Do Fans Impact Sports Outcomes? A COVID-19 Natural Experiment *

Jeffrey Cross†
Richard Uhrig‡

November 10, 2020

Abstract

This paper studies the effect of fan attendance on home field advantage in top European soccer leagues. We exploit exogenous variation in the level of fan attendance driven by COVID-19 mitigation policies and find that the home field advantage, as measured by home goals minus away goals, is reduced by 57% across the English Premier League, German Bundesliga, Italian Serie A, and Spanish La Liga. We find that this leads to a decrease in probability for a home win, indicating that these goals are pivotal with respect to match outcomes.

Keywords: Home field advantage, soccer, COVID-19, fan attendance
JEL Codes: Z2

*The authors would like to thank Kelly Bedard, Antoine Deeb, Matthew Fitzgerald, Molly Schwarz, Ryan Sherrard, and Gonzalo Vazquez-Bare for their insightful discussion and comments on previous versions of the manuscript.

†University of California, Santa Barbara (email: jeffreycross@ucsb.edu)
‡University of California, Santa Barbara (email: rauhrig@ucsb.edu)

1 Introduction

Home teams generally outperform away teams in a wide variety of sports. This stylized fact is well-established and known as home field advantage. Over the course of the 2008-2009 through 2018-2019 season in four top European leagues, home teams outperformed away teams in goal difference by 0.38 and win percentage by 17 percentage points.[1] This advantage is driven by various factors that are typically broken down into three main channels: travel fatigue, venue familiarity,[2] and crowd support. It is an empirical challenge to separately identify each effect. While fan attendance is indeed measurable, there are endogeneity concerns regarding the impact of crowd attendance on team performance; better teams typically have bigger stadia and higher attendances, so the direction of causality is not necessarily clear ex ante.

In this paper, we exploit exogenous variation in the number of fans, driven by the COVID-19 pandemic. Specifically, we leverage the fact that four of the top European soccer leagues implemented no-fans policies for the last quarter or so of the 2019-2020 season. The imposition of this no-fans policy provides a large decrease in fans that is plausibly exogenous to the outcome of soccer games. We find that these no-fans policy led to a 57% decrease in home field advantage relative to games in the same matchweek and league from previous seasons, as measured by goal difference, indicating that fans play a significant role in home field advantage in European soccer. These results are robust to inclusion of various fixed effects as well as controls for weather and local exposure to COVID-19.

We explore both the causes and implications of this change in home field advantage. Using the expected goals metric, we show that the decrease in home field advantage is driven by home teams actually playing worse relative to away teams after the implementation of the

[1] Home teams win approximately 47% of the time, draw 25% of the time, and lose 29% of the time

[2] In addition, there may be some advantage regarding field preparation and style of play. In particular, FC Barcelona was known in the mid 2010's to keep their grass especially short and somewhat wet, such that short passes were easier to complete. Teams hosting FC Barcelona would sometimes prepare their grass to be long and dry in order to hamper the visitor's short passing style. https://www.dailymail.co.uk/sport/football/article-2600454/Barcelona-complain-UEFA-length-grass-Vicente-Calderon-ahead-Champions-League-showdown-Atletico.html. However, evidence for this type of advantage remains anecdotal to this point and is unidentified relative to the venue familiarity effect. Therefore, if one prefers to think of this more generally, then the venue familiarity effect can also include such factors as pitch preparation, the colors in the locker room, and the exact jerseys worn by each team.

no-fans policy, as opposed to scoring relatively fewer goals on the same quality and quantity of goal-scoring opportunities. Our results show that the change in home field advantage manifests in fewer wins for the home team. Specifically, home teams without fans win games 5.4 percentage points less often relative to the control group. Interestingly, we find that there is no effect on the probability of a draw between the home team and the away team as decreases in wins led to a direct increase in losses for the home team. This finding is very much in line with the idea that fans have a linear impact on the outcome of games - if we assume that games are either shifted from home win to draw, home win to home loss, and draw to home loss, then approximately the same number of games are shifted from home win to draw as are shifted from draw to home loss.

Of all the potential sports leagues with which to examine the impact of fan attendance on home field advantage, the European soccer leagues discussed here provide the best natural experiment. The rules of the game were not materially affected in any way that might plausibly impact home field advantage. While games were delayed between two and three months, no games were cancelled or relocated. This is in stark contrast to both other soccer leagues (Major League Soccer in the United States, and European club tournaments such as the UEFA Champions League and the UEFA Europa League), international soccer matches, and North American leagues in other sports (National Basketball Association, Major League Baseball, National Football League, etc.).

This paper contributes to a rich literature focused on identifying the effect of crowd support, travel fatigue, and venue familiarity on the home field advantage. Broadly speaking, the literature on crowd support can be split into two strands. The first strand uses variation in turnout by country and league and generally finds larger advantages for the home team associated with higher fan attendance (Pollard and Pollard, 2005; Clarke and Norman, 1995; Agnew and Carron, 1994). To overcome endogeneity concerns, two recent papers have leveraged quasi-experimental variation in fan attendance with mixed results. Ponzo and Scoppa (2018) provide convincing evidence for the impact of crowd support by using same-stadium derbies. In a same-stadium derby, both teams playing are familiar with the stadium and do not suffer travel fatigue, but tickets are allocated differently to the home and away teams. Using 20 years of Serie A data and 5 same-stadium derbies, they find that the home team is 15 percentage points more likely to win than the away team, significantly lower than the

25 percentage point advantage in normal matches. Belchior (2020), on the other hand, uses randomized game times in Brazilian football and finds that attendance has no effect on home field advantage. Our results are much more similar to Ponzo and Scoppa (2018) in that we find home field advantage decreases by 57%.

Boyko, Boyko and Boyko (2007) and Ponzo and Scoppa (2018) find that referees play a role in the home field advantage, and that this "bias" is at least partially driven by the fans. In our context, it is unclear whether our estimated effects are due to changes in player performance or referee decision-making. We are unable to decompose our estimated effect due to the simultaneous nature of each component. It may be the case that fans primarily impact the outcome in soccer matches through influencing the referee, leading to yellow cards, red cards, and penalty kicks which then affect the outcome. However, it may also be the case that fans affect home team performance in ways that lead to more dangerous dribbling and scoring opportunities, which in turn affects those same referee actions. Both of these channels would appear identical ex-post to the researcher, so we focus on estimating the total effect rather than exploring the exact mechanism through which fans impact games.

Our results indicate that approximately 43% of the home field advantage can be attributed to non-fan factors of travel fatigue and venue familiarity. Oberhofer, Philippovich and Winner (2010) analyzes the effect of distance on home field advantage in the Bundesliga and find that performance decreases with distance. Similar effects have been found in the English Premier League (Clarke and Norman, 1995), U.S. National Football League (Nichols, 2014), and Australian Football League (Goumas, 2014), which have shown that crossing time zones is important in addition to distance. Barnett and Hilditch (1993) finds that distance can also suppress turnout of fans the away team, in addition to the established direct effects of distance on team performance. Stadium familiarity also likely plays a role in the residual home field advantage as players on the home team are more familiar with their own stadium. This has shown to have a particularly large effect when the field when the field is unusually large or small or has an artificial surface (Barnett and Hilditch, 1993; Clarke and Norman, 1995). To causally identify the effect of familiarity, researchers have used team stadium moves that reduce familiarity for the home team. Evidence on this effect has generally been mixed with some papers finding a drop in home field advantage after moving (Pollard, 2002), while others have found no effect (Loughead et al., 2003). In this paper, we are not able to separately disentangle these two factors, but our work does illustrate that the largest component of the home field advantage is due to fan attendance and not travel fatigue or venue familiarity.

