Analysis of Potential Gender Bias in Athlete Performance Versus Salary in the NBA and WNBA

Yi Lei Feng

12/22/2020

Abstract

In this paper, we attempt to determine whether gender bias occurs in sports, focusing specifically on basketball and the two gender-specific leagues in North America: the WNBA and NBA. This will be accomplished through analyzing potential discrepancies in athlete salary numbers relative to their performance on the court assessed by key metrics. A propensity score matching approach was adapted in order to classify the differences between NBA and WNBA player groups, with the use of nearest neighbor as the matching technique. This was followed up with a multiple linear regression model conducted on the newly-matched outcome based on propensity scores. The linear regression model focused on mapping the relationship of our predictors variables based on performance and the outcome of interest, player salaries. A discussion regarding the findings of this model is detailed where we determine whether gender bias is a problem when it comes to athlete salaries and provide justification for why this may or may not be the case.

Introduction

When thinking about the most iconic legends in the sport of basketball, some of the athletes that come to mind include the likes of Michael Jordan, Kobe Bryant, LeBron James, Magic Johnson, Stephen Curry, the list goes on and on. It should come as no surprise that most of these legends in basketball are male, as the majority of professional sports leagues around the world feature male athletes exclusively and these leagues are the ones receiving extensive news coverage for large audiences. Gender inequality has been one of the most prominent global issues rooted in our culture and this problem sees similar biases when it

comes to the sports industry. However, the discrepancy in the gender ratio is not only present in athletes, but is even more egregious in coaching staffs, front office management, and media. Shockingly, a mere 3.2% of US sports media coverage is dedicated to female sports, one of the many reasons that highlight why this issue persists (Cooky, 2015). The purpose of this report is to evaluate the consequences of these gender biases beyond just the sheer ratio discrepancies, by analyzing factors such as how athlete performance metrics translate to salary numbers for both genders.

To meet our goal, it is not a practical approach to attempt to analyze a multitude of different sports, so instead this observational study will focus purely on basketball statistics for both men (NBA) and women (WNBA). The variable that is most indicative of a basketball player's worth is their salary, and this has been a skyrocketing trend as the rise of player salaries coincides with the ascension of the league salary cap over the history of basketball. To put this into perspective, the NBA salary cap started at a value of $3,600,000 in the 1984-1985 season and is now at $109,140,000 for this upcoming 2020-2021 season ("NBA Salary Cap History", n.d). Our secondary goal in this study will be to analyze which performance metrics contribute significantly as predictors to these rising player salaries. A recent study regarding NBA player salaries from The Sport Journal found scoring, rebounding, and personal fouls to be meaningful contributors to a player's value (Lyons Jr, 2018), but a very different set of predictor variables will be used in this study, based more around how the modern basketball game is evolving. More discussion regarding the player statistics that will be used as predictors is featured in the methodology section of the report. The hope is that after analyzing how the various performance metrics affect player salaries, any significant results for WNBA players versus NBA players can be identified and discussed.

In order to make a causal inference such as gender bias being a problem in basketball regarding player salaries, it seems appropriate that propensity score matching could be a viable strategy. Propensity score matching is a common method that pairs well with observational data, and having the treatment variable being whether or not a basketball player is in the NBA or WNBA would be a logical step to infer if gender bias is existent. However, comparing WNBA and NBA players alone with regards to their salary is not very meaningful analysis, so a good candidate to further explore the on-court performance of these two groups of athletes is to run a multiple linear regression model to pair with propensity score matching. This should provide us with a solid foundation of results for analysis to view which predictors stand out.

The datasets that will be used for this study is from Kaggle, containing statistics for WNBA and NBA players in various basketball metrics as well as Spotrac, for supplying the player salary numbers. In the following sections of this report, the first piece of description will feature the simulation and model that was run on the dataset as part of the methodology section. This includes a more in-depth description of the propensity score matching with the multi regression model that was previewed. Following this, a formatted version of the results will be displayed from the model simulation using tables/graphs and any significant statistical results are interpreted. Then, there will be a discussion section summarizing the simulation as well as providing justification for any interpretations of the results. The final

parts of the discussion will address weaknesses and future steps to be taken regarding this topic.

