CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Home advantage in football during COVID-19 pandemic

Bachelor's thesis

Author: Filip Šnejdr

Study program: Economics and Finance

Supervisor: Mgr. Petr Polák, MSc. Ph.D.

Year of defense: 2022

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, January 4, 2022

Filip Snejdr

Abstract

The aim of this thesis was to investigate the change of home advantage in football during the pandemic of COVID-19. The effect of empty stadiums was analysed on data from top four European leagues with the highest average attendance. Collected match statistics from seasons 2016/17-2018/19 played with spectators were compared with statistics from the season 2020/21 played almost completely without spectators. The results of Welch's t-test showed that the differences in the match statistics between home and away teams varies across all the analysed leagues. The most substantial decline during the season 2020/21 of home advantage was found in the English Premier league. The difference in shots on target significantly dropped in all leagues, except German Bundesliga. The final part of thesis analysed the accuracy of betting odds proposed on home team victory. Brier score showed a significant drop in accuracy of betting odds in the matches ended in home victory across all four leagues. Afterwards, prediction accuracy of betting odds within equal sized groups with similar level of probability was analysed. However, any signs of systematically overrated betting odds proposed on home teams victory were not found.

Keywords	football, COVID-19, home advantage, match statistics, betting odds, attendance
Title	Home advantage in football during COVID-19 pandemic

Author's e-mail	`snejdr.f@seznam.cz`
Supervisor's e-mail	`polakpet@gmail.com`

Abstrakt

Cílem této práce bylo zjistit jak se změnila výhoda domacího prostředí ve fotbale během pandemie COVID-19. Na datech ze čtyř evropských lig s nejvyšší průměrnou návštěvností byl zkoumán vliv prázdných stadionů. Sesbírané zápasové statistiky ze sezón 2016/17-2018/19, které se odehrály s diváky, jsou porovnány se statistikami ze sezóny 2020/21 hrané téměř zcela bez diváků. Výsledky Welchova t-testu ukázaly, že rozdíly ve statistikách zápasů mezi domácími a hostujícími týmy se ve všech analyzovaných ligách liší. Nejvýraznější pokles výhody domácího prostředí během sezóny 2020/21 byl pozorován v anglické Premier League. Rozdíl ve střelách na bránu výrazně klesl ve všech ligách, kromě německé Bundesligy. V závěrečné části práce byla analyzována přesnost sázkových kurzů vypsaných na vítězství domácího týmu. Brierovo skóre ukázalo pokles přesnosti sázkových kurzů v zápasech, které skončily domácím vítězstvím ve všech čtyřech ligách. Poté byla analyzována přesnost předpovědí sázkových kurzů v rámci stejně velkých skupin s podobnou pravděpodobností. Nebyly však nalezeny žádné známky systematicky nadhodnocovaných sázkových kurzů vypsaných na vítězství domácích týmů.

Klíčová slova	fotbal, COVID-19, domácí výhoda, zápasové statistiky, kurzové sázky, návštěvnost
Název práce	Výhoda domácího prostředí ve fotbale během pandemie COVID-19
E-mail autora	snejdr.f@seznam.cz
E-mail vedoucího práce	polakpet@gmail.com

Acknowledgments

First of all, I would like to thank Mgr. Petr Polák, MSc. Ph.D., who has given me much useful advice regarding my thesis. I also thank to my family for all the unconditional support.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Snejdr, Filip: *Home advantage in football during COVID-19 pandemic.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2022, pages 53. Advisor: Mgr. Petr Polák, MSc. Ph.D.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Professional football leagues were paralysed by the outbreak of COVID-19 pandemic in the second half of the season 2019/20, which had to be finished after unprecedented two months interruption. Thereafter all matches were played with various attendance restrictions, mostly completely without spectators. In such circumstances, it is possible to assume that the advantage of home teams could be distorted. Hence the question arises: Did the home advantage change in matches played during the pandemic? In case the research finds the right answer before the end of pandemic restrictions, it might help in many decision making spheres. Teams might modify game strategy, new rules might be set in the game, new betting strategies might arise or preferences in football players market might be changed.

Although many recent articles related to change in home advantage could be found, most of them analysed only one particular country and did not provide a comparison of several leagues (Fischer & Haucap 2020), did not include matches played during pandemic at all (Sors *et al.* 2020; Reade *et al.* 2020) or included into the analysis only the matches played without spectators during the second part of the season 2019/20, and therefore the season was not balanced regarding the presence of spectators in both matches played between each two teams in the season (Scoppa 2021; McCarrick *et al.* 2021). This thesis analyses five years period of the top four European leagues since the season 2016/17. However, the season 2019/20 was excluded from the main part of the analysis in order to avoid unbalanced schedule of matches in one season.

The diminished home advantage during the mentioned period should imply that betting odds proposed on home teams should be relatively lower than odds proposed on matches played with spectators. The question is whether bookmakers could properly adjust betting odds or if the accuracy of their odds descended substantially. Tiitu (2016) and Kuypers (2000) also aimed on evaluating of odds accuracy, which is the aim of the final part of this thesis. The accuracy is evaluated by using Brier score and, moreover, groups of matches with similar level of odds are compared with the help of approximation of binomial distribution with the normal distribution.

The beginning of the work outlines a brief history and the development of home advantage in team sports, particularly in English football. After that, the key determinants of home advantage are presented with emphasis on the crowd related factors. Section of literature review is concluded with recent research regarding home advantage during the pandemic. The following section describes the process of collection and adjusting data and provides a summary of collected match statistics. Betting odds adjustment is presented at the end of this section. The first part of the main analysis tries to answer the question that home advantage diminished during COVID-19 pandemic in terms of detailed match statistics. The match statistics comparisons of every single league are presented in their own subsections. The second part of analysis then aims to ascertain whether the accuracy of betting odds proposed on home team victory are predicted well in comparison with previous seasons. Limitations section highlights that absence of spectators was not the only change during the pandemic. In the last section the work is concluded.

Chapter 2

Literature Review

This chapter is divided into three sections and should outline the literature regarding home advantage. Section 2.1 provides an overview of home advantage (in terms of win percentage or gained points) in the history of football leagues, starting with a brief comparison with other team sports. Section 2.2 presents many determinants of home advantage and suggests that the crowd could be the one of the most important. Section 2.3 brings an overview of the latest research concerning COVID-19 pandemic and home advantage in empty stadiums.

2.1 Home Advantage in the Past

Nevill & Holder (1999) analysed more than twenty studies of the major team sports in which the presence of the home advantage was indicated. Most of them were focused on basketball, American football or ice hockey. While the highest home winning percentage was reached in soccer, the lowest was reached in baseball. However, all drawn games were excluded from their analysis, hence the results could be uneasily compared with the results of this work. The assumption of no home advantage was clearly rejected by a simple binomial test in all analysed sports. That supports the conclusion of Schwartz & Barsky (1977) that home advantage exists in major team sports, however, also substantially varies among different sports.

Looking at home advantage development in English football, Pollard (1986) reveals that the home advantage in terms of gained points exists since the foundation of the Football league in 1888. The home advantage was calculated as a percentage ratio of gained points and maximum possible amount of gained points. His work shows that during 1888-1900 home advantage was nearly 68%

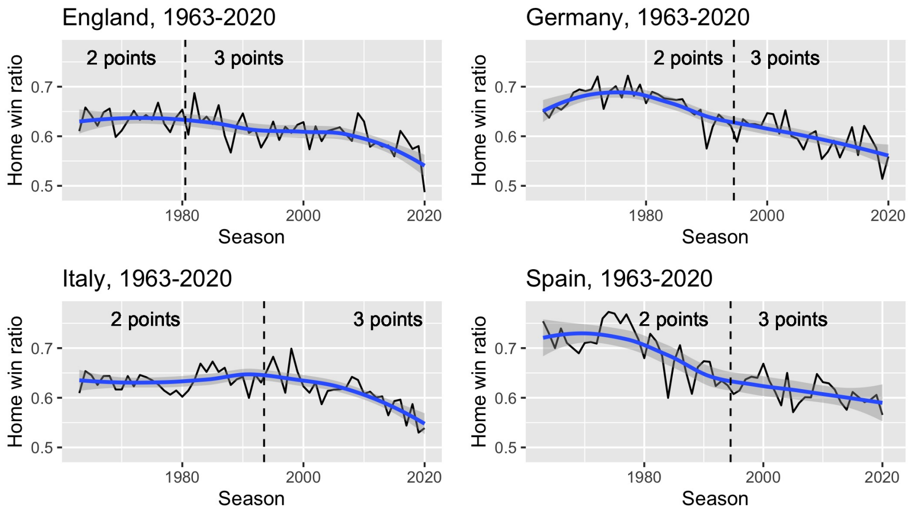
in the First Division, keeping this rate steady in the next four decades. After a pause caused by the World War II, home win rate significantly dropped to around 62.5%. Period since the season 1963 (for simplification, 'season 1963' will refer to a start year of the season, 'season 1963/1964' in this case) is already depicted in Figure 2.1. Clarke & Norman (1995) calculated an average home win percentage to be 62.1% from season 1981 to season 1990, which corresponds to the figure mentioned above.

Until the season 1981 the winning team in English league had gained two points, afterwards it was changed to three points. English league was the first to introduce the new rule. This was clearly one of the biggest rule change in modern football history and could possibly change the home advantage. Awarding a winner by one extra point could motivate both teams to risk more and rather gain three points then only one for a draw. This change of rules is depicted in Figure 2.1 by dashed vertical line.

The change also influenced Pollard's computation of home advantage ratio. When two points for a win were awarded, the home advantage ratio computed by Pollard was simply the ratio of points gained by home teams during whole season and the total number of points awarded among all teams - number of matches played in a season multiplied by two points. After the rule change, a win was relatively better awarded compared to a draw.

