

The Van Exel Effect: Adjusting for Scorekeeper Bias in NBA Box Scores

Matthew van Bommel and Luke Bornn
Simon Fraser University
mvanbomm@sfu.ca lbornn@sfu.ca

1 Introduction

Box score statistics are the baseline measures of performance in the National Basketball Association (NBA). These metrics, either in their raw form or as components of advanced metrics such as player efficiency rating (PER) [9] or win shares (WS) [3], are used by both fans and team officials to help measure and understand player performance. Thus, box score statistics play an influential role in determining playing time, salaries, trade negotiations, marketing potential, and public perception of players, so any inaccuracies or inconsistencies in their attribution can have far reaching impacts. Additionally, with the explosion in popularity of fantasy sports, and in particular daily fantasy, these inconsistencies could potentially have both legal and monetary impacts.

Box score statistics for each game are identified and recorded by a team of scorekeepers employed by the home team. Some statistics, such as points scored, are objective and there is little possibility of error by the scorekeepers. However, other statistics, such as assists and blocks, are more subjective in nature. For example, an assist “is awarded only if, in the judgment of the statistician, the last player’s pass contributed directly to a made basket” [10].

An example of the consequences of this subjectivity occurred in 1997 when the Vancouver Grizzlies hosted the Los Angeles Lakers. Laker Nick Van Exel was awarded 23 assists, including several that were “comically bad”, by a disgruntled Grizzlies scorekeeper, in protest of the inaccuracy of box score statistics [4]. The questionable score keeping went undetected, the scorekeeper unpunished, and the recorded box score unaltered.

While this example is extreme, it demonstrates that inconsistencies in the attribution of box score statistics can occur without notice. Something as simple as scorekeepers having differing views of statistic definitions can affect the comparability of the statistics and thus the perception of player performance.

In this paper, we seek to develop methods to quantify and adjust for inconsistencies in the recording of assists and blocks by the 30 team hired scorekeepers. A similar process for National Hockey League (NHL) statistics has previously been performed [13], in addition to related work which examined methods of adjusting statistics in Major League Baseball (MLB) based on the differing physical characteristics of individual ballparks [1].

The remainder of the paper is organized as follows. In the following section, we will conduct exploratory analysis at the game level into the tendency of scorekeepers to award assists and blocks. Then, in Section 3, we will introduce a model of assist and block attribution which accounts for the game location, the teams playing, and scorekeeper impacts. Section 4 introduces a new assist model

which takes advantage of optical player tracking data to predict the probability of individual passes being recorded as assists. The section also presents adjusted assist totals which correct for scorekeeper and other biases for a selection of players most affected by the adjustment process. Section 5 examines the impact of scorekeeper inconsistencies on daily fantasy contests. Finally, Section 6 presents conclusions from the results of the paper, as well as a discussion of possible future work and extensions.

2 Assists and Blocks: The Grey Zone of Basketball Analytics

According to a former NBA scorekeeper [4], scorekeepers are given broad discretion over two box score statistics: assists and blocks. Thus, we focus our attention on these metrics. Since assists are highly dependent on the number of made field goals, and blocks dependent on the number of opposing field goal attempts, we examine the assist ratio (AR) and block ratio (BR) rather than the raw totals. Here, the AR and BR for a team are defined as

$$AR = \frac{\text{Team Assists}}{\text{Team Field Goals Made}} \quad BR = \frac{\text{Team Blocks}}{\text{Opponent Field Goals Attempted}}$$

and can be computed for any given duration, such as a quarter, game, or season.

To examine the scorekeeper impact on these ratios, we use box score data from ESPN.com for the entire 2014-2015 NBA season to compute the season long AR and BR awarded by each scorekeeper to both their home team and the away teams. Figure 1 displays the results, demonstrating noticeable differences among scorekeepers. Note that since a scorekeeper is hired by an NBA team to record statistics for all of that team's home games, we use the team names and logos to reference the scorekeepers. Examining assist ratios, there are scorekeepers who awarded high ratios to both home and away teams (Los Angeles Clippers), low ratios to both home and away teams (Utah Jazz), and even high ratios for home teams but low ratios for away teams (Washington Wizards). Similar variability occurs with block ratios. For example, the home team block ratio of the New Orleans Pelicans awarded by the Pelicans scorekeeper is over twice that of the Brooklyn Nets ratio awarded by the Nets scorekeeper.

Examining the results for the Wizards scorekeeper, it may be the case that they have a bias (either intentional or unintentional) toward their team and thus are more generous when awarding assists. However, it may also be that the Wizards' offensive style focuses on ball movement resulting in the Wizards making more passes and earning more assists than the average NBA team. Similarly, we cannot be certain if the Pelicans scorekeeper seeks out blocks to award to their team more diligently than the Nets scorekeeper, or if Anthony Davis (a Pelicans player who led the league in blocks for the 2014-2015 season) is simply that much better of a shot blocker than any player on the Nets team, such that it inflates the Pelicans' ratio to well above that of the Nets. In the following section, we reduce this uncertainty through models of AR and BR which separate the influence of the teams from the influence of the scorekeepers.

3 Team Adjusted Models

We now introduce a regression model to estimate and isolate the effects of several influential variables. Such models are commonly used across several sports to adjust metrics for context [1, 8, 11, 13]. Our model examines the context surrounding the AR or BR for a single team in a given game. Aside from the actions of the scorekeeper, we suspect the primary variables influencing the assist or block ratio awarded by a scorekeeper in a game are the teams in that game. Both the style of play and the skill of

