *Week squared* suggest that dismissals occur more often as a season progresses, but such decisions are then made less often towards the end of the season.

The counterpart estimates for the treatment assignment of Treatment II, the second dismissal that hap-

[16]See Section 2 above for a review of related literature.

pens during the same season, are shown in Column (2), where we observe some inconsistencies. Particularly, the variables *Cumulative surprise* and *Point last four* are selected for this treatment assignment model, however, unlike that for the first dismissals, the predictors *Lost last*, *Lost last home*, and *Relegation* are not selected. This may indicate that the second dismissal decisions are not as sensitive to external pressure as the first dismissals.

Comparing the selected covariates in Columns (1) and (2), some of the managerial characteristics included in the prediction of first dismissals are also present in the treatment assignment model for second dismissals, whilst the effects of each characteristic are not consistent. For example, being a former player (*Player*) and having experience as a head coach abroad (*Experience abroad*) have opposite effects on the likelihood of second dismissals compared to those on the first dismissal decision. The second dismissals are further affected by managerial characteristic variables such as being a former vice coach of the club (*Former vice*), which is suggested to increase the probability of such events. It is not straightforward to decide which decisions are more sensible based on which managerial characteristics affect each decision. However, the fact that different managerial characteristics can influence the first and second dismissal decisions differently may suggest that the club's expectations or preference on particular managerial characteristics may be updated based on the outcome of the recent dismissal.

Whilst the general trend with respect to weeks is again similar to the first dismissal case, the number of days between matches does not seem to have an influence on the second dismissal decision.

Table 5: Logit estimates of treatment assignment models

	Dependent variable: New coach	
	(1) First dismissal	(2) Second dismissal
Cumulative surprise	-0.387^{***} (0.052)	-0.958^{***} (0.167)
Point last four	-1.262^{***} (0.318)	-1.235^{**} (0.564)
Lost last	0.962^{***} (0.285)	
Lost last home	0.680^{**} (0.290)	
Relegation	1.067^{***} (0.299)	
Experience Serie A	-0.627^{**} (0.296)	
Experience abroad	0.715^{**} (0.285)	-0.810^{*} (0.482)
Active Serie A		-1.020^{**} (0.456)
Player	-0.924^{**} (0.398)	1.567^{**} (0.650)
Player Serie A	0.712^{***} (0.266)	1.573^{***} (0.544)
Player GK/DF		-1.390^{**} (0.541)
Former vice		1.327^{*} (0.763)
Days	0.068^{**} (0.031)	
Week	0.299^{***} (0.058)	0.676^{***} (0.244)
Week squared	-0.008^{***} (0.002)	-0.013^{***} (0.005)
Constant	-2.934^{***} (0.741)	-9.311^{***} (2.922)
Observations	798	252
Log Likelihood	−225.789	−71.782
Akaike Inf. Crit.	477.578	165.565
Note:		*p<0.1; **p<0.05; ***p<0.01

Overall, the estimations of logistic regressions to predict the first and second dismissals indicate that the predictors of these two events could be quite different. The findings from the alternative classification model, GBM, are similar in that it also suggest that the factors affecting the single and multiple dismissals can differ. With GBM, the predictive power of each covariate is measured in terms of relative influence. As explained by Breiman (2001), this is computed by a percentage increase in misclassification rate when randomly introducing noise to one variable at a time. These values obtained for each variable are then normalised so that they add up to 100%.

Figure 3 illustrates the relative importance of the predictors that are assigned non-zero influence for the predictions of the first dismissal (A) and the second dismissal (B). The most influential predictor of the first dismissal is *Point last four* (36.9%), although *Cumulative surprise* is a close competitor (34.6%). On the other hand, the second dismissal has one dominating predictor, *Cumulative surprise* (53.9%), with the second most influential predictor being *Point last four* (12.4%). This indicates that the first dismissal is driven by both absolute measure of performance (*Four last point*) and the relative performance against expectation (*Cumulative surprise*), whilst the second dismissal is mostly influenced by the relative performance. It can be argued that the second dismissal decision is less sensitive to the mere fact that a club has poorly performed on average in the recent matches, but such a decision may be made by taking into account realistic expectations.

Consistent with the logistic regression presented above, the first dismissals are affected by the threat of demotion (*Relegation*), as well as the defeat in the most recent match (*Lost last*), which is not the case for the second dismissal. Furthermore, the sets of managerial characteristics that can affect the two events are again slightly different, although those predictors are shown to be of less importance.

The two classification models presented in this section confirm that the factors affecting the likelihood of the first and second within-season dismissals could be quite different. In addition, our findings suggest that the second dismissal decision may be less sensitive to the crude measure of performance, such as average performance in recent matches and defeat in the most recent match, but instead motivated by a fairer measure of performance i.e. performance against expectation. This is consistent with our hypothesis (H1).