In order to make both periods more comparable, the difference in home win ratio computation is explained by the following example. Considering 462 matches in a season (corresponds to twenty-four teams in the season 1981), and maximum possible amount of points 924, a win until the season 1981 added 0.216 percentage point regardless the number of draws in a season. However, since the season 1981, a percentage gain per each win depends on the overall number of draws in a season. Every draw decreases total amount of points awarded among all teams and, therefore, increases the percentage gain per win. The extreme case without a draw in the season would lead to the same percentage point gain per win as before (0.216). While 121 draws occurred in the season 1981, every three-point win added 0.237 percentage point. On the other hand, this also applies for the case that away team wins. Every draw then adds relatively less percentage points to home win ratio (0.079 instead of 0.108).

Figure 2.1: Home win ratio over time



Notes: Home win ratio expresses the percentage rate of points gained by home teams from the maximum possible amount of points per each season. Source of data: Curley (2016).

Altogether the difference in computation of home advantage caused by the change in awarding points in the season 1981 would be 0.9 percentage point he first season after the rule change. In addition, the average difference of 1.1 percentage points during the first ten seasons after the change could help to compare both periods captured in Figure 2.1. Although the rule change could affect both teams tactics, on average, home win ratio in the three points period is relatively overestimated from that perspective. Hence, the decreasing trend depicted in Figure 2.1 after the change would be even lower, if the home advantage was counted by two points system.

In other top European leagues the rule of three points was adopted more than ten years afterwards. Italy established the rule in 1994, Germany and Spain in 1995. The change is depicted by the vertical line in each plot. It was expected that teams would adjust to the new rule in several seasons. The lowest relative difference between two points and three points period in the first ten years after the change would be in Germany, on average 1.05 percentage points. The highest possible difference would be in France, on average about 1.44 percentage points.

Garicano & Palacios-Huerta (2005) compared results from Spanish La Liga from the season 1994-1995, the last one in two points period, and the season 1998-1999. They found increased offensive effort and number of fouls, however, no significant change in scored goals. Nevertheless, they concluded that the intention of raising the award to three points might not help to attract more attendance as the leading team tried to keep the lead only and played much more defensively.

Peeters & van Ours (2021) show in their analysis of attendance in English divisions that during years 1975-2020 the change in awarding points also might help to increase attendance. They found that home advantage increases with stadium attendance, however, the fluctuation from year to year was substantial. Moreover, they found secular decline in home advantage in past four decades. Both, the fluctuation and the decline can be seen in Figure 2.1.

2.2 Determinants of Home Advantage

As the presence of the home advantage was clarified, the focus of academic research was shifted to factors determining the advantage. Courneya & Carron (1992) divided them into 4 categories: a crowd factor, a familiarity/territoriality factor, a travel factor and a rule factor. Pollard (1986) added also special tactics or psychological factors. It should be noted that not all factors might be affected by pandemic situation.

Familiarity with team's own stadium and a need to defend its territory should remain unchanged during COVID-19 period. Both mentioned may help the team to increase probability of a win at home. Pollard (2002) analysed reduction in home advantage while a team moves to a new stadium in professional basketball, baseball and ice hockey, and suggested that approximately 25% of home advantage could be explained by familiarity of home stadium.

Travel fatigue may have a negative effect on away team result. Such evidence provided Pollard *et al.* (2008) based on Brazilian football league, where teams travel much larger distances than in European countries. It is assumed that neither travelling within one country might not be strongly changed due to COVID-19 restrictions for teams from top European leagues.

The new rule in substitution has been established due to COVID-19 pandemic. Instead of three substitutions per match, team could have two more due to busy schedule in the end of 2019/2020 season. The rule change remained the same in most European leagues during all season 2020/2021, only the English premier came back to three substitutions. This rule could eventually affect both teams.

All mentioned factors, except crowd factor, should not dramatically affect a potential change in home advantage. Therefore, the main factor affecting the home advantage during the COVID-19 pandemic could be a crowd factor.

Absolute crowd size might not be so important contribution to home advantage, according to Courneya & Carron (1992). That is in line with Pollard (1986) and Clarke & Norman (1995) who, despite the large differences in crowd size, also found only little variation in home advantage. Pollard (2006) then added that decline in home advantage seemed to be nonlinear over nine level of competitions in England. Among the top four levels the advantage remained over 60%, although the differences in crowd size were large. Remaining five levels, with average attendance below three thousand spectators, had home advantage around 55% regardless the level of competition. Interestingly, an average attendance even below one hundred spectators had very similar rate of home advantage. On the other hand, Goumas (2013) found no home advantage at level below one thousand spectators on a global scale. At least two thousand spectators were necessary to substantial (more than 55%) home advantage in all continents except for Europe, where the minimum number to reach the advantage was around ten thousand spectators.

Crowd density could be calculated as the ratio of attendance and the maximum capacity of a stadium. Dawson & Dobson (2010) analysed the impact of social pressure on referees and found that crowd density has significant impact on a referee's decision, measured in disciplinary points (one point for yellow card, two points for red card), and matters more than absolute crowd size. Although the impact on away team's punishment was larger, the home team was also significantly affected. According to Goumas (2013), the crowd density may have also positive impact on increasing of home advantage.

Crowd in derby matches is a special case where both teams either share the same home stadium or travel mostly within one city. Ponzo and Scopa (2014) focused on derbies played at the same stadium in order neutralize other factors as familiarity and travel fatigue. They found significant impact of the crowd on the home team advantage, although it was lower than in normal "no derbies" matches. The probability that a team won a home game was 13% higher than for away game and home teams also scored about 0.45 goals more.

2.3 Home Advantage during COVID-19 Pandemic

Fischer & Haucap (2020) aimed at the change of home advantage during COVID-19 period in German Bundesliga. Using the data from the last four seasons (2017/18-2020/21) they found the probability of a home win might have decreased by 12 percentages points.

The impact of empty stadiums during 2002-2020 investigated also Reade *et al.* (2020), however, matches during COVID-19 pandemic were excluded in the main part of their analysis. They observed that home advantage was on average lower without fans, mainly due to relatively more punishments for the home team. In conclusion, they also suggest that the effects of no fans could diminish if the participant become accustomed to the new conditions.

Sors *et al.* (2020) also analysed the change in home advantage and referee bias during COVID-19 pandemic among top European leagues. Analysis of two highest divisions from England, Germany, Italy and Spain during seasons 2016-2019 showed significantly lower number of wins of home teams in matches without fans. Their results did not reveal that the referee would favour any of the teams, hence they suggest that a referee bias might diminish during COVID-19 pandemic.

On the other hand, McCarrick *et al.* (2021) pointed out that team dominance (attacking or defending team setting) could also affect the referees' decisions. The more dominating the team was during a match played without fans, the more they were punished. In their analysis, fifteen different European leagues from eleven countries finishing the season 2019/2020 without fans were included. They showed that matches without fans had a significant negative impact on home teams, measured in points or goals per game.

Without fans, Scoppa (2021) found the home advantage to be almost halved in terms of points, goals or shots. His paper included data of two highest divisions from England, Germany, Spain, Italy and Portugal during seasons 2010-2020. He also found that the referees favour much less home teams during COVID-19 pandemic than before and provided evidence that more balanced decisions might result from the empty tribunes rather than the lower intensity and aggressiveness of play.

Chapter 3

Data

3.1 Data Selection

For the analysis, top four European leagues with the highest average attendance (2010/11 - 2016/17) according to Colombo & Batardière (2018) were chosen. The highest attendance implies the biggest possible difference between a standard situation with spectators and empty stadiums during COVID-19 period. The chosen leagues from the highest to the lowest average attendance: German Bundesliga, the English Premier League, Spanish La Liga and Italian Serie A. Data for the last five complete seasons (2016/2017 - 2020/2021) were collected from the sources described in Section 3.2.

Table 3.1 shows how the second half of the season 2019/2020 was forced to be played. 'Last match' refers to the last match played with spectators and also the last day of match played before the season had been interrupted. 'Restart 19/20' indicates how long the interruption lasted. 'End 19/20' and 'Start 20/21' indicates the end of season 2019/2020 and the end of the season 2020/2021, respectively.

Table 3.1: Season 2019/2020 - Start of COVID-19 Pandemic

Country	Last match	Restart 19/20	End 19/20	Start 20/21
England	2020-03-10	2020-06-17	2020-07-26	2020-09-12
Germany	2020-03-11	2020-05-16	2020-06-27	2020-09-18
Italy	2020-03-10	2020-06-20	2020-08-02	2020-09-19
Spain	2020-03-10	2020-06-11	2020-07-19	2020-09-12

All competitions were stopped in the same week in March during the season 2019/2020. German Bundesliga was restarted in May, as the first one without the fans, the other leagues were restarted one month afterwards.

After the restart, the rest of the season had to be played in much shorter time then was originally scheduled. Last nine rounds of Premier League and Bundesliga were played in five and six weeks, respectively, instead of scheduled ten weeks. Last eleven rounds of La Liga and thirteen rounds of Serie A were played in five and six weeks, respectively, instead of scheduled eleven weeks. The next season 2020/2021 started three weeks later than ordinarily in Germany and Spain and four weeks later in England and Italy. All competitions ended as usually in the season 2020/2021, in the third week of May, therefore the schedule of this season was also shorten.

Table 3.2: Descriptive Statistics - Number of Matches

	Germany	**England**	**Italy**	**Spain**
Teams	18	20	20	20
Rounds	34	38	38	38
Matches 16/17-18/19	918	1140	1140	1140
19/20 before pandemic	223	288	269	256
19/20 empty stadium	83	92	111	124
20/21 empty stadium	273	349	376	366
20/21 restricted	33	31	4	14
Matches Total	**1530**	**1900**	**1900**	**1900**
All leagues:				
with fans		5374		
restricted		82		
empty stadium		1774		
Total		**7230**		

In total, 7230 matches were involved into the analysis, of which 5374 were played before the COVID-19 pandemic. The English Premier League, Italian Serie A and Spanish La Liga consist of twenty teams, hence in total 1900 matches were played during five seasons. German Bundesliga is played by eighteen teams only, which results in 1530 matches in five seasons. At the end of the season 2019/2020, 410 matches were played behind closed doors, and subsequently the entire season 2020/2021 was forced to be played in empty stadiums (1774 matches) or only with a restricted number of spectators (82 matches). The number of spectators varied only between one and eleven thousand, with the mean of almost 4500 spectators, in the vast majority of restricted number of spectatorsâ€™ matches.

