MODELING HOME ADVANTAGE IN SOCCER WITHOUT FANS

OWOKO TREVOR ODUOL 092857

Submitted in partial fulfillment of the requirements for the Degree of Bachelors of Business Science Financial Economics at Strathmore University

Strathmore Institute of Mathematical Sciences Strathmore University Nairobi, Kenya

February, 2021

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Proposal contains no material previously published or written by another person except where due reference is made in the Research Project itself.

TREVOR ODUOL OWOKO ..... [Name of Candidate]

..... [Signature]

17/02/2021 ..... [Date]

This Research Proposal has been submitted for examination with my approval as the Supervisor.

MELEAH OLECHE ..... [Name of Supervisor]

..... [Signature]

17/02/2021 ..... [Date]

Strathmore Institute of Mathematical Sciences

Strathmore University

Table of Contents

ABSTRACT

The occurrence of COVID-19 in 2020, midway through the 2019/2020 soccer season brought rise to the stop and re-emergence of the English Football League. The re-emergence brought with it a new aspect in that football matches were being played but 'behind closed doors' in that no fans were allowed into the venue- because of the COVID-19 safety restrictions, the absence of fans in the stands to support their teams provided a niche for the study below. The purpose of this study is to investigate the effect between crowd presence and its relationship to Home Advantage, comparison between games when fans were present (pre-covid) and matches without fans (post-covid). The study was guided by research objectives that were seeking to identify the relationship between crowd presence and home advantage whether if present and significant towards the outcome of a soccer match and the effect of yellow cards and red cards on the outcome of a game. The study population consisted of 6 leagues: German Bundesliga, the Spanish la Liga Santander, the Italian Serie A, the French Ligue, the English Premier League and the English Championship. The quantitative secondary data was collected from a published website on shots, yellow cards, red cards and attendance. Regression analysis is used in data analysis with the application of R-studio. The study established that football matches played post-COVID period show that fan presence does not have a significant effect on the outcome of a game as previously thought. Yellow and red cards are seen to hold greater impact on the home games without fan presence. Based on the findings of the study referees need more help and support instead of admonishment in their decision making. What is deemed to be fairer by the crowd should not be treated with such great importance, to a point where trained referees' decisions are not supported. There is also need for further analysis with more data as the study was done with limited data and hence there might be room for further learning. The phenomena of home advantage is seen to continue eluding researchers with its exact causality not yet identified as observed in this paper that it is not crowd presence.

CHAPTER ONE: INTRODUCTION

1.1 Background Information

Conventionally, players, officials, commentators and fans all round argue that home advantage in professional sporting events is advantageous. Pollard and Gomez (2009) further discovered that home advantage persists across all leagues and continents. Home advantage is also seen to persist in the long horizon as seen by the study conducted by Inan(2018) who using Pollard's model found that home advantage persists across 30 seasons of the Turkish premier League.

It would be prudent to note that the exact causes of home advantage are still under investigation. The factors leading to the home field advantage and how these factors affect the outcome of a game have yet not been well understood.

Pollard (1986) came up with a method to analyse home advantage that has since commonly used to analyse home advantage. Pollard argued the likely causes of home advantage as local crowd support, travel fatigue, amongst other variables. (Pollard, 2008) added territoriality and rule factors to the model as influential variables. Other scholars have instead used goals scored, instead of points gained at home and proceeded to check for home advantage (Marek & Vavra 2018). Different studies have used different methods and have all accuracy to some degree.

Travel fatigue tends to affect teams negatively, (Pollard, Silva and Medieros, 2008) found that lengths of travel usually decrease performance in soccer games in Brazil. Familiarity with local conditions has been seen to be in favour with home team after a study (Loughead , Carron & Bray 2011). Territoriality is observed as subjective cause. This is because home team advantage might as well as be self-perpetuating phenomenon. Teams expect to do better at their home turf, or because they believe they should play well in their home city, in front of their own fans, they do better (Inan, 2019). They adopt different tactics or act more aggressively hoping that the referee, under crowd pressure, will be more lenient. This feeling of having to do better may be heightened because of political issues, usually analysed under territoriality. Referee bias has been studied and found to influence home advantage. Bokyo et al examined over 5000 matches and 50 referees and observed that some part of home advantage is influenced referee bias in the English Premier League. Home bias was observed

to differ among referees. (Nevill, Balmer, and Williams, 2002) conducted a study and established that the presence of crowd noise influenced the referee to make decisions that somewhat favour the Home team.

