Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operational Research

The effects of the VAR, Covid-19 and live broadcasting on home advantage in the Premier League

Anwar Alexander Ouni (583329ao)

Supervisor:	prof. dr. R. Paap
Second assessor:	prof. dr. ir. J. C. van Ours
Date final version:	2nd July 2023

Abstract

In this paper we analyse how seasonal home advantage in the Premier League changed due to external effects. We extend the analysis by Peeters & van Ours (2021) by looking at three specific effects; we look at the effects of the Virtual Assistant Referee (VAR), the Covid-19 pandemic and live broadcasting games on the level of home advantage. We study this in the period from 1974 to 2023 and do this by calculating the home advantage in terms of goal difference and point difference. By regressing home advantage on a set of variables, we find that a larger attendance at home games, having an artificial pitch and being promoted the season before increases the level of home advantage. We also find that after the VAR was introduced in the 2019/20 season, home advantage was significantly lower than the years before. We conclude that the VAR has made Premier League football more fair. In addition, we conclude that the lack of attendance during the pandemic in the 2019/20 and 2020/21 seasons significantly reduced the level of home advantage. Finally, our analyses show that when Premier League games were live broadcasted, since the 1992/93 season, home advantage was also significantly reduced. This is likely due to an increase in the quality of the referees. Namely, when games are recorded, referee decisions can be studied which makes it easier for the Football Association (FA) to select the best referees.

1 Introduction

The tendency of sporting teams to perform better at their home field than away is known as home advantage. Many studies consider the presence of home advantage in football and it has been established as a deciding factor in the outcome of football games. Factors that are known to influence the level of home advantage in football include crowd attendance, travel effects, stadium and pitch familiarity and referee bias (Pollard, 2008). For English professional football, Peeters & van Ours (2021) found that home advantage is influenced by crowd attendance, grass or artificial pitch, whether a club has just been promoted or relegated and the amount of wages paid. In addition, Boyko et al. (2007) demonstrated that with a higher attendance in the stadium, referees tend to favor the at home playing team. Moreover, a study by Thomas & Carmichael (2005) indicated that in the Premier League at home teams play more aggressively but are sanctioned less. For instance, they get fewer yellow cards and less penalties are given against them. Furthermore, research on referee bias in Spain concluded that referees grant more playing time to teams playing at home if they are down in score (Garicano et al., 2005). All these studies conclude that crowd attendance and especially pressure on the referee by the crowd results into an increased and unfair home advantage. Considering that football games should be as fair as possible, it is important to gain a deeper insight into which factors drive home advantage and how they influence it. In our study we extend the paper by Peeters & van Ours (2021) by looking at the effect of the introduction of the Video Assistant Referee (VAR), the effect of the Covid-19 pandemic and the effect of live broadcasting of games on home advantage in the Premier League.

In our first extension, we investigate the effect of the VAR on home advantage. In the 2019/20 season, the VAR was introduced to the Premier League with the intend to make the games more fair. The VAR is only used to help the referee with the most important decisions such as whether a goal should be disallowed or whether a red card should be given. If the VAR is effective, game-changing referee decisions should no longer be biased towards the home team. Regarding the German football league, the Bundesliga, findings by Dufner et al. (2023) indicated that the VAR influences the game to a fairer degree. Spitz et al. (2021) examined football matches across 13 countries and found that VAR intervention increases the odds of making a correct decision from 92.1% to 98.3%. Moreover, Peeters & van Ours (2021) found that in English professional football teams playing at home on average earn 0.59 to 0.64 more points per game because of home advantage. They used data until and including 2018, which is before the introduction of the VAR. If the VAR works, the extra points teams earn at home attributed to home advantage should be lower from the 2019/20 season onward. In extension of these prior findings, we focus on the Premier League and examine whether home advantage changed after the VAR was introduced in this league. Using the methodology of Peeters & van Ours (2021) we compare the home advantage of recent seasons in which the VAR was active with the home advantage in the seasons before the introduction of the VAR. If the VAR made games more fair, home advantage should have been decreased. We examine this by looking at whether the home advantage in terms of goal and point difference was reduced. We conclude that the VAR has indeed made Premier League football more fair, because home advantage decreased significantly after the introduction of the VAR.

In our second extension, we study how home advantage changed during the Covid-19 pandemic. The 2019/20 and 2020/21 seasons were unlike any other because of the Covid-19 pandemic. Halfway through the 2019/20 season, the Premier League was suspended and at the end of the season matches were played without any crowd attendance allowed. This was continued throughout the majority of the 2020/21 season. It is well known that crowd attendance has a positive effect on the level of home advantage (e.g., Boyko et al., 2007; Peeters & van Ours, 2021; Pollard, 2008). It has also been proven that crowd size influences referee decisions (Boyko et al., 2007). The Covid-19 pandemic is a unique period in which we can exclude crowd effects because no attendance was allowed. Research in the Portuguese major soccer league by Matos et al. (2021), indicated that the absence of the crowd during the pandemic did not effect home advantage. However, de Castro Ribeiro et al. (2022) found that in Brazilian professional football the home advantage was significantly smaller in the seasons where no crowd was allowed. Cross & Uhrig (2023) measured the level of home advantage by the number of home minus away (expected) goals and found that this was reduced by 50% during the pandemic in the top 5 European leagues, including the Premier League. In our study, we examine whether the lack of attendance during the pandemic reduced the level of home advantage by focusing on the Premier League and by using the same methodology used by Peeters & van Ours (2021). The effect the crowd has on the game can be examined by comparing the home advantage in full stadiums with the home advantage in empty stadiums. Effects such as pitch familiarity and travel fatigue will still play a role in home advantage when no fans are allowed, but factors concerning the crowd are filtered out. The crowd can increase home advantage by raising the morale of the home team and they can pressure the referee to favor the home team. Our findings show that the level of home advantage significantly decreased during the pandemic. This can possibly be attributed to the absence of the crowd but could also be due to the introduction of the VAR.

In our third extension, we investigate the effect live broadcasting of Premier League games has on the level of home advantage. Premier League games were live broadcasted since the 1992/93 season and this could have lowered the level of home advantage. The intuition behind this is that the quality of referees should increase when games are broadcasted live. Referee decisions can be reviewed and the Football Association (FA) can more carefully assess whether referees are biased when games are recorded. Therefore, only the best referees remain. The recordings also enable referees to study past decisions which increases their quality. The effect of factors that are known to pressure the referee into making biased decisions, such as crowd attendance, should be reduced when the quality of referees is increased. Prior research by Dawson et al. (2006) shows that referee behaviour in terms of disciplinary sanctions is not influenced by live broadcasting of matches in the Premier League. In our extension, we focus on the level of home advantage and attempt to find out whether this changed after games were live televised. If the quality of referees has increased since games were live broadcasted, home advantage should have decreased. Our findings confirm this as we conclude that after games were broadcasted live, the amount of home advantage significantly decreased. We believe that an increase in referee quality as a result of the live broadcasting of games is a probable explanation for this decrease.