3.2 Data Sources

Two main sources were *www.FiveThirtyEight.com* and *www.FootyStats.org*. The first one, *FiveThirtyEight*, is American website focused on analyses of politics, sports and science, which was used to obtain data about SPI ratings, expected goals and match importance. All three variables will be explained later in this chapter. *FootyStats* is a web page focussed on football match statistics and other data related to football betting market. Dataset downloaded from this website contains basic match statistics (score, shots, fouls, cards, possession or corners), match attendance and betting odds.

3.3 Data Adjustment

Both datasets were adjusted and merged in R studio and Excel. Both sources uses slightly different team names, hence they had to be unitized in one joint dataset. Some missing values in attendance during COVID-19 period were manually corrected and validated with help of *www.livesport.cz* website. Several variables had to be calculated, including home and away points, COVID variable (whether the match was played during COVID-19 pandemic or not) or shots-related statistics. The other calculated variables, shots accuracy and save percentage, are described in Subsection 3.4.2.

3.4 Data Characteristics

3.4.1 Match Statistics

Basic post-match statistics for all four leagues and five seasons together are summarized in Table 3.3 and should provide a broad overview. For all the matches in analysis, data for points, goals, expected goals, possession, shots, shots on target, corners, fouls, and yellow and red cards were collected. The identical statistics will be analysed later with respect to each league and season. Obviously, the table shows that home teams had almost all statistics better in terms of football productivity. On average, home teams gained 0.4 more *points* per game, scored 0.31 more *goals* per game and have 0.3 *expected goals* per game more in statistics.

Table 3.3: Match Statistics - Summary

	mean	**std.dev**	**min**	**max**	**home − away**
Home points	1.58	1.33	0	3	0.40
Away points	1.18	1.29	0	3	
Home goals	1.56	1.31	0	9	0.31
Away goals	1.25	1.19	0	9	
Home expected goals	1.56	0.90	0	7.07	0.30
Away expected goals	1.26	0,79	0	5.9	
Home possession	50.89	11.19	17	84	1.78
Away possession	49.11	11.19	16	83	
Home shots	13.03	5.10	0	37	2.08
Away shots	10.95	4.60	0	40	
Home shots on target	5.74	2.68	0	19	0.85
Away shots on target	4.89	2.40	0	17	
Home corners	5.43	3.00	0	20	0.91
Away corners	4.52	2.66	0	17	
Home fouls	12.43	4.15	0	29	− 0.19
Away fouls	12.62	4.18	0	32	
Home yellow cards	1.95	1.36	0	8	− 0.20
Away yellow cards	2.15	1.37	0	8	
Home red cards	0.08	0.29	0	3	− 0.02
Away red cards	0.10	0.32	0	2	

Notes: All leagues together, seasons 2016/17 - 2020/2021.

Possession, the percentage rate of keeping the ball under control by one team, also little favours the home teams (50.89% vs 49.11%). Team with higher possession should mostly control the game more, and therefore might have a higher chance to win. Overall number of *shots* and *shots on target* are statistics describing the offensive power of teams and more shots might imply a higher chance to win. Home teams have both statistics better, on average, home teams have 2.08 shots and 0.85 shots on target per game more than away teams. *Corners* is another statistics which suggests that home teams have certain advantage, every game they have 0.91 more corners than away teams. *Fouls*, *yellow cards* and *red cards* are the only statistics nearly identical for both teams. While away teams on average commit 0.19 fouls more per game, they are awarded only one yellow card more per five matches and one red card more in fifty matches than home teams.

Further analysis of match statistics will be brought in the next chapter, where all variables will be divided into two periods, pre-Covid and Covid, to compare the differences between both periods.

3.4.2 Shots-related Statistics

Expected goals is specially computed statistics based on FiveThirtyEight database, which tries to include not only exact number of goals or shots on target, but also the way a goal was (or was not) scored. How well the model of expected goals works is well described on their web page. The model assigns a probability of scoring a goal to each shot, depending on several circumstances. The main part of calculation depends on distance and angle from the point the shot was taken. The part of the body the shot was taken with is also a part of the model, together with player's ability to score a goal. For instance, a shot taken by Lionel Messi was on average converted into a goal 1.4 times more frequently than expected and, therefore, each shot taken by him multiplies the probability of scoring by 1.4. Each shot of the match is converted into a probability of scoring a goal by this method and all probabilities are summed up. The total sum corresponds to the number of expected goals by each team. This method might help to better understanding the course of the game. Significantly higher expected goals than scored goals could indicate that the team could have probably scored more goals in the match. One of the possible explanations of missing a shot, which was assigned a high probability of scoring a goal, could be the presence of spectators in the stadium. The difference in goals and expected goals in pre-Covid and Covid period will be compared in the next chapter, which might show if the presence of spectators negatively affected the shot productivity of away teams more than home teams.

Shots accuracy is calculated as the ratio of the total number of shots on target in the league during whole period (2016/2017-2018/2019 or 2020/2021) and the total number of shots off and on target together.

Goalkeeper save percentage is the ratio of total number of goals and the total number of shots on target. This statistics might be affected not only by goalkeepers' ability, but also by strikers' accuracy.

Expected goals/score ratio is the ratio of the total number of expected goals and the total number of goals. The ratio equal to one would mean the same number of expected goals and scored goals, therefore the scored goals would correspond to summed up probabilities assigned to taken shots. The ratio less than one would indicate that teams on average scored more goals than expected the probability model of expected goals. On the other hand, the ratio more than one would indicate that teams were supposed to score more goals than actually scored.

3.4.3 Betting Odds

Betting odds data of all matches were collected from FootyStats. Some matches had to be omitted from the analysis due to missing values, in total 7118 matches out of 7230 were included. Afterwards, all the odds were adjusted in order to obtain the actual probability hidden in the odds set by bookmakers. Evaluating of bookmaker's accuracy requires the sum of probabilities of all possible outcomes being equal to one. The format which the odds were collected in is called decimal odds (the standard European format). In this format the odds are usually presented by betting offices.

It could be explained on the example of the match between FC Liverpool and West Ham United from season 2019/2020, which ended in victory for the home team. The odds were 1.16 - 8.00 - 16.00 on home win, draw and away win, respectively. It implies that if a bettor placed a bet of €10 on home team, he would receive €11.6. For purpose of this analysis, decimal odds were converted into probabilities by calculation of reciprocal values of decimal odds. In case of former example, reciprocal values of the odds would be 0.862 - 0.125 - 0.063. Obviously, the sum of implied probabilities (1.05) is greater than one. The additional 5% above 100% threshold, so called bookmakers' margin, should be 'removed' from actual probabilities. Let $odds_{HW}$, $odds_D$ and $odds_{AW}$ denote decimal odds for home win, draw and away win, respectively. Then actual probability of home win ($prob_{HW}$) would be computed as $prob_{HW} = (1/odds_{HW})/(1/odds_{HW} + 1/odds_D + 1/odds_{AW})$.

After adjusting the betting odds of the latter example, the actual probabilities would be 0.821 - 0.119 - 0.060. According to García *et al.* (2017), the bookmaker's margin might not be divided equally among all three possible outcomes. However, for the purpose of this analysis the odds were recalculated as the margin would be equal regardless the outcome.

Chapter 4

Match Statistics Comparison

4.1 Methodology

This section aims to compare sample means of three complete seasons 2016/2017, 2017/2018 and 2018/2019 played with spectators with the last season 2020/2021 played almost completely without spectators (approximately 5% of matches was played with only restricted number of spectators) and test the hypothesis that home advantage dropped during matches played without spectators. In order to avoid comparison of matches in unbalanced schedule, where two teams would play the first match with fans and the second match without them in one particular season, the season 2019/20 was completely excluded from this part of analysis.

First of all, for each league the bar chart (Figures 4.1-4.4) comparing percentage rate of match outcomes in each season was created. The percentage rate was calculated on a basis of number of outcomes in the season over 380 matches (306 matches for German Bundesliga) played in every season. Season 2019/2020 was divided into two parts depending whether it was played before COVID-19 pandemic with spectators or during the pandemic in empty stadiums.

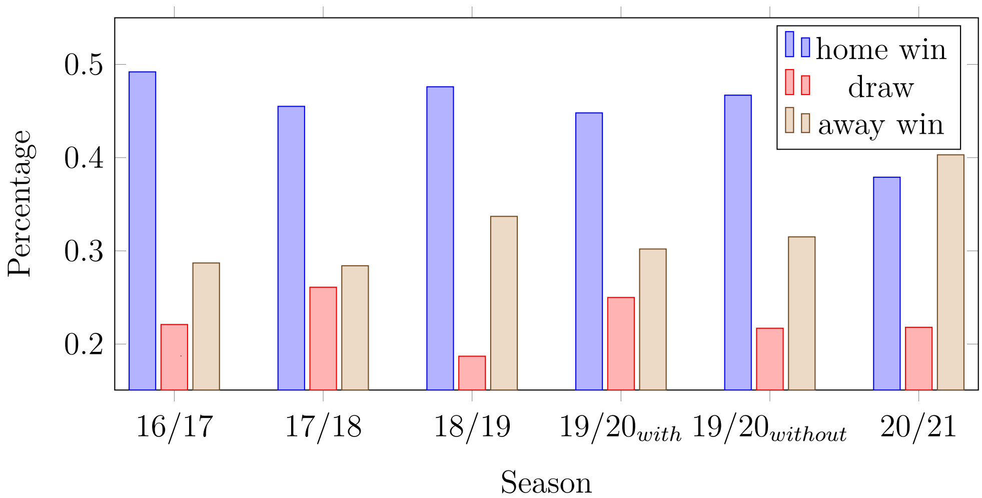
Consequently, match statistics of three complete seasons 2016/17-2018/19 and season 2020/2021 were compared. For every statistics in each match the differences between home and away team were computed. Afterwards, means of the differences in home and away statistics for both periods were used to perform Welch's t-test. Reade *et al.* (2020) and Sors *et al.* (2020) also used two-sided t-test to compare matches played with fans with those played behind closed doors, however the former work excluded matches from COVID-19 period. The latter analysed two highest divisions from Spain, England, Germany and Italy together, which could be compared to results of this section. Tables 4.1-4.4. then shows the results of t-tests and corresponding means in the tables denotes mean differences between home and away teams per match. Wilcox (2003) states that Welch's modification is suitable for comparison of means of two samples with possibility of unequal variances and unequal sample sizes. The null hypothesis of equality of both means (that there was no difference in means between pre-Covid and Covid period) was tested.

