2021

Baseball Home Field Advantage Without Fans in the Stands

Jeremy M. Losak
Syracuse University

Joseph Sabel
Syracuse University

International Journal of Sport Finance, 2021, **16,** 148–162, © 2021, West Virginia University

Baseball Home Field Advantage Without Fans in the Stands

Jeremy M. Losak and Joseph Sabel

Syracuse University

Jeremy Losak is an assistant professor teaching in the sport analytics undergraduate degree program at Syracuse University in the David B. Falk College of Sport and Human Dynamics. His research interests span topics in sports, including wagering markets, league rules and structure, spectator demand, and player labor markets.

Joseph Sabel is a recent graduate of the applied data science Master of Science program at Syracuse University. Prior to earning his master's degree, he earned a Bachelor of Science in sport analytics from the David B. Falk College of Sport and Human Dynamics at Syracuse University.

Acknowledgements: The authors are grateful to the anonymous reviewers whose feedback was critical in the final findings of this paper and to EV Analytics for providing timely data. The authors acknowledge the early contributions made by several authors in the field, which laid the groundwork for this article. Finally, the authors thank Shannon Graham for providing detailed and insightful copy edits.

Abstract

Home field advantage is universally accepted across most major sports and levels of competition. However, exact causes of home field advantage have been difficult to disentangle. The COVID-19 pandemic offers a unique, natural experiment to isolate elements related to home field advantage since all 2020 regular season Major League Baseball games were played without fans. Results provide no statistically significant evidence of a difference in home field advantage between the 2019 and 2020 seasons, evidence that home crowd support is not a driver of home field advantage. There does appear to be a statistical advantage by the home team batting second in the inning. Travel fatigue seems to have no impact on home field advantage, and while home field advantage seems to increase throughout the 2020 season, we chalk that up to small sample noise. Despite lacking historical precedence, betting markets seemingly respond efficiently to the new home conditions.

Keywords: home field advantage, market efficiency, baseball, ghost games

http://doi.org/10.32731/IJSF/163.082021.04

Introduction

The phenomenon of a home field advantage (HFA) is universally accepted across most major sports at all levels (Pollard & Pollard, 2005). While its existence is generally accepted, sources of that advantage have been difficult to detect, especially since many occur simultaneously. The literature identifies some potential causes, including crowd support, stadium and location familiarity, travel fatigue effects, referee bias, and psychological factors, among others.

The COVID-19 pandemic offers a unique natural experiment to isolate elements related to HFA. All 2020 regular season Major League Baseball (MLB) games were played in empty stadiums. Cancellations caused by individual team outbreaks were common, creating randomly assigned breaks in the season. For example, the St. Louis Cardinals had a stretch of 16 days between games with 12 postponements and two cancellations. Postponed games were made up with seven-inning doubleheaders instead of traditional nine-inning doubleheaders. Travel was reduced, as teams only played games within their regional divisions. For example, the New York Mets, who play in the National League East, only played games against teams in the American League East and National League East.

We use this natural experiment to analyze the effect of fans, travel fatigue, workload, and "last licks" on HFA. The 2019 MLB season is our control group, and 2020 is our treatment group. Fischer & Haucap (2020a) performed a similar analysis for German soccer, but to our knowledge, this is the first paper to analyze HFA in the era of COVID-19 for a professional North American league.

We also examine HFA in betting markets. Betting markets are often used to test the efficient market hypothesis (EMH) and analyze future markets. According to the EMH, a market is efficient if all available information is incorporated into the current price of an asset such that it is impossible to systematically outperform the market (Fama, 1970). Under semi-strong form market efficiency, all past and current information is included in the price. Jones (2018) identified an annual HFA in MLB between 3–5 percentage points between 2006 and 2015, normal years compared to 2020. The difficulty in pricing HFA in 2020 is the lack of historical precedence without fans in the stands to guide this updated pricing.

Initial inspection revealed an increase in HFA, not decrease, in 2020 compared to 2019. Similar to Fischer and Haucap (2020a) for first division German soccer, HFA increased throughout the sample period. On one hand, this may be evidence of players growing accustomed to playing without fans in the stands, improving HFA. Alternatively, this may be a consequence of small sample variance or other mechanisms related to differences between 2019 and 2020. Decreases in total travel load, and other general fatigue-related variables, do not seem to have influenced HFA in 2020. We do, however, identify positive benefits associated with "last licks." Analyzing betting market response to changes in HFA reveals that markets successfully dealt with changing HFA by not overreacting. When considering the possibility of statistical noise impacting the model, markets mostly appear to be efficient pertaining to changes made in 2020, a slightly different result than that found in Fischer and Haucap (2020b) for first division German soccer.

Literature Review[1]

There is vast interdisciplinary literature analyzing home field advantage across all levels of sport. It is generally agreed that HFA is a consequence of several factors that in a normal year would interact simultaneously (Pollard & Pollard, 2005; Jamieson, 2010). These factors can be categorized by (i) territoriality/psychological reasons, (ii) location familiarity, (iii) referee bias, (iv) crowd support, and (v) other physical reasons (Fischer & Haucap, 2020a).

This paper focuses on (iv) and (v) for MLB, but that does not dismiss other factors. For example, the Blue Jays played the majority of their home games in Buffalo, NY, during the 2020 season due to Canadian-imposed border restrictions, which one would expect to result in a decrease in HFA for the Blue Jays under (ii). Playing without fans may also influence referee bias (iii). While literature on referee influence on HFA is sparse for MLB, it has been shown countless times for European soccer for penalties (Sutter & Kocher, 2004; Dohmen, 2008; Goumas, 2014a) and stoppage time (Scoppa, 2007; Dohmen, 2008; Yewell et al., 2014). And of course, there are several psychological differences to playing games this season that differentiate them from games played in 2019.

We focus on the elements of HFA we observe and see variation in during the 2020 season. There is a vast literature for soccer identifying elements of HFA as it relates to crowd noise (Nevill et al., 1996, Goumas, 2014b; Pollard & Gomez, 2014; Van Damme & Baert, 2019; Peeters & van Ours, 2020), but crowd effects on HFA are still in question. The effects of travel and fatigue on HFA also seem to vary by league, with some papers identifying a positive relationship (Goumans, 2014b; Pollard & Gomez, 2014) and others identifying no relationship (Pollard & Pollard, 2005). A number of recent published and working papers find that HFA is impacted by reduced referee bias without crowd pressure (Bryson et al., 2020; Dilger & Vischer, 2020; Endrich & Gesche, 2020; Reade et al., 2020; Wunderlich et al., 2021).

While not as vast as professional soccer, there are several papers that examine HFA for MLB. Baseball seems to be less susceptible to the home field effect compared to other sports (Gomez et al., 2011), but there is a positive effect (Jones, 2015, 2018). Shmanske and Lowenthal (2009) failed to statistically identify an advantage to having "last licks" (the home team batting second in the inning). Smith and Groetzinger (2010) identified that a one standard deviation increase in attendance results in a 4% increase in the likelihood of a home win. Levernier and Barilla (2006) showed that a team's travel has no impact on home team win probability. Prior to the league settling on a 60-game schedule without fans in the stands, Ehrlich and Ghimire (2020) simulated the 2020 season using logit regression and a neural network, assuming that games played without fans would lead to a neutral environment and identifying that teams with normally strong HFAs would be worse off compared to teams with weaker home advantages.

