
# Does the crowd influence the home team's performance? Evidence from English soccer matches during COVID-19.

Student: Jesse Bijleveld (512799)

Sports Economics

Thesis supervisor: Sam Hoey

Second reader: prof.dr.ir. Jan van Ours

Keywords: Sports Economics, COVID-19, home advantage, crowd attendance, social pressure

# Abstract

This research investigates the influence of the crowd on the home team's performance in professional soccer. We do this by analysing soccer matches in the English Premier League before and during the COVID-19 pandemic. During the pandemic no spectators were allowed in the stadium. We find that, during COVID-19, home advantage decreased and almost completely disappeared. Besides this, we find no difference in referee bias before and after COVID-19. Therefore, this paper confirms that the crowd does have an influence on the home team's performance.

# Contents

# 1 Introduction

Home advantage (HA) is an established phenomenon in sports and especially in soccer. Jamieson (2010) found that HA is present in all sports competitions. However, HA in soccer is considerably higher than in other sports. If we look at soccer clubs, some clubs stand out in terms of atmosphere in the stadium and crowd fanaticism. We find a perfect example close by, namely the supporters of Feyenoord. They call themselves 'het legioen' (the legion) and are really seen as the 'twelfth man'. Therefore, the squad number 12 is never assigned to a player but reserved for 'het legioen'. They are famous for their fanaticism and for creating a supportive atmosphere for the players.

During the COVID-19 pandemic, many soccer enthusiasts and analysts complained about the lack of atmosphere in the stadium. Therefore, clubs came up with artificial crowd noises which were played through the stadium speakers. Still, this was not comparable to real crowd noises. Many researches, which we will discuss in the next section, found that HA decreased or even fully disappeared in all European soccer competitions during the COVID-19 restrictions. Besides this, soccer scores became more surprising and thus more unpredictable (Reade, 2020). Deutscher, Winkelmann, and Ötting (2020) researched an implication of this. They compare the last games in the German Bundesliga of the 2019/2020 season, which were played without spectators, with the last matches of the 2018/2019 season. They find that, especially in close competitions, away teams won considerably more matches than the bookmakers expected. They show that bookmakers experienced difficulties adapting to this and showed that opportunities for strategic betting strategies arose.

This paper will answer the following research question: "What is the influence of the crowd on the home team's performance in professional soccer?" We will conduct this research by comparing the probability of a home win and home advantage a season before the COVID-19 break and during the COVID-19 restrictions. This research will focus on the English Premier League (EPL), the top soccer division of England.

All recent researches were conducted at the end of 2020 or the beginning of 2021. Therefore, they had only data of approximately the last ten matches of the 2019/2020 season that were played without any spectators. To improve upon the reliability of the data analysis, this research will use the most recent match data for this paper. Because of this, this research will be based on data of approximately 50 matches per club without any spectators.

The results will answer questions about the fairness of sports and soccer. For example, whether banning fans at the stadium as a sanction for bad behavior is fair. Furthermore, it is also relevant for the club's managers to know if a filled stadium and great atmosphere has an effect on the home team's results. If this is the case, they could for example adjust their ticket pricing to this.

This research will contribute not only to current knowledge and research into sports and specifically soccer. This research will mainly contribute to the economic literature. Namely, this research could give an insight into social pressure and the influence a large group or crowd has on others regarding motivation, egoism, or concentration. More in general, this could give an insight in a person's, or group's decision-making while under (social) pressure. Furthermore, it could give insight into how groups perform when they are encouraged but also criticized by a large group or even public opinion.

However, little research has been done on the effect of social pressure on one's or a team's performance. Garicano, Palacios-Huerta, and Prendergast (2005) found that referee's favour the home team under social pressure. Mazer, Barnes, Grevious, and Boger (2013) found that a verbal aggressive coach has a negative influence on the motivation of athletes. Furthermore, Fransen, Boen, Vansteenkiste, Mertens, and Vande Broek (2018) find that not only coaches, but also athlete leaders can have a positive influence, through motivational feedback, on their fellow athletes' performance and intrinsic motivation. However, no research has been conducted which looked at the relationship between social pressure and a group's performance.

This paper is structured as follows. First, in the empirical evidence section, we will discuss and give an overview of relevant literature. Hereafter we will discuss the data set we used for this research. Besides this, we will explain our measures of home advantage and we will provide descriptive match statistics. After that, in the methodology section, we will discuss the methodology that we will use in this research and explain the regression formula's. Hereafter, we will show the results of the regression analyses and discuss these results. After this, the findings of this paper will be discussed in the discussion and conclusion section. We finish this paper with suggestions for further research.

# 2 Empirical evidence

Home advantage (HA) has been researched in many different sports and sports competitions. Schwartz and Barsky (1977) researched the four major sports leagues in the United States. They found that the home advantage is most pronounced in indoor sports such as basketball and ice hockey. In outdoor sports, such as baseball and (American) football, the home advantage is less pronounced. The relationship between HA and crowd presence has been researched by Agnew and Carron (1994). They found that, in Major Junior-A ice hockey, the crowd density was significantly related to HA. Not only clubs' HA has been researched, Balmer, Nevill, and Williams (2001) for example, examined whether host countries experienced HA in the Winter Olympics. For sports that were subjectively assessed by judges, they found more significant HA. They observed better home performance in sports that are subjectively judged than in sports that are not judged. For visiting athletes and teams, they observed no differences in performance between judged and non-judged sports. This suggested that judges were scoring home competitors disproportionately higher than away competitors.

