What are the effects of Covid- 19 on professional football?

**An analysis of the influences of “ghost games” on home advantage**

**Master thesis Marketing Analytics Spring 2020**

**Name: Alan Rijnders**

**SNR: 2063482**

**Date: 16-02-2021**

Contents

[1 Introduction 1](#_Toc534899587)

[1.1 Problem Indication 1](#_Toc534899588)

[1.2 Research Questions 3](#_Toc534899593)

[1.3 Research Approach 3](#_Toc534899598)

[1.4 Academic Relevance 3](#_Toc534899599)

[1.5 Managerial Relevance 4](#_Toc534899600)

[1.6 Structure of Thesis 5](#_Toc534899609)

[2 Theoretical Background](#_Toc534899614) 6

[2.1 Literature Review 6](#_Toc534899588)

2.1.1 The role of home advantage in sports........................................................................6

2.1.2 The role of home advantage in football.....................................................................6

2.1.3 Home advantage during covid-19..............................................................................9

[2.2 Theoretical Framework ..... 12](#_Toc534899593)

[2.3 Conceptual Model 15](#_Toc534899598)

[3 Method](#_Toc534899614) 16

[4 Analysis and Findings 17](#_Toc534899614)

[5 Conclusions and Recommendations 1](#_Toc534899614)9

[5.1 Conclusions 20](#_Toc534899610)

[5.2 Recommendations 20](#_Toc534899611)

[5.3 Future Research 20](#_Toc534899612)

[6 Bibliography 21](#_Toc534899614)

# Introduction

## Problem Indication

Football is the most popular sport in the world with millions of players across the Globe. According to a study conducted by FIFA in 2006, the amount of active football players was 265 million(FIFA 2007). In other words, roughly 4 percent of total world population at that time was regularly playing football. The amount of fans is even more staggering, with 3.5 billion people tuning in for the FIFA World Cup Final in 2018 between France and Croatia(FIFA, 2019) for example.

Many of these fans were left without their favourite pastime for a considerable time when the Covid pandemic struck Europe and other parts of the world in March 2020. Due to strict measures and even complete lockdowns in several countries, almost all sports games were cancelled or postponted until further notice. After a few months of lockdown, the Bundesliga was the first major league to restart their competition to finish the remaining quarter of the games. On the 16th of May the Bundesliga recommenced with a packed schedule featuring 6 matches behind closed doors. Other major european leagues such as the Premier League, Serie A and La Liga followed soon and there was even some space in the packed playing calendar to finish the remaining knock-out matches of the Champions League and Europa League. Global social distancing measures following the Covid-19 pandemic introduced the phenomenon “Ghost Matches” in the football world during the restart. So called “Ghost Matches” are football matches without any spectators attending. Without the fans attendance the matches seemed mere training games, which gives rise to the question what sort of influences this could have on the the outcome of matches. Is home advantage still present in games without a crowd? Is referee behaviour different without fans? Do Away teams perform better under the new circumstances? We aim to answer these questions in this paper.

Due to its popularity, an abundant base of research on football and each of its aspects currently exists. The same holds for the possible advantage of the team playing at home, with several papers trying to disentangle the dynamics at play that could lead to home advantage. (e.g. Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007; Pollard & Polla[rd, 2005](https://shapeamerica.tandfonline.com/doi/full/10.1080/02640410601038576?scroll=top&needAccess=true)). Examples of previous research topics include, does the crowd raise home team performance? Or is the performance of away teams significantly lower in hostile environments? (..) and (..) find that home teams on average win more games, collect more points and score more goals than away. However, whether this is a result from deteriorating away performances, better home performance or a combination of both is not clear yet. Another factor involved in home advantage is the referee. A referee has a large influence on the outcome of the games(..) Especially in a sport as football which has such a low scoring nature(..). A decisions to award a team a penalty in the 89th minute of the match with the score of 0-0 could completely change the match outcome. Or an early red card significantly alters the course of the match with both teams adjusting their tactics and strategy. (Neville and Holder, 1999 and Bokyo 2007) find that crowds could influence referee decisions subconsciously in favor of the home team. (Endrich & Gesche, ,2020) quantify this in their paper where they find that away teams on average receive 0.3 cards less and home teams 0.5 cards more per match when there are no spectators.

However, opinions on how and to what extent home advantage is shaped by crowd support are very mixed. (..) attribute home advantage solely as a result of the effect of the crowd on referee bias for example. Whereas, other papers perceive different tactics or different psychological and physiological states of players as a result of crowd support as important. In light of the uncertainty in the effect of crowd support, the current extraordinary circumstances thus provide a special occasion to increase our understanding of the relevance of crowd support and its influence on player and referee behaviour as well as performance.

## Research Questions

This natural experiment translates into the following research questions:

* + 1. What is the effect of the exclusion of home supporters on home advantage?
    2. To what extent is referee behaviour shaped by the home crowd?
    3. Is home and away team performance significantly different as a consequence of “ghost games”?