Methodology

Data

This report features an observational study with the majority of the data for WNBA and NBA athletes supplied from Kaggle. The exception was the player salary numbers, which was appended to each of the datasets with cap hit information provided by Spotrac. The player performance statistics were selected based on the 2019 season for the WNBA and NBA. The main reason behind this is that 2019 was the latest full basketball season before the pandemic year hit, which is important for providing us a fully representative sample size of player statistics for analysis. These samples notably included all "active" basketball athletes who played in the 2019 season, with an active player being defined as athletes who participated in a minimum of 2/3 of the games in the 2019 season. Having this requirement should help improve the precision of the data, as players who did not meet this minimum number of games played are likely to have been injured and would have been a misrepresented in their on-court performance.

As a brief overview of the data, let's go over some of the more essential data attributes for both the WNBA and NBA athletes separately. First, the NBA player data is shown below.

Table 1: NBA Player Data

Player	League	Position	PTS	3PT	TOV	PER	Salary
James Harden	1	G	36.1	4.8	5.0	30.6	0.2880183
LeBron James	1	F	27.4	2.0	3.6	25.6	0.2822718
Giannis Antetokounmpo	1	F	27.7	0.7	3.7	30.9	0.1948525

There are some modifications that were made to this data for these attributes that will now be explained. The "League" attribute is a binary classification that will come into play later on for propensity score matching. For NBA players, the league attribute will be set to 1, signifying that the player is in the NBA. A popular term that professionals use to describe the modern NBA is "positionless" basketball. What this means is essentially that coaches have become more and more creative in running 5 players on the floor who do not necessary fill the traditional basketball positional roles (e.g. PG, SG, SF, PF, C). An example of this can be playing 3 point guards on the floor at the same time or running a small-ball lineup with no center. This transitions into the main justification behind classifying the traditional basketball positions in the original data into three generalized positions: G (guard), F (forward), C (center). Three pointers made, points and turnovers are all metrics

measured on a per-game basis, representing the average number of 3-point field goals, points scored and turnovers committed by a player per-game. The PER metric stands for "Player Efficiency Rating", which is an overall rating system developed by NBA analyst John Hollinger, representing a player's per-minute statistical production across both negative (e.g. turnovers, poor shooting percentages, etc.) and positive (scoring, rebounding, assists, etc.) contributions ("Calculating PER", n.d).

Table 2: WNBA Player Data

Player	League	Position	PTS	3PT	TOV	PER	Salary
Elena Delle Donne	0	F	23.40	2.04	1.20	31.8	0.0900000
Brittney Griner	0	C	24.84	0.00	3.00	26.4	0.0884615
Jonquel Jones	0	F	17.52	1.56	2.28	24.3	0.0459369

For WNBA players, the league attribute is set to 0, indicating the player is part of the WNBA. Outside of this, the same modifications made to the NBA data also applies to the WNBA data as the variables are identical. An important distinction that must be noted is that the rule set of the WNBA specifies that games are played on 10 minute quarters, totaling a game length of 40 minutes ("Rules and Other Curiosities of WNBA", 1998) while the NBA is played on 12 minute quarters for a game length of 48 minutes. To make up for this difference, the WNBA player statistics were normalized by being scaled to adjust to a 48 minute game setting. This can be seen in the following example: WNBA player Jonquel Jones committed 1.9 turnovers per game in the 2019 season. To adjust to a 48 minute game setting, the following normalization is performed to scale the number by the hypothetical additional 8 minutes to match an NBA game: $1.9 + (1.9/40) * 8 = 2.28$ matching the value shown in Table 2. The normalization process is also required to be done for the 3PT and PTS categories in a similar fashion.

On a final note regarding data cleaning, the salary numbers for both NBA and WNBA athletes were converted from raw salary numbers to a percentage of the league's salary cap, representing the budget each of the league teams are able to spend on its roster. The reasoning for this is simply the gap between the NBA salary cap and WNBA salary cap is far too large, as the NBA cap comes out at a maximum of $109,140,000 ("NBA Team Salary Cap Tracker", n.d) while the WNBA salary cap is at $1,300,000 ("WNBA Team Salary Cap Tracker", n.d). To put this into perspective, the NBA salary cap is an absurd 84 times the WNBA salary cap and the league minimum salary that an NBA player with at least 1 year of experience can sign for is $1,445,697 (Adams, 2019), which alone is higher than an entire WNBA team's budget. To show the conversion, WNBA star Elena Delle Donne made a salary of $117,000 for the 2019 season, so the conversion to a percentage of the cap results in $117,000 / $1,300,000 = 0.09 or 9% of the cap. It should be quite clear that it is unrealistic to leave these raw player salary numbers untouched so it is a logical choice to make the conversion into a percentage of the cap to perform proper analysis.

