Home Team Advantage in the NBA: The Effect of Fan Attendance on Performance

La, Vincent

Dartmouth College

14 March 2014

Dartmouth College

Home Team Advantage in the NBA: The Effect of Fan Attendance on Performance

Vincent La[1]

March 18, 2014

Abstract

Our study aims to estimate the effect of fan attendance on performance in the National Basketball Association (NBA). We use game day and adverse weather as instruments for attendance. Using two-stage least squares, we fail to find a statistically significant effect of attendance on overall game outcomes. However, again using two stage least squares, we do find a statistically significant effect on away team's free throw percentage. We find that an increase in percent attendance by 10 percentage points is, on average, associated with a decrease in away team's free throw percentage by 1 percentage point. (JEL C36, D01, L83)

Author Notes: Special thanks to Patrick Yukman (Dartmouth College '14) for helping tremendously in collecting attendance data for NBA games. In addition, I would like to acknowledge Professor Douglas Staiger's helpful comments and suggestions.

[1]Contact Email: Vincent.La.14@dartmouth.edu

1 Introduction

Professional athletes around the world are considered to enjoy a significant advantage while playing at home over visitors. This advantage has been so pronounced among many different athletic associations around the world that it is commonly referred to as the "Home Court Advantage". In a less rigorous manner, Moskowitz (2011) studies the home-court advantage across multiple athletic associations and reports the following historical home win percentages shown in Table 1. Our study will focus on home-court advantage specifically within the National Basketball League (NBA).

Table 1: Historical Home Team Win Percentage Across Different Professional Athletic Associations

Association	Home Win %	Years	Home Win % (Last 10 Years)
NBA	**62.7%**	**1946 - 2009**	**60.5%**
WNBA	61.7%	2003 - 2009	61.7%
MLS (U.S)	69.1%	2002 - 2009	69.1%
English Premier (U.K)	63.1%	1993 - 2009	69.1%
MLB (U.S)	54.1%	1903 - 2009	53.9%
Nippon League (Japan)	53.3%	1998 - 2009	53.6%
NFL	57.6%	1966 - 2009	57.3%
NCAA (Football)	64.1%	1869 - 2009	63.0%

As Table 1 depicts, across multiple professional athletics associations, home teams win significantly more on average than away teams. This trend is present both historically and recently (within the past decade). In fact, the home court advantage is so well recognized, that during playoffs, the better teams are often given the opportunity to host the home game as a reward for performing better during the regular season. For example, in the NBA, playoffs are played in a best-of-7 playoff format, with the extra home game going to the superior team. Indeed, in the NBA, there is a substantial difference between home and away team win percentage, but it is unclear as to why this may be. This paper investigates how much of the home court advantage is due to fan attendance.

Many have attempted to explain the existence of the home court advantage. Some commonly cited explanations include familiarity with the playing grounds – home teams who play more often in their own stadium are more familiar with the unique quirks that may be associated with their playing grounds. Another explanation is that there might be rules that favor the home team. Three other commonly cited explanations are: travel-related factors, referee bias, and the psychological effect of fans. The last item is the focus of this paper.

As further motivation for this study, consider the relationship between a team's performance and profit. It is largely established that a more successful team on the court will be more profitable as people generally pay more to watch a winning team. If there is a significant effect of fans on performance, then it might be true that it is to the advantage of an NBA team's management to lower prices in order to raise attendance. If there is indeed a positive relationship between attendance and performance, this could actually increase performance and thus total profit for the team. Note, that this would only be the case for a team that can significantly affect attendance purely through changing ticket prices.

The strategy of this paper will be firstly, to quantify the causal effect of fans on the home-court advantage. This will be done by examining the effect of fans on game outcome. Secondly, this paper will also investigate the effect of fans on game outcomes through how fans affect other metrics of performance. These metrics include those commonly found on a box score, such as free throw percentage, rebounds, assists, steals, turnovers, etc.

2 Existing Literature

There is abundant existing literature which attempts to explain the home-court advantage across all professional sports. However, of the five most commonly cited reasons stated in the introduction, the first two are unlikely to apply for professional basketball. Firstly, the court dimensions in the NBA are standardized across all stadiums. Thus, familiarity with the playing grounds is probably more applicable in sports such as baseball, where each stadium has unique dimensions. Secondly, there are not any official rules in basketball that might favor the home team. Again, this may be more applicable in sports such as baseball, where the home team always bats last – which could lead to some strategical advantages.

Travel related factors, psychological effect of fans, and referee bias, remain to be the most relevant explanations for the home-court advantage in the NBA. However, researchers have generally had limited success in attributing the home-court advantage to travel related factors across several professional athletics associations. Pace and Carron (1992) examine the relative contributions of various travel-related variables to performance in the National Hockey League (NHL). Some travel-related variables included were: number of time zones crossed, distance traveled, preparation/adjustment time, direction traveled, and time of the season. They concluded that travel-related variables only explained 1.5% of the home-court advantage in NHL.

In addition, Courneya and Carron (1991) examine the effect of similar travel-related variables on the home-court advantage in Minor League Baseball. They find that travel-related variables only account for approximately 1.2% of the home-court advantage.

Entine and Small (2008) investigate the effect of travel-related variables on the home-court advantage in the NBA. They primarily use days of rest as their key independent variable in their analysis. First, they demonstrate that on average, away teams have less days of rest between games than home teams do. They then investigate the home-court advantage due to difference in days of rest with respect to both number of points scored and number of games won or lost. However, they ultimately conclude that while the high home-court advantage in the NBA can partially be explained through reduced rest for the away team, the bulk of the advantage arises from other, non-related factors.