This thesis proceeds as follows. In Section 2, we describe the data set we used and how we retrieved our data. We also explain which periods in the sample are of special importance. We continue by describing the methods we employ in Section 3. We apply the basic model used by Peeters & van Ours (2021) in Section 4. Here, we calculate home advantage in terms of goal difference and point difference. In Section 5, we expand on the replication part by presenting the results of our three extensions. We draw our conclusions from this research in Section 6. In this final section, we also give suggestions for further research.

2 Data

In this research, we use data on Premier League results from 1974 until and including 2023. Data from 1974 until and including 2018 are available from: `https://datarepository.eur.nl/articles/dataset/Replication_archive_Peeters_and_van_Ours_2021_De_Economist/13643738`. This data set includes detailed panel data of each game during this period. Variables such as goals scored, attendance and the number of points teams had at the end of the season are provided. As the sample period of this data set ends in 2018, we collected the remainder of the data to include results until and including the 2022-23 season. For data of most recent seasons, we use footballwebpages.co.uk (2023) to find data on attendance in the Premier League. However, on this website attendance data was not available for the seasons affected by Covid-19 and therefore we use worldfootball.net (2023) to find the attendances per game during the pandemic. To be able to calculate home advantage we gather data on the performance of Premier League teams at home compared to away. We use soccerstats.com (2023) to find data on points, goals and games split into home results and away results. For our third extension, we found the exact period when live broadcasting of Premier League games began on footballontv.com (2023). This is needed to analyse the effect live broadcasting of games has on home advantage.

Peeters & van Ours (2021) found that there is a decline in home advantage in the Premier League throughout their sample of 45 years (1974-2018). Furthermore, they found that the home advantage in terms of extra points per game was around 0.9 at the start of the sample (1974) and declined to around 0.5 at the end of the sample (2018). Since this sample ends before the the introduction of the VAR and the start of the Covid-19 pandemic, the decline can not be attributed to these factors. However, it could be due to the live broadcasting of games. We attempt to find causes for this decline. The average home advantage in goals is 0.464 in the sample from 1974 to 2018. Then, after the VAR was introduced and the Covid-19 pandemic happened, the average reduced to 0.236 for the period from 2019 to 2023. So, it could be that the introduction of the VAR and the Covid-19 pandemic are causes for the continuation of this decline.

Table 1 shows the descriptive statistics of the most important variables of the data set we use. The sample period is from 1974-2023. As we only look at the Premier League, we use 1037 observations in the full sample. It is important to have enough observations in our regressions to get an efficient estimator and trustworthy results in our regressions. We see that the average attendance in this period is 28,739. The highest attendance is recorded in Old Trafford, the stadium of Manchester United, with 75,826 spectators. The least amount of spectators was

equal to 0 during the Covid-19 pandemic. The average number of points per season scored at home is 33.19. The least amount of points a team scored at home is 7 and the highest amount is 59. The amount of games played in a season varies from 42 at the start of the sample to 38 towards the end of the sample. This means the maximum amount of points possible at home is equal to 63. The difference in goals scored at home and goals conceded at home is on average 8.85, the minimum is -31 and the maximum is 54.

Table 1: Descriptive statistics of our data set with full sample (1974-2023)

	Attendance	Homepoints	Homegd
Mean	28,739	33.19	8.85
Std. dev	13,404	9.15	12.98
Minimum	0	7	-31
Maximum	75,826	59	54

All data is from the Premier League. The variables presented in the table are: attendance, points scored at home (homepoints) and difference in goals scored at home compared to goals scored away (homegd). The mean, standard deviation, minimum and maximum of these variables are presented.

3 Methodology

In this research we make use of the same methods used by Peeters & van Ours (2021). They quantify home advantage by looking at the number of extra goals and extra points teams score when they play at home. Methods introduced by Clarke & Norman (1995) take into account the quality of teams while calculating the home point difference (HPD) and home goal difference (HGD). In our research we use the same methods to calculate home advantage. We regress these values of home advantage on the variables attendance, pitch type and a promotion dummy to see how home advantage is affected. In our extensions, we regress home advantage on these same variables but we add a dummy to study the three specific periods of our extensions (period in which VAR was used, period of the Covid-19 pandemic, and period in which games were live broadcasted). We use a t-test to examine whether effects are significant.

3.1 Calculating home advantage on a seasonal level

In order to calculate the home advantage in terms of points we need the home point difference and the away point difference. The home point difference for team i (HPD_i) can be calculated by:

$$HPD_i = (N-1)q_i - \sum_{j(j\neq i)}^{N} q_j + (N-1)h_i \tag{1}$$

In this equation, the home point difference depends on the quality of team i (q_i), the quality of team j (q_j), the home advantage of team i (h_i) and the number of teams in the league (N).

The away point difference for team i (APD_i) can be calculated by:

$$APD_i = (N-1)q_i - \sum_{j(j\neq i)}^{N} q_j - \sum_{j(j\neq i)}^{N} h_j \tag{2}$$

Again, q_i and q_j stand for the quality of team i and j respectively, h_j is the home advantage of team j and N is the number of teams in the league.

In the home point difference and the away point difference we take into account the quality of the team and its opponents. When a team wins a lot of games at home, it could be because they are a team of high quality and does not necessarily mean their home advantage is high. Therefore, it is important to take the quality of teams into account as well. The quality of team i (q_i) can be calculated by:

$$q_i = \frac{HPD_i - (N-1)h_i}{N} \tag{3}$$

The quality of the teams in the Premier League is normalized so that the average in the league is equal to zero: $\sum_{i=1}^{N} q_j = 0$. Moreover, H is defined as the total home advantage in the league. This means that $H = \sum_{i}^{N} h_i = N\overline{h}$. By applying these two characteristics we get:

$$\sum_{j(j\neq i)}^{N} q_j = -q_i \qquad \text{and} \qquad \sum_{j(j\neq i)}^{N} h_j = N\overline{h} - h_i \tag{4}$$

From this equation we can rewrite the home point advantage and away point advantage as:

$$HPD_i = Nq_i + (N-1)h_i \qquad \text{and} \qquad APD_i = Nq_i + h_i - N\overline{h} \tag{5}$$

By subtracting the away point difference from the home point difference as they are defined in equation (4), we find that the home advantage of team i can be calculated as follows:

$$h_i = \frac{HPD_i - APD_i - H}{N-2} \tag{6}$$

Home advantage measured by goal difference is calculated in the same way as home advantage measured by point difference. Only now the number of goals scored and conceded at home and away are used instead of the number of points. This gives us an insight in whether a team adapts a different playing style at home. A large goal difference could mean that a team takes more risk and mainly focuses on attacking.