## Research Approach

Providing legitimate answers to the aforementioned research questions requires a combination of theoretical and empirical analysis. Firstly, we examine the current and historical literature to obtain information on the relationship between crowd attendance and team performance. This will serve as our basis on which we can conduct the right analytical approach to analyze our data. We believe that such an aggregate analysis provides more information on home advantage than single league studies since a single league could perhaps simply be an anomaly with contrasting results compared to the general trend.

Historical data on football matches for the 11 main football leagues is available starting from the year 2000 until the latest season 2020/2021. The datasets are automatically updated every week to provide recent and accurate data. We then combine match data with a dataset from Fivethirtyeight to incorporate expected goals as extra measures of performance and home advantage. This dataset also contains team strength and match importance as control variables into the analysis. After data collection and construction of the final datasets we construct a regression model to apply to the data in order to draw valid conclusions on the existence of home advantage.

## Academic Relevance

The batch of matches without fans provides an unique opportunity to delve deeper in the impact of crowds on football matches since there is an abundance in new data for matches played without crowds. Past papers often required advanced econometric techniques to be able to discern the extent to which a crowd influences home team advantage. The difficulties in assessing drivers of home advantage lies in confouding variables effects(2008). Which make it unclear how much of home advantage can be attributed to what factors since most of the factors of home advantage are connected to eachother. However, in the current situation, suddenly there are substantial amounts of data on “ghost games” which can be used to directly compare matches with and without spectators. This simplifies the analysis of crowd effects on football matches significantly.

A few papers on home advantage in times of covid-19 have been or will be published already, including the work of (Fischer & Haucap, 2020;Deutscher & Winkelmann, 2020; Endrich & Gesche, 2020). They analyzed the home advantage during the covid-19 pandemic. However,each of these papers focused on German leagues in their analysis. We add to the current research by extending the preliminary research already completed to numerous other football leagues to obtain a comprehensive overview of the evolution of home advantage during the pandemic.

Another contribution of our paper is the extended availability of data with an entire season of games played behind closed doors available after the 2020/21 season. Papers from last year mostly only use data from the 2019/2020 season which was partly played under normal circumstances and partly played behind closed doors. By using the weekly updated datasets with the right timing we can incorporate the entirety of the 2020/21 season into our model.

Finally we present a comprehensive framework of the influence of crowd support on home advantage by using several mediating and moderating variables to assess how and to which extent crowd support influences home advantage. With our Structural equations modelling approach we try to use as much data as possible and combine different variables of home advantage, referee bias and intensity of crowd support(number of supporters and crowd density) to obtain more meaningful data on these measures. In our framework we analyze whether referee and team performance are significant mediators in the crowd support and home advantage relationships. This will give insight through which channels crowd support affects home advantage in football.

## Managerial Relevance

Covid affects the footballing world in multiple ways. The absence of crowds plus the extra heavy loaded schedule increase uncertainty surrounding team and individual performance(Metelski & Kornakov, 2021). Better results lead to better stock performance(Samagaio, Couto & Caiado, 2009) and increased market value for players(Müller, Simons & Weinmann, 2017). Similar findings are reported by (Galariotis, Germain & Zopounidis, 2018) and (He, Cachucho & Knobbe,2015). as they find positive relationships between revenues,position in the league table and individual performance.Therefore, estimating the influence covid-19 has on team performance and football results will help football clubs evaluating the effect of the pandemic on their marketing, financial and on-field performance.

Knowledge on what variables exactly drive home advantage and team performance provides great insight for football clubs in how to optimize their clubs environment and team to improve the chances of performing well. (Fischer & Haucap 2020) for example see preliminary findings of a significant effect of crowd occupancy on home team performance. By comparing home advantage during covid-19 between the German Bundesliga and the 2nd and 3rd level of German football, they find that the home advantage has only changed significantly in the Bundesliga. Whereas in the 2nd Bundesliga and 3rd Liga home advantage did not change significantly during ghost games. They account this difference to the differences in occupancy rates between these competitions. There are numerous small factors that could influence home advantage in football. (Attrill, Gresty, Hill & Barton 2008; Allen & Jones, 2014) find that even something trivial such as the colour of a football team’s shirt could moderately influence team performance, where teams in red have a higher ranking on average compared to teams wearing blue.

By examining the effect of variables such as crowd occupancy and team colours on home performance, we aim to provide marketeers direct tools to influence the performance of their clubs. Hypothetically speaking, knowing that improving the occupancy rates for their team increases team performance gives the marketing department a significant task in finding ways to attract more fans to the stadium. Perhaps, a decrease in ticket price with the associated lower per customer revenue could actually turn out as a smart investment with better team performance and an upwards positive spiral both on and off the field. We believe the current literature is lacking in this area. Most of the papers available on this topic refrain from applying their findings to managerial recommendations for football clubs. We add to the current body of literature by examining how factors that are to a certain extent under control of management of football clubs could influence home advantage and team performance and thus provide extra tools on how football clubs can improve their on field performance.