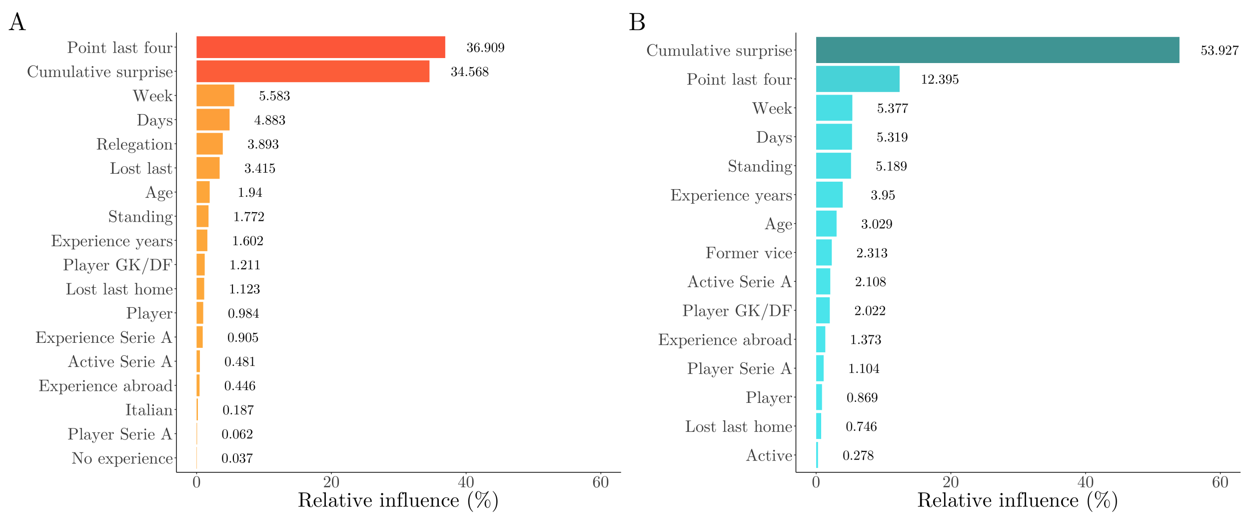


Figure 3: Relative importance of predictors with non-zero influences

Notes: Figure shows the relative influence of covariate assigned non-zero influence. GBM models are fitted for Treatment I (Panel A) and Treatment II (Panel B).

5.2 Estimation of propensity scores and covariate balance

In the previous subsection, we presented the fitted treatment assignment models to predict the first and second managerial dismissal events using the two classification models, logistic regression and GBM. The selected predictors vary across the two different dismissal types. On the other hand, the main predictors for each treatment are fairly consistent regardless of the choice of a classification model.

Note that we used synthesised samples to fit the classification models due to imbalanced classification. However, our aim is now to obtain the predicted probability of treatment for each individual unit. Therefore, we now obtain the fitted values using Sample I and Sample II, i.e. the entire samples rather than the SMOTE sample that are used to fit the models.

It turns out that the performance of the two classification models in terms of predictability is not significantly different from each other.[17] Meanwhile, it is found to be the case that GBM outperforms logistic regressions in terms of balancing underlying covariates differences between the treated and control groups for both treatment types. Given that our aim is "virtual randomisation" of each treatment, we, therefore, proceed with our analysis by obtaining the propensity scores with GBM.[18]

It is evident from the results presented in the previous subsection that managerial dismissals are not a random event, i.e. there are significant differences between the treated and control observations in terms of the pre-treatment conditions. This implies that any direct comparison between the treated and control observations for any treatment types examined here is subject to selection bias. Therefore, there is a need to adjust for the pre-treatment differences, and we do so using PSW described above.

The mean values and range of predicted values as well as the number of units in each treatment group are summarised in Table 6. The first two rows include the statistics for the whole sample (Sample I and Sample II for Treatment I and Treatment II, respectively). In addition, the bottom two rows include the statistics for observations within the common support, i.e. the range where the propensity scores within treated and control groups overlap. As expected, the predicted value of dismissal is higher in the treated groups compared to those in the control groups for the respective treatment types. Furthermore, only one treated observation is outside the common support for Treatment I, and all the treated cases are contained within the common support for Treatment II.

Table 6: Summary of estimated propensity scores

	Treatment I				Treatment II			
	Mean	Min.	Max.	N	Mean	Min.	Max.	N
Treated	.6904	.0270	.9908	114	.7176	.2376	.9769	36
Control	.1577	.0002	.9903	7890	.1648	.0006	.9884	1903
Treated (CS)	.6878	.027	.9900	113	.7176	.2376	.9769	36
Control (CS)	.2779	.027	.9903	4346	.5771	.2395	.9718	433

Notes: Table includes the summary statistics of estimated propensity scores for Treatment I and Treatment II. The first two rows are the statistics related to the full sample, and the last two include those for the observations within the respective common support (CS).