This paper is organized as follows: Section 2 provides background information and describes the data used. Section 3 discusses threats to identification. Section 4 explains the empirical strategy. Section 5 highlights the key results of our analysis. Section 6 concludes.

2 Background and Data

In this section, we first describe the five major European soccer leagues and their response to COVID-19. Next, we describe the data utilized in our analysis and present descriptive statistics for our sample.

2.1 Schedule Construction in European Soccer

Teams in the five soccer leagues discussed in this paper all play what is typically called a "balanced" schedule. Every team plays each other team in the same league exactly twice, once at home and once away. Therefore, each team plays $2 \times (n-1)$ games in a season, where n is the number of teams in that league.[3] Games against the same opponent are spaced throughout the season. In each of these leagues, a certain team will play each potential opponent once in the first half of the season and once in the second half. These two games are often called "reverse fixtures" since they contain the exact same teams but are played at opposite locations. In some leagues, the order of opponents is exactly the same in each half of the season.[4] For example, if Bayern Munich plays at Borussia Dortmund in each team's first game of the season, then Borussia Dortmund will play at Bayern Munich in the eighteenth game of the season.

Games are randomly allocated subject to some restrictions. These restrictions are in place to prevent long strings of home or away games for any given team, and in some cases they are also in place to avoid concurrent home games between two teams in the same metropolitan

[3]There are 20 teams in the English Premier League, Italian Serie A, and Spanish La Liga but only 18 teams in the German Bundesliga and French Ligue 1.

[4]This is certainly the case in the German Bundesliga but may also be true in other leagues.

area. If games were perfectly randomly allocated, then there would occasionally be long strings of home or away games for a given team. This could actually present problems for analysis, as results could be driven by randomness in the schedule rather than any changes in home field advantage. In the last ten matches of each league discussed here, every single team played between four and six home games, highlighting the benefits of non-random schedule construction in terms of balancing the dispersion of home and away games across the league schedule.

2.2 European Soccer and COVID-19

The top five ranked professional football leagues in the Union of European Football Associations (UEFA) are the Spanish La Liga, English Premier League, Italian Serie A, German Bundesliga and French Ligue 1. Over the last 20 years, these leagues have dominated both European soccer and global sports revenues.[5] In 2017-2018, these five leagues generated approximately 17.4 billion dollars (USD) in revenue, which is more than the total GDP of Jamaica and double the total revenue of the National Football League (NFL) over a similar time period, highlighting the importance of these leagues relative to other leagues and the rest of the world.

Seasons for all of the major European soccer leagues begin in August and last until May of the following year. In the 2019-2020 season, however, the normal schedules were interrupted by COVID-19 which was declared a pandemic by the World Health Organization on March 11, 2020. Amidst concerns for player and staff safety, all major European leagues halted play from that date until mid-May at the earliest, with some slight variation by country. [6]

Four of the five leagues resumed play with a restrictive set of protocols. [7] Many of the

[5] Over the last 23 seasons nearly every winner and runner-up in the most prestigious European club competition, the Champions League, has come from these five leagues. The only exception since the 1996-1997 seasons is Porto, who won the Champions league in 2003-2004.

[6] Each of these five major leagues paused their seasons within a few days of each other. The German Bundesliga was both the last league to stop play and the first to resume play. The last game before the hiatus was played on March 11th, and the first game after the restart was played on May 17th. The only league that did not resume was the French Ligue 1.

[7] The exception here is the French Ligue 1, which decided to cancel the rest of the 2019-2020 season rather than resume play after the hiatus.

policies instituted across the leagues were similar during the season,[8] but with some variation in the lead-up.[9] In addition to testing and quarantine policies, each league allowed for two extra substitutions per match and instituted water breaks in order to alleviate the burdens of the extended hiatus and warmer weather. No games were cancelled, but given concerns that the disease could spread during matches, all four of the leagues decided to resume play without any fans at the originally designated stadium for the match.

A brief summary of the timeline, on-field policy changes, and primary method of travel for each league is shown in Table 1. All of the leagues stopped at about the same time, but play resumed at different points of time for each league. The Bundesliga resumed play slightly over 2 months after, while the other leagues waited approximately 3 months before resuming play. In general, the policies instituted in all of the leagues were highly successful with no cancelled or rescheduled matches and very few missed games by players.[10]

2.3 Expected Goals

Due to randomness, human error, and occasional moments of athletic brilliance, the score of a match is a noisy signal for which team actually played better over the course of 90 minutes. Expected goals are a measure of team performance that accounts for both the quantity and quality of each team's chances to score. Expected goals have been shown to better predict future performance and more closely track team actual performance than realized goals (Rathke, 2017). They are calculated by summing the ex ante probabilities that each shot, based on its specific characteristics, is converted into a goal. These probabilities are estimated using historical data on similar chances.[11]

As a robustness check, we abstract from realized goals and instead use expected goals for each team as a measure of performance. We restrict the panel to years for which the

[8]The similarities across all of the leagues is not surprising since the heads of all five of the major European soccer leagues had weekly meetings starting in April.

[9]Spain, for example, instituted a four phase program in the lead-up to the season, which was slightly different than what was instituted in other leagues.

[10]The most publicized case of a player missing a game was Claudio Pizarro. He missed the game after the restart because his daughter tested positive. It is worth noting that there was an outbreak in a second division team in Germany, but no such flare ups happened in any of the top divisions.

[11]Chances are considered similar based on distance, angle, goalkeeper and defender position, and shot type (shot with the foot or the head).

expected goals data are available and estimate the effect of the no-fans policy on home field advantage as measured by expected goal difference to show that observed changes in home field advantage after the implementation of the no-fans policy are not driven entirely by chance conversion. We find that the decrease in the home field advantage is nearly identical when focusing on expected goals instead of goals. This result highlights the fact that our estimates of the decrease in the home field advantage is indicative of a similar change in the quantity and quality of goal-scoring opportunities.

2.4 Data Description

We use match data from the soccer statistics website FBref,[12] which provides information on all matches in the five major European soccer leagues going back at least 10 years. Our panel starts with the 2009-2010 season, giving us 10 years of data prior to the 2019-2020 season and 15,906 matches overall. The data are slightly more limited for the robustness checks with expected goals, which are only available on FBref starting in the 2017-2018 season.

In some specifications, we will control for daily weather conditions at the location of the match. To do this, we use NOAA Global Historical Climate Network daily (GHCN Daily) data (Menne et al. 2012), which includes information on daily precipitation as well as maximum and minimum temperature for weather stations around the globe. A key challenge with utilizing this, however, is that many of the stations do not have observations for every day (Auffhammer et al. 2013).[13] In general, we find that this is a much larger problem for weather stations in all England, Spain, and Italy, but not Germany. For this reason, all results utilizing weather data will focus only on the Bundesliga.[14]

To measure the effect of COVID-19 across space, we use one of the most spatially-resolved global dataset of daily confirmed COVID-19 cases from Carleton et al. (2020).[15] Combin-

[12] FBref.com launched in June 2018 with league coverage for six nations: England, France, Spain, Italy, Germany, and the USA.

[13] Appendix Figure A.1.

[14] An alternative would be to use reanalysis data, such as ERA5, rather than weather station data.

