Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and operational research

The causes and consequences of in-season wise and unwise coach changes and their effect on team performance

Job Jeeninga (566029)

Supervisor:	Dennis Fok
Second assessor:	Jan C. van Ours
Date final version:	30th June 2023

Contents

Abstract

A coach replacement is a crucial but drastic way of trying to improve a team's performance. It is important to understand what leads to a coach change and what consequences a change has on performance. This paper uses data from 6 consecutive seasons, from 2017/2018 until 2022/2023, from the major Dutch football league, the Eredivisie. I differentiate between coach replacements that are a result of poor performance on the pitch (wise change) or bad luck (unwise change), categorizing by means of the expected goals. As football is a low scoring game, the expected goals reflect the on pitch performance of a team better and are less vulnerable to random variation than the actual match result. The performance before and after a coach change is compared to a control group with similar poor performance or bad luck but without a coach change. The results show that poor performance, measured as the difference between the actual results and the expected results based on bookmaker odds in the previous couple of games, leads to a coach change. Moreover, replacing the coach when a club is expected to be ranked better, based on expected goals, does improve the performance.

1 Introduction

It is crucial to understand the effects of coach changes on team performance. In professional sports, when a team experiences bad results, head-coaches are often replaced to try and improve the performance. An in-season change of the head-coach is a drastic measure and can be expensive. Therefore, it is of great importance to know what leads to a coach quitting or being dismissed and what consequences this has on performance. This is of interest to the general manager of a sports club, who hires and fires the coach, but this may also provide information on managerial dismissals in non-sport-related companies. Namely, Kahn (2000) states that analysing professional sports data can provide valuable information for labour market research, as professional sports contain extensive and detailed data that is widely available, especially compared to microdata from businesses, which is often kept private. Pieper, Nüesch and Franck (2014) argue that the effect of a coach change on team performance can be related to the consequences a managerial dismissal has on a company.

For these reasons, this study tries to answer the question: What are the causes and consequences of in-season head coach changes on team performance? Whereas a considerable number of studies research the causes of coach dismissals or the effects of coach changes on performance, this study gives an overview of both the causes and consequences of coach replacements. The effects of coach changes on team performance are investigated using a relatively new and promising method introduced by Flepp and Franck (2021). This method studies the effects of wise and unwise coach dismissals on performance. The data used in my research is from the Eredivisie, the highest level of professional football in the Netherlands, from season 2017/2018 until season 2022/2023. This data set contains data from the most recent seasons and is, to my knowledge, never used before in this area of research. To answer the research question, multiple components are investigated based on three sub-questions.

The first sub-question this paper investigates is: *What are the determinants of an in-season head-coach change?* Multiple studies have used a duration analysis with a hazard model to determine what the causes are for coach dismissals and to estimate when a dismissal is going to happen (Bachan, Reilly and Witt (2008), Barros, Frick and Passos (2009), d'Addona and Kind (2014), Semmelroth (2022)). The use of a hazard model seems appropriate when analysing the duration until the next coach dismissal. d'Addona and Kind (2014) find that a change in league position is a determinant of a coach dismissal as well as multiple coach statistics. Consequently, my study incorporates a change in league position as a possible determinant of a coach change. Semmelroth (2022) discovers that team performance related to expected playoff qualification influences coach changes. Moreover, de Dios Tena and Forrest (2007) find that if a club is in the relegation zone, there is a higher chance of a coach change. Although the results differ as the papers look into different determinants of coach dismissals, most papers do point to recent bad team performance as a determinant of a coach dismissal. Thus, my study looks into the effect of recent performance, captured by the points scored, on the probability of a coach change. Pieper et al. (2014) and Van Ours and Van Tuijl (2016) discover that deviations from performance expectations, based on bookmaker odds, can influence the decision of firing a coach. Therefore, my paper investigates deviations from performance expectations, based on bookmaker odds, as a potential determinant of a coach replacement.

The second sub-question this paper investigates is: *What is the difference in performance before and after an in-season coach change?* A naive approach to researching the effect of a coach change on performance would be to compare the performance before and after the coach change. However, multiple studies show that it is important to account for the selectivity of head coach change which is often done by the use of a control group (Ter Weel (2011), Paola and Scoppa (2012), Van Ours and Van Tuijl (2016)). These papers use a difference-in-differences approach by comparing the team performance before and after dismissal to the performance under similar circumstances without a coach dismissal. Accordingly, this study uses a naive approach to analyse the difference in performance before and after a coach change and to emphasise the importance of a control group when making conclusions.

A considerable number of studies have investigated the effects that in-season coach changes have on the performance on the pitch. Most studies find that a coach dismissal has no positive effect on performance compared to a control group (Heuer, Müller, Rubner, Hagemann and Strauss (2011), Paola and Scoppa (2012), Van Ours and Van Tuijl (2016)). However, de Dios Tena and Forrest (2007) show that a coach replacement improves the performance of a club in home matches, and Muehlheusser, Schneemann and Sliwka (2016) discover that a dismissal enhances performances for homogeneous teams where the individual quality of the players is similar. The results of the studies differ, hence it is important to research the effect of an in-season coach change on performance. Therefore, the third sub-question this paper investigates is: *What is the effect of wise and unwise coach changes on team performance?* Flepp and Franck (2021) emphasise the challenge of appropriately dismissing a coach, as both bad luck and a low coaching ability can lead to poor performance. Their research revolutionises the analysis of the effect of coach changes on team performance by using expected goals as a way of categorising coach dismissals as wise or unwise. This useful categorization is supported by the results of Rocaboy and Pavlik (2020) which show that a coach dismissal can be beneficial for performance only if the team under-performed before the dismissal (which is not the case when the results are a result of bad luck). Expected goals are a good indicator of team performance as it is a measurement of performance, less vulnerable to random variation than actual match outcomes, which is backed up by Brechot and Flepp (2020). They find that wise coach dismissals improve the subsequent performance of a team compared to a control group where no coach is replaced. For this reason, to investigate the effects of a coach changes on performance, my paper uses the method from Flepp and Franck (2021) and categorises changes as wise and unwise to incorporate in the control group approach. Whereas Flepp and Franck (2021) use match-level data from the Premier League (England), Ligue 1 (France), Bundesliga (Germany), Serie A (Italy) and the La Liga (Spain), my study uses a different data set, which contains more recent data from a different league, the Eredivisie (Netherlands).

The results of this study show that underperformance compared to expectations based on bookmaker odds is a cause of in-season head-coach changes. Also, teams seem to perform better after a coach change than before the replacement when they are expected, based on expected goals, to be ranked better. However, after coach changes that take place when a club is expected to be ranked worse than their actual ranking, the performance seems to be worse than before the change. Nonetheless, when compared to a control group, the effect of such a coach change is not significant. The results indicate that, when a team is expected to be ranked better, coach changes have a significant positive effect on the performance of a team compared to a control group.

2 Data

To answer the main research question and the corresponding sub-questions, data from multiple data sources is used. Eredivisie match level data is used because Flepp and Franck (2021) have done a similar research using data from the five biggest professional football competitions, not including the Dutch football competition, the Eredivisie.