There is also significant existing literature regarding referee bias in professional sports. Parsons et al. (2011) investigate racial bias in Major League Baseball with respect to umpires calling balls and strikes. They find that umpires are less likely to call strikes if the pitcher does not match their own ethnicity. In the NBA, Price and Wolfers (2010) also investigate racial bias. In fact, they find significant referee bias due to race (Black vs. Non-Black). Players earn up to 4% fewer fouls and score up to 2.5% more points on nights in which their race matches that of the refereeing crew.

Price et al. (2012) investigates how referee bias might contribute to the home-court advantage in the NBA. This paper takes advantage of the differences between a dichotomy of turnovers: "discretionary" vs. "nondiscretionary" turnovers. Discretionary Turnovers (DTO's) are classified as turnovers resulting from an identifiable violation of the rules (such as traveling violations or fouls). Nondiscretionary Turnovers (NTO's) are turnovers purely caused by players (stepping out of bounds, bad passes, etc.) Observed statistics may be caused by either changes in referee bias or player behavior. While it is extremely difficult to determine whether referee or player behavior caused a turnover, this paper identifies bias by exploiting the fact that referees have varying degrees of discretion over DTO's. The results find that Referees do indeed favor the home team – however, there is no attempt at determining whether this bias increases or decreases with fan attendance.

This study will most closely resemble that of Smith and Groetzinger (2010). They investigate the effect of fan attendance on performance in Major League Baseball. Similar to Smith and Groetzinger, this study will attempt to find the marginal effect of fan attendance on performance, but in the NBA. Smith and Greotzinger find that increased attendance does have a significant effect on many game statistics, ultimately leading to an increased likelihood of winning for the home team. They account for the obvious endogeneity problem by utilizing instrumental variables regression. They find that game day/time and temperature to be valid instruments. They then used two-stage-least squares to find the effect of attendance on score difference and home win likelihood. The paper concludes that increasing percent attendance by one standard deviation (about 25 percentage points) around the mean increases the home win likelihood by 5.5%. They also find that a 38 percentage point increase in attendance as a percent of stadium capacity (i.e an increase in attendance from 40% of stadium capacity to 78% of stadium capacity) is associated with the home team scoring one additional run.

3 Data

Our analysis uses game-level data from the past seven NBA seasons: 2006-2007, 2007-2008, ... , 2012-2013. Normally, an NBA season contains 1230 total games. However, only 990 games were played during the 2011-2012 season due to a collective bargaining agreement dispute. Furthermore, during the 2012-2013 season, one game, Boston Celtics vs. Indiana Pacers, was canceled due to the Boston Marathon bombings. Finally, we exclude three games that were played internationally during the time frame of our analysis[2]. While the NBA did officially record a "home" and "away" team for these games, it is difficult to determine for which team the crowd was generally favoring. Thus, the effect of the crowd is inconclusive. In total, our data comprises of 8,366 games spanning over the past seven NBA seasons.

Attendance data is extracted from box scores hosted on www.basketball-reference.com. In addition, this paper utilizes game-level data obtained from www.nbastuffer.com. Table 2 below displays the overall means for various box score statistics for both Home and Away teams.

Table 2: Table Of Means. Home vs. Away Teams

Variable	**Home**	**Away**
Win %	60%	40%
Total Points	100.6	97.5
Field Goal Attempts	81.3	81.1
Field Goal Percentage	46.4%	45.2%
Three Point Attempts	18.2	18.3
Three Point Percentage	35.8%	35.1%
Rebounds	42.3	41.0
Defensive Rebounds	31.0	30.1
Offensive Rebounds	11.3	10.9
Fouls	20.4	21.2
Assists	22.2	20.6
Steals	7.5	7.3
Free Throw Attempts	24.5	23.6
Free Throw Percentage	75.9%	75.8%
Blocks	5.2	4.6
Turnovers	13.8	14.3
Point Spread	-3.37	3.37
N	8366	

As Table 2 shows, over the past seven seasons, the home team win percentage is approximately 60%. Notice that this is consistent with the historical percentage shown in Table 1. Furthermore, home teams score, on average, approximately three points higher than away teams per game.

The game-level data include one potential instrument for attendance – day of the week the game is played. More specifically, we will test if a weekend game[3] is a valid instrument for attendance. We also obtain weather data from the National Climatic Data Center (http://www.ncdc.noaa.gov/) that includes daily values for temperature, precipitation amount, and dummy variables for certain weather conditions, including but not limited to: snow, rain, hail, thunder, heavy fog, and high winds. We test whether any of these adverse weather conditions might also serve as a valid instrument for attendance.

[2]New York Knicks vs. Detroit Pistons (1/17/2013 London)
New Jersey Nets vs. Toronto Raptors (3/04/2011 and 3/05/2011 London)

[3]We define a game to be played on a weekend if it is played on a Friday, Saturday, Sunday, or Federal Holiday

4 Empirical Design

We utilize instrumental variables to proxy for attendance to avoid a clear endogeneity problem. For example, it could be the case that more fans attend games because the home team is performing well. Thus it may not be that greater attendance leads to better performance – in fact it could be the converse that is true. Using instrumental variables can isolate the effect of attendance on performance.

Using two-stage least squares, our model estimates

$$\Delta\text{Performance} = \beta_0 + \beta_1 \frac{\text{Attendance}}{\text{Stadium Capacity}} + \phi_{it} + \theta_{it}, \tag{1}$$

where Δ is Home - Away, ϕ_{it} are Home * Season Fixed Effects, and θ_{it} are Away * Season Fixed Effects.