(Pollard and Gomez, 2004) suggest that taking a closer look at crowd effects on home advantage and closely considering size, intensity, density and proximity can be expected to affect both teams and referees. When one considers size, according to (Pollard and Gomez, 2014) it has been observed that the home advantage varies a lot across different countries. Despite this, according to (Pollard, G and Pollard, R, 2005) home advantage across a league in the same country does not vary as much though crowd size does. When one considers proximity it is observed from a study by that stadiums that have running tracks hence proximity between crowd and referees has shown decreased home advantage (Armatas and Pollard. 2014)

1.2 Problem Statement

During this Covid-19 pandemic does the absence of fans in a stadium have a significant influence in the home advantage ability as compared to fans being present and therefore outcome of a match? In recent months due to the outbreak of the pandemic Corona virus, the footballing world has had to take drastic measures to ensure that matches are still being played. This brought about the idea of playing behind closed doors, meaning there are no fans in attendance. The results that have arisen since then have put in question the level of influence crowd presence has on home advantage.

Various models exist in place that consider home-field advantage as a factor but the presence/absent of fans has never been in question and now that fans are not available which is unprecedented, this paper would like investigate how home advantage is affected and the effects the lack of fans has on team performance.

1.3 Research Objectives

i) To model the Home Advantage in the presence and absence of fans.in the chosen leagues and teams.

ii)There is also investigation of the effect of Yellow cards and Red cards on match results.

1.4 Significance of the Research

Home advantage has been an elusive-phenomena in that previous papers and studies have not been able to pinpoint the exact influential variable behind Home Advantage, whether it be Crowd Size, Territoriality or even Travel Fatigue. The results from this study will allow the clarification of whether Home Advantage is directly linked to fan presence, or on the other hand if the crowd variable acts independently of the home factor.

The study will inform various clubs the level of revenue lost as a result of no fans present hence indicate the need for new ways for clubs to interact with fans hence diversification of fan base to offset their matchday deficits.

CHAPTER TWO: LITERATURE REVIEW

2.1. Introduction

The concept of home advantage was first brought up in soccer by Dowie (1982), this being the seminal paper, he came up with three possible causes of home advantage. He mentioned fatigue, familiarity and fans albeit he did not reach clear conclusions.

Pollard (2008) improved on his paper from 1986 and included territoriality and rule factors to the already existing variables influencing home advantage: travel fatigue, local crowd support and familiarity with local conditions, referee bias, special tactics and psychological factor. His paper is responsible to emergence of the most common way of quantifying "home side advantage" as percentage of points gained or total goals scored with respect of total points gained or total goals scored while home.

Recently Marek and Vavra (2018) applied the use of goals scored and conceded instead of points garnered and to describe home field advantage across different leagues. Using this approach, the home team advantage is a random variable that can be described by trinomial distribution, and it is possible to use Jeffrey divergence and test for homogeneity of parallel samples to compare and test home team advantage of different leagues.

2.2 Theoretical Framework

Three authors Pollard, Silva and Medieros (2008) after a study conducted in the Brazilian Premier League that travel fatigue influences home advantage as results showed teams located in the North and South of Brazil across 5 leagues all had better home advantage and decreased performance from teams that travelled long distances.

Goumas (2014) conducts a study that seeks to find the magnitude of home advantage in Australian football and home team crowd effect and away team contribute to it. He puts to use Poisson Regression. The results of the analysis showed that Home Advantage significantly increased number of time zones crossed by away teams. The results claim that football competitions where time zones are crossed, jet lag becomes a bigger factor as it results to fatigue and is observed to have a higher effect on home advantage as compared to home team crowd support.

Familiarity with conditions has also been considered as an influence of home advantage. With one study by Loughead, Carron & Bray (2011) shows that familiarity with conditions

influences home advantage in the case where team quality is introduced as a moderator, it also concludes that home side advantage is reduced when good teams shift stadiums.

Sanchez, Leo, Pollard and Gomez (2009) conducted a study to examine the effect of points system on home advantage in the two top Spanish Leagues in the form of a pooled cross section in that seasons observed are across 1980-1981 when two points were awarded for a win and the 2006-2007 season when rules had changed and a win was awarded with three points. The results observed were that home advantage reduced with the introduction of the three-point system. Key to note that in current times rule factors are considered obsolete as in the 1980s the point changed from 2 points for a win and became 3 points for a win and all the leagues use the same rules.

Bokyo et al (2007) conducted a study that seeks to establish whether various measures of home advantage are affected by the official for the match and by crowd size while controlling for team ability. Their results confirm that referees are responsible for some of the observed home advantage in the EPL and suggest that home advantage bias differs across different referees. Pollard and Gomez (2014) also show how referee bias towards the home team is established. Even though referee bias is deemed as a determinant for home-field advantage, yet that should be taken as a result at the intermediary stage in a sequential process which is triggered by the crowd.
The above is built upon by Anders and Rotthoff (2014) who established that home advantage was greater in teams with a greater prospect of fan violence, albeit further exploration needs to be done. When fans are hostile, match officials may feel their life is under threat and this may cast doubt on their ability to call a fair match.

