Winter 1-29-2021

The Effect of Attendance on Home Field Advantage in the National Football League: A Natural Experiment

Justin Ehrlich
Syracuse University

Joel Potter
University of North Georgia

Shane Sanders
Syracuse University

The Effect of Attendance on Home Field Advantage in the National Football League: A Natural Experiment

Justin Ehrlich[1], Joel Potter[2], Shane Sanders[1]

[1]Department of Sports Management, Syracuse University, Syracuse, New York, USA
[2]Department of Finance and Economics, University of North Georgia, Dahlonega, Georgia, USA

Abstract

While economists have previously noted that home field advantage is affected by crowd density, isolating this effect is difficult since crowd size is likely endogenous with team ability and game matchup. The COVID-19 pandemic has presented a unique natural experiment since local governments have introduced safety protocols that varies widely across the United States. These safety protocols have limited attendance in varying ways for live sporting events, including National Football League games. Given the differential (and exogenous) attendance restrictions, we were able to isolate three broad categories of attendance: 1) games without attendance restrictions (seasons 2016-2019), 2) games with limited in-person attendance, and 3) games without fans. We developed a model to predict the home team point differential, which allowed us to estimate home field advantage of these broad categories. We found that playing in partially full stadiums does not impact relative home team performance (compared to 2016-19 seasons). Conversely, we found that playing with no fans completely eliminates home field advantage. We also tested if fan density mattered in partial attendance game and found that attendance percentage does not statistically affect the home point differential. Hence, our results reveal that the presence of fans matters but the density of fans does not. Furthermore, we were able to compare these results with the 'predictions' offered by the betting market. We found that the betting market correctly predicted that games with partial attendance (due to social distancing) did not impact home field advantage in a statistically significant way and that home team performance would suffer in games played without attendance. Finally, we find evidence that the betting market exhibited a 'learning effect' since the market predicted an increasingly strong 'no fan' effect as the season progressed.

1. Introduction

Economists have noted that players are susceptible to performance pressure [e.g. Sanders and Walia, 2012)] and that crowd density likely influences home field advantage [e.g. Boudreaux et al (2017)]. However, previous researchers have also noted a fundamental problem in that crowd size is likely endogenous with team ability and game matchup [e.g. Smith and Groetzinger (2010)]. Given the differential (and exogenous) safety protocols brought about by the COVID-19 pandemic, we are able to directly test whether fan density influences home field advantage in the NFL. Since the betting market has historically recognized home field advantage in the NFL [e.g. Vergin and Sosik (1999)], we are also able to test whether the betting market correctly predicted the associated impact brought about by the COVID-19 safety protocols.

Individual NFL team policies attendance policies were dictated by local political conditions and local public health mandates. Due to differences in local political situations throughout the U.S. during the pandemic, 117 games were played with a small number of socially distanced fans in attendance while 148 games were played in the complete absence of fans.[1] For the 117 games played with a partial number of fans, the median attendance was 4% of total capacity while the maximum capacity game was at 33.6%. Conversely, during the pre-COVID years (2016-2019) the median game was played with a 99.6% attendance capacity while the minimum game had 61.2% capacity. In 2020: 1) NFL teams played the same number of games as in a typical season, 2) the NFL's season start date was not altered, 3) more than 98% of games were played in the home team's regular stadium; other major North American team sports do not meet these criteria.

As Figure 1 illustrates, we are able to isolate three broad categories of attendance: 1) games without attendance restrictions (seasons 2016-2019), 2) games with limited in-person attendance (shown by the 2020 Partial region), and 3) games without fans (shown by the 2020 No Attendance region). Figure 2 demonstrates the distribution of attendance for 2016-2020, and is color coded the same as in figure 1. Notice the disparity between the 2020 season and the other seasons analyzed in this study.

[1] For a summary of team policies, see https://www.espn.com/nfl/story/_/id/29910246/where-32-nfl-teams-stands-allowing-fans-stadiums