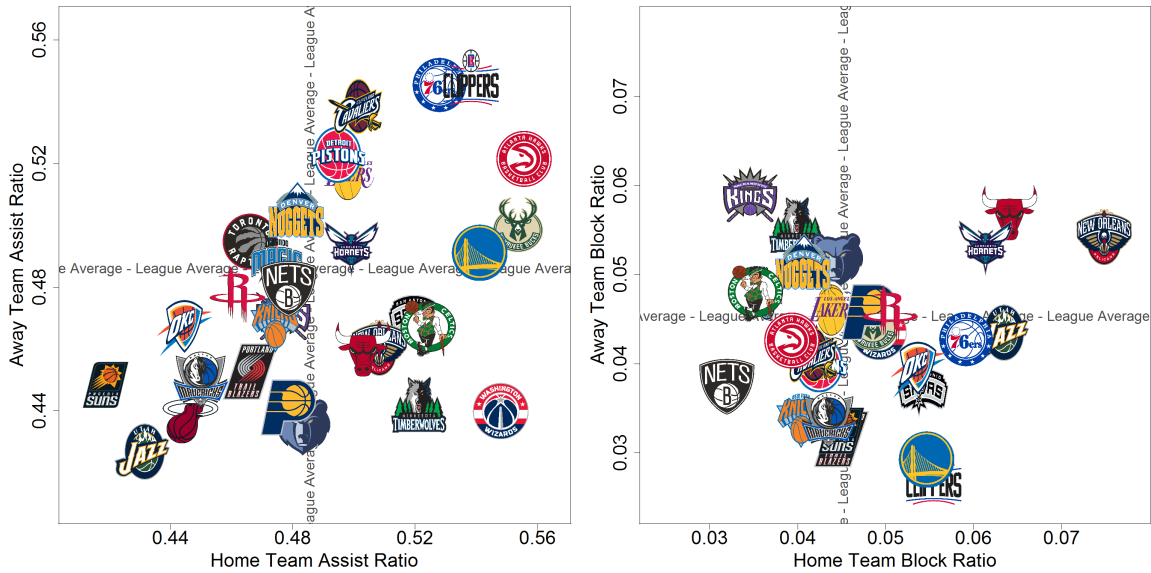


Figure 1: Home team and away team assist ratios (left) and block ratios (right) for the 2014-2015 NBA season for each scorekeeper

the players on a team influence its likelihood to record an assist or block. Similarly, the style and player skill of a team also influence its likelihood of allowing an assist or block. Thus, the models estimate a “team” (β_T) and “opponent” (β_O) effect for each team, corresponding to the likelihood of each team to respectively record or allow a given statistic. Our models also include a “home” (β_H) effect, common to all teams. This effect is present only when the team of interest is the home team and represents any possible league-wide home court advantage resulting in increased performance with respect to the given statistic.

The final two non-intercept effects estimated in our models are the “scorekeeper generosity” (β_G) and “scorekeeper bias” (β_B) effects corresponding to each team’s scorekeeper. β_G measures how likely a scorekeeper is to award assists or blocks to both teams while β_B measures how much more likely a scorekeeper is to award an assist or block to the home team compared to the away team. Isolated from the influence of the other previously mentioned effects, these effects will provide insight into the consistency of scorekeepers across the NBA.

Let R_i be the expected ratio of interest (either AR or BR) for a given team-game combination i . Our model for R_i takes the form

$$R_i = \beta_0 + H_i\beta_H + T_i\beta_T + O_i\beta_O + S_i\beta_G + S'_i\beta_B \tag{1}$$

where H_i is an indicator variable denoting if the team in i is the home team, T_i , O_i , and S_i are each 30×1 indicator (one-hot encoded) vectors for the team, its opponent, and the scorekeeper for team-game combination i respectively, $S'_i = S_i \times H_i$ is a 30×1 indicator vector denoting the scorekeeper if the team in team-game combination i is the home team (and is a zero vector otherwise), β_0 and β_H are estimated coefficients and β_T , β_G , and β_B are 1×30 row vectors of estimated coefficients. The coefficients measuring scorekeeper effects (β_G and β_B) are the coefficients of interest while the remaining coefficients account for the impact of other influential factors.

To generate the models, we again use box score data from ESPN.com for the entire 2014-2015 NBA season, but this time we compute the assist and block ratios of each team in every game. To compare

the consistency of the 30 scorekeepers for each ratio, we compute predicted ratios awarded by each scorekeeper to both the home team and an unspecified away team. Let the predicted home and away team ratios (either AR or BR) for scorekeeper s be denoted PR_{H_s} and PR_{A_s} respectively. Then

$$PR_{H_s} = \overline{LR} + (\beta_G^{(s)} - \bar{\beta}_G) + (\beta_B^{(s)} - \bar{\beta}_B)$$

$$PR_{A_s} = \overline{LR} + (\beta_G^{(s)} - \bar{\beta}_G)$$

where \overline{LR} is the season long league ratio (AR or BR), the $\beta^{(s)}$ are the entries in the β vectors corresponding to scorekeeper s , and the $\bar{\beta}$ are the average of the elements in the β vectors. The resulting PR_{H_s} and PR_{A_s} values for all 30 scorekeepers for both AR and BR are presented in Figure 2. Note that the scaled β_G values are the differences between the league average and the away team predicted ratios while the scaled β_B values can be determined by subtracting the away team predicted ratios from the home team predicted ratios. It can be noticed that some of the observations from Figure 1 still hold. For example, both AR figures indicate the Los Angeles Clippers scorekeeper is generous and unbiased. However, there are also substantial differences between the figures. The BR results for the Philadelphia 76ers scorekeeper are quite similar to several other scorekeepers (Utah Jazz, Oklahoma City Thunder, San Antonio Spurs) in Figure 1, but are distinct from all other scorekeepers in Figure 2.

The model results reveal several interesting characteristics. First, some scorekeepers are biased against their home team. Away teams are much more likely to be awarded an assist by the New York Knicks scorekeeper or a block from the Atlanta Hawks scorekeeper than the corresponding home teams are. Next, the effect of the scorekeeper on different statistics is not necessarily consistent. The 76ers scorekeeper is among the most likely to award an assist to either team but is among the least likely to award a block, particularly to the away team. Finally, the leaders in both assists (Chris Paul, Los Angeles Clippers) and blocks (Anthony Davis, New Orleans Pelicans) for the 2014-2015 season may have been aided by their home scorekeepers, as the Clippers scorekeeper was the most generous in awarding assists and the Pelicans scorekeeper was the most biased in awarding blocks to their home team.

Though we only present results for the 2014-2015 NBA season, it is interesting to note that these results are persistent across seasons. For example, when we estimate the team adjusted model using data from the 2013-2014 season we see that between the two seasons (holding the season long league assist ratios constant) the correlation of the predicted assist ratios awarded by each scorekeeper to an unspecified away team is 0.742 and the correlation of the predicted assist ratios awarded by each scorekeeper to their home teams is 0.792

While assists and blocks are both subjective statistics, the factors involved in their attribution differ. Blocks occur in an instant (the moment that a defender makes contact with an opposing player's shot) while assists involve two components (a pass and a made basket) which can develop over the course of several seconds and can include additional actions such as pivots, dribbles, and defender movements. Thus the context surrounding an assist is fundamental to its attribution. In the following section, we focus on assists and introduce a new model which uses spatio-temporal information to account for this context.

