Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis BSc^2 Econometrics and Economics

The Effect of Performance Related Football Coach Changes in the Dutch Eredivisie and the Use of Quasi-Exact Matching

Jan D.N. Tempel (497053)

Supervisor:	Jan C. van Ours
Second assessor:	Dennis Fok
Date final version:	June 28, 2023

Abstract

To analyse the difference between performance related (PR) in-season coach changes and non-performance related (NPR) in-season coach changes in the Dutch Eredivisie between season 2000/2001 and 2021/2022, I first explore a naive linear regression model approach. Results show that the positive coach change effect found for PR coach changes does not exist for NPR coach changes. Next, I use quasi-exact matching (QEM) on the PR coach changes, while using the Average Treatment Effect on Treated (ATT). This leads to a straightforward comparison between the treatment group and the different control groups, which are made using different matching criteria based on three coach change determinants: cumulative surprise, cumulative points and the number of wins in the last four matches before a coach change. By using seven different QEMs and thus seven different control groups, I obtain insight in the usefulness of the matching criteria, and the robustness of the found effects. It seems that the ATT is insignificantly different from zero, and the found positive naive effect of PR coach changes can be attributed to the regression-to-the-mean phenomenon.

1 Introduction

By examining four different areas of sport, including salary, performance, supervision and changes in rules, Kahn (2000) makes the argument that the sport labor market provides an outstanding environment to analyse the economic labor market. Due to the wide availability of detailed data in the sport labor market, better experiments and studies can be done. Instead of quarterly financial data in business, weekly data is usually provided for football. Furthermore, changes in the sport labor market or sport rules take less time to display its effects. Szymanski (2003) agrees with this argument and examines this labor market within sports with regards to restrictions, market efficiency, competition, and pay performance. Similarities between the effect of wages on performance and supervision in the sports labor market and real labor market are found. Top managers of companies in the labor market can be compared to football coaches in the sports labor market. Pieper et al. (2014) explore this analogy by looking at the German soccer league, the Bundesliga, and find that both the managers and professional football coaches are often male and in their late 40s to early 50s. They know how to deal with both the pressure to perform well and the media, they report to a higher body that determines their employment and they usually have multiple stakeholders. Thus, analysing professional football coaches is economically relevant due to their resemblance of top managers in business.

However, there is a debate in the current literature on the importance of CEO's in business; sometimes they have a considerable influence on the firm's performance (Westerberg et al. (1997)), while others find that for some firms the impact is negligible (Arnulf et al. (2012)). This debate also occurs in the literature on the importance of football coaches. Kuper & Szymanski (2018) argue that the coaches are overrated, and that the obsession with coaches is comparable with the "Great Man Theory of History", the idea that only a few men cause historical change. In their opinion, the coach has a more symbolic role than managerial. Anderson & Sally (2013) use as a counterargument the recent revival of the "Great Man Theory of History", and provide numerical studies done that provide evidence of the importance of CEO's.

An easier way to study the importance of coaches, and thus CEO's, is to look at the effects of a coach change.

In this paper, I will extent the research done by van Ours & van Tuijl (2016) on the effect of in-season head coach dismissals in the highest professional Dutch football league the Eredivisie. The paper will contribute to the existing literature for multiple reasons. First of all, this paper will perform a sample extension. van Ours & van Tuijl (2016) use 14 seasons in a sample period from 2000 to 2014. This can be extended to the Eredivisie season 2021-2022, which is the last finished season, giving insights on whether the found coach change effects in van Ours & van Tuijl (2016) changed in the past decade. Furthermore, increasing the sample size will also increase the number of coach changes, and therefore give more accurate estimator coefficients.

Sub-question 1: Have the effects of an in-season coach change found in van Ours & van Tuijl (2016) in the period 2000 to 2014 changed in the last decade (2014-2022)?

Secondly, I will make a distinction between coach changes that are performance related (PR), and non-performance related (NPR) coach changes like behaviour or health related issues. This is economically significant, because the cause of the change could potentially affect the effect of the coach change. For example, recently in December 2022, Henk Fraser, ex-coach of FC Utrecht, quit after he showed inappropriate behaviour to one of his players. The unexpected dismissal of this coach, who did not perform badly, could lead to upheaval among players and fans, leading to unexpected good or bad performances. In the last 22 years many coach changes happened, and for every coach change a reason why has to be found. Luckily, media coverage of in-season coach changes is heavy due to the extensive fan interest.

Sub-question 2: What is the effect, and difference in effect, of in-season performance related and non-performance related coach changes?

Thirdly and finally, to analyse the effect of using different control groups for PR coach changes, I will explore quasi-exact matching (QEM) while using the Average Treatment Effect on Treated (ATT), based on the idea of Arrondel et al. (2020). This leads to a straightforward comparison between the treatment group and the different control groups, which are made using different matching criteria based on multiple dismissal determinants. Other literature uses parametric propensity-score matching (PSM) to match control groups to treatment groups, with a logit or probit ((Audas et al. (2002), Paola & Scoppa (2012)) model, or parametric propensity-score weighting (PWS)(Narita et al. (2022)). However, most of those models provide unreliable results due to model misspecification or selection bias. That is why I prefer QEM on the observable characteristics of the teams. The dismissal determinants found in other literature will be explored, and used as criteria to match control groups on the treatment group (Paola & Scoppa (2012)). The criteria used in this paper are the three performance indicators: cumulative surprise, cumulative points, and the number of wins in the last four matches before a coach change (WL4M). On

those different criteria, I use QEM, meaning that we make some of the criteria for matching less strict, while keeping other matching criteria strict.

Sub-question 3: Do different types of control group matching criteria and strictness result in significant different results?

Combining the three different sub-questions on sample extension, type of coach change, and method of control group matching, leads to the research question defining this paper.

Research Question: What is the effect, and difference in effect, of in-season performance related and non-performance related coach changes, using quasi-exact matching based on the dismissal determinants cumulative surprise, cumulative points, and the number of wins in the last four matches, in the Dutch Eredivisie between 2000 and 2022?

This paper is divided into six sections. The next section, Section 2, gives a literature review on the different topics discussed in this paper. Section 3 discusses and describes the data used in this paper, while Section 4 discusses the methods that are utilized. Section 5 gives the results of the analyses and Section 6 will summarize the findings of this paper and discuss the limitations and possible extensions.

2 Literature Review

2.1 Coach Dismissal

An easier way to study the importance of coaches, and thus CEO's, is to look at the effects of a coach change. Football coaches and top managers have strong incentives to perform well, of which one is to avoid being dismissed. van Ours & van Tuijl (2016) give multiple possible reasons for a dismissal of a football coach after a period of poor performance; the hot-hand fallacy, media pressure, stakeholder satisfaction, conflicts between the coach and the board, and the idea that doing something is better than doing nothing at all. Different studies have been done trying to determine the causes of a coach dismissal, given in Table 1, by using models such as duration analysis, logistic regressions, probit and logit models, and linear probability models, using data of the English Premier league, the German Bundesliga, the Spanish La Liga, and the Dutch Eredivisie. Audas et al. (1999), Salomo & Teichmann (2000), Dobson et al. (2001), Tena & Forrest (2007), Frick et al. (2010), d'Addona & Kind (2014), Pieper et al. (2014) and van Ours & van Tuijl (2016) all conclude that past performance has the most impact on the probability of a dismissal, which seems logical. Audas et al. (1999), Bachan et al. (2008) and Pieper et al. (2014) also conclude that the current position of the club in the league has a negative impact. The better a club is performing, the lower the probability of a dismissal. Interestingly, Tena & Forrest (2007) find that managerial efficiency, which indicates the relative performance of a club, and whether the club is in the relegation zone matter. Change in league position (Dobson et al. (2001), d'Addona & Kind (2014)) and

relative points (Barros et al. (2009), Frick et al. (2010)) may also play a role. Pieper et al. (2014) and van Ours & van Tuijl (2016) are the first ones to use bookmaker odds in order to take expectations and surprise into account, and find that higher expectations increase the probability of a dismissal.

Table 1: Literature review on the Impact of Variables on Coach Dismissals; Mainly based on van Ours & van Tuijl (2016).

	Country	**Seasons**	***N***	**Performance Measures**	**Coach Characteristics**
Audas et al. (1999)	England	1972-1997	699	Results recent matches (–), league position (–)	Age; non-linear maximum at age 47
Salomo & Teichmann (2000)	Germany	1979-1998	194	Points last 4 matches (–)	None
Dobson et al. (2001)	England	1972-200	614	Results recent matches (–), change in league position (+)	Age, negative above 37
Tena & Forrest (2007)	Spain	2002-2005	20	Lost last match (+), club in relegation zone (+), managerial efficiency (–)	None
Bachan et al. (2008)	England	2001-2004	96	League Position (–)	None
Barros et al. (2009)	Germany	1981–2003	115	Relative points (–)	Career Win Percentage (–)
Frick et al. (2010)	Germany	1981–2003	115	Lost 3 last matches (+), relative points (–)	Salary (+), career win percentage (–)
d'Addona & Kind (2014)	England	1949–2008	1,213	Recent match results (–), change in league position (+)	Age (+), former player national team (+)
Pieper et al. (2014)	Germany	1998–2008	67	League position (–), points last 5 matches (–), points expected last 5 matches (+)	None
van Ours & van Tuijl (2016)	The Netherlands	2000-2014	42	Points last 4 matches (–), Cumulative Surprise (–)	Coach experience (–)

Note: N: Number of Coach Dismissals studied.

Furthermore, personal characteristics of a coach may influence the dismissal probability too. Audas et al. (1999), Dobson et al. (2001) and d'Addona & Kind (2014) find that age matters; the older the coach is, the more likely he gets fired. Barros et al. (2009) and Frick et al. (2010) find that the career win percentage of a coach has a negative impact. van Ours & van Tuijl (2016) conclude that the experience of a coach has positive impact on the probability of a dismissal, with as explanation an increased disappointment.