Our modeling strategy was guided by Fischer and Haucap (2020a). Like us, they examined the effect of ghost games—games played without fans—on HFA, utilizing the COVID-19 pandemic as a natural experiment to analyze the top three

German soccer divisions. German soccer was among the first team sports to return to play following league shutdowns worldwide in March. After restarting, games were played without fans in the stands, like MLB.

Fischer and Haucap (2020a) identified a reduction in HFA for ghost games played in the first division. However, this effect seems to go away over time as players grow accustomed to new playing conditions. This initial reduction in HFA seemed to be driven by reference-dependent deviations from previous occupancy levels. The initial decrease in HFA seemed to be greater for teams with higher occupancy levels prior to the COVID-19 shutdown. Another notable result is that there appears to be no effect of travel fatigue on HFA.

Our approach is similar to Fischer and Haucap (2020a) in that we compare ghost games with non-ghost games. However, MLB was not in progress when the shutdown occurred. Instead, all games played in the condensed 2020 season took place without fans in the stands, and we compare the effect of HFA in the 2020 season versus a normal 2019 season. We examine the persistence of any change in HFA brought on by the lack of fans in stands while examining travel fatigue and other baseball-specific attributes that may impact home field attendance.

We also examine the efficiency of betting markets as it relates to changes in HFA. There exists expansive literature on sports betting markets (Sauer, 1998) that consider various market inefficiencies and biases. Gander et al. (2001) found little evidence of HFA mispricing for basketball and baseball. A number of recent papers have identified that betting odds do not properly reflect the effects of no fans in the stands on HFA. Meier et al. (2020), using data from the most prominent European soccer leagues in Italy, Germany, England, and Spain, found that bookmakers systematically overestimated the home team's win probability during COVID-19-induced home games but that this effect gradually diminished over time. Fischer and Haucap (2020b) identified that betting markets underestimated the impact of the loss in HFA in German first division soccer while overestimating the impact for the German second division. They, along with Winkelmann et al. (2021), also showed a lack of updating expectations, indicating a violation of semi-strong market efficiency.

Similar to these papers, we consider the adaption process of match-related expectations due to new experiences, analyzing the market's semi-strong efficiency from a dynamic point of view. Specifically, we look to identify the extent to which markets adapt to unfamiliar and unexpected changes to HFA.

Methodology

First, we identify elements that impact home field advantage using the following model:

$$\begin{gathered} HomeWin_{ij} = \beta_0 + \beta_1 Season_j + \beta_2 GameDay_{ij} + \beta_3 Season_j \times \\ GameDay_{ij} + \beta_4 NoHomeDesign_{ij} + \beta_5 TravelDiff_{ij} + \beta_6 Season_j \times \\ TravelDiff_{ij} + \beta_7 GameLoadDiff_{ij} + \beta_8 GS_{ij} + \beta_9 OppGS_{ij} + \\ \beta_{10} OPS_{ij} + \beta_{11} OppOPS_{ij} + \varepsilon_{ij}. \end{gathered} \tag{1}$$

The outcome variable, $HomeWin_{ij}$,is the outcome of game i in season j from the home team's perspective (1 for a win and 0 for a loss). We estimate Equation 1 using logistic regression. Our sample includes all 2,397 regular season games from the 2019 season as our control group and all 895 regular season games from the 2020 season as our treatment group, a comparable data set to the 2,976 matches analyzed by Fischer and Haucap (2020a).[2] All regular season games played in 2020 took place without fans in the stands and are indicated by the binary $Season_j$ variable. The $GameDay_{ij}$ variable identifies the number of games prior to and including game i that the home team played (in 2020, most teams played 60 games in the season). This measure should be strongly correlated with the individual cumulative experiences of both the home and away team. We interact that variable with $Season_j$ to measure player adjustments to changes in HFA throughout the season (see Fischer & Haucap, 2020a). The $NoHomeDesign_{ij}$ variable identifies if the home-designated team was playing in the away team's ballpark. Due to the large number of games that needed to be rescheduled owing to positive COVID tests, there are 25 games in the dataset where, usually for one leg of a doubleheader, the traveling team was listed as the home team. This was meant to reduce travel and get more games completed in the condensed schedule. This circumstance was incredibly rare prior to 2020, so the variable effectively measures the effect for the 2020 season. Throughout the paper, we define the home team based on the ballpark where the game is being played and not the official home or away designation.

We also examine variables related to the condensed nature of the schedule, utilizing several different specifications that estimate travel fatigue effects. Team schedules for the 2020 season were designed such that each team only played games within their division and their counterpart division in the opposite league. Since divisions are generally regionally focused, the purpose of this initiative was to reduce the amount of team travel. This allowed teams to play more games in a shorter period, reduced travel expenses to cut total team costs, and restricted player geographic exposure to a subset of regions to minimize risk of league-wide COVID-19 exposure. This travel reduction could have influenced HFA in 2020. The condensed schedule may also have a negative effect in terms of player fatigue. The $GameLoadDiff_{ij}$ variable measures the difference in the number of games played by the home and visiting teams in the previous 10 days.[3]

Our main specification calculates $TravelDiff_{ij}$, the difference in travel loads over the previous five days between the home and away team (positive number indicates the home team traveled more than the away team and vice versa). To calculate this variable, first we identify the various ballpark locations the team played at over the previous five days. Location data includes longitude and latitude coordinates obtained from Google Maps. Using the geosphere package (v1.5-10; Hijmans, 2019) in R 4.0.3 (R Core Team, 2020), we calculate the straight-line mileage between stadiums. For example, on September 8, 2020, the Milwaukee Brewers played the Detroit Tigers in Detroit. Prior to playing the Tigers, the Brewers had no game on the September 3, 2020, then played in Cleveland on September 4, 5, and 6, 2020, followed by an off day on September 7, 2020. In this example, we assume that the Brewers traveled from Cleveland to Detroit after their game on September 6 and spent their off day in Detroit. Converting that travel itinerary to mileage, the team traveled zero miles between September 3 and 4, zero miles from September 4 to 5 and September 5 to 6, 91.22451 miles between September 6 and 7, traveling from Cleveland to Detroit, and zero miles between September 7 and 8. Recent travel is expected to have a greater effect on fatigue than older travel, so we weight travel such that travel on the most recent day is given the highest weight and travel on the fifth day counts the least. For our example, the travel fatigue variable for the away team would be 72.97961.[4] We calculate this for both the home and away team and take the difference. Different weighting schemes and other travel specifications are discussed later in the paper.

Finally, we include two measures of team ability, focusing on the starting lineup and the starting pitcher. These variables identify, within a particular season, the strength of these units for a particular game. The OPS_{ij} and $OppOPS_{ij}$ variables consider the recent on-base plus slugging average (OPS) performance of the players in the starting lineup. If a key player was injured, the team's lineup strength would be lower. To calculate these variables, we aggregated the last 15 game appearance box score statistics for each player included in the lineup and created a composite OPS measure.[5] Players with more plate appearances during their previous 15 appearances are likely players featured more prominently in the lineup and thus have a greater weighting on the team lineup strength measure. We prefer this method over team fixed effects, which would only capture average team offensive performance across the season.