Model

The main model that will be used in this study is **propensity score matching** to contrast the group of NBA players versus WNBA players followed by a **multiple linear regression model** to analyze the effects that chosen predictor variables has on the outcome of interest, basketball athlete salaries. This is broken down into two sections: the propensity score matching process using logistic regression and the multiple linear regression model afterwards.

Propensity Score Matching

In the first step, propensity score matching will be conducted where the treatment specifies whether a basketball athlete is playing for the NBA or WNBA, with 1 signaling the player is in the NBA and 0 representing that the player is in the WNBA. In order to accomplish this, the NBA and WNBA datasets that were discussed in the data section need to be merged together to form a collection of all player data. A logistic regression model will be used to estimate the propensity score. This is an appropriate model since we are estimating the likelihood of an athlete being in the NBA based on their on-court statistics, and this value will range between 0 and 1. The covariate variables used for this logistic regression include PTS (average points scored per game), FTA (free throws attempted), AST (average assists per game), 3PT (average 3-pointers hit per game) and FG (average field goals per game). As a quick justification, NBA players outperform WNBA players in nearly every statistical *counting* (non-percentage) category, due to the nature of the much higher scoring NBA games. These five selected categories represent the largest mismatches in these counting statistics and thus should serve well in estimating propensity scores of whether an athlete plays for the NBA. The propensity scoring estimation logistic regression model appears as follows:

$$log(\frac{y_i}{1 - y_i}) = \beta_0 + \beta_1 x_{PTS_i} + \beta_2 x_{FTA_i} + \beta_3 x_{AST_i} + \beta_4 x_{3PT_i} + \beta_5 x_{FG_i} + \epsilon$$

where y represents the likelihood of a particular player i being in the NBA, β_0 is the intercept of the model (likelihood of player i being an NBA player with no additional info provided), and $\beta_1 \ldots \beta_5$ are slopes for the PTS, FTA, AST, 3PT, and FG parameters respectively (i.e. an increase in any of the mentioned parameters causes a correlated change in the output likelihood).

Once the estimation of propensity scores is complete using the above logistic model, the nearest neighbor matching technique is applied to match an NBA player part of the treatment group to a WNBA player not part of the treatment group based on the similarity of their propensity scores. When the matching process is finished, the remaining goals are to evaluate the quality of the matching by comparing meaningful statistics between the two resulting groups before proceeding to evaluate the original outcome of interest in player salary numbers.

One of the main benefits behind using a propensity score matching approach is that it is crucial for reducing selection bias in observational studies. The matching helps make causal inference on the effect that being apart of the treatment group (NBA players) could have on the outcome of interest, their salary numbers, which is the purpose of the study. This will be highlighted in the next step of the process as multiple regression analysis will be introduced on the newly matched sample.

Multiple Linear Regression

Following the propensity score matching, a multiple regression model will be conducted on the final observed samples of the matching. The predictor variables selected that will be measured against player salaries for the linear regression include the following: player position, 3-point shooting, turnovers, player efficiency rating, and the binary treatment value from propensity score matching: whether a player plays in the NBA or WNBA. Each of these performance metrics were covered and described in the previous data section under Table 1, as well as the data cleaning process in the case where any of these variable values were adjusted. The overall regression model appears as:

$$y_i = B_0 + B_1 x_{Pos_i} + B_2 x_{3PT_i} + B_3 x_{TOV_i} + B_4 x_{PER_i} + B_5 x_{League_i}$$

where y_i = player i's salary, x_{Pos_i} = player i's position, x_{3PT_i} = number of 3-pointers made by player i on average per-game, x_{TOV_i} = turnovers committed by player i on average per-game, x_{PER_i} = player i's player efficiency rating, and x_{League_i} = whether player i plays in the NBA

Now it may be good to provide some justification behind why these particular performance metrics were picked as predictors. A lot of weight was assigned to basketball statistics that highlighted the strength of a player in the modern basketball game, which has evolved rapidly to become a much more fast-paced offensive style contest. Thus, this is something that all of the stats have in common, which is a heavy emphasis on efficiency in a player's game. For example, 3-point shooting is a highly valued skillset because players who are elite at this skill brings more efficient scoring to a game, as a 3-point shot is obviously worth more points than a normal 2-point field goal. Similar reasoning goes for turnovers and player efficiency rating, which both value players who contribute meaningfully to their teams on offense while remaining efficient to compensate for the modern basketball game. The more a basketball player excels in these metrics, the more efficient they are on offense which could potentially be a good translation to how much they are worth in salary. Lastly, the player position was thrown in as a more general variable to attempt to see if there were any meaningful results showing that players of a specific position were compensated higher, or if the modern "positionless" system reigns supreme.