[15] Spatial granularity varies for each country in this analysis. Broken down by the Nomenclature of Territorial Units for Statistics (NUTS): Germany is level 1 (16 states), Spain is level 2 (19 autonomous communities), and Italy is level 3 (107 provinces). For England, the data is at the national level and so we are not able to estimate differential cases by team.

ing the geographic location of each stadium with the geographic regions in the COVID-19 dataset, we calculate three different measures for the effect of COVID-19 on an area as of March 31, 2020: cumulative cases, new cases, and total cases per capita. For each of these measures, we calculate the difference between the home and away team and will use this as a measure of differences in the effect of the virus in the area of the home team compared to the away team.

2.5 Summary Statistics

Table 2 presents key summary statistics aggregated across all four of the "Big Five" European soccer leagues that resumed play following the pause in play as a result of COVID-19. Almost all match data is from the 2009 to 2010 season up to the 2019 to 2020 season. The lone exception is the expected goals metric, which is only available starting in the 2017-2018 season. As a measure of distance traveled, we calculate the linear distance between the home and away team stadiums.

There are two things worth highlighting from Table 2. First, home teams generally score more than away teams, which is reflected in a home minus away goal difference of 0.38. Throughout the paper we will focus on this measure as the primary outcome of interest. Second, there is a similar advantage to home teams in the home minus away expected goals difference. This highlights that the difference in actual goals is reflective of differences in the quality of goal-scoring opportunities. In other words, this table shows the core of what the home field advantage is - a difference in play between home and away teams, which then manifests in match outcomes.

The home field advantage, attendance, and distance between teams is different across each of the four leagues. Table 3 shows that the Bundesliga and Premier League have the highest average attendances and that La Liga and Serie A have teams that are the farthest apart. All of these leagues, have more goals scored, on average, by the home team compared to the away team. Although the exact home field advantage differs by league, they are all in the range of approximately 0.34 to 0.46.

3 Identification

We assume that any changes in home field advantage after the hiatus induced my COVID-19 are driven by the no-fans policies implemented in each of the four European soccer leagues discussed here. However, it is theoretically possible that other confounding factors may drive our results. In this section, we discuss the potential confounds and explain why our assumption is plausible. Where applicable, we test separate channels through which the home field advantage may have changed before and after the hiatus.

The most direct potential threat to identification is obviously that the COVID-19 pandemic affected home and away teams differently through infection. This is unlikely due to the low number of positive tests among soccer players. The players available to participate for the last quarter of the season were largely the same as those that would have been available if not for the hiatus, excepting of course for a small number of opt-outs and injuries. There is also the possibility that different regions had varying levels of exposure to COVID-19, and that these differences could affect the home field advantage orthogonal to the no-fans policy. As a robustness check, we show that neither controls for new cases nor total cases as a fraction of the local population affect our estimated coefficient of interest.

There is also the possibility that the long hiatus of two or more months affected home and away teams differently. Within this channel are two subchannels, those being the long break itself and the differences in weather relative to when these games were originally scheduled. Ekstrand, Spreco and Davison (2019) and Jamil, McErlain-Naylor and Beato (2020) describe the effects of long breaks on injuries and match outcomes. The German Bundesliga has the longest scheduled break each season, with no matches scheduled for approximately four weeks in January. While there is some effect of these breaks on injuries and the outcome of matches, there is no evidence that these breaks affect home and away teams differently. The primary impact of breaks on game performance is a decline in shot-to-goal conversion, but there is contradictory evidence that shot-to-goal conversion actually improved once play resumed from the COVID-19-induced hiatus in these four leagues (Cohen and Robinson, 2020). We also account for these potential differences by replicating our analysis using expected goals rather than realized goals and find similar results, indicating that differences in shot-to-goal conversion do not play a significant role. It is unlikely that differences in weather affected home field advantage during the resumed seasons. There is no evidence that weather affects home field advantage in general for European soccer matches. There were two mitigation strategies undertaken by these four leagues to limit the burden on players; teams were allowed five substitutions per game instead of the usual three, and there were two water breaks per game in addition to the halftime intermission. It is unlikely that these slight changes would affect home field advantage directly, but they could mitigate any differences in home field advantage driven by warmer weather than when the games were originally scheduled to be played. Our results are robust to the inclusion of controls for weather, indicating that the different weather conditions than normal do not affect our estimated coefficients.

The last potential confounding factor is that COVID-19 may have affected one of the other two channels of home field advantage, those being the familiarity and travel effects. In stark contrast to many American sports leagues, games were played in the normal team stadia and were not moved to better mitigate the spread of COVID-19. Thus there was no difference in the familiarity effect before and after the season was suspended. It may be reasonable to suspect that travel is more onerous now relative to before the pandemic; one might expect for away teams to perform worse relative to home teams due to more strenuous travel conditions. Indeed, if away teams are optimizing with respect to travel arrangements prior to the COVID-19 pandemic, then any additional restrictions placed upon those teams to avoid infection and outbreak would presumably negatively affect the performance of away teams. However, we observe that away teams perform better relative to home teams after the leagues resumed, and this potential difference in travel would in fact attenuate our observed effect of no fans towards zero.

In this section, we have examined each potential confound one-by-one. In every case, the proposed confound should have either no effect or a positive effect on home field advantage. Thus, we conclude that our estimated effect is indeed the portion of home field advantage attributable to the presence of fans, and that this estimated effect is if anything attenuated towards zero due to the increase strain of travel during the pandemic.

4 Empirical Framework

We estimate the following equation:

$$y_{imsct} - y_{jmsct} = \alpha + \beta Post_{msct} + \lambda_{sc} + \epsilon_{ijmsct} \tag{1}$$

where y_{imsct} is the goals scored by home team i in matchup m of season s, country c, and week t, y_{jmsct} is the goals scored by the away team j in the same game, α is the home field advantage under normal circumstances, capturing all three potential factors including fans, familiarity, and travel, $Post_{msct}$ is an indicator for matchup m occurring after the implementation of the no-fans policy, λ_{sc} is a country-by-season fixed effect, [16] and ϵ_{ijmsct} is the error term. The coefficient of interest β represents the change in home field advantage once these no-fans policies are implemented. Standard errors are clustered at the matchup-by-season level to allow for correlation in the error term between observations involving the same two teams in the same season. Note that for any matchup between two teams, there are two games played per season, one at each stadium. Thus, the home-away status in one game is perfectly negatively correlated with the home-away status of the same two teams in the reverse fixture. This analysis is replicated using expected goals for each team on the left hand side to see if any differences in home field advantage are driven by differences in creating or finishing goal-scoring opportunities.

Next, to examine how these changes in goal difference manifest in actual outcomes, we estimate the following two equations:

$$Win_{imsct} = \alpha + \beta Post_{msct} + \lambda_{sc} + \epsilon_{ijmsct} \tag{2}$$

where Win_{imsct} is an indicator variable taking 1 if the home team wins the match and 0 otherwise. Similarly, we estimate

$$Draw_{imsct} = \alpha + \beta Post_{msct} + \lambda_{sc} + \epsilon_{ijmsct} \tag{3}$$

[16] We also estimate Equation 1 with no fixed effects, matchweek fixed effects, and league fixed effects. Results are qualitatively similar, indicating that the choice of fixed effects do not drive our results.

where $Draw_{imsct}$ is an indicator variable taking 1 if the match is a draw and 0 otherwise. All subscripts and right hand side variables are the same as in Equation 1. The coefficient of interest in these equations, β, represents the change in the probability that a particular outcome occurs conditional on fixed effects, λ_{sc}.

