An Analysis of NBA Home Court Advantage

Simon Fettig, Megan Guldan, Michael Law, Ya'acov Ritov

Department of Statistics, University of Michigan

I. Abstract

On-court player performance and officiating decisions are predominant effects of a potential home court advantage in the National Basketball Association (NBA). To quantify these impacts, we use the 6150 games worth of data from the 2014-2018 NBA regular seasons. Using paired t-tests to compare home and away performance, we find significant statistical increases in multiple offensive metrics accompanied by playing at home, as well as foul calls favoring the home team. To analyze the impact of the crowd, we look particularly at attendance as a proportion of the home arena's maximum capacity, a concept we found under-explored in other research. Utilizing linear mixed-effects regression models to control for team strength from year to year, we perform model comparison with ANOVA to determine the impact of attendance on various facets of the game. This process provides us with evidence that attendance is significantly linked to home court advantage with respect to scoring and shooting performance.

Keywords: NBA, home court advantage, attendance, bias, basketball

II. Introduction and Literature Review

The existence of a home court advantage in basketball is frequently discussed by fans and the sports media alike. As a result, much of the research surrounding the theory addresses its impacts on different aspects of the game. Many studies examine the quantitative impact of attendance on a home-court advantage. A 2015 study isolated this impact by analyzing games played at the LA Staples Center, which hosts both the Lakers and the Clippers of the NBA, to account for court familiarity and travel differences (Boudreaux et al., 2015). This study found the effects of a sympathetic home crowd to be significant, which mirrors results found in other studies that claim crowd support can influence the game in many ways, such as impacting referee decisions in favor of the home team (Ponzo & Scoppa 2016).

Additional studies have quantified the effect of officiating bias in contributing to home court advantage. A 2009 study examined patterns of foul calls in the NCAA, determining a significant bias in the number of fouls called on the away team (Anderson & Pierce, 2009). Although this study acknowledged the general trend in basketball where the number of fouls called increases as the game progresses, it did not account for intentional fouling at the end of the game in its analysis.

More recent studies have used data from games played during the COVID-19 pandemic to remove many of the factors that contribute to a home-court advantage, such as the presence of a crowd. One study in particular analyzed data from the 2020 NBA playoffs, where games were played in a “bubble” that featured playing conditions without any fans or need for the teams to travel (Price & Yan, 2021). In this paper, Price and Yan performed a two-sample comparison using a z test (with an assumption of normality based on the central limit theorem) along with Fisher’s exact test and Wilcoxan’s rank-sum test as distribution free tests that analyzed differences in factors such as winning percentages between data from 2017-2019 and data in 2020’s COVID bubble. They found that, with the audience and travel impacts that tend to make up a home court advantage removed, the designated away teams in 2020 actually performed better than the home teams compared to previous years that took place under more normal circumstances.

In this paper, we focus on games played under normal conditions before the COVID-19 pandemic. First, we compare home team and away team relative performance with regards to eight factors - win rate, field goal percentage, three-point field goal percentage, free throw attempts, free throw percentage, points, personal fouls, and technical fouls - to quantify the impacts of home court advantage. Next, we analyze the impact of attendance on these eight key metrics. In particular, we use attendance proportions as opposed to total attendance to more accurately measure the energy of a crowd and its impact on performance. For each of the factors listed above, we compare a “baseline” mixed-effects model - one that accounts only for the strength of a team in the given year - to a model that also incorporates attendance proportion. This process allows to determine which box-score stats see a statistically significant home team improvement associated with greater attendance.

III. Data Collection Methodology

In order to dissect the influence of attendance in particular, we focus our analysis on games that occurred prior to the COVID-19 pandemic and the 2020 “bubble” that occurred thereafter. To gather data, we utilized the open-source R package *nbastatR*, which is maintained by Alex Bresler and retrieves data from the NBA Stats API. Functions within this package allowed us to access game logs for every regular season game from the 2014-2018 seasons (2014 represents the 2014-15 season, etc), as well as play-by-play data for each game within this same range. To supplement this dataset, we manually scraped information from basketball-reference.com to obtain game-by-game attendance figures.

Merging our data from these two sources formed an all-encompassing dataset for 5 seasons worth of NBA regular season games. Since we were particularly interested in comparing the performance of home teams against the away team, we adjusted the dataset to highlight the differences (in the form of home team - away team) with respect to key box score statistics available in the game logs, a full list of which can be found in Table 1.

In addition to these box score figures, we want to look at how other metrics were impacted by a potential home court advantage. In particular, we want to adjust the way that fouls are counted before looking for a home court officiating bias. Basketball teams often intentionally foul when they are losing towards the end of a game in order to prevent the other team from running out the clock; these foul calls lack the subjectivity and ability to be impacted by officiating bias that exists in a traditional foul call, as the intent to purposely commit a foul is obvious to the referees. To account for this, we filtered out personal fouls that occurred in the final two minutes of the fourth quarter or any overtime periods, creating a variable we named "valid fouls". Furthermore, while filtering through the play-by-play data, we are able to count the number of technical fouls committed by each team, which is kept separate from the number of valid personal fouls.

