Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis International Econometrics and Operations Research

Performance of football teams after a coaching change and determinants of head coach dismissals. A comparative analysis of Italy and The Netherlands.

Valts Nikitins (545239)

Supervisor:	prof.dr. Dennis Fok
Second assessor:	prof.dr.ir. Jan van Ours
Date final version:	2nd July 2023

Abstract

Changing of head coaches in football is a regularly observed occurrence in professional football. While it is common that coaches are changed amid spells of bad performance, some differences can be observed between countries. Head coaches in some leagues are changed more frequently compared to others. In this paper we compare a country with high managerial turnover (Italy) to one with a comparatively low (The Netherlands) one. We examine whether there are differences in effects of such change on performance and what are the most significant determinants of coaching change in football. For both countries we use the data on match results, bookmaker odds and coach characteristics in the time period 2000-2022. It is found that while performance in both countries tends to improve after a coaching change, no causal effect can be established. Furthermore, there is some indication that teams in Italy perform better if they opt not to change a coach after bad spells. Additionally, the results show that performance based indicators are the most significant determinants of dismissals in both countries, while foreign coaches are more likely to be dismissed in Italy.

1 Introduction

The most recent football season (2022/2023) has seen 14 different head-coaching changes (without taking caretaker managers into account) in the English Premier League, which is an unprecedented number, according to transfermarkt.com. However, that is not only the case with the English football, as coach dismissals are quite a common occurrence in every professional football league. Such changes are usually made to improve the performance of a team. (Paola & Scoppa, 2012) Even if coaching changes are made in all football leagues, some differences are still observed among different countries. More specifically, in some leagues coaches are changed more often than in others.

This leads us to our research question - *"How does a team performance after football head-coach dismissals differ between more volatile and less volatile environments?"* We aim to investigate whether the improvements in team performance after a coaching change differ when comparing countries, where coaching changes are made often to ones where such changes occur less. This research might be relevant to the decision makers in football clubs as parting ways with coaches may involve costs of terminating a contract. Moreover, such changes run a risk of disrupting the team. (de Dios Tena & Forrest, 2007; Audas, Dobson & Goddard, 2002) Furthermore, as shown by Pieper, Nüesch and Franck (2014), there are some similarities between football head coaches and top managers. As explained in Pieper et al. (2014), both football head coaches and top managers adhere to some expectations, which can affect their tenure at a specific job. Additionally, a manager may be changed after a poor performance of a firm, which also can happen with football coaches. Therefore, the findings of this research could be extended to manager turnover in other sectors. Studying performance improvements in football is advantageous as performance in sports is easier to quantify. (Ter Weel, 2011)

Table 1: In-season head-coach changes in 7 European football leagues from 2000/2001 to 2021/2022

League	Mean	Maximum	Minimum
Great Britain (Premier League)	6.70	14	3
Spain (La Liga)	8.35	13	4
Italy (Serie A)	9.53	16	5
France (Ligue 1)	5.76	13	3
Germany (Bundesliga)	7.11	12	4
Netherlands (Eredivisie)	4.58	7	2
Belgium (Jupiler Pro League)	7.85	13	3

Table 1 shows the summary statistics of coaching changes in 7 different European professional football leagues during the time period of 2000-2022. As can be seen, the highest football league in The Netherlands (further referred to as Eredivisie) on average sees 4.58 coaching changes during the season. Moreover, the maximum and minimum amount of coaching changes during the season are also the lowest amongst considered leagues. Such a result shows that amongst professional football leagues, Eredivisie can be considered as an environment of low managerial turnover. On the other hand, a league with the highest amount of coaching changes according to the summary statistics is Serie A in Italy with 9.53 coaching changes on average during the season. That is more than twice as much compared to Eredivisie. Therefore, in further analysis we consider Italy as a place with more volatile work environment or in other words it is an environment with high managerial turnover. However, it should be noted that the reported amount of changes is the total amount. This means that it does not account for differences in the amount of teams competing in each league. In the considered time period there have been 20 teams in Premier League, La Liga, Ligue 1 and for most of seasons in Serie A, while there are 18 in Bundesliga, Eredivisie and Jupiler Pro League. Even when taking into account the differences in the amount of teams competing, we still find The Netherlands and Italy as places with the lowest and highest managerial turnover. Therefore, in the following paper we will examine the differences between The Netherlands and Italy.

As shown in Table 1, coaching changes are inevitable in professional football. Therefore, in addition to examining whether the performance improves, we also distinguish between two different types of coaches that could be hired after a change is made. The first type is a coach, which is more experienced than the previous one and the second one is a coach, who is less experienced. Such a distinction is made in order to study whether the experience of a newly appointed coach plays a role in the performance of a team after a change. This leads to our first sub-question - *"Does replacing a less experienced coach with a more experienced one have a significant impact on team performance?"*. To answer this sub-question we analyze both countries together.

To further investigate the differences between the way coaching changes happen we form a sub-question - *"How do the determinants of in-season head coach changes differ between more and less volatile environments?"*. In this thesis we perform a similar analysis as in Van Ours and Van Tuijl (2016), where determinants of head coach dismissals are identified and research is performed on whether the performance of the teams improves after such changes. We extend their analysis by analyzing the differences between two countries and studying the effects of hiring a more experienced coach during the season.

Previous research mostly shows that there is no statistical evidence that football teams perform significantly better after appointing a new head coach. Van Ours and Van Tuijl (2016) show that when using a naive approach (not taking control group into account), a coaching change has a significant effect on team performance. However, their results also provide evidence that teams start performing better even if no change is made. Koning (2003) finds that performance of a team does not always improve after a change and in some cases a team performs even worse after appointing a new coach. Ter Weel (2011) report similar results, that when considering control groups, the effects of a coaching change are not significant. Regarding Italy, Paola and Scoppa (2012) show that a head coach firing has no significant effect on team performance. A paper by Narita, Tena and Buraimo (2022) examines the effects of multiple head-coach dismissals in a season. They determine that the first coaching change comes with some improvements in performance, while the second one does not seem to be reaping benefits. On the other hand, Bryson, Buraimo, Farnell and Simmons (2021) show that head coach quitting does not affect the team's performance, while a team dismissing its coach gives a small but statistically significant improvement in results when considering 8 different leagues in their analysis. Furthermore, Zart and Güllich (2022) find that a coaching change improves performance significantly in Bundesliga, Premier League and La Liga, however, the control group is not accounted for. When looking at previous research, it would be expected that in both The Netherlands and Italy coaching change has no effect on team performance.

When looking at the business sector, Huson, Malatesta and Parrino (2004) find that there are abnormal positive stock returns after an announcement of managerial change. It is found that performance of firms improves after managerial turnover. However, as noted by Ter Weel (2011), it is unclear whether this improvement is due to market beliefs or due to actual improvements in performance. Watrous, Huffman and Pritchard (2006) show that high management turnover coincides with a lower improvement of performance, although, no causal relationship is established.

Regarding the determinants of coaching changes, Van Ours and Van Tuijl (2016) find that in Eredivisie only performance-based measures have a significant effect on the dismissal rates, however, none of the considered regressors have significant effects on quit rate. Similar results in their research hold if a tenure of a coach instead of in-season duration is considered. In Italy Porro, Restaino, Ruiz-Castro and Zenga (2021) find that the performance of a team affects the rate at which coaches are dismissed, while coach-specific characteristics have no significant impact on the dismissal rate. Audas, Dobson and Goddard (1999) find that how well a team performs is a crucial determinant of dismissal rate in the English Premier League. Similarly, they find that a quit rate of coaches is affected by performance as well, which is not observed when studying the data on Dutch Eredivisie. Frick, Barros and Prinz (2010) find that experience of a coach has a negative effect on both the quit and dismissal rate. It would be expected for team performance prior to dismissal to be an important contributor to the hazard rate.

Pierce, Johnson, Krohn and Judge (2017) find that in NCAA women's basketball characteristics of a coach, such as past experience, have no significant effects on the performance of a team. Moreover, findings of White, Persad and Gee (2007) indicate that in the National Hockey League (NHL) prior experience of a coach does not explain team performance, when looking at mid-season changes. On the other hand, Pfeffer and Davis-Blake (1986) show that in the National Basketball Association (NBA) prior experience of a coach improves performance of teams. It should be noted that such differences in findings could occur due to the fact that different leagues and even different sports are considered.

This paper contributes to the existing literature by studying more recent data. For example, the research by Van Ours and Van Tuijl (2016) and Bryson et al. (2021) consider data up until the 2013-2014 season, while we add the most recent data up until 2021-2022. Moreover, in comparison to existing literature, this thesis will aim to consider the differences between two different countries instead of focusing on a specific one or adding the data from different countries together. Furthermore, we will do further analysis using bookmaker odds as done by Van Ours and Van Tuijl (2016) and Stadtmann (2003). Moreover, we investigate whether there are benefits of hiring a more experienced coach during the season in The Netherlands and Italy. Finally, we distinguish between more and less volatile work environments, which can provide valid insights into other areas outside of football. As stated by Kahn (2000) and Pieper et al. (2014), sports offers opportunities to research the labor market.

The data set required to perform this analysis contains data on all games during the seasons from 2000-2001 until 2021-2022 in both Eredivisie and Serie A. Data is collected on match results, bookmaker odds and different coach characteristics, such as age, nationality and whether the coach has played in the respective national team. In addition to the collected data we calculate performance measures, such as points earned in a game, goal difference etc.

We follow a largely similar methodology as described in Van Ours and Van Tuijl (2016). When estimating models of team performance after a change a simple linear model is employed, which is also in line with Van Ours and Van Tuijl (2016). The same performance indicators (points, win and goal difference) are used, while also adding whether a team outperforms their expectations. To establish whether there is a causal relationship between coaching change and improvement in team performance, we implement a matching algorithm, where we try to find cases, where team was performing similarly, but no coaching change happened. Furthermore, we compare the effects of actual and counterfactual coaching changes. For the determinants of coach dismissals, an exponential hazard rate model is employed. However, this model assumes a constant baseline hazard rate, which might be too restrictive. Therefore, we also consider the semi-parametric Cox proportional hazard model proposed by Cox (1972). Two types of duration models are considered - in-season and coaching spell duration models. To find whether a team performs better when hiring a more experienced coach during the season we also differentiate between cases where a more experienced was hired and where it did not happen. For linear models, the equality of coefficients is tested by the means of an F-test.

In our analysis it is found that in both The Netherlands and Italy the performance of a team improves after a coaching change. Moreover, the results show that the difference in effects of a coaching change between countries is not statistically significant. Furthermore, when analyzing the differences between actual and counterfactual coaching changes, we find that in both countries teams tend to perform better even if a coaching change does not happen. In the case of Italy, the results show that for some performance measures a team tends to perform significantly better if no coaching change happens. After estimating duration models, it is found that coaches in both countries are less likely to be fired if a team performs well in recent games and outperforms their expectations as based on bookmaker odds. In Italy we find some evidence that foreign coaches are more likely to be fired than local ones. Moreover, older coaches are more likely to be fired in The Netherlands while an opposite is seen in Italy. Results also show that performance indicators in most cases have no significant effect on quit rate. Finally, when considering the effect of hiring a more experienced coach, it is found that in the joint sample there are no statistical differences in improvement of performance. However, when considering The Netherlands individually, it is found that teams tend to perform better if a more experienced coach is hired mid-season compared to a hiring of a less experienced manager.

The setup of this paper is as follows. In Section 2 we explain further where the data is obtained and how it is processed. Section 3 explains and motivates the methods that are used. In Section 4 we report and discuss the obtained results and in Section 5 we give a conclusion of the most important results, mention some limitations and give some suggestions for future research.