The EPL has been studied many times in the field of HA. Pollard (1986) researched home advantage in the top English division from 1888 till 1984. He found that crowd support and travel fatigue are unimportant. The effect of referee bias, local conditions, and team tactics are unclear. Clarke and Norman (1995) researched HA in the EPL and designed a method to measure HA and disentangle this from the team's quality. The results showed no significant division effect but did show significant year effects. Besides this, he also found that London clubs had below-average HA. Recently, Peeters and van Ours (2021) researched HA in the top four English soccer divisions from 1974 until 2018. They build upon Pollard's (1986) research and they use Clarke and Norman's measures of HA, which they call relative and absolute HA. They find that the absolute HA is significant and equal among divisions and it declined over time. Relative HA is positively related to within-team variation in attendance and the use of an artificial pitch.

One of the possible causes of HA is a referee bias. If a referee bias exists, the favoured team is likely to benefit from this. Červenỳ, van Ours, and van Tuijl (2018) research the implications of this. They find that the team that receives a red card, has their goal-scoring rate decreased by 47%. At the same time, the unsanctioned team's probability to score a goal increases by 124%. Garicano et al. (2005) study the effect of social pressure on the referee's decision-making in the top Spanish soccer league. They argue that soccer referees favour home teams in order to satisfy the crowd in the stadium. They show that referees systematically favour home teams by shortening tight matches when the home

team is winning and lengthening tight matches when the home team is losing. A different way of referee bias is studied by Sutter and Kocher (2004). They study the behavior of referees in the German Bundesliga in the 2000/2001 season. They find that referees are more likely to award penalties to the home team and that away teams are refused a legitimate penalty more often. In line with Garicano et al. (2005), they find that in the German Bundesliga, referees add considerably more extra time when the home team is behind by one goal than when it is ahead by one goal or a draw after 90 minutes.

Furthermore, yellow and red cards awarded can be used as indicator for a referee bias. Dawson, Dobson, Goddard, and Wilson (2007) research yellow and red cards in the EPL from 1996 until 2003. They find that away teams received more cards. However, they argued that this could not be attributed to the HA but to the referee favouring the home team. Boyko, Boyko, and Boyko (2007) examine whether individual referees vary in their home bias or whether biased decisions contribute to overall home advantage in the EPL from 1992 until 2005. They find that referees vary significantly in their yellow card and penalty differentials. However, Johnston (2008) replicated the latter study for the 2006/2007 EPL season and found no statistical evidence for their claims. On the contrary, he confirmed Boyko and colleagues' finding that the away team's strength had a significant effect on HA. A recent study by Bryson, Dolton, Reade, Schreyer, and Singleton (2021) exploits the COVID-19 pandemic to research the effect of the crowd on refereeing decisions. They find that without a crowd, away teams receive 1/3 fewer cards. They suggest that these results are causally due to a lack of crowd pressure which usually influences the referee's decision-making in the favour of the home team.

Another source of home advantage can be familiarity with the stadium and crowd presence. Pollard (2002) researched familiarity with data from club's that moved to a new stadium. He used data from professional baseball, basketball, and ice hockey in North America. In all three sports, he found that HA was significantly lower in the first season in the new stadium than in the last season in the old stadium. Crowd factors such as crowd size and crowd density did not appear to have an effect. Boyko et al. (2007) find the opposite of Pollard. They show that, besides referees, crowd size significantly influences HA. However, after controlling for home team, crowd size was not significant. This suggests that the relationship between crowd size and HA is not as straightforward as is often assumed. Peeters and van Ours (2021) do not use total crowd size or density but use relative attendance in their research. They find a positive relationship between relative attendance and their a measure of relative HA in the EPL. Krumer (2020) researched whether different kick-off times affect the attendance and performance of teams in the

group stage of the UEFA Europa League. He found that lower-ranked teams who host games that begin at 21:05 have significantly lower attendance than games that kick off at 19:00. This results in lower HA at the latter games. However, the HA of the higher-ranked teams is not affected by the kick-off. This would suggest that lower-ranked teams depend more on the crowd's support than higher-ranked teams.

Yet another cause of HA can be found in fatigue and rest days between matches. Scoppa (2015) investigates the role of fatigue in soccer with data on matches in the FIFA World Cup and UEFA European Football Championship. The results show that there are no relevant effects of enjoying different days of rest on team performance. However, he shows that when the rest time of one of the opposing teams is equal to or less than three days, the advantage of additional days of rest is quite relevant. Therefore, it seems that there is a threshold of about three days under which rest time is essential. Krumer and Lechner (2018) build upon these rest periods between matches. They research the effect of midweek matches on the home advantage in the German Bundesliga. They find that the HA completely disappears in these games and thus the home teams perform significantly worse. Goller and Krumer (2020) find similar results. They research matches played on non-frequent days in the top German, Spanish, French, and English soccer leagues between 2007/2008 and 2016/2017. They find that all four leagues have a lower attendance in games that take place on four non-frequently played days than those on three frequently played days. Besides this, they also find a significantly lower home advantage for the underdog teams on non-frequent days.

Another interesting perspective for research into HA can be found in (same-stadium) derbies. Seckin and Pollard (2008) researched the effect of derbies matches on HA for clubs from Istanbul. They researched 12 seasons in the Turkish Super League. They find that home teams won 57.7% of the total points in derbies while home teams won 61.5% of the points in all matches. Ponzo and Scoppa (2018) take the research into derbies a step further. They investigate 128 same-stadium derbies in the Italian Serie A from 1991/1992 to 2012/2013. They find that, in derbies, the home team scores 0.45 goals more than the away team and the probability that the away team wins is 15 percentage points higher. They also confirm referee home bias. By taking this as a constant, they find that the crowd's support directly influences the home team's performance.