Another element we consider during data preprocessing is the distinguishment of attendance proportion versus raw attendance figures. We theorize that a full arena, with no gaps due to open seats, leads to a louder and more energetic crowd - and thus a more challenging environment for the visiting team to play in. To account for this, we calculate the attendance proportion for each game based on maximum capacity figures for the home team's arena at the time, and utilize this new variable in our forthcoming regression model analyses.

IV. Quantification of Home Court Advantage

Overall, we see that home teams performed better with regards to all box-score statistics we analyzed during the 2014-2018 regular seasons. We quantify this by computing the percent changes that accompanied playing at home - which we calculate by dividing the average difference (as described in Section II) by the overall averages for each metric (See Table 2). To determine the significance of these improvements, we perform paired t-tests on 8 of our variables of interest (See Table 3) using base R's *t.test* function. Using the Bonferonni correction to account for the multiple comparisons performed in our analysis, we obtain the following significance threshold:

$$\alpha = 0.05/16 = 0.003125$$

On average, home teams performed better in both aspects of field goal shooting efficiency, as seen by the statistically significant 2.25% increase in field goal percentage and 2.54% increase in three-point field goal percentage. As a result of these better shooting numbers, home teams scored an average of 2.48% more points per game over this 5 season period. These findings, in addition to the 58.4% overall win rate for home teams, confirm the notion that NBA teams generally perform better in their home arenas.

Additionally, we see that home teams have 3.1% fewer personal fouls called on them per game, as well as 3.03% fewer valid fouls (as defined in Section II), confirming the existence of a home-court advantage with regards to fouls and suggesting a possible officiating bias. This slight decrease in potential officiating bias when looking at valid fouls makes sense due to the nature of its calculation, and justifies our methodology - since home teams win more games on average, away teams commit more fouls while trailing late in games in order to stop the clock and regain possession. Although the difference in effect size appears slight, we see that excluding fouls with under 2 minutes remaining provides a more accurate quantification of home court advantage with respect to fouls.

V. Modeling Methodology and Results

To further analyze the impact of attendance proportion on home court advantage, we perform mixed-effects linear regression modeling in R utilizing the *lmer* function from the *lme4* package. These mixed-effects models combine and simultaneously incorporate fixed and random effects, where fixed effects are the general, population-level effects that persist across a full dataset, and random effects account for the fact that observations in certain groups differ from the overall average (Brown 2021).

In the context of our research, we set attendance proportion as a fixed effect, while we also want to control for the skill of both the home team and their opponent. To achieve this, we treat the intercept term in our regression models as a random effect and adjust it based on interaction terms between the season and both the home and away teams; this accounts for the fact that the annual "version" of each team (the 2015 Golden State Warriors, for example) has a certain talent level. Adjusting the intercept based on team strength allows us to more accurately determine if attendance proportion has a significant effect on our respective response variables, which include home team win rate, various measures of shooting efficiency, and fouling counts.

For each of our response variables (refer to Table 4), we perform a two-step modeling process to determine the significance of attendance proportion as a predictor. First, we build a baseline model, with only an intercept term and its corresponding adjustment for team strength described above.

Then, we add attendance proportion to the models as a fixed effect, giving us 8 models of the following form:

$$y_{ijkn} = \beta_0 + \beta_1 x_i + C_{jn} + D_{kn} + \epsilon_{ijkn}$$

where y_{ijkn} is the predicted response, β_0 is the intercept estimate, β_1 is the regression coefficient estimate for attendance proportion, x_i is the attendance proportion for game i, C_{jn} is the intercept adjustment for home team j in season n, D_{kn} is the intercept adjustment for away team k in season n, and ϵ_{ijkn} is the error term.

Next, we compare these to the baseline models with ANOVA tests. This comparison reveals if incorporating attendance into the model yielded a statistically significant improvement with respect to predicting each response variable. Results for each of these analyses can be found below in Table 4. Once again, we utilize the significance threshold calculated from the Bonferonni correction ($\alpha = 0.003125$). This process presents us with the following overarching takeaways:

1. Attendance proportion is significant when it comes to predicting shooting metrics (FG% and 3FG%) as well as points scored, suggesting that it has a strong association with on-court performance. As one would expect, the attendance proportion coefficient estimates are positive for each of these responses, indicating that a larger crowd accompanies better relative performance from the home team.
2. Despite the significant effect linked to shooting performance, attendance proportion is slightly insignificant with respect to home team win rate.
3. Attendance proportion is insignificant with regards to officiating-related statistics (Valid fouls, free throw attempts, and technical fouls), suggesting that while a home court officiating bias seems to exist (see section IV), it is not directly influenced by the presence of a larger crowd.

VI. Conclusion and Discussion

In this report, we have seen that in addition to winning 58.3% of games during the 2014-2018 NBA regular seasons, home teams performed significantly better in terms of various shooting metrics and benefited from a home court advantage with regards to receiving more foul calls. Possible explanations for the discrepancy in fouls certainly include an officiating bias, but other reasonings persist, such as the fact that away teams may play more aggressively (and in turn, commit more fouls) to establish a physical presence while in challenging environments away from their home court.