2 Data

Obtaining a data set and processing it in a meaningful way is an integral part of this research. The data set on Eredivisie for seasons 2000-2001 until 2013-2014 was provided by the authors of Van Ours and Van Tuijl (2016). Therefore, to perform our analysis we construct a data set in a similar fashion as the one provided for the most recent seasons of Eredivisie and the whole sample of Serie A.

The data set consists of each game played in each of the tournaments for every season. Each match is reported twice as to get the perspective of both teams in a game. For each entry it contains certain variables. Firstly, the round of a game, team from whose perspective this entry is constructed and its opponent in the match. Additionally, there is a binary variable which indicates whether the game is played as home, goals scored by both teams and the position of the opponent in the last season. Regarding the coaches, the name of the coach is reported together with his age. We also report whether there is a coaching change in that season for that club, the nationality of this coach, a binary variable which indicates whether this coach has played in his national team and, additionally, the number of appearances for the national team. For each entry we also report the bookmaker odds, which are the odds of a home win, a draw and an away win. These odds are displayed in a decimal format. The bookmaker odds are mostly from William Hill and for the games where they are not available, we use the ones from bet365 and if those are not reported, the ones from StanleyBet are used. However, it is noted that bookmaker odds should be quite similar between different sources. (Van Ours & Van Tuijl, 2016) There were 2 games in The Netherlands for which no bookmaker odds were found and 7 such instances in Italy. In that case for both entries of the game in the data, we set odds to 0. Finally, for each of the individual seasons of each team we assign an ID. An overview of used variables and their description is provided in Table 9 in Appendix A.

The data on matches - such as results and bookmaker odds are obtained via football-data.co.uk. The position of each team at the end of a season is found using transfermarkt.co.uk. To find data on managers we manually go through the information on transfermarkt.co.uk, where we go through the seasons of each team to see, who was the coach, whether there was a coaching change and if yes, then when did it occur. Additionally, we use the same source to find the previously mentioned characteristics of each coach. For each season where there was a coaching change, we define whether the coach was fired or quit. To find that, we searched for announcements of each team or relevant news articles. If the announcement mentions that there has been a mutual agreement to terminate the contract, we assume it to be a quit. As the research is mostly focused on team performance after coaching changes, if there is a caretaker manager, we assume that the change occurred when the newly appointed manager has his first match and we assume that the matches played under the caretaker manager are trained by the previous coach. If there is more than one coaching change in a season, we set the value of type of dismissal to one which happened first. For example, if one team had 3 coaches in a season - the first one quit and the second one was dismissed - we assume that an "unforced leave" happened in this season. Additionally, if multiple coaching changes occur in a season, only the first one is taken into account, when performing analysis.

Furthermore, we manually process the data to put everything in the correct order and to report the game from the perspective of both teams as explained previously. Also, we give a unique ID for each individual season of every team as done in Van Ours and Van Tuijl (2016). Furthermore, we find the number of points earned by the team, which are 3 for a win, 1 for a draw and 0 for a loss. We also set up a binary variable, which takes the value of 1 if the match was won and 0 otherwise. Moreover, we find the goal difference and the points obtained in the last 4, 5 and 6 games. An integral part of our research and its contribution is incorporating bookmaker odds to quantify expectations before the game. For each match we find the match surprise, which is the difference between expected and actual points (Stadtmann, 2003). It is calculated as explained in Stadtmann (2003). Firstly, we find the bookmaker markup by summing up the inverses of home win, draw and away win odds. After that, to find the probability of each outcome, we multiply the inverse of outcome odds with the inverse of bookmaker markup. Using the estimated probabilities and the points obtained with each outcome we can easily compute the expected points. Additionally, we compute the cumulative surprise by summing all of the previous surprises in the season. If the bookmaker odds are not available for the match, we set the surprise to 0.

The data is reported in separate data sets for each of the countries. However, using R it is quite straightforward to combine data together. Finally, we note some differences between both countries. In Eredivisie there have been 18 teams throughout the whole sample, which means that for each team there are 34 games played in a season. It is the case for every season considered, except for 2019-2020, where the season in The Netherlands was abandoned due to Covid-19 pandemic. 25 or 26 games were played by each team out of the possible 34. Regarding Italy, there were 18 teams competing in the competition until 2003-2004 season, which implies 34 games for each team. However, starting from 2004-2005 season, there have been 20 teams in Serie A and, therefore, 38 games for each team in a season. The amount of games in 2019-2020 season did not change due to Covid-19. 29 different teams have competed in Eredivisie in the given time period, while 47 teams have been in Serie A. Some summary statistics for variables considered in the analysis are shown in Table 10 in Appendix A.

3 Methodology

3.1 Impact of coaching change on team performance

3.1.1 Naive approach

To find the effect of a change of coach on team performance, we estimate the following simple linear model as in Van Ours and Van Tuijl (2016)

$$y_{i,j,k} = \alpha_{i,j} + \beta_1 H_{i,j,k} + \beta_2 P_{i,j,k} + \beta_3 C_{i,j,k} + \varepsilon_{i,j,k}, \tag{1}$$

where i denotes the team, j is the season, k is the match. $y_{i,j,k}$ denotes the performance indicator of team i in the season j in match k. Contrary to Van Ours and Van Tuijl (2016) we consider match surprise as an indicator of performance in addition to points, whether the match is won and the goal difference. Match surprise could be correlated with the regressors, as bookmaker odds take all of the possible predictors of performance into account, but we include it as actual points obtained in a game is still random. $\alpha_{i,j}$ denotes the fixed effects of team i in season j to account for heterogeneity between different teams, different seasons and countries. $H_{i,j,k}$ represents the binary variable on whether team i in season j played the match k at home. It is included in the regression as football teams tend to perform better when playing at home, as found by Carmichael and Thomas (2005). $P_{i,j,k}$ is a variable that informs on what position did the opponent of team i in season j and match k take the previous season. It is a variable indicating the strength of an opponent and, therefore, important when explaining performance. $C_{i,j,k}$ is a binary variable on whether there was a coaching change for team i in season j before match k. Therefore, parameter β_3 is of utmost importance to us as it reports on what is the effect of a coaching change on different performance indicators.

When comparing both countries we allow for differences in coefficients β_1 and β_2, as the effects of home advantage and differences in quality between higher and lower-ranked teams may differ between countries. Additionally, we use White's estimator for residual covariance matrix to account for possible heteroskedasticity. As this analysis is used to find the effects on team performance for in-season head coach changes, we only consider the seasons where such changes happened, also we only take into account the changes that happened after the fifth game of the season or at least 5 games before the end of the season. When performing this analysis, we distinguish between two separate cases. Firstly, all seasons where there was a coaching change and, secondly, only the seasons where the head coach was dismissed. Furthermore, an F-test is used to test whether the effects of coaching change on team performance are equal in The Netherlands and Italy. If the null hypothesis of equal coefficients is not rejected, then we can say that the effect of a coaching change in both countries is equal.

3.1.2 Matching treatment and control groups

However, even if all parameter estimates for the regression in Equation 4 are significantly different from 0, we cannot conclude on causal effects as we do not take into account the cases where a coach could have been changed, but was not, as explained in Van Ours and Van Tuijl (2016). To tackle this problem, we match actual coaching changes to a control group, which are individual seasons where head coach could have been dismissed, but was not.

To find the control group, for each of the coaching changes that have occurred we find the cumulative surprise in the last match of the coach that was dismissed. After finding that we search through all of the other seasons of the same team to find the closest match in terms of a cumulative surprise. After that this particular matched instance is called a counterfactual change. In this analysis we follow the research of Van Ours and Van Tuijl (2016) and set the maximum difference between a cumulative surprise of an actual and a counterfactual coaching change to 0.5. As a robustness check, we also consider other values for maximum difference - 0.25, 0.75, 1. The actual coaching changes that are matched to a counterfactual coaching change are called the treatment group, while the seasons of matched counterfactual coaching changes are called the control group. In contradiction to Van Ours and Van Tuijl (2016), we include all matched seasons in the control group and not only the unique instances. In other words, if one season is included in the control group more than once, then it is also included in further analysis more than once.

It should be noted that the counterfactual coaching changes are searched between the fifth and fifth to last game of the season to match the actual coaching changes in line with the actual coaching changes considered. These same steps are also performed strictly for head coach dismissals in which case we search for a counterfactual coach change in a different season, where a coach was not fired instead of in seasons where no changes happened. After matching the seasons, we estimate the model as in Equation 4 with adding an additional regressor of a counterfactual coaching change. That way we assure that the effects of home advantage and opponent strength stay constant in both cases - actual and counterfactual coaching changes. If the difference between these effects is not statistically significant, then it can be concluded that a coaching change has no effect on team performance and a team would improve its performance regardless of it.

3.2 Duration models of coach dismissals

Furthermore, we estimate duration models to find what are the possible determinants of coaching changes. In line with Van Ours and Van Tuijl (2016). We differentiate between two possible coaching changes - dismissals and quits.

We perform the analysis separately for both countries to find whether these determinants differ. Similarly as in Van Ours and Van Tuijl (2016) we consider two different types of covariates - time-varying and fixed. Time-varying covariates are variables characterising the performance of a team during the season, such as points in the last 4 games and cumulative surprise. Fixed regressors are variables concerning coach characteristics, such as age of the coach (divided by 10 in the analysis), whether he played in the national team and whether he is a foreigner in his respective league.[1] In our research we first consider the extended model with all previously mentioned possible determinants. Secondly, as done in Van Ours and Van Tuijl (2016) we consider a baseline model with using both time-varying covariates and the age of the coach as regressors. Thirdly, as a sensitivity analysis, we estimate the baseline model using points in last 5 and 6 games instead of points in last 4 games as a regressor. Duration models are estimated using Maximum likelihood estimation.

3.2.1 In-season duration

Similarly as in Van Ours and Van Tuijl (2016), firstly, we look at the in-season duration, where we look at the number of matches until a coach is dismissed in a particular season. We consider two different proportional hazard models in this analysis. The first one is a model used in Van Ours and Van Tuijl (2016), which assumes an exponential distribution for the baseline hazard, with the hazard rate defined as

$$\lambda_j(t) = \lambda_j \exp\left(x'\beta_j + z_t'\gamma_j\right), \tag{2}$$

[1] Contrary to the analysis of Van Ours and Van Tuijl (2016), the variables concerning experience are not considered in this particular analysis as those were not collected due to time constraints. However, an age of a coach could be a decent indicator of performance and in addition to possible determinants considered in Van Ours and Van Tuijl (2016) another binary variable, of whether the coach is foreign, is used. Gilfix, Meyerson and Addona (2020) have shown that, for example, in Major League Soccer (MLS) foreign coaches tend to be changed earlier.

where $j = 1, 2$ indicates the type of coach dismissal, with 1 referring to a dismissal and 2 referring to a quit. x are the fixed covariates and z_t are time-varying covariates. The hazard rate $\lambda_j(t)$ could be interpreted as a probability that a coaching change of type j happens at time t given that it has not happened before time t. However, this model has a limitation, namely the baseline hazard function is constant, which might be too restrictive. Therefore, we also consider a Cox proportional hazard model, proposed by Cox (1972), which allows for a change in baseline hazard function over time. This model has also been previously used in similar analysis. (Porro et al., 2021; Semmelroth, 2022; Gilfix et al., 2020). For the Cox proportional hazard rate model the hazard rate is defined as

$$\lambda_j(t) = \lambda_{0,j}(t) \exp\left(x'\beta_j + z_t'\gamma_j\right), \tag{3}$$

where $\lambda_{0,j}(t)$ is an unknown baseline hazard function for type of dismissal j. It should be noted that Cox proportional hazard model is semi-parametric, which in this case means that no assumptions are made regarding the distribution of baseline hazard function. However, in this analysis we make a few assumptions about the changes themselves. Firstly, in line with Van Ours and Van Tuijl (2016) every season, where a coach was still in his post at the last game of the season, is right-censored. Also coaching changes before the 5th or after 5th to last game of the season are right-censored. Additionally, when looking at rate of firing, all cases where coach quit, are right-censored and vice-versa.