4 A New Assist Model: Adjusting for Context

We now narrow our focus to the level of individual passes and examine passes with the potential to be recorded as assists. We define a potential assist to be a completed pass from a passer to a shooter who

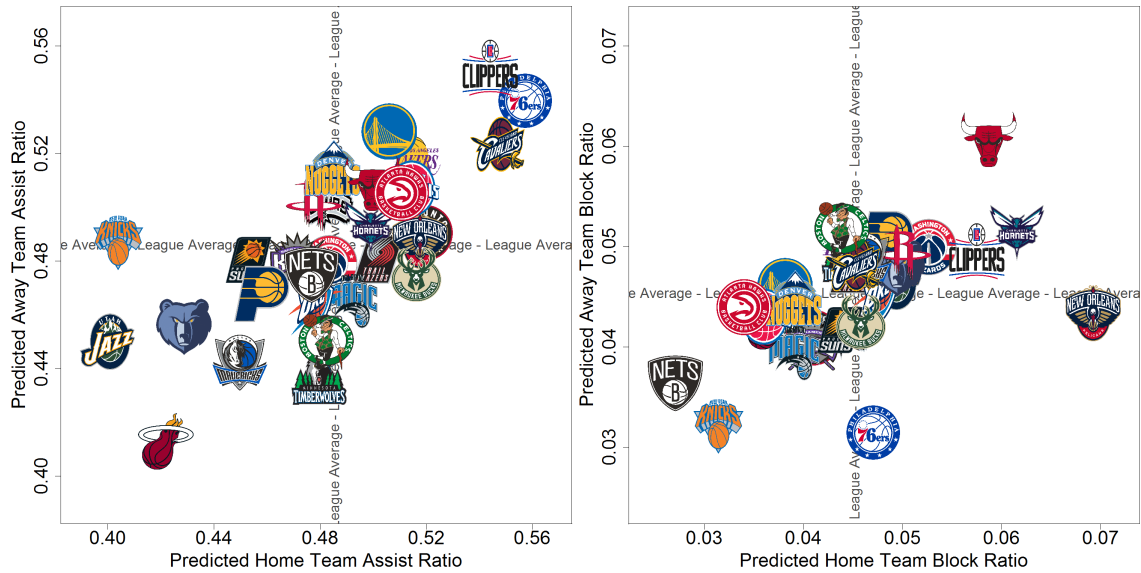


Figure 2: Predicted assist ratios (left) or block ratios (right) awarded by each scorekeeper to the home team and an unspecified away team based on the coefficients of the team-level model

Table 1: Court zone definitions

Name	Definition
Dunk	Area within a 3' radius from the center of the basket
Paint	Area inside the painted key, more than 3' away from the center of the basket
Long 2	Area within the three point line and outside of the painted key
Corner 3	Area behind the shorter 22' three point line (below the break)
Arc 3	Area behind the longer 23'9" three point line (above the break) but within 33'9" of the basket
Heave	Area at least 33'9" away from the basket

then scores a field goal within seven seconds of receiving the pass. The shooter is permitted to dribble and move after receiving the pass, as long as he maintains possession of the ball until the successful shot (no rebounds, turnovers, or additional passes may occur). Note that while an inbounds pass can be credited as an assist, for simplicity we will only examine passes made while the ball was in play.

4.1 Extracting Spatio-Temporal Features

When examining an individual potential assist, there are many contextual spatio-temporal factors that influence its probability of being recorded as an assist by the scorekeeper. Characteristics of the shooter's possession, such as possession length, number of dribbles, and distance traveled, are particularly relevant to the determination of assists, since the pass must be considered to lead directly to the made field goal. The locations of the passer and shooter may also influence the probability of a recorded assist. In order to measure these location impacts, we divide the court into the 6 distinct zones presented in Table 1.

To measure the contextual variables for potential assists, we use SportVu optical player tracking

Table 2: Spatio-temporal variables for the contextual model

Notation	Definition
$C_{(1)}$	Continuous variable denoting the time (seconds) of the shooter's possession
$C_{(2)}$	Discrete variable denoting the number of dribbles taken during the shooter's possession
$C_{(3)}$	Continuous variable denoting the distance (feet) traveled by the shooter during possession
$C_{(4)}$	Continuous variable denoting the distance (feet) traveled by the pass
$C_{(5)}$	Continuous variable denoting the distance (feet) of the nearest defender to the passer at the time of the pass
$C_{(6)}$	Continuous variable denoting the distance (feet) of the nearest defender to the shooter at the start of the shooter's possession
$C_{(7)}$	6×1 indicator vector denoting the court zone corresponding to the passer at the time of the pass
$C_{(8)}$	6×1 indicator vector denoting the court zone corresponding to the shooter at the start of the shooter's possession
$C_{(9)}$	36×1 indicator vector denoting the interaction of passer and shooter court zones ($C_{(7)}$ and $C_{(8)}$)

data from STATS LLC which contains the X- and Y-coordinates of each of the 10 players on the floor and X-, Y-, and Z-coordinates of the ball which are recorded 25 times per second throughout the course of a game. Annotations for events including passes, dribbles, and shots are also included, as well as additional information for player and team identification, dates and times, and game clock and shot clock times. The data set contains full game data for 1212 of the 1230 2014-2015 NBA regular season games. The 18 missing games show no noticeable patterns and of the 42 home games and 42 away games for each of the 30 teams, at least 39 of each are included in the data set and thus no team has fewer than 78 of its 82 games included. We examine the 70,948 potential assists contained in the data set, of which, 51,476 (72.55%) were recorded assists.

