University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 380

The role of boards' misperceptions in the relation between managerial turnover and performance: Evidence from European football

Raphael Flepp and Egon Franck

January 2019

---

http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Raphael Flepp

raphael.flepp@business.uzh.ch

Egon Franck

egon.franck@business.uzh.ch

University of Zurich

Department of Business Administration

Affolternstrasse 56, CH-8050 Zurich, Switzerland

The role of boards' misperceptions in the relation between managerial turnover and performance: Evidence from European football

Raphael Flepp* Egon Franck**

January 2019

Abstract

In this paper, we account for boards' misperceptions when replacing a top manager by differentiating between managerial turnovers following actual poor performance and managerial turnovers following seemingly poor performance due to bad luck in order to investigate their subsequent effects on performance. We focus on managerial changes within football organizations and analyze dismissals from the top European leagues. To account for the mean reversion of performance, we create a control group of non-dismissals using the nearest neighbor approach. To account for boards' misperceptions, we differentiate between dismissals and non-dismissals that occur either due to poor playing performance on the pitch or due to a sequence of bad luck, which is measured using "expected goals". We find that dismissals after poor playing performance on the pitch increase subsequent performance, while dismissals after a series of bad luck do not. Our results have important implications regarding the design of future turnover studies and the costs of boards' ineffective turnover decisions.

JEL Classification: J44, L83
Keywords: managerial turnover, performance, football

*Corresponding author. Email: raphael.flepp@business.uzh.ch
**Raphael Flepp and Egon Franck: Department of Business Administration, University of Zurich.

1 Introduction

Top managers are typically viewed as critical to the success or failure of organizations (Alexandridis, Doukas, & Mavis, 2018). Thus, replacing a top manager is one of the most important decisions made by boards of directors. However, the consequences for performance ensuing from top management turnover are widely debated (e.g., Giambatista, Rowe, & Riaz, 2005).

Empirically, it is challenging to properly investigate the effectiveness of replacing top managers because poor performance that triggers managerial turnover tends to coincide with bad luck *and* low manager ability (Huson, Malatesta, & Parrino, 2004). Thus, to account for mean reversion in performance, it is crucial to compare turnover events to a control group of non-turnover events with similarly poor prior performance (Giambatista et al., 2005). Moreover, Kaplan and Minton (2012) and Jenter and Kanaan (2015) show that corporate boards fail to filter out exogenous elements, and CEOs are dismissed after poor firm performance caused by factors beyond their control. This finding implies that boards falsely infer top management ability from exogenous performance components, which could disguise any positive replacement effect. While several studies considered the selectivity of management turnovers in large public firms (Huson et al., 2004) and sports teams (e.g., Paola & Scoppa, 2012; van Ours & van Tuijl, 2016), how boards' misperceptions when replacing a top manager affect the turnover-performance relationship has been neglected so far.

In this paper, we differentiate between managerial turnovers following actual poor performance and managerial turnovers following seemingly poor performance due to bad luck or other exogenous factors. To do so, we focus on managerial changes within football organizations. Football is a multibillion-dollar industry, and football clubs' boards frequently dismiss their head coaches which usually attracts high media attention. Furthermore, sports is an important context in which to study leader succession, and the use of data from sports organizations is well established in the literature (Giambatista et al., 2005).

Using the sports industry offers several advantages compared to conventional industries. First, the main objective of football clubs is to win as many games as possible, whereas the objectives of conventional firms are less clear (Hughes, Hughes, Mellahi, & Guermat, 2010). Second, team performance is publicly observed on a weekly basis, while accounting measures of firm performance are mostly annual. Finally, the industry is relatively homogeneous, with teams having similar organizational structures and technology constraints (Audas, Dobson, & Goddard, 2002).

In the football industry, exogenous factors are likely to play an even more important role than in conventional industries. Because football is a low-scoring sport, the random component in a single game is estimated to be approximately 50% (Anderson & Sally, 2013). Several studies have shown that decision makers underestimate the role of randomness in match outcomes and assign too much weight to the observed outcomes when they evaluate performance, which is commonly referred to as the outcome bias (e.g., Lefgren, Platt, & Price, 2015; Gauriot & Page, 2018). As a consequence, dismissal decisions of football club boards are prone to be based on misperceptions when exogenous factors shaped match outcomes (Brechot & Flepp, 2018).

This paper is the first that differentiates between dismissals and counterfactual dismissals that occur either due to poor playing performance on the pitch (*wise dismissals*) or due to a sequence of bad luck (*unwise dismissals*). If head coaches matter and replacing a bad manager with a more skilled one does have an impact, we expect that a team's performance would improve after wise dismissals when compared to a control group of teams with similar poor playing performance on the pitch but with no dismissal. After unwise dismissals, however, we expect that team performance would not improve compared to a control group of non-dismissal teams with a similar string of bad luck, due to the mean-reverting performance of both groups.