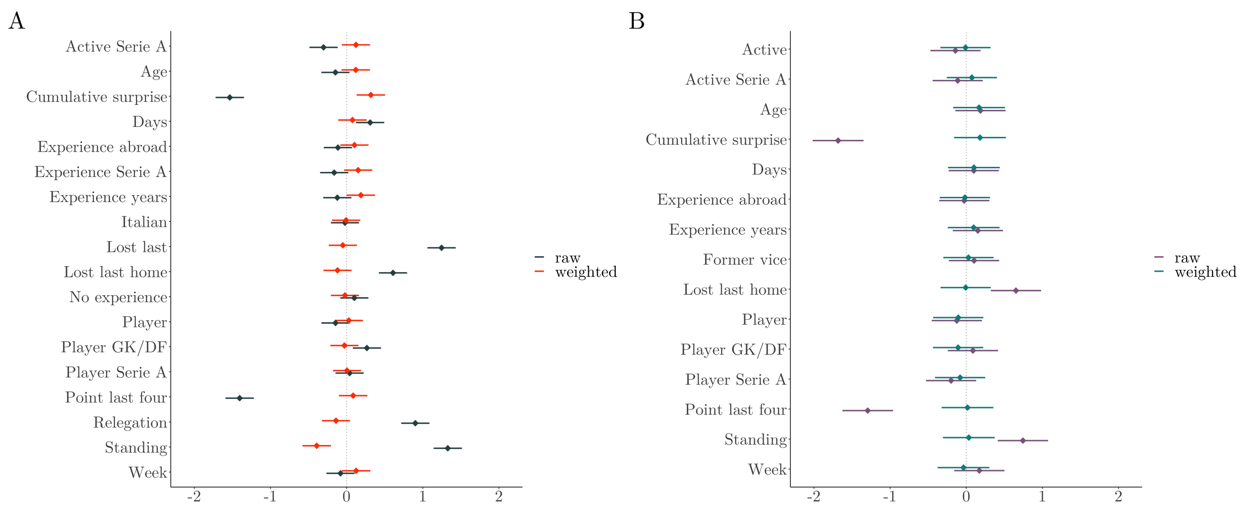
Now that the propensity scores for each treatment have been obtained, we examine how well PSW can reduce the pre-treatment differences between the treated and control groups. Following Austin and Stuart (2015) and Morgan and Todd (2008), we first obtain the standardised mean differences (SMD) in selected

[17] This is based on the evaluation metrics, AUC and balanced accuracy. See, Figure 6 and Table 9 in Appendix for details.

[18] We present the standardised mean differences of each covariate before and after applying PSW based on GBM later in this section and those based on logistic regression in Figure 7 in Appendix as a comparison.

covariates between the treated and control groups using a raw sample. Then, we obtain the weighted SMDs using the weights defined by Equation (3) and the estimated propensity scores and observations within the common support.[19] Figure 4 presents the raw/weighted SMDs in the relevant covariates for Treatment I (Panel A) and Treatment II (Panel B). The Figure suggests that weighting can significantly reduce the pre-treatment differences in relevant characteristics in both cases.

Figure 4: Standardised mean differences of selected covariates before and after weighting



Notes: Figure shows the standardised mean differences (SMD) in the selected covariates between treated and control groups for Treatment I (Panel A) and Treatment II (Panel B). Each panel includes SMDs for the respective raw and weighted sample. The horizontal line represents the 95% confidence intervals. The weights used here are based on the propensity scores estimated with GBM for each treatment.

5.3 Average treatment effects (ATE) on field performance

It is evident from the previous subsection that PSW can substantially mitigate the unbalancedness in the important pre-treatment characteristics between treated and controlled units for each type of treatment. As explained in Section 4.1, we therefore estimate our outcome models (Equation 2) by means of weighted regression with PSW in order to obtain the ATE of each treatment. As response variables, average performance in post-treatment matches within different time windows (subsequent one, five, and 10 matches, and all the remaining matches in the relevant season) are obtained for points (*Point*) and goal differences (*Goal dif*). In addition to the treatment and response variables, we control for the strength of a club and their opponent and home field advantage.

The estimation results of ATEs (the coefficients of our treatment variable *New coach*) for single and

[19] Whilst it comes with a cost (efficiency loss), using the observations within common support can further reduce the selection bias (Rosenbaum and Rubin, 1983).

multiple dismissals are presented in Table 7 and 8, respectively.[20] Table 7 suggests that managerial change resulting from the first dismissal in the season has no significant effects on the performance in the first match immediately after the change. This is the case for performance measured with both points obtained (*Point_1*) and goal difference (*Goal dif_1*). However, the first managerial change can improve performance when we consider the average performance up to 5 post-treatment matches. In particular, the ATE on the performance measured with average points (*Point_5*) are statistically significant at the 5% significance level; whilst the ATE on the performance measures with average goal differences (*Goal dif_5*) is only significant at the 10% significance level. These results are consistent when we include up to 10 post-treatment matches to obtain the average performance (*Point_5* and *Goal dif_5*). However, the positive ATE of the first dismissal is only significant at the 10% significance level with the response variable *Point_37*, and it is not statistically significant at any conventional significance level with the response variable *Goal dif_37*. Therefore, when we include all the post-turnover matches in the remaining season, the effects of the first dismissal are no longer robust. Overall, whilst the first dismissal does not immediately influence club performance, we observe a limited boost in performance shortly after.