## Structure of the Thesis

This first chapter serves as a background chapter for the rest of the thesis in which we outlined the concept to be researched in combination with the academic and managerial relevance of the concept. In the second chapter we construct the theoretical framework that will represent the basis of the empirical analysis in the later sections. First we analyze the current literature to In chapter 3, having laid out the theoretical framework, we thoroughly describe the dataset and the variables we use to define the concepts that we want to analyze. Furthermore, based on our variable selection and data structure, we select the most suitable method of analysis to attain the desired results and answers to our research questions. In the fourth chapter we summarize the analysis and findings of the research in order to answer the empirical questions. Furthermore, we extend our model with several robustness checks to ensure that our findings are actually meaningful and not a result of flaws in our methodology or reasoning. Finally, in the fifth chapter, we generate conclusions and recommendations based on the findings of the study. We use this chapter to provide football club management with deeper insights into the drivers of their team performance at home, and how these can be influenced by management. Additionally, we discuss the limitations of this study and provide a guideline for possible future research in this area to improve our understanding of home advantage even further.

## Theoretical Background

## Literature Review

## 2.1.1 The role of home advantage in sports

Home advantage has been widely studied in sports literature. One of the first to formally document the existence of a certain home advantage in sports were (Schwartz & Barsky, 1977). They find that home advantage exists in varying degrees across different sports. Furthermore in their research they suggest that the major contributor to home advantage is social support, opposed to familiarity and fatique for away teams. (Nevill & Holder, 1999) support this claim as they produce similar results in their analysis of home advantage in English and Scottish football matches. In a related study, (Carron & Agnew, 1994) examine crowd effects specifically to provide an overview of how crowds influence match outcomes. One of their main findings is a signicant positive relationship between home advantage and crowd density. In other words, more crowd support leads to a stronger home performance relative to away performance and consequently a higher chance of a home win than an away win.

One of the landmark studies in the area of home advantage in sports is the research conducted by (Courneya & Carron, 1992). In their paper they provide a comprehensive literature review on home advantage in sports, their framework has been used extensively in subsequent studies. Their framework on home advantage revolves around location factors that influence players, coaches and officials(referee) psychologically. Correspondingly behaviour of players, coaches and officials is influenced by their psychological state, the change in behaviour in turn alters the performance of players, coaches and referees. The four main factors behind home advantage in their framework are crowd factors, learning(familiarity) factors, travel factors and rule factors(specific rules that favor home team in some sports). (Carron & Hausenblas, 1997) apply this framework to assess the occurrence of home advantage. They claim that home advantage is relatively stable across time and is generalizable across team/individual sports, gender and professional and amateur sports.

Despite the increased complexity of models and data used, (Carron, Longhead & Bray, 2005) find that the framework is still relevant in the evolving situation, assessing that the framework still serves as bases for a majority of the research conducted on home advantage in sports.

## 2.1.2 The role of home advantage in football

(Pollard, 2008) provides an excellent overview of the consensus of home advantage in football. Similar to the work of (Courneya & Carron, 1992) he drafts a framework of the major factors of which home advantages in football sterns from. Some of the more interesting findings according to his review include those of (Naeve & Wolfson, 2003 & 2004) as well as (Anderson, Neave & Wolfsson, 2007) who find hormonal differences between home and away team players in match build up, with higher levels of testosterone for players when playhing a home match versus playing an a way match. Similarly to hormonal reaction when defending territory. This hormonal change could then be attributed to increased performance on the pitch for the home team. (Pollard, 2006; Pollard & Seckin, 2007) further scrutinize the sense of territoriality and its effect on home and away performance and find indeed evidence of a possible influence on home advantage.

Another widely recognized possible component of home advantage us familiarity. Familiarity with specific home turf conditions such as altitude(Mc Sharry, 2007) and climate (Pollard, Da Silva & Nisio, 2008; Pollard & Seckin, 2007)could play a role in improved home team performance. In a similar fashion, artificial pitches seem to play a role in home advantage(Barnett & Hidditch, 1993).

The findings on the significance of travelling for home advantage are more ambiguous. (Pollard, da Silva & Nísio, 2008 ) find that travel distance seem to play a role albeit a minor one in conceiving of home advantage. Especially since their research is focusing on Brazil, where distances between teams can be very large. Distances that would impossibly exist in European countries. This is similar to what (Nevill&Holder) found in their research, where they found that distance only mattered when distance traveled was of such magnitude that different time zones would be crossed. (Goddard, 2006; Clarke & Norman, 1995) however, do seem to find a significant effect for travel distance on away team performance even with reasonably small distances between 0-200 miles.

One of the factors that the literature seems to have reached consensus on is the role of referee bias in home advantage. According to (Nevill, Balmer & Williams, 1999 ; Nevill, Balmer & Williams, 2002; Garicano, Palacios-Huerta & Prendergast, 2005) there exists consistently find evidence of a referee bias in favour of the home team probably due to social pressure from the crowd. In more recent research. (Endrich & Gesche, 2020) find that referees give less cards to home teams and more cards to away teams on average, which could be interpreted as a sign of referee bias in favour of the home side. However the exact source of this referee bias remains somewhat unclear. Research like ours on the incidence of referee bias in ghost games settings could provide useful in discerning whether or not crowd support plays a major role in referee decision making.