We use match-level data from the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A and the Spanish La Liga in the five seasons from 2013/2014 to 2017/2018. During this sample period, we registered 144 involuntary in-season coach dismissals. Following van Ours and van Tuijl (2016), we create a control group in which a similar performance below expectations did not trigger a dismissal. Using nearest neighbor matching based on the same team and similar "cumulative surprise", i.e., the sum of the differences between the number of actual points and expected points derived from betting odds, we were able to match 59 actual dismissals to 56 non-dismissals.[1]

To differentiate between wise and unwise dismissals, we compare the ranking of the teams in the official league table to a ranking of teams based on *expected goals* at the time of the (non-)dismissal decision. Expected goals (xG) reflect the sum of quantified scoring chances within a match and thus form a performance evaluation metric that is less prone to randomness because scoring chances occur much more frequently than actual goals.[2] As a result, the rank in the xG league should reflect a team's playing quality on the pitch better than the rank in the official league table because the latter is fully subject to the random component in match outcomes.

We classify a dismissal as being wise if the rank based on xG is either worse, equal, or only one rank better than the ranking in the official league table. In such situations, teams fall short of expectations due to poor playing performance on the pitch, and the replacement of the coach appears justified. If the rank based on xG is more than one rank higher than the actual rank, we classify dismissals as being unwise because such teams likely performed below expectations due to bad luck. Conversely, we classify a non-dismissal as being unwise if the rank based on xG is more than one rank worse than the rank based on the official league table, and we classify it as wise otherwise. To account for biased club board decisions and mean reversion effects, we must compare dismissals after poor playing performance on the pitch to non-dismissals after poor playing performance

[1] Several non-dismissals in the control group are matched to multiple actual dismissals.

[2] We employ shots as proxies for scoring changes because shots are direct attempts to score goals. In the 2017/18 season of the English Premier League, the average number of total shots per match was 24.4 compared to the average number of total goals of 2.7.

on the pitch (wise dismissals vs. unwise non-dismissals) and dismissals after bad luck to non-dismissals after bad luck (unwise dismissals vs. wise non-dismissals).

Using a team-season fixed effects regression model, we find that team performance significantly increases after wise dismissals but not after unwise non-dismissals. Thus, the F-test for equality of parameters is rejected, which implies that replacing the coach has a positive effect if the team performed below expectations due to insufficient playing quality on the pitch. By contrast, we find that team performance similarly increases after both unwise dismissals and wise non-dismissals. Thus, in situations where the team performed below expectations due to bad luck, team performance reverts to the mean independently of whether or not the coach has been dismissed. These results are robust to various methods regarding the classification of dismissals, the calculation of expected goals, and the matching of the control group.

Our paper makes several major contributions to the literature on managerial turnover and performance. First, we account for boards' biased decision making and show that the post-turnover performance critically depends on the situation preceding the turnover decision. A replacement of the manager is beneficial only if poor quality was indeed the driver that led to a performance below expectations. If, however, a replacement decision is triggered when bad luck was the main cause of poor performance, a replacement of the manager has no subsequent effect on performance. Second, our results could explain the mixed empirical findings in the literature. Depending on the fraction of turnover decisions that are based on boards' misjudgments in the sample, the average post-performance effect of the managerial change is biased towards zero due to the effect of mean reversion. Finally, the disclosure of suboptimal turnover decisions that trigger only financial consequences with no performance effects is complex. Because the post-turnover performance also improves after unwise turnovers, boards may evaluate their decisions as justified even though the same result could have been achieved without replacing the manager. Thus, boards should complement their existing performance evaluation strategies with more data-driven approaches to become aware of exogenous factors that may have shaped organizational performance.

The remainder of this paper is structured as follows: In Section 2, we review the literature and derive the hypotheses. In Section 3, we present our empirical methods. In Section 4, we present the results and in Section 5 we conclude.

2 Related literature and hypotheses

The bulk of the empirical evidence on managerial turnover and subsequent firm performance comes from event studies investigating the stock price reaction to top management turnover news. While most studies find significantly positive abnormal returns around managerial turnovers (e.g. Weisbach, 1988; Bonnier & Bruner, 1989), some studies fail to find any stock market reaction (e.g. Reinganum, 1985; Warner, Watts, & Wruck, 1988). However, the results from event studies are difficult to interpret because stock price reactions only reflect investors' expectations around the news event (Denis & Denis, 1995).

Several studies thus examine the development of accounting data over several years before and after the turnover. Using a sample of 721 turnovers between 1985 and 1988, Denis and Denis (1995) find that operating performance, measured as the ratio of operating income to total assets (OROA), improves significantly after forced turnovers and to a lesser extent also after voluntary turnovers. Further, Chang, Dasgupta, and Hilary (2010) proxy CEO ability by pay and firm performance under the CEO. Using a sample of 298 CEO departures between 1992 and 2002, they find that the post-departure industry-adjusted return on assets is worse if prior performance and prior relative pay were higher. Thus, their results provide evidence that CEO ability positively contributes to firm value. By contrast, Wiersema (2002) finds that neither the operating earnings nor the return on assets significantly improved after the 83 CEO dismissals that occurred in the 500 largest public US companies between 1997 and 1998.