Contrarily, it is evident from Table 8 that the ATEs of the second managerial change are not significant at any conventional significance level. Unlike the first turnover, therefore, the second turnover does not influence the club's performance at all. Therefore, the consequences of the first and second managerial turnover are somewhat different, although neither type of dismissal has positive and significant effects in the long run. This partially supports our hypothesis (H2), although the difference is not as clear-cut as we initially expected.

Therefore, these results may justify a club's decision to replace a manager, provided that they have not done so already in the season, although the positive effects are not expected to be persistent. Although our results do not suggest that a consistent managerial change can be detrimental, it still indicates that undergoing yet another managerial change does not make any difference. This is in one sense ironic, given that our results suggest that the second managerial dismissal appears to be made with more caution.

[20]The OLS estimates of ATEs are provided in Table 10 and Table 11 in Appendix as a comparison. This "naive" approach, where the endogeneity of treatment assignment is not taken into account, generates somewhat different results from the ones presented in Table 7 and Table 8.

Table 7: PSW estimates: ATE of single dismissal on points and goal differences

	Dependent variable:							
	Point_1	Goal dif_1	Point_5	Goal dif_5	Point_10	Goal dif_10	Point_37	Goal dif_37
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New coach	0.235	0.409	0.361^{**}	0.483^{*}	0.348^{**}	0.470^{*}	0.291^{*}	0.429
	(0.240)	(0.387)	(0.158)	(0.250)	(0.161)	(0.253)	(0.174)	(0.263)
Observations	4,459	4,459	4,459	4,459	4,459	4,459	4,459	4,459
Log Likelihood	−7,990.227	−9,090.177	−5,456.790	−6,802.615	−4,979.050	−6,427.112	−4,759.587	−6,279.965
Akaike Inf. Crit.	15,990.450	18,190.350	10,923.580	13,615.230	9,968.101	12,864.220	9,529.173	12,569.930

Notes: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses. All models include control variables (club ability, opponent ability, and home advantage) associated with the response variables.

Table 8: PSW estimates: ATE of multiple dismissal on points and goal differences

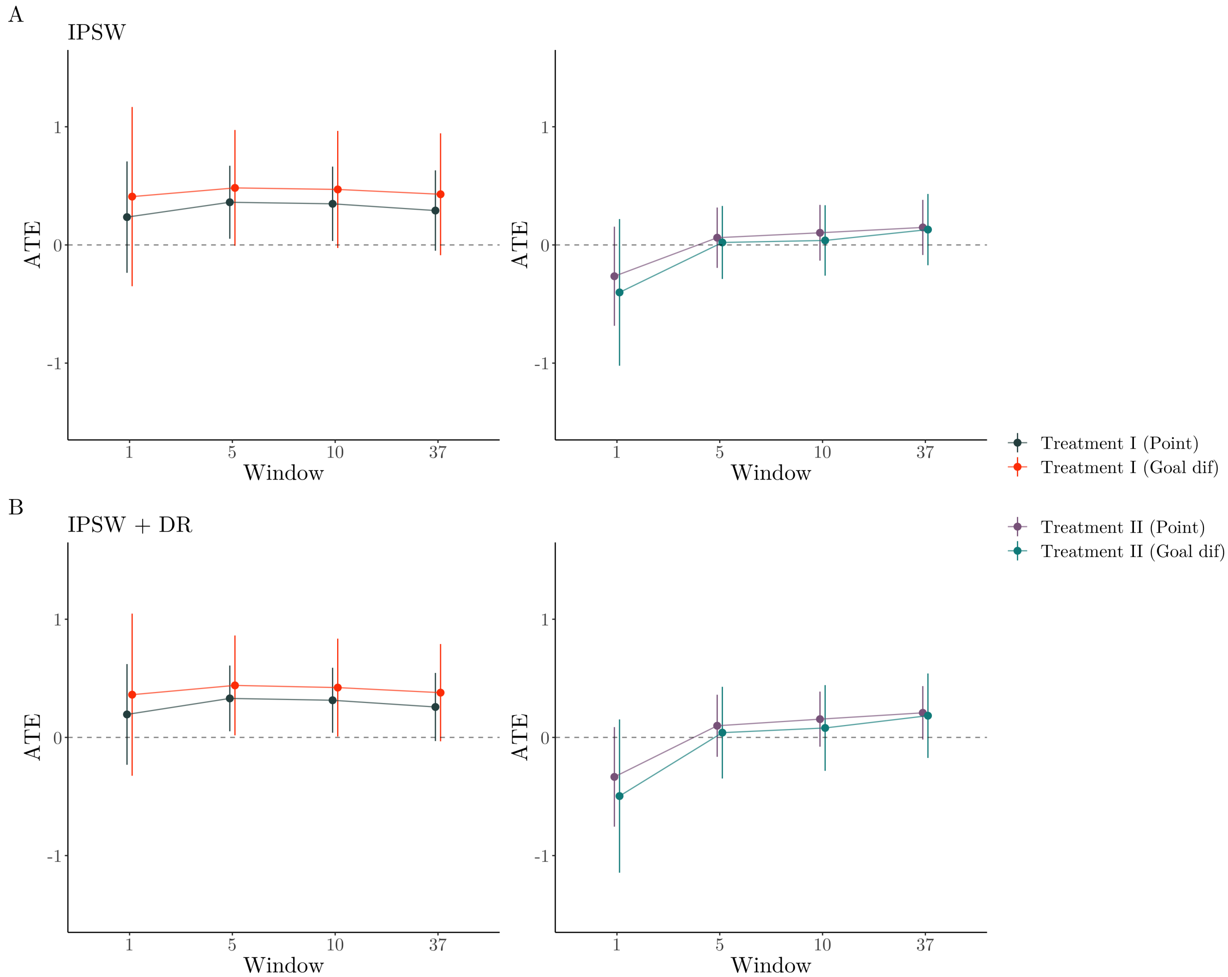
	Dependent variable:							
	Point_1	Goal dif_1	Point_5	Goal dif_5	Point_10	Goal dif_10	Point_37	Goal dif_37
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New coach	−0.265	−0.402	0.061	0.021	0.103	0.038	0.148	0.130
	(0.214)	(0.317)	(0.130)	(0.158)	(0.120)	(0.152)	(0.119)	(0.154)
Observations	469	469	469	469	469	469	469	469
Log Likelihood	−833.997	−964.951	−602.062	−775.277	−561.534	−735.339	−544.708	−720.081
Akaike Inf. Crit.	1,677.994	1,939.901	1,214.125	1,560.553	1,133.068	1,480.678	1,099.416	1,450.162