3.2 Explaining home advantage using regressions

To determine the driving factors of home advantage, we regress h_{it}, home advantage of club i during season t, on a set of club specific characteristics. We refer to this regression as the basic model and it looks as follows:

$$h_{it} = \alpha_i + \beta x_{it} + \epsilon_i \tag{7}$$

In this regression, α_i is a constant and x_{it} contains the time-varying club-specific variables. These variables include attendance, pitch type and whether a club has just been promoted.

We extend this basic model to analyse how home advantage is affected by the period in which the VAR was used, the period in which the Premier League was affected by Covid-19, and the period in which games were live broadcasted. In the extended model, we add an extra dummy for each specific period as an explanatory variable. This model looks as follows:

$$h_{it} = \alpha_i + \beta x_{it} + \beta_1 PeriodDummy_{it} + \epsilon_i \tag{8}$$

In this model, $PeriodDummy_{it}$ is equal to 1 depending on the period we study. It can be used as a dummy for the period in which the VAR was used, or it can be a dummy for the period in which Covid-19 affected the Premier League, or it can be a dummy for the period in which games were live broadcasted.

3.2.1 Dummy's

In equation (8) the model we use to analyse period specific effects is shown. As mentioned, we use three different dummy's. Each dummy is used for a different period and is used in its own regression. The first dummy is equal to 1 during the period in which the VAR was used in the Premier League and 0 otherwise. This period is from the 2019/20 season until and including the 2022/23 season. The second dummy we use is for the seasons in which the Premier League was affected by Covid-19. Specifically, the seasons in which a significantly lower amount of fans were allowed in the stadiums. This dummy is equal to 1 during the 2019/20 and 2020/21 seasons and 0 otherwise. Finally, we examine the effects of live broadcasting games on the level of home advantage by introducing a dummy for this period. The dummy for this period is equal to 1 from the 1992/93 season onwards and 0 before then.

3.2.2 Fifty-fifty sample

We also run the regression in equation 8 with a fifty-fifty sample. In such a regression, the dummy variable is equal to 0 for the first 50% of the sample and 1 for the second 50% of the sample. By doing this, we aim to study the direct changes in home advantage and try to exclude the substantial decline in home advantage over time found by Peeters & van Ours (2021). Dufner et al. (2023) research the effects of the VAR on home advantage in the Bundesliga and also do this by using the fifty-fifty sample technique. They used four seasons before the VAR was introduced and compared this with the home advantage during the four seasons after the VAR was introduced. The structural break which is due to live broadcasting of games would

be included if we study the effects of the VAR with a full sample. We exclude this structural break with the fifty-fifty sample. We also use a fifty-fifty sample when we study the effect of the Covid-19 pandemic and the effect of broadcasting games live.

3.2.3 T-test and standard error

We make use of a t-test to conclude whether our explanatory variables have a significant effect on home advantage. To test whether an explanatory variable x has a significant effect on the dependant variable y, we look at the estimated regression slope b_j. We test the null hypothesis H_0: $\beta_j = 0$ against the alternative H_1: $\beta_j \neq 0$ as shown in Heij et al. (2004). The t-statistic can be calculated as follows:

$$t_j = \frac{b_j - \beta_j}{s_j} \tag{9}$$

When we test the null hypothesis, t_j becomes: $t_j = \frac{b_j}{s_j}$. We reject the null hypothesis if t_j differs from significantly from zero. This happens when the estimated coefficient is large enough or the corresponding standard error is small enough. We use robust standard errors to make sure our errors are unbiased when we use the t-test. The (robust) standard error depends on the sample size and the precision of the estimated coefficient. When the total sample is smaller, there is more uncertainty in the estimation and the (robust) standard error will be higher. Which causes the variable to have an insignificant effect.

4 Basic model

The inspiration for this research stems from the findings by Peeters & van Ours (2021). Therefore, we use their findings as a starting point for our own research. It is important to replicate their method of calculating home advantage as we use the same methodology in our extensions. In our own research we also look at home advantage in terms of point and goal difference. In Peeters & van Ours (2021) the home advantage is regressed on factors such as attendance, a dummy for whether the pitch is artificial or real grass, a dummy for whether a team has been promoted last season and another dummy for whether a team was relegated last season. Our focus is solely on the Premier League whereas Peeters & van Ours (2021) include the top four leagues in England for their research. Consequently, we also focus solely on the Premier League in the replication part.

We adapted the data set in the replication archive to only include Premier League data. The sample is from 1974 until 2018. The average attendance throughout the sample is 28,642 fans. At the end of the sample period, in 2018, the average attendance is the highest it has ever been. This is because new stadiums are getting bigger and bigger. For example, West Ham has a new stadium since 2016 with a capacity of 62,500 people which is almost double the capacity of their old stadium. Since the artificial pitch dummy is one of the explanatory variables we regress home advantage on, it is relevant to mention that Queens Park Rangers and Luton are the only two teams that played on an artificial pitch in the Premier League. Both teams did so for five seasons in the late eighties. Since then, all teams in the Premier League have played

on real grass. A total of 49 different teams played in the Premier league during the sample period. Moreover, teams were promoted from the Championship to the Premier League 153 times during this period.

Following the example of Peeters & van Ours (2021), we start by calculating the seasonal home advantage for each team. Table 1 shows the average home advantage in terms of goal difference and points for each team that has played in the Premier League during the sample period. The number of seasons that each team played in the Premier League is also shown. What stands out is that relatively small teams that did not play many seasons in the Premier League have the largest home advantage. We see that Luton (0.913), Barnsley (0.906) and Oxford (0.810) have the highest home advantage in terms of goal difference while they only played 11, 1 and 3 seasons in the Premier League respectively. This could be due to the promotion factor. Peeters & van Ours (2021) have concluded that when a team has just been promoted, their home advantage is larger. Other teams in the division are less familiar with the pitch of a recently promoted team which makes playing there more difficult for them. Luton (1.243) and Barnsley (1.105) also have the second and third highest home advantage in terms of points. Only Oldham beats them with 1.307 extra points per home game. Again, the promotion factor seems to play a large role in this. Peeters & van Ours (2021) also concluded that a higher attendance has a positive effect on home advantage. However, the teams with the highest home advantages in table 1 have relatively small attendances. A reason why Luton had a substantially high home advantage could be due to the fact that they played on an artificial pitch. Peeters & van Ours (2021) also concluded that playing on an artificial pitch at home has a large and significant effect on the amount of home advantage teams enjoy. Luton played on an artificial pitch for five seasons in the Premier League and visiting teams that were not used to this likely performed worse.