The spatio-temporal context variables that will be included in our new contextual model are presented in Table 2. Note that the set of event annotations in the data set mark when a player releases or receives a pass and when the ball is dribbled or released for a shot. Thus, the methods of obtaining values for $C_{(1)}$, $C_{(2)}$, $C_{(4)}$, $C_{(7)}$, $C_{(8)}$, and $C_{(9)}$ are straightforward. Since the SportVu location data is noisy, summing the distance between each observation for a player's location over the range of time that player is in possession of the ball is likely to overestimate the total distance traveled by that player. To correct for this, we take advantage of the NBA traveling violation which prevents a player in control of the ball to move without dribbling the ball (with the exception of pivoting or of two steps allowed immediately after receiving a pass or concluding a final dribble) and define $C_{(3)}$ to be sum of the distances between the observations of a player's location when he receives possession of the ball, each time he dribbles the ball, and when he releases a shot. Finally, $C_{(5)}$ and $C_{(6)}$ are determined by calculating the distance between the corresponding offensive player and each of the five opposing players on the court (at the defined moment in time) and taking the minimum of those five distance values. The stages of a sample potential assist are displayed in Figure 3 to demonstrate the computation of the parameters in Table 2.

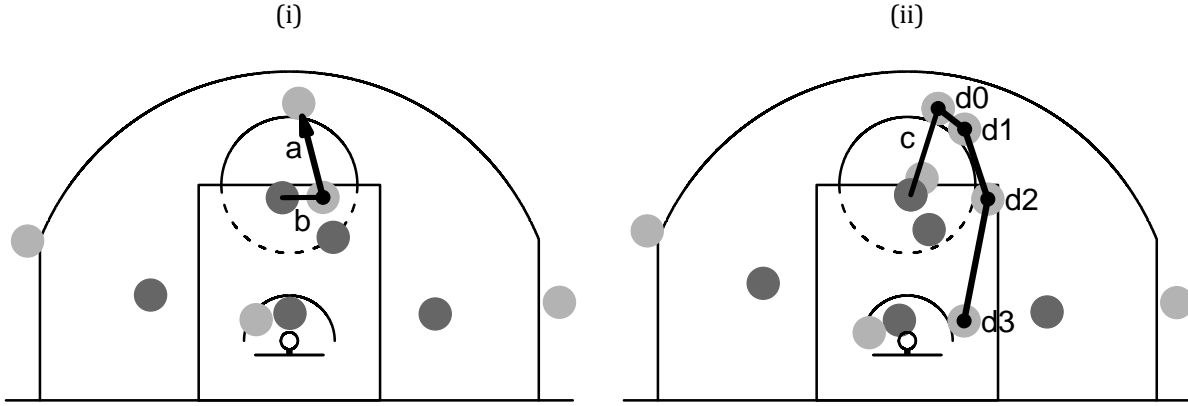


Figure 3: Optical tracking data for the stages of a potential assist which occurred in the first quarter of a January 7, 2015 game featuring the Los Angeles Clippers hosting the Los Angeles Lakers. Each point represents one of the 10 players on the floor, with the lighter points representing Clippers players (on offense) and the darker points representing Lakers players (on defense). The ball is marked by the black dot on top of the player who has possession. In (i), Clippers guard Chris Paul is in possession of the ball in the $C_{(7)}$ = “Paint” court zone and is about to make a pass of distance $C_{(4)}$ = 11.09 ft along arrow **a** to Clippers forward Blake Griffin. At the moment of the pass, the distance of the nearest defender to Paul is $C_{(5)}$ = 3.58 ft, measured by line **b**. In (ii), Griffin receives the pass in the $C_{(8)}$ = “Long 2” court zone. At this moment, the distance of the nearest defender to Griffin is $C_{(6)}$ = 13.63 ft, measured by line **c**. After receiving the pass (**d0**), Griffin drives to the basket, takes $C_{(2)}$ = 2 dribbles (**d1** and **d2**), and shoots the ball (**d3**) (the locations of the other players during the drive are held constant for simplicity). During this drive, Griffin travels $C_{(3)}$ = 20.41 ft over $C_{(1)}$ = 1.82 seconds. On this play, Griffin’s shot attempt was successful and the Clippers scorekeeper awarded Paul an assist.

4.2 Estimating the Contextual Assist Model

For the contextual assist model, we use logistic regression to predict the probability of the j^{th} potential assist being recorded as an assist, given a variety of contextual factors. The contextual model takes the form

$$P(A_j = 1) = \sigma \left(\begin{aligned} &\beta_0^* + H_j \beta_H^* + T_j \beta_T^* + O_j \beta_O^* + S_j \beta_G^* + S'_j \beta_B^* \\ &+ N_j \beta_N^* + P_j \beta_P^* + \sum_{k=1}^9 C_{(k)j} \beta_{C_k}^* \end{aligned} \right) \quad (2)$$

where $\sigma(x) = \exp(x)/(1 + \exp(x))$ and A_j is an indicator function equal to 1 when potential assist j is a recorded assist. The terms common to the team-level model in Equation 3 share the same definitions except the index j refers to a single potential assist achieved in a given team-game combination. For the new model terms, N_j is an indicator vector denoting the name of the passer (from the 486 unique passers in the data), P_j is an indicator vector denoting the primary position (point guard, shooting guard, small forward, power forward, or center) of the passer, and the $C_{(k)j}$ variables are the additional spatio-temporal context variables defined in Table 2. Finally, β_0^* and β_H^* are estimated coefficients and the β_l^* coefficient vectors are $1 \times n_l$ row vectors of estimated coefficients, where n_l is the number of rows of the observation vector which is multiplied by the corresponding coefficient vector. In estimating the model, we use logistic regression with an L2 penalty on the β coefficients,

learned through cross-validation. Thus, the model estimation solves

$$\min_{\beta} \left[- \left(\frac{1}{N} \sum_{j=1}^N A_j M - \log(1 + e^M) \right) + \lambda \|\beta^*\|^2 \right]$$

over a grid of values of λ covering the range of interest where β^* is a vector of all β^* coefficients estimated in the contextual model, $N = 70,948$ (the number of potential assists in the data set), and $\sigma(M)$ is the right hand side of Equation 2 describing the contextual model.

It is important to note that the results of the contextual assist model depend on the selected value of λ . In particular, the estimated coefficients for variables with relatively few observations (such as some passer effects in β_N^* and some court zone interaction effects in $\beta_{C_9}^*$) are more sensitive to shrinkage. Thus, while the relative order of the resulting effects associated with these coefficient values are largely unaffected, the magnitudes of the effects are impacted by the selected λ . In order to mitigate this impact we use 100-fold cross validation to select the optimal λ value and use this value to estimate the model used to generate the results presented in the remainder of this paper.

