Bias in NBA Scorekeeping:

Analyzing a Historic Season

Richard Francais (013404743)

Econ 485

Dr. Jake Meyer

21 December 2017

**Abstract**

The science of scorekeeping is imprecise. While box scores display definite results, the average viewer never sees the subjectivity behind the numbers. This subjectivity manifests itself in the traditional counting statistics of basketball: points, blocks, steals, assists, and rebounds. Whether by simple human error, or even malice, statistics are entered incorrectly all the time. The most common culprit of this are assists, due to the loose definition of an assist, a disproportionate amount of responsibility lies in the scorekeeper’s hands. In this paper, we explored the potential ramifications this responsibility could have had on Russell Westbrook’s MVP campaign. Although our results showed that Westbrook is less likely to get a triple double on the road, there was too much noise to make a definitive conclusion.

**Introduction**

This paper attempts to tackle the issue of scorekeeping bias in the NBA, focusing primarily on Oklahoma City Thunder point guard, Russell Westbrook. In the 2016-17 NBA season, Westbrook made history as only the second player in NBA history to average a triple double over the course of a season. Now what, in particular, makes Westbrook’s season have more to do with NBA scorekeeping than any other players? The answer has to do with the nature of Westbrook’s achievement.

A triple double is defined as simply reaching double digit totals in three of the five statistical categories. It’s a completely arbitrary achievement, meaning that reaching a triple double doesn’t indicate the quality of a performance. For example, two players could have identical triple doubles, say 30 points, 10 rebounds, and 10 assists, but one player could have 2 turnovers and the other could have 15 turnovers. Of course, this is an exaggerated scenario, but the logic holds true, the triple double is not an effective indicator of performance without context. Despite this flaw, the triple double carries an immense amount of weight in the fandom surrounding the NBA. It then follows that scorekeepers could feel obligated, or even pressured, to give Westbrook the nudge in the right direction. We will explore this further in the Literature Review section, but its safe to say that the NBA has been battling back against negative optics surrounding the subjectivity of the game ever since ex-referee Tim Donaghy admitted to fixing playoff games.

**Literature Review**

On January 5th, 1997, Los Angeles Laker’s point guard Nick Van Exel had an incredible game against the Vancouver Grizzlies. Van Exel finished with a stat line of 8 points, 8 rebounds, and a staggering 23 assists in their victory over the Grizzlies. On the surface, it may just seem like Van Exel had a career night, but, this game serves as an example of the control that NBA scorekeepers have over public perception. What really happened on January 5th was that one of the Vancouver scorekeepers was fed up with his job and decided to credit nearly every assist to Van Exel (Craggs 09). Due to the subjectivity of the statistic, this event went nearly unnoticed. This infamous event has inspired a multitude of papers that seek to analyze the impact that scorekeepers have on box scores. The Van Exel Effect: Adjusting for Scorekeeper Bias in NBA Box Scoresby Bommel and Bornn attempts to address this bias by quantifying the change in assist and block ratios between teams whether they’re home or away. The reason why this distinction is made is because the home team’s scorekeepers keep track of the scores for both teams, meaning they may have an incentive to inflate their own team’s stats. This stat inflation may have far reaching impacts in terms of how teams evaluate talent, the contracts that players may receive, or even the MVP race.

In the 2016-2017 NBA season, Oklahoma City Thunder point guard Russell Westbrook made history as only the 2nd player in NBA history to average a triple double stat line over the course a season. Westbrook’s final stat line of 31.6 points, 10.4 assists, and 10.7 rebounds per game was enough to secure him the Most Valuable Player award. A great accomplishment indeed, but due to the arbitrary nature of this achievement, it does raise the concern that scorekeepers may feel obligated, or even pressured into crediting Westbrook with assists he may not have earned. To further analyze this question, I will implement methods like the regressions used in the article previously mentioned. The main difference will be the scale in which the argument is made, the paper focused on the aggregate assists made by the team. In my approach, I want to implement a more player based regression model. The crux of the model will be the dummy variable that indicates whether the game is played at home or away. If the hypothesis is correct, then there should be a statistically significant difference in the predicted assists at home vs on the road.

**Theoretical Model**

Prob(TripleDouble)i = β0 + β1Awayi (+/-) β2USGi (+/-) β3TeamPPG-PTSi (+/-) β4PTSi(+/-) β5ASTi  (+/-) β6TRB

TripleDouble = 1 if triple-double, 0 if not.

Away = 1 if game is away, 0 if home

USG = Usage Rate

TeamPPG-PTS = Team’s points per game minus

points scored by Westbrook

PTS = Points scored by Westbrook

AST = Assists made by Westbrook

TRB = Total Rebounds made by Westbrook

By using a logit regression model, we can attempt to predict the likelihood that Westbrook reaches a triple double in any given game. The independent variables are Away, USG, TEAMPPG-PTS, PTS, AST, and TRB. Of course, the variable of most interest is the Away variable, which would be predicted to have a negative effect on the likelihood of a triple double. Each of the other variables would be expected to have positive coefficients, as the three counting statistics included are the categories in which Westbrook got his double digits, and TeamPPG-PTS is a representation of how effective Westbrook’s teammates are in converting possessions in to baskets. As for the functional form of the equation, it is just a simple linear OLS model, an argument could be made for logging PTS, however I felt that it would diminish the significance of the variable.