We use season * team fixed effects to control for other observable characteristics such as overall quality of the team, as well as unobservables, such as different psychological factors that might be associated with a given stadium. Although controlling for other variables related to within-season variation may not be necessary with a valid instrument(s), as robustness checks, we will also add controls for Home and Away year-to-date records along with Home and Away current win streak (number of games won or lost in a row).

We focus our analysis on two measures of overall performance: $Points_{Home} - Points_{Away}$ and a dummy for a Home Win. However, we also examine the effect of fan attendance on other commonly cited box score statistics, such as free throw percentage, turnovers, rebounds, etc. We limit our analysis to performance metrics and do not consider referee bias due to data limitations.

4.1 OLS Estimation

Before discussing how we selected our instruments for attendance, we present results from basic OLS regressions below in Table 3. Although these results may suffer from significant endogeneity bias, they allow us to gain more familiarity with the data and a sense of the relationship between attendance and performance.

As column (1) of Table 3 shows, there is a clear endogeneity bias. As $\frac{\text{Attendance}}{\text{Stadium Capacity}}$ is a variable going from 0 to 1, without any controls, the coefficient of 11.23 denotes that a 10 percentage point increase in attendance is associated with a 1.123 increase in score differential between Home and Away teams. As we noted in Table 2, the overall average Home - Away score differential is only about 3, so this coefficient is of both statistical and practical significance. Of course, this effect is most likely not due to the fans; rather, more fans are going because the Home Teams are more likely to win. However, notice that after adding in controls, as represented in columns (3) and (4), our result becomes statistically insignificant.

Table 3: Initial OLS Regressions. Dependent Variable: Home Score - Away Score. SE clustered on Home * Season.

VARIABLES	(1) Δ Scores	(2) Δ Scores	(3) Δ Scores	(4) Δ Scores
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	11.23***	-14.15***	-0.447	-0.109
	(1.998)	(1.616)	(1.670)	(1.668)
Home Team * Season FE	NO	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES
Month FE	NO	NO	NO	YES
Constant	-6.914***	10.69***	5.613**	5.894***
	(1.751)	(1.276)	(2.230)	(2.204)
Observations	8,366	8,366	8,366	8,366
R-squared	0.012	0.160	0.270	0.270

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

4.2 Selecting Instruments

Two instruments that we are considering are a dummy variable for weekend game and a dummy variable indicating presence of some form of adverse weather. Smith and Groetzinger (2010) use weekend game and temperature to instrument for attendance in Major League Baseball games. In addition, they found adverse weather to be an inappropriate instrument. While this may be the case in baseball, we expect that adverse weather and temperature to have a different effect on NBA attendance. This is because many Major League Baseball games are canceled if weather is significantly adverse, and because it is an outdoor sport, temperature may have more of an effect on attendance. However, as basketball is an indoor sport, we expect that temperature may not significantly affect attendance. Moreover, games will not be canceled due to adverse weather, but fans may be deterred from attending. Our first-stage results confirm this intuition.

Before conducting any instrumental variables (IV) regressions, we note that IV has limitations of its own. First, we must assume that the instruments we are using are uncorrelated with the error term. Second, the instruments must strongly predict attendance. While we can explicitly test the second issue, we must first address whether our instruments are uncorrelated with the error term.

In general, we believe that adverse weather should not be correlated to other unobservables that affect performance. While adverse weather is certainly more likely to occur in some cities rather than others, adding Home * Season Fixed Effects should account for this. However, we must address a potential issue with using Weekend Game as an instrument. For an ideal instrument, we would want the teams playing on weekends randomly distributed – with 30 teams, $\frac{1}{30}$ of all weekend games should be played at each team's home stadium. However, Table 4 shows that this might not be the case, and it seems that bigger market teams get more home games on weekends than smaller market teams.

Table 4: Proportion of Weekend Games by Teams and Market Size Rankings (2012-2013 NBA Season) Total Weekend Games: 622. Playoff teams marked in Blue.

Team Name	**Percentage of Weekend Home Games**	**Market Size Rankings (Measured by # TV Homes)**
1. New York Knicks	4.34%	1
2. Los Angeles Lakers	4.18%	2
3. Toronto Raptors	4.18%	N/A (Canada)
4. Washington Wizards	3.85%	11
5. Brooklyn Nets	3.70%	1 (Same as NYC)
26. Indiana Pacers	2.89%	23
27. Chicago Bulls	2.73%	5
28. Cleveland Cavaliers	2.73%	18
29. Oklahoma City Thunder	2.73%	27
30. Utah Jazz	2.25%	24
N	8369	

Market size rankings are taken from www.sportsmediawatch.com "NBA Market Size Number Game". As suggested earlier, Table 4 shows that it might be the case that scheduling for weekend games may not be entirely random. It seems that bigger market teams are getting more weekend home games than smaller market teams. While this may not inherently be a problem, if market size is correlated with team quality, then it may be the case that using a weekend dummy may not be suitable.

Thus, we are motivated to look at weekend game further to see whether it is uncorrelated with the error term. Indeed, Table 3 lists the top five teams receiving the most home weekend games and the bottom five teams receiving the least home weekend games during the 2012-2013 season. However, there are three playoff teams in each of these categories. From the top five teams, the New York Knicks, Los Angeles Lakers, and the Brooklyn Nets were playoff teams. From the bottom five teams, the Indiana Pacers, Chicago Bulls, and the Oklahoma City Thunder were also playoff teams.