Notes: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parentheses. All models include control variables (club ability, opponent ability, and home advantage) associated with the response variables.

The results discussed above were obtained using inverse propensity scores weighting (PSW). To check the robustness of our results, we compare these results with the double-robust estimator of ATEs. The double-robust estimation follows the PSW procedure, where the outcome model is estimated by means of weighted regression with the weights based on the estimated propensity scores. In addition, this estimation includes the covariates used to obtain the propensity scores in the outcome models, which can increase protection against model misspecifications (Funk et al., 2011).

Figure 5 compares the estimated ATEs of single dismissal (Treatment I, left-hand side of each panel) and multiple dismissal (Treatment II, right-hand side of each panel) for PSW (Panel A) and PSW with doubly-robust estimation (Panel B). Outcome variables are again measured with average points (*Point*) and goal differences (*Goal dif*), and the maximum number of post-treatment matches used to obtain the average performance are represented on the horizontal axis. The double-robust estimators of ATEs are very similar to those estimated with PSW. However, the significance of ATEs on some performance measures changes to some degree in that some become slightly more significant. For instance, the ATEs of first dismissals are now positive and significant at the 5% significance level for both measures of outcome (*Point* and *Goal dif*) when we consider up to 5 and 10 post-treatment matches. Overall, however, our conclusion remains: there are limited positive effects of first dismissals, whilst second dismissals barely influence club performance.

Figure 5: ATE under different estimation strategies



Notes: Figure presents estimated ATEs for Treatment I (left column) and Treatment II (right columns) with the 95% confidence intervals. The horizontal axis (window) represents the maximum number of post-treatment matches included to obtain response variables. The estimations are obtained with PSW (panel A) and PSW with a double robust estimator (panel B).

6 Conclusion

This study examines the causes and consequences of managerial dismissals by differentiating the first and second dismissals that occur within a season. Firstly, in order to identify the causes of the two types of dismissal, we employed logistic regression and GBM. Estimation results of the two classification models indicate that the motivating factors of the two can be different. A first dismissal is likely to have followed a sequence of poor results, measured by points-per-game in recent matches, and a series of matches where outcome has been disappointing relative to expectation. Consistent with previous findings, such as in Tena and Forrest (2007), the threat of relegation can also influence such a decision. On the other hand, the dominating factor of the second turnover is suggested to be performance against expectation, and the crude measure of performance such as the average points in recent matches has much less influence on a decision to dismiss another manager. In this sense, the second managerial dismissal appears to be considered with greater caution, instead of acting upon the mere streak of unfavourable results. Overall, whilst the predictors of the first dismissal are comparable with the findings from previous studies, for instance, van Ours and van Tuijl (2016), those of the second dismissal could be quite different.

