Home Sweet Home: Quantifying Home Court Advantages For NCAA Basketball Statistics

Matthew van Bommel Luke Bornn Peter Chow-White Chuancong Gao

Simon Fraser University

Abstract

Box score statistics are the baseline measures of performance for National Collegiate Athletic Association (NCAA) basketball. Between the 2011-2012 and 2015-2016 seasons, NCAA teams performed better at home compared to on the road in nearly all box score statistics across both genders and all three divisions. Using box score data from over 100,000 games spanning the three divisions for both women and men, we examine the factors underlying this discrepancy. The prevalence of neutral location games in the NCAA provides an additional angle through which to examine the gaps in box score statistic performance, which we believe has been underutilized in existing literature. We also estimate a regression model to quantify the home court advantages for box score statistics and compare the magnitudes to other factors such as increased number of possessions, and team strength. Additionally, we examine the biases of scorekeepers and referees. We present evidence that scorekeepers tend to have greater home team biases when observing men compared to women, higher divisions compared to lower divisions, and stronger teams compared to weaker teams. Finally, we present statistically significant results indicating referee decisions are impacted by attendance, with larger crowds resulting in greater bias in favor of the home team.

Keywords: Basketball, Scorekeeper, Referee, Bias, NCAA

1 Introduction

The home team advantage is an easily observed phenomenon in sports, particularly in National Collegiate Athletic Association (NCAA) basketball. At the end of the 2017-2018 season, 337 of 351 Division I mens basketball teams had a historical winning record at home, with a median home winning percentage of 67.7% (RPI Ratings). However, it is not only the win-loss record that is affected by playing at home; team and player box score statistics also differ between home and away teams. While many previous studies have examined home advantages in sports, few have examined the impact on box score statistics, and those that have tend to focus on professional leagues (Acharya et al. (2008), Schuckers and Macdonald (2014), van Bommel and Bornn (2017)). In this paper, we will conduct an examination of both the neutral and home court impacts on box score statistics in NCAA basketball, across both genders and all three divisions.

Box score statistics lie at the heart of evaluating the performance of NCAA basketball players and teams, thus it is important to understand the process behind them and any influential factors. Traditional views of home court advantage support the impact of travel, spectator support, and home team familiarity (Stefani). More recent attention has been focused on the impact of human biases through participants such as referees (Swartz and Arce, 2014) and scorekeepers (van Bommel and Bornn, 2017). Specific to the NCAA, previous research has concluded basketball referees call more fouls on away teams (Anderson and Pierce, 2009). Additionally, the teams of scorekeepers (or statisticians as they are sometimes referred) that track box score statistics are not a neutral party but are employed by the home teams of NCAA games. Given that the NCAA itself admits at times, statisticians have to use their judgment and knowledge on how to score a certain play (Bialik, 2014), the possibility of scorekeeper bias in favor of the home team, either intentional or otherwise, seems a relevant concern. Thus, this report will examine the human bias factors for referees and scorekeepers in greater detail.

NCAA data has unique challenges compared to data from professional leagues. Even when restricting to a single gender-division combination, NCAA basketball contains far more teams playing fewer games over the course of a season than its professional counterparts the National Basketball Association (NBA) and the Womens National Basketball Association (WNBA). This reality makes it difficult to draw conclusions at the individual team level. However, NCAA data also has its advantages. A feature of NCAA schedules is that teams play games each season at neutral locations: a location not considered to be the home of either team in the game. Such games provide a unique tool in examining home court advantage that has been underutilized in the existing literature. While occasionally used in quantifying an overall home court advantage (Harville and Smith, 1994), to our knowledge neutral location data has never been used in quantifying home court impact on box score statistics. Additionally, having multiple genders and divisions under the same organizational structure allows us to more easily examine factors such as player gender, team skill, and attendance. This report aims to utilize these advantages in presenting a full picture of the home court advantage on box score statistics in NCAA basketball.

The remainder of this paper is organized as follows. Section 2 outlines the data used in the paper and the statistics that will be examined. An overall view of home court advantage is presented using summary statistics in Section 3, and a selection of interesting observations are discussed in greater detail. Sections 4 and 5 present the results of statistical models that dive deeper into the home court advantages. The former section examines the impact of attendance on box score statistics and the latter quantifies the relative impact of home court and other factors on statistic totals. Finally, Section 6 ends with a summary of conclusions and a discussion.

2 Data

Our dataset contains box score information scraped from stats.ncaa.org for 117,897 games between the 2011-2012 and the 2015-2016 seasons across Division I (D1), Division II (D2), and Division III (D3) NCAA basketball for both women and men. The box score statistics examined in this report, and their abbreviations, are presented in Table 1. In addition to standard box score information, we also obtain the location of the game (including whether each team is playing at home, away, or at a neutral location) and the attendance at the game. Neutral location games in which neither team is playing at home, are comprised of games played in mid-season invitational tournaments and end of season NCAA tournaments and make up between 6% and 9% of the data for each gender-division combination. Such games tend to have slightly stronger teams participating (measured by rating percentage index) and slightly lower attendance, which the exception of D1 neutral location games, which tend to have notably higher attendance. Due to data quality issues (such as missing or conflicting data) our dataset contains 77% of the games played over the range of interest.