Figure 1: Attendance Variation from 2016-2020

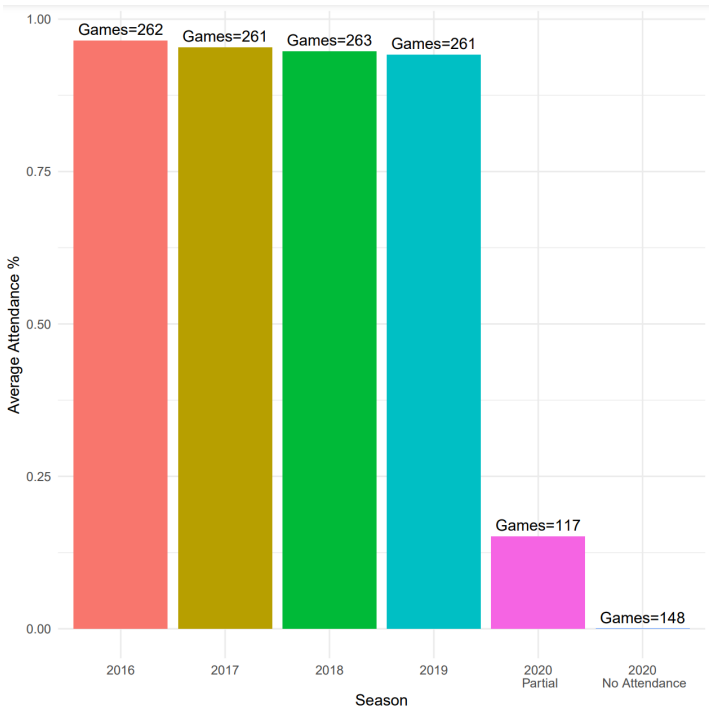
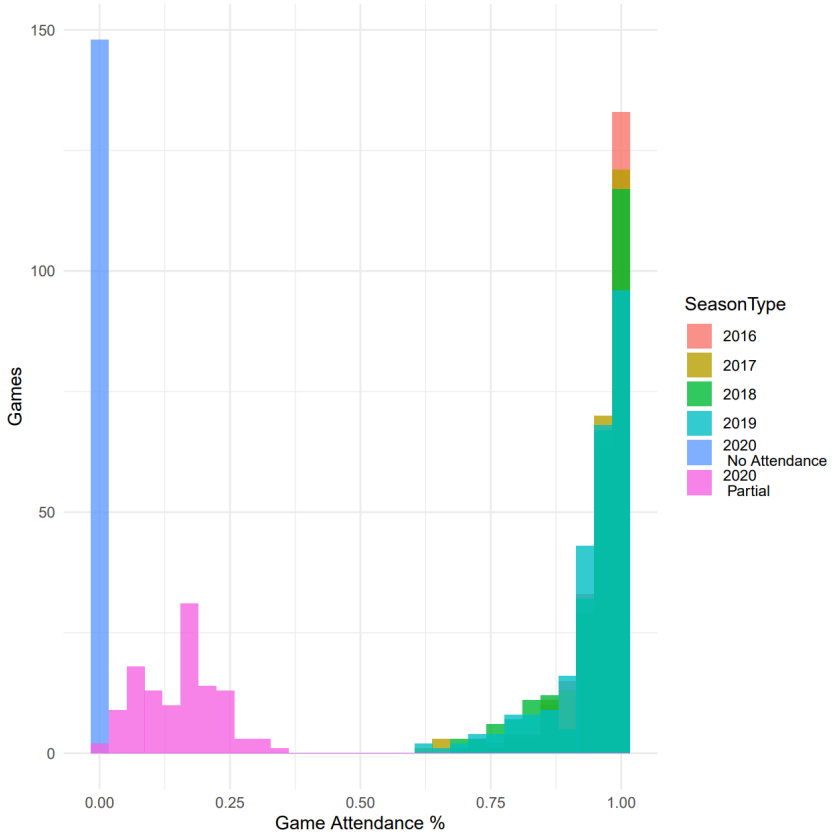


Figure 2: Histogram of Attendance of Games from 2016-2020



2. Data, Model, and Results

Data was obtained from pro-football-reference.com. The sample includes regular and post-season games for seasons 2016-2020 (we exclude all international games and three San Francisco 'home' games played in Arizona during the 2020 season). We extend the ordinary least squares (with team fixed effects) model first proposed by Nettleton (2019) as follows:

$$M_{ijnt} = H + S_{it} - S_{jt} + \beta_1 Partial_{nt} + \beta_2 No\ Fans_{nt} + \beta_2 Home\ QB\ Injury_{int} + \beta_2 Away\ QB\ Injury_{jnt} + \beta_4 Playoffs_{nt} + \beta_1 Rest_{ijnt} + \beta_1 Distance_{jnt} + \varepsilon_{ijnt}$$