After that, Chi-square tests verifying the difference in outcome frequencies between the two periods were conducted as a supplementary analysis. Similar approach. Match outcomes frequency from seasons 2016/17-2018/19 served for calculation of expected match outcomes frequency for season 2020/21. The null hypothesis of equality of both frequencies was tested and the results can be seen in Table A.2.

4.1.1 England - Premier League

The percentage rate per season of every possible outcome of match (home win, draw, away win) in English Premier League during seasons 2016/2017-2020/2021 is depicted in Figure 4.1. As percentage rate of draws remains on relatively same level across all seasons, the rate of away wins substantially ascended and the rate of home wins descended in season 2020/2021. Moreover, for the first time in history of the highest English football league more away wins than home wins occurred in one season.

Figure 4.1: England, Premier League



Notes: Seasons 16/17 - 18/19 and part of 19/20 played with fans, the second part of season 19/20 and season 20/21 without fans.

The results of Chi-square test in Table A.2 indicates that frequencies observed for season 2020/2021 were significantly different from the expected frequencies calculated from seasons 2016/17-2018/19. Obtained Chi-square value of 19.85 corresponds to p-value of less than 0.0005. Furthermore, standardised Pearson's residual of -4.36 with two degrees of freedom indicates the home wins frequency to be significantly lower than the expected value computed from seasons played with spectators. Significantly higher observed frequency than expected frequency of away wins indicates standardised Pearson's residual of 5.93. On average, away teams won exactly one match more than it was expected from a calculation based on seasons 2016/2017-2018/2019 in every round of the league.

Table 4.1: Differences (home − away), England, Premier League

	16/17-18/19		**20/21**		**Welch's t-Test**		
	Mean	SD	Mean	SD	t	df	p-value
points***	0.52	2.60	-0.07	2.65	3.75	637	<0.01
score***	0.36	1.90	0.01	1.89	3.16	654	<0.01
exp. goals**	0.34	1.32	0.15	1.39	2.31	623	0.021
possession	2.32	24.77	2.03	25.92	0.19	627	0.849
shots	2.18	7.36	1.39	8.62	1.60	574	0.111
shots target**	0.96	4.15	0.44	3.90	2.19	686	0.029
corners	1.10	4.73	0.92	4.57	0.64	669	0.523
fouls***	-0.29	4.57	0.54	4.57	-3.04	650	<0.01
yellow cards	-0.14	1.61	-0.03	1.55	-1.14	672	0.254
red cards	-0.01	0.33	-0.02	0.36	0.38	600	0.707

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 4.1 shows the means of differences between home and away teams. The number in 'Mean' column denotes the average difference per one match for particular statistics. A positive value of mean differences in these columns signs the advantage for home team in all variables except fouls, yellow cards and red cards. On the contrary, in these three variables a negative value corresponds to home advantage. In most cases the advantage of home teams expressed by the mean difference dropped. However, in two cases (points and fouls) the advantage was even turned to the side of away teams. This highly significant difference is confirmed by the results of Welch's t-test in the table above this paragraph, where both p-values are lower than 0.01. The drop in points per match statistic is in line with previous Figure 4.1. The change in committing fouls might have two explanations. It might be caused by overall performance improvement of away teams in matches without spectators or the referees themselves might be less affected by the absence home crowd and therefore favour less the home team. On average, the difference between the number of home and away fouls dropped by 0.83 fouls.

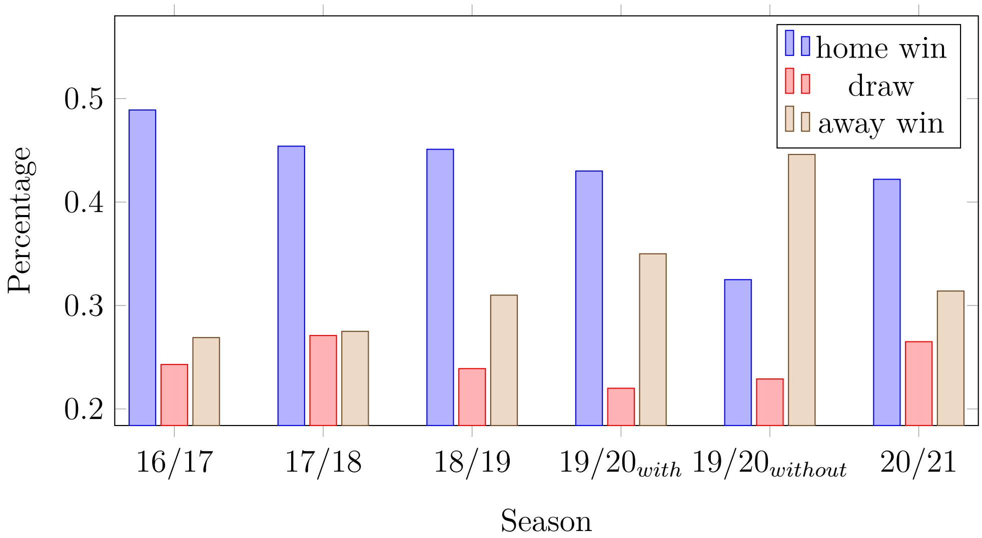
Another significant difference is evident in the cases of scored goals, expected goals and shots on target. All these statistics are related to each other. The number of shots on target directly affects expected goals as every shot in a match is multiplied by calculated probability of scoring a goal and resulting sum creates the number of expected goals. All three mean differences of these statistics are significantly lower in the season played without spectators at 5%

significant level. Moreover, the mean difference of scored goals was almost zero (0.01) in season 2020/2021, where in pre-Covid seasons home teams scored on average approximately one goal more than away teams in every three matches, which was in line with the findings of McCarrick *et al.* (2021). Shots accuracy dropped by 8.5 percentage points on both sides, however goalkeepers' save percentage for home teams dropped by 2.56 percentage points and increased by 1.59 percentage points for away teams. The biggest difference between home and away teams during Covid period can be seen in expected goals/score ratio. While the ratio for away teams remained unchanged, home teams had on average 0.13 expected goals per every scored goal more than in matches with spectators. If the probability behind the statistics of expected goals exactly reflected the reality, then home teams would missed one extra goal per every eight goals they scored in matches without spectators. This might suggest that home teams productivity dropped substantially during season 2020/21.

4.1.2 Germany - Bundesliga

Percentage rate of outcomes for each season for German Bundesliga is shown in Figure 4.2. In comparison with English Premier League, the substantial change in home advantage is noticeable immediately after the start of the COVID-19 pandemic, in the second part of season 2019/20. After the league interruption, 83 out of 306 matches were played until the end of this season.

Figure 4.2: Germany, Bundesliga



However, Chi-square test result in Table A.2 shows no significant change between the expected frequencies of outcomes based on seasons 2016/17-2018/19 and observed outcome frequency from season 2020/21. This could indicate that teams in German Bundesliga has got accustomed to the situation without spectators in new season.

Welch's t-test confirmed significant change in home advantage for three statistics. The difference between home and away teams in expected goals was significantly lower at 10% significance level. In contrast with Premier League, expected goals/score ratio dropped by 4.2 percentage points in matches without spectators, which might suggest home teams to be more productive than was expected in terms of successful shots taken in matches without spectators.

Table 4.2: Differences (home − away), Germany, Bundesliga

	16/17-18/19		20/21		Welch's t-Test		
	Mean	SD	Mean	SD	t	df	p-value
points	0.54	2.54	0.32	2.56	1.28	521	0.200
score	0.42	1.93	0.32	1.92	0.77	524	0.443
exp. goals*	0.37	1.30	0.21	1.31	1.89	520	0.059
possession	1.80	21.12	1.24	23.37	0.38	483	0.706
shots	1.84	6.66	1.12	8.00	1.42	454	0.156
shots target	0.79	3.87	0.52	3.92	1.07	517	0.286
corners	0.72	4.16	0.26	4.35	1.61	504	0.109
fouls*	-0.74	5.26	-0.19	4.72	-1.72	577	0.086
yellow cards**	-0.31	1.57	-0.10	1.53	-2.10	537	0.037
red cards	-0.01	0.37	-0.02	0.32	0.54	608	0.586