Table 1 Box score statistics and their abbreviations

Abbreviation	**Statistic**
3FGA	3 point field goal attempts
3FG%	3 point field goal percentage
AST	Assists
BLK	Blocks
DREB	Defensive rebounds
FGA	Field goal attempts
FG%	Field goal percentage
FTA	Free throw attempts
FT%	Free throw percentage
OREB	Offensive rebounds
PF	Personal fouls
PTS	Points
STL	Steals
TOV	Turnovers

3 Home Court Advantage

In all seasons between 2011-2012 and 2015-2016, NCAA basketball teams performed better at home compared to on the road in nearly all statistical categories across both genders and all three divisions. The magnitudes of these performance advantages are displayed in Figure 1. The percent increase value for a given statistic-gender-division combination is computed by filtering the data to include only games in the gender-division combination, subtracting the average away team statistic value from the average home team statistic value, and dividing the difference by the overall average statistic value in the filtered dataset. The home advantages appear quite stable across seasons, implying consistent underlying factors (and not randomness) as the cause. We can further examine home-away differences by splitting them into home-neutral differences and neutral-away differences (computed using the same logic as above), the distributions for which are displayed in Figure 2. Note that summing the home-neutral and neutral-away values for each statistic would reproduce the results of Figure 1.

The combination of these figures presents many interesting insights. Here we list several such insights, and the subsections below dive deeper into two deserving a closer look:

- AST and BLK, largely considered the most subjective statistics, havenoticeablygreater home advantages compared to the other more objective statistics

- the effect magnitudes for nearly all statistics examined are noticeably impacted by division, and many statistics are also impacted by gender

- the boost in FG% and 3FG% is driven by being the home team, while away teams perform more similarly to those playing at a neutral location

- FT% is negatively impacted for away teams compared to neutral teams, but home teams do not receive an additional boost compared to neutral teams (neutral D1 teams actually have