5 Results

Table 4 presents our primary results. We find that the home goal difference decreased by between 0.21 goals and 0.25 goals, depending on the inclusion of various fixed effects, from a pre-COVID-19 level of 0.39 more goals for the home team than the away team. In other words, the home field advantage decreased by between 55 and 64 percent of the baseline. There is a positive but small and statistically insignificant effect on total goals per game, shown in Table 5; even the largest estimated coefficient would only indicate a 3.6% increase in total goals per game. Therefore, the shift in goal difference is derived from approximately equal parts of fewer goals for the home team and more goals for the away team, relative to the control group.

Results are qualitatively similar for the home field advantage in terms of expected goal difference, presented in Table 6. We find that the home expected goal difference decreased by between 0.20 expected goals and 0.23 expected goals from a pre-COVID-19 level of 0.31 more expected goals for the home team than the away team. In other words, the home field advantage decreased by between 64 and 73 percent of the baseline. While the absolute effects are slightly smaller for the expected goals, they constitute a larger percentage of the baseline home field advantage than the realized goals difference shown in Table 4. There is a positive but small and marginally statistically significant effect on total expected goals per game, shown in Table 7. As was the case with realized goals, this effect only indicates a 3.7% increase in total expected goals per game and is much lower in magnitude than the effect on home minus away expected goals shown in Table 6.

These shifts in goal difference due to the lack of fans manifest in fewer wins for the home teams, as is shown in Table 8. There is no effect on the probability of a draw for the home team, shown in Table 9. This null effect is both close to zero and precisely estimated. Between 4.8 percentage points and 6.4 percentage points are thus shifted from the probability of winning to the probability of losing, depending on the inclusion of various fixed effects. Assuming that the implementation of the no-fans policy cannot benefit the home team, there are three possible shifts in home-draw-loss outcomes for each game. A home win could become a draw, a home win could become a home loss, and a draw could become a home loss. By estimating a precise null effect of the no-fans policy on the probability of a draw, we show that the shifts from home win to draw are approximately equal to the shifts from draw to home loss. This is evidence for the linearity of effect of fans on the home field advantage, at least locally around the goals that are pivotal for match outcomes.

One might be concerned that COVID-19 itself plays some role in home field advantage after these European soccer leagues resumed. Specifically, it is possible that more COVID-19 cases in a teams region could negatively affect their performance on the field, and that an away team from a high-infections region playing a home team from a low-infections region could suffer a performance penalty which might be confused with changes in home field advantage. Table 11 shows that inclusion of controls for prevalence of COVID-19 in a team's region, as measured by either total cases or new cases, do not affect our estimated coefficients for the effect of the no-fans policy on home field advantage. Table A.1 highlights that the results presented earlier are robust to the inclusion of controls for both distance and weather, indicating that neither of these potential confounds affect identification.

In summary, we find that European soccer matches after the break in the 2019-2020 season experienced a sharp decrease in the home field advantage across three key metrics. First, the actual difference between home and away goals decreased by between 55 and 64 percent relative to the baseline. Second, this drop in actual goals is representative of changes in the home field advantage in terms of chance creation, as measured by expected goals, which decreased by approximately the same amount. Lastly, the changes in the home minus away goal difference represent a change in pivotal outcome-affecting goals, as is shown by the shift from the probability of a home win to to the probability of a home loss.

6 Conclusion

Using exogenous variation in attendance due to the COVID-19 pandemic, we explore the impact of fans on home field advantage in top European soccer leagues. We find that the absence of fans leads to 57% decrease in home field advantage as measured by home minus away goals, with the estimated home effect decreasing from 0.387 to 0.167 goals per game. The absence of fans leads to a 68% decrease in home minus away expected goals, indicating that these changes in home field advantage are not driven by better or worse finishing but are instead indicative of changes in chance creation.

This paper contributes to the large literature aimed at estimating home field advantage in sports. Our results very similar to Ponzo and Scoppa (2018), which apportions 60% of the home field advantage to differences in fans by holding travel and familiarity constant across the home and away team. We take the opposite approach, instead eliminating the effect of fans and observing what remains and find that home field advantage decreases by 57% when fans are removed. Additionally, our results are based on matches from four different European leagues at various levels of distance and robust to differences in weather, exposure to COVID-19, and seasonality in home field advantage. We thus provide evidence that fans account for a sizable portion of the home field advantage throughout all of the European soccer leagues and not just same-stadium matchups in Serie A.

A major limitation of this paper is that we measure the effect of going from the "standard" number of fans down to zero. However, this standard level of fan attendance varies across leagues, teams, and even games. Similarly, the impact of fans may be different for different types of competitions. Pollard and Pollard (2005) suggests that home advantage is smaller in national cup competitions, like the FA Cup in England, but larger in continental club competitions, like the UEFA Champions League. These different types of games also experience different levels of travel and fan attendance and may provide future evidence for the marginal effects of fans, as opposed to the total effect measured in this paper. The marginal effect of fans is a promising area of future research with real-world applications regarding optimal stadium capacity and ticket-pricing schema, as soccer teams try to optimize on-field performance in addition to match-day revenue.

References

Agnew, Gary A, and Albert V Carron. 1994. "Crowd effects and the home advantage." *International Journal of Sport Psychology*.

Auffhammer, Maximilian, Solomon M Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using weather data and climate model output in economic analyses of climate change." *Review of Environmental Economics and Policy*, 7(2): 181–198.

Barnett, V, and S Hilditch. 1993. "The effect of an artificial pitch surface on home team performance in football (soccer)." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 156(1): 39–50.

Belchior, Carlos Alberto. 2020. "Fans and Match Results: Evidence From a Natural Experiment in Brazil." *Journal of Sports Economics*, 1527002520930812.

Boyko, Ryan H, Adam R Boyko, and Mark G Boyko. 2007. "Referee bias contributes to home advantage in English Premiership football." *Journal of sports sciences*, 25(11): 1185–1194.

Carleton, Tamma, Jules Cornetet, Peter Huybers, Kyle Meng, and Jonathan Proctor. 2020. "Ultraviolet radiation decreases COVID-19 growth rates: Global causal estimates and seasonal implications." *Available at SSRN 3588601*.

Clarke, Stephen R, and John M Norman. 1995. "Home ground advantage of individual clubs in English soccer." *Journal of the Royal Statistical Society: Series D (The Statistician)*, 44(4): 509–521.

Cohen, Ben, and Joshua Robinson. 2020. "The world's best athletes are now better at shooting; NBA stars in the bubble and european soccer players in empty stadiums have one thing in common: They have improved at shooting." *Wall Street Journal (Online)*.

Ekstrand, Jan, Armin Spreco, and Michael Davison. 2019. "Elite football teams that do not have a winter break lose on average 303 player-days more per season to injuries than those teams that do: a comparison among 35 professional European teams." *British journal of sports medicine*, bjsports–2018.

Goumas, Chris. 2014. "Home advantage in Australian soccer." *Journal of Science and Medicine in Sport*, 17(1): 119–123.

Jamil, Mikael, Stuart A McErlain-Naylor, and Marco Beato. 2020. "Investigating the impact of the mid-season winter break on technical performance levels across European football–

Does a break in play affect team momentum?" *International Journal of Performance Analysis in Sport*, 1–14.