Our pitching measures $AdjGS_{ij}$ and $OppAdjGS_{ij}$ are the average game scores for the team's and opponent's starting pitcher over their previous five games, adjusted for the quality of the opposing lineups in those games. Game score is a statistic created by Bill James that measures a pitcher's performance in any given game started and is highly correlated with winning percentage. We use Tom Tango's definition of game score that is more in line with how we view a pitcher's ability to influence the game outcome (more strongly penalizing walks and home runs allowed).[6] For starting pitchers without five previous starting appearances, we take the average game score for all starting pitchers in that group and apply that estimate. We also identify games that may be considered bullpen days—games in which the team only plans on throwing relief pitchers a few innings each and not a conventional starting pitcher—and apply an aggregate game score for all bullpen days.[7,8]

After examining how changes to HFA impact win percentage in 2020, we examine if these changes are reflected in market prices:

$$\begin{gathered} HomeWin_{ij} = \beta_0 + \beta_1 P_{ij} + \beta_2 Season_j + \beta_3 GameDay_{ij} + \beta_4 Season_j \times \\ GameDay_{ij} + \beta_5 NoHomeDesign_{ij} + \beta_6 TravelDiff_{ij} + \beta_7 Season_j \times \\ TravelDiff_{ij} + \beta_8 GameLoadDiff_{ij} + \beta_9 GS_{ij} + \beta_{10} OppGS_{ij} + \beta_{11} OPS_{ij} + \\ \beta_{12} OppOPS_{ij} + \varepsilon_{ij}. \end{gathered} \tag{2}$$

Most of the variables are similar to those in Equation 1. The variable P_{ij} is the implied win probability from the home team given consensus betting market money lines. Consensus money line data was provided to us by a private company, EV Analytics, for most regular season games in 2019 and 2020.[9] Their consensus lines consider lines offered by various national sport books. We convert their consensus money lines to implied win probabilities using the method given in Sauer (2005).[10] If markets are efficient, the implied win probabilities should be an unbiased estimator of the game outcome, such that the expected value of $HomeWin_{ij}$ equals P_{ij}. All information should be captured by the price, such that none of the coefficients included in Equation 2 should return statistically significant. If any of them do, there is evidence that the information contained in that variable is not efficiently priced into the betting odds.

One notable change between Equations 1 and 2 is the definition of the $GameDay_{ij}$ variable. In the betting model, $GameDay_{ij}$ measures the number of days prior to and including the day of game *i* when a game has occurred in MLB. During the season, odds setters gain new information with each passing game. While β_2 identifies the extent to which HFA is mispriced overall, β_4 captures changes in mispricing of HFA in 2020 throughout the season.

Summary statistics are included in Table 1, broken down by season and by favorite/underdog. Implied home win probability according to consensus betting data was slightly greater in 2019 (52.9%) than 2020 (51.4%). Play also appears

Table 1. Summary Statistics

(Full Data: N = 3,292; Home Implied Win Probability Available: N = 3,124)

Variable	Mean	SD	Min	Max	Variable	Mean	SD	Min	Max
Implied Home Win Probability					***Not Home Designated***				
Full	0.525	0.098	0.254	0.830					
2019	0.529	0.101	0.254	0.830					
2020	0.514	0.090	0.269	0.748	2020	0.028			
Home Favs	0.589	0.063	0.500	0.830	Home Favs	0.025			
Home Dogs	0.429	0.051	0.254	0.500	Home Dogs	0.031			
Travel Diff (5 days, miles, weighted)					***Games Played Last 10 Diff***				
Full	-270.6	635.1	-5969.2	2826.5	Full	0.049	1.308	-6	8
2019	-310.6	688.9	-5969.2	2826.5	2019	0.061	1.038	-4	4
2020	-163.5	443.9	-2647.1	1506.1	2020	0.016	1.847	-6	8
Home Favs	-262.8	613.2	-4663.8	2216.3	Home Favs	0.033	1.303	-5	8
Home Dogs	-279.0	656.1	-5969.2	2826.5	Home Dogs	0.045	1.347	-6	5
Home Travel (5 days, miles, weighted)					***Away Travel (5 days, miles, weighted)***				
Full	328.6	472.6	0	3118.6	Full	599.2	525.8	0	5969.2
2019	357.6	503.7	0	3118.6	2019	668.2	548.3	0	5969.2
2020	250.9	366.0	0	2096.7	2020	414.4	406.3	0	2647.1
Home Favs	325.1	474.2	0	2583.1	Home Favs	587.9	504.7	0	4663.8
Home Dogs	324.7	462.1	0	3118.6	Home Dogs	603.7	554.5	0	5969.2
OPS Home					***OPS Away***				
Full	0.726	0.077	0.473	1.020	Full	0.722	0.079	0.479	0.976
2019	0.731	0.079	0.473	1.020	2019	0.727	0.080	0.491	0.976
2020	0.713	0.072	0.480	0.948	2020	0.708	0.074	0.479	0.947
Home Favs	0.741	0.077	0.473	1.020	Home Favs	0.707	0.076	0.479	0.976
Home Dogs	0.702	0.073	0.480	0.899	Home Dogs	0.741	0.079	0.500	0.960
AdjGS Home					***AdjGS Away***				
Full	47.82	8.835	18.61	81.40	Full	47.92	8.963	16.56	79.95
2019	47.96	8.718	18.61	81.40	2019	48.12	8.890	16.56	79.95
2020	47.46	9.137	21.55	80.18	2020	47.39	9.141	17.81	75.39
Home Favs	49.28	9.002	18.61	81.40	Home Favs	45.97	8.129	16.56	76.99
Home Dogs	45.59	7.997	19.71	75.71	Home Dogs	50.70	9.358	19.34	79.95

to be more condensed in 2020, with a larger range and standard deviation for the $GameLoadDiff_{ij}$ variable. This is not surprising as there were many double headers scheduled during the 2020 season. Travel was also down significantly in 2020, with the mean of the weighted home travel variable 106.7 miles less than that in 2019 and the mean of the weighted away travel variable 253.8 miles less than that in 2019. Considering the travel difference variable, the distribution is much tighter in 2020 compared to 2019, suggesting that travel fatigue should be less of a factor in 2020. Variation in these distances is very geography dependent. For example, teams in the Eastern divisions and Central divisions are closer on average compared to teams in the Western divisions. Since 2020 included only within-regional division games, Table 2 provides an overview of the number of games and win percentage by division broken down by short, medium, and long home and away travel fatigue levels. These classifications are based on the positive (home fatigue) and negative (away fatigue) tertiles of the travel difference variable, such that the tertiles closest to zero are considered short, the middle tertiles away from zero are medium, and the furthest from zero tertiles are large. As expected, the West region featured games with greater fatigue differences while the Central and East regions played more short and medium distance games. In terms of home win percentage, there does not seem to be a discernable pattern. While there is an increase in home win percentage when the away team has greater fatigue in the Central region, that trend is flipped in the East and non-existent in the West.

Our offensive OPS measure was slightly lower in 2020 compared to 2019, consistent with league-wide non-pitcher OPS when comparing the two seasons (0.769 for 2019; 0.740 for 2020).[11] Our starting pitcher game scores are also lower in 2020, which makes sense since starting pitchers threw fewer innings. In 2019, starting pitchers accounted for 57.9% of innings pitched, while in 2020 starting pitchers accounted for 55.5% of innings pitched. The game score statistic is greater, all else equal, when the starting pitcher throws more innings. When the home team is the betting favorite, OPS and AdjGS measures tend to be higher, and vice versa when the road team is the favorite.