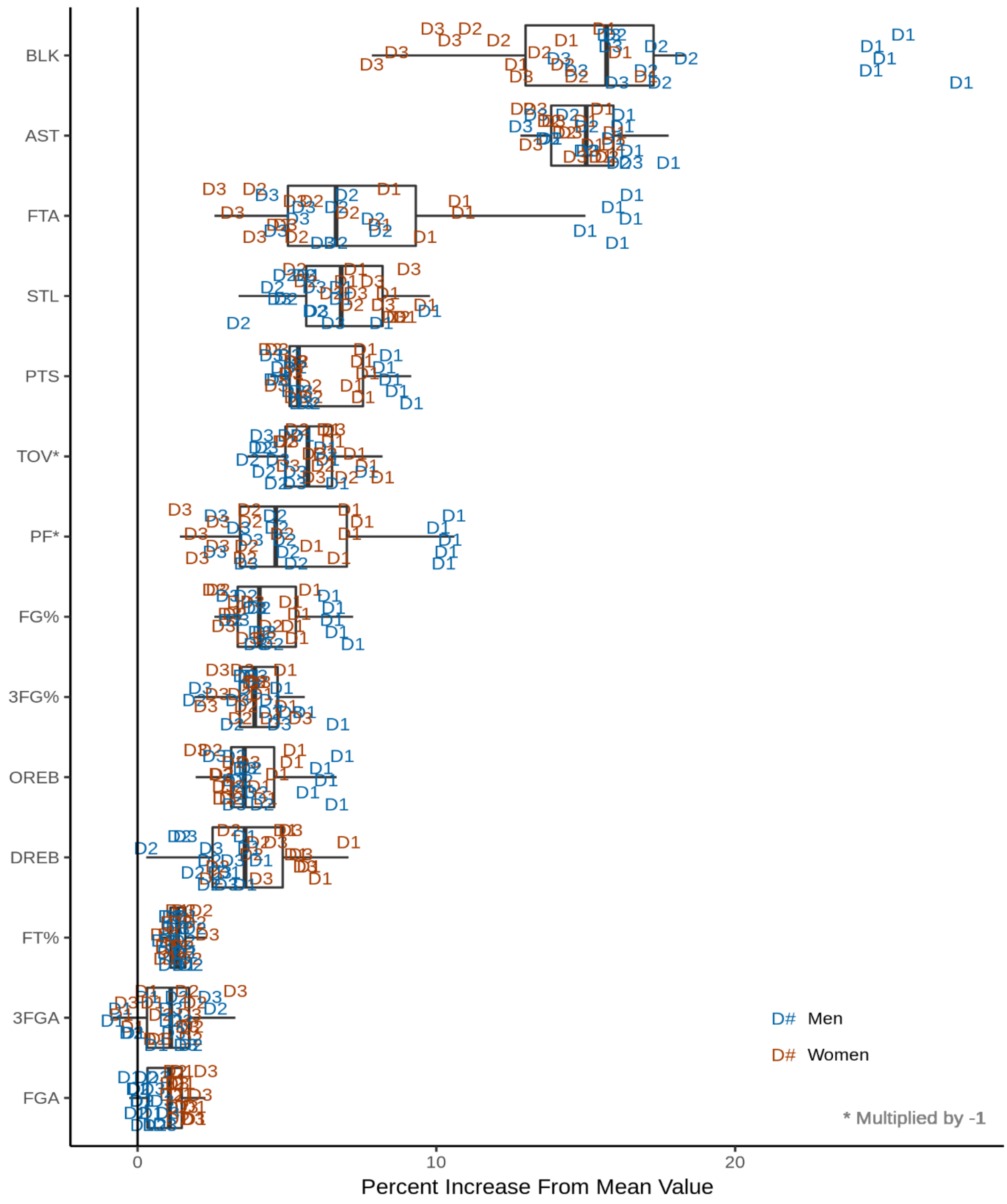

Figure 1 Distributions of the percent increase for home teams compared to away teams for a variety of box score statistics, across all gender-division-season combinations, sorted by the mean percent increase. The individual observations represent a specific combination of gender (color), division (label), and season (ordered vertically with 2011-2012 at the top and increasing down until 2015-2016 at the bottom). Note that positive values indicate an improvement for each statistic (decrease in PF and TOV and increase in all other statistics).