2.2 Effect of Coach Change

The question remains on whether a coach dismissal is effective. Some studies performed already investigated the effect of a coach dismissal, and the results differ per study (Table 2). However, an important distinction in the existing literature is whether the study accounted for selectivity of the chosen coach dismissals (van Ours & van Tuijl (2016)). When only using a "naive" approach, there is no control group to see whether the performance after a coach change that did not happen went up. It is possible that the performance would have gone up (or down) even if there was no coach change. Salomo & Teichmann (2000) and Dobson et al. (2001) performed as one of the first this naive regression, and concluded that the coach change effect is (slightly) negative. Poulsen (2000) also performed used the naive approach but did not find any effect. To get results that actually have significant meaning, this selectivity must be accounted for.

Koning (2003), Bruinshoofd & Ter Weel (2003), Balduck, Buelens, & Philippaerts (2010), Balduck, Prinzie, & Buelens (2010), ter Weel (2011), Paola & Scoppa (2012), Hentchel et al. (2012), Scelles & Llorca (2020), Scelles & Llorca (2021), Arrondel et al. (2020) and van Ours & van Tuijl (2016) perform studies

Table 2: Literature review on Effect of Coach Dismissals; Mainly based on van Ours & van Tuijl (2016).

	Country	Seasons	*N*	S	Team Performance	Effect
Poulsen (2000)	England	1993–1998	59	No	Points	No
Salomo & Teichmann (2000)	Germany	1979–1998	194	No	Points	Negative
Dobson et al. (2001)	England	1972–2000	614	No	Points	Initially significantly negative
Koning (2003)	The Netherlands	1993–1998	28	Yes	Goal Difference	No
Bruinshoofd & Ter Weel (2003)	The Netherlands	1988–2000	27	Yes	Four-game point average divided byseasonal average points per game	No
Tena & Forrest (2007)	Spain	2002–2005	20	No	Points	Improvement in homematches
Balduck, Buelens, & Philippaerts (2010)	Belgium	1998–2003	72	Yes	Points	No
Balduck, Prinzie, & Buelens (2010)	Belgium	1998–2005	45	Yes	Goal Difference	No
ter Weel (2011)	The Netherlands	1986–2004	81	Yes	Four-game point average divided by seasonal average points per game	No
Paola & Scoppa (2012)	Italy	1997–2009	84	Yes	Points, goals scored, goals conceded	No
Hentchel et al. (2012)	Germany	1993–2010	146	Yes	Points	No effect heterogeneous teams, positive effect homogeneous teams
van Ours & van Tuijl (2016)	The Netherlands	2000-2014	42	Yes	Points, Victories, Goal Difference	No
Arrondel et al. (2020)	France	1998-2018	103	Yes	Points	Initially significantly positive
Scelles & Llorca (2020)	France	2000-2016	79	Yes	Points, Victories, Goal Difference	Positive when controlling for expected performance
Scelles & Llorca (2021)	France	1994-2016	103	Yes	Goal Difference	No

Note: S: Selectivity accounted for, N: Number of Coach Dismissals studied.

investigating the effect of coach dismissals, while accounting for the selectivity of the coach dismissals. Those studies are done for football leagues in Germany, The Netherlands, Belgium, France, and Italy, and while the team performance measure may differ between points, goal difference, goals scored and goals conceded, almost no effects are found that only apply for the coach changes. An exception is the study done by Hentchel et al. (2012), which concludes that there is a positive effect, but only for homogeneous teams in which the players individually perform on the same level, which might trigger competition after a coach change. Bruinshoofd & Ter Weel (2003), Balduck, Buelens, & Philippaerts (2010), and van Ours & van Tuijl (2016), using a control group with counterfactual coach changes, find that the performance of a club declines pre-dismissal, and increases again post-dismissal. However, the counterfactual coach dismissals follow the same pattern, indicating a regression-to-the-mean-phenomenon. In contrast, Scelles & Llorca (2020) also uses a control group, but finds positive impact only for the coach change when controlling for expected performance. Scelles & Llorca (2021) build further on their previous research and look at whether different types of board presidents and owners have an impact on the effect. They conclude that the presidents longevity and the geographic orientation of the owners matter on the dismissal probability and effect.

Studies by Tena & Forrest (2007) (not accounting for selectivity, Spanish La Liga) and Arrondel et al. (2020) (accounts for selectivity, French Ligue 1) show that a coach change has a small, positive and statistically significant effect in the short run, but only during home matches. Instead of being the result of quality difference between the old and new football coach, this could be the result of fan pressure.

Peeters & van Ours (2021) investigate home advantage in English professional football, and conclude that home advantage exists through crowd pressure and stadium attendance, psychological benefits from playing at home, less fatigue due to less traveling, and even artificial pitches. This may be economically relevant, due to the fairness of competition.

3 Data

van Ours & van Tuijl (2016) mention that they got their data from various internet sources. For the period 2000/2001 - 2013/2014, I use the data set provided by van Ours & van Tuijl (2016). For the Eredivisie period 2014/2015 - 2021/2022, I use the data set by Gebala (2022), which gives data on multiple leagues from 2002 to 2022. For the Dutch Eredivisie, data is given from 2004 to 2022.

To modify the data set, I use JupyterLab in Python. I first filter only on the Eredivisie, and only use the rounds 1 - 34, meaning no Champions League, Conference League, Intertoto, Relegation and other play offs. I remove most of the columns, only keeping the relevant match participants, round, date, scores and betting odds. Furthermore, I transform the dates into seasons, in order to sort the matches on WNR (match number). I also rename some of the columns, club names and coach names, in order to match them with the data set of van Ours & van Tuijl (2016). One point that needs to be addressed too is to match home matches and away matches with the corresponding clubs, so all matches can be sorted on club name, and not only the home matches. This can easily be solved by duplicating the data, then switch within the duplicated data the club names, and the score, and add a zero as binary variable for home match. This way, for every club you will get 34 matches, while being in the first column. Finally, I merge the two data sets together to get the complete Eredivisie match data from season 2000/2001 till season 2021/2022.

For the bookmaker data, van Ours & van Tuijl (2016) mostly used William Hill (97%), Ladbrokes (2%) and Gamebookers (1%). The odds are also already in the database of Gebala (2022), coming from Fortuna Group or Star-Typ Sport, but they slightly differ from the odds used in the paper by van Ours & van Tuijl (2016). However, due to the lack of arbitrage opportunities in betting, the differences in the betting odds are minimal.

TransferMarkt (2023) and Voetbal.com (2023) indicate which coach changes happen in clubs (forced or unforced leave) and when, and I use individual news articles to find the reason why the change happened (PR or NPR). Important to look here is for when the coach change occurred; coach changes occurring during the summer break are not counted as in-season coach changes, as well as coach changes that happened within the first four matches of the season or the last four matches of the season. Media are covering the coach changes heavily, which results in an easy distinction between a PR coach change and a NPR coach change.

After manually adding the new information in Excel and its sources, I again use Python to upload the final rankings of the Eredivisie seasons found on VoetbalInternational (2023). Per game, I map the final ranking of the opponent of previous season. Clubs that were promoted that year get value -1 assigned.

Also, again by using Voetbal.com (2023), I add all coaches' birth dates. By taking the difference in time between the dates of the matches and the trainers' birth dates, I get the trainers' ages.

3.1 Descriptive Data: Coach Changes

Table 3: In-Season Coach Replacements per Season in the Dutch Eredivisie; 2000/2001 - 2021/2022.

Season	D	Q	PR	NPR	T	Clubs
2000-2001	3	1	4	0	4	AZ, De Graafschap, Fortuna Sittard, Sparta Rotterdam
2001-2002	3	1	3	1	4	Ajax, Fortuna Sittard, Roda JC, Vitesse
2002-2003	3	1	3	1	4	AZ, FC Groningen, RBC Roosendaal, Vitesse
2003-2004	2	2	1	3	4	ADO Den Haag, FC Volendam, FC Zwolle, Willem II
2004-2005	3	1	4	0	4	Ajax, FC Den Bosch, NEC, RBC Roosendaal
2005-2006	3	2	4	1	5	FC Twente, NAC Breda NEC, RBC Roosendaal, Willem II
2006-2007	1	2	2	1	3	ADO Den Haag, RKC Waalwijk, Roda JC
2007-2008	3	2	3	2	5	Ajax, Heracles Almelo, PSV, Sparta Rotterdam, Willem II
2008-2009	6	1	5	2	7	De Graafschap, FC Utrecht, Feyenoord, PSV, Roda JC, Vitesse, Willem II
2009-2010	4	1	3	2	5	ADO Den Haag, AZ, NEC, SC Heerenveen, Willem II
2010-2011	2	1	3	0	3	Ajax, Vitesse, VVV-Venlo
2011-2012	3	2	3	2	5	De Graafschap, FC Twente, FC Utrecht, PSV, VVV-Venlo
2012-2013	2	0	2	0	2	FC Twente, NAC Breda
2013-2014	4	0	2	2	4	ADO Den Haag, AZ, Cambuur, Roda JC
2014-2015	4	1	4	1	5	Dordrecht, Go Ahead Eagles, Heracles Almelo, NAC Breda, AZ
2015-2016	1	3	3	1	4	FC Twente, Cambuur Leeuwarden, SC Heerenveen, Vitesse
2016-2017	2	0	2	0	2	ADO Den Haag, Go Ahead Eagles
2017-2018	3	2	3	2	5	Ajax, FC Twente, Sparta Rotterdam, FC Utrecht, Willem II
2018-2019	3	2	4	1	5	FC Utrecht, FC Zwolle, SC Heerenveen, Excelsior, NAC Breda
2019-2020	2	3	5	0	5	PSV, VVV-Venlo, ADO Den Haag, Feyenoord, Vitesse
2020-2021	6	0	5	1	6	ADO Den Haag, AZ, FC Zwolle, Fortuna Sittard, VVV-Venlo, Willem II
2021-2022	3	2	4	1	5	FC Utrecht, SC Heerenveen, Willem II, Cambuur Leeuwarden, FC Zwolle
Total	66	30	72	24	96	

Note: D: Dismissals, Q: Quits, PR: Performance Related Coach Changes, NPR: Non-Performance Related Coach Changes, T: Total number of coach changes.