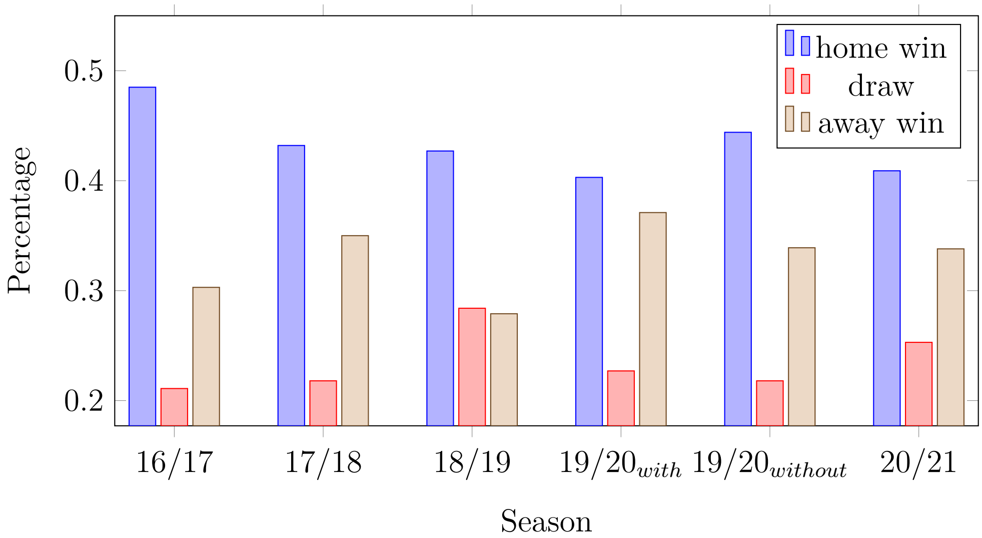
Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

During the period played with spectators, referees awarded approximately one yellow card more to away teams per every three matches played. After the closure of stadiums, this advantage of home teams significantly decreased on the level of one more yellow card for away teams per every ten matches. The mean difference in fouls dropped from -0.74 to -0.19, which corresponds to one foul less committed by away teams per every two matches, on average. Mean differences between home and away teams of all other variables also decreased, although not significantly. For example, home advantage at taken corners dropped from 0.72 to 0.26 corners per match more than away team.

4.1.3 Italy - Serie A

Neither the result of Chi-square test of Italian Serie A shows significant difference in expected and observed values for season 2020/2021. Low value of Pearson's residual (-2.07, df = 2) of home wins in the Table A.2 indicates that the observed value of 155 home wins per season is significantly lower than expected value at 5% significance level.

Figure 4.3: Italy, Serie A



Italian Serie A seems to be the least affected leagues from the all analysed. Looking at the Figure 4.3 or the Table 4.3, the differences between pre-Covid and Covid period do not appear to be as substantial as in the English Premier League or Spanish La Liga. Only two significantly different statistics are shots on target and yellow cards. While the difference in shots on target might not be so important (home teams advantage changed from five more shots on target to three shots per every five matches, approximately), the difference in yellow cards disappeared during matches without fans. In pre-Covid period away teams had disadvantage of one yellow card per every three matches, on average.

Shots accuracy dropped on both sides again (6.79 and 7.15 percentage points), save percentage increased also similarly in both sides (3.47 and 3.74 percentage points). The positive change of expected goals to scored goals ratio on both sides (growth of 2.19 and 4.45 percentage points) caused that both teams had on average more expected goals than scored goals during Covid period. That could sign the lack of productivity during matches without fans, although the differences are small.

Table 4.3: Differences (home − away), Italy, Serie A

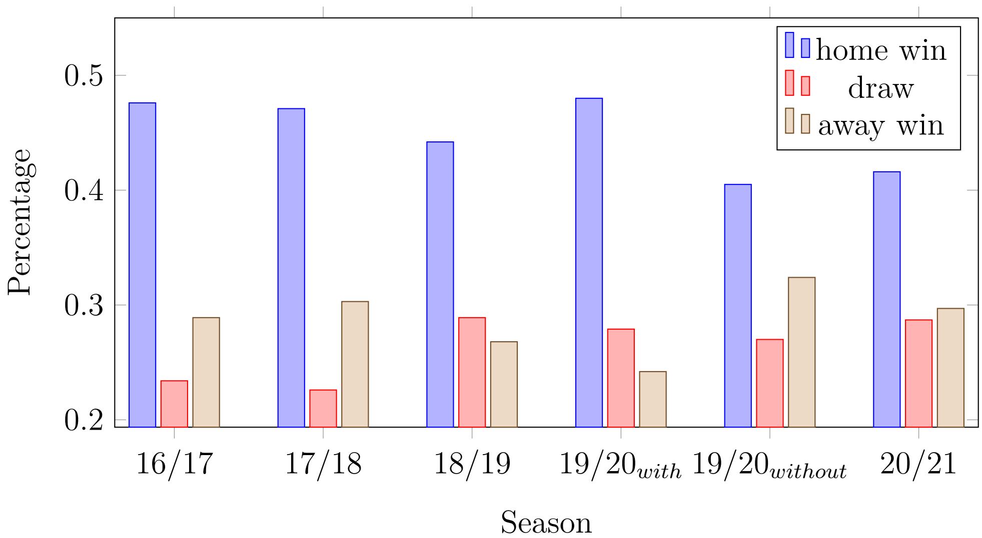
	16/17-18/19		20/21		Welch's t-Test		
	Mean	SD	Mean	SD	t	df	p-value
points	0.42	2.59	0.21	2.59	1.35	647	0.176
score	0.30	1.81	0.20	1.83	0.90	640	0.366
exp. goals	0.29	1.26	0.16	1.37	1.55	606	0.122
possession	1.47	20.15	-0.12	19.59	1.36	666	0.176
shots	2.23	7.60	1.46	8.61	1.56	587	0.119
shots target*	1.04	4.00	0.63	3.69	1.80	696	0.072
corners	0.86	4.71	0.43	4.41	1.61	687	0.108
fouls	-0.14	5.40	-0.17	4.99	0.07	695	0.944
yellow cards***	-0.30	1.66	-0.01	1.68	-2.91	642	0.004
red cards	-0.05	0.48	-0.03	0.41	-0.52	763	0.605

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

4.1.4 Spain - La Liga

Pearson's residuals of home wins (-2.32, df = 2) and draws (2.01, df = 2) both indicates significant differences of observed and expected values at 5% level, however, overall value of Chi-square test is not beyond critical value for rejecting the null hypothesis. Therefore, the difference in frequency of match outcomes between seasons 2016/17-2018/19 and 2020/21 cannot be rejected. Home teams together lost on average one more match than it was expected per every two rounds (20 matches) played in the league.

Figure 4.4: Spain, La Liga



Significant differences of several variables indicates that teams in Spanish La Liga might be affected by empty stadium more than those in German Bundesliga or Italian Serie A. Statistics of expected goals, shots and shots on target are significantly different in season 2016/17-2018/19 and 2020/2021 even at 1% significant level. Moreover, home teams' advantage was diminished by one expected goal per every five matches. Home advantage in shots and shots on target was decreased by approximately 50% in both cases. The difference in shots dropped by almost four shots per every three matches and in shots on target by almost two shots per every three matches. Shots accuracy dropped again at both sides by 7.67 and 8.97 percentage points. Every thirteenth and eleventh shot, respectively, which would head on target headed off target instead. Goalkeepers' save percentage also increased by 2.08 and 1.68 percentage points, respectively. Expected goals to scored goals ratio remained almost unchanged for home teams, however for away teams it dropped by 6.25 percentage points. If the logic behind expected goals works then away teams miss one extra goal per every sixteen goals scored.

Home teams' control of the ball significantly decreased by 2.5 percentage points per every game and home advantage in corners was reduced from three corners per every two matches to almost zero (one corner per every eight matches).

Table 4.4: Differences (home − away), Spain, La Liga

	16/17-18/19		**20/21**		**Welch's t-Test**		
	Mean	SD	Mean	SD	t	df	p-value
points	0.53	2.54	0.36	2.51	1.16	657	0.245
score	0.37	1.80	0.23	1.62	1.40	712	0.161
exp. goals***	0.42	1.22	0.22	1.25	2.74	637	0.006
possession*	3.26	20.33	0.74	24.87	1.78	560	0.075
shots***	2.75	5.92	1.45	7.07	3.23	566	<0.01
shots target***	1.18	3.42	0.57	2.98	3.31	739	<0.01
corners***	1.52	4.18	0.13	4.24	5.55	642	<0.01
fouls	0.09	5.92	0.03	5.38	0.16	709	0.872
yellow cards	-0.28	1.97	-0.13	1.78	-1.41	712	0.159
red cards	-0.02	0.44	0.00	0.44	-0.57	655	0.567

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

4.1.5 All leagues summary

Overall, the differences in home advantage in terms of particular match statistics among all four leagues varies. The most substantial changes in home advantage during season 2020/21 are noticeable in the English Premier league, where significantly lower number of home victories was confirmed by both performed tests. Observed frequencies of outcomes in other countries were not substantially different from the expected outcomes. However, the largest change between home and away victories can be found in German Bundesliga during the end of season 2019/20 (see Figure 4.2). Hence diminished home advantage was not confirmed by results in the following season, it could signify that teams in Bundesliga got accustomed to matches without audience.

The advantage in terms of scored goals completely disappeared in Premier league. Regarding the number of committed fouls, the advantage was even reversed from home to away side. Shots-statistics showed a substantial drop of xg/score ratio, which may suggest the drop in home teams' productivity. In terms of shots on target, the advantage for home teams was significantly lower for all leagues except Bundesliga. In case of Premier league and La Liga, the advantage diminished by more than 50% in comparison with period before the pandemic, in Serie A the advantage declined by almost 40 %. In Bundesliga and Serie A, significantly less number of yellow cards was awarded, therefore the advantage of home teams nearly disappeared.

Chapter 5

Betting Odds Comparison

5.1 Methodology

The goal of this section should be to describe the methods suitable for comparison of the betting odds (adjusted to the actual probabilities according to the description in Subsection 3.4.3), from seasons 2016/17-2018/19 with those from the season 2020/21. Moreover, this section aims to answer the second research question of this thesis, whether the proposed betting odds corresponds to the change in home advantage or whether bookmakers overrated the odds on home teams victory.