Table 5 below further suggests that even though whether or not a team gets a home weekend game is perhaps correlated with market size, it may still be uncorrelated with other unobservables that determine on-court performance.

Average attendance as a percentage of stadium capacity is almost 4 percentage points higher on weekend games than weekday games. In addition, we find that on average, home teams win almost 1 percentage point more during weekend games relative to weekday games. However, looking at other observable characteristics, there seems to be very little difference between weekend and weekday games. For example, average home team end of season records are 50% for both weekend and weekday games. The corresponding statistic for away teams only differ by 1 percentage point. To put that in perspective, there are only 82 games in a normal NBA season – one additional win increases a team's end of season record by $1/82 = 1.22\%$. Thus, 1 percentage point is not even a one game impact on end of season records.

Similarly, all other observable characteristics seem to be almost equal for weekend games versus weekday games. We conclude that although weekend games may not be scheduled completely at random, we believe that it is still uncorrelated with the error term (other unobservables that may affect performance).

Table 5: Comparing mean statistics between Weekend and Weekday Games. $\frac{1}{82} = 1.22\%$

Variable	**Weekend Game**	**Weekday Game**	**Weekend - Weekday**
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	91.16%	87.04%	4.12%
Home Team Win	60.4%	59.5%	0.9%
Home Team End of Season Record	50%	50%	0
Away Team End of Season Record	49.5%	50.5%	-1%
Home Team End of Previous Season's Record	49.8%	50.2%	-0.4%
Away Team End of Previous Season's Record	49.7%	50.3%	-0.6%
Home Team Year-To-Date Record	49.7%	49.8%	-0.1%
Away Team Year-To-Date Record	49.6%	50.6%	-1%
Home Assists	22.02	22.37	-0.25
Away Assists	20.69	20.60	0.09
Home Total Rebounds	42.34	42.22	0.12
Away Total Rebounds	41.06	41.04	0.02
Home Free Throws Made	18.88	18.80	0.08
Away Free Throws Made	17.84	17.98	-0.14
Number of Observations	4141	4225	

We now test the second assumption that our instruments strongly predict attendance. A priori, although we have many different indicators of adverse weather, we believe that the ones which most strongly predict attendance are ones which significantly affect commuting – such as snow or heavy precipitation. Indeed, we find that snow has a statistically significant negative effect on attendance. In addition, presence of Heavy Fog also has a statistically significant negative effect. We run first-stage regressions to confirm our intuition, with the first-stage equation being:

$$\frac{\text{Attendance}}{\text{Stadium Capacity}} = \beta_0 + \beta_1 Instruments + \beta_2 Controls + \mu. \tag{2}$$

First stage results are shown in Table 6 below. Weekend games are clearly the strongest predictor of attendance. A game played on a weekend game is on average associated with 4.87 percentage points higher percent attendance (i.e if a given team's average attendance on weekdays were 80%, on a weekend, that team would have an average attendance of approximately 84.87%). Snow and heavy fog, although they statistically significantly predict attendance, both have a much smaller effect in magnitude. Snow and heavy fog are on average associated with approximately 0.789 and 0.938 percentage points lower percent attendance, respectively. An F-test on the instruments, shown at the bottom of column (4) produce an F-statistic of 70, largely driven by the Weekend Game variable. Other indicators of adverse weather such as rain, thunder, heavy winds, etc. prove to be insufficient predictors of attendance. Lastly, we use an over-identification test to help validate our instrument choice, and indeed we find that weekend game, snow, and heavy fog perform well in over-ID tests.

Thus, we run all IV regressions using weekend game, snow, and heavy fog as instruments for attendance.

Table 6: First Stage Regressions. Dependent Variable: $\frac{\text{Attendance}}{\text{Stadium Capacity}}$. SE clustered on Home * Season.

VARIABLES	(1) %Attendance	(2) %Attendance	(3) %Attendance	(4) %Attendance
Weekend Game	0.0488***			0.0487***
	(0.00340)			(0.00339)
Snow		-0.00959**		-0.00789**
		(0.00382)		(0.00367)
Heavy Fog			-0.0113***	-0.00938**
			(0.00402)	(0.00374)
Home Team * Season FE	YES	YES	YES	YES
Away Team * Season FE	YES	YES	YES	YES
Month FE	YES	YES	YES	YES
Observations	8,366	8,366	8,366	8,366
F-Stat	205.84	6.30	7.94	70.00
R-squared	0.688	0.653	0.653	0.8407

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

4.3 IV Regressions – Second Stage

We first run IV regressions using score differential (Home - Away Score) as our dependent variable. These results will tell us whether or not fans do indeed contribute to the home-court advantage in the NBA. The first set of IV regressions are shown below in Table 7. Each column in the table represents separate regressions with Home - Away score as the dependent variable. Fixed Effects which are included are indicated by "YES" or "NO". We find that regardless of the specification, we obtain insignificant results. Surprisingly, in Column (4), which includes all of our chosen fixed effects, higher percent attendance seems to be negatively associated with Home - Away score differential. As $\frac{\text{Attendance}}{\text{Stadium Capacity}}$ is a variable ranging from 0 to 1, the coefficient of -6.732 in Column (4) means that a 10 percentage point increase in percent attendance is associated with 0.6732 decrease in score differential. The reported 95% confidence interval for the same coefficient is (-17.467, 4.003). This means at the top of the 95% confidence interval, a 10 percentage point increase in percent attendance is only associated with a 0.4 increase in score differential. At the bottom end, a 10 percentage point increase in percent attendance is associated with a 1.7467 decrease in score differential. Both of these estimates imply an impact of less than one scored basket in the overall outcome. However, as Table 2 reports, the overall average difference between Home and Away scores is only about three points. Thus, a 1.7467 score differential, if valid, would be of strong practical significance. Furthermore, we report the OverID and Hausman statistic for the specification model including all the fixed effects. This is shown in column (4) of Table 7 below. We notice that the Over ID statistic of 0.2942 (p-value = 0.8632) is not statistically significant, which provides further confidence in the validity of our instruments. However, the Hausman statistic of 1.7949 (p-value = 0.1804) is also not significant at any conventional levels. While this means we fail to reject exogeneity, we certainly believe that attendance is not exogenous. In this case, it may be that the standard errors are simply too big in order to reject exogeneity.