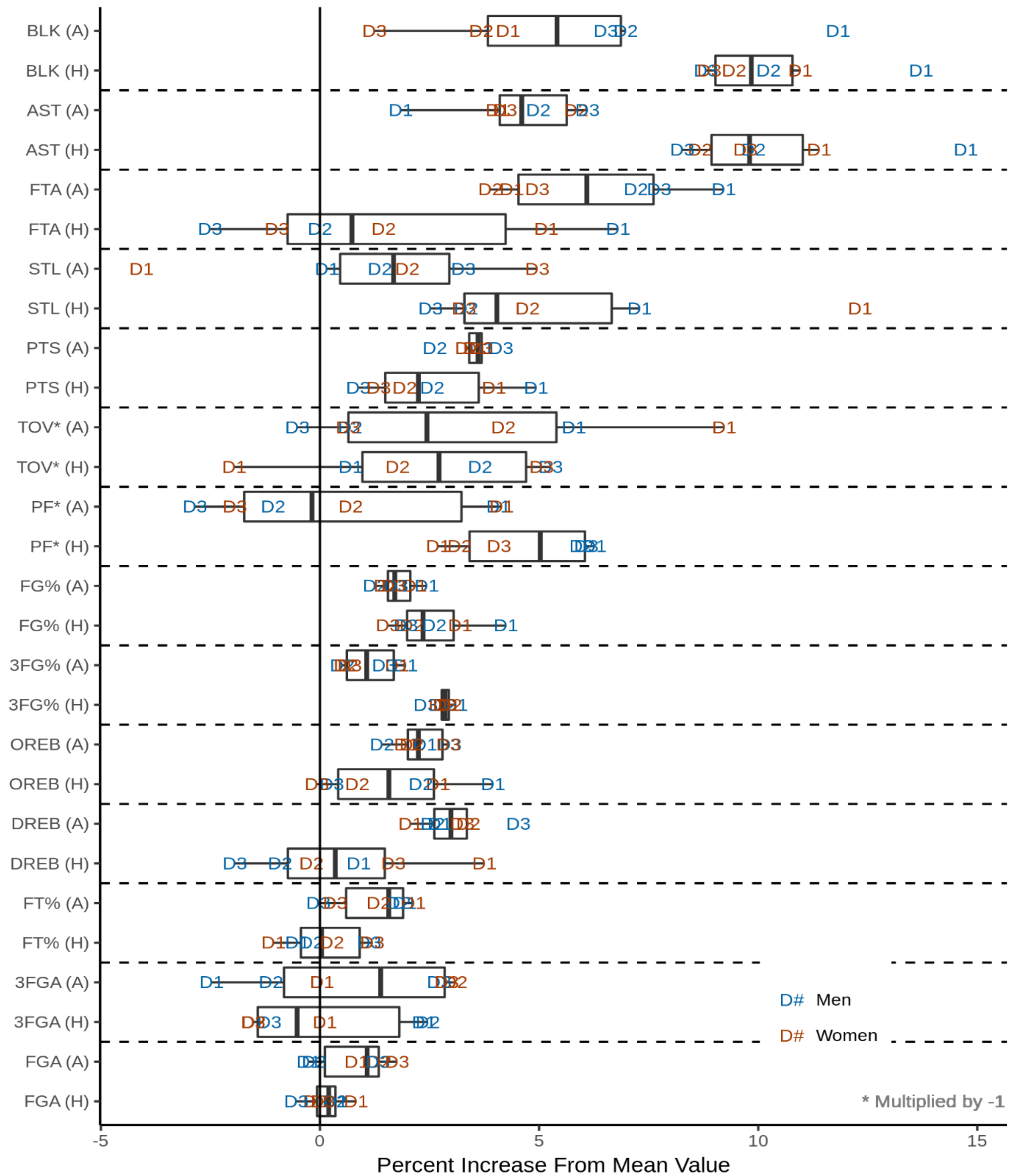

Figure 2 Distributions of the percent increase for home teams compared to neutral teams (H) and for neutral teams compared to away teams (A) for a variety of box score statistics, across all gender-division combinations. The individual observations represent a specific combination of gender (color) and division (label) across all seasons. Note that positive values indicate an improvement for each statistic (decrease in PF and TOV and increase in all other statistics) in the expected direction (home team improvement for home-neutral and neutral team improvement for neutral-away).

a higher FT% than home teams)

- as would be expected given their relationship, the home-neutral distribution of PF mirrors the neutral-away distribution of FTA, and vice-versa (the same is also true for STL and TOV)

3.1 The Impact of Scorekeeper Subjectivity

There are likely several factors contributing to the home advantage observed for box score statistics, including superior on-court performance. However, such a performance improvement does not explain why AST and BLK havenoticeablygreater home advantages compared to the other statistics. According to a former NBA scorekeeper,scorekeepers are given broad discretion over two categories: assists and blocks(Craggs, 2009). This statement aligns with previous observations on the subjectivity of statistics at both the NBA (Biderman (2009), Moore (2015)) and NCAA(Bialik, 2014) levels, including model-quantified patterns of inconsistent scorekeeper behavior for these statisticsin the NBA (van Bommel and Bornn, 2017). Additionally, the statistics with the least subjectivity (including FGA, FT%, and DREB) are all among the statistics with the least advantage for the home team. Thus, these results support the hypothesis that scorekeeper biases are a major factor in the home court advantages for box score statistics.

From Figure 2 we also see that most of the advantage for AST and BLK comes from being the home team, and there is less of a difference between neutral and away teams. This result supports the idea that scorekeepers are more biased in favor of the home team than they are biased against the away team. For the more objective statistics such as FGA, FT%, and DREB the trend reverses and the boost provided to neutral teams over away teams is greater than that for home teams over neutral teams, implying an increase in the relative impact of on-court performance.

3.2 The Impact of Gender and Division

Table 2 presents the statistics for which the order of the percent increase values (from Figures 1 and 2) is dictated by gender or division. Even when restricting to perfect sorting, division orders at least one of the home-away, home-neutral, or neutral-away distributions for 10 of the 14 statistics, including the 8 statistics with the greatest home-away differences. Only 3 of these statistics (STL, TOV, and FT%) have a distribution ordered such that D3 teams receive the greatest boost and D1 the least (the rest have the reverse order). Of those 3 statistics, only STL have consistent ordering across distributions and none of the statistics see D3 have the greatest boost for the overall home-away distribution. We also observe similar trends for gender: 9 of the 14 statistics have at least one distribution sorted by gender, though only 3 have gender as a primary sort for a distribution. Of the 9 total statistics, only 3 (TOV, PF, and FGA) have a distribution for which women receive the greater advantage and only 1 (FGA) has women receive the greater boost for its home-away distribution. Overall, men and higher divisions tend to see greater impact in the expected directions (home over neutral over away) and this observation is especially true when comparing box score statistics between home teams and away teams.

Interestingly, of the 3 statistics for which gender and/or division impact all distributions, one (BLK) is the statistic with the greatest overall home-away impact and the other two (FTA and PF) are the referee driven statistics. Specifically, the home-away impacts on both FTA and PF are both sorted by division, with D1 teams receiving a noticeably greater boost. Knowing that average attendance values are also ordered according to division, we raise the hypothesis that referee bias

Table 2 The division and gender ordering of statistic home percent increase effects, with top lines signifying the primary sort and bottom lines (if applicable) signifying the secondary sort.

	Home - Away	**Home - Neutral**	**Neutral - Away**
BLK	M > W D1 > D2 > D3	D1 > D2 > D3	M > W D1 > D2 > D3
AST	D1 > D2 > D3 M > W		
FTA	D1 > D2 > D3 M > W	D1 > D2 > D3	M > W
STL		D3 > D2 > D1 M > W	D3 > D2 > D1
PTS	D1 > D2 > D3	D1 > D2 > D3	
TOV		D3 > D2 > D1 M > W	D1 > D2 > D3 W > M
PF	D1 > D2 > D3 M > W	M > W	D1 > D2 > D3 W > M
FG%		D1 > D2 > D3 M > W	
3FG%			
OREB		D1 > D2 > D3 M > W	
DREB			
FT%		D3 > D2 > D1	D1 > D2 > D3
3FGA			
FGA	W > M		

is impacted by attendance. Such a hypothesis aligns with previous studies claiming home win probability advantages are primarily due to referee bias (Moskowitz and Wertheim, 2011), and that soccer referee bias in favor of the home team increases as attendance increases (Garicano et al., 2005). The following section examines this hypothesis in greater detail, by examining the impact of attendance on all box score statistics.