Huson et al. (2004) criticize the use of unadjusted and industry-adjusted operating performance measures because poor performance tends to coincide with bad luck and low manager ability. Thus, an increase in subsequent accounting performance could be either due to the higher ability of the new manager or due to mean reversion of the operating performance measures. To account for this, Huson et al. (2004) construct a control group of similar firms without a turnover and adjust each sample firm's performance by subtracting the median performance of its comparison firm. In particular, they match each sample firm to a comparison firm with the same two-digit SIC code and a similar performance of $\pm 10\%$ over the year before the turnover. Using 883 voluntary and 119 forced CEO turnovers of large public firms during the 1971–1994 period, the authors find that the average change in the control group-adjusted OROA is positive and significant and that this effect is stronger for forced turnovers. Thus, Huson et al. (2004) conclude that managerial quality increases after a CEO turnover, which translates into higher operating performance.

In a recent paper, Alexandridis et al. (2018) examine the effect of forced CEO replacements on merger and acquisition performance. Using a final sample of 110 forced turnovers between 1994 and 2011, they find that the investment decisions of new CEOs improve relative to their predecessors. Further, the authors construct a control group of similar voluntary turnovers using the propensity score matching approach and show that mean reversion is not the driver of their results. Overall, the empirical evidence from turnovers in traditional industries remains mixed, with the bulk of studies finding a positive performance effect.

Several issues, however, remain unsolved. First, prior studies employed annual data but most turnovers take place within a given financial year of a firm. Thus, the operating or investment performance in the turnover year is not clearly attributable to either the old or the new manager. Second, operating performance measures are sensitive to managerial discretion. For example, Denis and Denis (1995) note that new managers might sell underperforming assets under value to boost operating performance. Finally, long-term strategies of the new manager might only be reflected in operating performance after several years (Ter Weel, 2011). To overcome these problems, several studies investigate the impact of managerial turnover in the sports industry. More specifically, these studies mostly concentrate on the dismissals of head coaches of football teams and the subsequent changes in match results on the pitch. Match results are thereby straightforward, observed weekly and clearly attributable to either the old or the new coach. Furthermore, decisions of the head coach with respect to the player line-up and playing strategy are effective immediately (Ter Weel, 2011).

The sports literature also acknowledged the selectivity of coach dismissals after a sequence of results below expectations and the need to form an appropriate control group. To find a counterfactual dismissal for each actual dismissal, Paola and Scoppa (2012) employed a nearest neighbor matching based on the ranking difference, the number of points obtained in the most recent four matches and the period in a particular season. Using 12 seasons from the Italian *Serie A* between 1997/98 and 2008/09, Paola and Scoppa (2012) find that the increase in team performance is solely due to mean reversion. Thus, changing the coach does not causally affect team performance. This finding is in line with several other studies using a control group of counterfactual dismissals (e.g., Balduck, Buelens, & Philippaerts, 2010; Ter Weel, 2011).

More recently, van Ours and van Tuijl (2016) formed the control group by matching the same team to itself based on the "cumulative surprise". The cumulative surprise measures how far teams perform below expectations and is calculated as the sum of the differences between the number of actual points and the expected points derived from betting odds. Using 36 dismissals and 33 matched counterfactual dismissals from the Dutch *Eredivisie* in the 14 seasons between 2000/01 and 2013/14, the authors also find that forced coach replacements do not improve team performance. Besters, van Ours, and van Tuijl (2016) replicate the methods of van Ours and van Tuijl (2016) using 45 dismissals and 34 counterfactual dismissals from the English Premier League between

2000/01 and 2014/15. Similarly, they conclude that, on average, performance does not improve following a coach's dismissal and that a successful managerial turnover remains highly unpredictable. Overall, most empirical evidence from professional football shows that coach dismissals have no causal effect on subsequent team performance.[3]

However, the effect of boards' misperceptions when replacing a top manager on performance has been neglected in the literature so far. While standard economic theory suggests that the board of directors should ignore the components of firm performance that are exogenous, several studies show that boards' CEO retention decisions are affected by exogenous shocks (Jenter & Kanaan, 2015). For example, Kaplan and Minton (2012) find that board-driven CEO turnover is significantly related not only to firm-specific performance but also to industry performance and the performance of the overall market which are beyond the control of the CEO. Further, Jenter and Kanaan (2015) use 3,365 CEO turnovers from 1993 to 2009 and find that corporate boards are more likely to fire a CEO after negative exogenous performance shocks. Thus, boards falsely infer CEO quality from exogenous performance components. Because the exogenous component is even more considerable in football match outcomes, it appears likely that club boards similarly attribute bad luck to poor coach quality when making dismissal decisions.

Thus, to properly test whether a coach's dismissal is beneficial for a football club, we differentiate between dismissals and non-dismissals that occur due to either poor performance on the pitch or just bad luck. We refer to a dismissal as being *wise* if a team performed below expectations due to poor playing performance on the pitch and to a dismissal as being *unwise* if a team performed below expectations due to a sequence of bad luck. Conversely, we refer to a counterfactual non-dismissal as being *wise* if a team performed below expectations due to a sequence of bad luck and a counterfactual non-dismissal as being *unwise* if a team performed below expectations due to poor playing

[3] Madum (2016) and Bryson, Buraimo, and Simmons (2018) are two notable exceptions. Madum (2016) finds a positive performance effect, but only for home games, and Bryson et al. (2018) find a positive performance effect after dismissals, but no effect after voluntary quits.

performance on the pitch. Table 1 illustrates this theoretical decomposition of dismissals and non-dismissals.