Where M_{ijnt} is the difference between home team *i*'s score and away team *j*'s score in game *n* played during season *t*. When the home team loses, $M_{ijnt}<0$. S_{it} and S_{jt} are the respective strengths of team *i* and team *j* that are estimated by the model. The constant, *H*, measures home field advantage from 2016-19. $Partial_{nt}$ is an indicator variable equal to 1 when game *n* was played with a partial number of fans due to COVID-19 social distancing protocols. $No\ Fans_{nt}$ is an indicator variable equal to 1 when game *n* was played without any fans. $Home\ QB\ Injury$ is an indicator variables that is equal to 1 if the home team *i*'s starting quarterback did not play while $Away\ QB\ Injury_{jnt}$ is an indicator variable equal to 1 if the away team *j*'s starting quarterback did not play. $Rest_{ijnt}$ is the difference between the number of days since team *i*'s last game and the number of days since team *j*'s last game. $Distance_{ijnt}$ is the number of miles that team *j* traveled from their city to team *i*'s city. ε_{ijnt} is the unobserved error term.

Table 1: Summary statistics

	mean	std.dev	min	median	max
Margin	1.654	14.106	-49.000	3.000	44.000
Vegas Margin	1.913	6.085	-18.000	3.000	22.000
Partial	0.089	0.285	0.000	0.000	1.000
No Fans	0.113	0.316	0.000	0.000	1.000
Rest	-0.014	2.621	-8.000	0.000	8.000
Distance	969.129	672.517	0.000	829.017	2722.781
Home QB Injury	0.091	0.288	0.000	0.000	1.000
Away QB Injury	0.087	0.282	0.000	0.000	1.000
Playoffs	0.040	0.195	0.000	0.000	1.000
Partial · Attendance	.014	.048	0	0	.336

Our hypothesis is that if fan density matters to home field advantage, then $\beta_2 < \beta_1 < 0$. If home field advantage was completely eliminated by no attendance protocols, then $|\beta_2| = H$.

2.2 Primary Results

Table 2 (model 3) presents the primary results from the above specified model. The estimated constant term for the full specification is 1.748, demonstrating a statistically significant home field advantage for home teams from 2016-19.

Table 2: Dependent Variable: Margin

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Partial	-0.886	-0.846	-0.591	0.717
	(1.373)	(1.376)	(1.373)	(3.202)
No Fans	-2.362**	-2.411**	-2.354**	-2.298**
	(1.102)	(1.097)	(1.104)	(1.109)
Rest		0.0384	0.0236	0.0230
		(0.119)	(0.121)	(0.121)
Distance		0.000319	0.000302	0.000307
		(0.000512)	(0.000517)	(0.000517)
Home QB Injury			-3.744***	-3.722***
			(1.236)	(1.236)
Away QB Injury			0.909	0.902
			(1.337)	(1.337)
Playoffs			0.457	0.448
			(1.754)	(1.756)
Partial · Attendance				-9.038
				(20.33)
Constant	1.822***	1.516**	1.748***	1.742***
	(0.379)	(0.630)	(0.666)	(0.666)
Observations	1,312	1,312	1,312	1,312
R-squared	0.336	0.336	0.341	0.341

Robust standard errors in parentheses[2]
*** p<0.01, ** p<0.05, * p<0.1

Importantly, the coefficient estimates confirm our above hypothesis that $\beta_2 < \beta_1 < 0$. However, the $Partial_{nt}$ coefficient is not statistically significant (the associated p-value is 0.67 in model 3); therefore, playing in partially full stadiums did not impact relative home team performance (compared to 2016-19 seasons). Conversely, the $No\ Fans_{nt}$ coefficient was -2.354 and statistically significant at the 5 percent level, more than completely eliminating home field advantage; a Paternoster test for coefficient equality does not reject the null hypothesis that $|\beta_2| = H$ (see Paternoster (1998) for more information on the test). Finally, $Partial_{nt} \cdot Attendance_{nt}$ was included in model 4 to test if fan density mattered in partial

[2] Tables 2-4 also include team fixed effects for each of the 160 teams in the sample (32 teams in the league across 5 years). Without team fixed effects, results were largely the same as tables 2-4. These unreported results are available from the authors upon request.

attendance games, but was not statistically significant. Hence, our results reveal that the *presence* of fans matters but the *density* of fans does not.

2.3 Did the Betting Market Successfully Predict the Home Field Effects?

In order to test this question, we use the Las Vegas point spread (from pro-football-reference.com) as the dependent variable instead of the actual score margin. Figure 3 visualizes the difference between the actual point differential (Margin) and the betting market predicted point differential (Vegas Margin). If the market predicted the home team to lose (win) a particular game, then the point spread was negative (positive). The independent variables in table 3 are identical to table 2.