5.1.1 Groups by Probability

This method of evaluating the accuracy of betting odds with help of dividing the matches into groups could be found in work of Tiitu (2016) or Kuypers (2000).

All the matches were classified into groups of the same size, depending on the actual probabilities derived from betting odds. Two different grouping were accomplished, with ten and twenty same size groups. Initially, data were sorted in descending order by actual probabilities of home win, draw and away win. Every match was assigned by three group numbers, according to its actual probability of home win, draw and away win. The lower the group number was, the higher was the actual probability of corresponding outcome. For example, the match with actual probabilities of outcomes 65,5%, 22,5% and 13% (on home win, draw and away win, respectively) was sorted into the third highest "home win group", the fifteenth highest "draw group" and eighteenth highest "away win group" from all twenty groups. Afterwards, the hypothesis of equality of the mean of actual probabilities and the mean of corresponding actual outcomes within each group was tested.

The probabilities could be referred to as the subjective probability and objective probability, according to Tiitu (2016). He calculated the mean of subjective probabilities (derived from betting odds) as the average of probabilities derived from the odds, and the mean of objective probabilities (actual outcomes) of each group as a percentage rate of winning bets in the group. Therefore, the average subjective probability and the estimator for the objective probability is formally defined as

$$\overline{\rho}_h = \frac{\sum_{i=1}^{n} \rho_{hi}}{n_h} \text{ and} \tag{5.1}$$

$$\overline{\rho}_h = \frac{\sum_{i=1}^{n} \rho_{hi}}{n_h} \ , \tag{5.2}$$

where $\overline{\rho}_h$ corresponds to the mean subjective probability of the group h, ρ_{hi} corresponds to the subjective probability of the match i in the group h, n_h corresponds to the total number of matches in the group h, $\overline{\pi}_h$ corresponds to the mean objective probability of the group h, and Y_{hi} corresponds to the outcome of the bet i in the group h ($Y_{hi} = 1$ for the winning bet and $Y_{hi} = 0$ for the losing bet).

The fact that the matches could be perceived as independent binomial trials is used for the testing of the hypothesis of equality of means. Hence Y_{hi} follows a binomial distribution, $E(Y_{hi}) = \overline{\pi}_h$ and $Var(Y_{hi}) = \frac{(\overline{\pi}_h)(1-\overline{\pi}_h)}{n_h}$. The size of sample size is sufficient for using the Central Limit Theorem, therefore the binomial distribution could be approximated with the normal distribution, and the test statistics of

$$z_h = \frac{\overline{\rho}_h - \overline{\pi}_h}{\sqrt{\frac{\overline{\pi}_h(1-\overline{\pi}_h)}{n_h}}} \sim N(0,1) \tag{5.3}$$

could be used to test the null hypothesis of $H_0 : \overline{\rho}_h - \overline{\pi}_h = 0$, for all $h = 1, ..., k$, where k is the total number of groups.

5.1.2 Brier Score

One possible approach for comparison of betting odds is Brier score. This method was first used for evaluating the accuracy of weather forecasts (Brier

1950). The formula of Brier score for match i could be expressed as

$$B_i = \sum_{k=1}^{N} (\rho_{ik} - o_{ik})^2, \tag{5.4}$$

where N is the number of possible outcomes (three in this case: home win, draw and away win), ρ_{ik} is subjective probability derived from betting odds of k^{th} outcome of the match i, and o_{ik} is the result of corresponding outcome ($o_{ik} = 1$ if the outcome k occurs in the match i, $o_{ik} = 0$ otherwise). To obtain the overall Brier score for entire season, the average score is computed from the scores of all individual matches.

Using the odds from the example in Subsection 3.4.3 (betting odds: 1.16 - 8.00 - 16.00, corresponding adjusted probabilities: 0.862 - 0.125 - 0.063, match ended in home win) it could be shown how the score is computed: $(0.862 - 1)^2 + (0.125 - 0)^2 + (0.063 - 0)^2 = 0.039$. However, if this match ended in a draw the resulting Brier score would be 1,513, and if the away team won the Brier score would be 1,637.

From the definition, the minimum value of Brier score for three outcomes could occur in case of probabilities "1.000 - 0.000 - 0.000" and home win: $(1.000-1)^2+(0.000-0)^2+(0.000-0)^2 = 0$, and the maximum values could occur in case of probabilities "1.000 - 0.000 - 0.000" and draw: $(1.000-0)^2+(0.000-1)^2 + (0.000 - 0)^2 = 2$. If the odds were proposed uniformly ("0.333 - 0.333 - 0.333"), the Brier score would be $(0.333-1)^2+(0.333-0)^2+(0.333-0)^2 = 0.666$, regardless the outcome.

5.2 Results and discussion

5.2.1 Brier Score

The first part of betting odds analysis aims to evaluate the overall accuracy of proposed odds with Brier score method. Both periods should be compared regardless the direction of accuracy bias, hence Brier score counts for absolute values of bias. The rows of Table 5.1 show the values computed for all leagues together and for each one separately. The columns differ between the accuracy of all matches together and those, which ended in victory of home teams, draws and victory of away teams, respectively.

Table 5.1: Brier Score, Dividing by Leagues

		All	**Home win**	**Draw**	**Away win**
All	16/19	0.558	0.379	0.892	0.567
	20/21	0.578	0.410	0.878	0.552
	diff.	-0.020	-0.031	0.014	0.015
England	16/19	0.540	0.383	0.893	0.526
	20/21	0.597	0.416	0.897	0.604
	diff.	-0.057	-0.033	-0.003	-0.078
Germany	16/19	0.591	0.416	0.904	0.601
	20/21	0.586	0.443	0.900	0.513
	diff.	0.005	-0.026	0.004	0.088
Italy	16/19	0.570	0.371	0.889	0.611
	20/21	0.581	0.415	0.829	0.573
	diff.	-0.011	-0.044	0.060	0.038
Spain	16/19	0.537	0.351	0.884	0.542
	20/21	0.549	0.373	0.898	0.501
	diff.	-0.013	-0.022	-0.015	0.040

Notes: 16/19 is an abbreviation for 2016/2017-2018/2019

The scores of all matches together indicates that bookmakers' prediction have worsened, and, looking at single leagues, it holds for all leagues except German Bundesliga. The results for particular outcomes are more diversified. The most unpredictable outcome seems to be a draw, however a draw is usually the second most probable outcome for each match in which one team is the favorite, and therefore every "surprising" draw is valued by worse Brier score. This might be explanation of the fact that the score assigned to matches ended in a draw is higher than if the betting odds were chosen uniformly for all outcomes (33% per each).

Whereas the differences in accuracy of draws and away wins varies across all four leagues, and the accuracy of away wins have even improved in all leagues except for the English Premier league, the accuracy of home wins appears to be consistently worse across all leagues during season 2020/2021. This finding could signify that prediction of home wins was more difficult for bookmakers during matches played without fans, and therefore less accurate.

Table 5.2: Brier Score, Dividing Into 20 Groups by Probability

	Home			**Draw**			**Away**		
h	16/19	20/21	diff.	16/19	20/21	diff.	16/19	20/21	diff.
1	0,27	0,40	-0,13	0,66	0,67	-0,01	0,38	0,48	-0,09
2	0,39	0,41	-0,02	0,66	0,64	0,01	0,50	0,45	0,05
3	0,38	0,51	-0,13	0,65	0,64	0,01	0,61	0,51	0,10
4	0,52	0,49	0,03	0,65	0,65	0,01	0,61	0,62	0,00
5	0,52	0,56	-0,04	0,63	0,65	-0,02	0,65	0,62	0,03
6	0,54	0,55	-0,01	0,64	0,67	-0,03	0,67	0,65	0,01
7	0,60	0,64	-0,04	0,62	0,67	-0,05	0,67	0,68	-0,01
8	0,60	0,64	-0,03	0,64	0,64	0,00	0,66	0,67	-0,01
9	0,60	0,69	-0,09	0,63	0,65	-0,02	0,66	0,67	-0,01
10	0,65	0,65	0,00	0,62	0,64	-0,02	0,66	0,66	0,00
11	0,65	0,68	-0,03	0,61	0,56	0,05	0,64	0,66	-0,02
12	0,66	0,66	-0,01	0,57	0,61	-0,05	0,62	0,67	-0,05
13	0,67	0,67	0,00	0,58	0,55	0,03	0,59	0,66	-0,06
14	0,67	0,67	-0,01	0,56	0,48	0,08	0,59	0,61	-0,02
15	0,67	0,65	0,02	0,55	0,52	0,03	0,56	0,58	-0,02
16	0,64	0,64	0,01	0,45	0,49	-0,04	0,52	0,51	0,01
17	0,63	0,57	0,06	0,45	0,53	-0,08	0,53	0,58	-0,06
18	0,59	0,55	0,04	0,36	0,36	0,00	0,38	0,48	-0,10
19	0,50	0,50	0,00	0,36	0,55	-0,19	0,38	0,38	0,00
20	0,40	0,42	-0,03	0,27	0,38	-0,11	0,27	0,42	-0,15

Notes: The differences had been computed before all values were rounded to two decimal places. All 4313 matches from seasons 2016/2017-2018/2019 were divided into the groups so that each group contains 215 or 216 matches and all 1444 matches from season 2020/2021 were divided into the groups so that each group contains 72 or 73 matches. The lower is the group number h the higher is the probability.

5.2.2 Grouping by Probability

The goal of dividing into groups was to find whether the betting odds predicted well the real outcome of the match within each group. Afterwards, groups were also compared to find differences in various levels of odds.