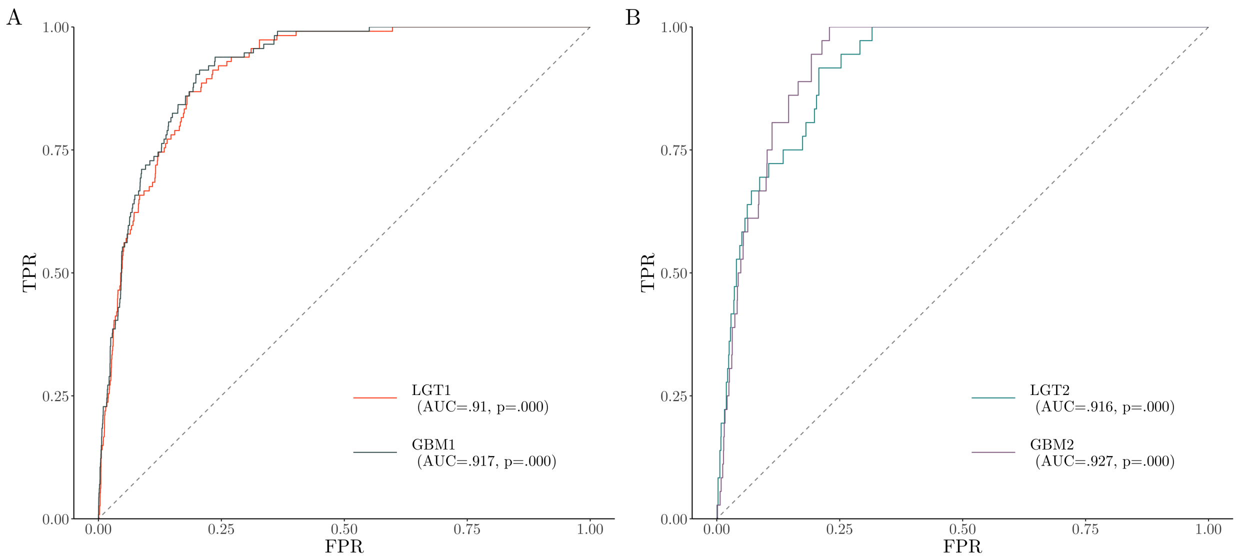
From the methodological point of view, the difference in the factors affecting the likelihood of the two types of dismissal is a relevant concern. Specifically, in order to estimate the effectiveness of managerial dismissals, one needs to take into account the endogeneity of such decisions. To do so, we employ PSW, where the weights based on the propensity score, i.e. the probability of a club dismissing a manager, are applied in order to make the treated and control groups comparable. Estimating the ATEs of the first and second dismissals with PSW, therefore, requires obtaining the probability of the two types of dismissal separately, using the relevant predictors for the respective treatment type.

We find that the consequences of the first and second managerial dismissals are also not identical. The immediate effect of the first turnover is not statistically significant at any conventional significance level, whilst some boost in performance has been observed when we consider up to 5 and 10 matches following the turnover. The effect on the average performance for the rest of the season is, however, non-significant. Despite our findings that indicate the second managerial change is likely to be made with greater caution, the impact of the second dismissal is not statistically significant at any conventional significance level on both immediate and extended post-treatment performance.

These findings add to the previous research on the causes and consequences of managerial dismissals in professional football. The existing studies focus on establishing whether within-season replacements are in general effective or not, and therefore they do not always provide guidance on how to make such decisions. Our findings are more informative since they suggest that managerial dismissals should not be enforced so soon after a recent dismissal since this is unlikely to make any significant difference. This can also add to the understanding of the leadership succession effect in broader management literature. Our findings provide some support for Boyne and Dahya (2002) and Gordon and Rosen (1981)'s hypothesis that frequent leadership change within a short period of time is unproductive. Indeed, whilst it could be tempting to replace leaders of an organisation whenever its performance is not favourable, constant changes in management are unlikely to be beneficial.

Appendix

Figure 6: Receiver operating characteristics (ROC) curves and area under the curve (AUC)