Table 3 gives a summary of all the coach changes that occurred in the seasons 2000/2001 - 2021/2022. Important to know is that coach changes that happened within the first four matches (WNR 1-4), and within the last four matches (WNR 31-34) are not counted as a coach change due to their often unnatural nature, and lack of performance data at that moment or following match result data. In Table 3, it can be seen that there have been a total of 96 coach changes, of which 66 dismissals and 30 quits, and 72 PR coach changes and 24 NPR coach changes. On average, 4.4 coach changes happen per season, with a maximum of 7 in season 2008/2009, and a minimum of 2 in seasons 2012/2013 and 2016/2017. Willem II has the most coach changes, with a total of eight, followed by ADO Den Haag with a total of seven coach changes.

Table 4: Coach Change Types and their expected Number in the Dutch Eredivisie; 2000/2001 - 2021/2022.

	PR	NPR	Total
Unforced Leave	15 (22.5)	15 (7.5)	30
Forced Leave	57 (49.5)	9 (16.5)	66
Total	72	24	96
$\chi^2(1)$		12.7	

Note: Expected number in brackets, $\chi^2(1)$: Chi-Squared Test with Yates correction, and one degree of freedom. PR: Performance Related, NPR: Non-Performance Related.

Table 4 also shows the number of total coach change per type. Of the 72 PR coach changes, 57 were forced leaves. The other 15 changes were coaches who did not want to continue coaching the team due to bad performances. Of the 24 NPR coach changes, only 9 were forced leaves, most of which were caused by friction between players or the board and the coach, or inappropriate behaviour. The other 15 unforced leaves are majorly caused by health problems or transfers to other clubs. It seems that PR coach changes are over-represented by forced leaves, and NPR coach changes are over-represented by unforced leaves. To test for an association between the two factors, we use the two-by-two contingency chi-squared test, with the expected frequencies depicted in the brackets. The Chi-Squared Test with Yates correction and one degree of freedom provides a test statistic of 12.7. The p-value is smaller than 0.05, which means that we reject the null hypothesis of the observation frequencies being similar to the expected frequencies. The differences did not happen by chance, and there is a relation between forced and unforced leaves, and PR and NPR coach changes.

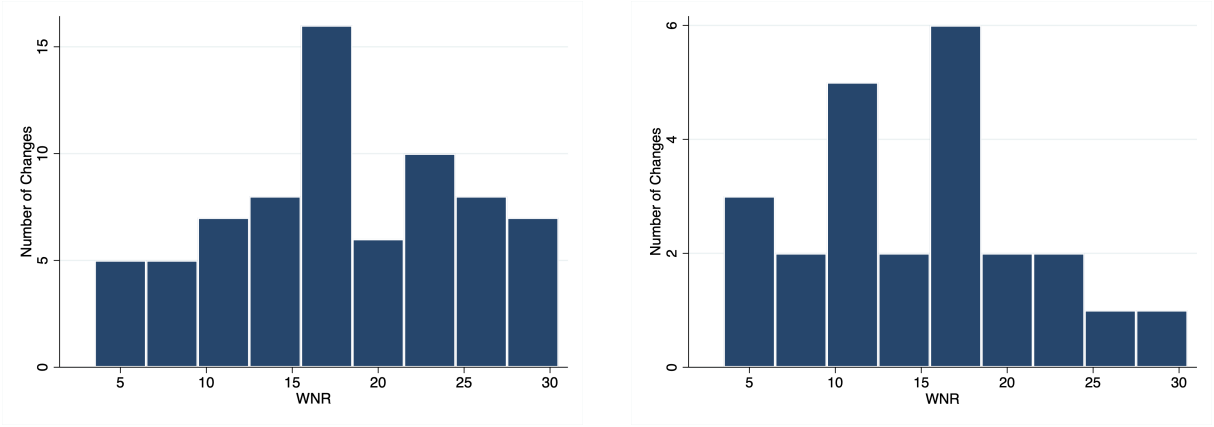

Figure 1: Performance Related (left) and Non-Performance Related (right) Coach Changes per WNR (bins of 3) in the Dutch Eredivisie; 2000/2001 - 2021/2022.
Note: Winter break is usually between WNR 17 & 18.

Figure 1 shows two histograms picturing the PR (left) and NPR (right) coach changes per WNR. As can be seen from the two histograms, most coach changes happen during the winter break. In the Dutch Eredivisie, the winter break usually lasts one month, from half December till half January. During this period, there is also a transfer window or registration period. In this time, clubs have time to find a new coach and to let the new coach train with the players. It is therefore logical that a relatively large number of coach changes happen during this period. However, a difference between PR and NPR coach changes, is the PR coach changes have a tendency to happen in the later stages of the season, when more of the results and performance is known. For NPR coach changes, the performance does not matter, and they actually tend to happen in the early stages of the season.

A coach spell is the period that one coach stays at the same club, without any interruptions. This can thus also be multiple seasons, shown in Figure 2. There is one outlier worth mentioning, which is Ron Jans at FC Groningen, who stayed for 264 matches from season 2002/2003 - 2009/2010. A declining

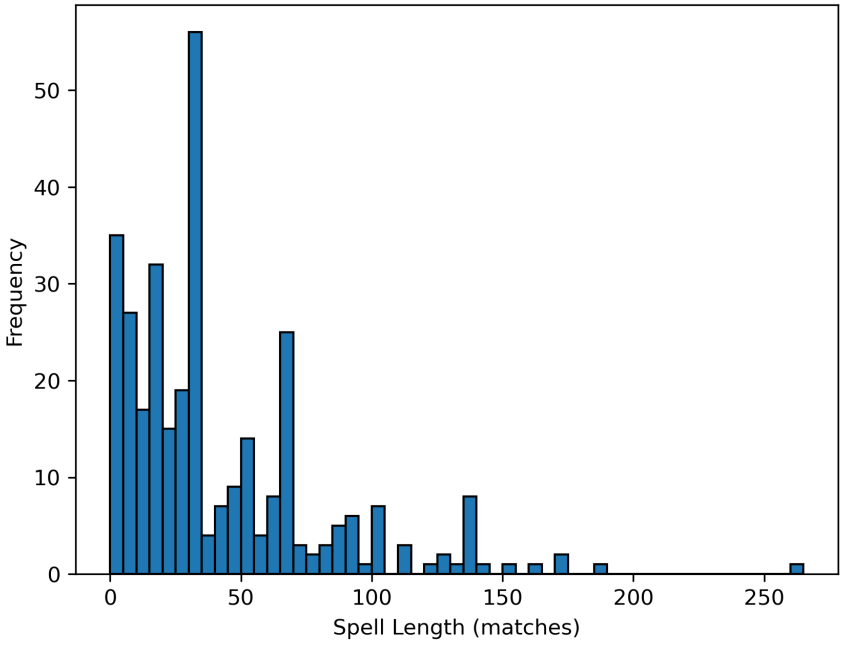

Figure 2: Histogram of Coach Spell Lengths in the Dutch Eredivisie; 2000/2001 - 2021/2022.
Note: All coach spells in season 2000/2001 are assumed to have been started in that season.

Table 5: Coach Spells in the Dutch Eredivisie; 2000/2001 - 2021/2022.

	All Coach Spells	**Coach Spells $\geq$ 5 WNR**	**Coach Spells $\geq$ 35 WNR**
Average	41,5	46,3	79,9
Min	1	5	35
Max	264	264	264
n	321	286	120
N	13,316	13,243	9,582

Note: All coach spells in season 2000/2001 are assumed to have been started in that season. n: Number of Spells, N: Number of Matches.

trend can be seen, with a spike on WNR 34. This is when a coach stays only for one full season, after which a coach change happens. Again, we see a spike around WNR 68, 102, 136, which is when a coach stays for 2, 3, or 4 full seasons respectively. Table 5 shows descriptive data about those coach spells. Interesting to see is that out of the 321 coach spells in the sample period, only 120 coach spells are longer than 1 season. Also, there are 35 coach spells that are less than or equal to four matches. Most of those small period coach spells can be attributed to interim coaches who just coach till a new coach is found.

For the performance indicators in this paper, three variables are used: the cumulative surprise, the cumulative number of points, and WL4M. Whereas the cumulative number of points is just the total number of points a club obtained in a season up till that moment, the cumulative surprise requires a more extensive explanation. On each match, there is bookmaker data; the odds on home team winning, away team winning or a draw. With those betting odds, following Equations 1, 2, 3 and 4, an expectation can be made on which team is going to win that game, and an expectation can be made on how much points a club is going to obtain. If the actual result differs from the expectation, there is "surprise",

which is the difference in points between actual result and expectation. The cumulative surprise is the total surprise that occurred for a club in a season up till that moment.

$$\text{Premium Bookmaker} = \left(\frac{1}{\text{Odds Home Winning}}\right) + \left(\frac{1}{\text{Odds Away Winning}}\right) + \left(\frac{1}{\text{Odds Draw}}\right) \tag{1}$$

$$P(\text{Home Winning}) = \frac{\left(\frac{100}{\text{Odds Home Winning}}\right)}{\text{Premium Bookmaker}} \tag{2}$$

$$P(\text{Away Winning}) = \frac{\left(\frac{100}{\text{Odds Away Winning}}\right)}{\text{Premium Bookmaker}} \tag{3}$$

$$P(\text{Draw}) = \frac{\left(\frac{100}{\text{Odds Draw}}\right)}{\text{Premium Bookmaker}} \tag{4}$$

Figure 3 pictures the kernel densities of those two performance indicators, at the end of the season for no coach change, for all coach changes and for four different types of coach changes club seasons: coach dismissals (forced leaves), coach quits (unforced leaves), PR coach changes and NPR coach changes. As can be seen from cumulative surprise (left) kernel density, the no coach change density has a peak around zero cumulative surprise. All the coach changes, coach dismissals and PR coach changes are all left of the no coach change line. This means, that for those coach changes, there is a more negative cumulative surprise than for the no coach change club seasons. This seems logical, as clubs that perform worse, have a higher probability of a PR coach change. However, interesting is the NPR coach change density and the quit density, which are more to the right of the no coach change density, have a peak in the positive cumulative surprise. It seems that NPR coach change club seasons and quit seasons perform unexpectedly well. One explanation for this might be that coaches that underwent coach transfers to other more prominent clubs caught the eye of those prominent clubs due to their unexpected good performance.