A final factor considered in the literature are tactical approaches to football matches as a driver of home advantage. There seems to be a difference in tactical approach between away and home teams according to (Pollard 1986; Pollard 2006) . For instance a more cautious approach from the away side might induce dynamics such as the beforementioned territorial factors influencing home and away performance. However, evidence on the significance of tactical approaches in determining home advantage is not conclusive. (Carmichael & Thomas, 2005; Tucker, Mellalieu, James & Taylor,2005) and should therefore be examined further.

Somewhere where the literature appears to agree on is the difficulty of disentangling each of the various forces driving home advantage. Pollard(2008) cites that struggles concerning the unraveling of individual factors effect on home advantage sterns from the phenomenon that multiple psychological and physiological influences involved all interact with eachother and possibly reinforce eachothers significance. For exactly this reason do “ghost games” provide such an unique opportunity to specifically study changes in home advantage under ghost games in order to disentangle crowd effects from other drivers of home advantage.

Our natural experiment setting in the covid pandemic provides an excellent opportunity disentangle different effects of home advantage. Previous research could only analyze detached cases of matches without supporters spread across leagues and time. However, the significant number of games played without a crowd following the covid-19 pandemic provide an excellent opportunity to systematically scrutinize crowd effects in isolation. Furthermore, examining whether a home advantage still exists without crowds could also shed light on the significance of the other before mentioned factors. Hypothetically, if home advantage is completely absent or even turned into a home disadvantage during absence of crowds, other factors such as travel effects or territorial effects are clearly not important drivers for the establishment of home advantage.

## 2.1.3 Home advantage during Covid-19

A few preliminary studies attempted to investigate the influence of crowd absence on home advantage by analyzing “ghost games” played in the German leagues after the restart of the season 2019-2020. For example (Thilp & Taller, 2020) find that home advantage has actually turned into a home disadvantage in case of “ghost games”. (Fischer & Haucap, 2020) also find that there seems to be a significant alteration in the strength of home advantage in the Bundesliga when crowd support is nonexistent. However, the German 2nd and 3rd leagues seem to be less affected by “ghost games” according to their analysis, due to lower occupancy rates on average in the lower tier leagues compared to the Bundesliga.

Early evidence thus seems to point to a significant drop in home advantage for teams without crowd support to back them in home games. And that the effect of crowd support is especially dependent on occupancy rates within a stadium, rather than absolute number of supporters. This makes sense as a large stadium only filled one thirdly seems less a motivating atmosphere than a small compact stadium completely filled with supporters close on the pitch.

Compared to these studies we choose to operate on an aggregate level where we aggregate data from matches in different countries into one comprehensive dataset to measure overall effects of covid-19 on home advantage and bookmaker margins. This is contrary to previously mentioned studies of home advantage in Corona which gravitated towards country specific analysis. We believe that such an aggregate analysis provides richer information on home advantage in general than single league studies where the results in that league might simply be a special outlier.

Additionally, by aggregating the analysis, single country anomalys and specialties such as scheduling bias will have less influence on the final conclusions. Scheduling bias could occur in a league where due to coincidence a disprorpotionate amount of strong home teams play versus weak away teams or vice versa. This could impact the proportion of home and away wins while being unrelated to covid-19. Furthermore by analyzing the extended period of “ghost games” up to now, thus including more data, we further reduce the impact of confounding variables such as scheduling bias. Previously conducted studies that analyzed partial seasons had serious problems with scheduling discrepancies. Currently it seems that for most league, the rest of the current season will also be played behind closed doors, which enables us to analyze an entire season worth of data behind closed matches.

## Theoretic Framework

## 2.2.1 The impact of crowd absence on team performance

1: We forecast that the exclusion of home crowds will have detrimental effects on the home advantage for home teams. A home crowd can be a positive stimulus for home team players and can create an intimidating and hostile environment for the opposition.( Ponzo & Scoppa, 2018). Every football fan will know about the famous European nights at Anfield where teams regarded superior in terms of quality found their waterloo against an inspired Liverpool side backed by the roar of the crowd. Similarly in Belgium teams often loathe the away game against Standard Liege because of the extremely passionate home fans creating a very aggressive atmosphere. These sentiments in football are confirmed by (Carmichael & Thomas, 2005), who find that home advantage is significantly higher in compressed and intense atmospheres compared to more open and wide areas where crowds might be more distant from the pitch. For example an athletic track around stadiums might decrease the effect of supporters on team performance. Additionally, (Carron & Agnew, 1994) report a positive relationship between crowd density and home performance. In a situation of “ghost games”, crowd density will of course be 0. Additionally(Tilp & Thaller, 2020) hypothesize that awareness of the absence of crowd support could boost away team morale and increase confidence in retrieving a result. This could as outlined before by(Pollard 1986; Pollard 2006) alter tactical approaches and mentality of both home and away teams. Therefore, based on previous findings of similar studies, we expect home performance to suffer significantly in the new situation of “ghost games” and as a result we predict a sizeable drop in the probability of a home win, all other things being equal.