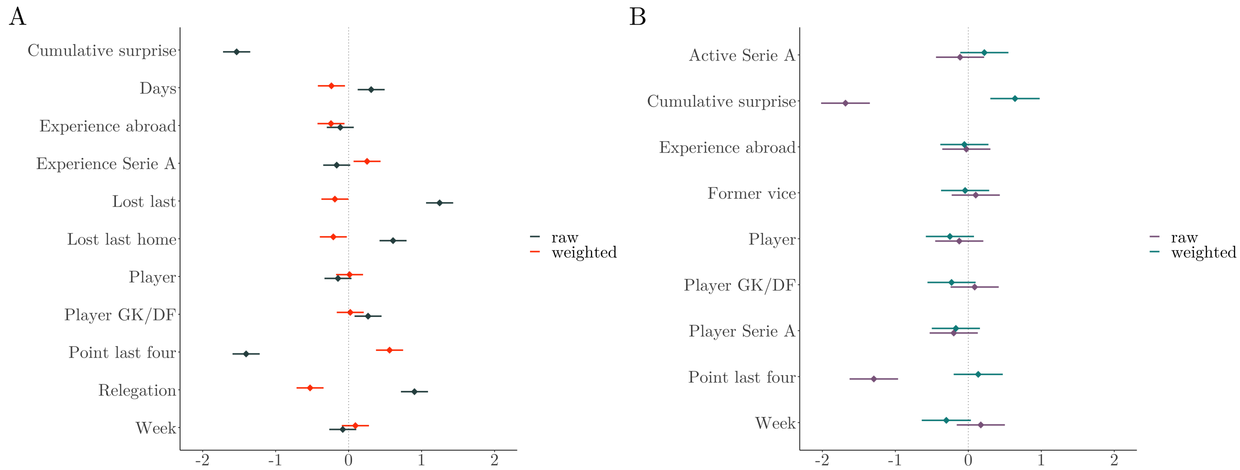
Notes: Figure plots ROC curves for logistic regression for treatment I (LGT1, Panel A) and GBM for Treatment I (GBM1, Panel A), logistic regression for Treatment II (LGT2, Panel B) and GBM for Treatment II (GBM2, Panel B). ROCs represent the true positive rate and the false positive rate at a given discrimination threshold for binary classification problems. A 45-degree line represents the ROC when the binary classification was done by a random selection. Therefore, ROCs with the area under the curve (AUC) significantly larger than that of the ROC of random selection, i.e. 0.5, exhibit a significant separability of the respective models. Also presented in Figure are the values for AUCs as well as p-values associated with the Mann-Whitney-Wilcoxon (MWW) test with $H_0 : AUC = 0.5$ against $H_1 : AUC > 0.5$ for each curve. The AUCs and MWW test suggest that all the models show significant separability. The statistical significance of the differences between the two AUCs between the two competing models for each treatment is tested using DeDong's test, similar to MWW. The associated p-values are $p = 0.162$ for the comparison between LGT1 and GBM1, and $p = 0.2833$ for the comparison between LGT2 and GBM2.

Table 9: Balanced accuracy of classification models

	Treatment I	Treatment II
Logit	0.841	0.854
GBM	0.853	0.885

Notes: Balanced accuracy is defined by the arithmetic mean of true positive rate and true negative rate.

Figure 7: Covariate balance with logistic regression



Notes: Figure shows the standardised mean differences (SMD) in the selected covariates between treated and control groups for Treatment I (Panel A) and Treatment II (Panel B). Each panel includes SMDs for the respective raw and weighted sample. The horizontal line represents the 95% confidence intervals. The weights used here are based on the propensity scores estimated with logistic regressions for each treatment.

Table 10: OLS estimates: ATE of single dismissal on points and goal differences

	Dependent variable:							
	Point_1	Goal dif_1	Point_5	Goal dif_5	Point_10	Goal dif_10	Point_37	Goal dif_37
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New coach	−0.136	−0.151	−0.034	−0.099	−0.059	−0.123*	−0.075	−0.153**
	(0.113)	(0.145)	(0.061)	(0.080)	(0.054)	(0.071)	(0.050)	(0.068)
Observations	8,004	8,004	8,004	8,004	8,004	8,004	8,004	8,004
R^2	0.156	0.176	0.229	0.252	0.269	0.290	0.286	0.300
Adjusted R^2	0.156	0.176	0.228	0.252	0.269	0.290	0.285	0.300

Notes: *p<0.1; **p<0.05; ***p<0.01. All models include control variables (club ability, opponent ability, and home advantage) associated with the response variables.

Table 11: OLS estimates: ATE of multiple dismissal on points and goal differences

	Dependent variable:							
	Point_1	Goal dif_1	Point_5	Goal dif_5	Point_10	Goal dif_10	Point_37	Goal dif_37
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New coach	−0.333*	−0.366	−0.046	−0.131	−0.025	−0.109	0.037	−0.013
	(0.202)	(0.253)	(0.111)	(0.149)	(0.101)	(0.135)	(0.097)	(0.129)
Observations	1,939	1,939	1,939	1,939	1,939	1,939	1,939	1,939
R^2	0.116	0.147	0.148	0.155	0.170	0.181	0.181	0.193
Adjusted R^2	0.115	0.145	0.146	0.153	0.168	0.179	0.179	0.191

Notes: *p<0.1; **p<0.05; ***p<0.01. All models include control variables (club ability, opponent ability, and home advantage) associated with the response variables.

References

Audas, R., Dobson, S., and Goddard, J. (1999). Organizational Performance and Managerial Turnover. *Managerial and Decision Economics*, 20(6):305–318.

Austin, P. C. and Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, 34(28):3661–3679.

Berns, K. V. and Klarner, P. (2017). A Review of the CEO succession Literature and a Future Research Program. *Academy of Management Perspectives*, 31(2):83–108.

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

Boyne, G. and Dahya, J. (2002). Executive succession and the performance of public organizations. *Public Administration*, 80(1):179–200.

Breiman, L. (2001). Random Forests. *Machine Learning 2001 45:1*, 45(1):5–32.

Bruce, P. and Bruce, A. (2017). *Practical statistics for data scientists*. O'Reilly, California, USA.

Bruinshoofd, A. and Ter Weel, B. (2003). Manager to go? Performance dips reconsidered with evidence from Dutch football. *European Journal of Operational Research*, 148(2):233–246.