Peeters & van Ours (2021) found an average home advantage of 0.64 points and 0.46 goals in the Premier League. The replication turns out to be accurate as we find an average home advantage of 0.648 points and 0.457 goals in the Premier League. Figure 1 shows the average home advantage in the Premier League over time in terms of goals and points. Again, our results look similar to those found by Peeters & van Ours (2021). Comparing results for each team individually is more difficult because we only look at results in the Premier League, whereas Peeters & van Ours (2021) look at four leagues. The average home advantage in terms of goal and point difference for each team in the Premier League can be found in Table 7 (see Appendix). The number of seasons each team played in the Premier League is also denoted. Results for teams with few or sometimes even just one season in the league are therefore not as accurate. The home advantages of Barnsley and Oldham are among the highest of all teams that have played in the Premier League. However, Peeters & van Ours (2021) found that when results of seasons they played in lower leagues are included, their home advantages are barely above average. Neither Barnsley nor Oldham have played more than three seasons in the Premier League and therefore, due to the lack of data, the results are heavily overestimated. Logically, teams with many seasons in the Premier League get results that are closer to the results found by Peeters & van Ours (2021).

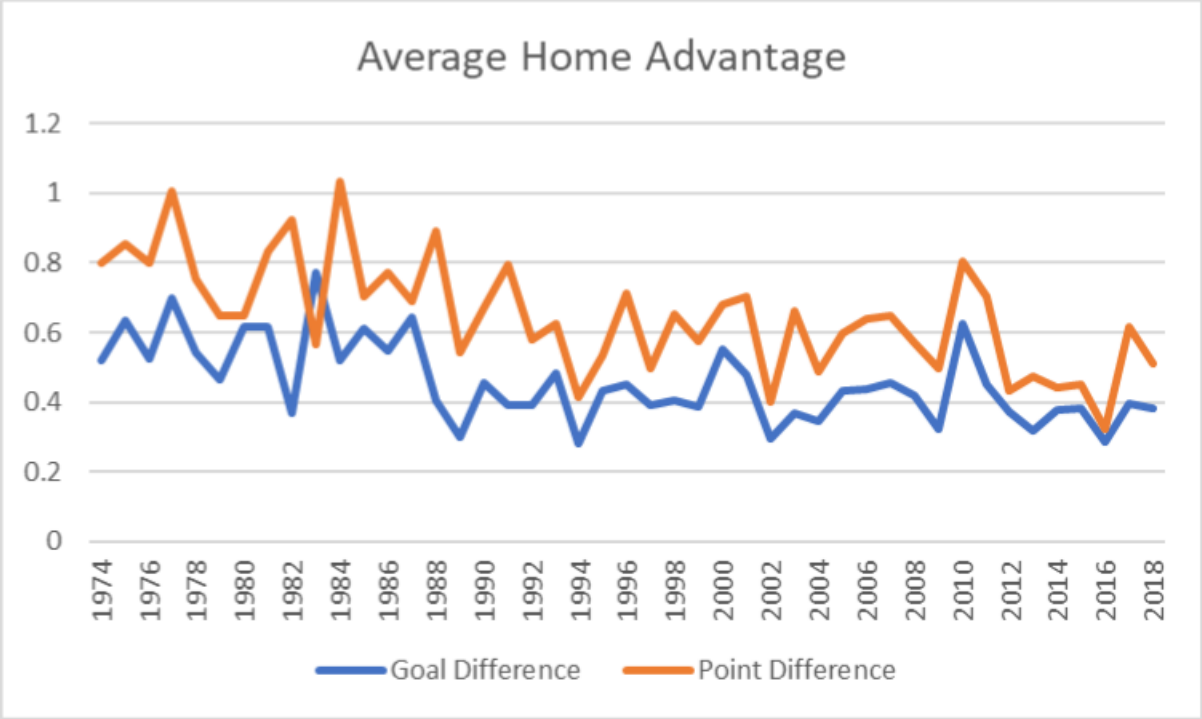

Figure 1: Average home advantage in terms of difference in goals or points in the Premier League (1974-2018)

The next part in the paper by Peeters & van Ours (2021) is where they regress relative home advantage in terms of points and goal difference on several factors. By doing this they found out which factors have a significant effect on home advantage. The factors used for these regressions include attendance, a dummy for whether a team plays on an artificial pitch, and a dummy for whether a team was promoted last season. Peeters & van Ours (2021) also include wages, and a dummy for whether a team was relegated as explanatory variables in their regression. As whether wages have an effect on home advantage is not the focus of our research we exclude this variable. Moreover, as a team can not be relegated to the Premier League because it is the highest possible league in English football, we also exclude this dummy from our regressions. Table 2 shows the regression output when we regress relative home advantage in terms of goal difference and point difference on attendance, a dummy for artificial pitches and a dummy for promotion. As in Peeters & van Ours (2021), we run the regression including club fixed effects as well. These effects account for time-invariant differences between clubs which include the nature and size of the pitch, the distance between the crowd and the pitch, the shape of the stadium etc. In Table 2, we see that attendance has a slightly positive but insignificant effect on home advantage in all regressions. However, having an artificial pitch at home increases the home advantage with a positive and significant effect (at 1% level). With an artificial pitch the home team increases the goal difference with 0.5 goals per game on average. The home team is also awarded 0.8 points more at the end of the game on average. When a team is promoted to the Premier League their home advantage is increased by a very small but insignificant amount.

Peeters & van Ours (2021) found similar results when we look at the size of the coefficients. However, the difference is that they found significant results (at 5% level) for attendance in

most regressions. They also found a few significant coefficients (at 10% level) for the promotion factor. An explanation for the fact that they got more significant coefficients could be that they included more observations. As they use data of the four highest leagues in English professional football and not just the Premier League, they include much more data. A larger number of observations leads to a larger t-statistic and therefore more significant results as mentioned in Section 3.2.3.