3.2.2 Coaching-spell duration

When we use in-season duration models, a crucial assumption is made - each season starts without any history. (Van Ours & Van Tuijl, 2016) This assumption might be too strong as coaches develop relationships with the board, owners and players and it would be naive to assume that all of it resets at the start of a new season. Therefore, the same as in Van Ours and Van Tuijl (2016) we also estimate the coach-spell duration models. In this case we look at the whole tenure of a coaches spell in a club. We use the same models as for in-season duration models. Additionally, for the extended model we include two binary variables *second_season* and *third_season*, which take values 1 if it is the second or third season of coaches tenure respectively. It should be noted that variables *second_season* and *third_season* are not included when estimating the Cox proportional hazard model, as the effect of variables that only depend on time is already included in the baseline hazard function $\lambda_{0,j}(t)$.

For the coaching-spell duration models some of our assumptions differ compared to in-season duration models. Firstly, coaching changes before the fifth or after fifth to last game of the season are considered to be right-censored only if a change occurs in the first season of the tenure. However, when estimating quit rate a dismissal is considered to be right-censored and vice versa, as for in-season duration models. Secondly, we assume that all coaching spells start at the time a team is first featured in the data. This means that there are no left-censored observations and spells which have started before 2000-2001 are assumed to start at 2000-2001 for the purposes of this research. Thirdly, if a coach is changed between seasons, this change is assumed to be a quit. Fourthly, we only consider spells where a team is in the highest football division in the respective country. For example, if a coach has trained a team for a season, his team is relegated and then promoted the next season with the same coach, his tenure is considered to continue as if the team was not relegated - the seasons in the lower are leagues ignored. Moreover, all cases, where a team no longer appears in the sample and the coach was not dismissed in the last season, are considered right censored. Finally, all coaching spells are assumed to start at the start of a season (Gameweek 1). This means that if a coaching change occurs during the season, all remaining matches in the respective season are ignored and the tenure of the new coach is assumed to start at the beginning of next season.

3.3 Impact of hiring a more experienced head coach during the season

Finally we perform an analysis to find whether replacing a current coach with a more experienced coach is beneficial compared to hiring a less experienced one. In order to do that we estimate the following linear model

$$y_{i,j,k} = \alpha_{i,j} + \beta_1 H_{i,j,k} + \beta_2 P_{i,j,k} + \beta_3 C_{more,i,j,k} + \beta_4 C_{less,i,j,k} + \varepsilon_{i,j,k}, \tag{4}$$

where $y_{i,j,k}, H_{i,j,k}, P_{i,j,k}$ are the same regressors as in Equation 4. $C_{more,i,j,k}$ is a binary variable, which takes value 1 if for team i in season j prior to match k, there was a coaching change and a more experienced coach than the last one was appointed. $C_{less,i,j,k}$ represents the case were there was a change, but the newly appointed coach is not more experienced than the previous one. In this analysis we consider age of the coach as an indicator of his experience - an older coach is considered more experienced.

For this research we consider the combined sample of The Netherlands and Italy. Additionally, we perform this analysis separately for both countries to investigate whether the conclusions differ. A similar analysis is done as for the model in Equation 4. After finding which type of coaching change provides the highest improvement to performance, an F-test is performed for the equality of coefficients β_3 and β_4 to find whether this difference is statistically significant.

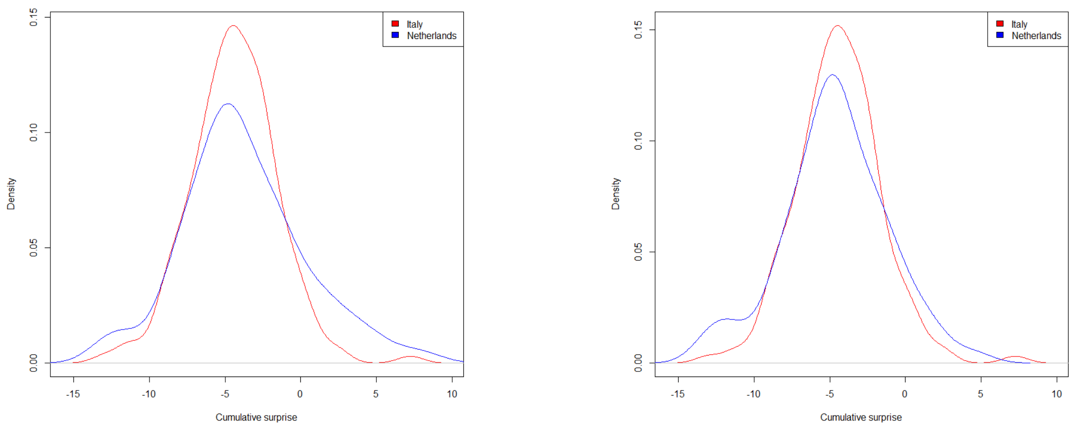
4 Results

In this section we report the results. Firstly, we look at the effects of coaching change on team performance by considering results for the naive approach model, which estimates whether teams perform better after a coaching change. Secondly, we compare the effects of a coaching change in both the cases where an actual coaching change happened and where it could have happened based on in-season performance. Furthermore, we look at duration models to find what are the most significant determinants of coach dismissals and quits. Models for in-season and coaching spell hazard rates are considered. Finally, we look at the effect of replacing a head coach with a more experienced one during the season.

4.1 Impact of coaching change on team performance

4.1.1 Naive approach

In Figure 1 we can see the kernel densities of cumulative surprise in the last game before a coaching change. Plots show that, when considering all changes, coaches in both The Netherlands and Italy tend to change when the value of cumulative surprise is around -5. It can be seen that kernel densities of all coaching changes have ”fatter tails” in the case of The Netherlands compared to Italy. That could indicate that in The Netherlands coaches continue working even if a season is not going as well as expected and they are given more time to improve team performance. On the other hand, the density plot in Figure 1a shows that, also, proportionally more coaching changes occur in The Netherlands when a cumulative surprise is positive. That could be explained by the fact that there are determinants other than performance of a team that would indicate a coaching change or that coaches themselves have higher expectations of results the team should achieve. This difference is not as pronounced in Figure 1b, which shows that the latter is more likely.



(a) All coaching changes

(b) Coach dismissals

Figure 1: Kernel densities of the last game of a coach before a change in The Netherlands and Italy (2000-2022).

In Table 2 and Table 3 we see the parameter estimates for the seasons where coaching changes happened for both Eredivisie and Serie A. In total there were 97 coaching changes in The Netherlands and 162 in Italy. Regarding dismissals, there were 71 in The Netherlands and 156 in Italy. This would explain why parameter estimates for Italy are quite similar when comparing the results in Table 2 and Table 3, as the sample is similar in both cases. We also see that all parameter estimates are positive when considering *Points*, *Wins* and *Goal difference* as dependent variables. That is in line with previous analysis by Van Ours and Van Tuijl (2016). However, when we look at the parameter estimates for *Surprise* as a dependent variable in both Table 2 and Table 3, we see that coefficients for *Home* and *Pos Opponent* are statistically insignificant for Italy. As explained before in Section 3 that is to be expected due to the fact that bookmakers would take the effect of home advantage and facing a ”weaker” opponent into account when setting the odds. These coefficients are also quite small for The Netherlands, but still significantly higher than 0. When looking at both countries, we see that the estimated standard errors are smaller for Italy which is to be expected due to the fact that the sample for Italy is larger.

Furthermore, to explain the interpretation of coefficients, we look at the case of The Netherlands in Table 2 which considers all seasons where any coaching changes happened. All things kept equal, when playing at home teams earn 0.53 points more, its probability of winning increases by 17%, the goal difference in a game increases by 0.87 and the team earns 0.06 points more than expected. When a team plays a team, which placed one spot lower in the standings last season, it earns 0.07 points more, probability of win increases by 2%, the goal difference increases by 0.12 goals and the team earns 0.01 points more than expected. And most importantly for our research, after a change of a coach, team earns 0.25 points more, probability to win increases by 8%, the goal difference in a game increases by 0.32 and the team earns 0.29 points more than expected. We find that for both countries, the performance of the team tends to improve after a coaching change based on the 4 performance indicators considered in this analysis. The same interpretation as described can be applied to both countries for the parameter estimates in Table 2 and Table 3.

Next, we compare the estimated parameters for a coaching change between The Netherlands and Italy. The results in Table 2 show that the effects of a coaching change are almost equal for all metrics and the reported F-statistics support that. That shows that when considering all coaching changes, the improvement in team performance after a change is the same in both countries. Furthermore, we look at the parameter estimates for all head coach dismissals in Table 3. It shows that the effect of a coaching change is higher in The Netherlands compared to Italy, however, the F-test for equality of coefficients shows that the difference is not statistically significant. Taking all of this into account, we can conclude that teams in both The Netherlands and Italy perform better after after a coaching change and the difference in the effect of a coaching change is equal in both countries. Therefore, making a coaching change in Italy is not more beneficial to the team performance compared to a change in The Netherlands.

Table 2: Parameter estimates for effects of a head coach change on team performance for both the Eredivisie and Serie A in the span of 2000-2022 for all coaching changes.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
Home	0.53(0.04)***	0.49(0.03)***	0.17(0.01)***	0.15(0.01)***	0.87(0.06)***	0.72(0.04)***	0.06(0.04)*	0.02(0.03)
Pos Opponent	0.07(0.00)***	0.05(0.00)***	0.02(0.00)***	0.01(0.00)***	0.12(0.01)***	0.07(0.00)***	0.01(0.00)***	0.00(0.00)
Coaching change	0.25(0.05)***	0.26(0.03)***	0.08(0.02)***	0.09(0.01)***	0.32(0.08)***	0.30(0.04)***	0.29(0.05)***	0.29(0.03)***
F-statistic	0.02		0.12		0.03		0.01	

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season-fixed effects. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for a coaching change in both countries.

Table 3: Parameter estimates for effects of a head coach change on team performance for both the Eredivisie and Serie A in the span of 2000-2022 for head coach dismissals.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
Home	0.55(0.05)***	0.49(0.03)***	0.18(0.02)***	0.15(0.01)***	0.91(0.07)***	0.72(0.04)***	0.09(0.05)*	0.01(0.03)
Pos Opponent	0.06(0.00)***	0.05(0.00)***	0.02(0.00)***	0.01(0.00)***	0.12(0.01)***	0.07(0.00)***	0.01(0.00)***	0.00(0.00)
Coaching change	0.30(0.05)***	0.26(0.03)***	0.11(0.02)***	0.09(0.01)***	0.40(0.09)***	0.30(0.04)***	0.34(0.05)***	0.30(0.03)***
F-statistic	0.30		0.40		0.90		0.52	

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season-fixed effects. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for a coaching change in both countries.

