

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

**An analysis of the influences of “ghost games” on home advantage and odds in betting markets**

**Research Proposal**

**Master thesis Marketing Analytics**

**Fall 2020**

**Name: Alan Rijnders**

**SNR: 2063482**

**Date: 29-11-2020**

# 1 Introduction

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

Many of these fans were left without their favourite pastime for a considerable time when the Covid pandemic struck Europe and other parts of the world in March 2020. Due to strict measures and even complete lockdowns in several countries, almost all sports games were cancelled or postponed until further notice. After a few months of lockdown, the Bundesliga was the first major league to restart their competition to finish the remaining quarter of the games. On the 16<sup>th</sup> of May the Bundesliga recommenced with a packed schedule featuring 6 matches behind closed doors. Other major European leagues such as the Premier League, Serie A and La Liga followed soon and there was even some space in the packed playing calendar to finish the remaining knock-out matches of the Champions League and Europa League.

As a consequence ongoing safety procedures, each match was to be completed behind closed doors with all the players and involved staff would be subject to multiple tests to confirm the safety of everyone participating. Fans around the world were happy to see their favourite teams playing again but at the same time noticed the peculiar and dreary atmosphere surrounding the “ghost games”. Without the fans attendance the matches seemed mere training games, which gives rise to the question what sort of influences this could have on the the outcome of matches. Previous research by (Agnew, G. A., & Carron, A. V. (1994) indicated that home advantage increased significantly with the density of the crowd. Therefore, is home advantage still relevant in the current situation or has it completely disappeared such that playing away can be regarded as beneficial?

## 2 Problem statement

Due to its popularity, an abundant base of research on football and each of its aspects currently exists. The same holds for the possible advantage of the team playing at home, with several papers trying to disentangle the dynamics at play that could lead to home advantage. (e.g. Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007), Pollard & Pollard, 2005). Does the crowd raise home team performance, or is the performance of away teams significantly lower in hostile environments. Additionally, (Neville and Holder, 1999 and Bokyo 2007) find that crowds could influence referee decisions subconsciously in favor of the home team. (Endrich, M., & Gesche, T., 2020) quantify this in their paper where they find that away teams on average receive 0.3 cards less and home teams 0.5 cards more per match when there are no spectators.

The batch of matches without fans provides an excellent occasion to delve deeper in the impact of crowds on football matches since there is an abundance in new data for matches played without crowds. Comparing the scores and detailed statistics of matches played across Europe pre and post covid may serve as a baseline model for the effect of crowds on home team advantage. Questions such as: is home advantage still present in games without a crowd and whether referee behaviour is different without fans will aid in grasping the effect of home supporters on the outcome of football matches.

An interesting extension of this model can be applied to the effect of “ghost games” on price setting behaviour of bookmakers. Sports Betting is heavily on the rise (Markham, F., & Young, M. 2015). In the Market research report on the gambling industry(Online Gambling Market - Global Outlook and Forecast 2018-2023) a projected growth rate of around 10 percent per year is reported. Subsequently, more and more countries across the globe now have legalized sports betting completely( Rodríguez, P., Humphreys, B. R., & Simmons, R. (Eds.). 2017) Partly due to its sheer size, the influence of the gambling industry on football is growing(Sharman, S. (2020). In the Premier League of 2017/18 for example 9 out of 20 clubs featured a bookmaker prominently as kit sponsor. By comparing the odds data for ghost games with the odds for previous seasons it can be checked if the pricing strategy of bookmakers has remained the same or that they adjusted their odds to the current developments.

The aim of the present research is therefore to utilize the current extraordinary circumstances to increase our understanding of the significance and dynamics of home crowd cheering for the home team on player and referee performance. Subsequently, the effect of “ghost games” on price setting behaviour of bookmakers can be examined by analyzing the odds data both during historic seasons as well as during the current games without spectators.

## 2.1 Research questions

Following the aim, the following research questions will be investigated:

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

## 3 Theoretical Framework

### 3.1 Conceptual framework

The main variables of interest will be the venue(home vs away) to see whether there are significant differences in goals and probabilities of winning when playing at home or away. Secondly a variable indicating whether a game was played pre or post covid will be of interest. There are several papers on the effect of “ghost games” on home advantage scrutinizing the German Bundesliga. What limits the interpretation of results in this natural experiment is the incompleteness of the season, if differences in home advantages and team performance between matches with and without crowds are examined there might be a scheduling bias. Tilp, M., & Thaller, S. (2020). control for this by incorporating beforehand expected probabilities of winning for each team and therefore rule out systematic scheduling bias on the proportion of home and away wins. McCarrick, D., Bilalic, M., Neave, N., & Wolfson, S. (2020). Home Advantage during the COVID-19 Pandemic in European football account for the scheduling of the league also through team strength but use the SPI index. Due to availability of data and convenience the

SPI data and associated probabilities will be used to control for scheduling in analysing the home advantage effects.