Table 1
Theoretical decomposition of dismissals and non-dismissals.

	dismissal	non-dismissal
wise	Below expectations due to poor playing performance on the pitch	Below expectations due to a sequence of bad luck
unwise	Below expectations due to a sequence of bad luck	Below expectations due to poor playing performance on the pitch

To gain insight from this decomposition, dismissals after poor playing performance on the pitch must be compared to non-dismissals after poor playing performance on the pitch, and dismissals after bad luck must be compared non-dismissals after bad luck. We expect that a dismissal is beneficial for the team if a coach whose team played poorly on the pitch is replaced. Thus, we formulate the first hypothesis as follows:

H1. *Wise dismissals improve performance relative to a control group with a similar poor performance on the pitch but with no dismissals (unwise non-dismissals).*

We expect the performance following an unwise dismissal to increase in a manner similar to that following wise non-dismissals because of simple mean reversion. Thus, we hypothesize the following:

H2. *Unwise dismissals do not improve performance relative to a control group with similar bad luck but with no dismissals (wise non-dismissals).*

In the next section, we present our methods and outline how we classify (non-)dismissals as wise or unwise in order to test our hypotheses.

3 Methods

3.1 Data

We employ data from the English Premier League, the French Ligue 1, the German Bundesliga, the Italian Serie A and the Spanish La Liga for the five seasons from 2013/2014 to 2017/2018. In total, our data set contains 9,130 matches for which we collected the date, the teams playing, the final match score, and the head coaches of the teams. For every in-season coaching change, we recorded whether it was a dismissal or a voluntary quit.[4] As a result, we registered 144 in-season coach dismissals during our sample period. Additionally, we collected betting odds from the bookmaker *B365* for each match and the final rank of all clubs in the previous season.[5]

3.2 Treatment and control group

Because dismissals are not exogenous, we construct a control group of teams without a dismissal but otherwise identical characteristics to infer the causal effect of dismissals on team performance. We replicate the methods of van Ours and van Tuijl (2016) and match dismissals (treatment group) to non-dismissals (control group) based on the same team and the closest "cumulative surprise" (CS) using the nearest neighbor approach. The CS measures deviations from expectations and is defined as the sum of the differences between the actual number of points won in a match and the expected number of points based on

[4] The data mainly stems from www.transfermarkt.com and www.football-data.co.uk. Following van Ours and van Tuijl (2016), we ignore coaching changes within the first four and last four match weeks. Further, following Besters et al. (2016), we only consider the first coach dismissal of a team within a season. For each coaching change, we documented at least one source that unambiguously states that the change was involuntarily. To validate our data collection, we compared our list of dismissals in the Premier League for the seasons 2013/14 and 2014/15 to the list of dismissals from Besters et al. (2016). The lists are identical, with the only exception being Crystal Palace in 2014/15 where we ignored the dismissal of Keith Millen in matchweek 2, but considered the dismissal of Neil Warnock in matchweek 18 instead.

[5] We downloaded the betting odds from www.football-data.co.uk. For two matches where the betting odds from *B365* were missing, we used the odds from *Bwin*.

betting odds.[6] To qualify for the control group, a non-dismissal must stem from another season without managerial change, and the CS values must not differ more than 0.5.[7] A matching based on the same team accounts for unobserved heterogeneity among teams because some teams might be more likely to dismiss a coach under similar circumstances than other teams (van Ours & van Tuijl, 2016). Further, matching controls for time-constant seasonal aspirations of teams, such as qualifying for the UEFA Champions League or avoiding relegation. At the same time, matching counterfactual situations based on the CS ensures that the control group team similarly performed below expectations without dismissing the coach afterwards.

Following this matching procedure, we were able to match 59 out of 144 actual dismissals to 56 non-dismissals.[8]

3.3 Wise and unwise (non-)dismissals

The teams from both the treatment and control group performed similarly below expectations around the (non-)dismissal. However, disappointing results below expectations could either be due to poor team performance on the pitch or simply due to bad luck. Because football is a low-scoring game, a team might occasionally lose a game or end a match in a draw against expectations, even though the team played well on the pitch. Thus, we need a measure of performance that is less prone to the randomness associated with match results.

To do so, we draw on the concept of "expected goals" (xG) based on the sum of quantified scoring chances.[9] Scoring chances are the second to the last step in the goal production process, and all teams try to create valuable chances in order to score goals. However, scoring chances occur much more often than goals, and thus, the expected goal

[6] The expected number of points is calculated as [(probability of win)·3] + [(probability of draw)·1]. For more detailed information, see van Ours and van Tuijl (2016).

[7] As for actual dismissals, we ignore non-dismissals within the first four and last four matches within a season. Alternative maximum CS differences of 0.1, 0.25, 0.75 and 1.0 do not alter our main conclusions.