The research into same-stadium derbies is expanded by Van de Ven (2011). He looks at both same-stadium derbies and empty stadiums, due to safety regulations, in the Italian Serie A and Serie B in the 2006/2007 season. He found that HA advantage exists for

the home team, regardless of whether a crowd is present. Furthermore, he found that HA does not exist in derby games, even if the home team has the most crowd support. These findings suggest that crowd support is not a necessary condition for a HA to occur and that other factors cause HA. Pettersson-Lidbom and Priks (2010) exploit the same situation as Van de Ven (2011) in Italy where, due to safety regulations, spectators were not allowed in the stadium. They find surprisingly large and significant evidence that the home teams are favoured in games with spectators compared to games without spectators. This bias varies between 20% to 70% depending on the type of punishment. Besides this, they argue that home and away teams are not affected by pressure from the spectators. Reade, Schreyer, and Singleton (2020) investigate matches that are played behind closed doors in the UEFA Champions League, UEFA Europa League, French Ligue 1, the top three Italian leagues, and the Coppa Italia (Italy Cup). They find that HA almost completely disappeared when there was no crowd present. Besides this, away teams received less yellow cards per foul. This suggests that normally the referee is affected by social pressure, but without a crowd this pressure disappears.

Since the COVID-19 pandemic, many researches has been conducted on sports, but specifically soccer, matches that were played in an empty stadium. Fischer and Haucap (2020) researched the effect of the crowd on home advantage during the COVID-19 pandemic. They did this by researching the last games, so-called 'ghost' games, in the 2019/2020 season in the top German leagues, where no crowd was present. They find a reduced home advantage in the first division, but they found no significant decrease in home advantage in the second and third league. They also argue that this decrease in home advantage has been less dramatic for clubs that always had lower attendance rates before the COVID-19 break. They show that the 'ghost' game effect decreased over time and they suggest that players adapt and get used to the new situation. Tilp and Thaller (2020) continue on Fischer and Haucup's findings. They argue that the home advantage in the German Bundesliga turned into a home disadvantage during the COVID-19 restrictions. They found that there were more home losses than home wins. They also found a referee bias. Home teams received significantly fewer yellow and red cards in matches without spectators compared to the matches with spectators.

R. Matos et al. (2021) used, among other methods, Pollard's (1986) method to research the effect of the absence of a crowd on home advantage in the first division of Portuguese soccer during the pandemic. They only found a significant difference between the last games in 2017/2018 and 2019/2020 while using their method for home advantage. They emphasize that other researchers should also use complementary methods for determining

the home advantage. McCarrick, Bilalic, Neave, and Wolfson (2020) researched all 15 European leagues that finished the 2019/2020 season without spectators in the stadium. They found that playing in an empty stadium had a significant negative effect on the home team performance. This is for example illustrated by the significant decrease in the home team's attacking opportunities. Besides this, they found that referees awarded significantly fewer cards against away teams.

# 3 Data

## 3.1 Data set

For this research, we use match-level data of the English Premier League (EPL) from the 2018/2019 until the 2020/2021 season. We obtain this from football-data.co.uk. This website and their data sets are designed for analysing soccer matches and using this data to predict match results for betting. The data set contains three seasons in which 380 matches were played, this makes the data set 1,440 observations.

Each match observation has variables on match events distinguished between home and away. Examples of these match events are the result (home win, draw, or home loss), goals scored, shots, shots on target, fouls committed, and cards received. In this data set, we add a dummy variable for the COVID-19 restrictions. The last match played before the COVID restrictions was on the 9th of March 2020. The EPL season restarted on the 17th of June 2020. From this date, COVID-19 restrictions were in force. Therefore, no crowd was allowed in the stadiums and the COVID dummy variable became 1.

The second data set that we will use for this research, is data of the final standings of the EPL from the 2018/2019 until the 2020/2021 season. We will split these final standings into home final standings, for all of the matches played at home, and away final standings, for all of the matches played away. With this data set, we will calculate and analyse different measures of home advantage. We will further explain these measures in the methodology section.

## 3.2 Measures of home advantage

As we mentioned above, this research will use three methods to measure and quantify home advantage. HA is not measurable on match level. Therefore, we will use seasonal home advantage as HA measure, this measure will be calculated on the final standings data as discussed in the data section.

The first method is the method for seasonal relative home advantage that Peeters and van Ours (2021) used in their research into HA and that is based on a method designed by Clarke and Norman (1995). This method first determines the absolute home advantage H for the whole league. First, we calculate the home goal difference (HGD) for each club. This is the goals scored at home minus the goals conceded at home. Then we divide the sum of all clubs' HGD's by the number of clubs in the league minus one. This gives us

the absolute leagues home advantage $H$.

$$H = \frac{\sum HGD_i}{(N-1)}$$

After this, the relative home advantage $h_i$ will be determined. For this, we subtract the away goal difference (AGD), which is is the goals scored away minus the goals conceded away, and H from HGD. We divide this value by the number of clubs in the league minus two. This gives us the club's relative home advantage $h_i$.

$$h_i = \frac{HGD_i - AGD_i - H}{(N-2)}$$

The second method that we will use to determine the home advantage will be Pollard's (1986) method. With this method we divide the points won at home by the total amount of points won. Here, 50% represents an equal amount of points won in home and away games and thus no home advantage or disadvantage. In this paper, we call this measure $ha_1$.