Loughead, Todd M, Albert V Carron, Steven R Bray, and Arvin J Kim. 2003. "Facility familiarity and the home advantage in professional sports." *International Journal of Sport and Exercise Psychology*, 1(3): 264–274.

Menne, Matthew J, Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, Ron Ray, Russell S Vose, Byron E Gleason, et al. 2012. "Global historical climatology network-daily (GHCN-Daily), Version 3." *NOAA National Climatic Data Center*, 10: V5D21VHZ.

Nichols, Mark W. 2014. "The impact of visiting team travel on game outcome and biases in NFL betting markets." *Journal of Sports Economics*, 15(1): 78–96.

Oberhofer, Harald, Tassilo Philippovich, and Hannes Winner. 2010. "Distance matters in away games: Evidence from the German football league." *Journal of Economic Psychology*, 31(2): 200–211.

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

Pollard, Richard, and Gregory Pollard. 2005. "Home advantage in soccer: A review of its existence and causes."

Ponzo, Michela, and Vincenzo Scoppa. 2018. "Does the home advantage depend on crowd support? Evidence from same-stadium derbies." *Journal of Sports Economics*, 19(4): 562–582.

Rathke, Alex. 2017. "An examination of expected goals and shot efficiency in soccer." *Journal of Human Sport and Exercise*, 12(2): 514–529.

Tables

Table 1: Covid Summary by League

	Bundesliga	La Liga	Premier league	Serie A
Last Game before Pause	March 11	March 10	March 9	March 9
Days between Games	66	93	100	103
Games Cancelled	0	0	0	0
On-Field Changes	<——— 5 substitutes, Water Breaks ———>			
Away Travel	Multiple buses	Travel organized in-house	Fly, bus, or personal vehicles	Multiple buses

Notes: This table includes some basic information on how COVID-19 impacted each of the leagues. All of the leagues stopped at about the same time, but restarts were staggered. No games were cancelled or postponed after the restart in any of the four major leagues.

Table 2: Summary Statistics for All Leagues

	Mean	Standard Error	Minimum	Maximum
Attendance (thousands)	32.52	18.74	0.01	259.30
Stadium Distance (km)	339.05	260.14	0.00	2259.93
Home Team Goals	1.57	1.32	0.00	10.00
Home Team Expected Goals	1.48	0.79	0.00	5.80
Away Team Goals	1.19	1.17	0.00	9.00
Away Team Expected Goals	1.19	0.71	0.00	5.30
Home-Away Goals Difference	0.38	1.83	-9.00	8.00
Home-Away Expected Goals Difference	0.29	1.15	-5.10	5.50

Notes: This table shows the summary statistics for all of the leagues. An observation is a game and stadium distance for each game is the distance between the two home stadia. Expected goals are only for the 2017-2018 season onwards for all leagues.

Table 3: Summary Statistics by League

	Bundesliga	La Liga	Premier	Serie A
Attendance (thousands)	42.98	27.96	37.29	23.72
	(17.66)	(20.64)	(16.03)	(13.93)
Stadium Distance (km)	293.74	487.62	187.76	378.25
	(146.78)	(341.51)	(110.38)	(251.45)
Home Team Goals	1.64	1.59	1.57	1.52
	(1.36)	(1.36)	(1.32)	(1.25)
Home Team Expected Goals	1.59	1.43	1.43	1.48
	(0.82)	(0.75)	(0.79)	(0.80)
Away Team Goals	1.30	1.13	1.18	1.18
	(1.23)	(1.15)	(1.16)	(1.13)
Away Team Expected Goals	1.32	1.06	1.17	1.23
	(0.76)	(0.63)	(0.72)	(0.72)
Home-Away Goals Difference	0.34	0.46	0.38	0.34
	(1.94)	(1.85)	(1.84)	(1.69)
Home-Away Expected Goals Difference	0.28	0.37	0.26	0.25
	(1.23)	(0.98)	(1.21)	(1.19)

Notes: This table shows the differences between leagues across some key dimensions. Each cell represents the mean for that league, while the standard deviation is shown in parentheses. When calculating these statistics, an observation is a game and stadium distance for each game is the distance between the two home stadia. Expected goals are only for the 2017-2018 season onwards for all leagues.

Table 4: Impact of No Fans on Home Minus Away Goals

	(1)	(2)	(3)	(4)
Baseline	0.387***	0.388***	0.388***	0.387***
	[0.013]	[0.013]	[0.013]	[0.013]
No Fans	-0.213**	-0.248**	-0.245**	-0.220**
	[0.100]	[0.103]	[0.103]	[0.103]
Week FE	No	Yes	Yes	No
League FE	No	No	Yes	No
Week*League FE	No	No	No	Yes
Observations	15,906	15,906	15,906	15,906
Games without Fans	408	408	408	408
Clusters	7,953	7,953	7,953	7,953

Notes: This table shows the change in home minus away goals when games are played with no fans (behind closed doors. This analysis uses data for all seasons from 2009-2020 in the following leagues: Bundesliga, Premier League, La Liga, and Serie A. The second row shows the estimated fan effect and the first row shows the home field advantage prior to when all games were played without fans. Each column shows a separate specification. The first column has no controls. The second column controls for seasonality in home field advantage with week fixed effects. Building on this, the third column controls for differences in the home field advantage across each of the five leagues. The last column controls for differences in levels and seasonality through week by league fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Impact of No Fans on Total Goals per Game

	(1)	(2)	(3)	(4)
Baseline	2.766***	2.767***	2.767***	2.767***
	[0.014]	[0.014]	[0.014]	[0.014]
No Fans	0.099	0.047	0.039	0.062
	[0.079]	[0.083]	[0.083]	[0.082]
Week FE	No	Yes	Yes	No
League FE	No	No	Yes	No
Week*League FE	No	No	No	Yes
Observations	15,906	15,906	15,906	15,906
Games without Fans	408	408	408	408
Clusters	7,953	7,953	7,953	7,953

Notes: This table shows the change in total goals when games are played with no fans (behind closed doors. The second row shows the estimated fan effect and the first row shows the total goals per game prior to when all games were played without fans. Each column shows a separate specification. The first column has no controls. The second column controls for seasonality in home field advantage with week fixed effects. Building on this, the third column controls for differences in the home field advantage across each of the five leagues. The last column controls for differences in levels and seasonality through week by league fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Impact of No Fans on Home Minus Away Expected Goals

	(1)	(2)	(3)	(4)
Baseline	0.307***	0.310***	0.310***	0.308***
	[0.015]	[0.015]	[0.015]	[0.015]
No Fans	-0.197***	-0.226***	-0.226***	-0.210***
	[0.065]	[0.073]	[0.073]	[0.073]
Week FE	No	Yes	Yes	No
League FE	No	No	Yes	No
Week*League FE	No	No	No	Yes
Observations	4,336	4,336	4,336	4,336
Games without Fans	408	408	408	408
Clusters	2,169	2,169	2,169	2,169

Notes: This table shows the change in home minus away expected goals when games are played with no fans (behind closed doors. This analysis uses data for all seasons which have expected goals calculated (2018-2020) in the following leagues: Bundesliga, Premier League, La Liga, and Serie A. The second row shows the estimated fan effect and the first row shows the home field advantage prior to when all games were played without fans. Each column shows a separate specification. The first column has no controls. The second column controls for seasonality in home field advantage with week fixed effects. Building on this, the third column controls for differences in the home field advantage across each of the five leagues. The last column controls for differences in levels and seasonality through week by league fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Impact of No Fans on Total xGoals per Game