[8] Several non-dismissals in the control group are matched to multiple actual dismissals.

[9] See Brechot and Flepp (2018) for a detailed description of expected goals as measure of performance.

metric is less prone to random variation. Indeed, Brechot and Flepp (2018) show that the xG metric contains more relevant information with regard to future team performance than do match outcomes.

Following Brechot and Flepp (2018), we consider shots as scoring chances and estimate their scoring probability based on the distance, the angle, the rule setting of the shot (i.e., open play, free kick or penalty kick), and the body part used. Additionally, we include team fixed effects in the logistic regression to account for unobserved team quality characteristics. In total, we estimate the scoring probability of 214,194 shots from all 9,130 matches in our sample.[10] Finally, we aggregate the quantified scoring chances for each team within each match to derive the number of expected goals per match.

The expected goal metric allows us to determine the better team on the pitch in terms of creating valuable scoring chances. Instead of awarding the number of points based on the actual goals within a match, i.e., three for a win, one for a draw and zero for a loss, we award the points based on expected goals.[11] A team wins the match and gains three points based on expected goals if it scores 0.5 or more expected goals than its opponent. If the expected goals of both teams lie within 0.5, we consider it a draw, and each team receives one point. Otherwise, the team loses the match based on expected goals and no points are gained. Finally, we rank the teams according to their points based on expected goals and construct an xG league table.

The rank in the xG league table should reflect a team's playing quality on the pitch more accurately than the rank in the official league table (OLT), because the OLT is solely based on actual match outcomes where bad luck fully translates into fewer points and a lower rank. For example, a team could play well on the pitch and win a match in terms of expected goals because that team created scoring chances of higher total value than its opponent. However, this team might actually lose the match in terms of outcomes because the scoring chances did not translate into actual goals. In such situations, the rank in the

[10] The shot data are provided by the measurement and data analytics company Nielsen.

[11] Note that points based on expected goals differ fundamentally from expected points based on betting odds. The first are based on actual performance on the pitch *after* the game whereas the latter are based on expectations *prior* to the game.

xG table should be higher than the rank in the OLT. In other situations where the team played poorly on the pitch, the rank in the xG table should be equal to or worse than the rank in the OLT. Consequently, we use the difference between the rank in the OLT and the rank in the xG table to differentiate between wise and unwise (non-)dismissals.

Figure 1 plots the rank in the OLT against the rank in the xG table at the time a (non-)dismissal took place. We categorize a dismissal as being wise if the rank based on