$$ha_1 = \frac{H}{H+A} * 100\%.$$

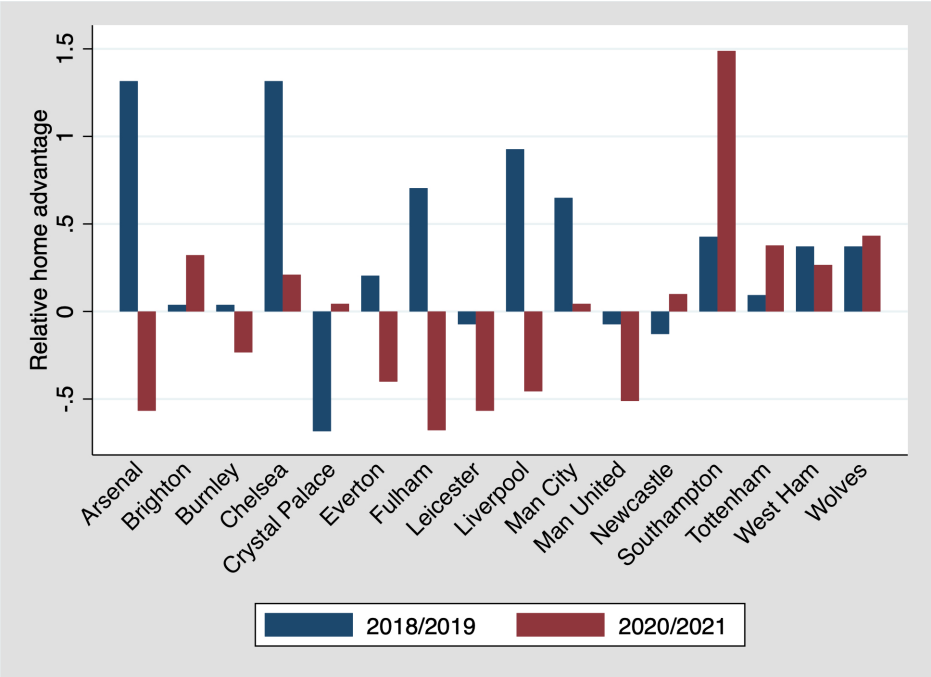
The third method is designed by R. M. Matos, Amaro, and Pollard (2019). They created this method because they noticed that the value obtained with Pollard's method does not properly relate to the advantage of playing at home. They preferred a method that gives a value of 0% if there is no home advantage, a negative value if there is a home disadvantage, and a positive value if there is home advantage. With this method, we divide the difference between the points won at home and way by the points won away. In this paper, we call this measure $ha_2$.

$$ha_2 = \frac{H-A}{A} * 100\%.$$

## 3.3 Descriptive statistics

Figure 1 shows the values of the home advantage measure $h_i$, by Clarke and Norman (1995), per team in the 2018/2019 and 2020/2021 season. Sixteen clubs have played in both season and are therefore suitable to compare. We can see that the larger clubs as for example, Arsenal, Chelsea, Liverpool, and Manchester City saw a decrease in HA. Some smaller clubs as for example, Southampton, Brighton, and Wolverhampton Wanderers seem to have benefited from playing in an empty stadium.

Figure 1: Measure of relative HA by Clarke and Norman (1995) compared between teams




Notes: Figure 1 shows the HA that the different teams in the EPL experienced before and after the COVID-19 break. The measure of HA we used is the one designed by Clarke and Norman (1995).

Table 1 column 1 shows the means of 668 matches played from the 2018/2019 season and 2019/2020 season before the COVID-19 break. Column 2 shows the means of 472 matches played after the COVID break in the 2019/2020 season and the 2020/2021 season. Column 3 shows the difference between columns 1 and 2 and the significance of this difference. From this, we can conclude that after COVID, significantly fewer matches were won by the home team. We find no significant evidence that there were more or less draws. However, there is a significant increase in home losses after the COVID break.

Looking at the goals statistics, we see that the home team did score less ($\alpha = 10\%$) goals after the COVID break. For away teams, the goals scored did not significantly change. Therefore, the goals difference is also slightly significantly significant, namely 0.225 ($\alpha = 10\%$) in favour of the away team. Regarding the home team's shots made, home shots have increased significantly ($\alpha = 1\%$) with 1.343 shots made less after the COVID break than before. For the away team, there is no significant evidence that this has changed. Looking at the shots at target, we see that the home team made slightly less shots on target ($\alpha = 10\%$). Again, there was no evidence that the shots at target changed for the away team.

Looking at the fouls statistics, we can see that the home team made significantly more fouls per game. The away team did not see any significant difference here. Remarkably, both the home team and away team received less significantly fewer yellow cards, respectively 0.140 ($\alpha = 5\%$) and 0.324 ($\alpha = 1\%$). Regarding red cards, we observe no significant difference. Looking at the total cards per foul received, an indicator for the referee's strictness, we see that significantly fewer cards per foul have been awarded. Remarkably, the coefficients for both the home and away team are the same, implying that there has been no change in referee's bias towards one of the teams. However, this referee bias should be taken into account and researched more thoroughly.

Table 1: Descriptive match statistics

| | Before COVID-19 | After COVID-19 | Difference |
|---|---|---|---|
| Home win | 0.464 | 0.396 | 0.068** |
| | (0.019) | (0.023) | (0.030) |
| Draw | 0.214 | 0.218 | -0.004 |
| | (0.016) | (0.019) | (0.025) |
| Home loss | 0.322 | 0.386 | -0.064** |
| | (0.018) | (0.022) | (0.029) |
| Home goals | 1.542 | 1.390 | 0.152* |
| | (0.049) | (0.062) | (0.079) |
| Away goals | 1.237 | 1.309 | -0.073 |
| | (0.046) | (0.057) | (0.074) |
| Goals difference | 0.305 | 0.081 | 0.225* |
| | (0.073) | (0.089) | (0.115) |
| Home shots | 13.981 | 12.638 | 1.343*** |
| | (0.022) | (0.249) | (0.332) |
| Away shots | 11.262 | 11.360 | -0.098 |
| | (0.189) | (0.231) | (0.298) |
| Home shots at target | 4.744 | 4.460 | 0.284* |
| | (0.100) | (0.120) | (0.156) |
| Away shots at target | 3.991 | 4.019 | -0.028 |
| | (0.092) | (0.105) | (0.140) |
| Home fouls | 10.207 | 11.206 | -0.999*** |
| | (0.133) | (0.157) | (0.206) |
| Away fouls | 10.563 | 10.682 | -0.119 |
| | (0.139) | (0.163) | (0.214) |
| Home yellow cards | 1.570 | 1.430 | 0.140** |
| | (0.049) | (0.050) | (0.070) |
| Away yellow cards | 1.766 | 1.443 | 0.324*** |
| | (0.047) | (0.054) | (0.071) |
| Home red cards | 0.052 | 0.051 | 0.002 |
| | (0.009) | (0.011) | (0.014) |
| Away red cards | 0.070 | 0.068 | -0.003 |
| | (0.010) | (0.013) | (0.016) |
| Home cards per foul | 0.162 | 0.134 | 0.028*** |
| | (0.005) | (0.005) | (0.007) |
| Away cards per foul | 0.180 | 0.153 | 0.028*** |
| | (0.005) | (0.007) | (0.009) |
| Delta cards per foul | -0.018 | -0.019 | 0.000 |
| | (0.007) | (0.009) | (0.011) |
| Observations | 668 | 472 | 1,140 |