Table 2: Regression results where relative home advantage (RHA) in terms of goal difference (GODIF) and points is regressed on a set of independent variables

	RHA_GODIF	RHA_GODIF	RHA_POINTS	RHA_POINTS
Attendance	0.06	0.005	0.02	0.05
	(0.05)	(0.1)	(0.07)	(0.2)
Artificial Pitch	0.5***	0.5***	0.8***	0.7***
	(0.1)	(0.007)	(0.2)	(0.1)
Promotion	0.01	0.03	0.01	0.02
	(0.05)	(0.05)	(0.07)	(0.07)
R^2	0.011	0.007	0.013	0.007
Club fixed effects	No	Yes	No	Yes
Observations	937	937	937	937

Independent variables are: attendance, a dummy for whether a team plays on an artificial pitch and a dummy for whether a team was promoted last season. P-values are denoted in parentheses. Data is from the Premier League from 1974 to 2018. Significance levels of 10%, 5% and 1% are denoted by *, ** and *** respectively.

5 VAR, Covid-19 and live broadcasting: the effects on home advantage

We perform three extensions on the replication part. First, we analyse how home advantage changed after the VAR was introduced to the Premier League. Then, we look at whether the amount of home advantage was reduced during the Covid-19 pandemic when no supporters were allowed in the stadiums. Finally, we also analyse whether live broadcasting of Premier League games changed the amount of home advantage. For each analysis we run two regressions. One in which we set home advantage in terms of goal difference as dependant variable and another in which home advantage in terms of point difference is used. The explanatory variables always include a constant, attendance, an artificial pitch dummy and a promotion dummy. However, we also include a dummy for the period we want to analyse to see whether home advantage changed during that period.

5.1 Explaining home advantage

In order to analyse the effects of the VAR and Covid-19, we extend the data sample to include most recent seasons. The sample ends after the 2022/23 season. For these results we use the full sample period to have a more efficient and significant estimator in our analysis. Peeters & van Ours (2021) used a sample period from the 1974/75 season until and including the 2017/18

season and found an average home advantage of 0.64 extra points and 0.46 extra goals. Both the VAR and Covid-19 pandemic are expected to reduce the amount of home advantage. We find that the average home advantage from the 2018/19 season until and including the 2022/2023 season is 0.252 extra goals and 0.265 extra points; a substantially smaller home advantage during this period. This is likely due to the effects of both the VAR and the Covid-19 pandemic. Namely, the VAR ensures important refereeing decisions are checked to be as fair as possible. And during the Covid-19 pandemic no supporters were allowed in the stadiums, whereas crowd attendance is known to increase the level of home advantage (e.g., Boyko et al., 2007; Peeters & van Ours, 2021; Pollard, 2008)). Consequently, the pandemic could play a significant role in the reduction of the amount of home advantage Premier League teams enjoyed during this period. The average home advantage during the two years affected by the Covid-19 pandemic was 0.16 extra goals and 0.06 extra points. In the two years before the pandemic, these averages were 0.35 extra goals and 0.47 extra points. To study whether the effects of the introduction of the VAR and Covid-19 pandemic were significant, we regress home advantage on dummies for these periods and the independent variables used in the basic model.

Since the level of home advantage has been declining over the years (Peeters & van Ours, 2021) it makes sense that the level of home advantage is lower near the end of the sample. So, it would be interesting to see whether our results hold when we focus on specific periods around the dummy period instead of the full sample. Therefore, we also do the regression analysis with a fifty-fifty sample for each dummy. To analyse the seasons affected by Covid-19 for example, we use two seasons before the pandemic and the two seasons during the pandemic. By doing this, we hope to exclude the structural decline of home advantage.

5.1.1 VAR effect analysis

First, we analyse the period after the VAR was introduced. The VAR has been a part of the Premier League from the 2019/20 season onward. Its purpose is to help the referee make game-changing decisions and thereby making the game more fair. Therefore, referees should be less biased in favor of the home team and home advantage should be reduced if the VAR is effective. Table 3 shows the results of a regression where home advantage is regressed on a constant, attendance, an artificial pitch dummy, a promotion dummy and a dummy for the period in which the VAR was used. We find that the constant is significant and that having an artificial pitch still has a large and significant effect on the amount of home advantage in goals as well as points. Moreover, we find that the VAR dummy has a considerably negative and significant effect on home advantage in terms of goals and points. This result could mean that the VAR is a success and referees have become less biased towards home teams. However, there could be other underlying factors that reduced the amount of home advantage during this period as well. The pandemic, for example, overlaps with this period. Furthermore, we see that a recently promoted team has a slightly increased home advantage but this result is not significant. In contrast to our findings in Section 4, the attendance factor is no longer significant and reduced to a negligible amount. The reduction of the effect of attendance on home advantage can be attributed to the VAR. Boyko et al. (2007) found that a higher attendance results in the referee favoring home teams but our analysis shows that with the VAR, the effect of attendance on

home advantage is reduced.

We also use a fifty-fifty sample for our analysis. The relevant sample for this is from 2015/16 to 2022/23. During this part of the sample, no games were played on an artificial pitch. To prevent issues with multicollinearity, we exclude the artificial pitch dummy as an explanatory variable in the regression for this analysis. Table 3 also shows the results of the fifty-fifty sample regression in the last two columns. What stands out in comparison to the same regression in which the full sample is used, is that the VAR dummy is no longer significant. We also see that the constant is no longer significant. The number of observations is a mere 160 which results into higher standard errors and less significance. We can not draw any clear conclusions from these results because there are not enough observations. The full sample regression uses 877 more observations to estimate the time period before the VAR which yields us far more accurate results.

Additionally, we study the cross effects between attendance and VAR. The idea behind this is that the positive effect of attendance on home advantage should be reduced when the VAR is used. This is expected because with the VAR, referee decisions should no longer be influenced by the crowd. To study whether this is the case, a cross effects variable is added, which is a multiplication of the dummy variable for the VAR and the attendance variable. Our results show that this cross effect variable is significant and positive. The attendance variable becomes close to zero and more insignificant while the VAR dummy becomes more negative and more significant when the cross effect variable is added to the regression. The coefficient of the cross effect variable is 0.010 and it is significant at 1% level. This goes against our intuition as this means that the effect of attendance on home advantage is increased when the VAR is used. A possible explanation could be that the VAR is also biased towards the home team or that the VAR is influenced by the crowd. However, further research is needed to investigate this.

Table 3: Regression results using the full sample and a fifty-fifty sample to analyse how home advantage in goals and points changes when the VAR is used

	HA_GODIF	HA_POINTS	HA_GODIF50	HA_POINTS50
Constant	0.389***	0.561***	0.038	-0.067
	(0.000)	(0.000)	(0.725)	(0.709)
Attendance	0.001	0.002	0.008***	0.013***
	(0.225)	(0.290)	(0.001)	(0.001)
Artificial Pitch	0.549***	0.890***		
	(0.001)	(0.001)		
Promotion	0.053	0.077	0.148	0.287
	(0.235)	(0.252)	(0.227)	(0.159)
VAR dummy	-0.224***	-0.406***	-0.049	-0.122
	(0.000)	(0.000)	(0.588)	(0.412)
R^2	0.026	0.032	0.083	0.094

Independent variables include: attendance, a dummy for whether a team plays on an artificial pitch, a dummy for whether a team was promoted last season, and a dummy for the period in which VAR was used. The last two columns show the results when a fifty-fifty sample is used. P-values are denoted in parentheses. Data is from the Premier League from 1974 to 2023. Significance levels of 10%, 5% and 1% are denoted by *, ** and *** respectively.