Boudreaux, Sanders and Walia (2017) carried out crowd effect upon home field advantage and found that crowd effects are sizable in influencing a home win, and they went onto say that the home advantage in its entirety should be attributed to crowd effects. Pollard and Gomez (2014) argue that if we take a closer look to crowd affect in home advantage, its density, proximity, size and intensity are likely to influence referees and teams.

Pollard (2005) conducted a study that indicated that home advantage across the same league does not differ much but crowd sizes differ a lot. Goumas (2013) discovered that there was a significant relation between crowd size and home advantage in European competitions. The

author also considers the use robust multivariate techniques to be able to observe independent effects of the numerous factors that influence home advantage.
Proximity is reduced in stadiums that have running tracks and as compared to stadiums that that do not have running tracks according to a study by Armatas and Pollard (2014) that home advantage is reduced when proximity is reduced. The paper goes on to indicate that the best indicator of home advantage as measured by goal difference, was the difference between home and away teams when considering shots taken within the penalty area.

Leite (2017) considers the application of home advantage as a key variable in the outcome of a football match. Analysis was carried out across over 3000 games and it was obtained that home advantage has an impact on the final outcome of a football encounter. The values obtained were lower as compared to similar indexes hence indicating gradual decrease of the variable over time.

Inan (2020) considers the size of crowd effect and crowd density and seeks to see whether they have an effect on home advantage. The author applies a new theoretical framework and a binary logistic regression. Analysis of the results revealed that crowd support and crowd density are of great importance for home field advantage.

2.3 Gaps in research

From the papers mentioned we observe that authors have considered a number of variables that influence home advantage. Each paper coming up with an analysis that provides further knowledge into the elusive "home advantage". There have been tests on crowd size effect, travel fatigue, referee bias just to mention a few.
Currently and since March of 2020 football games have been played with no fans allowed in attendance, and this is bound to continue for the foreseeable future, this is due to the presence of a pandemic by the name Covid-19. By virtue of this the question whether home field advantage is significantly influenced by fans being in the stadium or not can be answered. The study seeks to contribute to the gap by establish whether crowd presence has a significant effect on home advantage or not, and the effect this might have on a team's performance. The variable is important because crowd effect and crowd size in previous papers have been observed to influence home advantage but yet has a paper been able to compare in a case where fans are present and fans are entirely not present in the stadium across the top 6 European leagues.

2.4 Conceptual Framework

Teams get points in football for winning and tying, three and one respectively, but not for losing. Since there is no difference in winning by one goal or by multiple goals, we use a binary probit model following Anders and Rotthoff (2011). The research analyzes the impact of attendance on match outcome and also the effects of yellow and red cards.

Each individual match has a home (h) and away (a) team. Let:

$$U_h = X_h\beta_h + \varepsilon_h \quad (1)$$

Be the match-winning potential of team h.

X_h is a vector of match-specific, observable characteristics which are; shots on goal, corner kicks, off-sides, yellow cards, yellow-to-red cards, red cards, and fouls committed.

β_h is a vector of coefficients and ε_h is the error term that accounts for unobservable characteristics such as luck and ability.

The error terms are assumed to be distributed normally for both $U_h$and U_a. In this research we concentrate on the home team winning; U_h

CHAPTER THREE: METHODOLOGY

3.1 Introduction

This segment is divided into the research design, the population of the study as well as data collection methods and procedure of analysis.

3.1 Research Design

Research design refers to the structure of an enquiry. It ensures that evidence collected enables one to answer questions as unambiguously as possible (De Vaus, 2001). This research uses an Ordered Probit model to measure differences between the league matches before Covid-19 and after the Covid-19 restrictions were set.

3.2 Population and Sampling

The study aimed at analysing six European Leagues. These included the German Bundesliga in the first division, the Spanish la Liga Santander (first division), the Italian Serie A, the French Ligue 1, the English Premier League and the English Championship. Analysis was across the end 2019/2020 seasons and beginning of 2020/2021 when fans were not present compared to results from 2018/2019 and 2019/2020 seasons when fans were present.

3.3 Data collection

The main source of data was through secondary collection on the internet. The nature of the data is secondary and hence collected and compiled by other individuals or sources. The result of all this would be that web articles were an appropriate tool and source for my study.

3.4 Data Analysis

Concentrating on the home team winning, we regress the match-specific variables on the Out-come of the match:

$$Game\ Outcome_h = \propto + \beta_1 HYellowC + \beta_2 HRedC + \beta_3 HFouls + \beta_4 HCKicks + \beta_5 HShots + \beta_6 AYellowC + \beta_7 ARedC + \beta_8 AFouls + \beta_9 AShots + \beta_{10} Ln(Attendance) + \varepsilon \qquad (2)$$

Game Outcome equals 0 for the away team's victory, 1 for a tie and 2 for a home team's win. To measure any impact the fans themselves have on the outcome of the match, focus is on the impact of attendance at a given match.