When looking at the impact of attendance proportion on these metrics, we saw that its influence on officiating and home team win rate were insignificant over our 6150 game sample. However, we saw that larger crowds are linked to significant improvements in points scored - as well as field goal percentage and three-point field goal percentage - for home teams relative to their opponents. It is important to note that we cannot assign causality to attendance proportion in this context; attendance is likely correlated with a team's ability, as teams who are bad frequently attract fewer fans in the first place. While we can not isolate the direct impact of attendance on player performance, these associations remain a key takeaway of our research.

Having performed this process, we hypothesized ways to perform further statistical analysis related to home court advantage. Ideas include increased filtering of the play-by-play data to calculate valid free throw attempts (a parallel to valid personal fouls, described in Section II) in order to take a more in-depth look at a possible officiating bias, as well as searching for additional data, which would potentially allow for modeling with other predictors (travel/rest days, for example).

The existence of home-court advantage carries implications off the basketball court as well. As sports gambling becomes increasingly prevalent in the United States, home-court advantage and a possible officiating bias can cause large amounts of money to shift hands, potentially having large effects on people's livelihoods. Additionally, the door is opened for further pyschological research into the phenomenon that people may perform better when more comfortable with their environment, as potentially suggested by the link between higher shooting percentages and playing at one's home court.

Home-court advantage remains a fascinating concept for sports fans and researchers alike as sports continue to evolve and technology progresses. Due to the subjective nature of foul calls and human aspect of officiating, a possible refereeing bias in basketball is difficult to remove, but could potentially be reduced with a thorough re-training process. The potential addition of automated home plate umpires in Major League Baseball, however, could create exciting possibilities for analogous research regarding officiating and home team advantage in the future.

References

Anderson, K. J., & Pierce, D. A. (2009). Officiating bias: the effect of foul differential on foul calls in NCAA basketball. Journal of sports sciences, 27(7), 687–694. https://doi.org/10.1080/02640410902729733

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. https://doi.org/10.1177/1527002515595842

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. Advances in Methods and Practices in Psychological Science. https://doi.org/10.1177/2515245920960351

Ponzo, M., & Scoppa, V. (2018). Does the Home Advantage Depend on Crowd Support? Evidence From Same-Stadium Derbies. Journal of Sports Economics, 19(4), 562–582. https://doi.org/10.1177/1527002516665794

Price, Michael. (2021). "The Effects of the NBA COVID Bubble on the NBA Playoffs: A Case Study for Home-Court Advantage". Honors Scholar Theses. 823. https://opencommons.uconn.edu/srhonors_theses/823

Table 1 *Definitions of Statistics*

Abbreviation	Definition of Statistic
attProportion	Number of People in Attendance / Arena's Maximum Crowd Capacity
difFGPer	Difference (Home - Away) in Field Goal Percentage
dif3FGPer	Difference (Home - Away) in 3-point Field Goal Percentage
difFTAtt	Difference (Home - Away) in Free Throw Attempts
difFTPer	Difference (Home - Away) in Free Throw Percentage
difPts	Difference (Home - Away) in Points Scored
difFouls	Difference (Home - Away) in Personal Fouls
difValidFouls	Difference (Home - Away) in Valid Personal Fouls
difTechFouls	Difference (Home - Away) in Technical Fouls
homeWin	Home Team Win Indicator (1 if win, 0 if loss)

Table 2 *Numerical summaries of box score statistics*

Variable Name	Average Difference	Average % Difference	Standard Error
difFGPer	0.0103229	2.2546299	0.0009592
dif3FGPer	0.0089953	2.5387210	0.0017370
difFTAtt	0.8170732	3.5816646	0.1183192
difFTPer	0.0016494	0.2095994	0.0018516
difPts	2.6121950	2.4839379	0.1749080
difFouls	-0.6273171	-3.1011766	0.0648411
difValidFouls	-0.5692683	-3.0342931	0.0610939
difTechFouls	0.0138211	3.0644521	0.0120925

Table 3 *Home vs. Away performance T-test results*

Variable Name	p-value	Home Improvement Significant?
difFGPer	< 2.2 x 10^-16	Yes
dif3FGPer	2.305 x 10^-7	Yes
difFTAtt	5.499 x 10^-12	Yes
difFTPer	0.3731	No
difPts	< 2.2 x 10^-16	Yes
difFouls	< 2.2 x 10^-16	Yes
difValidFouls	< 2.2 x 10^-16	Yes
difTechFouls	0.2531	No

Table 4 *Model Comparison ANOVA Results*

Variable Name	p-value	Attendance Proportion Significant?
homeWin	0.01672	No
difFGPer	0.0003067	Yes
dif3FGPer	0.00236	Yes
difFTAtt	0.9708	No
difFTPer	0.009656	No
difPts	0.003082	Yes
difValidFouls	0.9379	No
difTechFouls	0.176	No