4.1.2 Matching - treatment and control groups

To create treatment and control groups of actual and counterfactual coaching changes, we set the maximum difference of cumulative surprise between actual and possible last games of a coach to 0.5. Additionally, as a sensitivity analysis we consider different maximum differences in cumulative surprise - 0.25, 0.75 and 1 to see whether the conclusions differ. For the maximum difference of 0.5 we find that from all 97 coaching changes in The Netherlands 82 could be matched to a counterfactual coaching change. For Italy we matched 162 actual coaching changes to 114 counterfactual ones. When looking strictly at dismissals, 71 dismissals in The Netherlands could be matched to 61 counterfactual ones and for Italy 156 dismissals were matched to 114 counterfactual ones.

Figure 2 in Appendix B.1 shows the kernel densities of cumulative surprise at the last matches before coaching changes compared to the last matches before a coaching change could have happened. As the plots show, the cumulative surprise is matched quite accurately and, therefore, the comparison between treatment and control groups refer to similar situations in different seasons.

Table 4 displays the comparison of effects for actual and counterfactual coaching changes on team performance for both the Netherlands and Italy. Results show that for The Netherlands the effects of actual coaching changes seem to be higher compared to possible ones, however, the F-tests show that these differences are not statistically significant. This result is in line with Van Ours and Van Tuijl (2016). On the other hand the results show that for Italy performance improvements are more sizeable when a coaching change does not occur. This difference is statistically significant when considering $Wins$. Therefore, in Italy, when there is an opportunity to change a coach, a probability to win a game will be higher if no change is made. One possible explanation could be that a coach who has known the team longer is better able to make the necessary adjustments than a newly appointed coach.

Table 4: Parameter estimates for effects of all actual and counterfactual coaching changes on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.25(0.05)***	0.26(0.03)***	0.08(0.02)***	0.09(0.01)***	0.32(0.08)***	0.30(0.04)***	0.29(0.05)***	0.29(0.03)***
Matched: Treatment	0.25(0.06)***	0.26(0.04)***	0.09(0.02)***	0.08(0.01)***	0.34(0.09)***	0.32(0.05)***	0.27(0.05)***	0.26(0.03)***
Matched: Control	0.20(0.05)***	0.30(0.04)***	0.07(0.02)***	0.12(0.02)***	0.18(0.08)**	0.29(0.05)***	0.26(0.05)***	0.34(0.04)***
F-statistic	0.55	0.74	0.55	3.12*	1.82	0.15	0.06	2.13

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season-fixed effects. A maximum difference of 0.5 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

Table 5 shows the parameter estimates for coaching change in treatment and control groups when considering the head coach dismissals in both The Netherlands and Italy. The conclusions are the same as made before for all coaching changes. We determine that there is no causal effect between a coaching change and improvement in team performance for The Netherlands. Although, it should be noted that the values of F-statistics are higher for The Netherlands when only considering dismissals, which implies lower p-values and, therefore, a lower probability that effects from actual and counterfactual coach dismissals are equal. However, the null hypothesis of equal coefficients does not get rejected for any of the performance indicators. Regarding Italy, we can see that the coefficients for the control group are higher with the difference in effects for $Wins$ being statistically significant. Similarly as for all coaching changes, it shows that in Italy a team is more likely to win if it does not fire the head coach after a bed spell, but sticks with the current one.

Table 5: Parameter estimates for effects of actual and counterfactual head coach dismissals on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.30(0.05)***	0.26(0.03)***	0.11(0.02)***	0.09(0.01)***	0.40(0.09)***	0.30(0.04)***	0.34(0.05)***	0.30(0.03)***
Matched: Treatment	0.28(0.06)***	0.25(0.04)***	0.10(0.02)***	0.08(0.01)***	0.39(0.11)***	0.32(0.05)***	0.32(0.05)***	0.27(0.03)***
Matched: Control	0.24(0.06)***	0.29(0.04)***	0.08(0.02)***	0.11(0.02)***	0.22(0.10)**	0.27(0.06)***	0.26(0.05)***	0.33(0.04)***
F-statistic	0.22	0.49	0.19	2.78*	1.37	0.40	0.75	1.42

Note: Significance levels: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season-fixed effects. A maximum difference of 0.5 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

Finally, we perform a sensitivity analysis by setting different values of maximum difference in cumulative surprise between actual changes and changes that could have happened. In Appendix B we can see the corresponding parameter estimates and kernel densities for each of the tried maximum differences. In Appendix B.2, Appendix B.3 and Appendix B.4 the results can be viewed for maximum differences of 0.25, 0.75 and 1 respectively. When we compare our results to the ones obtained with different values of maximal difference, the conclusions stay roughly the same. For The Netherlands we find no statistical evidence to suggest that team performs better after actual coaching changes than it does after counterfactual ones even when changing the maximum difference. For Italy a similar trend is observed as described before - teams perform significantly better after counterfactual changes compared to actual ones according to some performance indicators. Table 11 in Appendix B.2 even shows that the difference is statistically significant for two different measures of team performance - $Wins$ and $Surprise$. When the maximum difference is increased, the improvement in the case of counterfactual change is more significant when considering dismissals instead of all changes.

4.2 Duration models

Furthermore, we look at duration models to find what are the determinants of coach dismissals and quits in both The Netherlands and Italy. In this section we report the results for the models with exponential hazard rate, which assumes a constant baseline hazard rate. In Appendix C the parameter estimates for Cox proportional hazard model can be found.

4.2.1 In-season duration

Firstly, we look at in-season duration models, which assume that each season is a fresh start for a coach and T_0 or a start for each new duration is Gameweek 1 of every season. The parameter estimates for in season exponential hazard rate model are shown in Table 6. To explain how each of the coefficients can be interpreted, we look at column on dismissal rate in The Netherlands. In the extended model, coefficient for the variable $Points\ last\ 4$ is -0.23. This means that all other things kept equal and a team earns 1 point more in the span of last 4 games the probability of a coach being dismissed changes by the factor of $\exp(-0.23) = 0.79$. This means that when the coefficient is negative, the increase in the corresponding variable corresponds with a decrease in a probability of a coaching change occurring. For positive coefficients a similar intuition can be applied.

When considering the extended model, the results in Table 6 show that the performance of a team in past games and cumulative surprise have a significantly negative effect on the coach dismissal rate in both The Netherlands and Italy. This result is in line with Van Ours and Van Tuijl (2016) and also makes sense intuitively as football teams are less likely to fire a coach if a team is performing well. A difference between The Netherlands and Italy is that in The Netherlands an older coach is more likely to be fired, while in Italy in the case of an older coach, the probability of dismissal decreases. It should be noted that in our analysis covariates differ from the ones considered in Van Ours and Van Tuijl (2016). However, they found that experience of a coach has a positive effect on dismissal rate, which corresponds with our findings, as it is assumed that age is a substitute for experience. The results show that in The Netherlands teams have less patience when employing an experienced coach, while in Italy the rationale differs. In contrary to Van Ours and Van Tuijl (2016), we find that a coach being capped in his respective national team has a significant negative effect on the dismissal rate in both The Netherlands and Italy and, therefore, teams are less likely to fire a coach if he has been a high-level player before becoming a coach. Regarding the effect of a coach being foreign, results show that in The Netherlands a coach being foreign has no significant effect on the dismissal rate, while in Italy a foreign coach is more likely to be fired if his team performs the same as it would with an Italian head coach.

Furthermore, we consider the restricted model with only performance-based measures and age of a coach as regressors. It points to the same conclusion that better-performing teams are less likely to fire their head coach while the effects of age are similar to ones described before for the extended model and they differ between countries. Moreover, as we change the performance measure of *Points last* 4 to consider a longer time frame, the conclusions are still the same as shown in panels c) and d) of Table 6. However, it should be noted that the coefficients decrease as a longer time before dismissals is considered, which can be explained by the fact that proportionally more points are earned in last 4 than 5 or 6 games, which would make the effect of one additional point earned smaller.

Next, we look at the quit rate in Table 6. It should be noted that there were only 6 in-season quits in Italy during the whole sample and none of the coaches who quit were capped in their national teams, therefore, the variable *Foreign* is not included as a regressor when estimating quit rate in Italy. Similarly to results in Van Ours and Van Tuijl (2016) we find that most of the coefficients are not statistically significant. However, we find that cumulative surprise has a significant negative effect on quit rate in both The Netherlands and Italy. Most of the coefficients being statistically insignificant could be explained by the fact that a coach might quit his job based on some unobserved circumstances, such as personal problems or unsatisfactory relationship with the management. (Semmelroth, 2022)

Finally, we look at Table 17 in Appendix C.1 to find whether the results are sensitive to the choice of a baseline hazard function. The results in Table 17 show that both performance-based measures have a significant negative effect on the dismissal rate. However, it is found that age of a coach is not as significant of a determinant for dismissal rate when considering Cox proportional hazard model. Furthermore, we find that whether a coach is foreign has no significant effect on dismissal rate in either of countries. Regarding rate of quitting, the conclusions are roughly the same as for the exponential hazard model - most coefficients are not statistically significant and,

Table 6: Parameter estimates of in-season exponential hazard model for dismissals and coach quits in both Eredivisie and Serie A in the span of 2000-2022.

	Dismissals		Quits	
Variable	NL	IT	NL	IT
a) Points last 4	-0.23(0.06)***	-0.28(0.04)***	0.02(0.08)	-0.12(0.18)
Cumulative surprise	-0.19(0.03)***	-0.26(0.02)***	-0.08(0.05)*	-0.14(0.09)*
Age	0.34(0.18)**	-0.26(0.12)**	0.09(0.31)	0.49(0.53)
Capped	-0.49(0.29)**	-0.72(0.24)***	0.16(0.41)	
Foreign	0.38(0.36)	1.12(0.27)***	-0.07(0.62)	0.79(1.10)
Log-likelihood	-377.81	-678.40	-185.76	-50.56
b) Points last 4	-0.23(0.06)***	-0.29(0.04)***	0.02(0.08)	-0.12(0.18)
Cumulative surprise	-0.19(0.03)***	-0.23(0.02)***	-0.08(0.05)*	-0.14(0.09)*
Age	0.34(0.18)**	-0.18(0.12)*	0.08(0.31)	0.47(0.53)
Log-likelihood	-379.92	-687.48	-185.84	-50.77
c) Points last 5	-0.18(0.05)***	-0.21(0.04)***	0.06(0.06)	-0.11(0.16)
Cumulative surprise	-0.19(0.03)***	-0.23(0.02)***	-0.09(0.05)**	-0.14(0.09)*
Age	0.34(0.18)**	-0.18(0.12)*	0.07(0.31)	0.47(0.53)
Log-likelihood	-381.39	-694.40	-185.45	-50.74
d) Points last 6	-0.13(0.04)***	-0.17(0.03)***	0.08(0.06)*	-0.12(0.14)
Cumulative surprise	-0.20(0.03)***	-0.24(0.02)***	-0.10(0.05)**	-0.13(0.09)*
Age	0.33(0.18)**	-0.18(0.12)*	0.07(0.31)	0.48(0.53)
Log-likelihood	-384.27	-696.35	-184.89	-50.62

Note: Significance levels: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$. In the brackets are the reported standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. Points last 4, points last 5 and points last 6 are time-varying variables denoting the number of points a team has earned in the last 4, 5 or 6 games. Cumulative surprise is the surprise accumulating over the length of the season. Age is the coaches age at the start of the season divided by 10. Capped is a binary variable indicating that the coach has played for his national time. Foreign is a binary variable indicating whether a coach is a foreigner in the country where he works. Panel a) refers to the extended model, b) is the restricted one and c), d) are for sensitivity analysis. Intercepts not reported.

therefore, have no significant effect on the quit rate. To sum up, there are some differences when comparing models with different baseline hazard functions, which is mostly on which variables are considered statistically significant, while the signs of coefficients are the same in both models. It could be explained by the fact that in Cox proportional hazard model, the baseline hazard function $\lambda_{0,j}(t)$ would explain some of the effects in regressors.