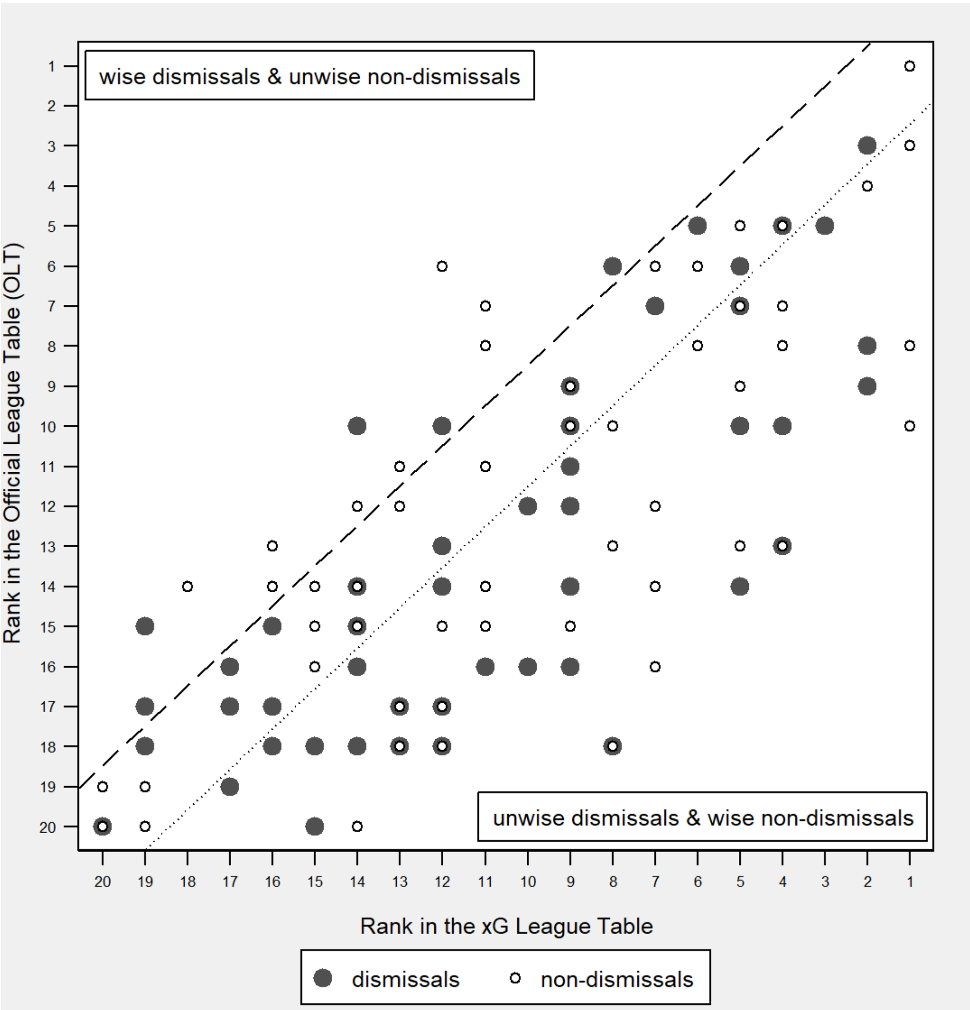

Figure 1
Rank in the OLT vs. rank in the xG table of all matched dismissals and non-dismissals. The dotted line marks the cutoff for wise and unwise dismissals. The dashed line marks the cutoff for wise and unwise non-dismissals.

xG is equal to or lower than the rank in the OLT. Additionally, we allow for a margin in favor of wise dismissals because we do not know the detailed circumstances that triggered the decision to dismiss the coach. In particular, we still categorize a dismissal as being wise even if the rank based on xG is one rank better than the rank in the OLT. Only if the ranking based on xG is more than one rank better we do categorize a dismissal as being unwise.[12] Such teams performed below expectations due to bad luck, because their ranking based on xG indicates a higher playing quality on the pitch. The dotted line in Figure 1 marks the cutoff for wise and unwise dismissals. Dismissals above the dotted line are categorized as being wise, whereas dismissals below the dotted line are classified as being unwise. Conversely, we categorize a non-dismissal as being unwise if the ranking based on xG is more than one rank worse than the ranking based on the OLT, and we categorize it as wise otherwise. The dashed line in Figure 1 marks the cutoff for wise and unwise non-dismissals. As a result, all dismissals and non-dismissals in the area between the dashed and the dotted line are categorized as being wise.

3.4 Econometric specification

Following van Ours and van Tuijl (2016), we estimate the following regression model:

$$\begin{aligned} \gamma_{ijk} = \quad & \alpha_{ik} + \beta_1 \cdot \textit{wise dismissal}_{ijk} + \beta_2 \cdot \textit{unwise non-dismissal}_{ijk} \\ & + \beta_3 \cdot \textit{unwise dismissal}_{ijk} + \beta_4 \cdot \textit{wise non-dismissal}_{ijk} \\ & + \gamma_1 \cdot \textit{home}_{ijk} + \gamma_2 \cdot \textit{rank opponent}_{ijk} + \epsilon_{ijk} \end{aligned} \tag{1}$$

where i denotes the team, j indicates the match and k refers to the season. For the dependent variable γ_{ijk}, we employ the same performance measures as in van Ours and van Tuijl (2016), namely, the number of points gained in a match (*points*), a dummy variable that indicates whether a match was won or not (*win*) and the goal difference in a match (*goal diff*). The (non-)dismissal variables are dummies that indicate whether or not there has been a (non-)dismissal of the relevant type earlier in the season. We are mainly interested in the difference between the coefficients β_1 and β_2 as well as in the

[12] In Section 4, we test the robustness of our results based on an alternative margin.

difference between the coefficients β_3 and β_4. According to our hypotheses, we expect β_1 to be significantly larger than β_2 (H1) and β_3 to be equal to β_4 (H2).

To account for unobserved differences in team quality within a particular season, we include team-season fixed effects α_{ik}. Furthermore, we control for home field advantage by including the dummy variable *home* and proxy the strength of the opponent by controlling for the final rank of the opponent in the previous season (*rank opponent*).[13]

4 Results

4.1 Main results

Table 2 presents our main results for the four (non-)dismissal groups.[14] The total number of observations is 4,054 and consists of 2,198 matches from teams with a dismissal and 1,856 matches from teams with a non-dismissal.[15] Column (1) shows that after a wise dismissal, the number of points per match is 0.24 higher than before the dismissal. By contrast, the effect for the relevant control group of unwise non-dismissals is -0.09 points per match. The F-test for equality shows that these parameters are significantly different from each other ($F = 4.76$***). The findings for *win* and *goal diff* as performance measures in Columns (2) and (3) are equivalent. Thus, our first hypothesis that wise dismissals improve performance relative to unwise non-dismissals is confirmed.

Column (1) further shows that the number of points per match after unwise dismissals also significantly improves by 0.36. However, wise non-dismissals similarly improve team performance by 0.34 points per match. Indeed, the F-test for equality of these parameters cannot be rejected ($F = 0.09$). This result suggests that the improvement in points per

[13] As in van Ours and van Tuijl (2016), we assign the rank 20 (18 for the Bundesliga) to promoted teams.

[14] We also replicated the approach of van Ours and van Tuijl (2016) by pooling all dismissals in one treatment group and all non-dismissals in one control group. Like van Ours and van Tuijl (2016), we fail to find any significant performance difference between the treatment and the control group.