Bryson, A., Buraimo, B., Farnell, A., and Simmons, R. (2021a). Special Ones? The Effect of Head Coaches on Football Team Performance. *IZA Discussion Paper*, No. 14104.

Bryson, A., Buraimo, B., Farnell, A., and Simmons, R. (2021b). Time To Go? Head Coach Quits and Dismissals in Professional Football. *De Economist*, 169(1):81–105.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357.

D'Addona, S. and Kind, A. (2014). Forced Manager Turnovers in English Soccer Leagues: A Long-Term Perspective. *Journal of Sports Economics*, 15(2):150–179.

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151–161.

Desai, M. N., Lockett, A., and Paton, D. (2018). Information Asymmetries in the Hiring Process and the Risk of New Leader Dismissal: Insights from English Premier League Soccer Organizations. *British Journal of Management*, 29(1):26–42.

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.

Farah, B., Elias, R., De Clercy, C., and Rowe, G. (2020). Leadership succession in different types of organizations: What business and political successions may learn from each other. *The Leadership Quarterly*, 31(1):101289.

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

Frick, B., Barros, C. P., and Prinz, J. (2010). Analysing head coach dismissals in the German "Bundesliga" with a mixed logit approach. *European Journal of Operational Research*, 200(1):151–159.

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.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., and Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, 173(7):761–767.

Gamson, W. A. and Scotch, N. A. (1964). Scapegoating in baseball. *American Journal of Sociology*, 70(1):69–72.

Giambatista, R. C., Rowe, W. G., and Riaz, S. (2005). Nothing succeeds like succession: A critical review of leader succession literature since 1994. *The Leadership Quarterly*, 16(6):963–991.

Gilfix, Z., Meyerson, J., and Addona, V. (2020). Longevity differences in the tenures of American and foreign Major League Soccer managers. *Journal of Quantitative Analysis in Sports*, 16(1):17–26.

Gordon, G. E. and Rosen, N. (1981). Critical factors in leadership succession. *Organizational Behavior and Human Performance*, 27(2):227–254.

Grusky, O. (1963). Managerial Succession and Organizational Effectiveness. *American Journal of Sociology*, 69(1):21–31.

Hill, G. C. (2009). The effect of frequent managerial turnover on organizational performance: A study of professional baseball managers. *Social Science Journal*, 46(3):557–570.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to Statistical Learning with applications in R.* Springer, New York.

Jensen, M. C. and Murphy, K. J. (1990a). CEO Incentives : It's Not How Much You Pay, But How. *Harvard Business Review*, 3(May-June):138–153.

Jensen, M. C. and Murphy, K. J. (1990b). Performance Pay and Top-Management Incentives. *Journal of Political Economy*, 98(2):225–264.

Khaliq, A. A., Thompson, D. M., and Walston, S. L. (2006). Perceptions of hospital CEOs about the effects of CEO turnover. *Hospital topics*, 84(4):21–27.

Kim, Y., Jeong, S. S., Yiu, D. W., and Moon, J. (2021). Frequent CEO Turnover and Firm Performance: The Resilience Effect of Workforce Diversity. *Journal of Business Ethics*, 173(1):185–203.

Kwon, I. (2005). Threat of Dismissal: Incentive or Sorting? *Journal of Labor Economics*, 23(4).

Morgan, S. L. and Todd, J. J. (2008). 6. A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects. *Sociological Methodology*, 38(1):231–282.

Muehlheusser, G., Schneemann, S., and Sliwka, D. (2016). The impact of managerial change on performance: The role of team heterogeneity. *Economic Inquiry*, 54(2):1128–1149.

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.

Olmos, A. and Govindasamy, P. (2015). A Practical Guide for Using Propensity Score Weighting in R. *Practical Assessment, Research & Evaluation*, 20(13).

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.

Peeters, T. L., Szymanski, S., and Tervii, M. (2017). The Inefficient Advantage of Experience in the Market for Football Managers. *SSRN Electronic Journal*.

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

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

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.

Scelles, N., Llorca, M., et al. (2020). Head coach change and team performance in the French men's football Ligue 1, 2000-2016. *Economics Bulletin*, 40(2):920–937020.

Sparks, R. (1986). A Model of Involuntary Unemployment and Wage Rigidity: Worker Incentives and the Threat of Dismissal. *Journal of Labor Economics*, 4(4):560–581.

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.

ter Weel, B. (2011). Does Manager Turnover Improve Firm Performance? Evidence from Dutch Soccer, 1986-2004. *De Economist*, 159(3):279–303.

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.

Wang, H., Zhao, S., and Chen, G. (2017). Firm-specific knowledge assets and employment arrangements: Evidence from CEO compensation design and CEO dismissal. *Strategic Management Journal*, 38(9):1875–1894.

Westreich, D., Lessler, J., and Funk, M. J. (2011). Propensity score estimation: machine learning and classification methods as alternatives to logistic regression. *Journal of Clinical Epidemiology*, 63(8):826–833.