4 The Crowd Effect: Attendance Impact on Statistics

This section seeks to determine the importance of fans in a game, specifically by answering the question: do larger crowds have a positive impact on box score statistics for the home team? To test the impact of attendance on box score statistics for a gender-division combination, we compare the home advantages (home team statistic value away team statistic value) in high attendance games and low attendance games for a variety of box score statistics. Specifically, we perform two-sample t-tests for significant differences in the mean home advantage values. Here, we define low and high attendance games as games with attendance values in the bottom 25 percent and top 25 percent respectively of the corresponding gender-division attendance values. The cutoffs for each gender-division combination are presented in Table 3.

From the data, we observe an intuitive trend: home teams with higher attendance tend to be stronger teams, as measured by rating percentage index (RPI). More sophisticated measures of team strength have been developed but we choose RPI for its interpretability and relevance in NCAA history (Wikipedia). We calculate the RPI for a team in a given season as:

$$RPI = (WP \times 0.25) + (OWP \times 0.50) + (OOWP \times 0.25)$$

where WP is team winning percentage, OWP is opponents' winning percentage, and OOWP is opponents' opponents' winning percentage. For a measure of relative team strength in a game, we compute the home team RPI advantage (RPI) as:

$$RPI' = Home\ Team\ RPI - Away\ Team\ RPI$$

Across all gender-division combinations, the KolmogorovSmirnov test (KS test) concludes the distributions of RPI in high attendance games are significantly different than those for low attendance games. A visual example of the RPI difference for D1 women is presented in Figure 3. Thus, any potential difference in home advantages between high attendance and low attendance games caused by attendance, will be confounded by RPI differences.

To remove the RPI impact on the home advantage results, we use statistical matching techniques to transform our high attendance samples so that their RPI distributions match the corresponding low attendance RPI distributions. Specifically, given distributions of low attendance and high attendance games, we perform the following algorithm:

- Define 25 equal width bins $b_1, \ldots, b_{25}$ across the range of RPI for the low attendance distribution
- For each bin b_i:
  - Determine the number of low attendance observations with RPI values in b_i and define this value as n_i
  - Randomly sample with replacement from the high attendance observations with RPI values within b_i until n_i values have been sampled and define this set of values as H_i. If there are no high attendance observations with RPI values within b_i, define H_i to be the empty set.
- Combine the values of $H_i, i = 1, \ldots, 25$ to form a new distribution of high attendance game observations, replacing the original distribution of high attendance game observations

Sample results of the above algorithm for 2015-2016 D1 women are presented in Figure 3. Matching visual intuition, the KS test concludes there are no significant differences in distribution between the low attendance observations and the adjusted high attendance observations resulting from the above algorithm, across all gender-division combinations.