Table 8 shows results of similar regressions. The only difference is that the dependent variable, Home - Away score differential, is bounded at an absolute value of 6. We did this just to check whether some games with large score differentials were driving our results. For example, if the Home - Away score differential was -8, we manually changed it to -6. Similarly, if it was +10, we manually changed it to +6. As expected, our numerical results are different from Table 6 – because score differentials are manually bounded. However, we again see no statistically significant results.

4.4 Further IV Regressions: Score Differential

Although we fail to initially find any significant impact of attendance on score differential, we run additional regressions controlling for other characteristics. Although we have added Home * Season and Away * Season fixed effects, there might also be within team and season characteristics that are biasing are results. Table 9 displays results with additional controls for teams' year-to-date record and teams' current win/loss streak. For these regressions, we must exclude any games that are the first for any team in a given season. For the first game of the season, there is neither a valid year-to-date record nor win/loss streak.

Column (1) of Table 9 below is the same as column (4) of Table 7. This is just there for comparison purposes. However, notice that as we add controls for within-season characteristics, the slope coefficient of $\frac{\text{Attendance}}{\text{Stadium Capacity}}$ virtually remains the same. While we still do not obtain any statistically significant slope coefficients, this is still somewhat encouraging to see these results as it further validates our choice of instruments.

We now examine the possibility of a non-constant effect of attendance on score differential. The motivation for this is that perhaps for big-market teams, fans are likely to go to games no matter what the conditions are. Thus, there is not much variation in fan attendance regardless of game characteristics. However, for teams where average attendance across the season was much lower, then a significant change in attendance on one game may have more of an impact on performance. For example, a big-market team like the New York Knicks might have a sell-out crowd almost every game. However, for a smaller market team, such as the Indiana Pacers where the average attendance might be lower, a sharp change in attendance for one particular game might have more of an effect. We capture this by running similar regressions, but restricting our observations where average attendance for a team in a given season falls in a certain range. Table 10 below shows these regressions with each column indicating what range of average attendance as a percent of stadium capacity we are including. Again, we do not find any slope coefficients to be statistically significant. In fact, numerically, the slope coefficients have increased in magnitude, but standard errors have also increase as well leading to extremely large 95% confidence intervals.

Table 7: Initial IV Regressions. Dependent Variable: Home Score - Away Score. Instruments are: Weekend, Snow, Heavy Fog. SE clustered on Home * Season.

VARIABLES	(1)	(2)	(3)	(4)
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	3.322	-1.646	-6.710	-6.732
	(6.464)	(6.206)	(5.594)	(5.477)
Home Team * Season FE	NO	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES
Month	NO	NO	NO	YES
Constant	0.133	0.812	10.31**	10.78**
	(5.754)	(4.902)	(4.515)	(4.370)
Observations	8,366	8,366	8,366	8,366
Over ID Stat				0.2942
Hausman Stat				1.7949

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

Table 8: IV Regression. Dependent Variable: Adjusted Home-Away Score Differential. Differential is capped at maximum of 6 points. Instruments are: Weekend, Snow, Heavy Fog. SE clustered on Home * Season.

VARIABLES	(1)	(2)	(3)	(4)
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	3.596	1.253	-0.721	-0.663
	(2.442)	(2.410)	(2.155)	(2.088)
Home Team * Season FE	NO	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES
Month	NO	NO	NO	YES
Constant	-2.044	-1.429	1.498	1.536
	(2.176)	(2.086)	(1.948)	(1.892)
Observations	8,364	8,364	8,364	8,364

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

Table 9: IV Regression. Dependent Variable: Home-Away Score Differential. Instruments are: Weekend, Snow, Heavy Fog. SE clustered on Home * Season. Additional controls, Year-To-Date Record and Win Streak account for within season characteristics.

VARIABLES	(1)	(2)	(3)	(4)
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	-6.732	-6.895	-7.066	-6.903
	(5.477)	(5.449)	(5.478)	(5.459)
Home Year To Date Record	NO	YES	NO	YES
Away Year To Date Record	NO	YES	NO	YES
Home Win Streak	NO	NO	YES	YES
Away Win Streak	NO	NO	YES	YES
Home Team * Season FE	YES	YES	YES	YES
Away Team * Season FE	YES	YES	YES	YES
Month	YES	YES	YES	YES
Observations	8,366	8,240	8,240	8,240

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

Table 10: IV Regression. Dependent Variable: Home-Away Score Differential. Instruments are: Weekend, Snow, Heavy Fog. SE clustered on Home * Season. In each regression, we restricted the observations to those where the average attendance for the home team in the given season fell in a certain range. These ranges are indicated at the top of each column.