Most of the data used in this study is from *www.football-data.co.uk* (2023) which contains a significant number of variables about matches such as full- and halftime results, match statistics and bookmaker odds. In this study, the 12 variables used are the season, date of the match, home team, away team, full-time home goals, full-time away goals, full-time result (home win, draw or away win), BET365 bookmaker odds (home win, draw or away win), and the shots on target for the home team and away team. The data from seasons 2017/2018 until 2022/2023 are used because those seasons include the home and away shots on target of the Eredivisie matches, which are crucial variables because these are used in the categorization of wise and unwise coach changes. The dataset contains the match-level data of the 1743 matches in the six seasons, ordered by date. The Eredivisie consists of 34 rounds in which 9 games are played between 18 clubs, resulting in 306 matches each season. Note that season 2019/2020 contains fewer matches (232) due to the early cancellation of the season at round 27 because of COVID-19, resulting in 25 or 26 games played per club. Also, season 2022/2023 contains fewer matches (287) as the season was not finished when doing the study, resulting in 31 or 32 matches played per club. To study the determinants and effects of coach changes on performance, the data is transformed such that it is ordered per team-season, which contains all the matches of a certain club in a season resulting in double the number of observations (3486) as each match is viewed from the perspective of the home team as well as the away team.

Season	Changes	Clubs
2017/2018	6	Ajax, Sparta Rotterdam, FC Twente, FC Utrecht, Vitesse, Willem II
2018/2019	5	Excelsior, SC Heerenveen, NAC Breda, FC Utrecht, Zwolle
2019/2020	5	ADO Den Haag, Feyenoord, PSV Eindhoven, Vitesse, VVV Venlo
2020/2021	7	AZ Alkmaar, ADO Den Haag, Fortuna Sittard, FC Utrecht, VVV Venlo, Willem II, PEC Zwolle
2021/2022	6	Cambuur, SC Heerenveen, Sparta Rotterdam, FC Utrecht, Willem II, PEC Zwolle
2022/2023	5	Ajax, Cambuur, FC Groningen, FC Utrecht, Vitesse
Total	34	

Table 1: The in-season coach changes in the Eredivisie.

The coach change data is from *www.transfermarkt.com* (2023) which contains all the coach changes in the Eredivisie. This study uses the in-season coach changes from Eredivisie seasons 2017/2018 until 2022/2023. If a club experiences multiple coach replacements in one season, only the first coach change is used. The coach changes during the first 4 rounds and the last 4 rounds of the season are ignored to be able to compare the performance of a team before and after a coach change. Table 1 shows the coach changes for each season after filtering and shows the clubs that experienced an in-season coach change in a particular season. The number of in-season coach replacements does not vary much across the different seasons, as the minimum number of coach replacements in a season is 5 and the maximum number of changes is 7 (in season 2020/2021).

To compare the performance of a club before and after a coach change, it is important to account for the quality of the opponents. This is done by using the rankings of the clubs at the end of the Eredivisie seasons 2016/2017 until 2021/2022 which are provided by *www.espn.nl* (2023). The final ranking of the opponent in the previous season is used to capture the quality of that opponent in the current season.

To research the effect of wise and unwise coach changes on performance, the expected goals are used to determine the expected match result based on the performance of the involved teams. These expected results are then used to categorise coach changes as wise or unwise and to create a control group for those coach changes. This method is explained in detail in Section 3.3. The expected goals are calculated using data from *www.fbref.com* (2023) which contains data on the goals per shot on target for each club in each season. Note that, by construction, this variable takes on values on the interval [0, 1] and describes the probability of a shot on target going in. The goals (G) per shot on target (SoT) variable, G/SoT, is used to calculate the expected goals (xG) for club i in match j in season k with $xG_{ijk} = SoT_{ijk} \cdot (G/SoT)_{i,k-1}$. Here, the goals per shot on target are from the previous season of that club, hence $k-1$, to prevent a forward-looking-bias of using an average of a season that is not finished yet. This way, only statistics are used that are available at the time of the match. The expected goals represent the number of goals expected by a team in a certain match, conditional on the average number of shots on target needed to score a goal. Note that clubs that were promoted to the Eredivisie in the previous season lack the goals per shot on target statistic. For those clubs, the G/SoT for season k are calculated by taking the average of the G/SoT of the clubs promoted a year before which therefore played in the Eredivisie in season $k-1$.

The characteristics and correlations of multiple match statistics of team-seasons with a coach change are given in Table 2. The characteristics and correlations of the match statistics of team-

		Mean	SD	Min.	Max.	1	2	3	4	5	6	7	8
1	Win	0.32	0.32	0	1	1							
2	Points	1.20	1.30	0	3	0.95	1						
3	Goal difference	-0.29	2.10	-13	8	0.74	0.83	1					
4	Goals scored	1.41	1.31	0	8	0.65	0.67	0.73	1				
5	xPoints	1.22	1.24	0	3	0.50	0.56	0.60	0.47	1			
6	xGoal difference	-0.15	1.39	-7.36	7.00	0.50	0.57	0.69	0.50	0.85	1		
7	xGoals scored	1.31	0.85	0	7.6	0.41	0.45	0.52	0.59	0.71	0.79	1	
8	Cumulative surprise	-0.47	2.38	-7.09	7.37	0.46	0.47	0.34	0.30	0.15	0.15	0.11	1

Table 2: The descriptive statistics and correlations of various variables for the team-seasons in which a club experienced an in-season coach change.

seasons without a coach change are shown in Table 7 in Appendix A. The difference between the actual results and the expected results based on bookmaker odds is captured in the cumulative surprise statistic, whose calculation is described in Section 3.1. The expected statistics (indicated with x...) are based on the expected goals, and the exact calculations of xPoints are given in Section 3.1. Table 2 shows that teams that have an in-season coach change only win 32 percent of their matches, on average, in that season. Also, they score fewer goals and are expected to score fewer goals than their opponents, given that the average (expected) goal difference is negative. There is a strong correlation between the variables, especially between the actual match statistics (1, 2, 3 and 4) with a minimum correlation of 0.65 and between the expected statistics (5, 6 and 7), with the lowest correlation being 0.71. The cumulative surprise has a weak to moderate correlation with the real match statistics, with values ranging from 0.30 to 0.47, and very weak correlations with the expected statistics.

3 Methodology

In this section the methods and techniques used to answer the main question and the three sub-questions are described. To research what effects a coach change has on team performance, this study uses two approaches: a naive approach (see Section 3.2) and a control group approach (see Section 3.3).

3.1 Determinants of a coach change

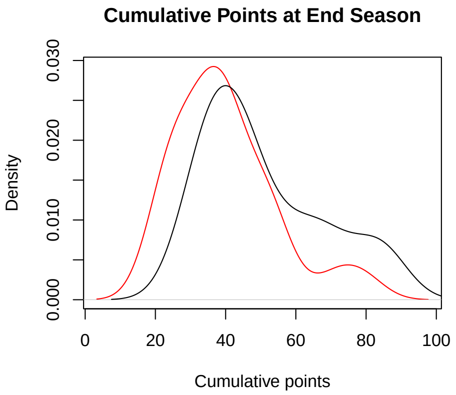
To research what determinants lead to a coach dismissal and answer sub-question 1 an exponential hazard model is used to estimate the duration until the next coach dismissal. The determinants that are investigated are the cumulative surprise, the cumulative points and a change in league position. These three potential determinants are performance indicators of a team and are time-varying indicators. This study follows the method of Van Ours and Van Tuijl (2016) to research the causes of in-season coach changes. However, the change in league position is incorporated as a possible determinant of a coach change instead of the coach characteristics used by Van Ours and Van Tuijl (2016).