A multiple linear regression model is an appropriate choice for this situation as the goal is to assess the direct linear relationship between these predictor variables and the response variable. This will be further supported in the next section through independent analysis measuring each of the factors individually to see its correlation with player salary. Ultimately, the hope is to identify if any distinction exists between the results when

running the model for the NBA treatment group versus the non-treatment WNBA group which could be signal that gender bias in salaries exist between the two.

Results

The first step of the procedure is to carry out the propensity score matching procedure, so the first thing done was combining WNBA and NBA athlete data into a single dataset. From this one dataset, the propensity score is able to be estimated based on the logistic regression model described in the model section. The command used to run the logistic regression is `glm()` with `family = binomial` and the propensity scores are then computed using the `predict` function passed with the flag `type = response` required for logistic models. The graph below will display the results of the estimated propensity scores separated by the original treatment group:
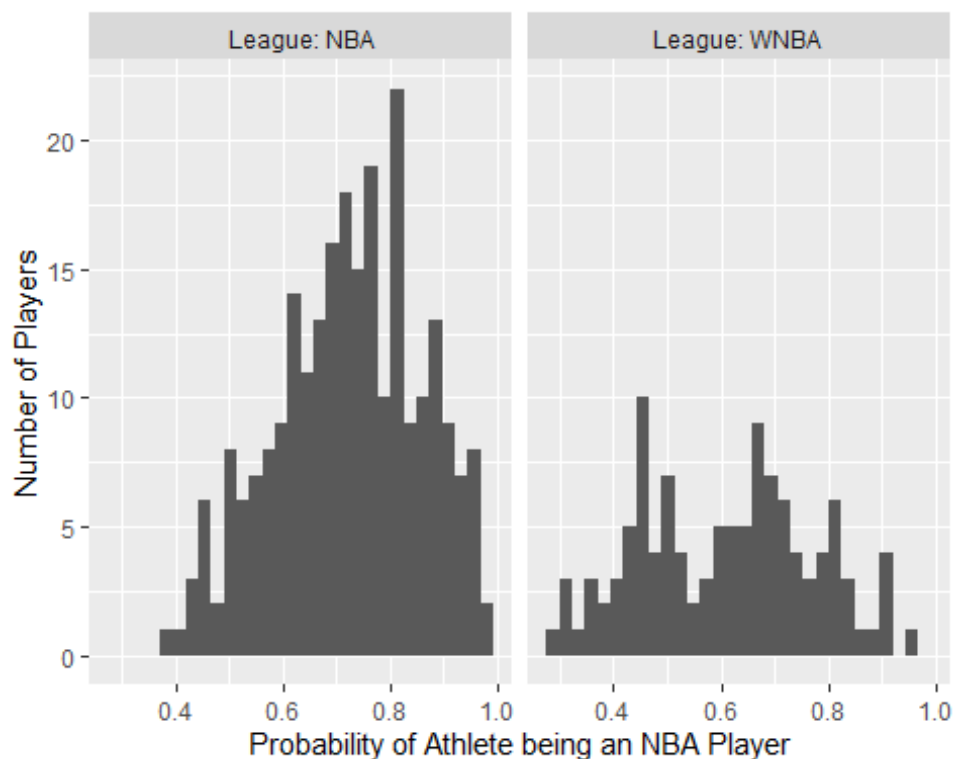
From these side-by-side plots, we can see that while the estimated propensity scores is not a perfect prediction of whether a basketball athlete is correctly classified into the treatment group of being an NBA player, it serves as a good baseline as the majority of NBA players have fairly high propensity scores while there is a much wider range of propensity scores for WNBA players. Some of the WNBA players with higher propensity scores comparable to the actual treatment group indicates that they impressively perform at an extremely high

level in the performance metrics in the WNBA league, at a similar level to their NBA counterparts.