One variable we do not include in our main models is previous attendance levels, as was done in Fischer and Haucap (2020a) for first league German soccer. They showed that the initial decrease in HFA seemed to be greater for teams with higher occupancy levels prior to the COVID-19 shutdown. A major consideration for us when comparing our approach to that of Fischer and Haucap (2020a) was the fact that the German soccer season was suspended mid-season and then resumed later, while the entirety of the MLB season was played without fans in the stands. For Fischer and Haucap (2020a), pre-shutdown attendance levels were used when coming to their conclusions regarding occupancy rates.

Figure 1 examines the relationship between attendance levels and changes in home team win percentage between 2019 and 2020. A ratio greater than 1 means the team did better at home compared to on the road in 2020 versus 2019. If teams with greater pre-COVID attendance levels face greater reductions in HFA, you would expect teams with higher 2019 attendances to have ratios less than 1. Panel A considers average attendance per home game, Panel B considers stadium capacity, and Panel C considers the percent of stadium capacity filled. In all three cases, there was no clear negative trendline. So ultimately, while we could have used 2019 attendance levels, it did not appear to have any effect on HFA, and that variable would be ignoring changes in rosters that would have occurred that may be less problematic with the

Table 2. 2020 Home Team W% by Travel Distance Difference by Division

	Count of Games Played				Home Win%			
Travel Difference (weighted miles range)	***East***	***Central***	***West***	***Total***	***East***	***Central***	***West***	***Total***
Large Home Fatigue (301.0, 1506.1)	29	6	49	84	51.72%	33.33%	53.06%	51.19%
Medium Home Fatigue (119.0, 301.0)	20	37	27	84	70.00%	43.24%	66.67%	57.14%
Short Home Fatigue (0.1, 119.0)	36	32	17	85	66.67%	46.88%	58.82%	57.65%
No Fatigue Advantage	25	17	19	60	44.00%	47.06%	63.16%	50.82%
Short Away Fatigue (-0.1, -165.4)	73	81	40	194	58.90%	61.73%	55.00%	59.28%
Medium Away Fatigue (-165.4, -414.6)	56	93	44	193	41.07%	54.84%	52.27%	50.26%
Large Away Fatigue (-414.6, -2647.1)	58	32	104	194	46.55%	65.63%	61.54%	57.73%

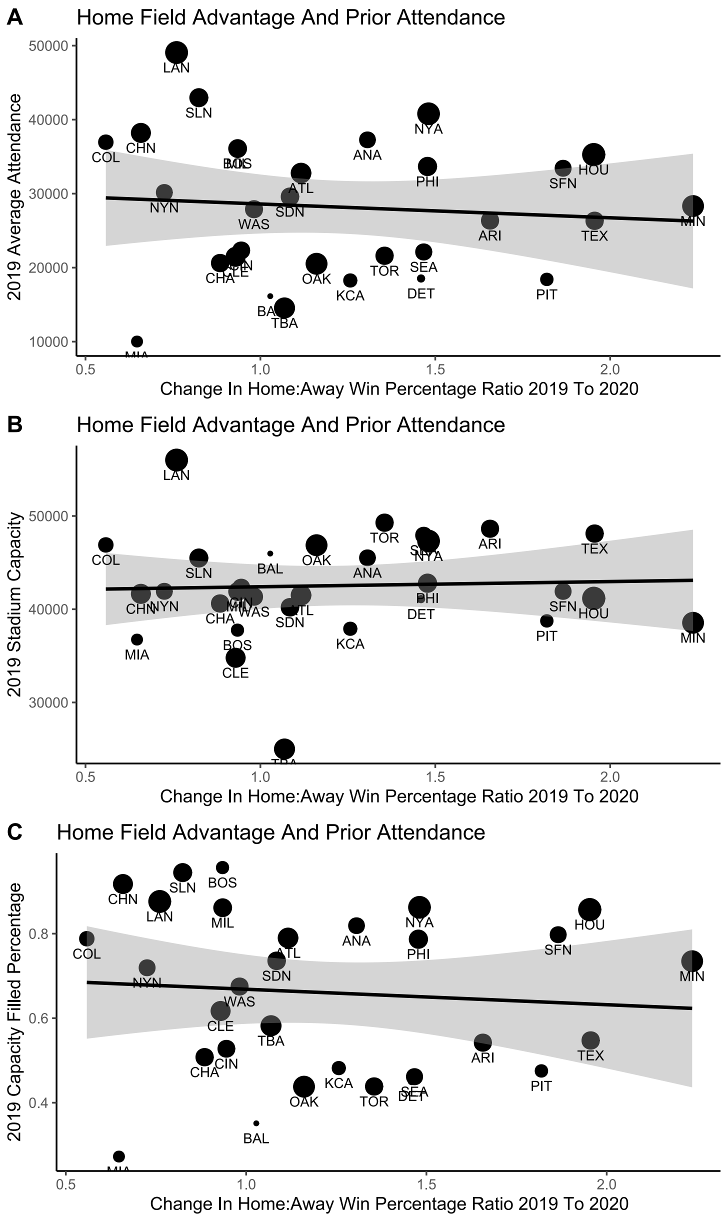

Figure 1.

Changes in the ratio of home to away win percentages in 2020 versus 2019 (x-axis) measured against 2019 average attendance (Panel A), 2019 stadium capacity (Panel B), and 2019 average capacity filled percentage (Panel C). Ratios greater than one indicate an improvement in 2020 of home win percentage relative to away win percentage. The dot size indicates team win percentage in 2019; better teams in 2019 have larger dots. This visualization was made in R 4.0.3 (R Core Team, 2020) using the ggplot2 package (v3.3.2; Wickham, 2016) and cowplot package (v1.1.1; Wilke, 2020).

mid-season suspension in German soccer. We also did not feel comfortable incorporating fixed effects as was done in Fischer and Haucap (2020a) for similar reasons.

Home Field Advantage in 2020

In 2019, the home team won 52.9% of the time (1,286 wins and 1,143 losses). In 2020, the home team won 55.7% of the time (500 wins and 398 losses). Both win percentages are statistically greater than 50% (p-values of 0.004 and 0.001 for 2019 and 2020, respectively), the expected win percentage if home field advantage did not exist, assuming each team had a balanced home and away schedule. The greater home win percentage in 2020 compared to 2019 is also statistically significant at the 10% level (one tailed test p-value of 0.086). In fact, the 2020 home win percentage was at its highest level since 2010. Of course, this simple use of a two-proportions z-test does not control for the various circumstances that differentiate 2019 and 2020. For that, we estimate Equation 1.

Table 3 provides results for variations of Equation 1 using logistic regression. Column 1a is our preferred model specification. Across specifications, the coefficients for OPS and game score are statistically significant and have expected signs: The home team wins more if it has better offensive and pitching talent and loses more if the opponent has better offensive and pitching talent. Not home designated is negative and statistically significant, suggesting that part of the HFA in baseball comes from batting second in an inning and having "last lick" advantages. Differences in game load over the previous 10 days is not statistically significant, failing to provide evidence that greater disparity in game loads impacted home win percentage.

Our main variable of interest, the 2020 season indicator variable, is not statistically significant. Thus, after controlling for differences between 2019 and 2020, there is no statistical evidence that overall home team win percentage in 2020 was different than that of 2019. With that said, players may have needed time to gain familiarity with the new 2020 playing conditions, so HFA may not have been homogenous throughout the

2020 season. Similarly, with increased travel burdens related to maintaining health and safety during the 2020 season, we may expect the effects of travel fatigue to be different in 2020 compared to 2019. Interacting the home game number and travel fatigue difference variables with our season variable examines these possibilities. The non-interacted variable captures the effect of the variable in 2019, while the interaction term captures changes in the effect of the variable on home winning percentage in 2020.