The cumulative points of a club are the total points earned over the last four matches, and therefore can take on values on the interval $[0, 12]$. Also, the results when using the cumulative points over the last five and last six matches are studied to analyse the sensitivity of the findings. The change in league position is the difference between the league position of a club after the previous round and their current league position. If this value is positive, this means that a club climbed the rankings and vice versa. First, a match surprise is defined to explain what a cumulative surprise is. A match surprise is the difference in the expected number of points and the actual number of points a team receives for the match result (win=3, draw=1, loss=0 points). The expected number of points is based on the bookmaker odds, where the implied probability of a certain result is calculated by the formula

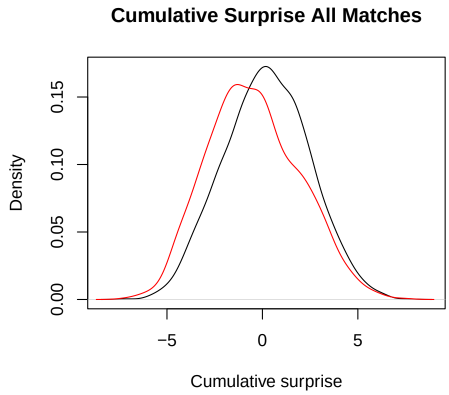
$$p_i = \frac{1}{O_i \cdot \text{harmonic sum}} \tag{1}$$

where O_i are the odds for $i = H, D, A$ (home win, draw and away win) and the harmonic sum is given by $\sum_{i=1}^{3} 1/O_i$. The total implied probability of the events happening is greater than 1, which is called the overround, this is the bookmaker's profit margin. By dividing by the harmonic sum, this calculation corrects for the bookmakers overround and sets the total probability of the results equal to 1. The expected number of points ($xPoints$) is calculated using the bookmakers implied probabilities with the formula $xPoints_{home} = 3p_H + p_D$ for the home team and $xPoints_{away} = 3p_A + p_D$ for the away team. The match surprise of match i is defined as the difference in the actual number of points received and the expected number of points in match i and given by $MS_i = Points_i - xPoints_i$. The cumulative surprise of a team is the summation of the last four match surprises and is given by $CS_i = \sum_{j=0}^{3} MS_{i-j}$. If the cumulative surprise is positive, a club has performed better than the bookmakers expected. If the cumulative surprise is negative, the club has performed worse than expected by the bookmakers. Figure 1a shows the kernel densities of the cumulative points at the end of a finished season, thus not including seasons 2019/2020 and 2022/2023 (see Section 2), for team-seasons with and without a coach change. The cumulative points at the end of a team-season with a coach change is generally lower than that of a team-season without a coach change, as the estimated density function is shifted to the left. Also, a high number of cumulative points (60 or higher) seems to be more scarce for team-seasons with a coach change. Figure 1b shows the kernel densities of the cumulative surprise of the previous four matches for team-seasons with and without a coach change. The cumulative surprise is generally lower for team-seasons with a coach change, meaning that those teams generally have perform worse, compared to the expected results based on the bookmaker odds, than the clubs without a coach change in that season.

For sensitivity analysis, the points over the last five and six matches are used as a potential determinant of coach changes. The models cannot be directly compared because the number of observations differs per model. Therefore, the fit of the models is compared with the Bayesian Information Criterion (BIC), which is used because it accounts for the number of observations used in a model. The criterion is given by $BIC = k\ln(n) - 2\ln(\hat{L})$, where k is the number of parameters, n the number of observations and $\hat{L}$ the maximised value of the likelihood function. The lower the BIC value, the better a model fits the data.



(a) Kernel densities of the cumulative points at the end of a season for seasons with (red) and without (black) a coach change.



(b) Kernel densities of the cumulative surprise, using the previous 4 matches, for seasons with (red) and without (black) a coach change.

Figure 1: Kernel densities of cumulative points and cumulative surprise.

3.1.1 Proportional hazard model

The proportional hazard model gives the duration until a coach change conditional on the time-varying variables mentioned in Section 3.1 and is given by the formula

$$\theta(t|z_t) = \exp(z_t'\gamma_j) \tag{2}$$

where z_t indicates the time-varying variables: cumulative surprise, cumulative points and a change in league position. Here, γ is the vector of the coefficients of the different variables and represents the relationship between the variables over the in-season duration of a head-coach spell. A proportional hazard model is used because it can handle time-varying covariates. I assume that the proportional hazard assumption holds: the effects of the covariates (potential determinants) on the hazard rate are constant over time. This model assumes season homogeneity: the hazard functions across different seasons are the same, or put differently, the probability of a coach change happening does not differ across seasons. For sensitivity analysis purposes, this paper also studies the results allowing for season heterogeneity: the hazard rate varies across different seasons, meaning the chance of a coach being replaced can differ across the seasons. This study looks into the in-season duration of a head coach and the time of a spell is measured in matches played until a coach change. Therefore, a duration spell is right censored if there is no in-season coach replacement. The maximum likelihood estimator is used to estimate the coefficients (γ) of the hazard model.

3.2 Effects of a coach change on performance: Naive approach

The naive approach compares the difference in team performance before and after the coach replacement and provides an answer to sub-question 2. This approach is not meant to prove causality between a coach change and the difference in team performance after the coach change. However, this approach does give an overview of the observed team performance around a coach change and provides a logical step to the control group approach by highlighting the importance of a control group when trying to study the causality between coach replacement and team performance.

3.2.1 Linear regression

The naive approach uses the linear regression from Van Ours and Van Tuijl (2016) to expose the relationships between multiple performance indicators and explanatory variables, such as a coach change. The linear regression is given by the formula

$$y_{ijk} = \eta_{ik} + r'_{ijk}\beta + \delta_1 Wise\ change_{ijk} + \delta_2 Unwise\ change_{ijk} + \epsilon_{ijk} \tag{3}$$

where y_{ijk} is the performance indicator for club i in match j of season k. In this paper, three performance indicators are used: a dummy variable for whether a team won or not (win=1 and loss=0), the goal difference and the number of points. To account for the quality of a club in a certain season, the constant η_{ik} is included. This is a fixed-effects model that factors out the team-season quality effects to focus on the team performance over time within a season. Here, r_{ijk} represents the explanatory variables for team performance: the position of the opponent in the league table at the end of the previous season and a dummy variable for if a match was a home match or an away match (home=1 and away=0). If the opponent has been promoted to the Eredivisie in the previous season, the rank for that club is set to 18. Whether a (un)wise coach change has happened is indicated by the dummy variables $Wise\ change_{ijk}$ and $Unwise\ change_{ijk}$ which are 1 for every match of a team-season after a coach change has happened for that type and 0 otherwise. The categorization of coach changes as wise or unwise is done in Section 3.3.2. The β and the δ are the coefficients of the corresponding explanatory variables, which represent the linear relationship between the independent variables and the three team performance indicators. For instance, if δ takes on a large positive value, this means that there is a positive correlation between a coach change and the team performance. It is important to note that this does not prove causality between coach replacement and team performance. For instance, assume that after a coach change the performance improved, but this improvement could also have taken place in the hypothetical situation of the same circumstances but with no coach change. Therefore, this naive approach does not prove causality, and to investigate this further, the difference-in-differences method is used for which a control group is needed, which approach is described in Section 3.3. The unobserved random error term is represented by ϵ_{ijk} and the coefficients are estimated using ordinary least squares (OLS).

3.3 Effects of a coach change on performance: Control group approach

As explained before, the naive approach can not give sufficient information about the direct effect of a coach change on team performance because the cases in which a coach is replaced is not compared to a control group where the coach does not get replaced. Namely, a club can perform poorly and as a consequence replace their coach. If team performance has improved after a coach change, this could be because of the change or as a consequence of different factors. A possibility could be that after a series of bad results, a team will start to perform better independently of

	Change	**Non-change**
Wise	Below expectations due to poor performance	Below expectations due to bad luck
Unwise	Below expectations due to bad luck	Below expectations due to poor performance

Table 3: Theoretical decomposition of wise and unwise coach changes and coach non-changes.

a coach change. Therefore, the difference-in-differences approach with the control group is used to research the direct effect between coach replacement and team performance. The idea of this approach is to categorise coach changes as wise and unwise, study their effects on performance and consequently answer sub-question 3. Table 3 shows the theoretical decomposition of wise and unwise coach changes and non-changes.