Attendance is analyzed in two distinct ways: as attendance and attendance squared for the purpose of controlling for any curve-linear possibilities.

The natural log of attendance is used to see if a percentage change in attendance or the lack of attendance impacts the probability of the home team winning the match. The model also controls for year level fixed effects. RStudio was used for this analysis.

CHAPTER FOUR: RESULTS AND ANALYSIS

4.1 Introduction

This chapter presents the results of the data analysis, interpretation, presentation and discussion of findings.

4.2 Model Summary

This study sought to analyze the significance of fan presence on home games as compared to the influence of other independent variables that affect a game outcome.

4.2.1 English Premier League

From the post-covid model;

```
Call:
polr (formula = EP.G. O1 ~ engpremierpostcovid1$Fl +
engpremierpostcovid1$Y + engpremierpostcovid1$R +
engpremierpostcovid1$Sh + engpremierpostcovid1$ShT +
engpremierpostcovid1$LnAttend)

Coefficients:
                                Value   Std. Error   t-value
engpremierpostcovid1$Fl          0.01011    0.02251    0.4493
engpremierpostcovid1$Y          -0.17839    0.12052   -1.4802
engpremierpostcovid1$R          -1.02828    0.47854   -2.1488
engpremierpostcovid1$Sh         -0.02462     0.02617   -0.9408
engpremierpostcovid1$Sh          0.24642     0.07133    3.4549
engpremierpostcovid1$LnAttend    0.34413     0.18009    1.9109
```

Here, the results show the number of red cards to have the highest impact on the game, with a coefficient of -1.02828. This indicates perfect negative correlation between the number of red cards obtained and the possibility of a win in the home game.

Attendance in the games; post-covid, does not hold much significance on the turnout of the match.

Looking at the pre-covid results;

```
Call:
polr(formula = EP.G. O ~ engpremierprecovid1$Fl +
engpremierprecovid1$Y + engpremierprecovid1$R +
engpremierprecovid1$Sh + engpremierprecovid1$ShT +
engpremierprecovid1$LnAttend)
```

```
Coefficients:
                                Value    Std. Error t value
engpremierprecovid1$Fl          0.003549   0.01488    0.2384
engpremierprecovid1$Y          -0.037412   0.06587   -0.5680
engpremierprecovid1$R          -0.395180   0.36220   -1.0910
engpremierprecovid1$Sh         -0.002239   0.01334   -0.1679
engpremierprecovid1$ShT         0.103744   0.03455    3.0028
engpremierprecovid1$LnAttend    0.488802   0.22726    1.9308
```

The results show that before covid, level of attendance had the highest impact on the game, with a coefficient of 0.488802. This indicates positive correlation between attendance and the possibility of a win in the home game.

Number of red cards are seen to have had a lower effect before covid when they had higher attendance levels.

4.2.2 German Bundesliga

From the post-covid model;

```
Call:
polr(formula = GB.G. O1 ~ g.bundesliga1postcovid1$Fl +
g.bundesliga1postcovid1$Y + g.bundesliga1postcovid1$R +
g.bundesliga1postcovid1$Sh + g.bundesliga1postcovid1$ShT +
g.bundesliga1postcovid1$LnAttend)

Coefficients:
                                    Value Std. Error  t value
g.bundesliga1postcovid1$Fl       -0.023794    0.01909 -1.24656
g.bundesliga1postcovid1$Y        -0.002702    0.12634 -0.02139
g.bundesliga1postcovid1$R        -0.345097    0.55944 -0.61686
g.bundesliga1postcovid1$Sh       -0.018475    0.04161 -0.44403
g.bundesliga1postcovid1$ShT       0.314916    0.12086  2.60571
g.bundesliga1postcovid1$LnAttend  0.713033    0.26057  2.73643
```

In the German League, the results show attendance to have the highest impact on the game, with a coefficient of 0.713033. This indicates high positive correlation between attendance and the possibility of a win in the home game.

Looking at the pre-covid model;

```
Call:
polr(formula = GB.G.O ~ g.bundesliga1precovid1$Fl +
g.bundesliga1precovid1$Y + g.bundesliga1precovid1$R +
```

```
g.bundesliga1precovid1$Sh + g.bundesliga1precovid1$ShT +
g.bundesliga1precovid1$LnAttend)

Coefficients:
                                  Value Std. Error t value
g.bundesliga1precovid1$Fl         0.005218    0.01430  0.3649
g.bundesliga1precovid1$Y         -0.047328    0.06902 -0.6857
g.bundesliga1precovid1$R          0.301666    0.36538  0.8256
g.bundesliga1precovid1$Sh        -0.007701    0.01539 -0.5005
g.bundesliga1precovid1$ShT        0.118336    0.04161  2.8437
g.bundesliga1precovid1$LnAttend   3.999803    0.23988 16.6738
```

Attendance before covid still had the highest impact on the outcome of the game, with a coefficient of 3.999803, as compared to other variables.