[15] Dismissal observations consist of 11×34 matches from the German Bundesliga and 48×38 matches from the other leagues. Non-dismissal observations consist of 11×34 matches from the German Bundesliga and 39×38 matches from the other leagues. Nine out of the 50 non-dismissal team-seasons serve as the control group for two dismissals. Of those, 3 have the same non-dismissal date, whereas 6 have a different non-dismissal date. For the latter, we employ the earliest non-dismissal date for the main analysis. However, if we double these observations and use the exact non-dismissal dates, the results remain unchanged.

Table 2
Main results.

	points (1)	*win* (2)	*goal diff* (3)
wise dismissal	0.24***	0.08***	0.29***
	(0.059)	(0.023)	(0.084)
unwise non-dismissal	-0.09	-0.01	-0.14
	(0.137)	(0.052)	(0.120)
unwise dismissal	0.36***	0.13***	0.37***
	(0.062)	(0.027)	(0.112)
wise non-dismissal	0.34***	0.13***	0.39***
	(0.055)	(0.020)	(0.086)
home	0.45***	0.15***	0.74***
	(0.034)	(0.010)	(0.047)
rank opponent	0.05***	0.016***	0.08***
	(0.003)	(0.001)	(0.005)
team-season fixed effects	Yes	Yes	Yes
R^2 overall	9.30%	7.65%	11.30%
N	4,054	4,054	4,054
F-test $\beta_1 = \beta_2$	4.76**	2.96*	4.03**
F-test $\beta_3 = \beta_4$	0.09	0.01	0.02

Notes: The table reports the coefficients estimated from a team-season fixed effects regression model with standard errors corrected for heteroskedasticity. Standard errors are reported in parentheses. In all models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

match after unwise dismissals would also have occurred if the coach had not been dismissed after all. The F-tests in Column (2) and (3) confirm this finding. Overall, our second hypothesis that unwise dismissals do not improve performance relative to the relevant control group of wise non-dismissals is supported.

4.2 Robustness checks

In this subsection, we address potential concerns regarding the robustness of our main results. First, a margin of one rank between the OLT and the table based on xG might not be sufficient to identify unwise (non-)dismissals. Thus, we extend the margin to -2 ranks for wise dismissals and to +2 ranks for wise non-dismissals. Column (1) of Table 3 shows the results for this alternative categorization. Again, the F-tests reveal that wise dismissals significantly increase the number of points gained within a match compared to unwise non-dismissals, whereas unwise dismissals do not improve performance compared to wise non-dismissals.

Table 3
Robustness checks.

	points		
	-2/+2 margin (1)	alt. xG model (2)	alt. matching (3)
wise dismissal	0.22*** (0.048)	0.23*** (0.063)	0.31*** (0.049)
unwise non-dismissal	-0.36*** (0.094)	-0.08 (0.152)	-0.10 (0.107)
unwise dismissal	0.44*** (0.079)	0.48*** (0.083)	0.39*** (0.038)
wise non-dismissal	0.33*** (0.053)	0.28*** (0.067)	0.33*** (0.048)
home	0.45*** (0.034)	0.46*** (0.042)	0.47*** (0.023)
rank opponent	0.05*** (0.003)	0.05*** (0.004)	0.04*** (0.002)
team-season fixed effects	Yes	Yes	Yes
R^2 overall	9.21%	9.53%	8.77%
N	4,054	2,862	8,504
F-test $\beta_1 = \beta_2$	29.89***	3.40*	12.20***
F-test $\beta_3 = \beta_4$	1.33	3.51*	1.11

Notes: The table reports the coefficients estimated from a team-season fixed effects regression model with standard errors corrected for heteroskedasticity. Standard errors are reported in parentheses. In all models, *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

Second, our analysis relies on the accuracy of xG as a performance evaluation measure. To test whether our results are robust to an alternative xG model other than the one employed by Brechot and Flepp (2018), we collected the xG values and the table ranks based on the xG from www.understat.com. Understat calculates expected goals based on a neural network prediction algorithm, including information on goal distance, angle, situation, last action, shot type, attack type, errors and "big chances", pass characteristics, dribbled players, and game state. Unfortunately, the data on www.understat.com are only available from the season 2014/15 onward. Thus, we drop all matches from the 2013/14 season and re-run our complete analysis with data from 2014/15 to 2017/18. This results in a matching of 41 dismissals to 39 non-dismissals. Column (2) of Table 3 shows the results if we employ the xG values from the alternative xG model of www.understat.com to categorize wise and unwise (non-)dismissals. The F-test rejects the equality of parameters of wise and unwise non-dismissals at the 10%-level. Surprisingly, the F-test that compares unwise dismissals and wise non-dismissals is also significant, indicating that teams also perform slightly better after unwise dismissals.

In a third robustness check, we omit the restriction that each dismissal team-season has to be matched to a non-dismissal team-season of the same team. Thus, for every dismissal, we now match a non-dismissal only based on CS. This alternative, less restrictive matching procedure ignores the heterogeneity across teams but allows us to match a non-dismissal to all 144 dismissals in our sample. As a consequence, the number of observations in the regression more than doubles to 8,504. Column (3) of Table 3 shows that the results remain very similar even if we apply this alternative matching procedure.[16] Overall, the robustness checks show that our main findings also hold under different alternative methods.