The game number coefficient is statistically significant for 2019. The interaction term, however, is positive and statistically significant, suggesting that HFA increased throughout the 2020 season. Figure 2 uses the model in Table 3 Column 1a and plots predicted home win probabilities for 2019 and 2020 over 60 games (the typical number of games played by teams in 2020), holding other covariates at their means. As can be seen in the plot, the predicted probabilities are not statistically different for the majority of the 2020 season until approximately the 55-game mark. On one hand, this result supports the theory by Fischer and Haucap (2020a) that players needed time to grow comfortable with new playing

Table 3. Home Field Advantage Models

Logistic Regression; Dependent Variable: Home Win (1/0)				
Model #	(1a)	(1b)	(1c)	(1d)
Season (1 = 2020)	-0.172 (0.164)	-0.009 (0.267)	-0.175 (0.164)	-0.089 (0.173)
Not Home Designated	-0.743* (0.423)	-0.739* (0.423)	-0.731* (0.423)	-0.623 (0.442)
Game Number, Home Team	0.0001 (0.001)	0.001 (0.002)	0.003 (0.003)	0.0001 (0.001)
X Season	0.010** (0.004)	0.010* (0.005)	0.015** (0.006)	0.010** (0.004)
Games Played Last 10, Diff	-0.012 (0.027)	-0.014 (0.027)	-0.012 (0.027)	-0.007 (0.027)
Travel Diff 5 (100s of M)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)	-0.006 (0.006)
X Season	0.004 (0.017)	0.005 (0.017)	0.005 (0.017)	0.005 (0.017)
Home OPS	1.504*** (0.469)	1.460*** (0.474)	1.533*** (0.471)	1.490*** (0.470)
Away OPS	-1.898*** (0.459)	-1.930*** (0.465)	-1.884*** (0.461)	-1.879*** (0.460)
Home Adj GS	0.017*** (0.004)	0.017*** (0.004)	0.017*** (0.004)	0.017*** (0.004)
Away Adj GS	-0.010*** (0.004)	-0.010*** (0.004)	-0.010*** (0.004)	-0.010*** (0.004)
Additional Variables		***First Cut*** -0.133 (0.164)	***Second Part*** -0.111 (0.158)	***Double Header*** 0.148 (0.291)
		Second Cut -0.103 (0.184)	***Third Part*** -0.246 (0.266)	***X Season*** -0.294 (0.363)
		Third Cut -0.172 (0.209) ***Fourth Cut*** -0.162 (0.236)	***Fourth Part*** -0.356 (0.381)	***Interleague*** -0.224* (0.125) ***X Season*** -0.066 (0.192)
Constant	0.043 (0.509)	-0.083 (0.541)	-0.031 (0.516)	0.058 (0.512)
N	3,292	3,292	3,292	3,292
Cox & Snell R^2	0.021	0.021	0.021	0.023
AIC	4,501.599	4,508.713	4,506.676	4,501.920

Notes: Statistical significance is defined at the * 10%, ** 5%, and *** 1% levels.

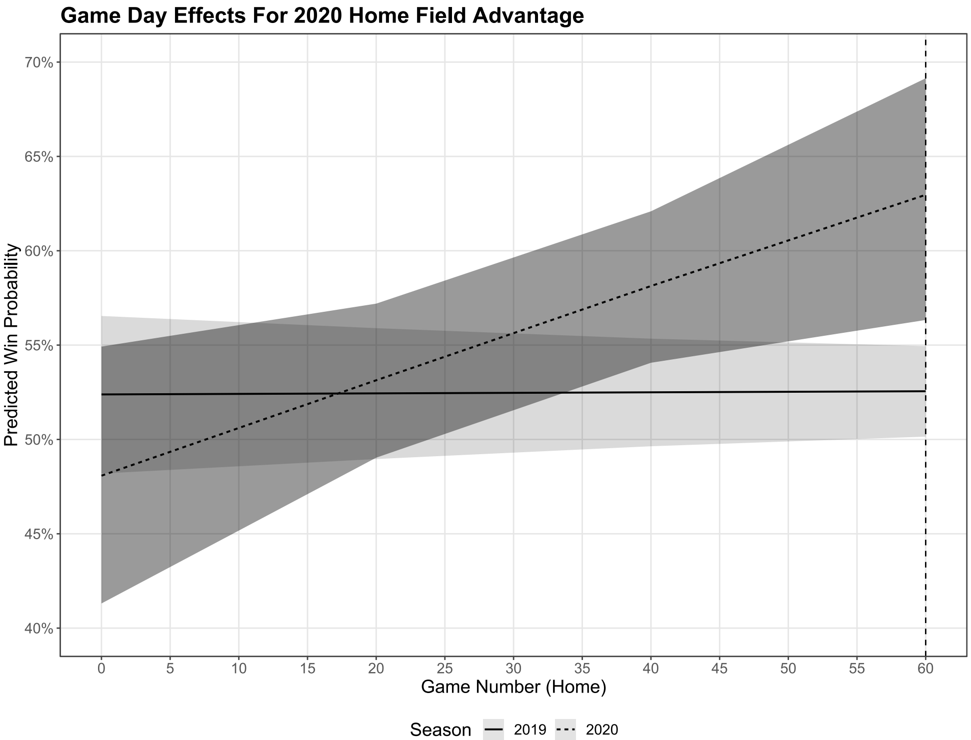

Figure 2.

Predicted win probability from Table 3 Column 1a based on home game number holding all other variables at their means. This identifies, relative to 2019 (light gray), the changing effect of home field advantage throughout the 2020 (dark gray) regular season. The plot shows predicted home win probabilities through 60 games, the number of games played by teams during the 2020 season. Confidence intervals for the predicted win probabilities are given at the 95% confidence level. The visualization was made in R 4.0.3 (R Core Team, 2020) using the sjPlot package (v2.8.7; Lüdecke, 2021).

conditions; any notable declines in HFA went away as the season progressed. Other theories for this result are discussed later in the paper.

Neither the travel difference nor the interaction coefficients returned statistically significant. This fails to provide evidence to support the conventional wisdom that travel fatigue impacts winning. This also fails to support the claim that travel effects on winning may have been different in 2020 due to increased travel precautions and other potentially mentally taxing travel burdens. Robustness of these results are examined in more detail later in the paper.

Columns 1b and 1c consider alternative theories regarding the game day variable results. The 2020 MLB season began in late July, much later than the typical season, which begins in late March/early April. Games played in April and May are more likely to suffer from less-than-ideal weather conditions such as cooler temperatures, and in some cases snow, that may have varying effects on HFA. To control for this, Column 1b synchronizes the season based on cuts of July 23, August 15, September 1, and September 15. The 2020 MLB regular season began on July 23, so the pre-cut variable is a binary variable for games taking place prior to that date during the 2019 season. The first cut variable indicates any game played in either 2019 or 2020 between July 23 and August 15. The final cut, the fourth cut, includes all games played after September 15. These cut variables should control for any seasonal specific elements that may impact HFA, such that the game number variable is only capturing the progression of the season. As shown in Table 3, including the cut variables has negligible impact on the game number coefficients, and none of the season cut variables return statistically significant.