	(1)	(2)	(3)	(4)
Baseline	2.661***	2.658***	2.660***	2.656***
	[0.016]	[0.016]	[0.016]	[0.016]
No Fans	0.049	0.074	0.051	0.097*
	[0.051]	[0.060]	[0.058]	[0.057]
Week FE	No	Yes	Yes	No
League FE	No	No	Yes	No
Week*League FE	No	No	No	Yes
Observations	4,336	4,336	4,336	4,336
Games without Fans	408	408	408	408
Clusters	2,169	2,169	2,169	2,169

Notes: This table shows the change in total expected goals when games are played with no fans (behind closed doors). The second row shows the estimated fan effect and the first row shows the total goals per game prior to when all games were played without fans. This analysis uses data for all seasons which have expected goals calculated (2018-2020) in the following leagues: Bundesliga, Premier League, La Liga, and Serie A. Each column shows a separate specification. The first column has no controls. The second column controls for seasonality in home field advantage with week fixed effects. Building on this, the third column controls for differences in the home field advantage across each of the five leagues. The last column controls for differences in levels and seasonality through week by league fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Impact of No Fans on Probability of Winning

	(1)	(2)	(3)	(4)
Baseline	0.463***	0.463***	0.463***	0.463***
	[0.004]	[0.004]	[0.004]	[0.004]
No Fans	-0.048*	-0.064**	-0.064**	-0.054**
	[0.025]	[0.026]	[0.026]	[0.025]
Week FE	No	Yes	Yes	No
League FE	No	No	Yes	No
Week*League FE	No	No	No	Yes
Observations	15,906	15,906	15,906	15,906
Games without Fans	408	408	408	408
Clusters	7,953	7,953	7,953	7,953

Notes: This table shows the change in the probability of a win for the home team when games are played with no fans (behind closed doors). The second row shows the estimated fan effect and the first row shows the probability that the home team wins prior to when all games were played without fans. All estimates are from a linear probability model with specifications varying by column. The first column has no controls. The second column controls for seasonality in home field advantage with week fixed effects. Building on this, the third column controls for differences in the home field advantage across each of the five leagues. The last column controls for differences in levels and seasonality through week by league fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9: Impact of No Fans on Probability of a Draw

	(1)	(2)	(3)	(4)
Baseline	0.247***	0.247***	0.247***	0.247***
	[0.003]	[0.003]	[0.003]	[0.003]
No Fans	-0.012	0.002	0.002	-0.003
	[0.021]	[0.022]	[0.022]	[0.022]
Week FE	No	Yes	Yes	No
League FE	No	No	Yes	No
Week*League FE	No	No	No	Yes
Observations	15,906	15,906	15,906	15,906
Games without Fans	408	408	408	408
Clusters	7,953	7,953	7,953	7,953

Notes: This table shows the change in the probability of a draw when games are played with no fans (behind closed doors). The second row shows the estimated fan effect and the first row shows the probability of a draw prior to when all games were played without fans. All estimates are from a linear probability model with specifications varying by column. The first column has no controls. The second column controls for seasonality in home field advantage with week fixed effects. Building on this, the third column controls for differences in the home field advantage across each of the five leagues. The last column controls for differences in levels and seasonality through week by league fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10: Impact of No Fans on Home Field Advantage by League

	Germany (1)	Spain (2)	England (3)	Italy (4)
Baseline	0.353***	0.465***	0.385***	0.340***
	[0.031]	[0.025]	[0.025]	[0.024]
No Fans	-0.673***	-0.256	-0.071	-0.042
	[0.254]	[0.176]	[0.229]	[0.180]
Week FE	Yes	Yes	Yes	Yes
Observations	3,366	4,180	4,180	4,180
Games without Fans	82	110	92	124
Clusters	1,683	2,090	2,090	2,090

Notes: This table shows the change in home field advantage, defined as the difference between home and away goals, for each league when games are played with no fans (behind closed doors). The second row shows the estimated fan effect and the first row shows the home field advantage prior to when all games were played without fans. Each column shows a separate regression by league. All regressions include week fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

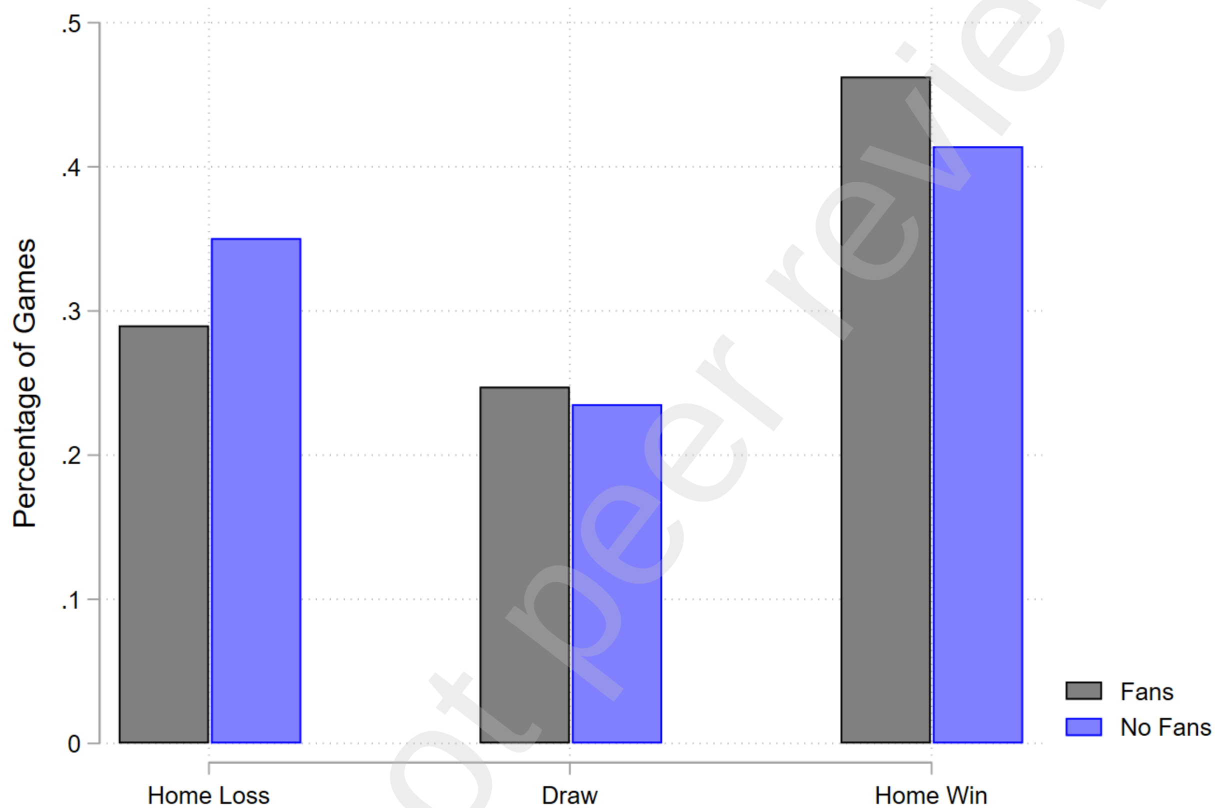
Table 11: Impact of No Fans and COVID-19 Cases on Home Field Advantage