This leads us to generate the following hypothesis regarding the effect of crowd absence on home advantage.

*H1: Home advantage ise less pronounced or possibly turns into a home disadvantage in a siutation where no crowd is allowed to attend the match.*

## 2.2.2 The impact of crowd absence on referee performance

2: Based on previous studies we expect that referee bias will be lower in matches without the influence of the home crowd. ((Endrich & Gesche, 2020; Nevill, Balmer & Williams, 2002; )Referees can be heavily influenced in their decision making by the heavy cheering of the crowd favouring the home team.(Unkelbach & Memmert , 2010). A few crucial decisions such as a controversial penalty or red card can significantly alter the course of a football game, and if the home team gets benefit of the doubt it could significantly home teams win their games. Other work by (Sutter & Kocher, 2004; Garicano, Palacios-Huerta & Prendergast, 2005) suggest that referees tend to award more extra time at the end of the first and second half if the home team is behind. The infamous “Fergie time” illustrates this perfectly. During the tenure of Sir Alex Ferguson at the helm of Manchester United, his side often received longer extra time in home matches, which more than once led to a late escape for the home side. In a similar fashion,(Neville & Holder, 1999; Bokyo, 2007) in their paper pose that crowds compel referee’s such that their decisions subconsciously favor the home team. (Endrich & Gesche, 2020) find that referees give more penalties to home teams, and more red and yellow cards to away teams. Therefore, in the case of “ghost games” we expect that referees will be less subject to social pressure of the crowd and consequently, less biased in decision making. This could benefit away teams in their chase to seize a result away from home. This leads to the following hypothesis on the effect of the absence of crowds on referee performance.

*H2: Crowd absence will positively influence referee performance, or in other words, reduce referee bias towards home teams.*

A schematic overview of the current body of literature and our contribution to it is presented in Table 1.

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Paper |  |  |  |  |  |
|  | General home advantage | Natural experiment covid-19 | Referee Bias | Crowd Occupancy | Country aggregate |
| Boyko, Boyko, & Boyko, (2007) | ✓ | 🗶 |  |  | 🗶 |
| Pollard(2008) | ✓ | 🗶 |  |  | ✓ |
| Schwartz & Barsky, 1977) | ✓ | 🗶 |  |  | ✓ |
| (Courneya & Carron, 1992). | ✓ | 🗶 |  |  | ✓ |
| Inan, T. (2020) | ✓ | 🗶 |  | ✓ | ✓ |
| Nevill&Holder(1999) | ✓ | 🗶 |  |  | 🗶 |
| Pollard(2006) | ✓ | 🗶 |  |  | 🗶 |
| Tilp&Thaller (2020) | ✓ | ✓ |  |  | 🗶 |
| Fischer & Haucap, (2020) | ✓ | ✓ |  |  | 🗶 |
| Endrich& Gesche, (2020). | ✓ | ✓ |  |  | 🗶 |
| This study | ✓ | ✓ | ✓ | ✓ | ✓ |

## 2.3 Conceptual Model

Table 2 shows the conceptual model we establish based on the current literature and hypotheses. This conceptual model will be used in later stages to build the correct model to analyze the data.

Table 2

Referee Bias

Team Performance

Occupancy rate

Crowd Support

Player psychology/physiology

## Data and Methodology

## Data collection

The website: <http://www.football-data.co.uk/data.php> has weekly updated datasets for all the important leagues around the world. The datasets include statistics on full-time and half-time results, amount of shots and shots on targets of both home and away teams as well as the number of yellow and red cards and corners for each team. For each of the major 11 leagues in Europe there is historical data available up to the season 1994/1995. All the datasets are split per season, per country and per league level in countries with multiple tiers. We aggregate all these seperate datasets into a large final dataset since we are primarily interested in overall home advantage not necessarily country level differences.

We need to include data on team strength in order to control for scheduling bias in our analysis. The most comprehensive index of team strength as fas as we know is the SPI index used by : project538. Their SPI index is constructed by ….. Their website provides weekly updated dataset on historical matches with SPI data, which can be merged with our previously constructed dataset with historical match outcomes and betting odds. Furthermore, we need to control for … and …. in our model, which can be done by including a measure of … into our analysis. Adding … and … into our dataset will aid in controlling for … and … . Finally a proxy for … will be incorporated into the analysis to increase the … of our analsyis.

## Variable operationalization

**SPI\_team\_strength:** this variable is calculated by … and … and is an advanced metric of team strength of both the home and away teams. The metric incorporates recent results and the strength of the teams played previously. It therefore is also a measure of recent form, with a string of bad results against lesser rated teams lowering the rating of a team possibly.

**Last 3 games:** This number captures the recent form of teams involved by calculating the number of points collected in the last 3 matches. Minimum value is 0, maximum value is 9.

**Last 3 games(home):** this number the captures the recent form of teams involved by calculating the number of points collected at home(away) by the home(away) team in the last 3 matches.

**Last 5 games:** This number captures the recent form of teams involved by calculating the number of points collected in the last 5 matches. Minimum value is 0, maximum value is 15.