Notes: Table 1 shows the means of the match statistics in columns 1 and 2. Column 3 shows the difference between these means and the significance. The standard error is showed between brackets. The significance is categorized as follows: *p<0.1; **p<0.05; ***p<0.01

# 4 Methodology

## 4.1 Home advantage regression analysis

To research the effect of the crowd on a team's performance, first, an analysis of the match statistics before and during the COVID-19 restrictions is necessary. A two-sided t-test in the data section showed whether the differences in match statistics before and during the COVID-19 pandemic. The matches that we will use for this will be from the 2018/2019, 2019/2020, and 2020/2021 seasons.

First, we will perform a simple linear regression, with robust standard errors, of the COVID dummy on the probability of a home win. The regression formula is as follows:

$$Y_{ijmt} = \beta_0 + \beta_1 COVID_{ijmt} + \alpha_i + \gamma_t + \rho_j + \epsilon_{ijmt}$$

Here, $Y_{ijmt}$ gives the probability of the home team winning the match. $COVID_{ijmt}$ is a dummy which is 1 if a match is played during the COVID-19 restrictions and 0 if it is played before the COVID-19 restrictions. $\alpha_i$, $\gamma_t$, and $\rho_j$ are fixed effects, capturing respectively team, season, and referee variances. The subscripts are: i for home team, j for referee, m for match, and t for season. The constant $\beta_0$ takes the value of the mean of the first club in our data set, in this case Arsenal, if only team fixed effects is included. If for example, team and season fixed effects is included in the regression, the constant takes the mean value of the 'first' club (Arsenal) in the 'first' season (2018/2019). This applies in the same way for other combinations of fixed effects and also applies to the regressions discussed below.

Additionally, we will perform a similar simple regression with team-season fixed effects instead of team and season fixed effects separately. The regression formula is as follows:

$$Y_{ijmt} = \beta_0 + \beta_1 COVID_{ijmt} + \rho_j + \tau_{it} + \epsilon_{ijmt}$$

Here, $Y_{ijmt}$ gives the probability of the home team winning the match. $COVID_{ijmt}$ is a dummy which is 1 if a match is played during the COVID-19 restrictions and 0 if it is played before the COVID-19 restrictions. $\rho_j$ and $\tau_{it}$ are fixed effects variables and capture respectively referee and team-season fixed effects. The subscripts are: i for home team, j for referee, m for match, and t for season.

We will use these three measures of home advantage as the dependent variables in a simple linear regression with team fixed effects and robust standard errors. The respective regression formula is as follows:

$$Y_{i_t} = \beta_0 + \beta_1 COVID_{it} + \alpha_i + \epsilon_{it}$$

Here, $Y_{i_t}$ gives the measure of HA for team i in season t. $COVID_{i_t}$ is a dummy which is 1 if a match is played during the COVID-19 restrictions and 0 if it is played before the COVID-19 restrictions. Team fixed effects team will be captured by $\alpha_i$.

This analysis will give a comprehensive and solid explanation of the effects found with the first method. Here, we will exclude the final standings from the 2019/2020 season from the regression. This is because this season can not be equally and correctly measure seasonal home advantage compared to the other seasons.

## 4.2 Referee's decision-making analysis

Home advantage can be caused by two factors: the influence of the crowd on the team's performance and the influence of the crowd on the referee's decision-making. To filter out the latter, we will analyse the referee's decisions before and during the COVID restrictions. We will do this by calculating the number of cards given per foul for the home team and away team. We will split this variable into yellow cards per foul and total cards per foul. This variable serves as indicator for the referee's strictness. By analysing the difference between this variable between the home and away team, we can make assumptions about the referee bias and whether the effect of the COVID variable is due to the crowd's influence on the home team's performance. For this analysis we used the following regression.

$$Y_{jt} = \beta_0 + \beta_1 COVID_{jt} + \gamma_t + \rho_j + \epsilon_{jt}$$

Here, $Y_{i_t}$ gives the (yellow) cards per foul for team i in season t. $COVID_{i_t}$ is a dummy which is 1 if a match is played during the COVID-19 restrictions and 0 if it is played before the COVID-19 restrictions. Fixed effects for season and referees are captured respectively by $\gamma_t$ and $\rho_j$.

# 5 Results

## 5.1 Regression on home wins

Table 2 shows the results of the simple regression of the COVID dummy on the probability of a home win. This regression has been performed on the seasons from 2018/2019 until 2020/2021. Columns 1-3 show the results with different combinations of team, season, and referee fixed effects. In columns 4-6, the referee fixed effects are added. Here, it is important to remember the results of Table 1, namely that the mean of the probability of a home win is 0.465 before the COVID break.

The results in column 1 show a significant negative coefficient. This is in line with the results shown in the descriptive statistics in Table 1. Column 2 also shows a significant negative coefficient for the COVID dummy. This means that after the COVID break while controlling for team fixed effects, home teams won 8.1 percentage points fewer matches. When controlling for team and season fixed effects (column 3), we see no significant effect of COVID on the probability of a win by the home team. We can explain this by the fact that there are little observations on matches played with and without spectators in the 2019/2020 season. Here, approximately 28 matches were played without and ten were played with spectators. Therefore, we obtain a larger standard error and an insignificant COVID coefficient.