	(1)	(2)	(3)	(4)
No Fans	-0.237**	-0.236**	-0.237**	-0.236**
	[0.097]	[0.097]	[0.097]	[0.097]
No Fans x Total Cases		0.016*		
		[0.009]		
No Fans x New Cases			0.154	
			[0.120]	
No Fans x Total Cases/Pop.				62.572
				[44.709]
League Week FE	Yes	Yes	Yes	Yes
Home Team FE	Yes	Yes	Yes	Yes
Observations	15,906	15,906	15,906	15,906
Games without Fans	408	408	408	408
Clusters	4180	4180	4180	4180

Notes: This table shows the change in home field advantage, defined as the difference between home and away goals, when the seasons continued with no fans interacted with hard COVID-19 impacted the region where the home team is compared to where the away team is. All specifications include league by matchweek and home team fixed effects. The second, third, and fourth rows show estimates of how the home field advantage changed differently based on how hard the home and away team regions were hit by the disease. Row 2 shows the effect of a variable which is the interaction between an indicator for no fans (1 if yes) with the difference in the total number of cases between the home and away team regions as of March 31, 2020. Rows 3 and 4 estimate a similar relationship, but instead uses new cases or total cases per capita as measures for the effect of the disease on a location. If there is a large direct effect of COVID-19 on match outcomes, then one would expect these interactions to be negative, reflecting that the home team is performing worse due to more cases. The results in this table, however, show this is not the case. Each column shows a separate regression by league. All regressions include week fixed effects. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

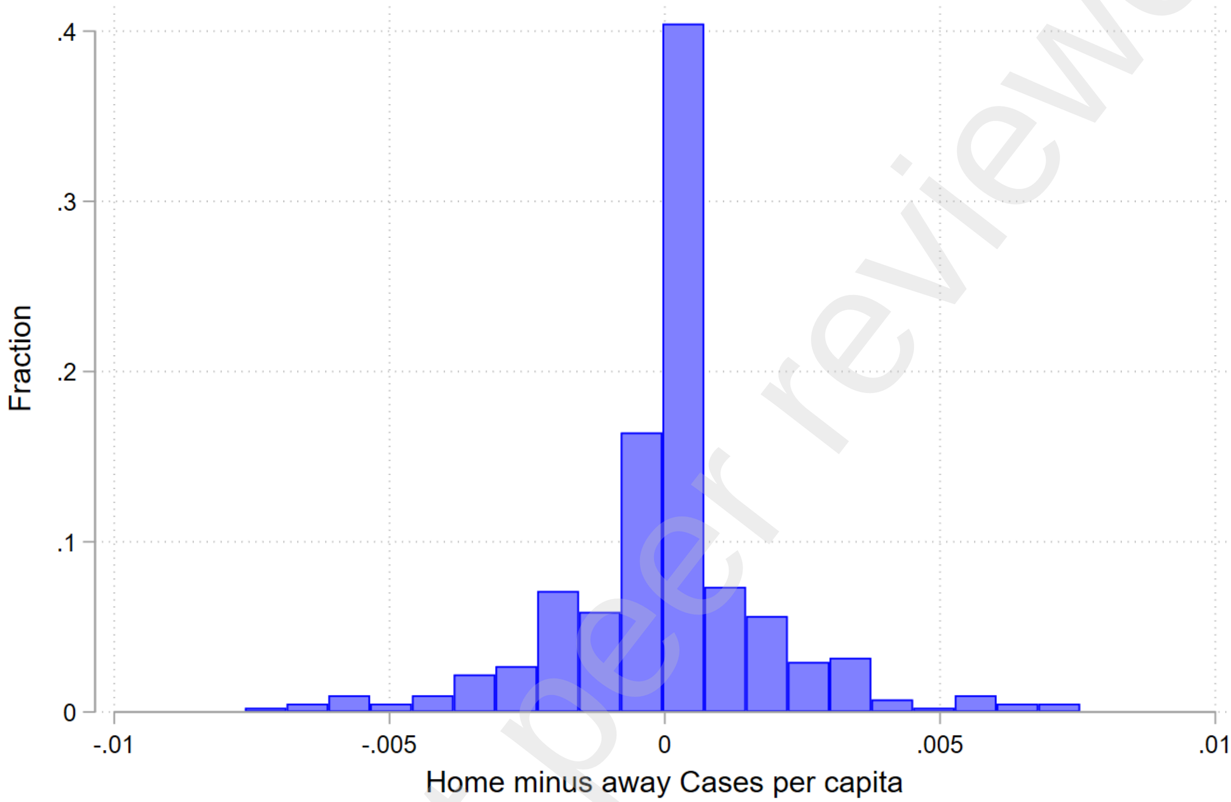
Figures

Figure 1: Change in Match Outcomes with No Fans



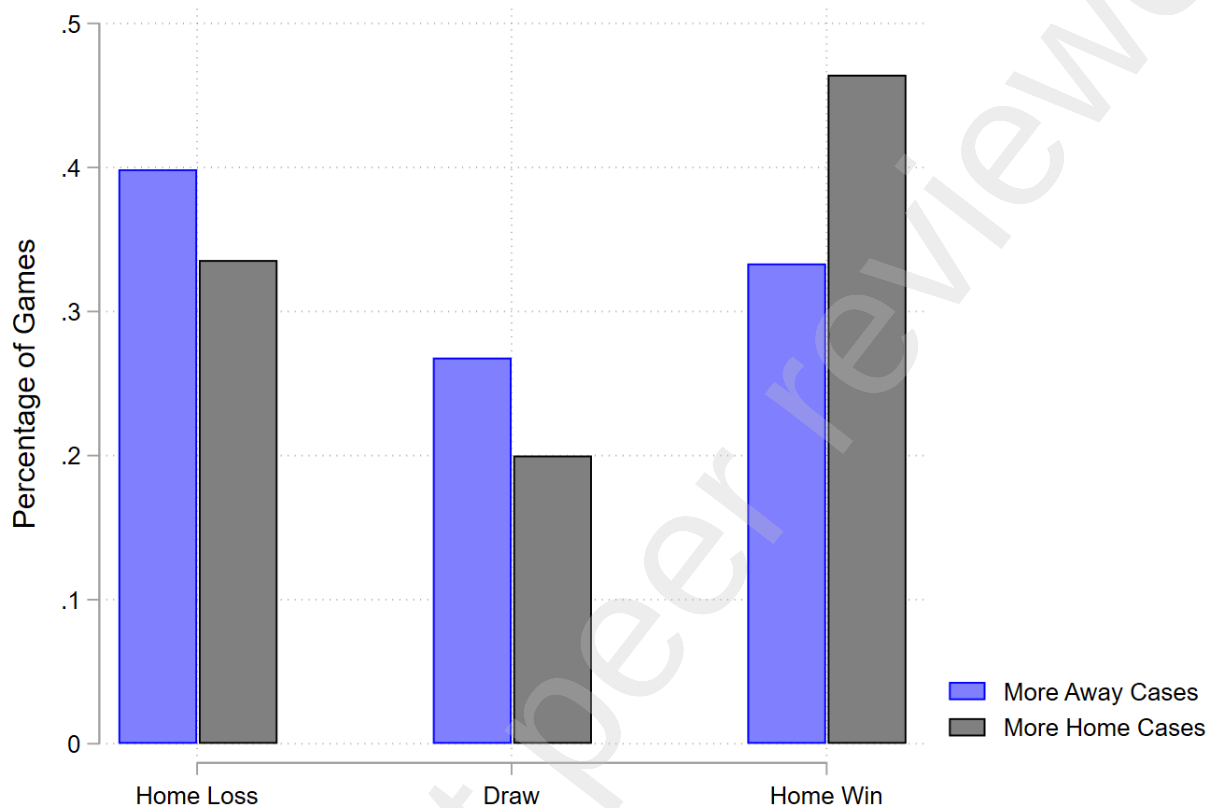
Note: This figure shows the distribution of match outcomes when there are no fans compared to when there are fans. There is little change in the percentage of games that end as a draw, but there are large differences for wins and losses.