IV Regression. Dependent Variable: Home Score - Away Score

VARIABLES	(1) [0.9, 1)	(2) [0.8, 0.9)	(3) [0.0, 0.8)
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	-37.36	-12.36	-2.080
	(23.68)	(17.52)	(11.72)
Home Team * Season FE	YES	YES	YES
Away Team * Season FE	YES	YES	YES
Month	YES	YES	YES
Constant	19.77	-0.551	16.35
	(31.04)	(16.20)	(16.92)
Observations	3,552	1,714	911

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

4.5 IV Regressions: Home Win

We now examine the effect of fan attendance on home wins. We use an IV Probit regression to estimate the effect of fan attendance on the probability of the home team winning. Unfortunately, computational limitations do not allow us to include Home * Season and Away * Season fixed effects at the same time. However, since we believe our instruments are valid, we do not believe that our slope coefficients would change significantly if we were able to add both fixed effects into our model. Table 10 displays the results of these IV Probit regressions. As these are probit regressions, and the magnitudes of the slopes are difficult to interpret, we also include IV OLS regressions below with each column corresponding to the same specification. Although the functional form for the bottom table may not be correct, it gives an idea for the interpretation of the magnitude of the slopes presented in the probit regressions.

We find that the slope coefficients under the probit regressions are not statistically significant either. However, it is encouraging to see that the signs of the slope coefficients are positive, which is what we expected prior to conducting these analyses. In addition, the two stage least squares estimates, although statistically insignificant, yield results that are practically feasible. For example, in Column (5), a slope coefficient of 0.199 denotes that increasing percent attendance by 10 percentage points is associated, on average, with a 1.99 percentage point increase in the probability of the home team winning.

Table 11: IV Probit Regression. Dependent Variable: Dummy for Home Win. Instruments are: Weekend, Snow, Heavy Fog. SE Clustered on Home * Season.

VARIABLES	(1) Home Win	(2) Home Win	(3) Home Win	(4) Home Win	(5) Home Win
$\frac{\text{Attendance}}{\text{StadiumCapacity}}$	1.099	0.6439	0.5681	0.5998	0.5396
	(0.7574)	(0.7046)	(0.6860)	(0.7557)	(0.7446)
Home Team * Season FE	NO	YES	YES	NO	NO
Away Team * Season FE	NO	NO	NO	YES	YES
Month	NO	NO	YES	NO	YES
Observations	8,366	8,366	8,366	8,366	8,366

IV Two Stage Least Squares Regression. Dependent Variable: Dummy for Home Win

VARIABLES	(1) Home Win	(2) Home Win	(3) Home Win	(4) Home Win	(5) Home Win
$\frac{\text{Attendance}}{\text{StadiumCapacity}}$	0.427*	0.174	0.152	0.223	0.199
	(0.247)	(0.241)	(0.234)	(0.226)	(0.223)
Home Team * Season FE	NO	YES	YES	NO	NO
Away Team * Season FE	NO	NO	NO	YES	YES
Month	NO	NO	YES	NO	YES
Observations	8,366	8,366	8,366	8,366	8,366

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

4.6 IV Regressions: Free Throw Percentage and Other Metrics

Although we fail to find any statistically significant result of fan attendance on score differential or home wins, we examine the possibility of fan attendance affecting other aspects of performance. We believe that of all the available box score metrics, free throw percentage is the metric where we can best isolate the effect of fans on performance. Almost any other event in basketball: field goal made, rebound, assist, turnover, etc. are not only influenced by the crowd but also the players and referees. Thus, for most of these other metrics, it is difficult to isolate the effect of the crowd on performance. However, during a free throw, the shooter is left completely unguarded and thus there is no physical interaction between the shooter and any other player or referee. Since interaction between the shooter and other players or referees is minimal, we believe we can best observe the fans' effects during a free throw. Table 12 below displays results for regressions similar to the ones shown previously, except the dependent variable is now free throw percentage. The first table shows results with Away Free Throw Percentage as the dependent variable. The second and third table show results with Home Free Throw Percentage and the difference between the Home and Away Free Throw Percentage as the dependent variable, respectively.

Note that the coefficient (and standard errors) on percent attendance remains virtually the same within each of these tables. This is expected as we believed that free throw percentage should be less noisy than overall score or home wins. Furthermore, percent attendance is statistically significant at the 1% level with away free throw percentage as the dependent variable. In Column (4) of Away Free Throw percentage table, the coefficient on percent attendance is -0.110, meaning that on average, a 10 percentage point increase in percent attendance is associated with a 1.1 percentage point decrease in away free throw percentage. Although this is highly statistically significant, we note that this is of very little practical significance. A 1.1 percentage point decrease in free throw percentage would unlikely change the outcome of a NBA game. This is consistent with our earlier findings that there does not seem to be a statistically significant effect of fan attendance on overall game performance. In addition, the instruments perform well in an Over-ID test, with a $\chi^2(2) = 2.0932$ (p-value =0.3511). Furthermore, the Hausman statistic is 4.9432 (p-value = 0.0262). Thus, unlike with overall game performance as the dependent variable, we can reject exogeneity at the 5% significance level. This is because as free throw percentage is a much tighter metric, our standard errors are much more confined.

Although percent attendance statistically significantly affects away free throw percentage, we observe that it does not statistically significantly affect home free throw percentage. If anything, the effect is negative. This might be because more fans may still add some pressure to the players of the home team, although not as much as on away team. Lastly, we note that the effect of percent attendance on the difference in free throw percentage between home and away teams is statistically insignificant.