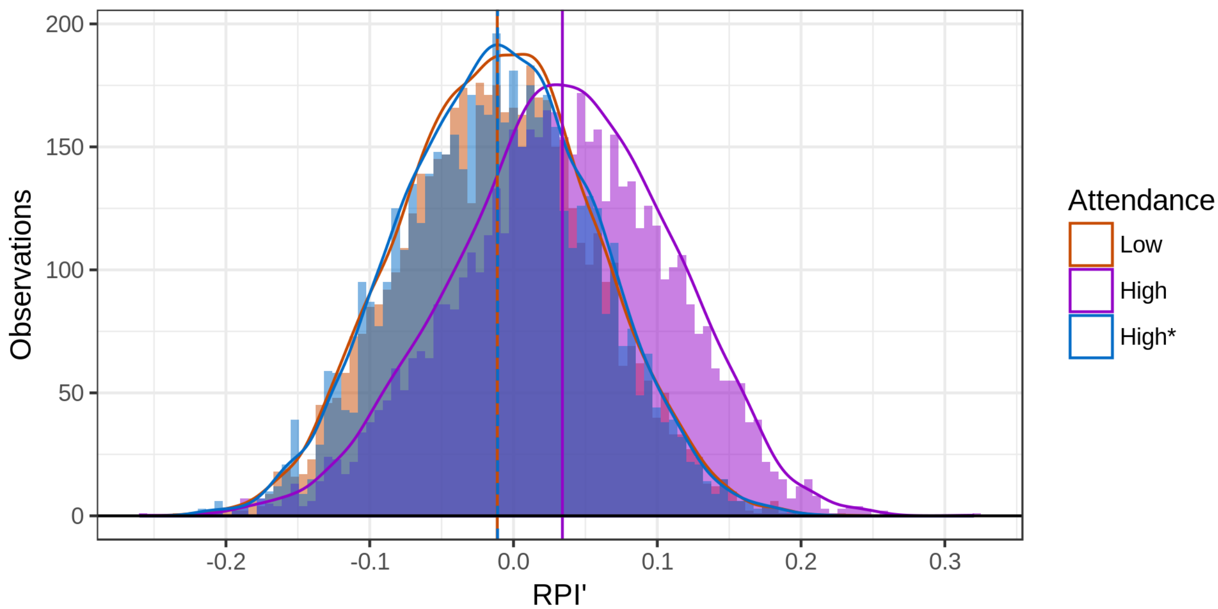

Figure 3 Results of RPI statistical matching for 2015-2016 D1 women data. The Low and High attendance distributions are those observed in the data and the High* distribution is the result of the matching algorithm. The vertical lines label the mean of each distribution.

With the impact of RPI mitigated, we are able to perform two-sample t-tests to test for significant differences in the mean home advantage values. Since the matching algorithm has a random component, we repeat the process 1000 times for each statistic, fitting 1000 distinct models. The average magnitudes of the resulting p-values are displayed in Table 3, with negative values representing combinations implying negative home advantage values. Since we examine 14 statistics and 6 gender-division combinations, we perform 84 tests and thus use the Bonferroni correction to select a level of significance of $\alpha = 0.05/84 = 0.0006$. The two statistics which have significantly different home advantage means in the low attendance and high attendance subsets for the most division-gender combinations are PF and FTA: the two referee driven statistics (no tests for D2 women produce significant differences, though PF and FTA represent 2 of the 3 lowest p-values for that combination). This result supports our hypothesis from Section 3, providing strong evidence that referee bias in favor of the home team increases as attendance increases.

The statistic which is next most impacted by attendance is PTS. Since PTS are also influenced by referees (through FTA), this result is likely a consequence of the home advantage in PF and FTA. Additionally, the PTS p-values for D2 and D3 are lower than those for D1, implying that referee decisions may have greater impact on outcomes at levels in which the teams and players are less skilled. Attendance has significant impact for two additional statistics for D1 men: BLK and FGA. Like PTS, FGA is influenced by referees, since possessions ending in FTA cannot end in FGA. The increase in BLK is the only significant result without a clear tie to referees and may indicate that players are more likely to attempt to block shots (and elicit a crowd response) when attendance is especially high. Finally, FT% is not impacted by attendance in any gender-division combination. Thus, while a home crowd has a positive effect on FT% (see Figure 1), the size of that crowd does not seem to have a significant impact.

Table 3 Results of two-sample t-tests for significant differences in the mean home advantage values between low attendance and high attendance games. The Low and High rows present the cutoffs for low and high attendance games respectively for each gender-division combination. The remaining rows present mean p-values from 1000 iterations of t-tests, multiplied by -1 if the test concluded a negative home advantage. P-values significant at the level $\alpha = 0.0006$ are highlighted in bold. The Overall column presents the average magnitude of the p-values for each statistic across all gender-division combinations.

		Division I		**Division II**		**Division III**	
	Overall	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Low	384	1507	427	281	181	181	112
High	1584	7042	1824	897	527	500	300
PF	3.67×10^{-4}	$\mathbf{9.20\times10^{-15}}$	$\mathbf{5.76\times10^{-9}}$	$\mathbf{3.53\times10^{-4}}$	0.002	$\mathbf{4.64\times10^{-7}}$	$\mathbf{1.01\times10^{-12}}$
FTA	0.003	$\mathbf{3.71\times10^{-10}}$	$\mathbf{2.80\times10^{-4}}$	$\mathbf{2.04\times10^{-4}}$	0.024	$\mathbf{2.08\times10^{-4}}$	$\mathbf{1.17\times10^{-6}}$
PTS	0.031	0.056	0.127	$\mathbf{7.79\times10^{-6}}$	0.015	0.001	0.020
3FGA	0.068	-0.014	0.108	-0.025	-0.227	-0.009	0.096
AST	0.071	-0.098	0.130	0.069	0.034	0.006	0.162
TOV	0.072	0.046	-0.086	0.095	0.229	0.026	0.023
BLK	0.092	$\mathbf{2.17\times10^{-4}}$	-0.091	0.001	0.235	0.044	-0.272
FG%	0.111	0.144	0.266	0.002	0.049	0.048	0.269
FGA	0.125	$\mathbf{-1.95\times10^{-4}}$	-0.093	0.080	-0.177	-0.234	-0.289
OREB	0.133	0.312	0.250	0.003	0.035	0.234	-0.094
DREB	0.138	-0.015	-0.150	0.063	0.252	-0.261	-0.227
3FG%	0.158	0.163	-0.041	0.245	0.183	0.188	0.284
FT%	0.164	0.284	-0.252	0.071	0.150	0.241	0.151
STL	0.169	0.169	-0.272	0.183	0.230	0.055	0.277