Column 1c considers another way in which the game number variable may not be solely capturing the effects of the progression of the season on home win probability. While game one of the 2019 season and game one of the 2020 season both refer to the beginning of their respective seasons, game 60 in 2019 is still in the later part of the first half of the season, while game 60 in 2020 is the last game of the season. Timing-specific elements including playoff considerations, trade deadlines, and roster adjustment periods such as September call-ups occur at different game numbers between the two seasons. To control for the possibility that the game number interaction variable is capturing some of these differences, we include indicator variables if the game occurred during the second quarter, third quarter, or fourth quarter of the season (compared to the excluded first quarter), where quarters are season-specific. These variables should control for any timing-specific elements that may impact HFA such that game number is only capturing the progression of the season. Again, these control variables have a negligible impact on the game number variables, and none of the quarter variables return statistically significant.

Column 1d introduces two additional elements that control for the new seven-inning double header rules and the universal designated hitter rule in 2020. All the not home designated observations occurred as part of double headers, so it is important to isolate the impact of double headers on HFA in order to properly measure "last lick" advantages. Including the double header variable causes the not home designated variable to lose its statistical significance, although its negative coefficient and *p*-value of 0.159 offer highly suggestive evidence that there is still a batting second advantage. As for the universal designated hitter, American League teams would often be at a disadvantage when playing in National League ballparks because they would not be allowed to use the designated hitter, which in many cases would be one of their better hitters. Likewise, National League teams would be at a disadvantage playing in American League ballparks because their roster construction strategies typically mean they are playing with a lower-quality designated hitter compared to their American League counterparts. According to Column 1d, HFA was less in interleague games overall, an effect that was consistent between the two seasons. This result is not intuitive as one would expect a positive HFA in this situation. Although not listed, we ran an additional model that includes indicators if the interleague home game was in an American League ballpark and if the interleague home game was in a National League ballpark. It appears the negative coefficient on interleague play is being entirely driven by interleague games in American League ballparks, again a result that is not very intuitive.

We also consider various specifications for the travel variable as robustness checks. Table 4 includes the travel fatigue related variables in each specification; changes to other coefficients were negligible and are thus not reported here. The base specification is as described previously, with a weighting scheme of {0.2, 0.4, 0.6, 0.8, 1} applied to travel of the previous five days with the most recent day getting the one weight. The next specification extends the travel under consideration to ten days, with a weighting scheme of {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}.

Next, we consider that greater levels of travel differences observed in 2019 are not observed in 2020, thus the travel variable may be capturing more long-distance travel, while the interaction term is capturing short distance travel. Instead, we split travel fatigue difference into two variables: short travel fatigue and long travel fatigue. Short travel is defined as any travel difference between the 20th percentile (-431.39 miles) and the 80th percentile (109.10 miles) for 2020 travel distance. Travel distances outside this range will skew predominantly towards 2019 observations, while those inside the short travel range will be strongly represented by observations from both seasons. Different thresholds for close and far travel differences were tried, but results were robust.

We incorporate a specification that differentiates home travel fatigue versus away travel fatigue, with the idea that the impact of travel fatigue may be more pronounced if you are actively the visiting team. We also check if time zone changes impact travel fatigue by including an indicator variable if the team changed time zones. While differentiating the exact magnitude of the time zone change would be more granular, we felt this overall indicator variable should be sufficient to detect an overall effect, especially since the expected effects or direction of a hypothetical directional time zone change (east-to-west or west-to-east) are unclear. In our final two specifications, we consider different weighting mechanisms for the five-day travel difference variable, the first putting more weight on more recent travel {0.0625, 0.125, 0.25, 0.5, 1}, and the second putting more weight on earlier travel {0.6, 0.7, 0.8, 0.9, 1}.

The overall conclusion from these robustness checks is that our initial base model results are robust to different specifications of travel fatigue. The inclusion of these various specifications had negligible effects on the other covariates, and none of the alternative measures of travel fatigue were statistically significant. It seems safe to conclude that travel fatigue was not a factor of HFA in 2019, and this did not change in 2020.

Table 4. Travel Fatigue Alternative Specifications

Logistic Regression; Dependent Variable: Home Win (1/0)				
Variation	**Base Model**	**Longer Travel Range**	**Close and Far Travel**	**Home and Away Travel**
Travel Diff Variable 1	Travel Diff 5 -0.006 (0.006)	Travel Diff 10 -0.006 (0.004)	Dis Diff 5 -0.003 (0.024)	Home Travel 5 -0.005 (0.008)
X Season	0.004 (0.017)	0.012 (0.011)	0.005 (0.049)	0.031 (0.022)
Travel Diff Variable 2			Long Dis Diff 5 -0.007 (0.010)	Away Travel 5 -0.007 (0.008)
X Season			0.003 (0.032)	0.015 (0.020)
N	3,292	3,292	3,292	3,292
Cox & Snell R_2	0.021	0.021	0.021	0.022
AIC	4501.599	4500.159	4505.576	4500.860

Variation	**Time Zones**	**More Weight Recent Travel**	**Less Weight Recent Travel**	
Travel Diff Variable	Travel Diff 5 -0.006 (0.006)	Travel Diff 5a -0.012 (0.008)	Travel Diff 5b -0.003 (0.005)	
X Season	0.004 (0.017)	0.013 (0.022)	0.004 (0.013)	
Travel Diff Variable 2	Time Zone Δ 0.017 (0.075)			
X Season				
N	3,292	3,292	3,292	
Cox & Snell R_2	0.021	0.021	0.021	
AIC	4,503.546	4,500.578	4,502.046	

Notes: Statistical significance is defined at the * 10%, ** 5%, and *** 1% levels. All other variables from Table 3, Column 1a are estimated in each of these models; their coefficients are not included here.

Betting Markets and Home Field Advantage

In this section, we analyze how betting markets reacted to the perceived changing landscape of home field advantage. Table 5 examines if elements of HFA are efficiently priced in consensus game money lines. All three columns estimate our base HFA model using logistic regression. The first column analyzes the full data set, and the other two columns split up the data by home favorites and home underdogs.

If betting markets are efficient, all available information should be captured in the price. The price is represented by the implied home team win probability obtained from consensus money lines. Under market efficiency, none of the other covariates should return statistically significant since that information should already be captured in the win probability.

As expected, the OPS and game score measures are no longer statistically significant, as information related to the ability of the offense and pitching is captured in the money lines. Not home designated is no longer statistically significant, possibly suggesting that markets did correctly adjust for the loss in "last licks" advantage. The season variable is not statistically significant, suggesting no systematic mispricing as it relates to fans in the stands. In the previous section, we illustrated that fans likely did not have an impact on HFA. The betting results here suggest that bookmakers did not overreact to the possibility of changes in HFA.