The results in Table 5.3 and **??** show that not many means of subjective and objective probability across all the groups are significantly different. Although the number of matches in each group should be sufficient for approximation with the normal distribution, more degrees of freedom would help to reject the null hypothesis of equal means in more groups. Tiitu (2016) rejected the null hypothesis of equal means in approximately fifty percent of groups regardless the overall number of groups. While the matches were divided into twenty groups, thirty two out of sixty groups (together for home win, draw and away win) were significantly different in means. However, almost 100 000 matches from seasons 2009/10 - 2013/14 were included in his analysis. Each group of Tiitu's work contained almost two thousand matches, whereas this analysis included 215 or 216 matches in each group before COVID-19 pandemic and 72 or 73 matches during COVID-19 pandemic. As the defined test statistics depends only on the averages of subjective and objective probability and the number of matches in the group, if the averages remained the same with higher number of matches in each group, the absolute values of z-scores would grow. With approximately five times more matches, this analysis would also reach the similar number of groups with statistically significant mean differences, if the differences in probability averages stayed on the same level.

Table 5.3: $\overline{\rho}_h$ vs. $\overline{\pi}_h$, 20 groups, 2016/2017-2018/2019

	Home			**Draw**			**Away**		
h	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score
1	0.82	0.85	-1.36	0.31	0.35	-1.25	0.73	0.77	-1.20
2	0.76	0.76	0.23	0.30	0.32	-0.49	0.62	0.67	-1.28
3	**0.71**	**0.77**	**-1.83***	0.30	0.29	0.22	0.55	0.53	0.52
4	0.66	0.64	0.51	0.29	0.31	-0.62	0.48	0.53	-1.40
5	0.61	0.64	-0.86	0.29	0.25	1.09	0.42	0.45	-0.73
6	0.57	0.62	-1.37	0.28	0.27	0.33	0.38	0.38	0.25
7	0.54	0.54	-0.16	0.28	0.25	1.15	0.35	0.30	1.57
8	0.50	0.54	-1.12	0.27	0.25	0.67	0.33	0.32	0.23
9	**0.47**	**0.55**	**-2.25****	0.27	0.28	-0.25	0.31	0.29	0.46
10	0.45	0.44	0.36	**0.27**	**0.21**	**2.06****	0.28	0.28	0.08
11	0.43	0.45	-0.67	0.26	0.26	-0.17	0.26	0.22	1.49
12	0.40	0.41	-0.27	0.25	0.23	0.87	0.25	0.24	0.33
13	0.38	0.37	0.50	0.24	0.29	-1.43	0.22	0.19	1.06
14	0.36	0.33	0.78	0.23	0.25	-0.56	**0.20**	**0.15**	**1.99****
15	0.33	0.33	0.02	0.22	0.22	0.01	0.18	0.15	1.14
16	0.29	0.28	0.50	**0.21**	**0.16**	**1.80***	0.16	0.14	0.77
17	0.25	0.30	-1.49	0.19	0.19	-0.22	0.13	0.15	-0.65
18	0.20	0.22	-0.75	0.16	0.14	0.85	0.10	0.08	1.33
19	0.15	0.16	-0.41	0.14	0.14	-0.25	**0.09**	**0.06**	**2.08****
20	0.10	0.08	1.13	0.10	0.12	-0.65	**0.07**	**0.04**	**2.35****

Notes: All 4313 matches were divided into the groups so that each group contains 215 or 216 matches, the lower is the group number h the higher is the probability. $^{***}p < 0.01;^{**} p < 0.05;^{*}$ p < 0.10

Negative z-score signifies higher objective probability than subjective probability proposed by bookmakers (originally proposed decimal betting odds are higher than they should be according to real probability within a group), which implies that bookmakers overrated these odds and systematic betting on correspond outcome within these groups might be profitable. Reversely, higher subjective probability than objective probability could mean that the actual outcomes occur less than it was expected by bookmakers and proposed betting odds were underrated.

Almost all groups of Tiitu's analysis with average probability of home win higher than 0.55 were significantly different in probability means, the same hold for away win with probability higher than 0.44, draws with probability lower than 0.21 and away wins with probability lower than 0.12. In groups with high probability (low number h), subjective probability was systematically lower than objective probability and vice versa in groups with low probability (high number h).

In contrast with the results mentioned above, which revealed significant differences in average probabilities on both ends of odds spectrum for all outcomes, Table 5.3 and Table 5.4 shows such differences only for two groups of away win with probability lower than 0.1 during season before pandemic and for two groups of away win with probability higher than 0.55. The rest units of groups with significantly different means are spread across different levels of probability.

The groups with ninth highest probability for home win during seasons 2016/17-2018/19 and 2020/21 could be the prime example, which could have supported the hypothesis of overrated of betting odds on home win during matches played without fans. Before pandemic, the difference between the objective and the subjective probability within ninth group was significantly different, with the objective probability being much higher than the subjective probability derived from proposed betting odds. On the contrary, during the pandemic the subjective probability within the ninth group was significantly lower than objective probability. If this happened systematically in more groups, it would support the hypothesis of overrated betting odds. However, none of other groups had significantly different means of the subjective and the objective probability, neither any group with significantly different means changed the sign of its z-score between both periods. Overall, the comparison of betting odds divided into groups might not considerably help to specify whether empty stadiums affected bookmaker's calculation of betting odds.

Table 5.4: $\overline{\rho}_h$ vs. $\overline{\pi}_h$, 20 groups, 2020/2021

	Home			**Draw**			**Away**		
h	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score
1	0.80	0.76	0.72	0.32	0.35	-0.41	0.73	0.68	0.82
2	0.72	0.75	-0.58	0.31	0.35	-0.73	**0.63**	**0.72**	**-1.75***
3	0.66	0.65	0.17	0.30	0.29	0.10	**0.57**	**0.67**	**-1.65***
4	0.62	0.68	-1.12	0.29	0.32	-0.50	0.51	0.51	-0.05
5	0.57	0.60	-0.41	0.29	0.28	0.17	0.46	0.51	-0.85
6	0.54	0.61	-1.25	**0.28**	**0.39**	**-1.87***	0.43	0.46	-0.55
7	0.51	0.47	0.57	0.28	0.28	-0.02	**0.40**	**0.31**	**1.71***
8	0.47	0.47	-0.04	0.27	0.31	-0.64	0.37	0.40	-0.55
9	**0.44**	**0.30**	**2.64*****	0.27	0.36	-1.60	0.34	0.32	0.50
10	0.41	0.44	-0.42	0.26	0.19	1.48	0.32	0.29	0.61
11	0.39	0.34	0.88	0.25	0.23	0.45	0.30	0.29	0.19
12	0.37	0.41	-0.79	0.25	0.21	0.90	0.28	0.36	-1.44
13	0.34	0.32	0.31	0.24	0.28	-0.70	0.25	0.24	0.31
14	0.32	0.25	1.28	0.23	0.22	0.22	0.23	0.17	1.43
15	0.29	0.28	0.26	0.22	0.19	0.64	**0.21**	**0.14**	**1.69***
16	0.26	0.22	0.83	0.21	0.18	0.70	0.18	0.15	0.70
17	0.23	0.21	0.46	0.20	0.22	-0.46	0.16	0.19	-0.82
18	0.19	0.14	1.22	0.18	0.13	1.45	0.13	0.14	-0.15
19	0.15	0.14	0.35	0.16	0.22	-1.30	0.11	0.13	-0.47
20	0.11	0.07	1.24	0.13	0.13	0.00	0.07	0.11	-1.09

Notes: All 1444 matches were divided into the groups so that each group contains 72 or 73 matches, the lower is the group number h the higher is the probability. $^{***}p < 0.01;^{**} p < 0.05;^{*} p < 0.10$

Chapter 6

Limitations

The absence of crowd in the stadium was not the only change during the season 2020/2021, however it was probably the only significant change influencing much more the home teams than the away teams during the match.

Another significant factor influencing negatively the performance could be absences of players from the starting lineup due to mandatory quarantine. It could affect both teams, however probably not equally at both sides. Unfortunately, this factor could not be included into the analysis due to lack of data about missing players.

All the other changes influenced either both teams equally or might not considerably change the outcome of the match. One of the changes was a the possibility to substitute more than three players in one match. This new rule was established due to fewer days off between the matches in the final part of the season 2019/2020 and has remained valid in the most of the European leagues except for the English Premier league also during entire season 2020/2021. Five instead of three players can be substituted in one team during the match, however teams have only three possibilities to interrupt the game. Therefore more substitutions have to be realised either at once or during half-time.

Chapter 7

Conclusion

This thesis aims to investigate the impact of COVID-19 pandemic restrictions on home advantage in selected European football leagues. Three complete seasons of the English Premier league, German Bundesliga, Italian Serie A and Spanish La Liga finished before the pandemic were compared with season 2020/21 played during the pandemic. Attendance in all matches of 2020/21 was restricted and almost 95 percent of matches were played completely without spectators.

The analysis of collected match statistics showed that differences in home advantage varies among all four leagues. The most substantial changes across fundamental match statistics can be found in the English Premier league, where the Welch's t-test confirmed significantly lower number of points, scored goals, expected goals, shots on target and fouls. Home advantage in terms of scored goals completely disappeared in Premier league during the season 2020/21. In terms of gained points or committed fouls home advantage was even reversed into away team advantage. A substantial drop in expected goals/shots ratio may sign a decrease in home teams' productivity.

An extraordinary decrease of home wins during the final part of the season 2019/20 in German Bundesliga was not confirmed in the season 2020/21, hence home teams might become accustomed to new conditions, as suggested also by Reade *et al.* (2020). In terms of shots on target, the advantage for home teams was significantly lower for all leagues except Bundesliga, and significantly lower number of yellow cards was awarded in Bundesliga and Serie A. Preceding findings are supported also by a recent work of McCarrick *et al.* (2021), who also found that empty stadiums had negative effect on goals per game of home teams.

The final part of this thesis aimed to investigate the accuracy of betting odds and whether the betting odds proposed on victory of home teams showed signs of overestimation. Brier score showed overall decline of betting odds accuracy for all four analysed leagues and the matches together in the COVID-19 period compared to pre-COVID-19 period. Although the accuracy of single leagues varies, the accuracy computed only for the matches ended in home victory declined for all of them. However, the comparison of prediction accuracy within the groups divided by the level of probability did not reveal any signs of systematically overrated betting odds proposed on home teams' victory.