4.3 Contextual Assist Model Results

The primary focus of this section is to examine the results corresponding to the new variables introduced to the contextual model that were not included in the team-level model. However, we first examine the impact of these additional variables on the scorekeeper effects. Comparing the β_G and β_B coefficients of the team level model to the β_G^* and β_B^* coefficients of the contextual model, both pairs of estimated coefficients are positively correlated. However, the correlation of β_G and β_G^* is 0.864 compared to only 0.321 for β_B and β_B^* . We propose several possible explanations for this difference. First, it may be the case that some teams have a slightly different playing style (intentional or otherwise) when playing at home - one that induces a different average quality of potential assists. These differences would be captured in the contextual model but not in the team-level model, and, since they effect only home teams, would impact only the bias coefficient values. Another explanation is that the range of generosity coefficient values compared to the bias values is much greater in the contextual model. Thus, it may be the case that the variation explained by β_B^* is better explained by β_G^* or other new coefficients in the contextual model. Finally, since bias coefficients apply only to the home team of each game and the generosity coefficients apply to both teams, the bias values are estimated using fewer observations, making them less reliable compared to the generosity values.

Shifting focus to the new variables introduced in the contextual model, we can isolate the impact of a single variable on the probability of a potential assist being recorded as an assist by examining an average potential assist, and observing how its recorded assist probability changes as we manipulate the value of the variable of interest. An average potential assist is a potential assist with no impact from the indicator variables or vectors, and the average values over all potential assists of the other variables. Thus an average potential assist is a pass that travels 15.66 feet from a passer whose nearest defender is 5.29 feet away, to a shooter whose nearest defender is 9.58 feet away when he catches the ball and who then takes 1.07 dribbles, traveling 10.42 feet, over 1.95 seconds before scoring a field goal. The model predicts that this average potential assist has a 78.72% chance of being a recorded assist.

Let V be the sum of the estimated intercept coefficient and the average potential assist values multiplied by their corresponding estimated coefficient values, that is $V = \beta_0^* + 1.95\beta_{C_1}^* + 1.07\beta_{C_2}^* + 10.42\beta_{C_3}^* + 15.66\beta_{C_4}^* + 5.29\beta_{C_5}^* + 9.58\beta_{C_6}^*$. Also, let I be any variable of interest from the contextual

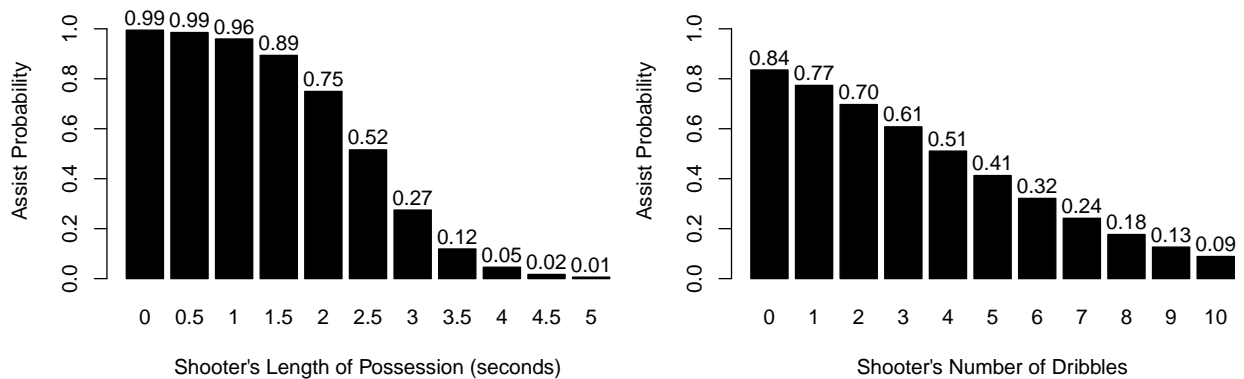


Figure 4: Select predicted recorded assist probabilities of the average potential assist as a function of the continuous possession length (left) or the discrete number of dribbles (right)

assist model, with corresponding estimated coefficient vector β_I^* . If I is a variable included in V , re-define V without the corresponding variable term. The influence of a given value I^* of the variable of interest I has the following effect (E) on the probability of an average potential assist being recorded as an assist:

$$E = \sigma(V + I^* \beta_I^*) - \sigma(V)$$

where E is measured in units of change in probability.

We first use the above effect computation expression to examine the impact of some of the contextual variables in the model. Figure 4 demonstrates the effects of changing the possession length or the number of dribbles of the average potential assist. Both factors have a substantial impact on the probability of a potential assist being a recorded assist. As the number of dribbles increases from 0 to 10, each additional dribble decreases the probability by an average of 7.5%. As for possession length, an increase of only 2 seconds (from 1 second to 3 seconds) decreases the probability of a recorded assist by a remarkable 69%.

We also examine the impact of the passer and shooter locations. Here we ignore passes to or from the “heave” zone as they have relatively low probabilities of being recorded assists. The resulting impact of each remaining pair of zones, and the frequency of passes between them, are presented in Figure 5. Accounting for the other contextual model factors (including player positions, which are discussed in the following paragraph), for all passer locations, passes to the corner 3 zone are the most likely to be recorded assists while passes to the dunk zone are the least likely to be recorded assists. Also, a pass from the paint zone to the corner 3 zone has the highest probability of being a recorded assist while a pass within the dunk zone has the lowest such probability.

Next, we shift our focus to the non-contextual variables of the contextual model, beginning with the primary position of the passer. The results of the position effects are displayed in Table 3 and show that even after accounting for all contextual variables in the model, the probability of an average potential assist being a recorded assist is greatest for point guards and least for centers, with a 7.88% difference in probabilities between the two positions. We propose two possible explanations for this discrepancy. First, the average pass by a point guard may have a high probability of being a recorded assist due to characteristics beyond the scope of the contextual model. However, if the model captures all important characteristics, then the discrepancy may be the result of bias from scorekeepers with respect to passer positions.