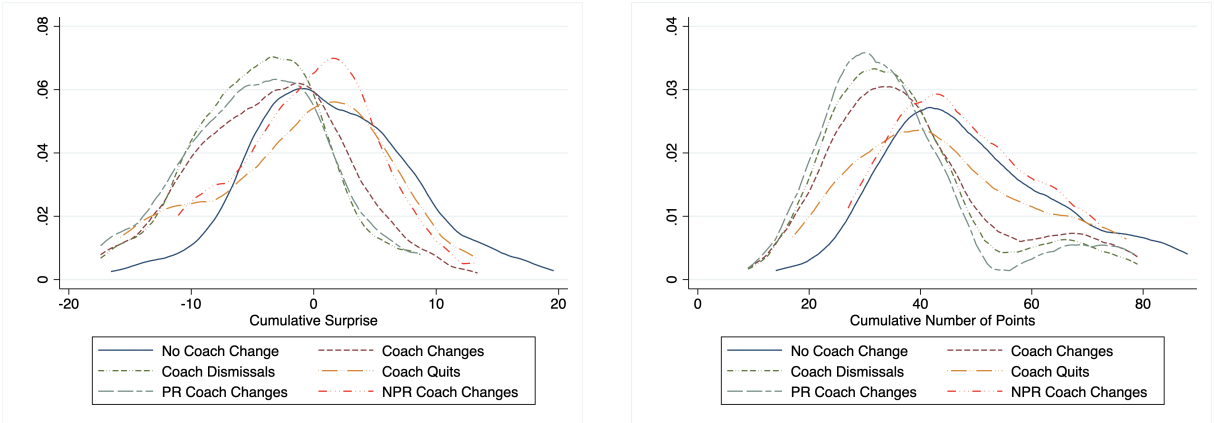

Figure 3: Kernel Densities of Cumulative Surprise (left) and Cumulative Points (right) at end of the Dutch Eredivisie Season, for different types of coach changes; 2000/2001 - 2021/2022

The right kernel density of Figure 3 depicts the cumulative points for the same five coach change scenarios. Again, we see that all coach changes, coach dismissals, and PR coach changes club seasons perform significantly worse than the average no change club season, with a peak around 30 cumulative points. This time, NPR coach change and quit densities are similar to the no coach change density.

Table 6 provides the summary statistics of the variables used in this paper, for all coach changes and four different types of coach changes. The average position of the opponent is not always 9.5 as expected, due to the promotion of multiple clubs (getting value 18 assigned) in some seasons. The averages for Points, Victories, Goal Differences, Cumulative Surprise and Cumulative Points seem relatively similar for All Coach Changes, Coach Dismissals and PR Coach Changes, although somewhat decreasing in that order. However, Quits and NPR Coach Changes have averages that are much higher, indicating that club seasons in those categories perform better. This also confirms the findings found in Figure 3 mentioned and described before. As shown in Table 4, the two following a somewhat similar pattern is expected as 50% of the total amount of quits are NPR.

Table 6: Summary Statistics of Variables used in the Analysis on the Dutch Eredivisie; 2000/2001 - 2021/2022; Means, Minimum, Maximum.

	All Coach Changes			**Dismissals**			**Quits**			**PR Coach Changes**			**NPR Coach Changes**		
	Mean	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Home Match	0.50	0	1	0.50	0	1	0.50	0	1	0.50	0	1	0.50	0	1
Position(k-1)	9.53	1	18	9.49	1	18	9.62	1	18	9.51	1	18	9.59	1	18
Points	1.16	0	3	1.10	0	3	1.31	0	3	1.08	0	3	1.41	0	3
Victories	0.31	0	1	0.29	0	1	0.36	0	1	0.29	0	1	0.39	0	1
Goal Diff.	-0.37	-13	8	-0.47	-13	8	-0.15	-9	7	-0.50	-13	8	0.01	-9	7
Cum. Surprise	-2.48	-17.41	13.37	-3.11	-17.41	8.71	-1.08	-15.78	13.37	-3.29	-17.41	8.78	-0.08	-11.06	13.37
Cum. Points	19.51	0	79	18.32	0	79	22.15	0	77	17.68	0	79	24.90	0	72
n		96			66			30			72			24	
N		3,223			2,228			995			2,407			816	

Note: PR: Performance Related, NPR: Non-Performance Related, n: Number of Seasons, N: Number of Matches. Position(k-1) is the position of the opponent, in the previous season. If club was promoted last year, it gets position 18.

3.2 Descriptive Evidence: Relegation

Table 7: Promoted and Relegated Clubs in the Dutch Eredivisie per season; 2000/2001 - 2021/2022.

Season	Relegated Clubs	Promoted Clubs	N
2000-2001 → 2001-2002	RBC Roosendaal	FC Den Bosch	1
2001-2002 → 2002-2003	Sparta Rotterdam, Fortuna Sittard, FC Den Bosch	Excelsior, RBC Roosendaal, FC Zwolle	3
2002-2003 → 2003-2004	Excelsior, De Graafschap	FC Volendam, ADO Den Haag	2
2003-2004 → 2004-2005	FC Volendam, FC Zwolle	FC Den Bosch, De Graafschap	2
2004-2005 → 2005-2006	FC Den Bosch, De Graafschap	Sparta Rotterdam, Heracles Almelo	2
2005-2006 → 2006-2007	RBC Roosendaal	Excelsior	1
2006-2007 → 2007-2008	ADO Den Haag, RKC Waalwijk	VVV-Venlo, De Graafschap	2
2007-2008 → 2008-2009	VVV-Venlo, Excelsior	FC Volendam, ADO Den Haag	2
2008-2009 → 2009-2010	FC Volendam, De Graafschap	VVV-Venlo, RKC Waalwijk	2
2009-2010 → 2010-2011	Sparta Rotterdam, RKC Waalwijk	Excelsior, De Graafschap	2
2010-2011 → 2011-2012	Willem II	RKC Waalwijk	1
2011-2012 → 2012-2013	Excelsior, De Graafschap	Willem II, FC Zwolle	2
2012-2013 → 2013-2014	VVV-Venlo, Willem II	Go Ahead Eagles, Cambuur Leeuwarden	2
2013-2014 → 2014-2015	Roda JC, NEC, RKC Waalwijk	Excelsior, Dordrecht, Willem II	3
2014-2015 → 2015-2016	Go Ahead Eagles, NAC Breda, Dordrecht	Roda JC, NEC, De Graafschap	3
2015-2016 → 2016-2017	Cambuur Leeuwarden, De Graafschap	Sparta Rotterdam, Go Ahead Eagles	2
2016-2017 → 2017-2018	Go Ahead Eagles, NEC	NAC Breda, VVV-Venlo	2
2017-2018 → 2018-2019	Sparta Rotterdam, Roda JC, FC Twente	Fortuna Sittard, FC Emmen, De Graafschap	3
2018-2019 → 2019-2020	NAC Breda, Excelsior, De Graafschap	Sparta Rotterdam, FC Twente, RKC Waalwijk	3
2019-2020 → 2020-2021			0
2020-2021 → 2021-2022	VVV-Venlo, ADO Den Haag, FC Emmen	Go Ahead Eagles, Cambuur Leeuwarden, NEC	3
Total			43

Note: N: Number of Promoted and Relegated Clubs. Due to the Covid-19 pandemic, there were no promotions and relegations at the end of season 2019-2020.

Table 7 provides an overview of all the relegated and promoted clubs in the sample period, with a total of 43 relegations and promotions. A significant season is the 2019/2020 season, which was cancelled at match 25/26 due to the Covid-19 pandemic, and therefore no relegations happened. On average, without season 2019/2020, 2.2 relegations happen per year. De Graafschap is the club with the most relegations (six) and promotions (five).

Table 8: Connection Managerial Change and Relegation in the Dutch Eredivisie; 2000/2001 - 2021/2022.

	Managerial change		
Relegation	**No**	**Yes**	**Total**
No	265 (254.35)	70 (80.65)	335
Yes	22 (32.65)	21 (10.35)	43
Total	287	91	378
$\chi^2(1)$		14.8	

Note: Expected number in brackets, $\chi^2(1)$: Chi-Squared Test with Yates correction, with one degree of freedom.