Table 3: Matched Observational Data

	Player	League	Pos	PTS	3PT	TOV	PER	Salary	Prop Score
110	Brittney Sykes	0	G	12.24	0.96	2.40	11.3	0.0412023	0.8454412
111	Betnijah Laney	0	G	6.72	0.72	1.68	8.7	0.0557692	0.7904111
112	Kaela Davis	0	G	7.20	0.60	2.64	4.6	0.0367677	0.4653874
113	James Harden	1	G	36.10	4.80	5.00	30.6	0.2880183	0.8185645
114	LeBron James	1	F	27.40	2.00	3.60	25.6	0.2822718	0.9794014
115	Giannis Antetokounmpo	1	F	27.70	0.70	3.70	30.9	0.1948525	0.8886863

Using the estimated propensity scores from the logistic regression model, the next step was to perform matching using the nearest neighbors technique. As there was an uneven number of data points that received treatment over no treatment (i.e. there were more NBA players than WNBA players in the original data), the matched dataset output contained 224 observations, a notable reduction from the original dataset as there were only 112 non-treatment observations (number of WNBA players) that could be matched to treatment observations (NBA players). An additional column in the newly matched data set also contains the estimated propensity scores.

The next step is to proceed to evaluate the outcome of interest by running our multiple regression model on the matched dataset. We will begin by analyzing the correlation of each of our predictors independently against a player's salary to get a visualization of how much each metric stacks up. The first variable to be measured is a player's position.

Figure 1: Player Position vs Salary
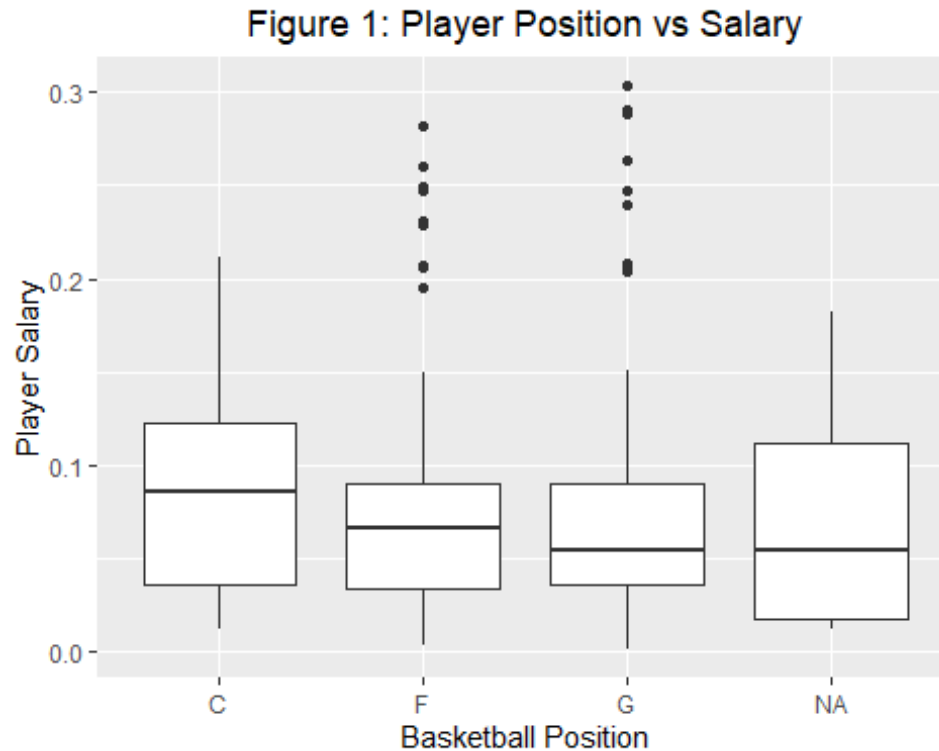
The player position versus salary relationship actually shows some interesting results, contrary to expectations. The expectation coming in was that a player's position would have a negligible effect on salary earned as there should be evenly talented players valued relatively the same across all roles on average but this is not exactly the case as the plot shows. The top-end players at the forward and guard positions earn a substantially higher percentage of the cap compared to the top-end centers. When doing some research around the NBA, perhaps this is justified as the small number of franchise players for teams are dominated by players who handle the ball the most, which happens to be at these two positions. Overall, the majority of players across the NBA and WNBA are paid a similar percentage of the cap, outside of these top-end talents.