4.2.3 Spanish La Liga

From the post-covid model;

```
Call:
polr(formula = LL.G. O ~ laliga1precovid$Fl + laliga1precovid$Y +
    laliga1precovid$R + laliga1precovid$Sh + laliga1precovid$ShT +
    laliga1precovid$LnAttend)

Coefficients:
                           Value Std. Error t value
laliga1precovid$Fl         0.012890    0.01383  0.9318
laliga1precovid$Y         -0.006289    0.04684 -0.1343
laliga1precovid$R         -0.245324    0.22582 -1.0864
laliga1precovid$Sh        -0.018009    0.01733 -1.0393
laliga1precovid$ShT        0.103014    0.04166  2.4727
laliga1precovid$LnAttend   1.216374    0.53239  2.2848
```

In La Liga, the results show attendance to have the highest impact on the game, with a coefficient of 1.216374. This indicates high positive correlation between attendance and the possibility of a win in the home game.

The pre-covid model;

```
Call:
polr(formula = LL.G.O ~ laliga1precovid$Fl + laliga1precovid$Y +
    laliga1precovid$R + laliga1precovid$Sh + laliga1precovid$ShT +
    laliga1precovid$LnAttend)

Coefficients:
                           Value Std. Error t value
laliga1precovid$Fl         0.012890    0.01383  0.9318
laliga1precovid$Y         -0.006289    0.04684 -0.1343
```

```
laliga1precovid$R          -0.245324    0.22582 -1.0864
laliga1precovid$Sh         -0.018009    0.01733 -1.0393
laliga1precovid$ShT         0.103014    0.04166  2.4727
laliga1precovid$LnAttend    1.216374    0.53239  2.2848
```

From the results of the pre-covid analysis, attendance still holds greatest significance on the outcome of the game, with a coefficient of 1.216374.

4.2.4 Italian Serie A

The post-covid model analysis;

```
Call:
polr(formula = SA.G.O1 ~ serieApostcovid1$Fl + serieApostcovid1$Y +
    serieApostcovid1$R + serieApostcovid1$Sh + serieApostcovid1$ShT
+ serieApostcovid1$LnAttend)

Coefficients:
                             Value Std. Error t value
serieApostcovid1$Fl       -0.01655    0.01886 -0.8775
serieApostcovid1$Y         0.04841    0.07657  0.6322
serieApostcovid1$R        -0.77121    0.49823 -1.5479
serieApostcovid1$Sh       -0.01166    0.03539 -0.3295
serieApostcovid1$ShT       0.26073    0.06238  4.1795
serieApostcovid1$LnAttend  0.26096    0.17536  1.4881
```

The results show the number of red cards to have the highest impact on the game outcome, with a coefficient of -0.77121. This indicates high negative correlation between the number of red cards obtained and the possibility of a win in the home game. Attendance does not hold much significance on the turnout of the match.

Looking at the pre-covid results;

```
Call:
polr(formula = SA.G.O ~ serieAprecovid1$Fl + serieAprecovid1$Y +
    serieAprecovid1$R + serieAprecovid1$Sh + serieAprecovid1$ShT +
    serieAprecovid1$LnAttend)

Coefficients:
                           Value Std. Error t value
serieAprecovid1$Fl       0.04336    0.01383  3.1353
serieAprecovid1$Y       -0.11513    0.06786 -1.6966
```

```
serieAprecovid1$R          -0.07473    0.25761 -0.2901
serieAprecovid1$Sh         -0.04006    0.02284 -1.7537
serieAprecovid1$ShT         0.14896    0.04496  3.3135
serieAprecovid1$LnAttend    1.91154    0.24357  7.8482
```

However, pre-covid data shows attendance to hold greater significance on the outcome of the game, with a coefficient of 1.91154.

4.2.5 French Ligue

The post-covid league analysis;

```
Call:
polr(formula = FL.G.01 ~ f.liguepostcovid1$Fl + f.liguepostcovid1$Y
+ f.liguepostcovid1$R + f.liguepostcovid1$Sh + f.liguepostcovid1$ShT
+ f.liguepostcovid1$LnAttend)

Coefficients:
                             Value Std. Error t value
f.liguepostcovid1$Fl       -0.01149    0.02959 -0.3884
f.liguepostcovid1$Y         0.21097    0.15321  1.3770
f.liguepostcovid1$R        -0.16420    0.47797 -0.3435
f.liguepostcovid1$Sh        0.05715    0.03616  1.5808
f.liguepostcovid1$ShT       0.06831    0.07598  0.8990
f.liguepostcovid1$LnAttend  0.74735    0.42229 -1.7697
```

Post-covid, the results show attendance to have the highest impact on the game, with a coefficient of 0.74735. This indicates positive correlation between attendance and the possibility of a win in the home game.