Tena & Forrest (2007) already showed that being in the relegation zone increases the probability of a coach change. Adding the simple analysis of whether a club relegates or not when there was a coach change or not, based on the idea of Arrondel et al. (2020), may provide quick additional insight into the effect of a coach change. Just like the Chi-Squared Test performed on Table 4 in the previous section, I perform the same Chi-Squared Test with Yates correction on Table 8. As Table 8 shows, of the 378 transitions (18 clubs and 21 season changes), 43 were relegations, while 335 clubs stayed in the Dutch Eredivisie. There were 91 managerial changes, as the last five coach changes of season 2021/2022 are not taken into account here. With the expected frequencies in the brackets next to the observed frequencies,

it seems that relegations are over-represented with managerial changes, whereas no relegations seem under-represented by managerial changes. The Chi-Squared Test gives a test statistic of 14.8, with one degree of freedom. The p-value is smaller than 0.05, which means that we reject the null hypothesis of observed frequencies being similar to the expected frequencies. There seems to be a relation between relegating and managerial change.

4 Methodology

4.1 Naive Approach

The performance of a club in a specific season can be measured by multiple different performance indicators. Most studies performed in the past only observe the number of points obtained in a match as performance indicator. To gain more insights, we add similar to van Ours & van Tuijl (2016) a binary variable for a victory (1) and a loss (0), and the goal difference. To determine the effect of a coach change without any control group, the "naive" approach, a simple regression that includes a dummy variable for a coach change is used (van Ours & van Tuijl (2016)):

$$y_{ijk} = \eta_{ik} + \beta r_{ijk} + \delta d_{ijk} + \varepsilon_{ijk} \tag{5}$$

The performance is measured by y_{ijk} (dependent variable), being the number of points, or being a victory or loss, or the goal difference of team i in match j of season k. As club-season fixed effect, we have η_{ik}, which represents the quality of a team in a specific season. This intercept is allowed to vary between clubs and seasons. Then for the independent variables, r_{ijk} represents a vector of potential determinants of the performance; in this naive regression, I use the position of the opponent in previous season as determinant and whether the game is a home match. d_{ijk} is a dummy variable for whether or not a coach change occurred. β and δ are the corresponding parameter estimates, and ε_{ijk} is the error term. In this regression, δ indicates the naive effect of a coach change, and is thus the most interesting variable we look at.

This naive regressions is first performed for all coach changes and coach dismissals, to compare the results of the Dutch Eredivisie seasons 2000/2001 - 2021/2022 with the results of van Ours & van Tuijl (2016), who study the Dutch Eredivisie seasons 2000/2001 - 2013/2014. In addition, the sample will be split up in two different samples: seasons 2000/2001 - 2010/2011, and seasons 2011/2011 - 2021/2022. This split shows us how the coach change effect developed over the last two decades. Furthermore, this naive regression will also be performed for PR and NPR coach changes, over the whole sample period.

4.2 Treatment and Control Group

To account for the selection bias, we include control groups in the regression. After all, the performance of a club could possibly have followed exactly the same pattern even if there was no coach change. Including

a dummy variable for being in the control group, gives a somewhat similar regression as before:

$$y_{ijk} = \eta_{ik} + \beta r_{ijk} + \delta d_{ijk} + \mu m_{ijk} + \varepsilon_{ijk} \tag{6}$$

where now d_{ijk} is a dummy variable for a coach change (treatment group), and m_{ijk} is a dummy variable for a counterfactual coach change that did not happen (control group). Again, r_{ijk} represents a vector of potential determinants of the performance, being the position of the opponent in previous season and whether the game is a home match. In this regression, δ indicates the effect of a coach change on the treatment group, and μ indicates the effect of a coach change that did not happen (counterfactual coach change). It is interesting to see how much those two coefficients vary.

By performing a F-test, we can test equality of the parameter estimates of the dummy variables for treatment (δ) and control group (μ). For high enough test statistics, we reject the null hypothesis of the equality between the effect of actual and counterfactual coach changes, and then the effect of an actual coach change differs significantly. If not, the effect of a coach change is not significantly different, and we can attribute any possible positive effects to the regression to the mean phenomenon. As PR coach changes are the most interesting treatment group, I will only focus on this treatment group.

4.3 Quasi-Exact Matching

To form a control group, multiple methods can be used. One of the most popular methods is the "Nearest Neighbour Method"(Paola & Scoppa (2012), Abadie & Imbens (2006), Abadie et al. (2004), Bruinshoofd & Ter Weel (2003), Balduck, Buelens, & Philippaerts (2010), and van Ours & van Tuijl (2016)). It is impossible to know what would have happened with the performance of a team that had a coach change, if that coach change had not occurred. However, it is possible to look for similar teams with similar characteristics that did not perform a coach change. In this method, every member of the treatment group with a coach change, is matched, if possible, to teams with almost the same covariate values that did not have a coach change, forming the control group. This control group now has counterfactual coach changes; coach changes that did not occur. The treatment group and control group are comparable due to their similar characteristics. The matching can be done on one specific observable criterion, such as van Ours & van Tuijl (2016) did on cumulative surprise, but can also be done on multiple observable variable criteria like Paola & Scoppa (2012) did on points obtained in the last four matches, rank difference, dummies for season periods, and seasons itself.

In this study, three matching criteria will be used. Those are all performance indicators, namely cumulative points, cumulative surprise based on bookmaker odds, and WL4M. Literature, as shown in Section 2, indicates that past performance is a determinant of coach dismissal, which is why cumulative points is taken into account. van Ours & van Tuijl (2016) and Pieper et al. (2014) prove that cumulative surprise is an interesting and quite recent found determinant of coach dismissal. The win-rate of the past four matches criterion is based on the studies performed by Salomo & Teichmann (2000), Frick et al. (2010), Pieper et al. (2014) and van Ours & van Tuijl (2016), that showed the importance of recent match

results in coach dismissal probability. To validate the use of those three coach change determinants, I will regress those determinants on the PR coach changes. If the coefficients are significant, they have a substantial impact on the PR coach change probability. An advantage of using matching criteria, is that you only use treatment and control group teams that are sufficiently similar, with common support, and ignore the outliers. To explore QEM (Arrondel et al. (2020)), we allow seven different combinations of the three matching criteria, and thus seven different control groups. Every member in the control groups is the same club as the club of the corresponding member in the treatment group.

4.4 ATT

Whereas Paola & Scoppa (2012) use Average Treatment Effect (ATE), Arrondel et al. (2020) use the Average Treatment Effects on the Treated (ATT). As we are mostly interested in the effect of the clubs in specific seasons that had a coach change, and not on all clubs, ATT is preferred. For the ATT, the following equation can be used (Acquah (2020)):

$$\begin{aligned} \text{ATT} &= E\left[\delta_{ik} \mid X_{ik} = 1\right] \\ &= E\left[Y_{ik}^{1} - Y_{ik}^{0} \mid X_{ik} = 1\right] \\ &= E\left[Y_{ik}^{1} \mid X_{ik} = 1\right] - E\left[Y_{ik}^{0} \mid X_{ik} = 1\right] \\ &= \sum_{i=1}^{N_T} \left(y_{ik}^{1} \mid x_{ik} = 1\right) - \sum_{i=1}^{N_T} \left(y_{ik}^{0} \mid x_{ik} = 1\right) \end{aligned} \tag{7}$$

Where $X_i = 1$ means that the club i in season k is in the treatment group, and δ_{ik} is the non-naive effect of a coach change. y_{ik}^{1} represents the performance indicator of that team with a coach change, whereas y_{ik}^{0} is the performance indicator of that same team, where the coach change did not occur, thus being the corresponding counterfactual coach change in the control group. ATT is the result of a coach change minus the result if that coach change did not happen, and is thus the "gain" in performance a team received because of their coach change.

5 Results

5.1 Naive Approach

In the Table 9, the results of the naive regression can be seen for all coach changes, coach dismissals and coach quits, for three different time periods: 2000/2001 - 2010/2011, 2011/2012 - 2021/2022 and 2000/2001 - 2021/2022. Coincidentally, all number of seasons in the full sample are split perfectly over the two independent smaller time samples. To start off with all coach changes, it can be seen that the effect of a coach change is present and that it is positive, and that all coefficients are significant at a 5% level. In the full time period (2000/2001 - 2021/2022), the effect of a coach change on points is an increase of 0.24, for victory 0.08 and for goal difference 0.33. With the averages given in Table 6, for

Table 9: Naive Effects of Coach Changes in the Dutch Eredivisie, over 3 different Time Periods.

	2000/2001 - 2010/2011			**2011/2012 - 2021/2022**			**2000/2001 - 2021/2022**		
	P	**V**	**Goal Diff.**	**P**	**V**	**Goal Diff.**	**P**	**V**	**Goal Diff.**
All coach changes									
Position(k-1)	0.07 (12.0)**	0.02 (9.4)**	0.12 (13.6)**	0.06 (12.9)**	0.02 (10.7)**	0.12 (16.3)**	0.07 (17.6)**	0.02 (14.2)**	0.12 (20.8)**
Home Match	0.62 (13.9)**	0.20 (13.1)**	1.02 (11.9)**	0.45 (6.9)**	0.14 (6.3)**	0.74 (8.2)**	0.53 (13.3)**	0.17 (12.3)**	0.88 (13.9)**
Coach Change	0.28 (4.9)**	0.09 (4.3)**	0.37 (4.1)**	0.20 (2.6)**	0.07 (2.3)**	0.29 (2.5)**	0.24 (5.12)**	0.08 (4.5)**	0.33 (4.6)**
N		1,632			1,591			3,223	
n		48			48			96	
Coach dismissals									
Position(k-1)	0.06 (10.8)**	0.02 (8.3)**	0.12 (12.1)**	0.06 (10.4)**	0.02 (8.4)**	0.12 (13.6)**	0.06 (15.1)**	0.02 (11.9)**	0.12 (18.1)**
Home Match	0.66 (14.2)**	0.21 (13.6)**	1.14 (12.1)**	0.45 (5.6)**	0.16 (5.3)**	0.70 (6.5)**	0.56 (11.7)**	0.18 (10.9)**	0.92 (12.1)**
Coach dismissals	0.35 (5.9)**	0.12 (5.1)**	0.53 (5.2)**	0.21 (2.6)**	0.08 (2.6)**	0.29 (2.0)*	0.28 (5.7)**	0.10 (5.4)**	0.41 (4.8)**
n		33			33			66	
N		1,122			1,106			2,228	
Coach quits									
Position(k-1)	0.07 (5.9)**	0.02 (5.0)**	0.12 (6.4)**	0.07 (7.6)**	0.02 (6.7)**	0.12 (8.6)**	0.07 (9.3)**	0.02 (7.9)**	0.12 (10.4)**
Home Match	0.50 (5.3)**	0.19 (5.1)**	0.76 (4.5)**	0.45 (3.9)**	0.12 (3.4)**	0.83 (5.0)**	0.48 (6.5)**	0.16 (5.9)**	0.79 (6.8)**
Coach Quits	0.11 (1.0)	0.03 (0.8)	0.03 (0.2)	0.19 (1.1)	0.05 (0.7)	0.30 (1.4)	0.15 (1.5)	0.04 (1.0)	0.16 (1.2)
n		15			15			30	
N		510			485			995	

Note: Absolute t-statistic given in brackets, based on robust standard errors. P: Points, V: Victories, n: Number of Seasons, N: Number of Matches
** (*) indicates significance at 5%(10%).

points being 1.16, victories 0.31 and goal difference -0.37, the positive effect of a coach change seems substantial; increasing both points and victory chance with around 25%, and almost neutralizing the goal difference.