Figure 2: Home minus Away Difference in Cases per Capita



Note: This figure plots the distribution of home minus away cases per capita. There are slightly more games where the home team has more cases per capita.

Figure 3: Outcomes Based on Differences in Cases per Capita

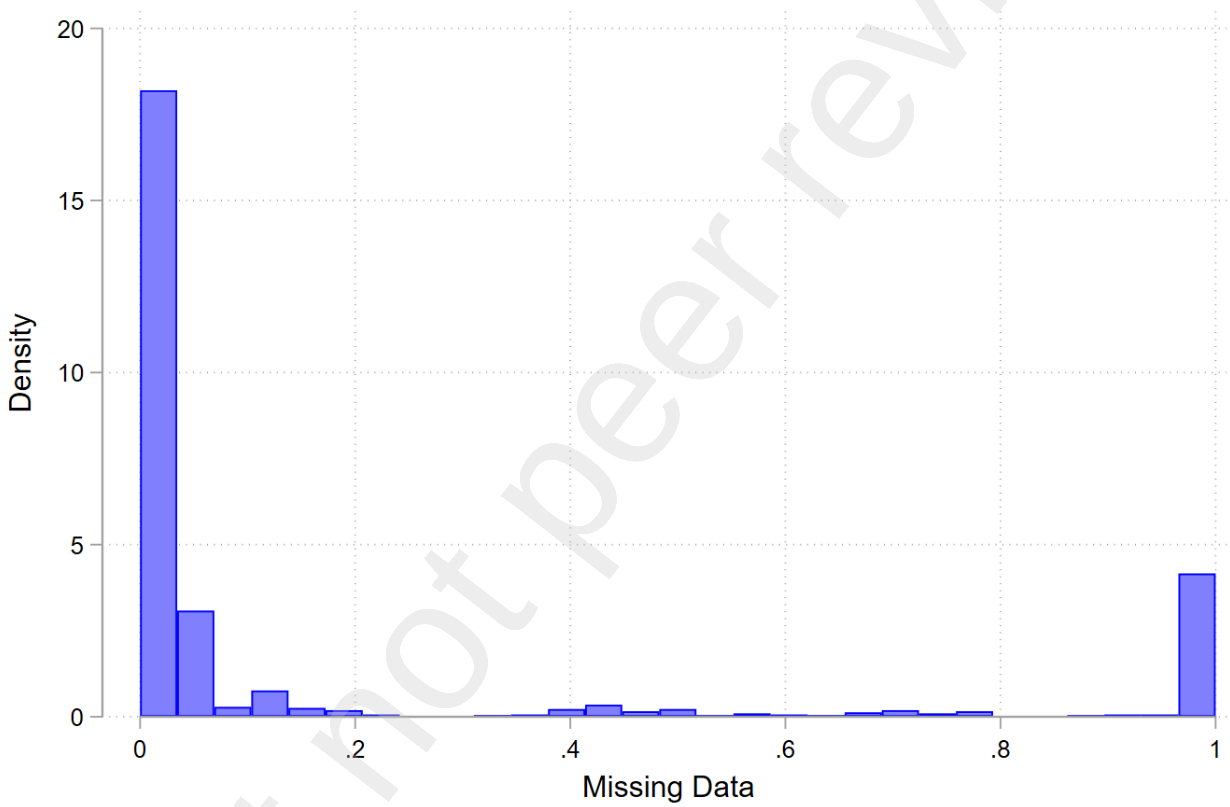


Note: This figure shows the distribution of match outcomes when the away team has more cases per capita compared to when the home team has more cases per capita. Home teams appear to be slightly more likely to lose when the away team has more cases per capita, which is the opposite direction one would expect. One potential reason for this would be that cases could be correlated with innate home team skill.

Appendices

A Additional Tables and Figures

Figure A.1: Missing Days by Weather Station



Note: This figure plots the fraction of observations that are missing for the weather stations in all four countries that the leagues are in. Although many of the weather stations have observations almost every day, many are missing a large fraction of daily observations.

Table A.1: Impact of No Fans on Home Field Advantage Controlling for Distance and Weather

	All Leagues		Germany Only		
	(1)	(2)	(3)	(4)	(5)
No Fans	-0.220**	-0.217**	-0.673***	-0.672***	-0.730***
	[0.103]	[0.103]	[0.254]	[0.254]	[0.269]
Week FE	Yes	Yes	Yes	Yes	Yes
Distance Controls	No	Yes	No	Yes	Yes
Weather Controls	No	No	No	No	Yes
Observations	15,906	15,906	3,366	3,366	3,362
Games without Fans	408	408	82	82	82
Clusters	4180	4180	4180	4180	4180

Notes: This table shows the change in home field advantage, defined as the difference between home and away goals, for each league when games are played behind closed doors. Each column shows a separate regression. The first two columns show the results for all leagues while the last 3 columns focus on Germany. Week fixed effects for all leagues allow for differences across leagues. Distance controls amounts to controlling for distance (km) and distance squared. Weather controls, column 5, are shown only for Germany due to inconsistencies in weather station reporting in all other countries. Specifically, when controlling for weather factors, we allow for a quadratic relationship between the weather (daily precipitation, maximum temperature, and minimum temperature) and the home field advantage. Standard errors in brackets are clustered at the matchup by season level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The inclusion of weather controls does not qualitatively effect the point estimate or standard errors for the effect of no fans policies in Germany, potentially due to the relative homogeneity in climate over the whole country.

B Comparison to North American Sports Leagues

The impact of COVID-19 on the four European soccer leagues studied here differ in some important respects compared to the major North American sports leagues. After a hiatus similar to that in European soccer, North American leagues that were already in the midst of their respective seasons all resumed play in a bubble format designed to minimize exposure to the disease.[17] Games were played without fans and at neutral sites, as opposed to the originally scheduled stadia. In some cases, planned games were cancelled due to the lost time. These bubbles were largely successful in preventing COVID-19 spread, although some whole teams (primarily in Major League Soccer) did experience outbreaks that led to cancellations. Even in the National Basketball Association bubble, many more players missed many more games due to COVID-19 than all four European soccer leagues combined, either from the actual sickness or refusal to play under the bubble conditions.

North American sports leagues that had not yet begun their seasons did not implement a bubble format but instead adopted a structure somewhat similar to European soccer. National Football League and Major League Baseball games were played in originally scheduled stadia, but the National Football League did actually allow some fans to attend. Major League Baseball drastically shortened the season[18] and implemented non-trivial rule changes to actually gameplay, adjusting extra innings, doubleheaders, and designated hitter usage. Additionally, there were many outbreaks in some teams that led to cancelled and postponed games, in stark contrast to the European soccer leagues.

[17] These leagues included the National Basketball Association, National Hockey League, and Major League Soccer.

[18] MLB teams played at most 60 games instead of the usual 162.