Finally, we run similar regressions on various performance metrics. The results are shown in Table 13. In this table, all regressions are run with Home * Season, Away * Season, and Month Fixed Effects. The coefficients on percent attendance are reported. The labels on the left most column indicate what the dependent variable of that regression is. The first column contains slope coefficients on regressions where the dependent variable is only for the Away Team, the second column contains slope coefficients on regressions where the dependent variable is only for the Home Team, and the third column contains slope coefficients on regressions where the dependent variable is for the difference between the Home and Away teams.

For example, for the row labeled "Free Throw Attempts" in Table 13, the regression reported in this row and column (1) estimates similar models as we have shown previously, but with Free Throw Attempts for the Away team as the dependent variable. Column (2) has Free Throw Attempts for the Home team as the dependent variable, and column (3) has the difference between the Home and Away team as the dependent variable. In addition, column (1) of row "Free Throw Attempts" is -0.935. This denotes that a 10 percentage point increase in percent attendance is associated, on average, with the Away Team receiving 0.0935 less free throw attempts. Similarly, in column (2) of the same row, a coefficient of 3.384 denotes that a 10 percentage point increase in percent attendance is associated, on average, with the Home Team receiving 0.3384 free throw attempts more. Although these slope estimates are not only statistically insignificant, but also arguably practically insignificant, it is interesting to note that Home Teams' number free throw attempts are found to be positively correlated with attendance while Away Teams' number free throw attempts are found to be negatively correlated with attendance. If nothing else, this may serve as further motivation to investigate the possibility of referee bias, especially in the context of fan attendance.

Furthermore, Table 14 shows results from additional regressions run using the Home Point Spread as the dependent variable. These coefficients tell us the impact of attendance as percent of stadium capacity on the point spread.

The point spread serves as a handicap to be placed on one team, purely for betting purposes. It is calculated such that a bettor should have equal chance of winning the bet regardless of which team he picks. A point spread on a team reflects how much of a favorite that team is to win. For example, suppose the New York Knicks are playing the Miami Heat, and the New York Knicks are the home team. Suppose the point spread on the Knicks (in this case the Home Point Spread) is +3. Perhaps unintuitively, this means that the Knicks are expected to lose by 3 points[4].

Table 14 shows that even after adding all controls (as shown in column (5), a 10 percentage point increase in percent attendance is associated with 0.2391 increase in the Home Point Spread – bettors expect that the Home - Away score differential will be 0.2391 less. Although this direction is surprising and statistically significant, it is consistent with our earlier results that the effect of percent attendance on actual Home - Away score differential may be negative.

[4]The point spread is designed in this manner purely for betting purposes. In this example, if one were to place a bet on the Knicks, as long as the Knicks lose by less than 2 points (or wins the game outright), one would win that bet.

Table 12: IV Regression. Dependent Variable: Away Free Throw Percentage. Instruments are: Weekend, Snow, Heavy Fog. SE clustered on Home * Season. AFTP: Away Free Throw Percentage, HFTP: Home Free Throw Percentage, DHAFTP: Difference between Home and Away Free Throw Percentage.

VARIABLES	(1) AFTP	(2) AFTP	(3) AFTP	(4) AFTP
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	-0.119**	-0.0971**	-0.118***	-0.110***
	(0.0471)	(0.0487)	(0.0447)	(0.0431)
Home Team * Season FE	NO	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES
Month	NO	NO	NO	YES
Constant	0.864***	0.819***	0.838***	0.829***
	(0.0420)	(0.0415)	(0.0388)	(0.0373)
Observations	8,366	8,366	8,366	8,366
Over ID Stat				2.0932
Hausman Stat				4.9432

VARIABLES	(1) HFTP	(2) HFTP	(3) HFTP	(4) HFTP
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	-0.0267	-0.0665	-0.0590	-0.0593
	(0.0457)	(0.0462)	(0.0436)	(0.0423)
Home Team * Season FE	NO	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES
Month	NO	NO	NO	YES
Constant	0.783***	0.825***	0.792***	0.793***
	(0.0407)	(0.0397)	(0.0395)	(0.0383)
Observations	8,366	8,366	8,366	8,366
Over ID Stat				0.636
Hausman Stat				1.924

VARIABLES	(1) DHAFTP	(2) DHAFTP	(3) DHAFTP	(4) DHAFTP
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	0.0922	0.0306	0.0589	0.0508
	(0.0647)	(0.0667)	(0.0620)	(0.0600)
Home Team * Season FE	NO	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES
Month	NO	NO	NO	YES
Constant	-0.0811	0.00614	-0.0467	-0.0362
	(0.0576)	(0.0574)	(0.0553)	(0.0535)
Observations	8,366	8,366	8,366	8,366
Over ID Stat				1.318
Hausman Stat				0.381

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

Table 13: IV Regressions comparing Away, Home, and Home - Away. $\frac{\text{Attendance}}{\text{Stadium Capacity}}$ coefficient is reported for each regression. All regressions include Home * Season FE, Away * Season FE,and Month FE. Standard errors clustered on Home * Season. Instruments are: Weekend, Snow, Heavy Fog. Total Number of Obs: 8366.