This approach, invented by Flepp and Franck (2021), consists of multiple steps. First, create a league table on the basis of the expected goals to compare to the official league table. Second, compare the expected goals league table to the official league table to categorise coach replacements as wise or unwise. Third, match the coach changes to a control group of counterfactual coach changes (non changes). Finally, perform a linear regression with the coach changes and the control group of counterfactual coach changes as explanatory variables and evaluate their differences. These steps of the approach of Flepp and Franck (2021) are described in detail in order of execution in Sections 3.3.1, 3.3.2 and 3.3.3.

3.3.1 Creating an expected goals league table

Football is a low-scoring game, with an average of 3.09 goals per game over the seasons 2017/2018 until 2022/2023. Therefore, the match results are sensitive to randomness because the match outcome does not perfectly represent the performance of the teams. A team can lose due to bad performance on the pitch or because they experienced bad luck. This study uses the expected goals as a team performance measure because they are less sensitive to random outcomes than the actual goals scored. Therefore, to locate instances in which performance on the pitch is misrepresented by match outcomes, the first step is to create a league table based on the expected goals (xG), which is called the expected goals league table (XGLT).

The expected goals, as calculated in Section 2, are used to determine the result of a match. If the difference in the expected goals of both teams involved in the match is larger than 0.5 the team with the most expected goals is considered the winner and earns 3 points. If the difference in expected goals is between 0.5 and -0.5 the match is considered a draw and both teams earn 1 point. If the difference in expected goals is smaller than -0.5 this is considered a loss and the corresponding team earns 0 points. Then, a league table is created based on the points earned with the expected goals: the expected goals league table.

This expected goal league table is an addition to the official league table (OLT) which is based on the actual match results. If a club is ranked better in the XGLT than in the OLT this means that according to their performance (measured in expected goals), the club should have had better results and be higher ranked than it actually is. Therefore, bad results of a club are the result of bad luck and not poor performance. If a club is ranked worse in the XGLT than in the OLT this means that bad results are due to poor performance on the pitch and not bad luck. The expected goals league table and the official league table get created for every round, resulting in a total of ($2 \cdot 34$=) 68 league tables per season. The league tables get created after every club has played the same number of matches, so there is no bias when matches get postponed.

3.3.2 Categorization of coach changes

The second step is the categorization of the coach changes into two categories based on the expected goals league table: wise changes and unwise changes. This categorization determines whether a coach replacement happened because of bad performance or because of bad luck. By categorising the coach replacements, the effects on performance of wise and unwise coach changes can be analysed separately.

If a club's results are below expectations, a coach can be replaced. This coach change is categorised as wise if the results are due to poor performance. In other words, a coach change is categorised as wise if a club's rank in the XGLT is equal to or worse than its rank in the OLT. If a club's rank is worse in the XGLT than in the OLT this means that with their performance, it is expected for them to be ranked worse than their actual ranking and therefore their ranking is a result of poor performance. This coach change is wise in the sense that the club's results are below expectations due to poor performance.

A coach change is categorised as unwise if the results are due to bad luck. Or, put differently, a coach change is categorised as unwise if a club's rank in the XGLT is better than its rank in the OLT. If a club's rank is better in the XGLT than in the OLT this means that with their performance, it is expected for them to be ranked better than their actual ranking and therefore their ranking is a result of bad luck. This coach replacement is unwise in the sense that the club's results are below expectations due to bad luck and not poor team performance. Figure 2 shows the XGLT and OLT rankings of a team at the time their coach is replaced, with unwise changes at the top left, where the OLT ranking is worse than the XGLT ranking, and wise changes at the bottom right, where the OLT ranking is equal to or better than the XGLT ranking. Of the total of 34 coach replacements, 11 are categorised as wise changes and 23 as unwise changes.

Note that this categorization method, introduced by Flepp and Franck (2021), does not include the absolute rank of a team but only the difference in the expected results and the actual results. For example, it is possible that a coach change at Ajax, a high performance club, is categorised as unwise when they are ranked 18th (last) in the OLT and 17th in the XGLT. It can be more intuitive to rank this as a wise coach change as the results of Ajax are way worse than normal. This is an extreme example, but a consequence of this is that it is possible that the categorization of a change as 'wise' or 'unwise' is counter-intuitive. However, only the name of the label of categorization ('wise' or 'unwise') can sometimes be counter-intuitive but the results are still as valuable. Moreover, keeping the notation the same as that of Flepp and Franck (2021) makes it easier to compare the results. More on this in the discussion part of the conclusion in Section 5.

Also, coach non-changes, which are all observations without a coach change, are categorised into two categories: wise non-changes and unwise non-changes. This is done to create control groups to which the coach changes are matched, which is done in the next step of the control

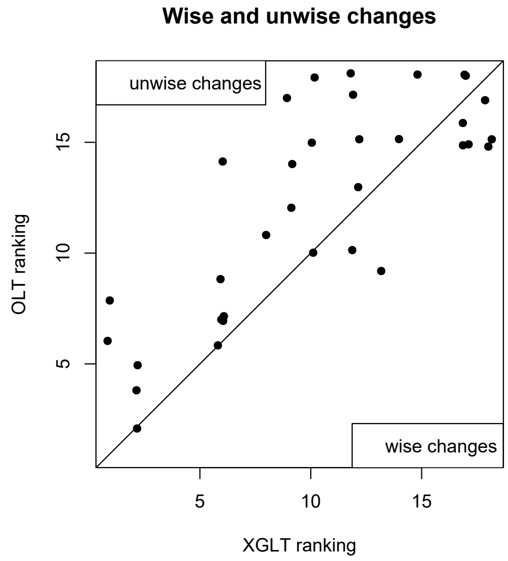


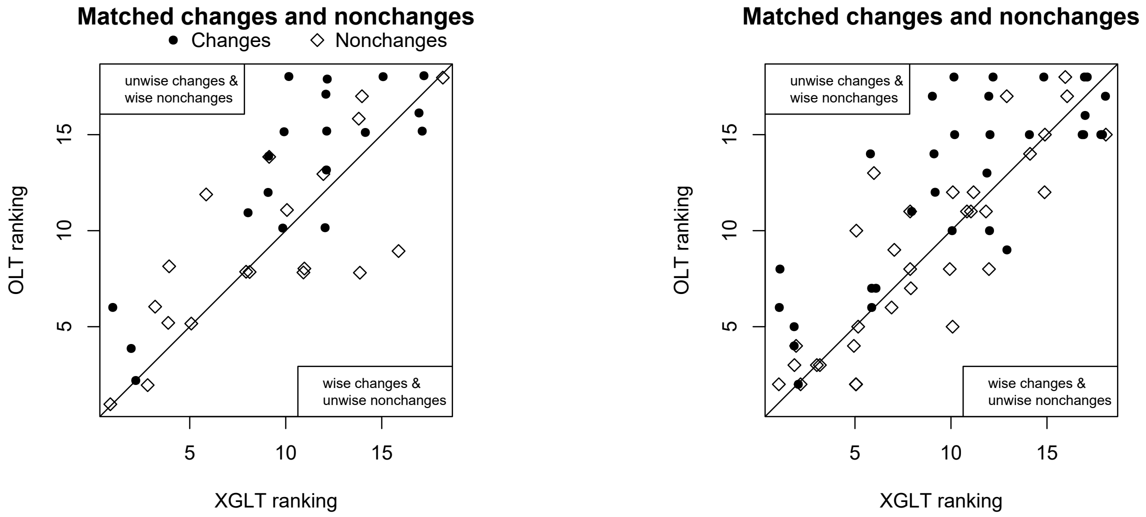
Figure 2: The expected goals league table (XGLT) and the official league table (OLT) rankings at the time of the coach changes. The line is the cutoff between wise coach changes and unwise coach changes. Coach changes that lay on the line are categorised as wise. Note that some combinations of OLT/XGLT rankings occur more than once.

group approach. A coach non-change is categorised as wise if a club's rank in the XGLT is equal or better than its rank in the OLT, following the same reasoning as before. A coach non-change is categorised as unwise if a club's rank in the XGLT is worse than its rank in the OLT.