4.2.2 Coaching spell duration

As discussed before in Section 3, in-season duration models are used under the assumption that each season is a fresh start for a coach and therefore, there is no history. This might be an unrealistic setting in real life (Van Ours & Van Tuijl, 2016) and, therefore, we also apply duration models based on the whole tenure of a coach in a specific club. The parameter estimates for such a setting with an exponential hazard rate are shown in Table 7. The coefficients can be interpreted in the same way as they were before for the in-season duration models.

Firstly, we look at the parameter estimates for the extended model of dismissal rates in both The Netherlands and Italy in panel a) of Table 7. It can be seen that for both countries time-varying performance measures (points in last 4 games and cumulative surprise) have significant negative effects on the dismissal rate. As explained before for the in-season duration models it makes intuitive sense as coaches are less likely to be dismissed when a team is performing well. In contradiction to the estimated parameters for in-season duration models in Table 6 we can see that when the full tenure of a coach is considered, his age and whether the coach was capped in the national teams are insignificant determinants of the dismissal rate. However, we can still observe the previously explained dynamic that foreign coaches are more likely to be dismissed in Italy compared to Italians. When we look at the parameter estimates for variables *Second season* and *Third season* the results show that in The Netherlands a coach is more likely to be dismissed in his second season than his first or third. However, in Italy a different dynamic is observed. If a coach is in the third season of his tenure, he is significantly less likely to be fired. It means that if a coach reaches the third season of his tenure, he is quite likely to finish the season in his post. It shows that management of a team in Italy is less prone to make changes if a coach has already worked at the team for a significant amount of time. This is in contradiction of results in Van Ours and Van Tuijl (2016), where it was found that whether a coach is in the second or third season of his tenure has no significant effect on the dismissal rate.

Secondly, we look at the restricted model of dismissal rates in panel b) of Table 7. As it can be seen the results are similar to ones obtained for in-season duration models. Points earned in last 4 games and cumulative surprise have significant negative effects on the dismissal rate for both countries. However, we can see that when considering the whole tenure of a coach, his age still has significant positive effects on the dismissal rate in The Netherlands, while in Italy the effects of age are no longer statistically significant. The results also do not change when considering a different number of past games for one of the performance measures, as shown in panels c) and d) of Table 7.

Thirdly, we consider the parameter estimates for quit rate in both countries. In contradiction to Van Ours and Van Tuijl (2016) we find that coaches are more likely to quit if they are in the second or third season of their tenure, while in Van Ours and Van Tuijl (2016) it was found that none of the considered variables have a significant effect on the quit rate. However, when discussing the quit rate of coaches in both countries it must be noted that we made a crucial assumption, due to which the coefficients may be biased. For the purposes of this analysis we assumed that all tenures of coaches, which end at the end of the season, are quits. This means that while we find these variables to be statistically significant it may not be in line with reality. This should especially be noted in the case of Italy where we found that out of 162 total in-season changes only 6 are recognized as quits, while this approach adds a sizeable number of such changes. However, when further examining the extended model in panel a) of Table 7 we find that all other potential explanatory variables have statistically insignificant effects on the quit rate, which is in line with the results of Van Ours and Van Tuijl (2016). As for restricted model in panel b) and model for sensitivity analysis in panel c) of Table 7, the results show that number of points earned in last 4 or 5 games, cumulative surprise and age of the coach have insignificant effects on the quit rate, similarly as in Van Ours and Van Tuijl (2016). However, when considering panel d) of Table 7, it can be seen that points in last 6 games have a significant positive effect on the quit rate in The Netherlands. Although, it should be noted that the parameter estimate is quite small at 0.04.

Furthermore, we look at the parameter estimates for coaching spell Cox proportional hazard model. As noted before in Section 3, for the Cox model variables $Second\ season$ and $Third\ season$ are not included in the regression. Some of the differences are such that an age of the coach has significant positive effects on the dismissal rate in The Netherlands. Additionally, a coach being capped is found to have significant negative effects in both countries. It is also found that a coach being foreign has significant positive effects on the quit rate in Italy. This means that similarly as in the last section we find that results are sensitive to the choice of the baseline hazard function.

Table 7: Parameter estimates of coaching spell exponential hazard model for dismissals and coach quits in both Eredivisie and Serie A in the span of 2000-2022.

	Dismissals		Quits	
Variable	NL	IT	NL	IT
a) Points last 4	-0.53(0.10)***	-0.34(0.05)***	-0.03(0.04)	0.00(0.06)
Cumulative surprise	-0.22(0.03)***	-0.20(0.03)***	0.02(0.03)	-0.01(0.04)
Age	-0.06(0.24)	-0.03(0.14)	-0.02(0.18)	0.19(0.22)
Capped	-0.40(0.36)	0.06(0.25)	0.07(0.23)	0.24(0.35)
Foreign	0.23(0.50)	0.51(0.32)*	0.14(0.32)	0.21(0.46)
Second season	0.73(0.33)**	0.09(0.24)	1.35(0.25)***	1.53(0.34)***
Third season	-0.05(0.53)	-1.26(0.48)***	1.43(0.31)***	1.88(0.37)***
Log-likelihood	-385.57	-657.59	-483.02	-319.41
b) Points last 4	-0.23(0.06)***	-0.18(0.03)***	-0.02(0.04)	0.04(0.06)
Cumulative surprise	-0.20(0.03)***	-0.16(0.02)***	0.02(0.03)	-0.01(0.04)
Age	0.40(0.18)**	-0.03(0.11)	0.01(0.17)	0.12(0.20)
Log-likelihood	-369.33	-681.17	-502.26	-336.62
c) Points last 5	-0.17(0.05)***	-0.10(0.03)***	0.02(0.04)	0.04(0.05)
Cumulative surprise	-0.21(0.03)***	-0.17(0.02)***	0.01(0.03)	-0.02(0.04)
Age	0.39(0.18)**	-0.03(0.11)	-0.01(0.17)	0.12(0.20)
Log-likelihood	-371.54	-688.60	-502.22	-336.50
d) Points last 6	-0.12(0.04)***	-0.11(0.03)***	0.04(0.03)*	0.05(0.04)
Cumulative surprise	-0.21(0.03)***	-0.17(0.02)***	-0.01(0.03)	-0.03(0.04)
Age	0.38(0.18)**	-0.02(0.12)	-0.01(0.17)	0.12(0.20)
Log-likelihood	-374.40	-685.14	-501.35	-336.17

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. Points last 4, points last 5 and points last 6 are time-varying variables denoting the number of points a team has earned in the last 4, 5 or 6 games. Cumulative surprise is the surprise accumulating over the length of the season. Age is the coaches age at the start of the spell divided by 10. Capped is a binary variable indicating that the coach has played for his national time. Foreign is a binary variable indicating whether a coach is a foreigner in the country where he works. Second season and third season are duration dependence in the second and third season respectively. Panel a) refers to the extended model, b) is the restricted one and c), d) are for sensitivity analysis. Intercepts not reported.

Moreover, we compare the parameter estimates for in-season duration models in Table 6 with coaching spell duration models in Table 7 some differences can be noticed. Firstly, regarding the dismissal rate, fewer coach-specific characteristics are found to have a significant effect on the dismissal rate when considering the coaching spell duration models. However, some characteristics, such as a coach being foreign in Italy have significant positive effects on dismissal rate for both approaches. Secondly, for the coaching spell duration models we find that performance-based measures have no significant effects on the quit rate compared to in-season models, where cumulative surprise was found to have significant negative effects. That would also seem to be more in line with real life, as coaches are likely to be fired when a team is performing badly and quitting could occur due to reasons not related to performance. (Van Ours & Van Tuijl, 2016; Audas et al., 1999)

4.3 Impact of hiring a more experienced head coach during the season

Finally, we investigate the cases where teams do decide to make an in-season coaching change. We distinguish between two possible scenarios - a team hiring a more experienced head coach than the previous one or hiring a coach who is the same age or younger. In this analysis we consider the joint sample of both The Netherlands and Italy. In total there were 259 coaching changes between both countries in the observed time period. It has been found that in 126 of these cases teams hired a coach with more experience after a change, while in 133 remaining instances less experienced coaches were hired. It should be noted that if multiple coaching changes occur in the season, then only the first one is considered in further analysis.

Parameter estimates for the joint sample of The Netherlands and Italy are shown in Table 8. The interpretation of parameters is the same as explained in Section 4.1.1. However, in this case the variable *Coaching change* is split into two different variables corresponding to two previously mentioned scenarios. The results show that for all considered indicators, when hiring a more experienced coach, an improvement in performance is higher compared to the case where a less experienced coach is appointed. However, this difference is not statistically significant for any of the considered performance measures, as shown by the F-test. Therefore, there is not enough statistical evidence to suggest that a team is better off hiring a more experienced coach after making a change.

Table 8: Parameter estimates for effects of hiring a more or less experienced coach on team performance in the combined sample of The Netherlands and Italy in the span of 2000-2022.

Variable	Points	Wins	Goal difference	Surprise
Home	0.51(0.02)***	0.16(0.01)***	0.78(0.03)***	0.03(0.02)
Pos Opponent	0.05(0.00)***	0.01(0.00)***	0.08(0.00)***	0.01(0.00)***
Coaching change more XP	0.29(0.04)***	0.10(0.01)***	0.36(0.05)***	0.32(0.03)***
Coaching change less XP	0.21(0.04)***	0.08(0.01)***	0.25(0.06)***	0.26(0.03)***
F-statistic	2.04	1.79	1.85	2.06

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. All estimates contain club and season-fixed effects. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for hiring a more and less experienced coach.

Furthermore, we consider each of the countries separately. The parameter estimates for Italy can be seen in Table 20 in Appendix D. For Italy the results show that the difference between hiring a more or less experienced coach is even smaller than the one for the joint sample. Therefore, we can conclude that the experience of a hired coach after an in-season change does not have a significant effect on improvement in performance. However, when we consider the parameter estimates for The Netherlands in Table 19 in Appendix D, we can see that appointing a more experienced coach improves performance significantly compared to the alternative. Results show that when a more experienced coach is hired during the season, team earns significantly more points and also earns significantly more points than expected. While the probability to win or the goal difference also increases in the scenario of appointing a more experienced coach, this difference is not statistically significant. However, it should be noted that in The Netherlands only in 35 out of 97 coaching changes a more experienced coach was appointed.

5 Conclusion

This paper examines how does the performance of football teams differ after an in-season coaching change in environments of high (Italy) and low (The Netherlands) volatility. Furthermore, duration models are used to find how do the determinants of head coach dismissals differ between both countries. Moreover, an analysis is performed on how does the change in performance differ when an in-season coaching change is made and a more experienced coach compared to the last one is appointed.