In the pre-covid model;

```
Call:
polr(formula = FL.G.0 ~ f.ligueprecovid1$Fl + f.ligueprecovid1$Y +
f.ligueprecovid1$R + f.ligueprecovid1$Sh + f.ligueprecovid1$ShT +
f.ligueprecovid1$LnAttend)

Coefficients:
                             Value Std. Error t value
f.ligueprecovid1$Fl       -0.02524    0.01070 -2.3586
f.ligueprecovid1$Y         0.03492    0.07251  0.4816
f.ligueprecovid1$R        -0.10334    0.23775 -0.4346
f.ligueprecovid1$Sh        0.02986    0.01636  1.8254
f.ligueprecovid1$ShT       0.07870    0.03413  2.3061
f.ligueprecovid1$LnAttend  1.53194    0.69146  2.2155
```

From the results, attendance is seen to still have a high significance on the outcome of the game for this league, with a coefficient of 1.53194, before the covid pandemic.

4.2.6 English Championship

The post-covid model;

```
Call:
polr(formula = EC.G.O1 ~ engcpostcovid1$Fl + engcpostcovid1$Y +
engcpostcovid1$R + engcpostcovid1$Sh + engcpostcovid1$ShT +
engcpostcovid1$LnAttend)

Coefficients:
                          Value Std. Error  t value
engcpostcovid1$Fl       -0.01100    0.01767 -0.62264
engcpostcovid1$Y         0.26206    0.12776  2.05115
engcpostcovid1$R         0.01807    0.40205  0.04494
engcpostcovid1$Sh       -0.02419    0.02131 -1.13512
engcpostcovid1$ShT       0.27215    0.06377  4.26762
engcpostcovid1$LnAttend  0.32231    0.11874  2.71435
```

The model results show none of the variables to have significant effect post-covid.

```
Call:
polr(formula = EC.G.O ~ engcprecovid1$Fl + engcprecovid1$Y +
engcprecovid1$R + engcprecovid1$Sh + engcprecovid1$ShT +
engcprecovid1$LnAttend)

Coefficients:
                          Value Std. Error t value
engcprecovid1$Fl        0.007241    0.01017  0.7121
engcprecovid1$Y        -0.019585    0.05047 -0.3880
engcprecovid1$R         0.396732    0.29201  1.3586
engcprecovid1$Sh        0.008281    0.01322  0.6264
engcprecovid1$ShT       0.067105    0.03314  2.0247
engcprecovid1$LnAttend  3.662204    0.21902 16.7209
```

However, pre-covid results show attendance to have great impact on the game outcome, with a coefficient of 3.662204.

4.3 Correlation Analysis

The English premier league;

	Fl	Y	R	Sh	Attendance	rate_homewins
Fl	1.0000000	0.8569460	0.1917090	0.80412399	0.49439077	-0.26592526
Y	0.8569460	1.0000000	0.1999337	0.64054100	0.31620758	-0.34404435
R	0.1917090	0.1999337	1.0000000	0.10403254	0.14147786	-0.26394764
Sh	0.8041240	0.6405410	0.1040325	1.00000000	0.56271263	0.02743679
Attendance	0.4943908	0.3162076	0.1414779	0.56271263	1.00000000	0.01713451
rate_homewins	-0.2659253	-0.3440444	-0.2639476	0.02743679	0.01713451	1.00000000

From the results, there is very low correlation between attendance and the rate of home wins, at 0.01713451.

German Bundesliga;

	Fl	Y	R	Sh	Attendance	rate_homewins
Fl	1.00000000	0.7416560	0.01068497	0.7136688	0.40051212	-0.08527534
Y	0.74165604	1.0000000	0.27124505	0.4955688	0.16578130	-0.30482133
R	0.01068497	0.2712451	1.00000000	-0.1264603	-0.07768428	-0.26331146
Sh	0.71366881	0.4955688	-0.12646027	1.0000000	0.56760995	0.29706932
Attendance	0.40051212	0.1657813	-0.07768428	0.5676100	1.00000000	0.34507613
rate_homewins	-0.08527534	-0.3048213	-0.26331146	0.2970693	0.34507613	1.00000000

Here, the correlation is a bit higher but still below 0.5, as it's at;0.34507613.

Italian Serie A;

	Fl	Y	R	Sh	Attendance	rate_homewins
Fl	1.0000000	0.8092596	0.2782349	0.6309992	0.63578652	-0.19434103
Y	0.8092596	1.0000000	0.4738762	0.5374229	0.56560851	-0.10851359
R	0.2782349	0.4738762	1.0000000	0.2213787	0.16035481	-0.20266786
Sh	0.6309992	0.5374229	0.2213787	1.0000000	0.63625010	0.26130420
Attendance	0.6357865	0.5656085	0.1603548	0.6362501	1.00000000	0.03669899
rate_homewins	-0.1943410	-0.1085136	-0.2026679	0.2613042	0.03669899	1.00000000

For this league, the correlation between the rate of home games won and attendance is also very low, at 0.03669899.