Table 10: Naive Effect of Performance Related and Non-Performance Related Coach Changes in the Dutch Eredivisie; 2000/2001 - 2021/2022.

	2000/2001 - 2021/2022		
	P	**V**	**Goal Diff.**
PR coach changes			
Position(k-1)	0.06 (15.0)**	0.02 (12.3)**	0.12 (17.7)**
Home Match	0.52 (11.6)**	0.16 (10.7)**	0.85 (11.6)**
Coach changes	0.35 (7.3)**	0.12 (6.5)**	0.47 (6.0)**
N		72	
n		2,407	
NPR coach changes			
Position(k-1)	0.07 (8.8)**	0.02 (6.8)**	0.12 (10.3)**
Home Match	0.58 (6.7)**	0.20 (6.4)**	1.00 (7.9)**
Coach changes	-0.07 (-0.8)	-0.03 (-0.7)	-0.07 (-0.5)
n		24	
N		816	

Note: Absolute t-statistic given in brackets, based on robust standard errors. P: Points, V: Victories, n: Number of Seasons, N: Number of Matches.
** (*) indicates significance at 5% (10%).

Looking at the two decade samples, there is an interesting pattern. The coach change effect declined with almost 33% in the last decade (2011/2012 - 2021/2022) relative to the decade before (2000/2001 - 2010/2011) for every performance indicator (points, victories and goal difference). For example, for points, the coach change effect decreased from 0.28 to 0.20 in the next decade, and for victories there was a decrease from 0.09 to 0.07. Although the home advantage also had a smaller effect in the next decade, the effect of the position of the opponent remained the same. The 2000/2001 - 2010/2011 sample shows similar results to the results found in van Ours & van Tuijl (2016), with the only difference being a little smaller coach change effect on points (0.25), which is logical due to their use of a sample up to season 2013/2014. Later seasons have a decreased coach change effect on the performance indicators.

All coefficients of coach dismissals are significant at a 5% level, except for the coach dismissal effect for Goal Differences for the 2011/2012 - 2021/2022 sample period, which is significant at a 10% level. The naive positive effect of a coach dismissal is larger than the positive effect found for all coach changes. Comparing the effect in the full sample for points gives 0.28 for coach dismissals and 0.24 for all coach changes. Again, those effects seem substantial, even more than with all coach changes, as the averages found in Table 6 for coach dismissals are 1.10 for points, 0.29 for victories and -0.47 for goal difference. The decreasing positive effect found for all coach changes is also present for the coach dismissals, but even more heavily. Whereas both coach change effects in the last decade for goal difference is 0.29, the effect in the decade before is 0.53 for coach dismissals, and 0.37 for all coach changes. It could be argued that in the last decade (2011/2012 - 2021/2022), the effect of coach changes are almost the same for all coach changes as for coach dismissals.

Although all effects of coach quits are positive and heavily increasing through time, no conclusions can be drawn due to the fact that they are all insignificant. The number of seasons used for quits in the smaller samples is less than half of the number of dismissals used in those samples. As explained before, Figure 3 and Table 6 show that like NPR coach changes, only 50% of all coach quits are related to bad performances. The cumulative surprise and cumulative points of seasons with coach quits is not below average, and thus we do not expect a coach quit to have a significantly positive effect. With this logic, we should see in Table 10, that the effects of NPR coach changes are not significant either.

Now looking at the naive effect of PR coach changes and NPR coach changes over the whole sample period 2000/2001 - 2021/2022 in Table 10, we see some interesting results. Starting with the effect of the PR coach changes, extreme high positive effects can be found: 0.35 for points, 0.12 for victories, and 0.47 for goal difference. Interestingly, even though the number of matches and number of seasons used is higher here for PR coach changes, the coach dismissal effect values of 0.28 for points, 0.10 for points, and 0.41 for goal difference are smaller. One reason for this is that PR coach changes, logically, always have to do with performance. The team is performing relatively bad, and almost anything would be better than the current situation. For coach dismissals, there are also coach changes that have to do with friction between the players or the board and the coach, or inappropriate behaviour of the coach. Those particular coach changes have no connection with performance, and would likely decrease the positive effect of a coach change. The coefficients of the position of the opponent are similar to the other three coach change types, but the home advantage effects are smaller for all three performance indicators.

Investigating the effect of NPR coach changes gives the expected, but nevertheless interesting results. The effects on the three performance indicators are insignificant, just like the effects of the coach quits. Even in the naive regression, the effect of a NPR coach change does not matter in any way. The club performs the same as before. This could be explained by the nature of NPR coach changes. NPR coach changes are mostly associated with coaches transferring to other clubs, or coaches' health problems. The clubs are thus not performing (significantly) bad, and therefore it is logical and expected that the club does not suddenly perform better. However, one could expect a negative effect of the coach change, due to new relationship between the players and the new coach. This small negative effect of a NPR coach change can be seen in Table 10, although it is insignificant.

5.2 Treatment and Control Groups

By first performing a simple regression of the three performance indicators, and the personal coach characteristic age, on the different types of coach changes, insights are obtained on the impact of variables on coach change probability. As shown in Table 11, four different linear regressions are formed for the four different type of coach changes. Starting off with the personal coach characteristic Age, it can be seen that it has no impact on the probability of the coach changes, except for coach quits which is significant at a 10% level. The coefficient is negative, meaning that as age increases, the probability of a coach quit decreases. This could potentially have to do with the persistence of the older trainers, but more detailed

research has to be performed to find out why exactly.

Next, the NPR coach changes shows as expected non-significant performance indicators coefficients, except for the cumulative points coefficient which is -0.06x10^{-03}. One explanation for this very small but negative relationship could be underlying correlation between the cause of coach change, like team dynamics or coach health, and the performance of a team. If the dynamics or health are good, performance is better, and a thus less likely NPR coach change. For quits, the performance indicators are insignificant, except for cumulative surprise. High positive cumulative surprise leads to lower probability of a quit, which seems likely as a 50% of quits are PR.

Dismissals and PR seem to follow somewhat the same significance patterns, but the performance indicators' coefficients are larger for PR coach changes. As can be seen for PR coach changes, all three performance indicators have a significant impact on the probability of a PR coach change. With cumulative points being fairly small but positive, cumulative surprise and WL4M have a large negative effect on the probability.

Table 11: Regression Coefficients of Performance Indicators and Age on Different Types of Coach Changes

	Dismissals	**Quits**	**PR**	**NPR**
Cum Surp	-0.64 (2.9)**	-0.26 (2.0)**	-0.96 (4.3)**	0.06 (0.5)
Cum Points	0.01 (0.5)	-0.01 (0.4)	0.06 (2.2)**	-0.06 (2.6)**
WL4M	-1.89 (3.5)**	-0.07 (0.2)	-2.61 (4.9)**	0.65 (1.3)
Age	0.01 (0.0)	-0.59 (1.8)*	-0.54 (0.9)	-0.04 (0.1)

Note: All coefficient values are expressed in units of 10^{-03}. Absolute t-statistic given in brackets, based on robust standard errors. PR = Performance Related Coach Changes, NPR: Non-Performance Related Coach Changes, WL4M: Wins in Last Four Matches.
** (*) indicates significance at 5% (10%).

As explained before, only using a naive regression to assess the effect of a coach change would contribute little, due to unknowing of what would have happened without the coach change. Therefore, counterfactual coach changes are needed; club seasons with coach changes that did not happen that are similar to club seasons with coach changes that did happen. The former will be the control group, whereas the latter is our treatment group. There are many ways to form a control group and in this analysis, three matching criteria will be explored; cumulative surprise based on the bookmaker odds, cumulative points and the win rate of the last four matches before the coach change. Changing the strictness per criteria gives different types of QEM, and thus control groups. The use of different control groups and their results may give insights on the use of the three matching criteria.

Table 12 shows the criteria combinations for the QEMs used in this paper. For every matching, there is an additional criterion: the club in the control group has to be the same club as its corresponding treatment group member. There are a total of 72 treatment group members, which can all have (if possible) multiple control group members, depending on the matching criteria. QEM 1 and QEM 2 focus on cumulative surprise and cumulative surprise and win rate in last 4 matches, respectively. QEM 1, similar to the matching done in van Ours & van Tuijl (2016), although done in a different manner, only has one criterion: the cumulative surprise difference between the treatment group member and its

Table 12: Characteristics of the Quasi Exact Control Group Matching on the Performance Related Coach Changes in the Dutch Eredivisie; 2000/2001 - 2021/2022.