5.1.2 Covid-19 effect analysis

Next, we look at the effect of the Covid-19 pandemic on home advantage. The regression is the same as the analysis of the effect of the VAR but now we use a dummy for the period in which the Premier League was influenced by the pandemic instead of a dummy for the VAR. Table 4 summarizes the regression results. Similar to the previous regression of the VAR analysis, we find that having an artificial pitch has a large and significant effect and that attendance has a slightly positive but insignificant effect on home advantage. The promotion dummy is about the same size as in the VAR analysis but is significant now, even at a 1% level. Moreover, we find that the Covid-19 dummy, which is only equal to 1 for the seasons affected by the pandemic, is substantially negative and significant at 1% level. This means that home advantage was significantly lower during the Covid-19 pandemic. As Boyko et al. (2007) and Peeters & van Ours (2021) state, attendance increases home advantage. Due to the fact that no attendance was allowed during the pandemic, this could be a large factor in the reduction of the home advantage during this period. The average attendance in the 2020/21 season was 461 spectators which is extremely small compared to other seasons before Covid-19 (average attendance 28.642 spectators). Furthermore, the VAR was also first introduced in the 2019/20 season. Because no spectators were allowed in the stadiums and the VAR was introduced, the home advantage enjoyed by Premier League teams was reduced substantially.

The results of the fifty-fifty sample regression can be found in the last two columns of Table 4. The sample used for this regression is from 2017/18 to 2020/21. Again, no games were played on an artificial pitch during this period. Therefore, the artificial pitch dummy is excluded from the regression again. In this regression, even fewer observations are used which substantially increases the standard errors. Therefore, it is no surprise that the Covid dummy is not significant. Only the attendance variable is significant but the corresponding coefficient is extremely small.

Table 4: Regression results using the full sample and a fifty-fifty sample to analyse how home advantage in goals and points changed during the Covid-19 pandemic

	HA_GODIF	HA_POINTS	HA_GODIF50	HA_POINTS50
Constant	0.403***	0.594***	-0.051	-0.243
	(0.000)	(0.000)	(0.741)	(0.421)
Attendance	0.001	0.001	0.010***	0.017**
	(0.307)	(0.789)	(0.003)	(0.010)
Artificial Pitch	0.543***	0.873***		
	(0.001)	(0.001)		
Promotion	0.054	0.078	0.017	0.320
	(0.228)	(0.245)	(0.913)	(0.308)
Covid-19 dummy	-0.270***	-0.553***	0.051	-0.006
	(0.002)	(0.000)	(0.710)	(0.983)
R^2	0.023	0.031	0.142	0.128

Independent variables include: attendance, a dummy for whether a team plays on an artificial pitch, a dummy for whether a team was promoted last season, and a dummy for the period during the Covid-19 pandemic. The last two columns show the results when a fifty-fifty sample is used. P-values are denoted in parentheses. Data is from the Premier League from 1974 to 2023. Significance levels of 10%, 5% and 1% are denoted by *, ** and *** respectively.

5.1.3 Live broadcasting effect analysis

Finally, we also look at whether the amount of home advantage changed when Premier League games were live broadcasted since the 1992/93 season. The idea behind this is that the quality of the referees should have been increased because of this development. As referees and the English Football Association were able to look at recordings of the games, they were able to study refereeing decisions. Therefore, only the best referees remained and referees could not be as biased anymore. This resulted in an increase of the quality of referees in the Premier League which should reduce the bias towards home teams and therefore the amount of home advantage. In order to research this, we run another regression with the same independent variables but we use a dummy (TV dummy) for the period in which games we live broadcasted. Table 5 summarizes the regression results. We find that the TV dummy is substantially negative and significant. This means that the amount of home advantage was significantly reduced after Premier League games were broadcasted live. Like in the previous regressions, the coefficient corresponding to the artificial pitch dummy is large and significant. However, in this regression the attendance factor is minimal but significant as well (at 1% level for home advantage in goals and 5% level for points). The promotion dummy is still slightly positive but insignificant. As the TV dummy is negative and significant, we believe that after Premier League games were live broadcasted, the quality of the referees increased which resulted in a decrease of home advantage.

Again, we also use a fifty-fifty sample regression. The sample period is from 1974 to 2009 for this regression. Several teams played on artificial grass during this period so the artificial

pitch dummy is added to the regression again. Table 5 shows the results of this regression. A clear difference compared to the full sample regression is that the attendance variable is no longer significant. Because the coefficient of the attendance variable is extremely small, a small increase in the standard error is enough to cause this. In contrast to the other fifty-fifty sample analyses, we find significant and logical results because a large amount of observations are used in this sample. We conclude that after games in the Premier League were broadcasted live, the amount of home advantage in terms of goal difference and points decreased.

Furthermore, we look at the cross effects between attendance and live broadcasting of Premier League games. As live broadcasting of games reduces the level of home advantage, we expect that the effect of attendance should be lower when games are broadcasted live. However, the cross effect variable is not significant and close to zero so we can not draw a clear conclusion from these results.

Table 5: Regression results using the full sample and a fifty-fifty sample to analyse how home advantage in goals and points changes when the games are broadcasted live

	HA_GODIF	HA_POINTS	HA_GODIF50	HA_POINTS50
Constant	0.431***	0.611***	0.496***	0.724***
	(0.000)	(0.000)	(0.000)	(0.000)
Attendance	0.004***	0.005**	0.001	0.00
	(0.002)	(0.012)	(0.469)	(0.960)
Artificial Pitch	0.487***	0.811***	0.455***	0.756***
	(0.004)	(0.002)	(0.007)	(0.003)
Promotion	0.047	0.071	0.037	0.053
	(0.285)	(0.295)	(0.459)	(0.485)
TV dummy	-0.180***	-0.251***	-0.131***	-0.151**
	(0.000)	(0.000)	(0.001)	(0.010)
R^2	0.038	0.035	0.027	0.025

Independent variables include: attendance, a dummy for whether a team plays on an artificial pitch, a dummy for whether a team was promoted last season, and a dummy for the period in which Premier League games were broadcasted live. The last two columns show the results when a fifty-fifty sample is used. P-values are denoted in parentheses. Data is from the Premier League from 1974 to 2023. Significance levels of 10%, 5% and 1% are denoted by *, ** and *** respectively.