The results in columns 4-5 show a significant effect of COVID-19, and thus playing without a crowd, on the probability that the home team wins. These results also indicate that the difference in probability that the home team wins, is not due to the referee. Therefore, these results indicate that home advantage decreased while there was no crowd in the stadium.

Table 2: Regression of COVID on the probability of a home win 18/19-20/21 season

| | Probability of a home win | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| COVID | -0.068** | -0.081*** | 0.019 | -0.063** | -0.082*** | -0.368** | -0.312** | 0.022 |
| | (0.030) | (0.030) | (0.055) | (0.031) | (0.031) | (0.155) | (0.148) | (0.056) |
| Constant | 0.464*** | 0.594*** | 0.620*** | 0.419*** | 0.638*** | 0.737*** | 0.741*** | 0.678*** |
| | (0.019) | (0.067) | (0.068) | (0.093) | (0.114) | (0.105) | (0.138) | (0.115) |
| Team FE | No | Yes | Yes | No | Yes | No | No | Yes |
| Season FE | No | No | Yes | No | No | No | No | Yes |
| Team x Season FE | No | No | No | No | No | Yes | Yes | No |
| Referee FE | No | No | No | Yes | Yes | No | Yes | Yes |
| Obs. | 1,140 | 1,140 | 1,140 | 1,140 | 1,140 | 1,140 | 1,140 | 1,140 |

Notes: Table 2 shows the results of the regression of the COVID dummy on the probability of a home win. Team, season, team-season, and referee fixed effects are added to the regression in various combinations. The standard error is showed between brackets. The significance is categorized as follows: *p<0.1; **p<0.05; ***p<0.01

In columns 6-7, we added team-season fixed effects. The results correspond to the the results in the previous columns, except for column 4. Both with and without referee fixed effects, the COVID coefficient is significant. This confirms our previous findings that playing without a crowd has a negative effect on the probability that the home team wins.

The COVID coefficient in column 8 is not significant. With the results from column 4 in mind, this is something that we expected. We can, again, explain this with the same reason as for the results in column 4. When including season fixed effects, the matches played in the 19/20 season with and without spectators become the observations of interest. However, there are few observations, only ten matches are played without a crowd in this season. Therefore, the standard error becomes large quickly and therefore the results are not significant.

## 5.2 Regression on home advantage measure

To further investigate whether home teams have lost home advantage during the COVID restrictions, as the results in the previous section suggest, we will use the three measures of home advantage as discussed earlier. For this regression, we leave the 2019/2020 season out because this season was interrupted by the COVID restrictions. Therefore, the measures of seasonal home advantage do not correctly represent the actual measures.

Table 3: Regression on HA

| | $h_i$ | | $HA_1$ | | $HA_2$ | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| COVID | -0.305* | -0.352 | -0.086*** | -0.100** | -0.460** | -0.559** |
| | (0.169) | (0.225) | (0.025) | (0.035) | (0.177) | (0.254) |
| Constant | 0.316** | 0.550 | 0.583*** | 0.601*** | 0.506*** | 0.604** |
| | (0.120) | (0.891) | (0.017) | (0.052) | (0.154) | (0.260) |
| Team FE | No | Yes | No | Yes | No | Yes |
| Observations | 40 | 40 | 40 | 40 | 40 | 40 |

Notes: Table 3 shows the results of the regression of the COVID dummy on home advantage measure by Clarke and Norman (1995) in columns 1-2, Pollard (1986) in columns 3-4, and Matos et al. (2019) in columns 5-6. Team fixed effects is added to the regression. The standard error is showed between brackets. The significance is categorized as follows: *p<0.1; **p<0.05; ***p<0.01

The results in Table 3 columns 1-2 shows the effect of playing without a crowd on the home advantage measure designed by Clarke and Norman (1995). These results show a slightly significant ($\alpha = 10\%$) relationship between playing without a crowd and a lower home advantage when not controlling for team fixed effects. However, when controlling for team fixed effects, the significance disappears. This does not correspond with the results we found in the previous regression analysis. However, we should remember that this regression analysis has been performed on only 40 observations.

Table 3 columns 3-4 shows the effect of playing without a crowd on the measure of home advantage by Pollard (1986). Here, we see more significant results than in columns 1-2. This means that columns 3-4 show that playing in an empty stadium has a negative effect on HA. To be precise, the HA also (almost) completely disappeared. In this method, 0.500 or 50% means no home advantage. In both columns 3 and 4, the measure of home advantage approaches this value.

Table 3 columns 5-6 confirms the results we found in the previous two columns. Both models have significant coefficients, meaning that playing without a crowd has a negative effect on the home advantage. Here, a value of 0 means that there is no home advantage or disadvantage. Looking at the values of the coefficients and constants, we can conclude that the home advantage decreased to almost 0. This means that we only observe a small measure of HA during the COVID restrictions. We will discuss the causes comprehensively in the discussion section.

## 5.3 A deeper look into referees

As discussed in the methodology section, home advantage can be caused by two factors. The first one is the influence the crowd has on the team's performance in the form of motivation, concentration and aggressiveness. The second factor is the influence of the crowd on the referee's decision-making. The crowd creates the atmosphere in the stadium and could through this, affect the referee's bias towards the home or away team.

To determine the effect the crowd has on the performance of the home team, we must investigate whether any form of referee bias exists and whether it changed after the COVID-19 restrictions. For this, we included referee fixed effects in the regressions in the previous section. However, we must analyse if the referee's decision-making changed after the COVID break. Therefore, we will analyse this decision-making from the 2018/2019 season until the 2020/2021 season. As an indicator for this, we use the number of cards per foul awarded to both the home and away teams. Table 4 columns 1-2 shows that both the home and away teams received fewer cards per foul while playing without a crowd. However, the difference between home and away, as showed in column 3, is not significant. Therefore, we can say that referees did not structurally have a bias towards the home or away team during the pandemic. The same applies to the results in columns 4-6. For this regression, we controlled for both referee and season fixed effects. With season fixed effects, we filter out the effect of new game rules or new guidelines for referees on the decision-making. With referee fixed effects, we filter out the difference between the various referees because not all referees are the same, and some give more cards than the other.