	(1)	(2)	(3)
VARIABLES	Away	Home	Home - Away
Free Throw Percentage	-0.110***	-0.0593	0.0508
	(0.0404)	(0.0469)	(0.0596)
Free Throw Attempts	-0.935	3.384	4.319
	(2.872)	(3.037)	(4.180)
Three Point Percentage	-0.0496	-0.0370	0.0126
	(0.0522)	(0.0559)	(0.0722)
Rebounds	0.731	2.941	2.210
	(2.600)	(2.794)	(3.690)

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

	(1)	(2)	(3)
VARIABLES	Away	Home	Home - Away
Turnovers	-1.360	2.227	3.586*
	(1.564)	(1.486)	(1.981)
Offensive Rebounds	0.0601	0.785	0.724
	(1.690)	(1.746)	(2.592)
Blocks	0.741	-0.338	-1.079
	(0.963)	(1.033)	(1.446)
4th Quarter PTS	-3.041	-3.607	-0.566
	(2.691)	(2.575)	(3.417)

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

Table 14: IV Regressions with the Home Point Spread as the dependent variable. Instruments are: Weekend Game, Snow, and Heavy Fog. SE clustered on Home * Season.

VARIABLES	(1)	(2)	(3)	(4)	(5)
$\frac{\text{Attendance}}{\text{Stadium Capacity}}$	-8.276***	-3.346	2.271**	2.366**	2.391***
	(3.165)	(2.272)	(0.959)	(0.933)	(0.907)
Home Team * Season FE	NO	YES	YES	YES	YES
Away Team * Season FE	NO	NO	YES	YES	YES
Month	NO	NO	YES	YES	YES
Home Year To Date Record	NO	NO	NO	YES	YES
Away Year To Date Record	NO	NO	NO	YES	YES
Home Win Streak	NO	NO	NO	NO	YES
Away Win Streak	NO	NO	NO	NO	YES
Constant	4.001	3.826**	-4.292***	-4.155***	-4.153***
	(2.804)	(1.795)	(0.792)	(0.779)	(0.786)
Observations	8,366	8,366	8,366	8,240	8,240
R-squared	0.040	0.409	0.829	0.846	0.851
Over ID Stat					0.44167
Hausman Stat					10.2758

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

5 Conclusion

Ultimately, we find very little evidence that fan attendance has a significant impact on overall performance in a NBA game. Indeed, we fail to find any statistically significant result when looking at score differential or home wins as the dependent variable. Even after examining score differential through various perspectives and controlling for within season characteristics, we are unable to find a statistically significant effect. In addition, after running regressions using the Home Point Spread as the dependent variable, it seems that the market's expectation of the impact of attendance is also negative for the home team as well.

Contrasting this study with that of Smith and Groetzinger, we see that while the NBA enjoys a larger home team advantage overall, we fail to find that it is attributable to fan attendance. On the other hand, the MLB seems to have a lesser home team advantage in aggregate, but Smith and Groetzinger find a significant effect of attendance on performance. One key difference is that MLB has 81 home games per team, while the NBA only has 41. Thus, Smith and Groetzinger identify more variation in attendance through their attendance than this study does. Indeed, Smith and Groetzinger report that a weekend game in the MLB is associated with a 13.3 percentage point increase in attendance as a percent of stadium capacity. In the NBA, we find that a weekend game is only associated with a 4.87 percentage point increase in percent attendance (Table 5 Column (4)).

Although we fail to show that fans have an effect on overall performance, it may be because overall score differential and home wins are too noisy of a variable. There may be unobservables that are unaccounted for that are biasing out results – such as referee bias. However, we do observe a statistically significant effect of fans on free throw percentage. Almost every other metric suffers from a myriad of confounding variables, thus making it very difficult to isolate the effect of fans. However, as the free throw is an isolated event, it is much easier to observe the effect of fans on free throws. Furthermore, free throw percentage is much more consistent of a metric than score differential. This perhaps motivates a different approach to answer the question we initially posed. Do fans affect free throw performance? Based on both our intuition and the results we present here, it is plausible that the free throw is where the fans have the most effect within an NBA game. Utilizing game-level data by individual player as opposed to team would help us more accurately determine the psychological effect of fans on free throw performance. In addition, further data work regarding referee calls may lead to opportunities to investigate referee bias for the home team, and specifically how it relates to fan attendance. Lastly, we note that the statistically significant result we observed on away free throw percentage should be viewed with caution, especially in light of the number of regressions we have presented and it being the only one showing any statistical significance.

References

[1] Courneya, K. and Albert V. Carron (1991). Effects of Travel and Length of Home Stand/Road Trip on the Home Advantage *Journal of Sports & Exercise Psychology,* 13, 42-49.

[2] Entine, O. and Dylan S. Small (2008). The Role of Rest in the NBA Home-Court Advantage *Journal of Quantitative Analysis in Sports,* Vol. 4: Iss. 2, Article 6.

[3] Moskowitz, Tobias J and L. Jon Wertheim. (2011) *Scorecasting: The Hidden Influences Behind How Sports Are Played and Games are Won.* New York, NY: Crown Archetype.

[4] Pace, Andres and Albert V. Carron (1992). Travel and the Home Advantage *Canadian Journal of Sport Science,* 17, 60-64.

[5] Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh (2011). Strike Three: Discrimination, Incentives, and Evaluation. *American Economic Review,* 101 (4), 1410-35.

[6] Price, J. and Justin Wolfers (2010). Racial Discrimination Among NBA Referees *The Quarterly Journal of Economics,* 125 (4), 1859-1887. Relations.

[7] Price, J., Marc Remer, and Daniel F. Stone (2012). Subperfect Game: Profitable Biases of NBA Referees *Journal of Economics & Management Strategy,* 21 (1), 271-300.

[8] Smith, Erin E., and Jon D. Groetzinger (2010). Do Fans Matter? The Effect of Attendance on the Outcomes of Major League Baseball Games *Journal of Quantitative Analysis in Sports,* Vol. 6: Iss. 1, Article 4.