Figure 3: Density Plot Demonstrating the Difference between the Actual Margin and the Vegas Margin.

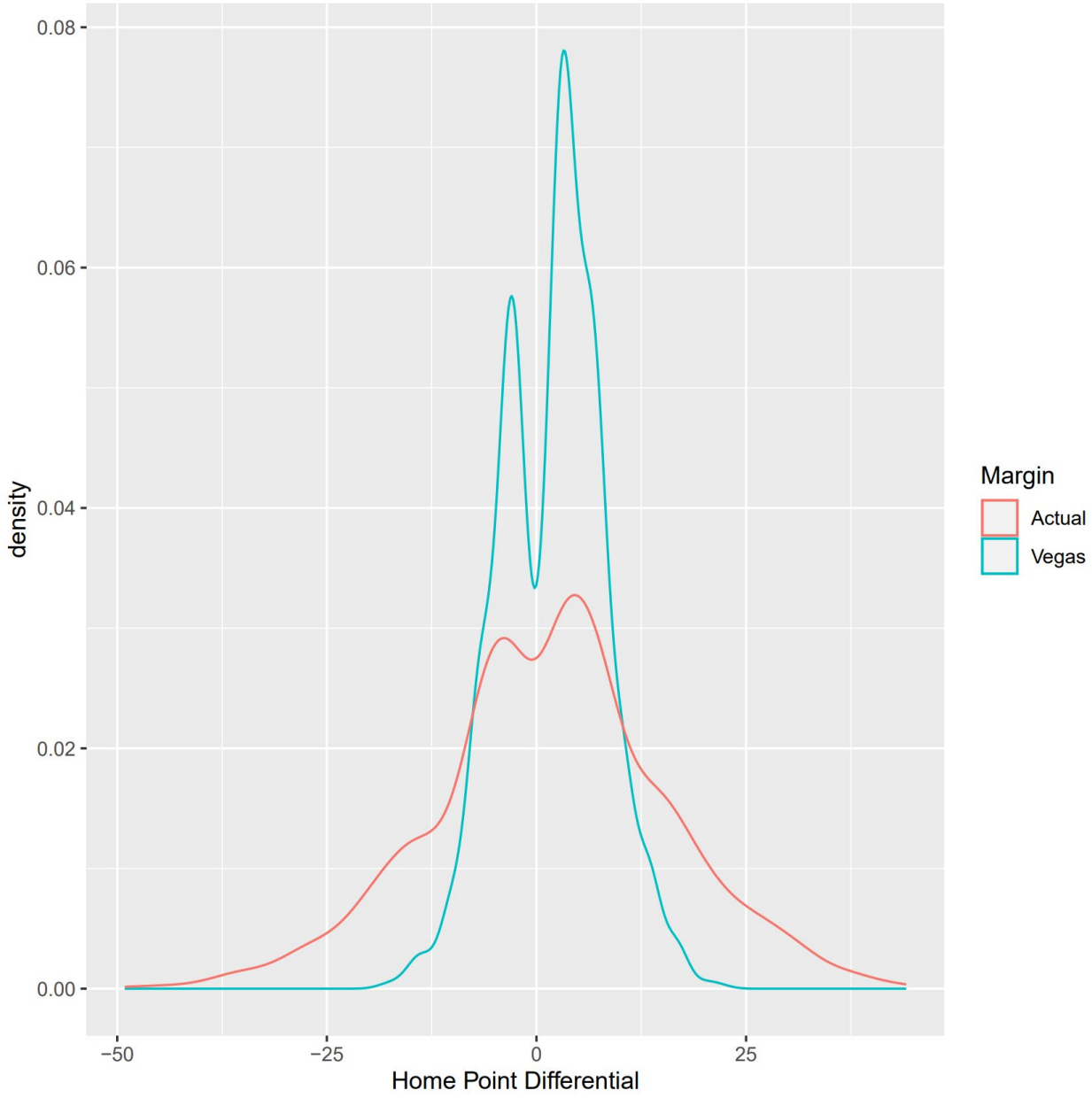


Table 3: Dependent Variable: Vegas Margin

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Partial	-0.558*	-0.536	-0.433	0.00197
	(0.336)	(0.332)	(0.313)	(0.665)
No Fans	-1.071***	-1.114***	-1.109***	-1.091***
	(0.317)	(0.309)	(0.304)	(0.303)
Rest		0.156***	0.130***	0.130***
		(0.0338)	(0.0332)	(0.0332)
Distance		0.000506***	0.000517***	0.000519***
		(0.000145)	(0.000141)	(0.000141)
Home QB Injury			-2.116***	-2.109***
			(0.360)	(0.360)
Away QB Injury			1.744***	1.741***
			(0.403)	(0.404)
Playoffs			1.483***	1.481***
			(0.417)	(0.415)
Partial · Attendance				-3.003
				(4.790)
Constant	1.996***	1.514***	1.481***	1.479***
	(0.103)	(0.172)	(0.168)	(0.168)
Observations	1,312	1,312	1,312	1,312
R-squared	0.737	0.745	0.759	0.759