Figure 2: Three Pointers Made vs Salary
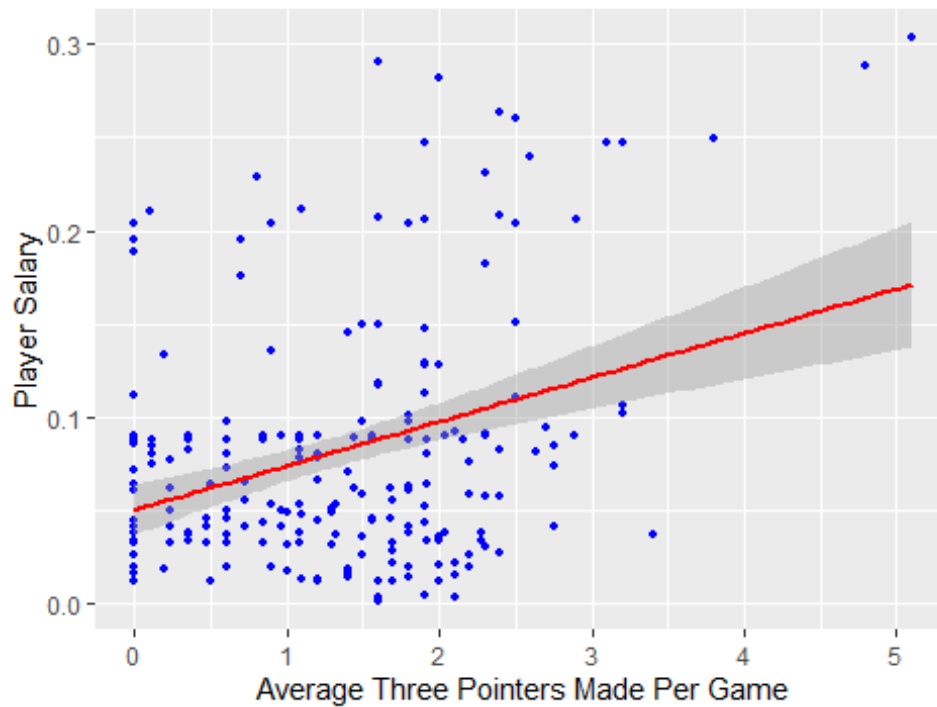
For the most part, three pointers made by a player has a less significant, but slightly positive correlation with a player's salary. This is not too surprising especially considering our previous relationship with centers earning the lowest salaries, as they are often the one position that does not shoot the ball from deep as often. Given that three-point shooting is only a specific skillset in a multi-skilled game on offense, specializing in just one skill alone does not automatically translate to higher salaries (but it is certainly a plus), as we can see from the high number of data points that were fit below the slope line.

Figure 3: Turnovers Committed vs Salary
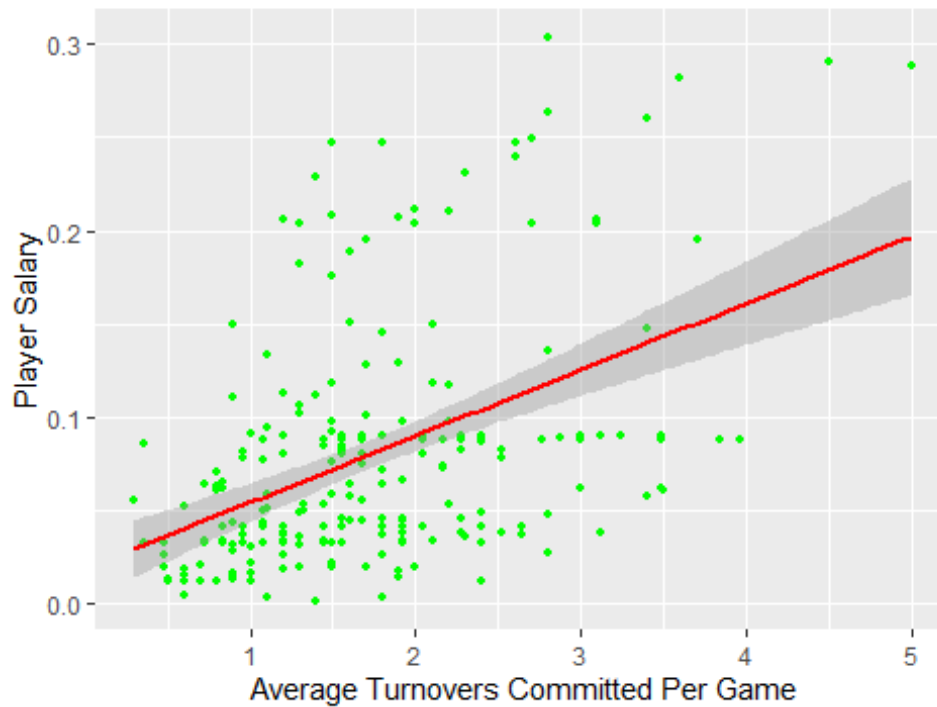
At first glance, the results from the graph appear to be contradicting intuition, as turnovers are seen as a negative stat for basketball players. Yet, it produces a seemingly very positively-correlated line of best fit. To explain this without considering the use of more advanced basketball analytics, generally the best players on a team will have the highest usage, or in other words, have the ball in their hands for the highest percentage of time. Having high usage rates will translate to higher turnovers as these star players are often the ones trying to run plays for their team. Of course, higher-usage star players should also be the ones that make the highest salaries relative to the team's cap. Given this context, it is far more reasonable that the turnovers and player salaries have a strong positive correlation.