3.3.3 Matching control groups

It is important to compare the cases in which a coach gets replaced to those where circumstances are similar but without a coach change. This way, the effect a coach change has on performance can be analysed conditionally on everything else being (mostly) the same. Thus, matching the wise coach changes to unwise non-changes and unwise changes to wise non-changes results in control groups which have a similar performance prior to the (non)change where the only difference is whether a coach replacement took place or not. As a consequence, these control groups make it possible to analyse the direct effect of (un)wise coach changes on team performance.

The wise changes get matched to unwise non-changes to create equal circumstances for the control group, as in both cases the team scored below expectations due to poor performance on the pitch. The unwise changes get matched to the wise non-changes, as in both cases the team scored below expectations due to bad luck. A coach change of a club gets matched to the non-change at the same club with the closest cumulative surprise in a season where no coach change happened. Note that when a coach non-change in a team-season gets matched, this team season gets removed from the pool of possible control group team-season. Applying this nearest neighbour approach ensures that the circumstances of a club's expectations and performance leading up to a coach replacement are similar to the circumstances of a control group where no coach was replaced. Matching to a similar situation at the same club accounts for the heterogeneity of coach changes among clubs. Namely, some clubs replace the head coach under circumstances where another club continues with the same coach. Also, this matching controls for the different seasonal aspirations of each club, such as avoiding relegation or qualifying for the Champions League. The matches using the restricted matching method, where a change gets matched to a non-change from the same club, are shown in Figure 3a.



(a) Restricted matching

(b) Unrestricted matching

Figure 3: The expected goals league table (XGLT) and the official league table (OLT) rankings at the time of the coach changes and control group matched nonchanges. The line is the cutoff between wise coach changes (and unwise nonchanges) and unwise coach changes (and wise nonchanges).

As an alternative matching approach, following that of Paola and Scoppa (2012) and Flepp and Franck (2021), only the cumulative surprise is used as a restriction and the coach changes can get matched to non-changes from all clubs. This unrestricted matching method, contrary to the restricted method, does not allow for club heterogeneity but can lead to better matches regarding cumulative surprise or more matches when the original method fails to match a change to a control group. The matches using the unrestricted matching method are shown in Figure 3b. For both matching methods, a maximum difference of 0.5 in cumulative surprise is used to ensure similarity in performance, which results in 19 non-changes (14 wise and 5 unwise) in the restricted control group and 32 (21 wise and 11 unwise) in the unrestricted control group. Only changes that get matched to a non-change are used in further analysis. Figure 4 shows the kernel densities of the cumulative surprise at the last match before a coach replacement. This is a probability density estimation and gives an overview of the similarities between the cumulative surprise before coach changes and matched non-changes. The kernel density of the cumulative surprise using the restricted matching approach, using only non-changes from the same club, seems to be fluctuating around that of the actual coach changes. This can be explained by the small number (19) of non-changes in the restricted control group. The kernel density of the cumulative surprise using the unrestricted matching approach, using non-changes from all clubs, is almost identical to that of the actual coach changes. This means that there is a close relationship between the cumulative surprise of the last match before a coach change and the matched non-change.

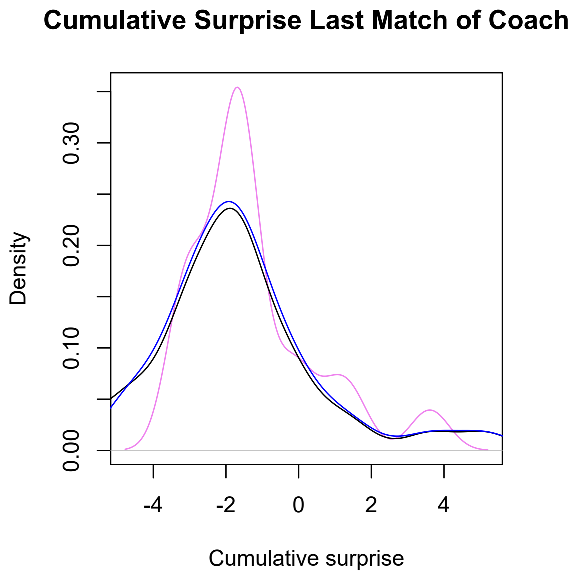


Figure 4: Kernel densities for the cumulative surprise, using the last 4 matches before a coach change (black), or a restricted (pink) or unrestricted (blue) matching approach.

3.3.4 Linear regression

The control group approach uses a modified version of the linear regression of the naive approach, given by Equation 3, to estimate the effect of un(wise) coach (non)changes on team performance. The regression used is given by

$$\begin{aligned} y_{ijk} = \eta_{ik} + r'_{ijk}\beta + \delta_1 Wise\ change_{ijk} + \delta_2 Unwise\ nonchange_{ijk} \\ + \delta_3 Unwise\ change_{ijk} + \delta_4 Wise\ nonchange_{ijk} + \epsilon_{ijk}. \end{aligned} \tag{4}$$

Here the explanatory variables $Unwise\ nonchange_{ijk}$ and $Wise\ nonchange_{ijk}$ are added to Equation 3. Here, $(un)wise\ change_{ijk}$ is 1 if a change of relevant type has happened in club i, before match j, in season k and 0 otherwise. $(Un)wise\ nonchange_{ijk}$ is 1 if a matched non-change of relevant type has happened in club i, before match j, in season k and 0 otherwise. The other variables in the regression are defined the same as in Equation 3: y_{ijk} are the performance indicators (win, goal difference and points), η_{ik} the team-season quality and r_{ijk} the explanatory variables $Home\ match_{ijk}$ and $Rank\ opponent_{ijk}$, where $Rank\ opponent_{ijk}$ the opponent's final rank is in the previous season. The unobserved random error term is represented by ϵ_{ijk}. The coefficients are estimated using ordinary least squares (OLS).

3.3.5 F-test

The F-test is used to test the significance of the effect of in-season coach changes on team performance. This is done by comparing the coefficients of the regression from the control group approach given in Equation 4. The 3 performance indicators are regressed on all explanatory variables, which give the estimated coefficients. For this research, the most interesting results are the differences between the coefficients δ_1 (wise change) and δ_2 (unwise non-change) and the differences between the coefficients of δ_3 (unwise change) and δ_4 (wise non-change), as these differences capture the causal effect of a coach replacement on team performance. The F-test is used to test the significance of the differences between these coefficients. The null hypotheses state that the coefficients are equal ($H_{0,1} : \delta_1 = \delta_2$ and $H_{0,2} : \delta_3 = \delta_4$) and the alternative hypotheses state that the coefficients of the variables are not equal.

4 Results

4.1 Determinants of a coach dismissal

First, this paper tries to answer sub-question 1: What are the determinants of an in-season head-coach change? In short, poor performance leads to in-season coach changes. The results show that points in the last 5/6 matches and the cumulative surprise are determinants of in-season coach changes. An increase in the cumulative points over the last 5/6 matches or the cumulative surprise over the last 4 matches leads to a decrease in the probability of a coach change taking place. A more detailed explanation of these results follows below.

Table 4 shows the results of the hazard model given in Equation 2, with on the left-hand side the three models with the three potential determinants. Also, the results of the sensitivity analysis are shown, where the hazard model is estimated using the covariates: points in the last 5 (A) and the last 6 matches (B). The p-values of the estimated coefficients are in parentheses. Under 'Season homogeneity' the estimated coefficients are shown when assuming hazard rate homogeneity across the seasons. In the column 'Season heterogeneity' the results are shown when allowing for hazard rate heterogeneity across different seasons.