So, to summarize the findings in our three extensions, we conclude that home advantage in terms of goals and points was significantly lower in the period in which the VAR was used, in the period in which Covid-19 affected the Premier League, and in the period in which games were live broadcasted. We also see that having an artificial pitch significantly increases the level of home advantage. Attendance also has a significantly positive effect in some of the regressions. When we use fifty-fifty samples for our regressions, we do not find any significant results for the period in which the VAR was used and for the period in which the Covid-19 pandemic affected the Premier League. This is likely because the number of observations is not large enough in these samples. We do find that live broadcasting has a significantly negative effect on home advantage using a fifty-fifty sample.

5.1.4 Which period has the strongest effect?

We also run a regression in which the dummies for all three periods are included. In this regression we include a dummy for when the VAR was used, another dummy for when the Covid-19 pandemic affected the Premier League, and a dummy for the period in which games were broadcasted live. We use the full sample for this regression. The results of this regression can be found in Table 6. What stands out is that the dummies for the VAR and the Covid-19 pandemic are no longer significant. Only the dummy for when games were broadcasted live is significant, even at 1% level. The corresponding coefficients of this dummy are also substantially negative. These results imply that live broadcasting of Premier League games has the strongest and most significant effect on reducing home advantage. This could be because the period in which games were live broadcasted is far larger than the period in which the VAR was used and the period of the Covid-19 pandemic. Therefore, more data was used to estimate the coefficient for this period which reduces the standard error and increases the significance.

Table 6: Regression results when the dummies of all three periods are included

	HA_GODIF	HA_POINT
Constant	0.442***	0.643***
	(0.000)	(0.000)
Attendance	0.003***	0.003*
	(0.009)	(0.095)
Artificial Pitch	0.482***	0.794***
	(0.004)	(0.001)
Promotion	0.046	0.069
	(0.298)	(0.303)
Dummy VAR	-0.133	-0.199
	(0.130)	(0.139)
Dummy Covid	-0.050	-0.243
	(0.683)	(0.195)
Dummy TV	-0.157***	-0.202***
	(0.000)	(0.000)
R^2	0.044	0.046

Independent variables include: attendance in thousands of fans, a dummy for whether a team was promoted last season, a dummy for the period in which the var was used, a dummy for when Covid-19 affected the Premier League and a dummy for the period in which games were broadcasted live. P-values are denoted in parentheses. Data is from the Premier League from 1974 to 2023. Significance levels of 10%, 5% and 1% are denoted by *, ** and *** respectively.

6 Conclusions

Home advantage is known to be a deciding factor in the outcome of sports games. Extensive prior research has analysed the presence of home advantage in professional football. However, to our knowledge, the effect of the VAR, Covid-19 and live broadcasting of games on home advantage has not been researched yet in the Premier League. Therefore, we did three extensions to investigate how the level of home advantage in the Premier League is effected by the introduction of the VAR, the Covid-19 pandemic, and by live broadcasting of games. First, we calculated home advantage following the methodology used by Peeters & van Ours (2021), with the focus on the Premier League. Then, we utilized a basic model to find out which factors drive home advantage. In this model, home advantage is regressed on attendance, a dummy for whether a team plays on an artificial pitch, and a dummy for whether a team was promoted the season before. For our three extensions, we introduced three new dummies to this model. The first new dummy is equal to 1 when the VAR was used, the second new dummy is equal to 1 during the seasons that were affected by Covid-19 and the last new dummy is equal to one for the seasons that were broadcasted live. By adding these dummies to our regressions, we were able to analyse the effects of these three periods on home advantage.

Our results using the basic model corroborate the findings of Peeters & van Ours (2021) as we conclude that the average amount of home advantage is 0.648 points and 0.457 goals in the Premier League from 1974-2018. Our regressions also show that home advantage is significantly increased by having an artificial pitch. A higher attendance and being promoted the season before also increases the level of home advantage, but these results are significant in only a few of the regressions in our study. The research of our extensions leads to significant conclusions as well. In all three extensions, the full sample is used to estimate the effects (1974-2023). This means that the periods in which the VAR was used and the pandemic affected the Premier League are included. We find that the VAR results in a higher level of fairness in the Premier League as home advantage has been reduced significantly after its introduction. Moreover, we see that during the Covid-19 pandemic, where games were played without attendance, home advantage was significantly lower. The regression where we analyse the effects of live broadcasting of games leads to the same conclusion. Therefore, we conclude that live broadcasting of games played a substantial role in the decline of home advantage over time. We also used a fifty-fifty sample in other regressions to exclude structural breaks caused by earlier periods. To accurately examine the effect of Covid-19 for example, we excluded the structural break caused by live broadcasting using this technique. However, these results are inconclusive and insignificant for the analysis on the effects of the VAR and the Covid-19 pandemic. This is because only a small part of the sample can be used to estimate these effects. When fewer observations are used, estimation errors are larger and results are less significant. We did find significant results using a fifty-fifty sample for the analysis on the effect of live broadcasting of games. Namely, that this leads to a significant reduction in the level of home advantage in the Premier League. In a completely fair league, the crowd should not be able to pressure the referee into making biased decisions. However, factors such as pitch familiarity and travel fatigue will still lead to some home advantage. Therefore, the reduction in home advantage makes football more fair but home advantage will continue to exist.

Further research is needed to examine whether these results hold in other leagues as well. Especially for the fifty-fifty sample regressions, more observations are needed to get significant results. If the data set can be expanded by including other leagues, more accurate results can be found. The decline in home advantage over the years could possibly also be explained by other events than the ones we found, for example, changes in the rules of the game. If more leagues are included, where the events we study occurred in different periods, it becomes more certain that the events that we studied were the cause for the decline in home advantage. This applies to the VAR for instance as it was introduced at different times in different leagues. In the future, more data on the VAR will become available as the VAR is likely to stay in the Premier League for years to come. Therefore, research on the effect of the VAR on home advantage in the Premier League can be more accurate in the future when more observations can be included. Moreover, we found that the effect of attendance on home advantage increased when the VAR was active. Thus, it would also be interesting to investigate whether the VAR is biased towards the home team or influenced by the crowd.