5 Quantifying Home Court Advantage

We now seek to quantify the impact of playing at home on the totals of team box score statistics by estimating a regression model for each statistic. Specifically, we estimate least absolute selection and shrinkage operator (LASSO) Poisson regression models, which allow for unimportant variables to be removed from consideration. Thus, the estimated coefficients for a given statistic take the form:

$$\hat{\beta} = \underset{\beta_0,\beta}{\arg\min} -\frac{1}{N}\sum_{i=1}^{N}\left(y_i\left(\beta_0+\beta x_i\right)-e^{\beta_0+\beta x_i}\right)+\lambda\sum_{j=1}^{p}|\beta_j|$$

where β_0 is an intercept term, y_i is the i^{th} observed value for the given statistic, x_i contains information about p input values for the i^{th} observation, including RPI, number of possessions in the game, gender-division combination, and home team, N is the total number of observations, and λ is a penalty term selected through 100-fold cross validation. Since RPI is not measured in a meaningful unit, we transform the RPI values to have mean 0 and standard deviation 1. Thus, the resulting coefficient values represent the effects of increasing RPI by one standard deviation, the equivalent of a mens D1 2015-2016 team switching from playing Iowa (#15 Associated Press rank) to Kansas (#1 Associated Press rank) (Nolan). For these models, we remove the neutral games from the dataset to directly target home-away differences. Also, since we are examining

Table 4 Percent impact of a variety of factors on box score statistics for the 2015-2016 NCAA season, estimated using LASSO regression, with columns sorted in order of overall home advantage. The empty cells indicate variables not selected in the corresponding modelling process. The Division factors account for the baseline differences in statistic occurrences among gender-division combinations. The Home factors measure the increase in frequency of each statistic for the home team compared to the away team both overall and within each gender-division combination. Note that all values represent improvements (decrease in TOV and increase in all other statistics).

		BLK	**AST**	**TOV**	**STL**	**DREB**	**OREB**	**3FGA**	**FGA**
Home	Overall	12.91	12.27	4.48	4.48	1.90	1.74	0.51	0.09
	Men D1	8.87	0.26		-0.34	3.08	0.27	-0.77	-1.00
	Men D2	3.70	1.09	0.89		0.89	-0.57	0.38	
	Men D3	3.55					-0.71	0.59	
	Women D1		0.90		0.22	1.99	1.58	-0.37	0.89
	Women D2	-1.69				-0.03	0.62	0.80	0.82
	Women D3	-3.70	-0.09	0.27	1.56	-0.87	1.83	1.55	1.45
Division	Men D1		2.04	7.02	-7.18	-1.85	-0.86	8.16	-0.89
	Men D2	-8.08	-0.52	4.49	-6.85	-1.39	-1.96	3.81	-0.86
	Men D3	-9.87		0.84	-0.79	1.90	-0.81	0.52	-0.12
	Women D1	5.45	1.27	18.34	19.45	-0.55	16.78	-2.60	1.63
	Women D2	-3.94	-0.81	20.70	19.10	2.46	13.50	-5.90	0.75
	Women D3	1.03	-1.50	30.87	32.16	5.92	22.03	12.36	1.99
	RPI'	10.14	12.27	6.31	10.01	5.61	5.90	0.90	1.71
	Possessions	0.30	0.97	-0.57	1.05	0.46	0.74	2.14	0.80

statistic totals, we do not estimate models for the percentage statistics (FG%, 3FG%, and FT%). Additionally, given that attendance was shown to be an influencing factor for PF and FTA, we remove those statistics from consideration.

Including a term for the gender-division combination allows the models to capture any specific differences in baseline statistic performance. For example, D3 womens teams averaged 8.94 STL per game in 2015-2016 while D1 mens teams averaged only 6.12. Section 4 demonstrated the impact relative team strength can have on statistics and so we include an RPI term to account for such effects. Finally, a possessions term is included for two reasons. First, the number of possessions in a game certainly influences the statistics totals, and second, the term allows us to make relative statements about the other model effects, using the unit of possessions as a baseline.

After estimating the models, the percentage impact values are easily computed as $e^{\beta_j} - 1$ and are presented in Table 4. Here we present results for the 2015-2016 season, but testing other seasons produces similar conclusions. Aligning with previous observations, BLK and AST have a much greater home team advantage compared to the other statistics, with home teams receiving over a 12% boost for each. In units of possessions, the home court boost for AST is the equivalent of receiving 12 additional possessions in a game, and for blocks the boost is equivalent to that of 43 possessions, just under half (47%) of an average game! Even with this impressive overall home court advantage, D1 mens teams receive an additional 8.87% boost in BLK when playing at home and D2 and D3 mens teams observe a similar but lesser increase.

The division coefficients also produce interesting results. Controlling for the other factors, while men receive a greater home advantage for BLK, they do not tend to record more BLK than women overall. Across all divisions, women record fewer turnovers but more steals than men, implying women record far fewer non-steal turnovers (throwing the ball out of bounds, shot clock violations, etc.). Women also tend to record fewer 3FGA and more OREB: two observations which are likely intertwined.