Firstly, similarly to Van Ours and Van Tuijl (2016); Paola and Scoppa (2012) it is found that, when considering the seasons where coaching changes are made, performance improves significantly after the coaching change. This result holds true for both The Netherlands and Italy, however, it is found that for none of the performance measures the difference in effect between The Netherlands and Italy is statistically significant. Therefore, we find that the effects of managerial turnover on performance in football is equal in both more and less volatile environment. Furthermore, we also match actual coaching changes to cases where a coach could have been changed but was not, using bookmaker odds. It is found that the effect of an actual coaching change is equal to one of a counterfactual change in The Netherlands, which is in line with findings of Van Ours and Van Tuijl (2016) and our expectations. However, for Italy there was some evidence to suggest that performance of football teams tends to be better if teams opt to keep the head coach after a bad spell. Therefore, our main finding is that changing managers in both environments of high and low volatility have no effects in performance, but in environment of high volatility it might be counter-productive to make such a change. This result could be extended outside of football. In industries or companies with high manager turnover, it might be beneficial to stick with the current manager after a bad spell instead of making a change.

Secondly, it is found that for both The Netherlands and Italy coaches are less likely to be fired if a team is earning more points and outperforming their expectations, which makes intuitive sense and corresponds to findings of Van Ours and Van Tuijl (2016) and our expectations formed in Section 1. Moreover, it is found that older coaches are more likely to be fired in The Netherlands, while the opposite holds true for Italy. Furthermore, a foreign coach is significantly more likely to be dismissed in Italy, while such dynamic is not observed in The Netherlands. Additionally, for both countries we find that used regressors mostly have no significant effect on the quit rate of coaches. When a coach spell duration is considered instead of in-season duration, it is found that individual coach characteristics are less important when estimating the dismissal rate. Additionally, we find that results are sensitive to the choice of a baseline hazard rate. Therefore, it might be beneficial in future research to consider a wider range of baseline hazard functions.

Finally, we find that in the joint sample of both countries hiring a more experienced coach during the season has no significant benefits on the team performance. However, when examining effects in both countries separately, it is found that in The Netherlands replacing a coach with a more experienced one improves performance of a team significantly.

For further research we suggest including more explanatory variables when explaining performance of a team, for example, money spent in the transfer window if applicable. Additionally, some more explanatory variables could be included for the duration models, such as previous achievements of a coach (Bryson et al., 2021) or social media pressure (Attié, Pacheco & Oliveira, 2023). Additionally, when matching treatment and control groups, a different metric of performance can be used to match different seasons to find whether the same conclusions hold. It should be noted that in this paper we equate age to experience, which might not be true and, therefore, more data on experience should be gathered to verify the results of our analysis, e.g. experience in years since the first coaching job or number of prior coaching spells as in Van Ours and Van Tuijl (2016). Furthermore, this analysis could be extended to more countries to see whether the differences between more volatile and less volatile environments still hold. For example, differences between Spain and France could be examined.

References

Attié, M., Pacheco, D. & Oliveira, M. (2023). Getting the boot? predicting the dismissal of managers in football. In *International workshop on complex networks* (pp. 132–140).

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.

Bryson, A., Buraimo, B., Farnell, A. & Simmons, R. (2021). Special ones? the effect of head coaches on football team performance.

Carmichael, F. & Thomas, D. (2005). Home-field effect and team performance: evidence from english premiership football. *Journal of sports economics*, *6*(3), 264–281.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

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.

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.

Gilfix, Z., Meyerson, J. & Addona, V. (2020). Longevity differences in the tenures of american and foreign major league soccer managers. *Journal of Quantitative Analysis in Sports*, *16*(1), 17–26.

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

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.

Narita, K., Tena, J. & Buraimo, B. (2022). Causal and consequences of multiple dismissals: Evidence from italian football league.

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.

Pfeffer, J. & Davis-Blake, A. (1986). Administrative succession and organizational performance: How administrator experience mediates the succession effect. *Academy of Management journal*, *29*(1), 72–83.

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

Pierce, D. A., Johnson, J. E., Krohn, B. D. & Judge, L. W. (2017). Who should we hire?: Examining coaching succession in ncaa division i women's basketball. *International Journal of Sports Science & Coaching*, *12*(2), 151–161.

Porro, F., Restaino, M., Ruiz-Castro, J. E. & Zenga, M. (2021). Determinants of football coach dismissal in italian league serie a. *Preface XIX 1 Plenary Sessions*, 805.

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.

Stadtmann, G. (2003). Frequent news and pure signals: Evidence of a publicly traded football club. *Scottish Journal of Political Economy, Vol53 (4)*, 485–504.

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.

Watrous, K. M., Huffman, A. H. & Pritchard, R. D. (2006). When coworkers and managers quit: The effects of turnover and shared values on performance. *Journal of Business and Psychology*, *21*, 103–126.

White, P., Persad, S. & Gee, C. J. (2007). The effect of mid-season coach turnover on team performance: The case of the national hockey league (1989–2003). *International Journal of Sports Science & Coaching*, *2*(2), 143–152.

Zart, S. & Güllich, A. (2022). In-season head-coach changes have positive short-and long-term effects on team performance in men's soccer—evidence from the premier league, bundesliga, and la liga. *Journal of sports sciences*, *40*(6), 696–703.

Appendices

A Variables used in the analysis

Table 9: Description of variables considered in the analysis

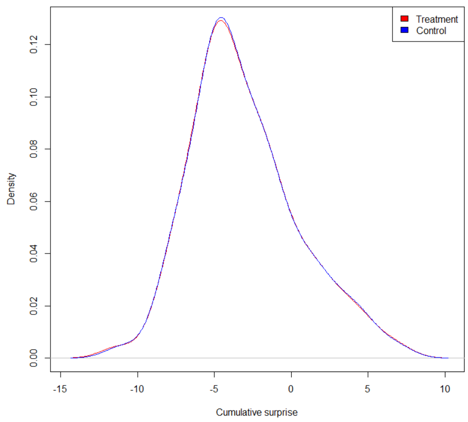
Variable	Description
WNR	Gameweek in which a game happens
Club	Name of the team considered
N	Name of the opposing team
H	Binary variable indicating where the game was played (0 - away, 1 - home)
V	Number of goals scored
T	Number of goals opponent scored
P	Position of an opponent in the last season (-1 for newly promoted teams)
changeT1	Type of change that happened in the considered season
NT	Name of the head coach
AGE	Age of the head coach
seizoen	Season in which the game happens
oddh	Odds of a home team winning the game
oddd	Odds of a draw in the game
odda	Odds of an away win in the game
nationality	Nationality of a coach
capped	Binary variable indicating whether the coach was capped in his national team (1 = has been capped, 0 = has not been capped)
caps	Number of caps in the national team
ID	Unique ID assigned to each individual season of each club

Table 10: Summary statistics of different variables considered in the analysis for both The Netherlands and Italy in the time period of 2000-2022

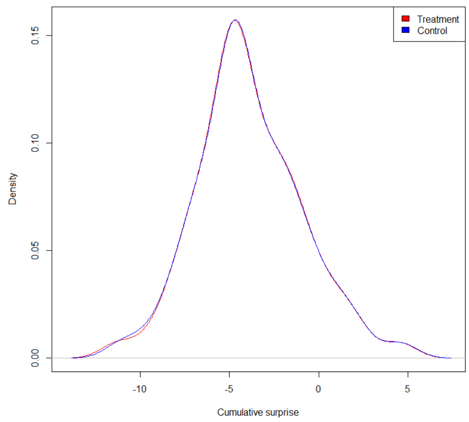
	Mean		Maximum		Minimum	
Variable	NL	IT	NL	IT	NL	IT
WNR	17.36	19.20	34.00	38.00	1.00	1.00
H	0.50	0.50	1.00	1.00	0.00	0.00
V	1.52	1.34	13.00	7.00	0.00	0.00
T	1.52	1,34	13.00	7.00	0.00	0.00
P	9.59	7.05	18.00	20.00	-1.00	-1.00
AGE	47.58	50.35	73.00	71.00	31.00	35.00
oddh	2.53	2.56	29.00	21.00	0.00	0.00
oddd	3.97	3.54	15.00	15.00	0.00	0.00
odda	4.54	4.53	71.00	34.00	0.00	0.00
capped	0.35	0.25	1.00	1.00	0.00	0.00
caps	11.45	9.62	112.00	125.00	0.00	0.00
win	0.38	0.37	1.00	1.00	0.00	0.00
points	1.38	1.37	3.00	3.00	0.00	0.00
surprise	0.00	0.00	2.72	2.71	-2.61	-2.58
cumulative surprise	0.05	0.01	20.75	20.33	-18.31	-18.39

B Matching - treatment and control groups with different maximum difference of cumulative surprise

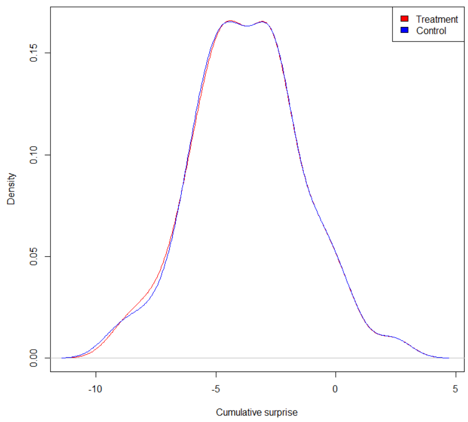
B.1 Maximum difference - 0.50



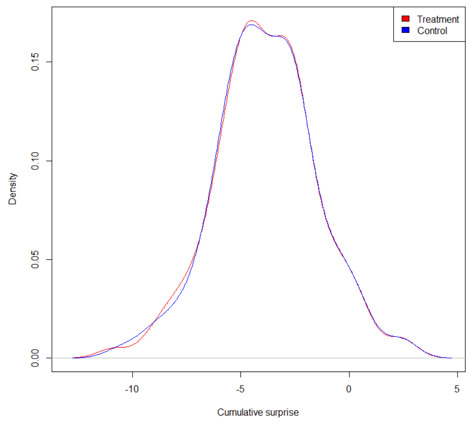
(a) All matched coaching changes in The Netherlands



(b) All matched coach dismissals in The Netherlands



(c) All matched coaching changes in Italy



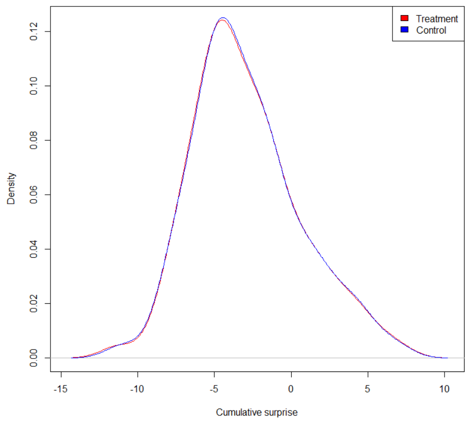
(d) All matched coach dismissals in Italy

Figure 2: Kernel densities of actual and counterfactual last games of a coach before a change in The Netherlands and Italy (2000-2022) with a maximum difference in cumulative surprise of 0.5.

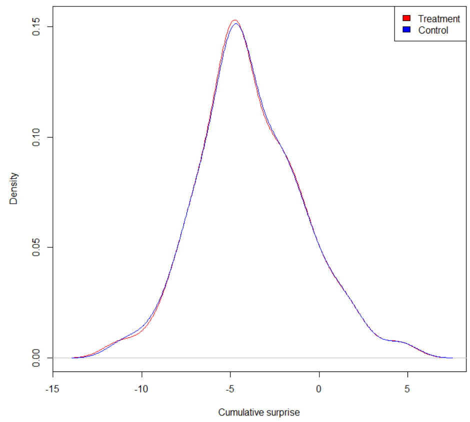
B.2 Maximum difference - 0.25

In this case the maximum difference of cumulative surprise between actual and counterfactual last game is set to 0.25. For all changes the size of treatment and control groups is 78 seasons for The Netherlands and 114 for Italy. When looking at dismissals, the size of treatment and control groups is 60 for The Netherlands and 109 for Italy.