		Matching Criteria		
	Control Group n	**Cum Points**	**Cum Surp**	**WL4M**
QEM 1 (CS)	88	unrestricted	+/- 0.05	unrestricted
QEM 2 (CS + WL4M)	45	unrestricted	+/- 0.05	exact
QEM 3 (CP)	220	exact	unrestricted	unrestricted
QEM 4 (CP + WP4M)	118	exact	unrestricted	exact
QEM 5 (CP + CS)	77	+/- 5	+/- 0.2	unrestricted
QEM 6 (CP + CS + WP4M)	36	+/- 5	+/- 0.2	exact
QEM 7 (CP + CS + WL4M)(Opposite)	38	+/- 2	+/- 0.5	exact

Note: All control group members are the same club as their corresponding treatment group member. Number in Treatment Group is 72. n = Number of Seasons, QEM = Quasi-Exact Matching, WL4M: Wins in Last Four Matches.

corresponding treatment group members may not have a cumulative surprise difference exceeding 0.05. Having this criterion gives a control group consisting of 88 members. QEM 2 adds an extra matching criteria: having exactly the same WL4M. This constraints reduces the number of control group club seasons by half. cumulative points remains unrestricted in both those QEMs.

QEM 3 and 4 focuses on cumulative points, and cumulative points and WL4M, respectively. Both matchings are exact matching, meaning that control group members are selected only if they have have a cumulative number of points equal to the treatment group member on the last match of the coach. Cumulative surprise is unrestricted.

QEM 5 explores the combination between matching on cumulative points (maximum difference of five) and cumulative surprise (maximum difference of 0.2). QEM 6 is equal to the matching done in QEM 5, but adds exact matching based on WL4M. To assess the effect of a trade-off in strictness between performance indicator criteria, QEM 7 makes the cumulative surprise criterion more lenient, while making the cumulative points criterion more strict. As can be seen, adding more constraints always results in a lower number of club seasons in the control group. The use of the seven different QEMs should shed some light on the patterns present in the use of matching criteria.

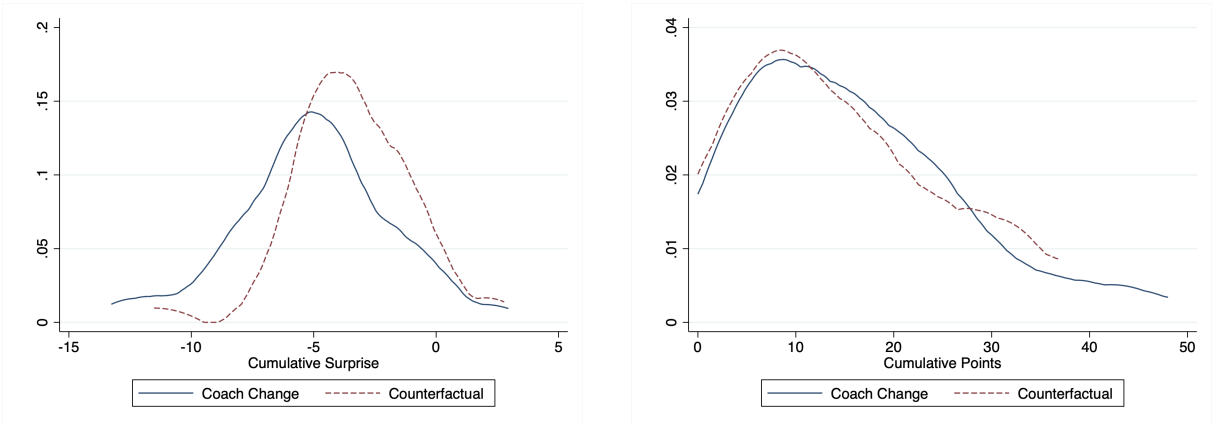

Figure 4: Kernal Density of Cumulative Surprise (left) and Cumulative Points (right) at Last Match of the Coach of QEM 7; Actual Performance Related Coach Changes and Counterfactuals in the Dutch Eredivisie; 2000/2001 - 2021/2022

Based on the matching criteria depicted in Table 12, control groups are made. For QEM 7, one of the most strict matchings, Figure 4 shows the treatment group's and control group's cumulative surprise (left) and cumulative points (right). As can be seen in the left figure, the control group's cumulative surprise follows the PR Coach Changes' cumulative surprise quite well, although the lower treatment group members with low cumulative surprise seem under-represented, and high cumulative surprise seems over-represented. Less negative cumulative surprise club seasons are less hard to match with similar club seasons, and often have more matchings than the treatment group club seasons with extremely low cumulative surprise. In the figure on the right, the counterfactual cumulative points kernel density follows the treatment group's kernel density perfectly.

5.3 Actual Effect of Coach Changes - Treatment and Control Groups

Table 13 contains the effects of a PR coach change and a counterfactual PR coach change on the three performance indicators points, victory, and goal difference. Notice that the effect values given in the column 'Matched - Treatment' are the same as the effect values given in Table 10 on the naive regression of PR coach changes. Per QEM, the counterfactual coach change effects, and the ATT, which is the Average Effect of Treatment on the Treated, are given. The values in this column are followed by F-statistics in brackets, constructed by a F-test investigating the similarity of the effects of the PR coach change and counterfactual PR coach change.

QEM 1, focusing only on the cumulative surprise constraint with a maximum difference of 0.05, provides significant results. The found counterfactual coach change effects are similar to the ones found in van Ours & van Tuijl (2016). For all three performance indicators, we have a F-test with $p \geq 0.05$, meaning that we can not reject the null hypothesis of similarity; the effect is similar for actual and counterfactual coach changes. When adding WL4M exact matching to this matching, resulting in QEM 2, we get even more significant similarity between the effects of the treatment and control group. The effects found in the control group are almost exactly the same as the effects found in the treatment group, leading to F-statistics of 0.0 for points and victories. Although the difference in effect is bigger for goal difference, the effects are still significantly similar.

When matching on the exact same number of cumulative points in QEM 3, we obtain non-significant negative counterfactual coach change effects. The numbers are insignificantly different from zero and it is therefore hard to not reject the null hypothesis of similarity of effect for any of the three performance indicators. In theory, this can be logically explained. Cumulative points as only criterion seems like a misused matching criterion. Clubs with a PR coach change usually have a low number of cumulative points after a number of lost games, and are therefore matched to club seasons that may be doing relatively well, but are still early in the season. This way, badly performing club seasons in the treatment group are matched to good performing club seasons in the control group, which is a mismatch. Adding WL4M exact matching to this matching leads to QEM 4, and gives mixed results. Although the counterfactual coach change effects for points and victories are significant, only the effects for victories between the

Table 13: Effects of Performance Related Coach Changes in the Dutch Eredivisie - Treatment and Control Groups; 2000/2001 - 2021/2022.

	Matched - Treatment	**Matched - Control**	**ATT**
QEM 1 (CS)			
P	0.35 (7.3)**	0.26 (5.6)**	0.09 (1.8)***
V	0.12 (6.5)**	0.09 (5.7)**	0.02 (1.0)***
Goal Diff.	0.47 (6.0)**	0.28 (3.6)**	0.19 (2.9)***
n	72	88	
N		5,382	
QEM 2 (CS + WL4M)			
P	0.35 (7.3)**	0.33 (4.6)**	0.02 (0.0)***
V	0.12 (6.5)**	0.12 (4.6)**	0.00 (0.0)***
Goal Diff.	0.47 (6.0)**	0.32 (2.5)**	0.15 (0.9)***
n	72	45	
N		3,920	
QEM 3 (CP)			
P	0.35 (7.3)**	-0.04 (1.1)	0.39 (41.8)
V	0.12 (6.5)**	-0.00 (0.2)	0.12 (28.1)
Goal Diff.	0.47 (6.0)**	-0.06 (1.1)	0.53 (31.1)
n	72	220	
N		9,805	
QEM 4 (CP + WL4M)			
P	0.35 (7.3)**	0.14 (2.7)**	0.21 (8.6)
V	0.12 (6.5)**	0.07 (3.8)**	0.05 (3.2)***
Goal Diff.	0.47 (6.0)**	0.12 (1.4)	0.35 (8.8)
n	72	118	
N		6,362	
QEM 5 (CS + CP)			
P	0.35 (7.3)**	0.28 (5.6)**	0.07 (0.9)***
V	0.12 (6.5)**	0.09 (4.9)**	0.03 (1.1)***
Goal Diff.	0.47 (6.0)**	0.35 (4.3)**	0.11 (1.0)***
n	72	77	
N		5,009	
QEM 6 (CS + CP + WL4M)			
P	0.35 (7.3)**	0.38 (5.7)**	-0.04 (0.2)***
V	0.12 (6.5)**	0.12 (5.0)**	0.00 (0.0)***
Goal Diff.	0.47 (6.0)**	0.49 (4.0)**	-0.02 (0.0)***
n	72	36	
N		3,631	
QEM 7 (CS + CP + WL4M (Opposite))			
P	0.35 (7.3)**	0.34 (5.2)**	0.01 (0.0)***
V	0.12 (6.5)**	0.12 (5.1)**	0.00 (0.0)***
Goal Diff.	0.47 (6.0)**	0.50 (4.3)**	-0.03 (0.0)***
n	72	38	
N		3,691	

Note: All estimates contain club-season fixed effects; absolute t-statistics based on robust standard errors in parentheses in columns Matched - Treatment & Matched - Control. F-statistics given in parentheses in column ATT. N = number of matches; n = number of seasons; P = points, V = victory, ATT = Average Effect of Treatment on the Treated, QEM = Quasi Exact Matching, CS: Cumulative Surprise, CP: Cumulative Points, WL4M: Wins in Last Four Matches.
** (*) indicates significance at 5% (10%)
*** indicates similarity of Treatment and Control coefficients at significance level 5%

treatment and control group are similar. For the goal difference, the counterfactual coach change effect is not significant, and therefore there is a significant ATT effect.