Continuing with the non-contextual variables, we end the results examination of the contextual

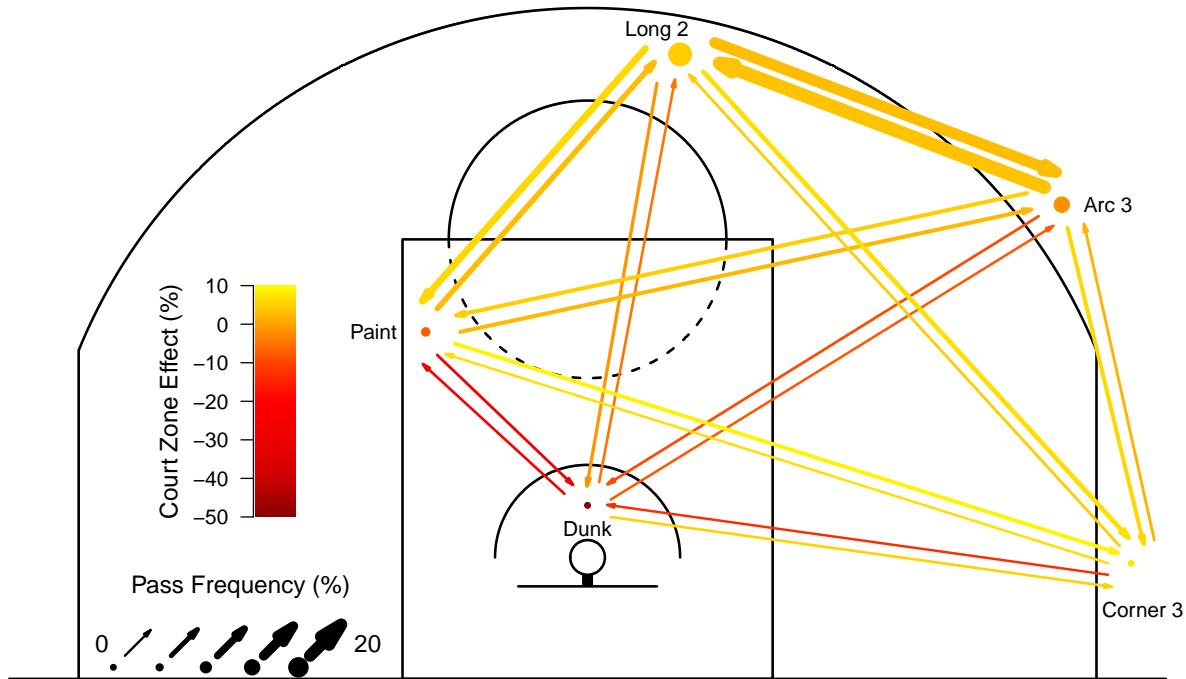


Figure 5: The impact of passer and shooter location on the recorded assist probability of an average potential assist and the frequency of passes between each location pair. For each pair of zones, the arrow points in the direction of the pass. Note that the point in each zone represents passes made within that zone.

model with the impact of the passer on the probability of a recorded assist. The results for the top and bottom ten passer effects are displayed in Table 4. Since passer positions are accounted for separately, they should not be the primary influence of passer coefficients. This idea holds true in the results as both the top and bottom ten include players from all five positions. The difference in effects between the player with the highest effect (James) and the player with the lowest effect (Mozgov) is a substantial 17.77%. Similarly to the position effect, we suspect the differing passer effects are the result of either characteristics of passes not picked up by the model, or bias from scorekeepers with respect to individual passers.

Table 3: Passer position effects where “Effect” is the isolated effect of the passer position on the recorded assist probability of an average assist

Position	Point Guard	Shooting Guard	Small Forward	Power Forward	Center
Effect (%)	3.07	0.84	0.99	-2.16	-4.81

Table 4: NBA players with the top and bottom ten passer effects where “Effect” is the isolated effect of the passer on the recorded assist probability of an average assist and “Rank” is the rank of “Effect” in decreasing order and ranges from 1 to 486

Rank	Passer Name	Position	Effect (%)	Rank	Passer Name	Position	Effect (%)
1	LeBron James	SF	7.68	486	Timofey Mozgov	C	-10.39
2	Chris Andersen	C	7.17	485	Tristan Thompson	PF	-8.75
3	Corey Brewer	SF	7.14	484	Luc Mbah a Moute	SF	-8.56
4	LaMarcus Aldridge	PF	6.58	483	Ramon Sessions	PG	-8.40
5	Klay Thompson	SG	6.20	482	Nate Robinson	PG	-8.09
6	Chris Bosh	PF	6.09	481	Enes Kanter	C	-7.94
7	Patrick Patterson	PF	5.95	480	Jae Crowder	SF	-7.85
8	John Wall	PG	5.90	479	Dahntay Jones	SG	-7.55
9	Stephen Curry	PG	5.86	478	Tony Snell	SG	-7.34
10	Al Horford	C	5.67	477	Andre Drummond	C	-6.88

4.4 Adjusting Player Assist Totals

Since we have isolated the impact of each variable in the contextual model on the probability of a potential assist being a recorded assist, we can compute adjusted assist totals for each player which remove the effects of all non-contextual variables in the model. Similar adjustments of box score statistics have previously been carried out in other leagues such as the NHL [13] and MLB [2]. We can also determine the expected number of recorded assists gained or lost by a player due to an individual factor, such as the passer effect for that player. The ten players with the greatest increase and the greatest decrease in total assists after this adjustment are presented in Table 5. The results show that the “Home Scorekeeper” effect, which is the sum of all assists gained or lost due to the generosity and bias of the home scorekeeper for a player, tends to be the greatest inflater or deflater of the players’ recorded totals. Conversely, the “Away Scorekeeper” effect, which is the sum of all assists gained or lost due to the generosity of the away scorekeepers for a player, tends to be relatively insubstantial. The position effect is a significant deflater for many of the top passers as those players tend to be point guards. Finally, the impacts of the passer effects vary substantially depending on the individual player.