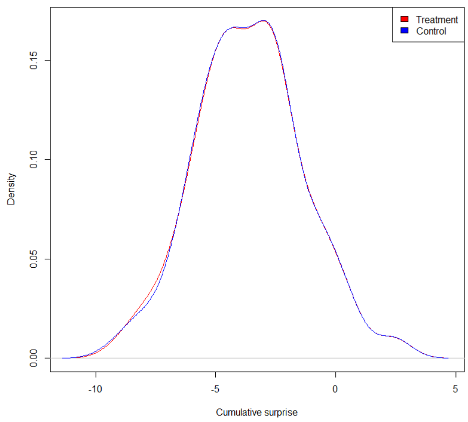
Figure 3 shows the kernel densities of cumulative surprise in matched actual and counterfactual last games of a coach. Table 11 reports the parameter estimates for effects of actual and counterfactual changes in team performance when considering all changes, while Table 12 shows the parameter estimates of these coefficients when only considering coach dismissals.



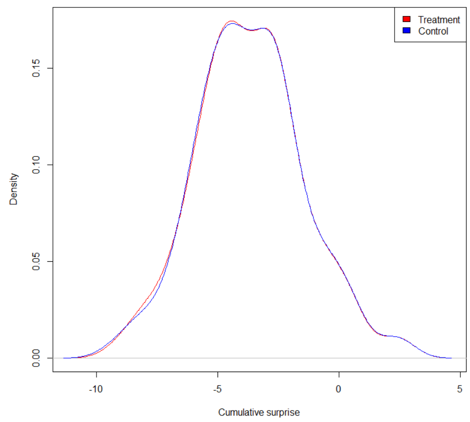
(a) All matched coaching changes in The Netherlands



(b) All matched coach dismissals in The Netherlands



(c) All matched coaching changes in Italy



(d) All matched coach dismissals in Italy

Figure 3: Kernel densities of actual and counterfactual last games of a coach before a change in The Netherlands and Italy (2000-2022) with a maximum difference in cumulative surprise of 0.25.

Table 11: Parameter estimates of effects of all actual and counterfactual coaching changes on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.25(0.05)***	0.26(0.03)***	0.08(0.02)***	0.09(0.01)***	0.32(0.08)***	0.30(0.04)***	0.29(0.05)***	0.29(0.03)***
Matched: Treatment	0.22(0.06)***	0.23(0.04)***	0.08(0.02)***	0.07(0.02)***	0.29(0.09)***	0.29(0.06)***	0.25(0.05)***	0.25(0.03)***
Matched: Control	0.18(0.05)***	0.31(0.04)***	0.06(0.02)***	0.12(0.01)***	0.17(0.09)**	0.30(0.05)***	0.23(0.05)***	0.34(0.03)***
F-statistic	0.21	1.93	0.22	5.15**	0.88	0.01	0.07	3.61*

Note: Significance levels: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season fixed effects. A maximum difference of 0.25 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

Table 12: Parameter estimates of effects of actual and counterfactual head coach dismissals on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

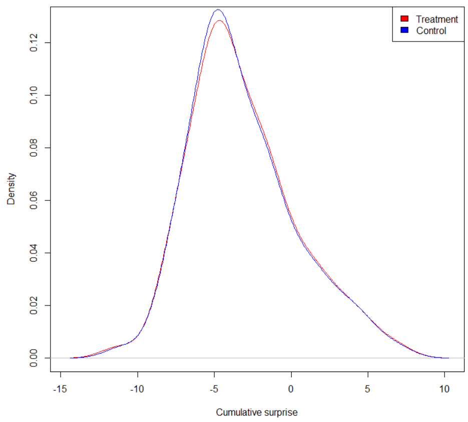
	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.30(0.05)***	0.26(0.03)***	0.11(0.02)***	0.09(0.01)***	0.40(0.09)***	0.30(0.04)***	0.34(0.05)***	0.30(0.03)***
Matched: Treatment	0.29(0.06)***	0.24(0.04)***	0.10(0.02)***	0.08(0.02)***	0.41(0.10)***	0.30(0.06)***	0.33(0.05)***	0.25(0.03)***
Matched: Control	0.22(0.06)***	0.30(0.05)***	0.08(0.02)***	0.12(0.02)***	0.20(0.09)**	0.28(0.06)***	0.25(0.06)***	0.34(0.04)***
F-statistic	0.75	0.89	0.66	3.40*	2.41	0.06	1.43	2.41

Note: Significance levels: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season fixed effects. A maximum difference of 0.25 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

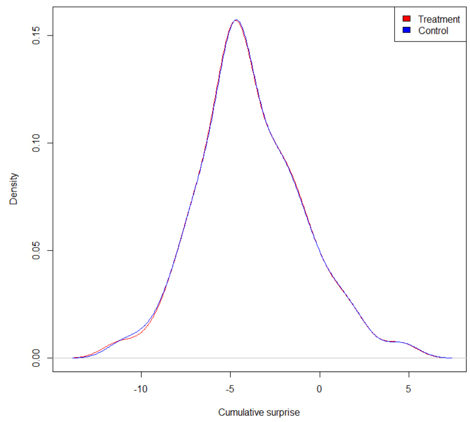
B.3 Maximum difference - 0.75

In this case the maximum difference of cumulative surprise between actual and counterfactual last game is set to 0.75. For all changes the size of treatment and control groups is 84 seasons for The Netherlands and 118 for Italy. When looking at dismissals, the size of treatment and control groups is 61 for The Netherlands and 118 for Italy.

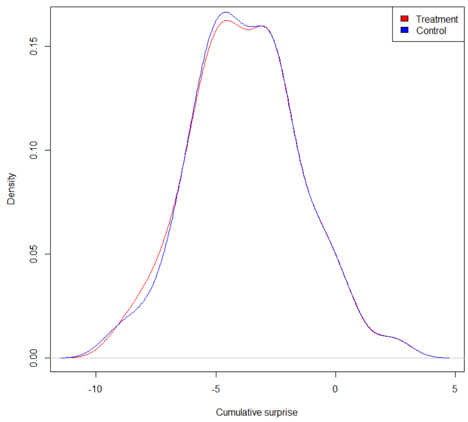
Figure 4 shows the kernel densities of cumulative surprise in matched actual and counterfactual last games of a coach. Table 13 reports the parameter estimates for effects of actual and counterfactual changes in team performance when considering all changes, while Table 14 shows the parameter estimates of these coefficients when only considering coach dismissals.



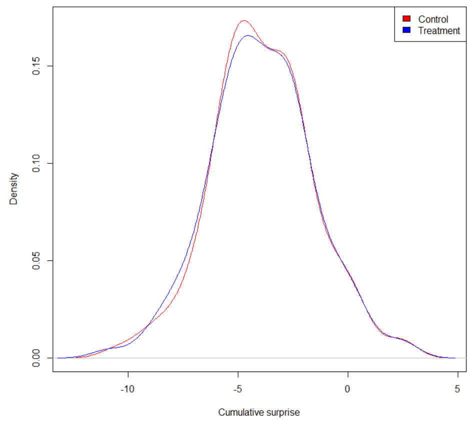
(a) All matched coaching changes in The Netherlands



(b) All matched coach dismissals in The Netherlands



(c) All matched coaching changes in Italy



(d) All matched coach dismissals in Italy

Figure 4: Kernel densities of actual and counterfactual last games of a coach before a change in The Netherlands and Italy (2000-2022) with a maximum difference in cumulative surprise of 0.75.

Table 13: Parameter estimates of effects of all actual and counterfactual coaching changes on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.25(0.05)***	0.26(0.03)***	0.08(0.02)***	0.09(0.01)***	0.32(0.08)***	0.30(0.04)***	0.29(0.05)***	0.29(0.03)***
Matched: Treatment	0.26(0.06)***	0.26(0.04)***	0.09(0.02)***	0.08(0.02)***	0.36(0.09)***	0.31(0.06)***	0.28(0.05)***	0.26(0.04)***
Matched: Control	0.21(0.05)***	0.30(0.04)***	0.07(0.02)***	0.12(0.01)***	0.20(0.09)**	0.28(0.05)***	0.25(0.05)***	0.32(0.04)***
F-statistic	0.39	0.46	0.39	2.52	1.50	0.10	0.16	1.39

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season fixed effects. A maximum difference of 0.75 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

Table 14: Parameter estimates of effects of actual and counterfactual head coach dismissals on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.30(0.05)***	0.26(0.03)***	0.11(0.02)***	0.09(0.01)***	0.40(0.09)***	0.30(0.04)***	0.34(0.05)***	0.30(0.03)***
Matched: Treatment	0.28(0.06)***	0.26(0.04)***	0.10(0.02)***	0.08(0.01)***	0.39(0.11)***	0.31(0.05)***	0.32(0.05)***	0.27(0.03)***
Matched: Control	0.24(0.06)***	0.31(0.05)***	0.08(0.02)***	0.12(0.02)***	0.22(0.10)**	0.29(0.06)***	0.26(0.05)***	0.33(0.04)***
F-statistic	0.22	0.74	0.19	3.28*	1.37	0.07	0.75	1.38

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season fixed effects. A maximum difference of 0.75 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

B.4 Maximum difference - 1

In this case the maximum difference of cumulative surprise between actual and counterfactual last game is set to 1. For all changes the size of treatment and control groups is 86 seasons for The Netherlands and 123 for Italy. When looking at dismissals, the size of treatment and control groups is 62 for The Netherlands and 123 for Italy.

Figure 5 shows the kernel densities of cumulative surprise in matched actual and counterfactual last games of a coach. Table 15 reports the parameter estimates for effects of actual and counterfactual changes in team performance when considering all changes, while Table 16 shows the parameter estimates of these coefficients when only considering coach dismissals.

Table 15: Parameter estimates of effects of all actual and counterfactual coaching changes on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

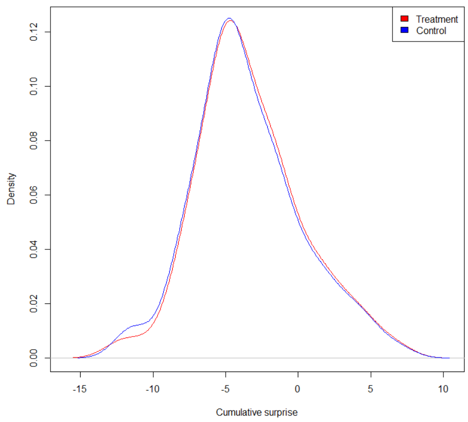
	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.25(0.05)***	0.26(0.03)***	0.08(0.02)***	0.09(0.01)***	0.32(0.08)***	0.30(0.04)***	0.29(0.05)***	0.29(0.03)***
Matched: Treatment	0.27(0.06)***	0.25(0.04)***	0.09(0.02)***	0.08(0.01)***	0.35(0.9)***	0.29(0.05)***	0.29(0.05)***	0.26(0.03)***
Matched: Control	0.21(0.05)***	0.29(0.04)***	0.07(0.02)***	0.11(0.02)***	0.20(0.08)**	0.28(0.04)***	0.26(0.05)***	0.32(0.04)***
F-statistic	0.67	0.77	0.52	2.29	1.50	0.02	0.21	1.46

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season fixed effects. A maximum difference of 1 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.