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

With full controls, results suggest the market successfully predicted the absence of a change in home field advantage for partial attendance games. In addition, the betting market correctly predicted that relative home team performance decreased in games played without fans. Although the market successfully predicted the sign of the $No\ Fans_{nt}$ coefficient, the effect was less than half the size (in terms of absolute value) compared to the table 2 results.[3]

In light of the results from tables 2 and 3, we hypothesize that given the novelty of zero fan games, the betting market may have updated its response to $No\ Fans_{nt}$ games as the season progressed. To test this potential learning effect, we included the following interaction term, $No\ Fans_{nt} \cdot Week$, to table 4, where $Week$ is equal to 1 for games played the first week of the regular season and equal to 20 for games played in the conference championship round.

[3] Interestingly, the signs of the controls were statistically significant in table 3 but a discussion of these results is beyond the scope of this paper.

Table 4: Testing for a market 'learning effect'

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3
Partial	-0.535	-0.420	0.625
	(0.336)	(0.313)	(0.734)
No Fans	-0.0575	-0.146	-0.164
	(0.540)	(0.510)	(0.514)
No Fans · Week	-0.116**	-0.111**	-0.111**
	(0.0526)	(0.0514)	(0.0516)
Partial · Week			-0.103
			(0.0734)
Rest		0.127***	0.126***
		(0.0333)	(0.0333)
Distance		0.000503***	0.000500***
		(0.000141)	(0.000141)
Home QB Injury		-2.048***	-2.023***
		(0.363)	(0.366)
Away QB Injury		1.826***	1.837***
		(0.400)	(0.402)
Playoffs		1.554***	1.714***
		(0.420)	(0.440)
Constant	1.996***	1.480***	1.474***
	(0.103)	(0.168)	(0.168)
Observations	1,312	1,312	1,312
R-squared	0.738	0.760	0.761

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The $No\ Fans_{nt} \cdot Week$ coefficient from table 4 shows the betting market predicted increasingly worse home team performance as the season progressed (in games without fans). In games without fans, the market predicted that home teams would perform 1.11 points worse in week 10, which is equivalent to the table 3 (model 3) estimate of 1.109. However, by week 20, the market predicted home teams would perform 2.22 points worse in games without fans. This is close to the estimate of -2.354 produced by model 3 (table 2).[4] Therefore, it appears the market 'learned' as the season progressed. Finally, $Partial_{nt} \cdot Week$ was included in Model 3 to demonstrate the progression of the 2020 season did *not* impact the market's evaluation of in-person games.

[4] We also added $Partial_{nt} \cdot Week$ and $No\ Fans_{nt} \cdot Week$ to the table 2 (model 3); there were no statistically significant 'week' effects and the signs of the original coefficients were not changed. These results are available upon request.

3. Conclusion

The safety protocols brought about by the COVID-19 pandemic allowed us to test if fan density affects home team performance. We found that home teams did significantly worse in games without in-person attendance, *ceteris paribus*. Furthermore, we were able to compare these results with the 'predictions' offered by the betting market. These findings suggest the betting market correctly predicted that games with partial attendance (due to social distancing) did *not* impact home field advantage in a statistically significant way. We find the betting market correctly predicted that home team performance would suffer in games played without attendance. Finally, we find evidence that the betting market exhibited a 'learning effect' since the market predicted an increasingly strong 'no fan' effect as the season progressed.

References:

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

Nettleton, D. (2018). "An Alternative to the Ratings Percentage Index for Ranking Iowa High School Football Teams." *Midwest Sports Analytics Meetings*. Pella, IA, November 22.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859-866.

Sanders, S., & Walia, B. (2012). Shirking and "choking" under incentive-based pressure: A behavioral economic theory of performance production. *Economics Letters*, *116*(3), 363-366.

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).

Vergin, R. C., & Sosik, J. J. (1999). No place like home: an examination of the home field advantage in gambling strategies in NFL football. *Journal of Economics and Business*, *51*(1), 21-31.