Finally, we can use the RPI coefficients to examine which statistics are most impacted by team strength. As expected, stronger teams have superior performance in all examined statistics. Interestingly, the RPI coefficients are highly correlated (0.85) with the overall home advantage coefficients. Thus, it may be the case that scorekeeper biases are also correlated with strength and stronger teams (and their players) record improved box score statistics not only due to their strength, but also due to their scorekeepers.

6 Conclusions and Discussion

In this paper we have presented an examination of the impact of home court on box score statistics in NCAA basketball. Teams playing at home received a boost in nearly all statistical categories across all gender-division combinations and these results have remained consistent over five seasons of data. We also made use of neutral location games to separate the impact of being the home team from the impact of being the away team, providing a unique view into home court advantage.

This paper has presented substantial evidence supporting the idea of scorekeeper and referee biases impacting box score statistics. In both cases, external factors seem to influence these decision makers. For scorekeepers, gender and division appear to impact their behavior, since men and teams in higher divisions receive a greater boost in most statistics at home compared to women and teams in lower divisions. While these biases may be unintentional, they are damaging to the very gender-division combinations that struggle the most with attention and engagement. Scorekeepers may also be influenced by the strength of the team that hired them, as the impacts of home court and team strength are highly correlated for box score statistics. For referees, we presented strong evidence that larger crowds have greater impacts on their decisions, specifically increasing their bias in favor of the home team. Through foul trouble and free throws, such biases have direct impacts on the games, and can steal wins from better performing teams, simply because they are playing on the road.

While removing all human biases from the game may be impossible, the NCAA would be well served to enact a proactive approach to limit such impacts. Until such steps are taken, our results indicate that it may be time to implement adjustments for gender, division, and game location in order for box score statistics and game results to truly reflect the on-court performance of teams and players.

References

Acharya, R. A., Ahmed, A. J., D'Amour, A. N., Lu, H., Morris, C. N., Oglevee, B. D., Peterson, A. W., and Swift, R. N. Improving Major League Baseball park factor estimates. *Journal of Quantitative Analysis in Sports*, 4(2), 2008. doi: 10.2202/1559-0410.1108.

Anderson, K. J. and Pierce, D. A. Officiating bias: The effect of foul differential on foul calls in NCAA basketball. *Journal of Sports Sciences*, 27(7):687–694, 2009. doi: 10.1080/02640410902729733.

Bialik, C. Its easier to get assists in Huntsville, Texas, than anywhere else. https://fivethirtyeight.com/features/its-easier-to-get-assists-in-huntsville-texas-than-anywhere-else, 2014. Accessed 14 November 2018.

Biderman, D. The NBAs most misleading number. http://www.wsj.com/articles/SB123855027541776617, 2009. Accessed 14 November 2018.

Craggs, T. The confessions of an NBA scorekeeper. http://deadspin.com/5345287/the-confessions-of-an-nba-scorekeeper, 2009. Accessed 14 November 2018.

Garicano, L., Palacios-Huerta, I., and Prendergast, C. Favoritism under social pressure. *Review of Economics and Statistics*, 87(2):208–216, 2005.

Harville, D. and Smith, M. The home-court advantage: How large is it, and does it vary from team to team? *The American Statistician*, 48(1):22–28, 1994. doi: 10.2307/2685080.

Moore, J. NBA court factors: How assists arent as unbiased as you think. http://deadspin.com/nba-court-factors-how-assists-arent-as-unbiased-as-you-1694917271, 2015. Accessed 14 November 2018.

Moskowitz, T. and Wertheim, L. *Scorecasting: The hidden influences behind how sports are played and games are won*. Crown Archetype: New York, 2011.

Nolan, W. 2015-2016 mens college basketball rating percentage index (RPI). http://www.warrennolan.com/basketball/2016/rpi. Accessed 14 November 2018.

RPI Ratings. Mens division I home court records. http://rpiratings.com/homecourtrec.php. Accessed 18 November 2018.

Schuckers, M. and Macdonald, B. Accounting for rink effects in the National Hockey League's real time scoring system. http://arxiv.org/abs/1412.1035v1, 2014. Accessed 2 December 2015.

Stefani, R. Measurement and interpretation of home advantage. In Koning, R. H. and Albert, J. H., editors, *Statistical Thinking in Sports*. Chapman & Hall/CRC.

Swartz, T. B. and Arce, A. New insights involving the home team advantage. *International Journal of Sports Science & Coaching*, 9(4):681–692, 2014. doi: 10.1260/1747-9541.9.4.681.

van Bommel, M. and Bornn, L. Adjusting for scorekeeper bias in NBA box scores. *Data Min Knowl Disc*, 31(1622), 2017. doi: 10.1007/s10618-017-0497-y.

Wikipedia. Rating percentage index. https//en.wikipedia.org/wiki/Rating_percentage_index,. Accessed 17 November 2018.