5 Impact of Scorekeeper Inconsistency on Daily Fantasy Sports

In daily fantasy contests individuals pay entrance fees, select a roster of players who generate fantasy points, and potentially receive a payout depending on the performance of their roster, all within the span of 24 hours. The popularity of such games is increasing, and so too is the amount of money at stake. In 2014, FanDuel Inc. and DraftKings Inc., currently the two largest daily fantasy operators in North America, together awarded over \$800 million in prizes across all sports and pledged to increase that number to over \$3 billion in 2015 [12]. Both operators offer daily fantasy contests for the NBA with the fantasy points scoring system relying exclusively on box score statistics - including assists and blocks [5, 7]. Because of this fact, scorekeeper inconsistencies have the potential to influence the outcomes of daily fantasy contests and participants in the daily fantasy community have noticed this influence. In a November 17, 2015 daily fantasy basketball article on ESPN.com, DraftKings analyst Peter Jennings recommends that participants select Anthony Davis in the New Orleans Pelicans home



Table 5: Comparisons of the total recorded and adjusted assists for the 10 players experiencing the greatest increases and the 10 players experiencing the greatest decreases. The “Assist Change” column measures the difference between the recorded and adjusted totals. The “Original Rank” and “Adjusted Rank” columns provide the players’ ranks (1-486) before and after the adjustment respectively. The four right-most columns display the estimated number of additional assists a player originally received due to the given effect (SK is short for scorekeeper) which were removed in the adjustment process. Note that not all factors in the adjustment are displayed, so the “Assist Change” column is not equal to the negative sum of the four displayed adjustment effects.

Player	Assist Change	Recorded Assists	Original Rank	Adjusted Rank	Position	Passer	Home SK	Away SK
Derrick Favors	18.32	109	162	143	-1.83	-2.03	-16.31	-0.03
Marc Gasol	16.98	295	36	30	-7.04	0.17	-13.45	-1.85
Enes Kanter	14.97	53	276	240	-2.34	-3.76	-6.07	-0.75
Gorgui Dieng	13.81	142	123	108	-3.61	-4.04	-4.19	-0.87
Rudy Gobert	11.99	106	167	150	-3.14	1.43	-13.33	-0.85
Dwyane Wade	11.66	277	43	36	0.92	0.57	-13.38	-0.76
Trevor Booker	11.46	81	211	185	-1.46	-1.07	-11.19	0.77
Goran Dragic	10.68	336	25	22	5.77	0.01	-15.94	-0.65
Eric Bledsoe	10.09	480	12	10	1.9	6.05	-18.43	1.11
Udonis Haslem	9.86	43	307	278	-0.86	-1.78	-5.56	-0.03
Blake Griffin	-17.35	340	23	28	-3.31	6.14	13.84	-1.76
Dennis Schroder	-17.94	315	30	34	3.84	2.67	7.42	0.46
Ty Lawson	-19.78	706	3	3	8.93	-4.34	5.06	-0.32
M. Carter-Williams	-19.93	411	19	19	5.65	1.49	14.22	-0.16
Jeff Teague	-25.39	503	9	11	5.20	7.70	9.93	-2.19
Kyrie Irving	-26.24	365	20	23	4.60	1.23	13.58	-0.30
Stephen Curry	-32.07	564	4	4	6.08	11.94	7.02	-1.75
LeBron James	-34.07	481	11	13	1.52	12.61	13.46	-2.14
John Wall	-35.10	750	2	2	8.14	16.12	1.85	-1.22
Chris Paul	-37.92	807	1	1	8.13	13.90	21.41	-3.46

game against the Denver Nuggets for their roster because “Davis was much better at home last season (scorekeeper might be a Davis fan) and the trend is continuing” [6]. This observation that the success of Davis at home may be due, at least in part, to the Pelicans scorekeeper aligns with our results in Figure 2 which showed that the Pelicans scorekeeper was biased toward the home team in awarding assists and was the most biased scorekeeper in the NBA when it came to awarding blocks.

In addition to the overall bias or generosity of scorekeepers, the variability of their behaviour is also important to daily fantasy participants. Depending on their selection criteria and the contest they enter, a participant may either seek or avoid a player in a game whose scorekeeper has a high level of variability in the recording of statistics, since this variability affects the overall variability of a player’s performance.

Using the contextual model, we can compute adjusted assist totals for each team in all 1212 games in the data set by computing the sum of the expected probabilities of the potential assists after remov-

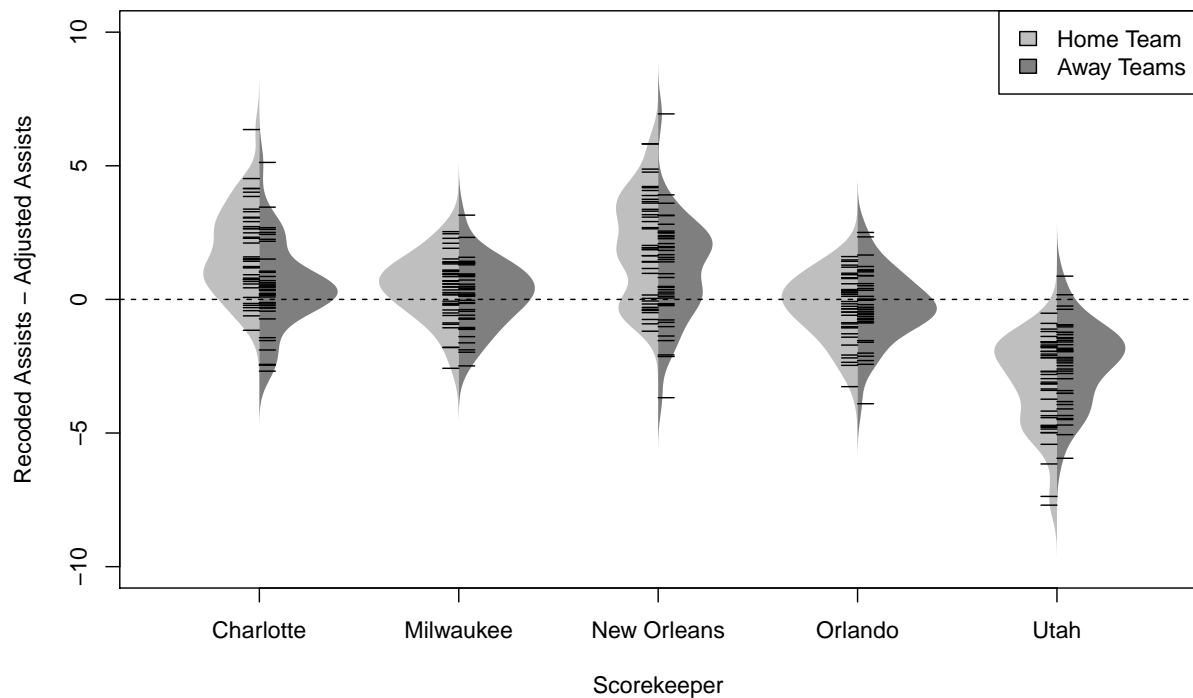


Figure 6: Scorekeeper bonus distributions of the home and away teams for 5 selected scorekeepers