Table 4: Referee's decision-making analysis

| | Cards per foul | | | Yellow cards per foul | | |
|---|---|---|---|---|---|---|
| | Home | Away | Δ | Home | Away | Δ |
| COVID | -0.029** | -0.050*** | 0.021 | -0.027** | -0.052*** | 0.025 |
| | (0.013) | (0.015) | (0.018) | (0.013) | (0.014) | (0.017) |
| Constant | 0.116*** | 0.147*** | -0.031 | 0.114*** | 0.143*** | -0.029 |
| | (0.021) | (0.020) | (0.026) | (0.020) | (0.019) | (0.024) |
| Referee FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Season FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,140 | 1,140 | 1,140 | 1,140 | 1,140 | 1,140 |

Notes: Table 7 shows the results of the referee's decision-making analysis. Columns 1 and 2 show the regression of the COVID dummy on respectively the home and away team's cards per foul. Column 3 shows the difference Δ. Columns 4 and 5 show the regression of the COVID dummy on respectively the home and away team's yellow cards per foul. Column 6 shows the difference Δ. Team and season fixed effects are added to the regression. The standard error is showed between brackets. The significance is categorized as follows: *p<0.1; **p<0.05; ***p<0.01

# 6 Discussion and conclusion

Although the COVID-19 pandemic serves as a natural experiment, there are still mechanisms that could bias the treatment effect. The first one is clubs' finances. In this regression analysis, we assumed that all clubs were equally affected by the decreased income due to the withdrawal of sponsors and missed ticket sales. In reality, this could affect clubs that are more dependent on stadium or match day revenues differently than clubs that are less dependent on these revenues. Clubs with for example, smaller stadiums or with higher sponsorship revenues could be less affected than clubs which have the opposite.

Another mechanism could be the fact that at the end of the 19/20 season, players had to play the approximately nine matches in a much smaller time frame because of the incurred delay. After a period without matches, this could have affected players' fitness and could have increased the risk of injuries. Clubs who have a small squad, could be more affected by this. The EPL is one of the few leagues that did not increase the allowed substitutes from 3 to 5 after the COVID break. Therefore this effect could be even higher in the EPL than in other leagues. However, this is something for further research.

The last mechanism that could influence the home team's performance is a player or multiple players missing a match due to testing positive for the coronavirus. This is the case if one team misses (many) players while the other team plays with their best squad. This is not considered in the data but could cause a substantial bias in the treatment effect.

The descriptive statistics showed us that after the COVID break, home teams won 6.8 percentage points less matches. Besides this We found no evidence for a difference in draws but saw an increase of 6.4 percentage points of away teams winning. Furthermore, after the COVID break, home teams made 1.3 ($\alpha = 1\%$) less shots and 0.3 ($\alpha = 10\%$) less shots on target. Besides this, after the COVID break, home teams made 1.0 ($\alpha = 1\%$) more fouls per game, received 0.14 ($\alpha = 10\%$) fewer yellow cards, and 0.03 ($\alpha = 1\%$) fewer cards per foul, which corresponds to a decrease of 17%.

The results of the simple regressions in Tables 2 confirm these findings. While controlling for team, referee fixed effects, and team-season fixed effect, the treatment effect is negative and significant. When including season fixed effects, the treatment effect became not significant. This means that after the COVID break while controlling for team fixed effects, home teams won 8.1 percentage points fewer matches.

The three measures of HA in Table 3, which quantified HA, are in line with above-mentioned results. The measure of relative HA by Clarke and Norman (1996) showed a negative, slightly significant ($\alpha = 10\%$) effect when not controlling for team fixed effects. This coefficient also indicated that HA almost completely disappeared. The measure by Pollard (1986) was significant and indicated that the HA disappeared or even slightly turned into a home disadvantage. The measure by Matos, Amaro and Pollard (2020) was also negative and significant. This measure also indicated that that HA almost completely disappeared.

The analysis of the referee's decision-making was carried out to check whether the loss of HA is only attributed to the absence of the crowd, i.e. the influence of the crowd on the home team's performance. In the descriptive statistics, we saw an identical decrease in yellow cards awarded to both the home and away team. This already indicated that after the COVID break, the referee's decision-making did not change. The results in Table 4 confirm this. The difference $\Delta$ between the total cards per foul and yellow cards per foul did not significantly differ from 0. Therefore, we have shown that the referee bias did not change and thus the crowd does causally influence the home team's performance.

This implies that, for example, banning fans at the stadium as a sanction for bad supporters behavior is effective. Besides this, clubs who suffer from lower attendance could, for example, adjust their ticket pricing to this. Furthermore, they could support supporter's associations to improve the atmosphere during matches.

More importantly, this research confirms that a large group or a crowd and social pressure can have an effect on others. In sports, this could be on motivation, egoism or concentration. More in general, this could have an effect on for example, a person's, or group's decision-making or cooperation. However, to confirm this, more research is necessary. Therefore, we hope that this research encourages others to research this topic through sports economics or other fields of research.

# 7 Suggestions for further research

In this, we will discuss several suggestions for further research. First, in a year, one more season of soccer will be played. It would be good to include this data into a new research, both if spectators are allowed and if they are not allowed. Next season, probably the first part of the season will not be played with full capacity, while the last part has a higher chance of being played in a full stadium. Additionally, stadium capacity and other stadium measurements should be included in the research and should be linked to HA.

In general, it would be best to do this kind of research on multiple (European) soccer clubs. This will yield more extensive and reliable results on HA. Additionally, adding other competitions, including this year's European Championship or the UEFA Champions League and UEFA Europa League would be an interesting expansion of this research.