Figure 4: PER vs Salary
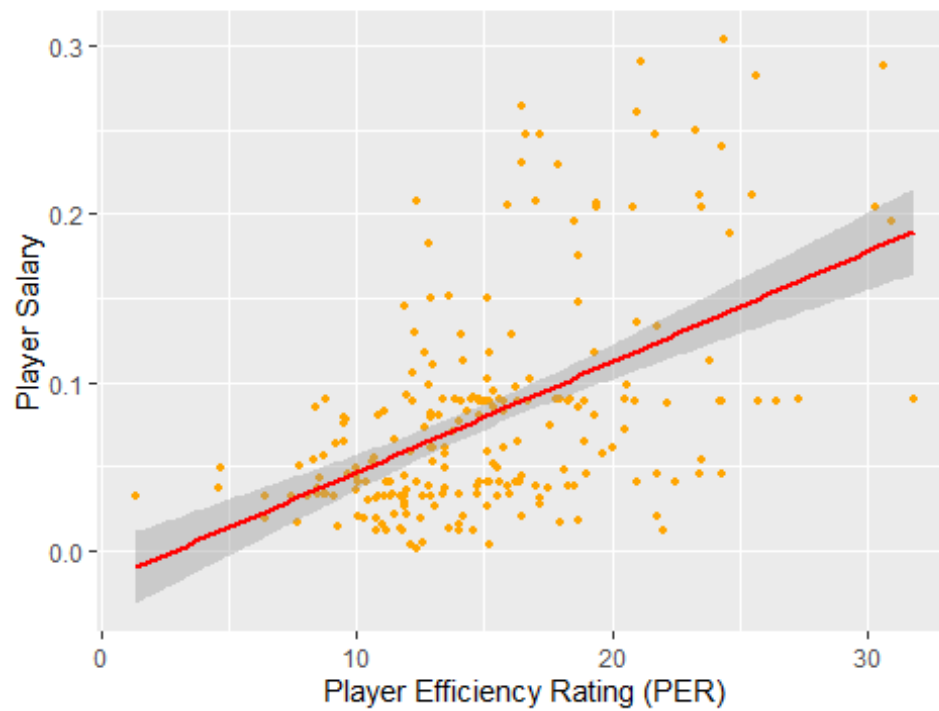
This appears to be one of the strongest predictors so far, as having higher PER is indicative of a player earning a higher salary as well. This is one of the most well-designed advanced basketball statistics that is widely used across both national leagues to evaluate players, so it is good to see players be rewarded for performing well in this metric.

Figure 5: League (Treatment Variable) vs Salary
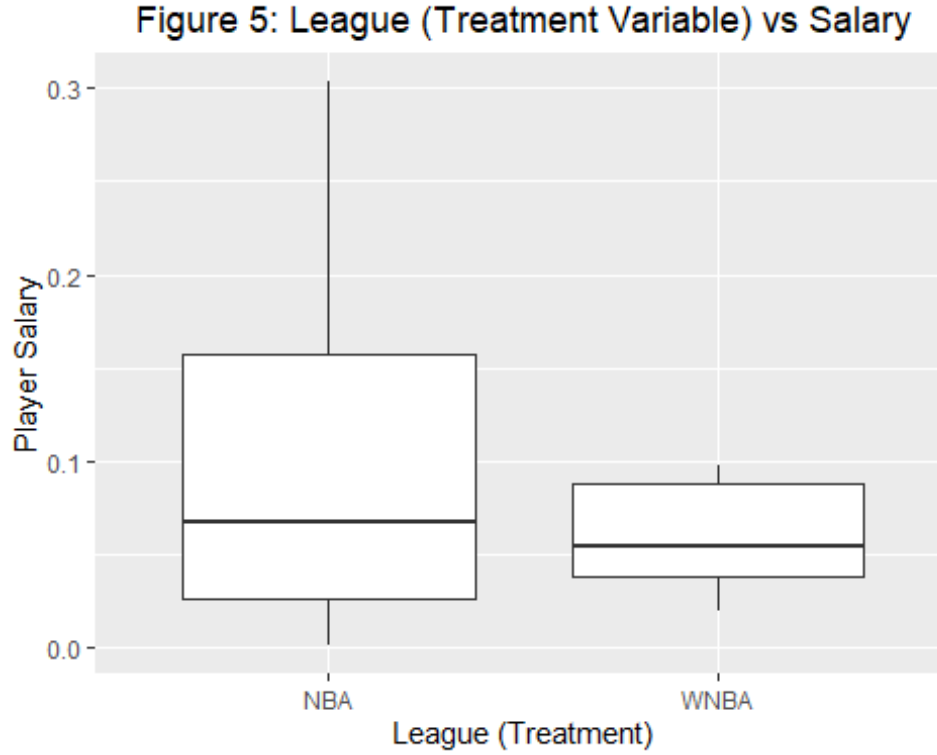
The result that we are most interested in is measuring the treatment variable (whether a player plays in the NBA or WNBA) against their salary. The results do indeed show that a high volume of NBA players are paid a much higher percentage of the cap when compared to WNBA players. The fact that the salary numbers were converted to a percentage of the cap masks how significant this difference truly is. Whether or not NBA players are justified to be earning much more than WNBA players is a topic for the discussion section.