Rather than calculating difficult statistics of dominance based on corners, shots and target etcetera the expected goals metric will be used to estimate the differences in team performance. The so called “Expected Goals” metric is a metric that calculates the expected goals of a team or player in a certain match based on the quality of chances created by the specific team or player. As (Brecht, M., & Flepp, R. ,2020) outline this metric can serve as a simplified measure for team performance. Similarly here, scheduling needs to be accounted for in the regression model.

Spreading the analysis to bookmakers price setting process, the most important variable are the odds set by bookmakers. Odds are the prices bookmakers set for each outcome associated with the match. By calculating the probability of the event happening plus applying a margin(Cortis, D. (2015)) over this probability the bookmakers try to set their odds such that regardless of the final result and betting behaviour of clients it will always remain profitable. To assess whether bookmakers adjusted their pricing strategy following the “ghost games” it is possible to analyze differences between the true probabilities of the event happening with the bookmaker’s margin provided. The closing odds at Pinnacle(Kaunitz, L., Zhong, S., & Kreiner, J. (2017). provide a very accurate estimation of the true probabilities and therefore are suitable to use as benchmark to compare different bookmakers odds to the true underlying probability.

## 3.2 Hypotheses

1: In general it is expected that the exclusion of home crowds will have detrimental effects on the home advantage for home teams. A home crowd can be a positive stimulus for home team players and can create an intimidating and hostile environment for the opposition.( Ponzo, M., & Scoppa, V. ,2018). Every football fan will know about the famous European nights at Anfield where teams regarded superior in terms of quality found their waterloo against an inspired Liverpool side backed by the roar of the crowd. Similarly in Belgium teams often loathe the away game against Standard Liege because of the extremely passionate home fans creating a very aggressive atmosphere.

2: It is expected that referee bias will be lower in matches without the influence of the home crowd. Referees can be heavily influenced in their decision making by the heavy cheering of the crowd favouring the home team.(Unkelbach, C., & Memmert, D. ,2010). The infamous “Fergie time” illustrates this perfectly. During the tenure of Sir Alex Ferguson at the helm of Manchester United, his side often received longer extra time in home matches, which more than once led to a late escape for the home side. (Endrich, M., & Gesche, T. ,2020) find that referees give more penalties to home teams, and more red and yellow cards to away teams.

3: As outlined before it is expected that both home and away team performance are affected by the exclusion of supporters onto the football grounds. Home crowds can raise the performance of the home team and intimidate the away team. It is expected therefore that in “ghost games” home teams on average will have less corners, shots and shots on target whereas the away team on average will see an increase in corners, shots and shots on target compared to before covid-19. First signs of these effects are shown by (Fischer, K., & Haucap, J. (2020). in their investigation of post covid Bundesliga games.) Therefore it is expected that expected goals for home teams will be significantly lower and for away teams significantly higher. A striking example of affected away performance is that of Jens Toornstra in 2016/2017. Toornstra as a midfielder managed to score an impressive 14 goals throughout the season, of which all 14 came at home in the Kuip.

4: Betting odds are initially set by bookmakers based on advanced statistics and predictive techniques to reflect the true probabilities as much as possible. These betting odds possess a certain home bias( Deutscher, C., & Winkelmann, D. 2020).which implies that the odds for home teams are higher than the actual underlying probability of a home victory. If this exceeds the margin kept by the bookmaker it might be a profitable strategy to bet on the home team. It is expected that in the new circumstances especially in the beginning bookmakers won’t adjust their odds accordingly to the new situation of “ghost games” and therefore will underestimate away team strength in their odds.

## 4 Methodology

### 4.1 Data Collection

Data on football matches is very abundant. Companies such as Opta are specialised in sports analytics and have a separate branch that focuses on producing huge amounts of data specifically for football. Therefore the main task is not to find data but choose the right type of data to be utilized. The website: <http://www.football-data.co.uk/data.php> has weekly updated datasets for all the important leagues around the world. The datasets start in the season 2000-2001. The datasets include statistics on full-time and half-time results, amount of shots and shots on targets of both home and away teams as well as the amount of yellow and red cards and corners for each team. Furthermore, betting odds at major bookmakers are provided. Which is very convenient for the subsequent odds pricing analysis. Data on expected goals is more difficult to find, Opta for example requires payment to receive access to their data. Fortunately through <https://projects.fivethirtyeight.com/> data on expected goals for matches in major football leagues around the world from 2016 onwards can be obtained. The datasets on their website are updated weekly as well.