Similar to our base model, the interaction term between game number and season is statistically significant, suggesting that money lines did not capture increases in HFA as the season progressed. If players grew comfortable to the new home field circumstances without fans in the stands, or if changing COVID protocols caused changes in HFA, and HFA increased, betting markets simply did not adjust to this new information. However, our theory is that this statistical effect

Table 5. Betting Markets and Home Field Advantage

Logistic Regression; Dependent Variable: Home Win (1/0)			
Model	*All*	*Home Favorites*	*Home Underdogs*
Implied Win Probability	4.051*** (0.457)	3.366*** (0.834)	5.407*** (1.283)
Season (1 = 2020)	-0.131 (0.170)	0.0004 (0.222)	-0.374 (0.277)
Not Home Designated	-0.616 (0.428)	-0.665 (0.581)	-0.556 (0.655)
Game Number, Season	0.00000 (0.001)	-0.0003 (0.001)	0.001 (0.001)
X Season	0.010*** (0.004)	0.003 (0.005)	0.020*** (0.006)
Games Played Last 10, Diff	-0.013 (0.028)	0.048 (0.037)	-0.105** (0.044)
Travel Diff 5 (100s of M)	-0.008 (0.006)	-0.002 (0.008)	-0.018* (0.010)
X Season	0.008 (0.017)	0.019 (0.023)	0.002 (0.025)
Home OPS	-0.036 (0.506)	-0.233 (0.641)	0.290 (0.835)
Away OPS	-0.764 (0.494)	-0.571 (0.655)	-1.082 (0.763)
Home Adj GS	0.007 (0.004)	0.009 (0.006)	0.003 (0.007)
Away Adj GS	0.001 (0.004)	-0.001 (0.006)	0.005 (0.007)
Constant	-1.858*** (0.579)	-1.402 (0.854)	-2.607*** (0.979)
N	3,124	1,870	1,253
Cox & Snell R_2	0.046	0.016	0.040
AIC	4,196.520	2,511.070	1,689.298

Notes: Statistical significance is defined at the * 10%, ** 5%, and *** 1% levels.

is the consequence of small sample randomness. Considering the interacted effect on home favorites and underdogs, the effect is not statistically significant for favorites, but it is strongly statistically significant for underdogs. Examining that further, home win percentage during the fourth quarter of 2020 was 60.7% and over 55.7% for underdogs, much larger than the 53.5% overall and 46.0% underdog win percentages during the first three quarters of the season. This could be random small sample noise (the home team was the underdog in 97 games during the 2020 fourth quarter, so it would only take a few extra wins to swing the win probability) or could be impacted by the 2020 elimination of the September expanded 40-man rosters.[12]

The only other statistically significant coefficients are in the underdog model. According to the model, it may be profitable to bet against home underdogs or in favor of road favorites when the home team faces heavier recent game loads. It also suggests a possible profitable strategy of betting for underdogs that have a travel load advantage. These results are counter to what we saw in the previous model, with the explanation unclear, save for a possible small sample randomness argument.

Discussion and Conclusion

While home field advantage is generally a universal concept, the specific drivers of that advantage have remained difficult to identify and quantify. This is partly due to the simultaneity of various factors, making it difficult to empirically differentiate their effects. COVID-19 was a major disruptor for the 2020 season across all sports, creating a unique natural experiment that would allow us to differentiate some of these HFA drivers. Specifically, without fans in the stands, we can identify the effect of HFA created by spectators.

Our results fail to show evidence of a change in HFA between 2019 and 2020 for MLB, suggesting that fans in the stand have little effect on home team win probability. We did identify clear "last lick" advantages resulting from teams playing at home but being designated as the visiting team. We also found that HFA increased throughout the 2020 season, which could be due to players needing time to adjust to changing circumstances but could also come from small sample noise introduced by home underdogs during the fourth quarter of the 2020 season. Travel fatigue and changes in travel fatigue conditions showed no impact on HFA.

Examining betting markets, we show evidence that implied win probabilities did not capture the effects of increasing HFA during 2020. While others in the literature (Meier et Al, 2020; Fischer & Haucap, 2020b; Winkelmann et al, 2021) conclude for European soccer that this result is indicative of a violation of semi-strong market efficiency, we are less definitive in that result. Yes, home win percentage in MLB is lower during the first quarter of 2020 (50.5%) compared to the first quarter (53.7%) and first 15 team games (55.2%) of 2019, but our results do not show an early statistically significant difference in reduction in home field advantage. The fact that betting markets do not adjust and HFA losses in previously cited papers tend to disappear as the season progresses leads us to believe that it is just as feasible that small sample variance drove early season results and that betting markets were correct in not overreacting. We show no further evidence that systematic mispricing existed in this market. Whether this means betting markets had correctly identified the negligible impact of fans on HFA or if they simply were lucky due to their slow reaction time is unclear.

There are a number of factors related to HFA that we did not consider in this paper. We do not consider changes in referee bias resulting from a lack of fans in the stands. Of course, this element only matters if there is in fact referee bias. With the implementation of instant replay, the only true element where referee bias may come into play is the ball and strike calls from the home plate umpire. To include referee bias would first require identifying its existence in 2019 and then comparing it to referee bias in 2020. This sort of analysis is outside the scope of this paper but may be an interesting topic for future research.

We also do not consider specific heterogenous elements related to the gameday experience for the home and away team. For instance, we do not observe factors related to artificial noise pumped into stadiums. We also do not consider state-by-state COVID-19 regulations, which may differentiate the travel experience for players in a way that may affect HFA.

MLB's HFA is unique compared to other professional leagues. For one, the literature has shown a much smaller HFA for baseball compared to basketball, football, hockey, and soccer. There is also the added strategic dimension of batting second, which we showed impacts HFA. While our results suggest that the lack of fans had no impact on HFA in MLB, it does not mean that result would hold for other sports. Baseball is much more of an individual sport compared to the aforementioned sports, so there may be team dynamics that draw on home crowd energy that baseball players are less likely to experience.

Given the format the National Basketball Association (NBA) and National Hockey League (NHL) used for their league resets—playing in bubbles without spectators at neutral sites—it would be difficult to test for this effect in those sports using data from the 2019/2020 season. The National Football League (NFL) and college football introduce even more interesting variability in 2020, with only some teams permitted to have fans in the stands based on state and local rules related to mass gatherings. Of course, MLB provided us with a much larger sample to work with (about 60 games for each team) compared to the NFL (16 games per team) and college football (anywhere between 8–11 games per school, depending on the conference). Analysis of different leagues would allow for a more universal understanding of the factors that drive HFA.

Endnotes

[1] We want to acknowledge the comprehensive and recent literature review of home field advantage compiled by Fischer and Haucap (2020a, 2020b). Many of the papers cited here were previously referenced in their paper, and the general direction and focus of the section was certainly influenced by their work. For a more detailed overview, we refer to their paper.

[2] We excluded all neutral site games from the 2019 season, including games that took place in Japan, Mexico, Omaha, London, and Williamsport. There are unique circumstances surrounding most of those games, and they offer unique experiences to a traditional home game.

[3] If a double header is being played, the first game of the double header will count when considering the game load variable for the second game of the double header, but not vice versa.

[4] The primary weighting scheme used is {0.2, 0.4, 0.6, 0.8, 1}. Based on that, the $TravelDiff_{ij}$ calculation for our example is $0.2 \times 0 + 0.4 \times 0 + 0.6 \times 0 + 0.8 \times 91.22451 + 1 \times 0 = 72.97961$.

[5] For the early games of the season, we include prior season games to get to 15 previous appearances. For players making their debuts, or players who are playing their first professional games since before 2017, we calculate the OPS for all players in those situations during their first 15 games and assign those average aggregate values for that player. While there is likely to be some noise in these estimates, as there is certainly heterogeneity in the type of players in this category, the overall variable should still be a good, unbiased approximation of lineup strength. Fifteen games offered us a sizeable sample of recent appearances to limit noisy estimates while not large enough to limit our ability to estimate these variables earlier on in our sample.

[6] See http://m.mlb.com/glossary/advanced-stats/game-score for more details on how the statistic is measured.

[7] Similar to our offensive measures, we expect there to be some noise in this measure. However, it should still do an adequate job over the full sample identifying the expected production of the starting pitching staff that day.