The smaller the negative log likelihood value is, the better a model fits the data. Table 4 shows that the hazard model fits better when allowing for season heterogeneity, as all -Log(Likelihood) values are smaller than those of the models assuming season homogeneity. This implies that the probability of a replacement taking place differs across seasons. The standard model, model A and model B, have different numbers of observations because the first 4, 5 or 6 observations of each team-season are needed to calculate the cumulative points variable. Therefore, each team-season misses the first couple of observations as not enough matches are played to calculate the points over the last 4, 5 or 6 matches. This results in 2644 observations for the standard model, 2536 observations for model A and 2429 observations for model B. Also, model B, using points in the last 6 matches, seems to be the best model, as the BIC values are smaller than for the standard model and model A. This means that the model using points in the last 6 matches better fits the data than the models using points in the 4/5 matches.

In all models, the three potential determinants of a coach change have negative values. Therefore, an increase in the covariates leads to a decrease in the hazard rate, meaning that the probability of a coach change happening decreases. This makes sense, as receiving more points, outperforming the bookmaker odds and climbing the league table are all indicators of a club performing well. It is expected that the probability of a coach changing, either by dismissal or as a quit, decreases as a team performs well on the pitch.

For the standard model, only the effect of the cumulative surprise is significant with p-values

		In-season duration	
		Season homogeneity	Season heterogeneity
Standard model			
	Points last 4 matches	-0.10 (0.35)	-0.11 (0.28)
	Cumulative surprise	-0.26 (0.03)**	-0.23 (0.06)*
	Change in rank	-0.28 (0.12)	-0.28 (0.13)
	-Likelihood	224.49	164.88
	BIC	454.97	335.76
Sensitivity analysis			
A	Points last 5 matches	-0.14 (0.08)*	-0.15 (0.07)*
	Cumulative surprise	-0.23 (0.04)**	-0.21 (0.07)*
	Change in rank	-0.29 (0.14)	-0.28 (0.15)
	-Likelihood	192.15	139.75
	BIC	437.96	332.13
B	Points last 6 matches	-0.15 (0.04)**	-0.16 (0.03)**
	Cumulative surprise	-0.23 (0.04)**	-0.21 (0.06)*
	Change in rank	-0.30 (0.15)	-0.28 (0.17)
	-Likelihood	176.8	127.77
	BIC	434.19	318.09

Table 4: The results of estimating the proportional hazard model (Equation 2) with the potential determinants of in-season coach changes as covariates. The number of observations of the standard model, mode A and model B are 2644, 2536 and 2429 respectively. The p-values are in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% level.

of 0.03 and 0.06, thus rejecting the null hypothesis of the coefficients being equal to zero at the 5% and 10% levels. This is different when looking at the models of the sensitivity analysis, where the points in the last 5/6 matches becomes a significant determinant of a coach change with p values ranging from 0.03 to 0.08. The effect of the 'change in rank' coefficient is not significant in any of the models, with p-values between 0.12 and 0.13.

These results are similar to those of Van Ours and Van Tuijl (2016) who also found that the points in the last 5 and 6 matches as well as the cumulative surprise are significant determinants of coach dismissals. (d'Addona & Kind, 2014) also have a similar result, which shows that a decline in the league position leads to a higher chance of a coach being dismissed, although my results are not significant.

4.2 Naive approach

Here, the results are presented, which are used to answer sub-question 2: What is the difference in performance before and after an in-season coach change? In short, the results show that the team's performance is significantly better after an unwise coach change than before the change. The performance of clubs after a wise change is slightly worse, on average, than before the change. A more detailed explanation of these results follows below.

Table 5 shows the results of the linear regression given by Equation 3. On top are the dependent variables, which are the three performance indicators: points, win and goal difference. On the left are the explanatory variables: rank opponent, home match, wise coach change and unwise coach change. The estimated coefficients of the explanatory variables are shown with the corresponding p-values in parentheses.

For the ‘unwise change’ variable all the coefficients are positive, as seen in Table 5. This means that the performance of a team, measured by the performance indicators, is generally better after an unwise coach change than before the change. On average, a club receives 0.50 points more per game after an unwise coach change and wins 20% more of their games compared to the matches of that club before the coach change. Also, the goal difference of a club is 0.64 higher after an unwise coach change than before. For all three performance indicators (points, win and goal difference), the difference in performance before and after an unwise coach replacement is significant at the 1% level. Note that this does not imply that the improvement in team performance is a causal effect of an unwise coach change. The direct effects of a coach change on performance are discussed in Section 4.3.

In contrast to an unwise coach change, all the coefficients are negative for the ‘wise change’ variable, as seen in Table 5. This means that the performance of a team, measured by the performance indicators, is generally worse after a wise coach change than before the change. Namely, a team receives 0.13 points less on average per game after a wise coach change took place. Also, after a wise coach change, a club loses 7% more of their games and their goal difference is 0.34 lower on average compared to the matches of that club before the wise coach change. The coefficients of the ‘wise change’ are not significantly different from zero as shown by the p-values in Table 5.

It seems counterintuitive that after an unwise coach change a team’s performance improves whereas after a wise coach change the team performance is worse. However, note that coach changes are categorised as wise or unwise on the basis of the difference between realised results and the expected outcomes, independent of the observed effect the coach changes have on the performance, following the method of Flepp and Franck (2021).

	Points	**Win**	**Goal difference**
Rank opponent	0.07 (0.00)***	0.02 (0.00)***	0.12(0.00)***
Home match	0.42 (0.00)***	0.14 (0.00)***	0.78 (0.00)***
Wise change	-0.14 (0.49)	-0.07 (0.38)	-0.34 (0.26)
Unwise change	0.50 (0.00)***	0.20 (0.00)***	0.64 (0.00)***

Table 5: The estimated coefficients of the naive approach regression. The p-values are in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% level.

Table 5 shows that the rank of the opponent in the previous season has a significant impact on the performance indicators. For every increase in the opponent’s rank (e.g.: rank 6 to rank 7), a team receives 0.07 more points, on average. For example, a club generally receives ($17 \cdot 0.07 =$) 1.19 more points playing against last place (rank 18) compared to playing against first place (rank 1). Also, with each increase in opponent’s previous season rank, a team wins 2% more of the games and has an average increase in goal difference of 0.12. Whether a match is a home match or not also has a significant effect on all three team performance indicators. A club that plays a match at home has a 14% higher probability of winning than an away match. Also, the points received and the goal difference are significantly higher for the home team.

4.3 Control group approach

Here, the results are presented, which are used to answer sub-question 3: What is the effect of wise and unwise coach changes on team performance? In short, coach changes categorized as unwise improve performance significantly when assuming club homogeneity. The results suggest that coach changes categorized as wise do not improve performance but do limit the worsening of team performance that would happen if no coach was replaced, but this is not significant. Below, the results are described in more detail.

Table 6 shows the results for the restricted (same club) and unrestricted (all clubs) control group matching of the linear regression given by Equation 4. On top, it shows the dependent variables, which are the team performance indicators: points, win and goal difference. The table shows the coefficients of the explanatory variables, with the corresponding p-values in parentheses. 'Rank opponent' is the final rank of the opponent in the previous season. The bottom two rows contain the p-values of the two F-tests.

Note that, by construction of the regressions, the coefficients and corresponding p-values of the variables Rank opponent and Home match for the control group regression are the same in the naive approach results shown in Table 5. Moreover, the coefficients of the variables Rank opponent and Home match are the same for the restricted and unrestricted control group regressions because those variables do not depend on the type of control group matching. Also, it seems that there is no correlation between those variables and the dummy (non)change variables, as the coefficients of Rank opponent and Home match are the same for the restricted and the unrestricted control group regressions.