References

Boyko, R., Boyko, A. & Boyko, M. (2007, 10). Referee bias contributes to home advantage in english premiership football. *Journal of sports sciences*, *25*, 1185-1194. doi: 10.1080/02640410601038576

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. Retrieved from `https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2348899` doi: https://doi.org/10.2307/2348899

Cross, J. & Uhrig, R. (2023, 01). Do fans impact sports outcomes? a covid-19 natural experiment. *Journal of Sport Economics*, 3-27. Retrieved from `https://journals.sagepub.com/doi/pdf/10.1177/15270025221100204` doi: 10.2139/ssrn.3705085

Dawson, P., Dobson, S., Goddard, J. & Wilson, J. (2006, 11). 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. Retrieved from `https://doi.org/10.1111/j.1467-985X.2006.00451.x` doi: 10.1111/j.1467-985X.2006.00451.x

de Castro Ribeiro, L., de Souza Fonseca, F., Costa, G. D. C. T., de Oliveira Castro, H., da Silva Santos, J. P. V. & Figueiredo, L. S. (2022). Did the absence of crowd support during the covid-19 pandemic affect the home advantage in brazilian elite soccer? *Journal of Human Kinetics*, *81*(1), 251–258. Retrieved from `https://doi.org/10.2478/hukin-2022-0047` doi: doi:10.2478/hukin-2022-0047

Dufner, A.-L., Schütz, L.-M. & Hill, Y. (2023). The introduction of the video assistant referee supports the fairness of the game – an analysis of the home advantage in the german bundesliga. *Psychology of Sport and Exercise*, *66*, 102386. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1469029223000109` doi: https://doi.org/10.1016/j.psychsport.2023.102386

footballontv.com. (2023, May). *History of football on tv.* Footballontv. Retrieved from `https://www.live-footballontv.com/history-of-football-on-tv-uk.html`

footballwebpages.co.uk. (2023, May). *Football web pages.* Footballwebpages. Retrieved from `https://www.footballwebpages.co.uk/premier-league/attendances`

Garicano, L., Palacios-Huerta, I. & Prendergast, C. (2005, 02). Favoritism under social pressure. *The Review of Economics and Statistics*, *87*, 208-216. doi: 10.2139/ssrn.274934

Heij, C., de Boer, P., Franses, P., Kloek, T. & van Dijk, H. (2004). *Econometric methods with applications in business and economics.* United Kingdom: Oxford University Press.

Matos, R., Monteiro, D., Antunes, R., Mendes, D., Botas, J. & Clemente, . A. N., J. (2021). Home-advantage during covid-19: An analysis in portuguese football league. *International journal of environmental research and public health*, *18*(7), 37-61. Retrieved

from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8038430/ doi: https://doi.org/10.3390/ijerph18073761

Peeters, T. & van Ours, J. C. (2021). Seasonal home advantage in english professional football; 1974–2018. *De Economist*, *169*, 107-126. Retrieved from https://link.springer.com/article/10.1007/s10645-020-09372-z doi: https://doi.org/10.1007/s10645-020-09372-z

Pollard, R. (2008, 06). Home advantage in football: A current review of an unsolved puzzle. *The Open Sports Sciences Journal*, *1*. doi: 10.2174/1875399X00801010012

soccerstats.com. (2023, May). *Soccer stats.* Soccerstats. Retrieved from https://www.soccerstats.com/homeaway.asp?league=england

Spitz, J., Wagemans, J., Memmert, D., Williams, A. M. & Helsen, W. F. (2021). Video assistant referees (var): The impact of technology on decision making in association football referees. *Journal of Sports Sciences*, *39*(2), 147-153. Retrieved from https://doi.org/10.1080/02640414.2020.1809163 (PMID: 32794432) doi: 10.1080/02640414.2020.1809163

Thomas, D. & Carmichael, F. (2005, 08). Home-field effect and team performance: Evidence from english premiership football. *Journal of Sports Economics*, *6*, 264-281. doi: 10.1177/1527002504266154

worldfootball.net. (2023, May). *world football.* WorldFootball. Retrieved from https://www.worldfootball.net/attendance/eng-premier-league-2020-2021/1/

7 Appendix

Table 7: The average home advantage in terms of goal difference and point difference for each team that has played in the Premier League from 1974-2018

Club	HA_GODIF	HA_POINTS	SEASONS
Arsenal	0.466	0.637	45
Aston Villa	0.383	0.571	40
Barnsley	0.906	1.105	1
Birmingham	0.523	0.839	18
Blackburn	0.478	0.609	18
Blackpool	0	-0.614	1
Bolton	0.420	0.699	15
Bournemouth	0.199	0.240	3
Bradford	0.705	1.066	2
Brighton	0.573	0.902	5
Bristol City	0.548	0.796	4
Burnley	0.444	0.733	7
Cardiff	0.246	0.675	1
Carlisle	0.202	0.107	1
Charlton	0.292	0.330	12
Chelsea	0.362	0.434	37
Coventry	0.505	0.635	28
Crystal Palace	0.120	-0.038	14
Derby	0.541	0.763	18
Everton	0.510	0.741	45
Fulham	0.523	0.9601	13
Huddersfield	0.240	0.430	1
Hull City	0.267	0.509	5
Ipswich	0.590	0.722	18
Leeds United	0.337	0.440	23
Leicester City	0.328	0.621	22
Liverpool	0.523	0.698	45
Luton	0.913	1.243	11
Manchester City	0.532	0.727	36
Manchester United	0.477	0.498	44

Club	HA_GODIF	HA_POINTS	SEASONS
Middlesbrough	0.457	0.722	24
Millwall FC	0.662	0.592	2
Newcastle United	0.719	0.936	33
Norwich	0.573	0.853	24
Nottingham Forest	0.368	0.527	20
Notts Country	0.310	0.784	4
Oldham	0.724	1.307	3
Oxford	0.810	0.845	3
Portsmouth	0.598	0.685	8
QPR	0.416	0.813	22
Reading	0.224	0.986	3
Sheffield United	0.515	0.689	8
Sheffield Wednesday	0.438	0.671	15
Southampton	0.63	0.888	34
Stoke City	0.632	0.851	20
Sunderland	0.363	0.469	23
Swansea	0.431	0.742	9
Swindon	0.338	0.293	1
Tottenham Hotspur	0.328	0.495	44
Watford	0.551	0.751	11
West Bromwich	0.354	0.392	22
West Ham	0.480	0.618	35
Wigan	0.042	0.088	8
Wimbledon	0.234	0.193	14
Wolverhampton	0.420	0.683	13

The last column shows the amount of seasons each team has played in the Premier League from 1974-2023.

8 Programming code

In this research we make use of three different types of programming software. We load our collected data in Stata to calculate the level of home advantage in terms of goals and points. Then, we regress this level of home advantage on different sets of explanatory variables in Eviews. Furthermore, we use Excel to calculate some simple statistics such as averages, standard deviations, minima and maxima. A ZIP-file has been made available with the exact programming code we used and a description of the code and variables.