One of the main contributions of this thesis lies in its topicality. Considering that restrictions due to pandemic of COVID-19 persist also in current season 2021/2022 and stadiums could be closed for spectators once again, this work might be extended by utilisation of more data related to home advantage during restricted matches. The extension of the analysis by data concerning the absences of significant players could show another incidence of the pandemic. Possible future research could also aim at returning of home advantage and the time necessary to recover from the matches played without spectators.

Bibliography

Brier, G. W. (1950): "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* **78**: pp. 1–3.

Clarke, S. R. & J. M. Norman (1995): "Home Ground Advantage of Individual Clubs in English Soccer." *Journal of the Royal Statistical Society. Series D (The Statistician)* **44(4)**: pp. 509–521.

Colombo, A. & K. Batardière (2018): "European Leagues Fan Attendance Report." *Technical report*, Association of European Professionall Football Leagues, Nyon.

Courneya, K. S. & A. V. Carron (1992): "The Home Advantage in Sport Competitions: A Literature Review." *Journal of Sport & Exercise Psychology* **14(1)**: pp. 13–27.

Curley, J. (2016): "engsoccerdata: English Soccer Data 1871-2016. R package version 0.1.5." .

Dawson, P. & S. Dobson (2010): "The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees." *Journal of Economic Psychology* **31(2)**: pp. 181–191.

Fischer, K. & J. Haucap (2020): "Betting market efficiency in the presence of unfamiliar shocks: The case of ghost games during the COVID-19 pandemic." *Technical Report 349*, Düsseldorf.

García, J., L. Pérez, & P. Rodríguez (2017): "Forecasting football match results: Are the many smarter than the few?" In "The Economics of Sports Betting," pp. 71–91.

Garicano, L. & I. Palacios-Huerta (2005): "Sabotage in Tournaments: Making the Beautiful Game a Bit Less Beautiful." *Technical Report 5231*.

GOUMAS, C. (2013): "Home advantage and crowd size in soccer: a worldwide study." *Journal of Sport Behavior* **36(4)**: pp. 387–399.

KUYPERS, T. (2000): "Information and efficiency: an empirical study of a fixed odds betting market." *Applied economics* **32(11)**: pp. 1353–1363.

MCCARRICK, D., M. BILALIC, N. NEAVE, & S. WOLFSON (2021): "Home advantage during the COVID-19 pandemic: Analyses of European football leagues." *Psychology of sport and exercise* **56**: p. 102013.

NEVILL, A. M. & R. L. HOLDER (1999): "Home Advantage in Sport." *Sports Medicine* **28(4)**: pp. 221–236.

PEETERS, T. & J. C. VAN OURS (2021): "Seasonal Home Advantage in English Professional Football; 1974-2018." *De Economist* **169(1)**: pp. 107–126.

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

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

POLLARD, R. (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* **29**: pp. 169–189.

POLLARD, R., C. DA SILVA, & N. C. MEDEIROS (2008): "Home advantage in football in Brazil: Differences between teams and the effects of distance traveled." *Brazilian Journal of Soccer Science* **1**: pp. 3–10.

READE, J. J., D. SCHREYER, & C. SINGLETON (2020): *Echoes: what happens when football is played behind closed doors?*

SCHWARTZ, B. & S. BARSKY (1977): "The Home Advantage." *Social Forces* **55**: pp. 641–661.

SCOPPA, V. (2021): "Social pressure in the stadiums: Do agents change behavior without crowd support?" *Journal of economic psychology* **82**.

SORS, F., M. GRASSI, T. AGOSTINI, & M. MURGIA (2020): "The sound of silence in association football: Home advantage and referee bias decrease in matches played without spectators." *European journal of sport science* pp. 1–9.

TIITU, T. (2016): *Abnormal returns in an efficient market? Statistical and economic weak form efficiency of online sports betting in European soccer*. Ph.D. thesis, Aalto University. School of Business, Aalto.

WILCOX, R. R. (2003): "Applying Contemporary Statistical Techniques." In R. R. WILCOX (editor), "Applying Contemporary Statistical Techniques," pp. 237–284. Burlington: Academic Press.

Appendix A

Match Statistics Comparison

Table A.1: Shots statistics

	shots accuracy		**save percentage**		**xg/score**	
	home	**away**	**home**	**away**	**home**	**away**
Prem. League						
16/17-18/19	48.36%	49.31%	27.30%	25.10%	99.29%	101.07%
20/21	39.86%	40.80%	24.74%	26.69%	112.14%	101.73%
difference	**-8.50%**	**-8.51%**	**-2.56%**	**1.59%**	**12.86%**	**0.66%**
Bundesliga						
16/17-18/19	48.11%	48.89%	28.38%	24.65%	97.66%	100.51%
20/21	40.85%	40.35%	30.32%	27.07%	93.45%	100.15%
difference	**-7.26%**	**-8.54%**	**1.94%**	**2.42%**	**-4.21%**	**-0.36%**
Serie A						
16/17-18/19	46.43%	46.35%	25.51%	24.89%	98.08%	98.38%
20/21	39.64%	39.20%	28.97%	28.63%	100.27%	102.83%
difference	**-6.79%**	**-7.15%**	**3.47%**	**3.74%**	**2.19%**	**4.45%**
La Liga						
16/17-18/19	47.49%	48.83%	26.51%	25.35%	101.42%	97.23%
20/21	39.81%	39.86%	28.59%	27.03%	102.11%	103.47%
difference	**-7.67%**	**-8.97%**	**2.08%**	**1.68%**	**0.69%**	**6.25%**

Table A.2: Comparison of Matches Outcomes

	Home wins	Draws	Away wins	Total
Premier League				
16/17 - 18/19				
Frequency	541	254	345	1140
Probability (%)	47.45	22.29	30.26	100.00
20/21				
Observed	144	83	153	380
Expected	180.31	84.70	114.99	380.00
Pearson's res.	**-4.36***	**-0.232**	**5.93***	χ^2 = **19.85***
Bundesliga				
16/17 - 18/19				
Frequency	426	230	261	917
Probability (%)	46.46	25.08	28.46	100.00
20/21				
Observed	129	81	96	306
Expected	142.15	76.75	87.10	306.00
Pearson's res.	**-1.91**	**0.66**	**1.39**	χ^2 = **2.37**
Serie A				
16/17 - 18/19				
Frequency	514	271	354	1139
Probability (%)	45.13	23.79	31.08	100.00
20/21				
Observed	155	96	128	379
Expected	171.03	90.18	117.79	379.00
Pearson's res.	**-2.07***	**0.82**	**1.42**	χ^2 = **2.77**
La Liga				
16/17 - 18/19				
Frequency	528	285	327	1140
Probability (%)	46.32	25.00	28.68	100.00
20/21				
Observed	158	109	113	380
Expected	176.00	95.00	119.00	380.00
Pearson's res.	**-2.32***	**2.01***	**0.55**	χ^2 = **4.06**

Notes: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Appendix B

Betting Odds Comparison

Table B.1: $\overline{\rho}_h$ vs. $\overline{\pi}_h$, 10 groups, 2016/2017-2018/2019

	Home			**Draw**			**Away**		
h	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score
1	0.79	0.81	-0.71	0.31	0.33	-1.21	**0.68**	**0.72**	**-1.79***
2	0.68	0.70	-0.83	0.29	0.30	-0.39	0.52	0.53	-0.54
3	0.59	0.63	-1.55	0.28	0.26	1.07	0.40	0.41	-0.40
4	0.52	0.54	-0.90	0.28	0.25	1.28	0.34	0.31	1.26
5	0.46	0.50	-1.33	0.27	0.24	1.19	0.30	0.29	0.38
6	0.41	0.43	-0.67	0.26	0.25	0.48	0.25	0.23	1.27
7	0.37	0.35	0.91	0.24	0.27	-1.42	**0.21**	**0.17**	**2.12****
8	0.31	0.30	0.33	0.21	0.19	1.17	0.17	0.15	1.34
9	0.22	0.26	-1.60	0.18	0.17	0.40	0.12	0.11	0.29
10	0.13	0.12	0.36	0.12	0.13	-0.60	**0.08**	**0.05**	**3.12*****

Notes: All 4314 matches were divided into the groups so that each group contains 431 or 432 matches, the lower is the group number h the higher is the probability. $^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.10$

Table B.2: $\overline{\rho}_h$ vs. $\overline{\pi}_h$, 10 groups, 2020/2021

	Home			**Draw**			**Away**		
h	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score	$\overline{\rho}_h$	$\overline{\pi}_h$	z-score
1	0.76	0.76	0.09	0.32	0.35	-0.81	**0.68**	**0.70**	**-0.62**
2	0.64	0.67	-0.66	0.29	0.31	-0.28	0.54	0.59	-1.15
3	0.56	0.60	-1.17	0.28	0.33	-1.25	0.45	0.49	-0.99
4	0.49	0.47	0.45	0.27	0.30	-0.60	0.38	0.35	0.83
5	0.43	0.37	1.39	0.26	0.27	-0.16	0.33	0.30	0.72
6	0.38	0.38	-0.02	0.25	0.22	0.90	0.29	0.32	-0.96
7	0.33	0.28	1.16	0.24	0.25	-0.32	**0.24**	**0.20**	**1.23**
8	0.28	0.25	0.75	0.22	0.19	0.95	**0.20**	**0.15**	**1.67***
9	0.21	0.17	1.13	0.19	0.17	0.54	0.14	0.17	-0.72
10	0.13	0.10	1.01	0.14	0.17	-1.01	**0.09**	**0.12**	**-1.10**

Notes: All 1444 matches were divided into the groups so that each group contains 144 or 145 matches, the lower is the group number h the higher is the probability. $^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.10$