Table 6 shows that after an unwise coach change the team performance is better than before the change as its coefficients are all positive, using the restricted as well as the unrestricted control group. This is counter-intuitive as it is an unwise coach change, meaning bad results are due to bad luck. However, despite the coach change being classified as 'unwise' it can still have a positive effect on the team performance. The coefficients of the unwise change are significantly different from zero, but this does not mean that an unwise coach change was the cause of better subsequent performance. The performance improvement could be a consequence of different factors explained in Section 3.3. For that reason, the difference with a control group is invest-

	Restricted control group			**Unrestricted control group**		
	Points	**Win**	**Goal difference**	**Points**	**Win**	**Goal difference**
Rank opponent	0.07 (0.00)***	0.02 (0.00)***	0.12 (0.00)***	0.07 (0.00)***	0.02 (0.00)***	0.12 (0.00)***
Home match	0.42 (0.00)***	0.14 (0.00)***	0.78 (0.00)***	0.42 (0.00)***	0.14 (0.00)***	0.78 (0.00)***
Wise change	-0.14 (0.49)	-0.07 (0.38)	-0.34 (0.26)	-0.05 (0.73)	-0.05 (0.31)	0.08 (0.71)
Unwise nonchange	-0.37 (0.09)*	-0.11 (0.16)	-0.65 (0.05)**	-0.11 (0.47)	-0.03 (0.58)	-0.11 (0.63)
Unwise change	0.50 (0.00)***	0.20 (0.00)***	0.64 (0.00)***	0.38 (0.00)***	0.14 (0.00)***	0.44 (0.00)***
Wise nonchange	0.20 (0.13)	0.07 (0.12)	0.14 (0.45)	0.21 (0.05)**	0.08 (0.04)**	0.06 (0.72)
F-test $\delta_1 = \delta_2$	0.43	0.67	0.50	0.75	0.78	0.54
F-test $\delta_3 = \delta_4$	0.09*	0.07*	0.08*	0.25	0.26	0.09*

Table 6: The estimated coefficients of the control group regression using a restricted (same club) and unrestricted (all clubs) control group matching corrected for team-season fixed effects. Rank opponent is the final ranking of the opponent in the previous season. The p-values of the F-tests are shown. The p-values of the estimated coefficients are in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% level.

igated by comparing the Unwise change and Wise non-change coefficients. Table 6 shows that the wise non-change coefficients are also positive in all cases, meaning team performance after a matched wise non-change is better than performance before the non-change. The coefficients of the wise non-change are not significantly different from zero using the restricted control group method, with p-values larger than 0.10. Using the unrestricted control group approach, the effect of a wise non-change is significant at the 5% level for the performance indicators Points and Win, with p-values of 0.05 and 0.04 respectively. The effect of an unwise change on performance seems to be larger than the effect of a wise nonchange, as all its coefficients are larger for both the restricted and the unrestricted control group method. To check if this difference is statistically significant the results of the F-test are used. The p-values of the F-test in the bottom row show that the difference in the coefficients of Unwise change (δ_3) and Wise non-change (δ_4) are significant at the 10% level for any of the team performance measures, using a restricted control group, with p-values of 0.09, 0.07 and 0.08. This implies that an unwise coach change improves the performance of a club when accounting for the heterogeneity of clubs. As a reminder, a coach change is unwise if according to the expected goals a team should be ranked better than they actually are. Bad results are therefore seen as a consequence of bad luck and not due to poor performance. A possible explanation of the positive effect of unwise coach changes could be that the results are in fact not due to bad luck but are a consequence of factors not (accurately) captured by the expected goals. It can also come from the fact that the expected goals statistic can vary across seasons, especially when a club drafts different players with different play styles. Given that the expected goals are calculated using the average goals per shot on target from the previous season of a club, this can lead to an incorrect number of expected goals used to categorise a coach change. This could be a reason why an 'unwise' coach change does have a positive effect on performance.

There is a big difference in the p-values of the F-test comparing δ_3 and δ_4 using the restricted control group or the unrestricted control group. Table 6 shows that, in contrast to the results using the restricted control group, there is no significant difference in the coefficients δ_3 and δ_4 for the Points and Win variables using the unrestricted control group method, according to the F-test. An unwise coach change does still have a significant positive effect on the goal difference at the 10% level, with a p-value of 0.09.

Table 6 shows that after a wise change the team performance is worse than before the change as its coefficients are all negative. This is as unexpected as it is a wise coach change, meaning bad results is due to bad performance on pitch. However, the coefficients are not significantly different from zero, following from the p-values, and take on values close to zero, especially using the unrestricted control group, with coefficients of -0.05, -0.05 and 0.08. Moreover, the results after an unwise non-change are even worse, as the coefficients are more negative compared to the results after a wise change. This suggests that a wise coach change does not improve performance but does limit the worsening of team performance that would happen if no coach was replaced. The coefficients of the unwise non-change variable for the performance indicators Points and Win are significantly different from zero using a restricted control group, at the 10% and 5% levels with corresponding p-values of 0.09 and 0.05, respectively. This means that after a matched unwise non-change, the points received per game are 0.37 points lower, and the goal difference of a club decreases by 0.64 on average. The F-test in Table 6 reveals that there is no significant difference in the wise change and unwise nonchange coefficients. This implies that, if a team has bad results due to poor performance, a wise coach change does not improve the performance, and by not replacing the coach, the same decrease in performance is observed.

It has to be noted that using the restricted matching method, only 5 unwise non-changes are in the control group (and 5 wise changes in the coach change group). This low number of observations comes with the problem of limited statistical power and increased sampling error, which can lead to biased estimates. Therefore, the results have to be considered carefully when drawing conclusions. Using the restricted matching, 11 unwise non-changes are in the control group, making this a more reliable result in terms of observations.

Interestingly, the results of the control group approach in this paper are different from those of Flepp and Franck (2021). They found that, instead of only after an unwise change, both the performance after a wise and an unwise coach change is better compared to preceding the performance. The results are similar in the sense that the performance after an unwise coach change is better than the performance after a wise coach change. However, Flepp and Franck (2021) do not find that an unwise coach change has a significant (positive) effect but instead find that a wise coach change does have a significant (positive) effect on subsequent performance. There are multiple possibilities leading to a difference in findings. For example, Flepp and Franck (2021) use only coach dismissals, whereas this paper uses all coach changes (quits and dismissals), which could have an impact on the results.

5 Conclusion

This paper examines the determinants of a coach change and the effects of wise and unwise changes on team performance. Data from six consecutive seasons, from 2017/2018 until 2022/2023, of the Eredivisie, the Dutch highest football league, is used to answer the main research question: What are the causes and consequences of in-season head coach changes on the team performance? This question consists of multiple components, which are investigated based on three sub-questions.

First, what are the determinants of an in-season head-coach change? The results show that poor performance, measured in the cumulative surprise, which is the difference in actual results and the expected results based on bookmaker odds, is a significant determinant of an in-season coach change. If a club performs worse than expected, the probability of a coach change increases. Also, the number of points accumulated over the last six matches is a significant factor in determining if a coach gets replaced or not. When a club receives a low number of points over the last six matches, there is a higher chance of a coach change.

Second, what is the difference in performance before and after an in-season coach change? The data suggests that, when accounted for home match advantage and the qualities of the opponents, a club performs significantly better after an unwise coach change, scoring more points, winning more games and attaining a bigger (positive) goal difference. Note that a change is categorised as wise when a club is ranked equally or worse based on the expected goals than in the official league table, and unwise if a club is expected to be ranked better based on expected goals. However, this result does not prove causality between a unwise coach change and improved performance, which leads us to the third sub-question.