**Last 5 games(home):** this number the captures the recent form of teams involved by calculating the number of points collected at home(away) by the home(away) team in the last 5 matches.

**Rating\_difference;** this variable is calculated by subtracting the away team spi rating from the home team spi rating. A large positive(negative) value entails a large quality difference in favour of the home(away) team. In case of a large quality difference it is thus more likely for a certain team to win the match.

**Team average age:** This variable captures the average age of the entire squad at the beginning of the season.

**Old or young:**  this variable captures the differences in age between the home and away team, where a positive value implies the home team have an older squad on average than the away team, and a negative value implies the away team have older squad players.

**Importance of match:** This variable is calculated by … and …. and represents the importance of the match outcome on the league developments for both teams their perspective. There are 2 values, one value for the away team and one value for the home team

**Spread importance:** this variable is calculated by subtracting the away importance value from the home importance. A positive value thus can be viewed as a match where the match is relatively more important for the home team whereas a negative value implies the contrary. Values close to 0 signify a situation where the match is of approximately equal importance to both teams.

**Home:** a dummy variable that states whether or not the match entry in the dataset is viewed from the home team perspective or the away team perspective. Each match within the data-set is double coded , once from the home-side perspective and once from the away-side perspective. A 1 is associated with an entry where the home team is the focal team whereas a 0 is associated with the mirrored entry where the away team is the focal team.

**Covid:** a dummy variable that indicates whether or not the game was played before or after the start of the covid-19 pandemic. A value of 1 represents matches that are played behind closed doors, with no spectators attending, whereas a 0 implies fans were present.

**Shirt colour:** A variable indicating the shirt colour of the home team.

**Crowd occupancy:** The ratio of the number of attendants to the number of seats in the stadium. Due to data collection and computation limits, the crowd occupancy is calculated as the total number of attendants within a league divided by the average capacity for all teams in the league. For example a league where the clubs on average have a stadium capacity of 50.000 with 4.000.000 fans in total visiting 200 matches means an average attendance of ((4000000/200))/50.000 = 0.4.

**Expected goals:** Expected goals are measured as the expected goals that would have been made from a teams chances within a match, based on average player and average situation. It is a sum of the quality of chances created by a team. A shot for open goal with no opponent player in tackling distance has a very high xg chance since it will be converted into a goal almost all the time. A 40 yard shot that gets deflected by a defender on the way and flies in with luck has a low xg value, as such shots have a low probability of producing a goal. Expected goals metric thus takes out luck out of the performance of a team and can be seen as a good metric of match performance and match outcomes.

**Spread in fouls:** The number of fouls committed by the home team - the number of fouls committed by the away team. A positive number means the home team has committed more fouls than the away team in the specific match, and a negative number implies the away team committed more fouls. 0 implies an equal number of fouls for both teams. Based on a non biased referee and controlling for team dominance and team quality we expect the number of fouls for home and away teams to not be significantly different from eachother.

**Spread in yellow cards :** The number of yellow cards for the home team - the number of yellow cards for the away team. A negative number can be interpreted as the away team receiving more yellow cards than the home team. A positive number means the number of yellow cards received by the home team exceeds those of the away team. 0 implies an equal number of yellow cards for home and away team. This could be 0 for both but also 1 or 2 or more for each side. Based on a non biased referee and controlling for team dominance and team quality we expect home and away teams to have no significant difference in the number of yellow cards.

**Spread in red cards:** The number of red cards for the home team - the number of red cards for the away team. A negative number implies the away team received more red cards than the home team. A positive number signifies the number of red cards received by the home team exceeds those of the away team. 0 implies an equal number of red cards for home and away team. This could be 0 for both but also 1 or 2 for each side. Based on a non biased referee and controlling for team dominance and team quality we expect home and away teams to have equal number of red cards on average

**Absolute crowd size:** The absolute number of supporters present at the game**.**

**Corners:** The amount of corner kicks for a team within the match. This can be seen as a measure of attacking dominance since corners are often results of defending clearances or goalkeeper saves.

**Shots on target:** The number of shots undertaken by a team that without interference of any player would be a goal. Examples include a shot that would have been on target but was blocked by an opponents defender, or a shot saved by the keeper that was going in the goal otherwise.

**Shots/shots on target ratio:** The ratio between the total number of shots undertaken by a team to the total number of shots on target from that same team. A higher ratio means better quality shooting of a team and more efficient use of shots.

**Shots:** The total number of shots undertaken by a team within the match. This includes all of shots wide, shots blocked, shots on target, shots on woodwork.

**Referee:** A variable that captures the referee present to officiate the game.

**Home Goals:** This variable denotes the number of goals scored by the home team in the specific games

**Away Goals:** This variable denotes the number of goals scored by the away team in teh specific game.

**Points home:** The number of points for the home team in this specific match.

**Points away:** The number of points for the away team in this specific match.

**Percentage Points home from total:** Home points as a fraction of total points obtained in the game. A value above 50 percent implies a home advantage as more points are collected on average at home

**Points away from total:** Away points as a fraction of total points obtained in the game. A value above 50 percent implies a home disadvantage as more points are collected away on average than at home

**Difference in goals home/away:** DIfferences in goals scored by the home and away team per match. A 4-0 home win implies a value of +4, whereas a 3-0 away win implies a value of -3. A draw signifies a value of 0. A negative value implies more goals scored by home teams than away teams, although not necessarily more home wins than away wins.