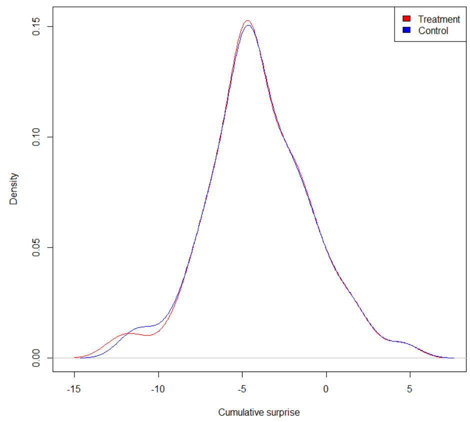
Table 16: Parameter estimates of effects of actual and counterfactual head coach dismissals on team performance using the Treatment and Control groups for Eredivisie and Serie A during the time span of 2000-2022.

	Points		Wins		Goal difference		Surprise	
Variable	NL	IT	NL	IT	NL	IT	NL	IT
All actual changes	0.30(0.05)***	0.26(0.03)***	0.11(0.02)***	0.09(0.01)***	0.40(0.09)***	0.30(0.04)***	0.34(0.05)***	0.30(0.03)***
Matched: Treatment	0.29(0.06)***	0.26(0.03)***	0.10(0.02)***	0.08(0.01)***	0.39(0.11)***	0.31(0.05)***	0.33(0.05)***	0.27(0.03)***
Matched: Control	0.23(0.06)***	0.31(0.04)***	0.08(0.02)***	0.12(0.02)***	0.21(0.10)**	0.29(0.05)***	0.26(0.05)***	0.33(0.04)***
F-statistic	0.46	0.82	0.35	3.66*	1.60	0.04	1.14	1.45

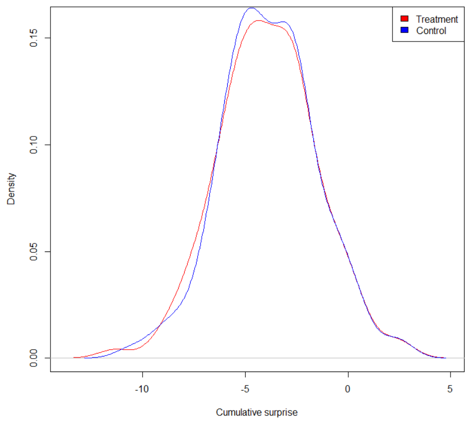
Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. All estimates contain club and season fixed effects. A maximum difference of 1 in cumulative surprise between actual and counterfactual changes during the last match is considered. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for actual and counterfactual coaching changes.



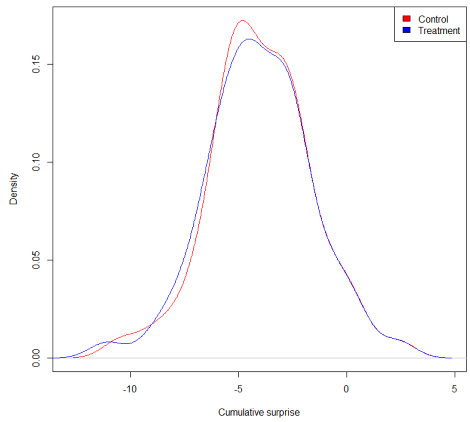
(a) All matched coaching changes in The Netherlands



(b) All matched coach dismissals in The Netherlands



(c) All matched coaching changes in Italy



(d) All matched coach dismissals in Italy

Figure 5: Kernel densities of actual and counterfactual last games of a coach before a change in The Netherlands and Italy (2000-2022) with a maximum difference in cumulative surprise of 1.

C Output for duration models with an exponential baseline hazard

C.1 In-season duration

Table 17: Parameter estimates of in-season Cox proportional hazard model for dismissals and coach quits in both Eredivisie and Serie A in the span of 2000-2022.

	Dismissals		Quits	
Variable	NL	IT	NL	IT
a) Points last 4	-0.27(0.06)***	-0.37(0.05)***	-0.01(0.08)	-0.16(0.19)
Cumulative surprise	-0.18(0.03)***	-0.24(0.02)***	-0.07(0.05)	-0.11(0.10)
Age	0.31(0.18)*	-0.19(0.12)	0.09(0.31)	0.46(0.53)
Capped	-0.49(0.29)*	-0.64(0.24)***	0.18(0.41)	
Foreign	0.42(0.36)	1.23(0.27)	-0.06(0.62)	0.78(1.10)
Log-likelihood	-355.91	-736.54	-153.03	-33.63
b) Points last 4	-0.28(0.06)***	-0.38(0.05)***	-0.00(0.08)	-0.16(0.19)
Cumulative surprise	-0.18(0.03)***	-0.21(0.02)***	-0.08(0.05)	-0.11(0.10)
Age	0.30(0.18)*	-0.13(0.12)	0.09(0.31)	0.46(0.53)
Log-likelihood	-358.08	-745.81	-153.13	-33.85
c) Points last 5	-0.24(0.05)***	-0.30(0.04)***	0.03(0.07)	-0.17(0.18)
Cumulative surprise	-0.17(0.03)***	-0.20(0.02)***	-0.10(0.06)*	-0.10(0.10)
Age	0.29(0.18)	-0.13(0.12)	0.08(0.31)	0.45(0.53)
Log-likelihood	-358.66	-751.20	-153.02	-33.73
d) Points last 6	-0.18(0.05)***	-0.29(0.04)***	0.06(0.06)	-0.20(0.16)
Cumulative surprise	-0.18(0.03)***	-0.19(0.02)***	-0.11(0.06)*	-0.08(0.11)
Age	0.28(0.18)	-0.11(0.12)	0.08(0.31)	0.46(0.52)
Log-likelihood	-362.02	-749.01	-152.75	-33.38

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.In the brackets are the reported standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. Points last 4, points last 5 and points last 6 are time varying variables denoting the number of points a team has earned in the last 4, 5 or 6 games. Cumulative surprise is the surprise accumulating over the length of the season. Age is the coaches age at the start of the season divided by 10. Capped is a binary variable indicating that the coach has played for his national time. Foreign is a binary variable indicating whether a coach is a foreigner in the country where he works. Panel a) refers to the extended model, b) is the restricted one and c), d) are for sensitivity analysis.

C.2 Coach spell duration

Table 18: Parameter estimates of coaching spell Cox proportional hazard model for dismissals and coach quits in both Eredivisie and Serie A in the span of 2000-2022.

	Dismissals		Quits	
Variable	NL	IT	NL	IT
a) Points last 4	-0.28(0.06)***	-0.29(0.05)***	-0.06(0.04)	-0.05(0.06)
Cumulative surprise	-0.20(0.03)***	-0.32(0.03)***	-0.01(0.02)	-0.02(0.03)
Age	0.48(0.20)**	-0.05(0.12)	0.06(0.18)	0.26(0.23)
Capped	-0.59(0.29)**	-0.44(0.25)*	0.04(0.23)	-0.07(0.35)
Foreign	0.42(0.37)	1.00(0.29)***	0.39(0.33)	0.86(0.49)*
Log-likelihood	-267.44	-577.40	-314.65	-164.49
b) Points last 4	-0.29(0.06)***	-0.28(0.05)***	-0.05(0.04)	-0.05(0.06)
Cumulative surprise	-0.20(0.03)***	-0.30(0.03)***	-0.01(0.02)	-0.02(0.03)
Age	0.47(0.20)**	-0.00(0.12)	0.09(0.18)	0.23(0.21)
Log-likelihood	-270.35	-582.88	-315.32	-165.96
c) Points last 5	-0.24(0.05)***	-0.22(0.04)***	-0.03(0.03)	-0.05(0.05)
Cumulative surprise	-0.19(0.04)***	-0.31(0.03)***	-0.01(0.02)	-0.02(0.03)
Age	0.44(0.20)**	-0.00(0.12)	0.09(0.18)	0.22(0.21)
Log-likelihood	-272.00	-587.19	-315.66	-165.71
d) Points last 6	-0.17(0.05)***	-0.20(0.04)***	-0.02(0.03)	-0.05(0.05)
Cumulative surprise	-0.20(0.04)***	-0.30(0.03)***	-0.02(0.02)	-0.01(0.03)
Age	0.44(0.20)**	0.01(0.12)	0.09(0.18)	0.23(0.21)
Log-likelihood	-275.66	-586.51	-315.99	-165.66

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported standard errors. NL and IT refer to the parameter estimates in each country - NL being The Netherlands and IT Italy. Points last 4, points last 5 and points last 6 are time varying variables denoting the number of points a team has earned in the last 4, 5 or 6 games. Cumulative surprise is the surprise accumulating over the length of the season. Age is the coaches age at the start of the spell divided by 10. Capped is a binary variable indicating that the coach has played for his national time. Foreign is a binary variable indicating whether a coach is a foreigner in the country where he works. Panel a) refers to the extended model, b) is the restricted one and c), d) are for sensitivity analysis.

D Impact of hiring a more experienced coach - separate analysis with The Netherlands and Italy

Table 19: Parameter estimates on effects of hiring a more experienced coach on team performance for Eredivisie in the span of 2000-2022.

Variable	Points	Wins	Goal difference	Surprise
Home	0.53(0.04)***	0.17(0.01)***	0.87(0.06)***	0.07(0.04)*
Pos Opponent	0.07(0.00)***	0.02(0,00)***	0.12(0.01)***	0.01(0.00)***
Coaching change more XP	0.34(0.07)***	0.11(0.03)***	0.41(0.12)***	0.37(0.06)***
Coaching change less XP	0.19(0.06)***	0.07(0.02)**	0.27(0.10)**	0.24(0.05)***
F-statistic	3.20*	2.64	0.86	3.90**

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. All estimates contain club and season fixed effects. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for hiring a more and less experienced coach. Of total 97 coaching changes in 35 instances a more experienced coach was hired, while in 62 a less experienced one was.

Table 20: Parameter estimates on effects of hiring a more experienced coach on team performance for Serie A in the span of 2000-2022.

Variable	Points	Wins	Goal difference	Surprise
Home	0.49(0.03)***	0.15(0.01)***	0.73(0.04)***	0.02(0.03)
Pos Opponent	0.04(0.00)***	0.01(0.00)***	0.07(0.00)***	0.00(0.00)
Coaching change more XP	0.27(0.04)***	0.09(0.01)***	0.34(0.06)***	0.30(0.03)***
Coaching change less XP	0.24(0.05)***	0.08(0.02)***	0.25(0.07)**	0.28(0.04)***
F-statistic	0.23	0.18	1.00	0.14

Note: Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. In the brackets are the reported robust standard errors. All estimates contain club and season fixed effects. F-statistics are the test statistics of an F-test with a null hypothesis of equal coefficients for hiring a more and less experienced coach. 91 instances of hiring a more experienced coach, 71 of less.

E Programming code

The programming for this thesis is done in *R*. All the codes are written originally by the author of this paper. Regressions to obtain parameter estimates for naive approach models, output of matching algorithm and effects of hiring a more experienced coach are performed using the *plm*() function of package *plm* in *R*. This choice is made as the function accommodates panel data and allows to calculate club and season fixed effects. Furthermore functions *coeftest*() and *linearHypothesis*() of *R* packages *sandwich* and *car* are used to test for significance and equality of coefficients respectively. When estimating the duration models for Cox proportional hazard model function *coxph*() of *R* package *survival* is used. To calculate the model with exponential baseline hazard we use function *flexsurvreg*() of package *flexsurvreg* in *R*. Such a choice is made due to the fact that function *survreg*() of package *survival* in R does not allow for time varying covariates, which is imperative in this particular analysis. Moreover function *tidy*() is used to test for statistical significance of coefficients in the exponential hazard model.