Third, what is the effect of wise and unwise coach changes on team performance? The findings indicate that, an unwise coach change does improve the performance of a team. This implies that when the goal is to improve the results, changing the coach is useful when a club is expected to be ranked better based on expected goals. Also, the results suggest that wise coach changes do not improve performance but do limit the worsening of team performance that would happen if no coach was replaced, but this is not significant. This implies that changing the coach when a club is expected to be ranked worse, based on expected goals, does not change the performance. Also, the results show that, when trying to research the direct effect a coach change has on performance, it is not sufficient to compare the performance before and after a coach change.

The results of this research contribute to our understanding of the causes and consequences of coach changes in professional football, but there are some limitations to this paper, leading to suggestions for further research. For instance, in this research, the expected goals are calculated with the shots on target and the average goals per shot on target from last season. To calculate the expected goals more precisely, shots statistics can be used, such as distance from target and angle from goal, which is done in Flepp and Franck (2021). This calculation could even be improved by using individual player skills and taking defensive pressure into account. Also, coach changes can be distinguished into dismissals and quits to investigate their differences. This study does use the most recent data, but only from one competition, limiting the number of observations. Therefore, in further research, it can be beneficial to research more football leagues or investigate the effects of coach changes in other sports, as done in McTeer, White and Persad (1995).

In forthcoming studies, besides using the difference in expected rank and actual rank of a club, the absolute rank can be taken into account when classifying coach changes as wise or unwise. The importance of this can be indicated with the use of an example. For instance, a coach is replaced at a club that normally ends up around the bottom of the league table but surprisingly stands in third place. In this paper, this coach change is categorised as wise if their expected rank, based on the expected goals, is worse (e.g. fourth or fifth place) than their actual rank. However, it makes more sense to categorise this coach change as unwise, because the club's results with the current coach are way better than usual. Therefore, it can be interesting to include the absolute rank of a club and their average final ranking when categorising coach changes as wise or unwise.

The effects of coach changes on performance in professional football can be linked to manager dismissals in non sport related companies, where the coach's position resembles the manager's position in a business. Using football data to investigate the effect of managerial dismissals is useful, as lots of data is available to analyse the effects. In future research, the link between coach changes in football and dismissals of company managers can be explored more, or direct research can be done on the effect of non-sports-related manager dismissals.

References

Bachan, R., Reilly, B. & Witt, R. (2008). The hazard of being an english football league manager: empirical estimates for three recent league seasons. *Journal of the Operational Research Society*, *59*(7), 884–891.

Barros, C. P., Frick, B. & Passos, J. (2009). Coaching for survival: The hazards of head coach careers in the german 'bundesliga'. *Applied Economics*, *41*(25), 3303–3311.

Brechot, M. & Flepp, R. (2020). Dealing with randomness in match outcomes: how to rethink performance evaluation in european club football using expected goals. *Journal of Sports Economics*, *21*(4), 335–362.

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

d'Addona, S. & Kind, A. (2014). Forced manager turnovers in english soccer leagues: a long-term perspective. *Journal of Sports Economics*, *15*(2), 150–179.

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

Heuer, A., Müller, C., Rubner, O., Hagemann, N. & Strauss, B. (2011). Usefulness of dismissing and changing the coach in professional soccer. *PloS one*, *6*(3), e17664.

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

McTeer, W., White, P. G. & Persad, S. (1995). Manager coach mid-season replacement and team performance in professional team sport. *Journal of Sport Behavior*, *18*(1), 58–69.

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

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.

Pieper, J., Nüesch, S. & Franck, E. (2014). How performance expectations affect managerial replacement decisions. *Schmalenbach Business Review*, *66*, 5–23.

Rocaboy, Y. & Pavlik, M. (2020). Performance expectations of professional sport teams and in-season head coach dismissals—evidence from the english and french men's football first divisions. *Economies*, *8*(4), 82.

Semmelroth, D. (2022). Time to say goodbye: A duration analysis of the determinants of coach dismissals and quits in major league soccer. *Journal of Sports Economics*, *23*(1), 95–120.

Ter Weel, B. (2011). Does manager turnover improve firm performance? evidence from dutch soccer, 1986–2004. *De Economist*, *159*, 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.

www.espn.nl. (2023). `https://www.espn.nl/voetbal/stand/_/competitie/NED.1/season/2021`. (Accessed: 2023-05-30)

www.fbref.com. (2023). `https://fbref.com/en/comps/23/Eredivisie-Stats`. (Accessed: 2023-05-20)

www.football-data.co.uk. (2023). `https://www.football-data.co.uk/netherlandsm.php`. (Accessed: 2023-05-30)

www.transfermarkt.com. (2023). `https://www.transfermarkt.nl/eredivisie/`

`trainerwechsel/wettbewerb/NL1`. (Accessed: 2023-05-20)

A Appendix

Table 7 shows that a team-season without a coach change generally has better performance statistics than a team-season with a coach change. On average, clubs perform better in seasons without a coach change compared to seasons with a coach change (see Table 2. Namely, they win more matches (42 percent vs. 32 percent), earn more points per match (1.47 vs. 1.20), have a better goal difference (0.14 vs. -0.29), and score more goals (1.47 vs. 1.31). Also, all expected statistics and the cumulative surprise of a team-season without a coach change are better than those of a team-season with a coach change. The correlations of the statistics in seasons with and without a coach change are very similar.

		Mean	**SD**	**Min.**	**Max.**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	Win	0.42	0.49	0	1	1							
2	Points	1.47	1.34	0	3	0.96	1						
3	Goal difference	0.14	2.14	-8	13	0.78	0.84	1					
4	Goals scored	1.61	1.42	0	13	0.66	0.67	0.78	1				
5	xPoints	1.41	1.28	0	3	0.55	0.58	0.61	0.48	1			
6	xGoal difference	0.07	1.46	-7.00	7.36	0.55	0.59	0.71	0.55	0.85	1		
7	xGoals scored	1.47	0.91	0	7.36	0.49	0.50	0.58	0.66	0.70	0.81	1	
8	Cumulative surprise	0.21	2.23	-7.16	7.58	0.42	0.44	0.32	0.25	0.14	0.14	0.10	1

Table 7: The descriptive statistics and correlations of various variables for the team-seasons without a coach change.

B Programming code

The data analysis in this paper is done in RStudio with multiple files that have to be run in order of the numbers at the beginning of the file names. First, *1.Sorting_data_by_club* is run to sort the data per team-season, resulting in double the observations as every match is viewed from the perspective of the home team and the away team. This code creates the Excel file Data_sorted, to which the club name column is added by hand. This data file is then used in *2.Data_.sorted_transformed* which calculates multiple statistical variables needed for the analysis and adds these to the data frame Data_sorted to create the Excel file Data_.sorted_transformed. Then the Data_.sorted_transformed data is used in *3.xGcalculations* which calculates the expected goals and corresponding variables based on the expected goals and creates the Excel file Data_complete, which contains all data transformations up to that point. The Data_complete data is then used in *4.LeagueTables* to create the official league table (OLT) and the expected goals league table (XGLT). After running *4.LeagueTables* the *5.ControlGroup_matching* file can be run, which matches the coach changes to coach non-changes and creates the dummy variables for (un)wise (non)changes and saves the data with the dummies as Excel file Data_PLM. The Data_PLM data set can then be used in *6.Regressions* which performs all the regressions, calculates the statistical characteristics of the data and estimates the kernel densities. The Data_PLM data set is also used in *7.HazardModel* which estimates the hazard models. The details of how a code works and what it does can be found in the code itself, with the use of comments. The Excel files needed to run a certain code are stated at the beginning of that code.