5 Conclusion

The findings of this paper show that the effect of managerial turnover on performance critically depends on whether boards falsely attributed bad luck to low manager ability or whether performance was indeed poor when deciding to replace the manager. Analyzing head-coach dismissals in European football, we show that dismissals after poor playing performance on the pitch increase subsequent team performance compared to a control group while, dismissals after a series of bad luck do not.

These results offer a reconciling explanation for the mixed empirical results regarding the effect of managerial turnover on subsequent performance. Depending on how pronounced the exogenous factors and the resulting boards' misperceptions are, the average

[16] Additionally, we run all alternative specifications with *win* and *goal diff* as dependent performance variables. The results remain unchanged except for the alternative xG model where the coefficient of wise dismissals is still larger than the coefficient of unwise non-dismissals, but equality of the coefficients cannot be rejected.

effect might be biased towards zero. Thus, our paper has important implications for the design of future studies that investigate the relationship between turnover and subsequent performance. Without accounting for the misperceptions of boards in their turnover decisions, the results might be misleading.

Furthermore, while corporate boards partially adjust for peer group performance when assessing CEOs, they appear to concentrate on the largest firms in their industry but fail to adjust for other, less obvious exogenous performance components (Jenter & Kanaan, 2015). Such less obvious components might only be filtered out by the use of systematic data-driven approaches. Similarly, football club boards should complement their existing decision-making strategies with more data-driven approaches, such as the use of expected goals to reduce ineffective decisions. Replacing a head coach is very costly, and the costs associated with unwise dismissals should better be invested in new players to increase the playing strength of the team.

References

Alexandridis, G., Doukas, J. A., & Mavis, C. P. (2018). Does firing a CEO pay off? *Financial Management*, 1-41.

Anderson, C., & Sally, D. (2013). *The numbers game: why everything you know about football is wrong.* Penguin UK.

Audas, R., Dobson, S., & Goddard, J. (2002). The impact of managerial change on team performance in professional sports. *Journal of Economics and Business*, *54*(6), 633–650.

Balduck, A.-L., Buelens, M., & Philippaerts, R. (2010). Short-term effects of midseason coach turnover on team performance in soccer. *Research Quarterly for Exercise and Sport*, *81*(3), 379–383.

Besters, L. M., van Ours, J. C., & van Tuijl, M. A. (2016). Effectiveness of in-season manager changes in English Premier League football. *De Economist*, *164*(3), 335–356.

Bonnier, K.-A., & Bruner, R. F. (1989). An analysis of stock price reaction to management change in distressed firms. *Journal of Accounting and Economics*, *11*(1), 95–106.

Brechot, M., & Flepp, R. (2018). Dealing with randomness in match outcomes: how to rethink performance evaluation and decision-making in European club football. *UZH Business Working Paper, Department of Business Administration, University of Zurich*(374).

Bryson, A., Buraimo, B., & Simmons, R. (2018). Special ones? The effect of head coaches on football team performance. *Working Paper*.

Chang, Y. Y., Dasgupta, S., & Hilary, G. (2010). CEO ability, pay, and firm performance. *Management Science*, *56*(10), 1633–1652.

Denis, D. J., & Denis, D. K. (1995). Performance changes following top management dismissals. *Journal of Finance*, *50*(4), 1029–1057.

Gauriot, R., & Page, L. (2018). Fooled by performance randomness: over-rewarding luck. *Review of Economics and Statistics, forthcoming.*

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

Hughes, M., Hughes, P., Mellahi, K., & Guermat, C. (2010). Short-term versus long-term impact of managers: evidence from the football industry. *British Journal of Management*, *21*(2), 571–589.

Huson, M. R., Malatesta, P. H., & Parrino, R. (2004). Managerial succession and firm performance. *Journal of Financial Economics*, *74*(2), 237–275.

Jenter, D., & Kanaan, F. (2015). CEO turnover and relative performance evaluation. *Journal of Finance*, *70*(5), 2155–2184.

Kaplan, S. N., & Minton, B. A. (2012). How has CEO turnover changed? *International Review of Finance*, *12*(1), 57–87.

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

Madum, A. (2016). Managerial turnover and subsequent firm performance: evidence from Danish soccer teams. *International Journal of Sport Finance*, *11*(1).

Paola, M. D., & Scoppa, V. (2012). The effects of managerial turnover: evidence from coach dismissals in Italian soccer teams. *Journal of Sports Economics*, *13*(2), 152–168.

Reinganum, M. R. (1985). The effect of executive succession on stockholder wealth. *Administrative Science Quarterly*, 46–60.

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

Warner, J. B., Watts, R. L., & Wruck, K. H. (1988). Stock prices and top management changes. *Journal of Financial Economics*, *20*, 461–492.

Weisbach, M. S. (1988). Outside directors and CEO turnover. *Journal of Financial Economics*, *20*, 431–460.

Wiersema, M. (2002). Holes at the top. why CEO firings backfire. *Harvard Business Review*, *80*(12), 70–7.