Spanish La Liga 1;

	Fl	Y	R	Sh	Attendance	rate_homewins0
Fl	1.0000000	0.7646357	0.11063865	0.52283821	-0.29206326	-0.30522935
Y	0.7646357	1.0000000	0.16845661	0.40331570	-0.18821543	-0.26702641
R	0.1106387	0.1684566	1.00000000	0.09699895	0.01629086	-0.03969438
Sh	0.5228382	0.4033157	0.09699895	1.00000000	0.32834420	0.17333690
Attendance	-0.2920633	-0.1882154	0.01629086	0.32834420	1.00000000	0.62081704
rate_homewins0	-0.3052294	-0.2670264	-0.03969438	0.17333690	0.62081704	1.00000000

In La Liga, the rate of home games won is seen to be highly correlated to attendance, where the coefficient is 0.62081704.

French Ligue 1;

	Fl	Y	R	Sh	Attendance	rate_homewins
Fl	1.0000000	0.35444499	-0.1499726	0.15000413	0.29652828	-0.1174825
Y	0.3544450	1.00000000	0.2094028	0.03893092	0.13642815	0.0677892
R	-0.1499726	0.20940279	1.0000000	0.30570909	-0.14722387	0.1635481
Sh	0.1500041	0.03893092	0.3057091	1.00000000	-0.02268963	0.7555611
Attendance	0.2965283	0.13642815	-0.1472239	-0.02268963	1.00000000	-0.1159506
rate_homewins	-0.1174825	0.06778920	0.1635481	0.75556111	-0.11595058	1.0000000

In this league, the correlation between attendance and rate of home games won is low.

English Championship;

	Fl	Y	R	Sh	Attendance	rate_homewins
Fl	1.000000000	0.946853922	0.4792605	0.9010228	-0.034398297	0.007387099
Y	0.946853922	1.000000000	0.4648242	0.8506330	-0.008600167	0.021153234
R	0.479260469	0.464824154	1.0000000	0.4336381	-0.144653546	-0.108468309
Sh	0.901022848	0.850632960	0.4336381	1.0000000	0.121706920	0.142982713
Attendance	-0.034398297	-0.008600167	-0.1446535	0.1217069	1.000000000	0.387087344
rate_homewins	0.007387099	0.021153234	-0.1084683	0.1429827	0.387087344	1.000000000

The rate of home wins correlation to attendance is a bit higher in this league, at 0.387087344, but still lower than 0.5, for it to be of much significance.

4.4 Discussion

The results in the analysis show that the Covid-19 restrictions have not highly impacted the game outcome. Home advantage has not significantly declined or changed in effect to the absence of fans as was initially expected. If anything, the yellow and red cards are seen to hold greater impact on the home games without fan presence. This is because officials seem to be showing fewer of these cards to the away teams. With the initial setting of restrictions, data showed results to have changed, but as the players got used to stepping out into empty stadiums, the results stabilized reducing the significance of fan presence.

The leagues have also incorporated sound effects in the game to mitigate the lack of fan presence and this has enabled the players to get used to the lack of fans in the stadiums. Therefore, the analysis turns us to reject the null hypothesis that fan absence has significantly reduced home advantage, since the results show that this has not been the case.

While Covid-19 may not have had a serious direct impact on the game itself, the requirements and changes made to ensure protection and safety have posed an economic challenge to commercial partners of the clubs. With a massive reduction in fan attendance and sport interaction by the consumers, matchday revenue has been affected. Clubs are faced with the struggle to find means to retain and also grow their revenues.

For many clubs and leagues, sponsors and commercial gains are the main source of income. For example, in the premier league they account for 28% of total revenue. For the leagues that have matchday revenues as their most significant source, the measures taken to prevent the spread of covid, will greatly impact their funding. While matchday revenues account for 13% of total revenue in the premier league, they account for 21% in the English Championship and an even higher 47% in the Scottish premiership. With the pandemic and quarantine measures, there is need to increase engagement with global fans has risen. The clubs will have to diversify their fan base, while still paying attention and focusing on their existing fan base in order to offset their matchday income deficits.

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

In this chapter, the study focuses on interpretations and implications of the study based on the research questions and objectives. The purpose of this study was to analyse the significance of fan presence on home advantage. The chapter is structured as follows; summary of the findings, conclusion, recommendations as well as the limitations of the study and areas of further study.