**Data and Results**

Data Sources:

* All data used was found on basketball-reference.com

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Summary Statistics | Observations | Mean | Std. Dev. | Min | Max |
| Away | 81 | .4938272 | .503077 | 0 | 1 |
| USG | 81 | 41.54815 | 7.794006 | 20.1 | 58 |
| TeamPPG | 81 | 106.679 | 10.90164 | 80 | 132 |
| PTS | 81 | 31.58025 | 10.49031 | 5 | 58 |
| AST | 81 | 10.37037 | 3.809345 | 0 | 22 |
| TRB | 81 | 10.66667 | 3.687818 | 3 | 18 |
| TEAMPPG-PTS | 81 | 75.09877 | 12.47658 | 50 | 104 |

Below is the first logit regression I ran to find the probability of Westbrook receiving a triple double. Admittedly, this was not the first regression I tried, my initial plan was to create a model that predicted assists as a function of the same independent variables, which I would then compare to a baseline metric to determine if Westbrook’s statistics were unique in a statistically significant fashion. Unfortunately, my grasp of econometric theory was not strong enough to answer the question in a convincing manner, so I searched for a different method. By using this logit model, we can get a more direct answer of the effect of road games on Westbrook’s triple doubles. Now this initial model had some glaring flaws, luckily, they were easy fixes and not structural problems. The most obvious one is the fact that Westbrook’s points are being double counted in both his PTS variable and in the total TeamPPG variable.



By fixing that problem, the signs return to what we would expect. Instead of PTS being very slightly negative, it is now slightly positive, in line with our expectations from the theoretical model. The independent variable of interest, Away, has a negative coefficient as expected, suggesting that Westbrook is indeed less likely to log a triple double on the road than he is at home. However, the magnitude of this coefficient is much smaller than expected, also suggesting that the effects of home scorekeeping does not have a drastic impact on Westbrook’s play. Another interesting thing of note with the coefficients is how large the coefficients on ASTs and TRBs are relative to PTS, even though they all count towards the same goal. Understanding this phenomenon requires a deeper understanding of basketball, but it makes perfect sense. Since Westbrook averaged over 30 points a game, he had no problem reaching 10 points in any given game, so additional points don’t increase the likelihood of a triple double. On the other hand, Westbrook barely averaged over 10 assists and rebounds, so each additional assist or rebound was a much more substantial hurdle in reaching double digits.



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Away | USG | TeamPPG | PTS | AST | TRB | TeamPP~S |
| Away | 1.0000 |  |  |  |  |  |  |
| USG | 0.1638 | 1.0000 |  |  |  |  |  |
| TeamPPG | -0.2670 | -0.2192 | 1.0000 |  |  |  |  |
| PTS | 0.0469 | 0.6403 | 0.3202 | 1.0000 |  |  |  |
| AST | -0.2206 | -0.5204 | 0.4842 | -0.1628 | 1.0000 |  |  |
| TRB | -0.0180 | -0.2105 | 0.2090 | 0.0965 | 0.4218 | 1.0000 |  |
| TeamPPG-PTS | -0.2727 | -0.7299 | 0.6046 | -0.5611 | 0.5599 | 0.1015 | 1.0000 |

Conclusion

The goal of this paper was to uncover potential bias in NBA scorekeeping with respect to Russell Westbrook’s MVP season. The data analyzed in here has proven to be ineffective at demonstrating any potential bias. Instead, it supports the significance of Russell Westbrook’s season and why this has not been achieved in nearly 60 years. In the model used, there was a lot of noise in differences between home and away statistics.

Further efforts to answer this question would do good to normalize the data between road and away. Like how Bommel and Bornn (2016) scaled their assist data by team field goals made and their block data by opponent field goals attempted. Doing this would eliminate noise like difference in number of possessions between home and away games. While this paper may have failed to uncover any significant bias, the case remains open. Especially with the significance of box scores on so many aspects of the NBA, ranging from public perception, accolades, and even contract negotiations all heavily rely on the statistics being accurate.

**Works Cited**

Bommel, Matthew Van, and Luke Bornn. "The Van Exel Effect: Adjusting for Scorekeeper Bias in NBA Box Scores." *MIT Sloan*, March 11, 2016.

Craggs, Tommy. "The Confessions Of An NBA Scorekeeper." Deadspin. August 26, 09. Accessed November 24, 2017. <https://deadspin.com/5345287/the-confessions-of-an-nba-scorekeeper>.