Besides this, a research on player specific level would be an interesting direction of further research. It would be valuable to research which players are affected by crowd support. Besides this, it would be interesting how these players are affected by crowd support and what it does to the other players in the team. For example, does an attacker shoot more frequently under crowd support. Or does a defender make more risky tackles and how does the team react to this? This is a whole new area of research but can be really valuable into the research of (social) pressure.

Furthermore, the earlier mentioned mechanisms, namely clubs finances, fitness & risk of injuries, and testing positive on the coronavirus. These mechanisms should be taken into account because they could bias the COVID treatment effect.

At last, grouping matches, for example, the first quarter of the season and comparing them in the regression analysis before and during the COVID-19 restrictions, could yield more reliable results. Especially if at the beginning of next season stadiums are completely filled, this method could yield compelling results.

# References


Agnew, G. A., & Carron, A. V. (1994). Crowd effects and the home advantage. *International Journal of Sport Psychology*.

Balmer, N. J., Nevill, A. M., & Williams, A. M. (2001). Home advantage in the winter olympics (1908-1998). *Journal of sports sciences*, *19*(2), 129–139.

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.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., & Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during covid-19. *Economics Letters*, *198*, 109664.

Červeňy, J., van Ours, J. C., & van Tuijl, M. A. (2018). Effects of a red card on goal-scoring in world cup football matches. *Empirical Economics*, *55*(2), 883–903.

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.

Dawson, P., Dobson, S., Goddard, J., & Wilson, J. (2007). Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanction in the english premier league. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(1), 231–250.

Deutscher, C., Winkelmann, D., & Ötting, M. (2020). Bookmakers' mispricing of the disappeared home advantage in the german bundesliga after the covid-19 break. *arXiv preprint arXiv:2008.05417*.

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.

Fransen, K., Boen, F., Vansteenkiste, M., Mertens, N., & Vande Broek, G. (2018). The power of competence support: The impact of coaches and athlete leaders on intrinsic motivation and performance. *Scandinavian Journal of Medicine & Science in Sports*, *28*(2), 725–745.

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. *Review of Economics and Statistics*, *87*(2), 208–216.

Goller, D., & Krumer, A. (2020). Let's meet as usual: Do games played on non-frequent days differ? evidence from top european soccer leagues. *European Journal of Operational Research*, *286*(2), 740–754.

Jamieson, J. P. (2010). The home field advantage in athletics: A meta-analysis. *Journal of Applied Social Psychology*, *40*(7), 1819–1848.

Johnston, R. (2008). On referee bias, crowd size, and home advantage in the english soccer premiership. *Journal of Sports Sciences*, *26*(6), 563–568.

Krumer, A. (2020). Testing the effect of kick-off time in the uefa europa league. *European Sport Management Quarterly*, *20*(2), 225–238.

Krumer, A., & Lechner, M. (2018). Midweek effect on soccer performance: Evidence from the german bundesliga. *Economic Inquiry*, *56*(1), 193–207.

Matos, R., Monteiro, D., Antunes, R., Mendes, D., Botas, J., Clemente, J., & Amaro, N. (2021). Home-advantage during covid-19: An analysis in portuguese football league. *International Journal of Environmental Research and Public Health*, *18*(7), 3761.

Matos, R. M., Amaro, N., & Pollard, R. (2019). How best to quantify home advantage in team sports: an investigation involving male senior handball leagues in portugal and spain.[la mejor manera de cuantificar la ventaja de jugar en casa en deportes de equipo: una investigación con ligas de balonmano senior masculino en portugal y españa]. *RICYDE. Revista Internacional de Ciencias del Deporte. doi: 10.5232/ricyde*, *16*(59), 12–23.

Mazer, J. P., Barnes, K., Grevious, A., & Boger, C. (2013). Coach verbal aggression: A case study examining effects on athlete motivation and perceptions of coach credibility. *International Journal of Sport Communication*, *6*(2), 203–213.

McCarrick, D., Bilalic, M., Neave, N., & Wolfson, S. (2020). Home advantage during the covid-19 pandemic in european football.

Peeters, T., & van Ours, J. C. (2021). Seasonal home advantage in english professional football; 1974–2018. *De Economist*, *169*(1), 107–126.

Pettersson-Lidbom, P., & Priks, M. (2010). Behavior under social pressure: Empty italian stadiums and referee bias. *Economics Letters*, *108*(2), 212–214.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of sports sciences*, *4*(3), 237–248.

Pollard, R. (2002). Evidence of a reduced home advantage when a team moves to a new stadium. *Journal of Sports Sciences*, *20*(12), 969–973.

Reade, J. (2020, Nov). *Yes, soccer scores are more surprising than usual this season.* Forbes Magazine. Retrieved from `https://www.forbes.com/sites/jamesreade/2020/11/09/yes-soccer-scores-are-more-surprising-than-usual-this-season/?sh=65aa89952d26`

Reade, J., Schreyer, D., & Singleton, C. (2020). Echoes: what happens when football is played behind closed doors? *Available at SSRN 3630130*.

Schwartz, B., & Barsky, S. F. (1977). The home advantage. *Social forces*, *55*(3), 641–661.

Scoppa, V. (2015). Fatigue and team performance in soccer: Evidence from the fifa world

cup and the uefa european championship. *Journal of Sports Economics*, *16*(5), 482–507.

Seckin, A., & Pollard, R. (2008). Home advantage in turkish professional soccer. *Perceptual and motor skills*, *107*(1), 51–54.

Sutter, M., & Kocher, M. G. (2004). Favoritism of agents–the case of referees' home bias. *Journal of Economic Psychology*, *25*(4), 461–469.

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.

Van de Ven, N. (2011). Supporters are not necessary for the home advantage: Evidence from same-stadium derbies and games without an audience. *Journal of Applied Social Psychology*, *41*(12), 2785–2792.