5.2 Conclusion

The Covid-19 pandemic created a unique opportunity to research key factors that are usually thought to be key contributors to home advantage in football, which are; crowd, travel fatigue and familiarity or territorial advantage. Previously, research on these factors led us to conclude that the crowd can be significantly impactful on team performance, arguing that their presence spurred the home team to perform better, mainly through increased confidence and motivation and in turn depressing the visiting team performance.

The results from the analysis on home advantage in the absence of fans over the games in the 2019/2020 season and the 2020/2021 league post-covid shows very little change in game outcome as a result of reduced attendance. The first few games may have shown significant change, but that can be attributed to all the restrictions and adjustments that were put in place as a precaution to enhance safety. Once the players grew accustomed to these changes, the game outcome stabilized. The change in performance, with increased games and time in the field over the post-covid period, reduced for the European leagues such that the difference in game outcome between pre-covid and post-covid matches became diminutive.

With an entire season being played with these restrictions and a higher probability of the next season being played under similar restrictions, more data will be obtained and further analysis can divulge the impact of these effects or changes even further. Also, as the disease becomes contained and restrictions are then lessened, further data from the seasons of those periods can be analysed to get more clarity on the contribution of key factors to home advantage.

5.3 Recommendations

From the analysis, the absence of fans in the field may not have significantly impacted game outcome but it has led to the reduction of the number of yellow and red cards issued to the away teams. Since with lesser attendance, fewer cards are doled out, the conclusion arises that the presence of a crowd comes with social pressure on referees to punish the visiting or away team more harshly. Therefore, referees need more support and help rather than admonishment on their decision making. What is deemed to be fairer by the crowd should not be treated with such great importance, to a point where trained referees' decisions are not supported.

There is need to maximize on digital marketing platforms. There will be need to research on fans. Finding new ways to interact with their fans and understand their demographic segments will be key in helping them learn their fans' priorities and the channels they use to engage with the clubs so that they can tailor their approach in engaging with them. This will give them much more authentic interaction with fans and have a key role on the motivation of the players. They can also create means to monetize these engagements through digital marketing.

The impending financial troubles may cause future setbacks for the various leagues and eventually affect game outcome in the long term. Therefore, other means must be created and enhanced to offset income deficit and increase digital reach. This will decrease the eventual long-term effect awaiting the leagues and clubs since the quarantine measures will be in place for quite a while.

5.4 Limitations of the study

While we may insist on the importance of increasing their digital presence, good football data was hard to find. There was a lot of missing information, required for this analysis, on the official websites.

References

1. Pollard, R., & Gómez, M. A. (2014). Components of home advantage in 157 national soccer leagues worldwide. *International Journal of Sport and Exercise Psychology*, *12*(3), 218-233.

2. Dowie, J. (1982). Why Spain should win the world cup. *New Scientist*, *94*(10), 693-695.

3. Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The open sports sciences journal*, *1*(1).

4. Marek, P., & Vávra, F. (2018). Comparison of Home Team Advantage in English and Spanish Football Leagues.

5. Pollard, R., Silva, C. D., & Medeiros, N. C. (2008). Home advantage in football in Brazil: differences between teams and the effects of distance traveled. *Revista Brasileira de Futebol (The Brazilian Journal of Soccer Science)*, *1*(1), 3-10.

6. Loughead, T. M., Carron, A. V., Bray, S. R., & Kim, A. J. (2003). Facility familiarity and the home advantage in professional sports. *International Journal of Sport and Exercise Psychology*, *1*(3), 264-274.

7. Inan, T. (2020). The Effect of Crowd Support on Home-Field Advantage: Evidence from European Football. *Annals of Applied Sport Science*, 0-0.

8. 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.

9. Anders, A., & Rotthoff, K. W. (2014). Is home-field advantage driven by the fans? Evidence from across the ocean. *Applied Economics Letters*, *21*(16), 1165-1168.

10. Boudreaux, C. J., Sanders, S. D., & Walia, B. (2017). A natural experiment to determine the crowd effect upon home court advantage. *Journal of Sports Economics*, *18*(7), 737-749.

11. Pollard, R., & Pollard, G. (2005). Home advantage in soccer: A review of its existence and causes.

12. Goumas, C. (2013). Modelling home advantage in sport: A new approach. *International Journal of Performance Analysis in Sport*, *13*(2), 428-439.

13. Armatas, V., & Pollard, R. (2014). Home advantage in Greek football. *European Journal of Sport Science*, *14*(2), 116-122.

14. Leite, WS (2017). Home advantage: Comparison between the major European football leagues. *Athens Journal of Sports* , *4* (1), 65-74.

15. Sánchez, P. A., García-Calvo, T., Leo, F. M., Pollard, R., & Gómez, M. A. (2009). An analysis of home advantage in the top two Spanish professional football leagues. *Perceptual and motor skills*, *108*(3), 789-797.

16. Goumas, C. (2014). Home advantage in Australian soccer. *Journal of Science and Medicine in Sport*, *17*(1), 119-123.