ing the estimated scorekeeper effects. These adjusted values represent the expected assist totals for each team in every game if an average scorekeeper had recorded the statistics. We can then compare the number of recorded assists to the number of adjusted assists and collect the difference values for both the home and away teams for each scorekeeper to obtain a home and away “scorekeeper bonus” distribution for each of the 30 NBA scorekeepers. The home and away team distributions for 5 selected scorekeepers are presented in Figure 6. The means of the distributions range from -3.15 for the home team of the Utah Jazz scorekeeper to 2.02 for the home team of New Orleans Pelicans scorekeeper. Given that teams averaged 21.24 assists per game over the 1212 games in the data set, this difference of over 5 assists per game is substantial. The consistency of individual scorekeepers also varies greatly among the 30 teams. The scorekeeper for the Milwaukee Bucks is among the most consistent in the league with a home distribution variance of 1.36 and an away distribution variance of 1.44. At the other end of the spectrum, the Pelicans scorekeeper has among the greatest scorekeeper bonus variance for both home and away teams with equal values of 3.82. Finally, the similarity of the home and away team distributions for a scorekeeper range from nearly identical (Orlando Magic) to drastically different (Charlotte Hornets).

Thus, even accounting for all other variables in the contextual model, over the course of the 2014-2015 NBA season, scorekeeper performance and behaviour varied greatly both among the 30 scorekeepers and among games for individual scorekeepers.

6 Discussion and Conclusion

In this paper we have presented evidence of inconsistencies in the awarding of box score statistics by the 30 team-hired scorekeepers. We have also presented methods of quantifying these inconsistencies for assists and blocks, which can then aid in adjusting the recorded values. Though we only presented adjusted results for the contextual model from Section 4, the results of the team-level model from Section 3 can also be used to adjust recorded assist and block totals.

In addition to demonstrating inconsistencies in the awarding of assists by the scorekeepers, both to all opposing teams and to their corresponding home teams, the results of the contextual model indicate scorekeepers may have biases in regard to both passer positions and the individual passers. Though this model attempts to include the coefficients we suspect have the greatest impact on the probability of a recorded assist, basketball is a complex system of positioning, events, and interactions, and it is impossible to include all potential factors in any model. As such, the difference in position and passer effects may be the result of characteristics that extend beyond the model. Further work must be completed in order to verify these results.

The same level of detail we used to examine assists could also be applied to the examination of other statistics. We have already presented evidence of scorekeeper inconsistencies in the recording of blocks, and the same may be true for other statistics such as rebounds or turnovers.

In addition to the inconsistencies among the 30 NBA scorekeepers, Section 5 provides evidence of the inconsistency of individual scorekeepers among different games. These inconsistencies have real world consequences, including the rising potential of monetary consequences for daily fantasy participants due to the growing popularity of the contests.

In light of the findings of this paper, the NBA and its players may be well served to adopt a more proactive stance towards monitoring the attribution of subjective box score statistics. While our approach provides an adjustment for players' assist totals, significant work remains to understand the impact scorekeeper inconsistencies have on aggregate metrics such as PER, WS, and other advanced metrics which rely heavily on box score statistics as inputs.

References

- [1] Rohit A. Acharya, Alexander J. Ahmed, Alexander N. D'Amour, Haibo Lu, Carl N. Morris, Bradley D. Oglevee, Andrew W. Peterson, and Robert N. Swift. Improving major league baseball park factor estimates. *Journal of Quantitative Analysis in Sports*, 4(2), 2008. doi: 10.2202/1559-0410.1108.
- [2] Baseball Reference. Neutralized and converted stats. http://www.baseball-reference.com/about/equiv_stats.shtml, 2015. Accessed 2 December 2015.
- [3] Basketball Reference. Calculating win shares. <http://www.basketball-reference.com/about/ws.html>, 2015. Accessed 2 December 2015.
- [4] Tommy Craggs. The confessions of an NBA scorekeeper. <http://deadspin.com/5345287/the-confessions-of-an-nba-scorekeeper>, 2009. Accessed 2 December 2015.
- [5] DraftKings. Daily fantasy basketball league rules. <https://www.draftkings.com/help/nba>, 2015. Accessed 20 January 2016.
- [6] ESPN.com Contributors. Daily fantasy basketball: Building blocks, fades for Nov. 17. http://espn.go.com/blog/fantasy-basketball/post/_/id/3764/daily-fantasy-basketball-building-blocks-fades-for-nov-17, 2015. Accessed 20 January 2016.
- [7] FanDuel. Rules and scoring. <https://www.fanduel.com/rules>, 2015. Accessed 20 January 2016.
- [8] Robert B. Gramacy, Shane T. Jensen, and Matt Taddy. Estimating player contribution in hockey with regularized logistic regression. *Journal of Quantitative Analysis in Sports*, 9(1):97-111, 2013. doi: 10.1515/jqas-2012-0001.
- [9] John Hollinger. *Pro basketball forecast 2004-05*. Brassey's, Washington, 2004.
- [10] NBA. Basketball U on assists. http://www.nba.com/canada/Basketball_U_on_Assists-Canada_Generic_Article-18072.html, 2013. Accessed 2 December 2015.
- [11] Douglas M Okamoto. Stratified odds ratios for evaluating nba players based on their plus/minus statistics. *Journal of Quantitative Analysis in Sports*, 7(2), 2011. doi: 10.2202/1559-0410.1320.
- [12] Kate O'Keeffe. Daily fantasy-sports operators await reality check. Washington Post. <http://www.wsj.com/articles/daily-fantasy-sports-operators-await-reality-check-1441835630>, 2015. Accessed 20 January 2016.
- [13] Michael Schuckers and Brian Macdonald. 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.