**Difference in points home/away:** the difference in points for the home and away team for a single match, calculated as points of the home team - points of the away team. A home win means a +3 difference, whereas an away win gives a -3 difference. A draw results in 0. Thus a overall positive number for this metric can be interpreted as more home than away wins, whereas a negative number implies more away than home wins

**Difference in expected goals home/away:** Difference in the number of expected goals for home and away teams. A match where the home team had chances created such that their expected goals totaled up to 2.93, and the away team had expected goals of 1.39 means a spread of 2.54, the higher this spread the more dominant the home team was and the more chance the home team had of winning the match. Vice versa, a large negative value signifies an away team that dominated the match and chances and had a high probability of winning the match.

**Difference in expected points home/away:** The expected points are calculated based on the expected goals of the match. These expected goals are then measured on a poisson distribution to calculate the probability of each result given these expected goals. For example, a 2.02/1 expected goals distribution for a certain match results in a .. probability for a home win, … probability for an away win and a … percent change of a draw.

**% of total expected points won at home:** The fraction of expected home points divided by the sum of expected home and away points. Over 50 percent implies a home advantage based on expected points.

**% of total expected points won away:** The fraction of expected away points divided by the sum of expected home and away points. Over 50 percent implies a home disadvantage based on expected points.

**Home goals HT:** The number of goals scored by the home team in the first halfe

**Away goals HT:** number of goals scored by the away team at half-time

**HT result:** whether the first half result would have resulted in a home win, draw or away win.

**Referee bias:** since referee bias is a latent concept and not easily capturable in a single number, we use SEM to construct a latent variable Referee bias: which takes in multiple numbers on the referee assessment of home and away teams. Based on a non biased referee and controlling for team dominance and team quality we expect home and away teams to have equal number of fouls/red cards and yellow cards.

**Team performance home advantage:** Team performance is difficult to capture in a single variable, goals can be result of pure luck. By combining, goals, expected goals, corners, shots, shots on targets into a single measure, we obtain a richer definition of team performance, to better capture the different dynamics behind team performance.

**Outcome home advantage:** a variable constructed by combining the different measures of match outcomes to obtain a better measurement of outcome. It combines, points, expected goals, goals and expected points to get a very rich definition of home advantage.

## Methods of analysis

## Descriptive statistics

Table … provides the descriptive statistics for the variables mentioned in section …

## Regression Equation

As mentioned in section 3.3, we deem … as most suitable method for analysis of our dataset. Therefore we use a … regression on our variables of interest. … is a statistical method that is based on the … estimation of .. . relationship between … and … … estimates the relationship between the .. Dependent variable(s) and … Independent variables by …

Our main model is illustrated by the following equation:

# 6 Bibliography

1. Agnew, G. A., & Carron, A. V. (1994). Crowd effects and the home advantage. International Journal of Sport Psychology
2. Allen, M. S., & Jones, M. V. (2014). The home advantage over the first 20 seasons of the English Premier League: Effects of shirt colour, team ability and time trends. *International Journal of Sport and Exercise Psychology*, *12*(1), 10-18.
3. Attrill, M. J., Gresty, K. A., Hill, R. A., & Barton, R. A. (2008). Red shirt colour is associated with long-term team success in English football. *Journal of sports sciences*, *26*(6), 577-582.
4. Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. Journal of sports sciences, 25(11), 1185-1194.
5. Brechot, M., & Flepp, R. (2020). Dealing with randomness in match outcomes: how to rethink performance evaluation in european club football using expected goals. Journal of Sports Economics,
6. Carmichael, F., & Thomas, D. (2005). Home-field effect and team performance: evidence from English premiership football. Journal of sports economics, 6(3), 264-281.
7. Carron, A. V., Hausenblas, H. A., & Eys, M. A. (2005). Group dynamics in sport. Fitness Information Technology.
8. Carron, A. V., Loughhead, T. M., & Bray, S. R. (2005). The home advantage in sport competitions: Courneya and Carron's (1992) conceptual framework a decade later. Journal of sports sciences, 23(4), 395-407.
9. Clarke, S. R., & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. Journal of the Royal Statistical Society: Series D (The Statistician), 44(4), 509-521.
10. Courneya, K. S., & Carron, A. V. (1992). The home advantage in sport competitions: a literature review. Journal of Sport & Exercise Psychology, 14(1).
11. Endrich, M., & Gesche, T. (2020). Home-bias in referee decisions: Evidence from “Ghost Matches” during the Covid19-Pandemic. Economics Letters, 197, 109621.
12. Fédération Internationale de Football Association. (2007). FIFA Big Count 2006: 270 million people active in football. Zurich: Fédération Internationale de Football Association.
13. Fédération Internationale de Football Association.(2019), FIFA activity report 2018. Retrieved from https://www.fifa.com/who-we-are/official-documents/
14. Fischer, K., & Haucap, J. (2020). Does crowd support drive the home advantage in professional soccer? Evidence from German ghost games during the COVID-19 pandemic.
15. Galariotis, E., Germain, C., & Zopounidis, C. (2018). A combined methodology for the concurrent evaluation of the business, financial and sports performance of football clubs: the case of France. Annals of Operations Research, 266(1), 589-612.
16. Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. Review of Economics and Statistics, 87(2), 208-216.
17. Goddard, J. (2006). Who wins the football?. Significance, 3(1), 16-19.
18. He, M., Cachucho, R., & Knobbe, A. J. (2015, September). Football Player's Performance and Market Value. In Mlsa@ pkdd/ecml (pp. 87-95).
19. Inan, T. (2020). The Effect of Crowd Support on Home-Field Advantage: Evidence from European Football. Annals of Applied Sport Science, 0-0
20. McSharry, P. E. (2007). Altitude and athletic performance: statistical analysis using football results. Bmj, 335(7633), 1278-1281.
21. Metelski, A., & Kornakov, K. (2021). Effect of lockdown owing to COVID-19 on players’ match statistics in Bundesliga. Journal of Physical Education & Sport, 21(1)
22. Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. European Journal of Operational Research, 263(2), 611-624.
23. Neave N, Wolfson S. Testosterone, territoriality, and the 'home advantage'. Physiol Behav 2003; 78: 269-75.
24. Neave N, Wolfson S. The home advantage: psychological and physiological factors in soccer. In: Lavalle D, Thatcher J, Jones MV, Eds. Coping and Emotion in Sport. New York, Nova Science Publishers 2004; 127-44.
25. Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport. Sports Medicine, 28(4), 221-236.
26. Nevill, A. M., Balmer, N. J., & Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. Psychology of Sport and Exercise, 3(4), 261-272.
27. Nevill, A., Balmer, N., & Williams, M. (1999). Crowd influence on decisions in association football. The Lancet, 353(9162), 1416.
28. Nevill, A.M., Holder, R.L.(1999). Home Advantage in Sport. Sports Med 28, 221–236
29. Pinnuck, M., & Potter, B. (2006). Impact of on‐field football success on the off‐field financial performance of AFL football clubs. Accounting & Finance, 46(3), 499-517.
30. Pollard R, Seckin A. Why is home advantage in South-east Europe the highest in the world? In: Pollard R. Worldwide regional variations in home advantage in association football. J Sports Sci 2006; 24: 231-240.
31. Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. Journal of sports sciences, 4(3), 237-248.
32. Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. The open sports sciences journal, 1(1)
33. Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876–2003). Journal of sports sciences, 23(4), 337-350.
34. Pollard, R., Silva, C. D., & Medeiros, N. C. (2008). Home advantage in football in Brazil: differences between teams and the effects of distance traveled. Revista Brasileira de Futebol (The Brazilian Journal of Soccer Science), 1(1), 3-10.
35. Pollard, Richard,(2006), Home Advantage in Soccer: Variations in its Magnitude and a Literature Review of the Inter-related Factors Associated with its Existence , Journal of Sport Behavior, Vol. 29, Iss. 2, 169-189.
36. Ponzo, M., & Scoppa, V. (2018). Does the home advantage depend on crowd support? Evidence from same-stadium derbies. Journal of Sports Economics, 19(4), 562-582.
37. Rodríguez, P., Humphreys, B. R., & Simmons, R. (Eds.). (2017). The Economics of Sports Betting. Edward Elgar Publishing.
38. Samagaio, A., Couto, E., & Caiado, J. (2009). Sporting, financial and stock market performance in English football: an empirical analysis of structural relationships. Journal Department of Management, ISEG/School of Economics and Management, Techincal University of Lisbon, Portugal.
39. Schwartz, B., & Barsky, S. F. (1977). The home advantage. Social forces, 55(3), 641-661.
40. Sutter, M., & Kocher, M. G. (2004). Favoritism of agents–the case of referees' home bias. Journal of Economic Psychology, 25(4), 461-469.
41. Tilp, M., & Thaller, S. (2020). Covid-19 has turned home-advantage into home-disadvantage in the German Soccer Bundesliga. Frontiers in Sports and Active Living, 2, 165
42. Tucker, W., Mellalieu, D. S., James, N., & Taylor, B. J. (2005). Game location effects in professional soccer: A case study. International Journal of Performance Analysis in Sport, 5(2), 23-35.
43. Unkelbach, C., & Memmert, D. (2010). Crowd noise as a cue in referee decisions contributes to the home advantage. Journal of Sport and Exercise Psychology, 32(4), 483-498.
44. Wolfson S, Neave N, Anderson M. Hormones and the home advantage in English football. In:
45. Wolfson S, Neave N. Preparing for home and away matches. Insight 2004; 8(2): 43-46.
46. Theodorakis Y, Goudas M, Papaioannou A, Eds. Book of long papers, 12th European Congress of Sport Psychology; 2007: Halkidiki, Greece: FEPSAC 2007; pp. 57-60.