### 4.2 Modelling

The following linear regression equations will be modeled to analyze the extent of home advantage, referee bias and team performance. Venue indicates(home vs away), COVID-19 indicates whether match was played with or without public and the interaction effect between them. SPI

**Equation 1:** Points\_at\_home:  $\beta_0 + \beta_1 \text{venue} + \beta_2 \text{SPI} + \beta_3 \text{COVID-19} + \beta_4 \text{Venue} * \text{COVID-19} + \beta_5 \text{Venue} * \text{SPI}$

**Equation 2:** Expected\_Goals:  $\beta_0 + \beta_1 \text{venue} + \beta_2 \text{SPI} + \beta_3 \text{COVID-19} + \beta_4 \text{Venue} * \text{COVID-19} + \beta_5 \text{Venue} * \text{SPI}$

**Equation 3:** Fouls:  $\beta_0 + \beta_1 \text{venue} + \beta_2 \text{SPI} + \beta_3 \text{COVID-19} + \beta_4 \text{Venue} * \text{COVID-19} + \beta_5 \text{Venue} * \text{SPI}$

**Equation 4:** Yellow Cards:  $\beta_0 + \beta_1 \text{venue} + \beta_2 \text{SPI} + \beta_3 \text{COVID-19} + \beta_4 \text{Venue*COVID-19} + \beta_5 \text{Venue*SPI}$

**Equation 5:** Red Cards:  $\beta_0 + \beta_1 \text{venue} + \beta_2 \text{SPI} + \beta_3 \text{COVID-19} + \beta_4 \text{Venue*COVID-19} + \beta_5 \text{Venue*SPI}$

Due to the categorical nature of betting odds linear regression is not deemed a suitable statistical tool for the final research question. As (Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998).outline in their book in the case of a categorical dependent variable a multinomial logit or probit model is preferred to analyze the odds data. In general the logit model is slightly preferred over probit in general applications(Dow, J. K., & Endersby, J. W. 2004).This combined with the non binary nature of our data makes the logit model the most convenient model choice.

The following logistig regression equation will be estimated: where Implied probability = 1/odds,

**Equation 6:** P-won(=1):  $\beta_0 + \beta_1 \text{Implied Probability} + \beta_2 \text{Venue} + \beta_3 \text{SPI} + \beta_4 \text{COVID-19} + \beta_5 \text{Venue*COVID-19} + \beta_6 \text{Venue*SPI}$



## 5 References

1. Agnew, G. A., & Carron, A. V. (1994). Crowd effects and the home advantage. *International Journal of Sport Psychology*.
2. 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.
3. Brechot, M., & Flepp, R. (2020). Dealing with randomness in match outcomes: how to rethink performance evaluation in european club football using expected goals. *Journal of Sports Economics*, 21(4), 335-362.
4. Deutscher, C., & Winkelmann, D. (2020). Bookmakers' mispricing of the disappeared home advantage in the German Bundesliga after the COVID-19 break. *arXiv preprint arXiv:2008.05417*.
5. Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial logit: a comparison of choice models for voting research. *Electoral studies*, 23(1), 107-122.
6. Endrich, M., & Gesche, T. (2020). Home-bias in referee decisions: Evidence from “Ghost Matches” during the Covid19-Pandemic. *Economics Letters*, 197, 109621.
7. Fédération Internationale de Football Association.(2019), FIFA activity report 2018. Retrieved from <https://www.fifa.com/who-we-are/official-documents/>
8. Fédération Internationale de Football Association. (2007). FIFA Big Count 2006: 270 million people active in football. *Zurich: Fédération Internationale de Football Association*.
9. Fischer, K., & Haucap, J. (2020). Does crowd support drive the home advantage in professional soccer? Evidence from German ghost games during the COVID-19 pandemic.
10. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice hall.
11. Kaunitz, L., Zhong, S., & Kreiner, J. (2017). Beating the bookies with their own numbers-and how the online sports betting market is rigged. *arXiv preprint arXiv:1710.02824*.

12. Markham, F., & Young, M. (2015). "Big gambling": the rise of the global industry-state gambling complex.
13. Nevill, A.M., Holder, R.L.(1999). Home Advantage in Sport. *Sports Med* **28**, 221–236
14. Online Gambling Market - Global Outlook and Forecast 2018-2023. (2018), Retrieved from <https://www.researchandmarkets.com/>
15. Pollard, Richard,(2006), Home Advantage in Soccer: Variations in its Magnitude and a Literature Review of the Inter-related Factors Associated with its Existence , *Journal of Sport Behavior*, Vol. 29, Iss. 2, 169-189.
16. Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876–2003). *Journal of sports sciences*, 23(4), 337-350.
17. Ponzio, M., & Scoppa, V. (2018). Does the home advantage depend on crowd support? Evidence from same-stadium derbies. *Journal of Sports Economics*, 19(4), 562-582.
18. Rodríguez, P., Humphreys, B. R., & Simmons, R. (Eds.). (2017). *The Economics of Sports Betting*. Edward Elgar Publishing.
19. Sharman, S. (2020). Gambling in football: How much is too much?. *Managing Sport and Leisure*, 1-8.
20. Tilp, M., & Thaller, S. (2020). Covid-19 has turned home-advantage into home-disadvantage in the German Soccer Bundesliga. *Frontiers in Sports and Active Living*, 2, 165
21. Unkelbach, C., & Memmert, D. (2010). Crowd noise as a cue in referee decisions contributes to the home advantage. *Journal of Sport and Exercise Psychology*, 32(4), 483-498.