[8] If the team uses an opener, a strategy in which the team begins the game with a bullpen arm and then follows with a traditional starter, we use the starter's game score measure. A more detailed explanation as to how we define the game score variable is available upon request.

[9] Consensus money line data for all games between August 2 and August 14, 2019 was missing due to technical difficulties in data aggregation from our data provider. We do not believe this missing data systematically biases our results.

[10] A general discussion of money line to implied win probability conversion techniques is available in Berkowitz et al. (2018).

[11] This also seems to be true for most offensive variables. For instance, the traditional measures seem to be higher for 2019 (0.256 AVG, 0.327 OBP, 0.443 SLG) compared to 2020 (0.245 AVG, 0.322 OBP, 0.418 SLG). Advanced metrics also support this conclusion (0.325 wOBA in 2019 versus 0.320 wOBA in 2020).

[12] In most seasons, team roster sizes expand from 25 players to 40 players in September, otherwise called September call-ups. Traditionally, teams out of the playoff hunt would use these spots on younger players and give them the opportunity to gain Major League experience in mostly meaningless games. In 2020, roster expansion was capped at 28. It is possible lower quality teams, those likely to be underdogs, were systematically better than they would be in 2019. This possibly explains at least some of the fourth quarter increase in HFA in underdogs in 2020.

References

Berkowitz, J. P., Depken, C. A., & Gander, J. M. (2018). The conversion of money lines into win probabilities: Reconciliations and simplifications. *Journal of Sports Economics, 19*(7), 990–1015.

Bryson, A., Dolton, P., Reade, J., Schreyer, D., & Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during COVID-19. *Economics Letters, 198,* Article 109664.

Dilger, A., & Vischer, L. (2020). No home bias in ghost games. *Discussion Paper of the Institute for Organisational Economics 2/2020.*

Dohmen, T. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic Inquiry, 46*(3), 411–424.

Ehrlich, J., & Ghimire, S. (2020). COVID-19 countermeasures, Major League Baseball, and the home field advantage: Simulating the 2020 season using logit regression and a neutral network [Version 1; peer review; awaiting peer review]. *F1000Research 2020,* 9:414. https://doi.org/10.12688/f1000research.23694.1

Endrich, M., & Gesche, T. (2020). Home-bias in referee decisions: Evidence from "ghost matches" during the COVID-19 pandemic. *Economics Letters, 197,* Article 109621.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417.

Fischer, K., & Haucap, J. (2020a). Does crowd support drive the home advantage in professional soccer? Evidence from German ghost games during the COVID-19 pandemic. *DICE Discussion Paper, No. 344,* ISBN 978-3-86304-343-8, Düsseldorf Institute for Competition Economics (DICE), Düsseldorf. http://hdl.handle.net/10419/222278

Fischer, K., & Haucap, J. (2020b). Betting market efficiency in the presence of unfamiliar shocks: The case of ghost games during the COVID-19 pandemic. *CESifo Working Papers.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3692914

Gander, J. M., Zuber, R. A., & Lamb, R. P. (2001). The home field advantage revisited: A search for the bias in other sports betting markets. *Journal of Economics and Business, 53*(4), 439–543.

Gómez, M. A., Pollard, R., & Luis-Pascual, J. (2011). Comparison of the home advantage in nine different professional team sports in Spain. *Perceptual and Motor Skills, 113*(1), 150–156.

Goumas, C. (2014a). Home advantage and referee bias in European football. *European Journal of Sport Sciences, 14*(sup1), S243–S249.

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

Hijmans, R. J. (2019). Geosphere: Spherical trigonometry. R Package Version 1.5-10. https://cran.r-project.org/package=geosphere

Jamieson, J. P. (2010). The home field advantage in athletics: A meta-analysis. *Journal of Applied Social Psychology, 40*(7), 1819–1848.

Jones, M. B. (2015). The home advantage in Major League Baseball. *Perceptual and Motor Skills, 121*(3), 791–804. https://doi.org/10.2466/26.PMS.121c25x1

Jones, M. B. (2018). Differences in home advantage between sports. *Psychology of Sport and Exercise, 34,* 61–69.

Levernier, W., & Barilla, A. G. (2006). The probability of winning and the effect of home-field advantage: The case of Major League Baseball. *Academy of Information and Management Sciences Journal, 9*(2), 61–77.

Lüdecke, D. (2021). *sjPlot: Data visualization for statistics in social science.* R Package Version 2.8.7. https://CRAN.R-project.org/package=sjPlot

Meier, P. F., Flepp, R., & Franck, E.P. (2021). Are sports betting markets semi-strong efficient? Evidence from the COVID-19 pandemic. *International Journal of Sport Finance, 16*(3), 111–126.
Nevill, A. M., & Newell, S. M. (1996). Factors associated with home advantage in English and Scottish soccer matches. *Journal of Sports Sciences, 14*(2), 181–186.
Peeters, T., & van Ours, J. C. (2020). Seasonal home advantage in English professional football; 1973–2018. *De Economist, 169,* 107–126. https://doi.org/10.1007/s10645-020-09372-z
Pollard, R., & Gomez, M. A. (2014). Components of home advantage in 157 national soccer leagues worldwide. *International Journal of Sport and Exercise Psychology, 12*(3), 218–233.
Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876–2003). *Journal of Sports Sciences, 23*(4), 337–350.
R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/
Reade, J. J., Schreyer, D., & Singleton, C. (2020). Echoes: What happens when football is played behind closed doors? *SSRN Working Paper,* No. 3630130. https://dx.doi.org/10.2139/ssrn.3630130
Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature, 36*(4), 2021–2064.
Sauer, R. D. (2005). The state of research on markets for sports betting and suggested future directions. *Journal of Economics and Finance, 29,* 416–426.
Scoppa, V. (2008). Are subjective evaluations biased by social factors or connections? An econometric analysis of soccer referee decisions. *Empirical Economics, 35,* 123–140.
Shmanske, S., & Lowenthal, F. (2009). Home field (dis)advantage and the "last-ups" effect. *NINE: A Journal of Baseball History and Culture, 18*(1), 130–146.
Smith, E. E., & Groetzinger, J. D. (2010). Do fans matter? The effect of attendance on the outcomes of Major League Baseball games. *Journal of Quantitative Analysis in Sports, 6*(1), Article 4.
Sutter, M., & Kocher, M. G. (2004). Favoritism of agents: The case of referees' home bias. *Journal of Economic Psychology, 25*(4), 461–469.
Van Damme, N., & Baert, S. (2019). Home advantage in European international soccer: Which dimension of distance matters? *Economics, 13*(1), 20190050. https://doi.org/10.5018/economics-ejournal.ja.2019-50
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Tidyverse. https://ggplot2.tidyverse.org
Wilke, C. O. (2020). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2.'* R Package Version 1.1.1. https://CRAN.R-project.org/package=cowplot
Winkelmann, D., Deutscher, C., & Ötting, M. (2021). Bookmakers' mispricing of the disappeared home advantage in the German Bundesliga after the COVID-19 break. *Applied Economics, 53*(26), 3054–3064. https://doi.org/10.1080/00036846.2021.1873234
Wunderlich, F., Weigelt, M., Rein, R., & Memmert, D. (2021). How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the COVID-19 pandemic. *PLOS ONE, 16*(3), Article e0248590. https://doi.org/10.1371/journal.pone.0248590
Yewell, K. G., Caudill, S. B., & Mixon, F. G. (2014). Referee bias and stoppage time in major league soccer: A partially adaptive approach. *Econometrics, 2*(1), 1–19.