Next, we combine a cumulative point matching criterion (maximum difference of 5) with a cumulative surprise matching criterion (maximum difference of 0.2) in QEM 5. For all three performance indicators,

the counterfactual coach change effect is substantially positively significant, also resulting in significant similarity between the effect of the treatment group and control group. There is no significant ATT here. Adding an exact matching criterion on WL4M to this (QEM 6) gives even more similar coach change effects, resulting in F-statistics of 0.2, 0.0 and 0.0. Undoubtedly, with this precise matching criteria, the treatment group looks characteristically the same as the control group, and has non-coincidentally almost the exact same effects. With those results, it can be concluded that the effect of a coach change, the ATT, is zero, and that the positive effect we found in the naive regression can be attributed to the regression-to-the-mean phenomenon. The club would have performed better, regardless of whether a coach change happened.

To explore whether it is possible to make a trade-off between the strictness in two performance indicators matching, the cumulative points criterion is made more strict (from max 5 to max 2 difference) and the cumulative surprise criterion is made more lenient (from max 0.2 to max 0.5 difference). QEM 7, with approximately the same number of club seasons in the control group as QEM 6, shows the results. The effects seem very similar to the effects found for QEM 6, although the counterfactual coach change effect for points decreased a bit, and for goal difference increased a bit. The ATTs are close to 0.00 and the F-statistics are all 0.0. It is safe to say that it is possible to make a trade-off in strictness between criteria, while keeping the results the same.

When adding an additional matching criterion, namely the personal coach characteristic age with a maximum difference of 3 to any mix, the counterfactual coach change effects remain exactly the same. This confirms the results shown in Table 11; Age does not have any impact on the probability and effect of a coach change. This is in line with what van Ours & van Tuijl (2016) say, namely that the age of a coach is not a significant factor in coach dismissal.

Summarizing the patterns found while using the seven different QEMs, it seems that cumulative surprise as sole matching criterion works well, as opposed to cumulative points that does not work well alone. Adding WL4M as extra matching criterion increases the similarity of the effects between the actual and counterfactual coach changes. Using a combination of the three different performance indicators as matching criteria gives the most likely effects, with an insignificant ATT. The positive effect we found in the naive regression can be attributed to the regression-to-the-mean phenomenon. Furthermore, a trade-off in strictness gives the same results and the personal coach characteristic age does not influence the effects in any way.

6 Conclusion

In this paper, the effect of coach changes, and in particular PR coach changes, in the Dutch Eredivisie in season 2000/2001 till 2021/2022 is analysed. The data used consists of all the regular Eredivisie matches played, all the corresponding bookmaker odds, to form cumulative surprise, and the coaches that managed the clubs in this time period. First, by analysing the data itself it becomes apparent that there is a connection between forced and unforced leaves, and PR and NPR coach changes. The same can

be said for relegations and managerial changes. Also, it seems that while NPR and quit coach changes club seasons have a more similar cumulative surprise and cumulative number of points as the no coach changes club seasons, the other club season types (all coach changes, coach dismissals and PR coach changes) have a lower cumulative surprise and cumulative number of points.

Secondly, I investigate the naive effect of a coach effect on three performance indicators; points, victory, and goal difference. For all coach changes and coach dismissals, this effect is substantially positive, although somewhat bigger for coach dismissals, but seems to have decreased for both with around 33% in the past decade compared to the decade before. Performing this naive regression for PR coach changes shows even more extreme effects, while NPR and quit coach changes have an insignificant effect on the three performance indicators. Evidence shows that focusing on PR coach changes instead of all coach changes or coach dismissals is more logical and gives larger effects.

Thirdly, to analyse the actual effect of a coach change, there has to be a control group which consists of counterfactual coach changes. While using the PR coach changes as treatment group, seven different control groups are formed using seven QEMs based on three performance criteria: cumulative surprise, cumulative points, and WL4M. Those three performance indicators have a proven significant effect on PR coach changes. By analysing the counterfactual PR coach change effects and its similarity with the actual PR coach changes, we obtain insights into the actual effect of a coach change, shown by the ATT and its significance. It seems that cumulative surprise as sole matching criterion works well, as opposed to cumulative points that does not work well alone. Adding WL4M as extra matching criterion increases the similarity of the effects between the actual and counterfactual coach changes. Using a combination of the three different performance indicators as matching criteria gives the most likely effects, with an insignificant ATT. Furthermore, a trade-off in strictness gives the same results and the personal coach characteristic age does not influence the effects in any way.

It seems that like most of the literature on this topic performed before, there is no real positive effect from a coach change. There is a positive effect on the PR treatment group, but this positive effect can also be found for the counterfactual PR coach changes based on a fitting QEM, which means that the effect can be attributed to the regression-to-the-mean phenomenon. The club would have performed better, regardless of whether a coach change happened.

There are limitations to my analysis, which can be solved in order for more accurate and complete results. More data can be used and explored to assess the consistency of the results obtained. For example, it would be interesting to see if the coach change effects found are also present in other European leagues, as well as other continental leagues. Furthermore, my data sample only used the data from season 2000/2001 onwards, while although harder to find, earlier data should be available too.

One specific limitation to the analysis performed in this paper is the constraint of having only a maximum of one control member per club season. This has advantages, like no control group members that are extremely similar, while also having disadvantages like less possible data. It would be interesting to see how the results and their significance would change, when this constraint is lifted. More members

in the control group would be present, which could allow for more strict matching criteria. Although already using seven different QEMs, more matching combinations and performance criteria could lead to better insights on the patterns. Other performance indicators that may be interesting to look at are position in the league, cumulative points and surprise in only the last few matches, whether a club is in the relegation zone, position change, and if losing the last match has influence.

Conclusions are drawn in this paper on the effect of a in-season coach change. As mentioned before in the introduction, the soccer world might be representative for the business world. However, the question still remains what the effect of a CEO change would be. To draw conclusions on those effects, although less frequent data exists on that topic, the study performed here could be performed on CEO change data.

References

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, *4*(3), 290–311.

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267.

Acquah, K. (2020). *Potential outcomes model.* Causal Flows. https://www.causalflows.com/potential-outcomes-model/.

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

Arnulf, J. K., Mathisen, J. E., & Hærem, T. (2012). Heroic leadership illusions in football teams: Rationality, decision making and noise-signal ratio in the firing of football managers. *Leadership*, *8*(2), 169–185.

Arrondel, L., Duhautois, R., & Zimmer, C. (2020). *Within-season dismissals of football managers: evidence from the French Ligue 1* (Tech. Rep.). HAL.

Audas, R., Dobson, S., & Goddard, J. (1999). Organizational performance and managerial turnover. *Managerial and Decision Economics*, *20*(6), 305–318.

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.

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.

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.

Balduck, A.-L., Prinzie, A., & Buelens, M. (2010). The effectiveness of coach turnover and the effect on home team advantage, team quality and team ranking. *Journal of Applied Statistics*, *37*(4), 679–689.

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.

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

Dobson, S., Goddard, J. A., & Dobson, S. (2001). *The economics of football* (Vol. 10). Cambridge University Press Cambridge.

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.

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

Gebala, S. (2022). *Football dataset +96k matches (18 leagues).* Kaggle. https://www.kaggle.com/datasets/bastekforever/complete-football-data-89000-matches-18-leagues?resource=download.

Hentchel, S., Muehlheusser, G., & Sliwka, D. (2012). The impact of managerial change on performance: the role of team heterogeneity. *CES-ifo Working Paper*, *3950*.

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

Koning, R. H. (2003). An econometric evaluation of the effect of firing a coach on team performance. *Applied Economics*, *35*(5), 555–564.

Kuper, S., & Szymanski, S. (2018). *Soccernomics: Why England loses, why Germany and Brazil win, and why the US, Japan, Australia, Turkey–and even Iraq–are destined to become the kings of the world's most popular sport.* Hachette UK.

Narita, K., Tena, J., & Buraimo, B. (2022). Causal and consequences of multiple dismissals: E evidence from Italian football league. *University of Liverpool*.

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.

Peeters, T., & van Ours, J. (2021). Seasonal home advantage in English professional football; 1974–2018. *De Economist*, *169*, 107–126.

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

Poulsen, R. (2000). Should he stay or should he go? Estimating the effect of firing the manager in soccer. *Chance*, *13*(2), 29–32.

Salomo, S., & Teichmann, K. (2000). The relationship of performance and managerial succession in the German premier football league. *European Journal for Sport Management (EJSM)*, *7*, 99–199.

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

Scelles, N., & Llorca, M. (2021). Leader dismissal or continuity, president longevity, geographic orientation of owners and team performance: insights from French men's football, 1994–2016. *Journal of Risk and Financial Management*, *14*(9), 439.

Szymanski, S. (2003). The assessment: the economics of sport. *Oxford Review of Economic Policy*, *19*(4), 467–477.

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

ter Weel, B. (2011). Does manager turnover improve firm performance? Evidence from Dutch soccer, 1986–2004. *De Economist*, *159*(3), 279-303.

TransferMarkt. (2023). *Managerial changes during the season.* https://www.transfermarkt.com/eredivisie/trainerwechselprosaison/wettbewerb/NL1/plus/1.

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.

Voetbal.com. (2023). *Trainerhistorie.* https://www.voetbal.com/teams/de-graafschap/9/.

VoetbalInternational. (2023). *Eredivisie 2021/2022.* https://www.vi.nl/competities/nederland/eredivisie/2021-2022/stand.

Westerberg, M., Singh, J., & Häckner, E. (1997). Does the CEO matter? An empirical study of small Swedish firms operating in turbulent environments. *Scandinavian Journal of Management*, *13*(3), 251–270.