By running a full regression analysis using lm(), our multiple regression model can be summarized by the following formula:

$$y_i = -0.0210637 - 0.0220043x_{PosF_i} - 0.0358648x_{PosG_i} + 0.0192840x_{3PT_i} + 0.0242272x_{TOV_i} + 0.0032668x_{PER_i} + 0.0219923x_{League_i}$$

where y_i = salary for player i, x_{PosF_i} = whether player i is a guard, x_{PosG_i} = whether player i is a forward, x_{3PT_i} = average 3-pointers made by player i per game, x_{TOV_i} = average turnovers committed by player i per game, x_{PER_i} = player i's PER, and x_{League_i} = whether player i plays in the NBA.

The fact that all these slope values are minimal is expected due to our conversion of player salaries to a percentage of the league cap, so slope values are expected to remain small. From the equation, we can see that the most significant predictor contributors to salary are a player's position, turnovers and our treatment variable of whether or not they play in the NBA.

Discussion

Summary

The intended goal of this observational study is to identify whether gender bias is present in the sport of basketball when it comes to how athletes are paid. The methodology selected to be used on the dataset was a propensity score matching approach to isolate NBA players from WNBA players with whether or not a basketball player is in the NBA being the binary treatment variable. In this process, a logistic regression model was required to find a propensity estimation between 0 and 1, representing the likelihood of a specific player being in the NBA based on how they performed in various basketball statistical categories. During this estimation, we learned that there were quite a number of WNBA players who performed comparably to NBA players, evident by their high propensity scores. For the next step, we matched treatment observations (NBA players) with non-treatment observations (WNBA players) based on nearest propensity scores. The result was a reduced dataset, but with matching pairs of data with a 1:1 ratio of female to male athletes in the remaining data. With this new dataset, a multiple linear regression model was run to estimate a player's salary based on the following predictors: basketball position, 3-point shooting, turnovers, Player Efficiency Rating (PER), and the binary treatment variable of whether a player plays in the NBA or WNBA. All of these predictors showed varying levels of correlation with a player's salary, and the results of the regression showed a significant effect between how much NBA athletes part of the treatment group make compared to WNBA athletes. More specifically, the slope of the treatment variable is 0.0219923, meaning that if a player is in the NBA, they will on average earn up to approximately 2.2% more of the salary cap than WNBA players, which becomes a significant amount when taking into account the NBA's $109,140,000 salary cap.

Conclusion

Outside of the binary treatment value, some observations can be made regarding the other predictor variables when it comes to estimating player salary. In terms of the predictors that measured on-court performance, all the predictors can be grouped and classified as efficient offensive statistics. Turnovers, 3-point shooting, and PER were the three categories part of this group and they all showed solid correlation with how much a player is worth, supporting the theory that the modern offensive game has put higher value on these metrics. For reference, the NBA has evolved to shoot almost twice the number of 3-pointers compared to 15 years ago ("Evolution of the NBA offense by Geography", 2020). More significantly, because we performed the regression analysis on the combined WNBA and NBA datasets with propensity matching, there is a sizable proportion of WNBA players who show strong performance relative to their counterparts when it comes to their on-court performance. Despite this finding, we have identified that the average WNBA player earns 2.2% of the salary cap less than NBA players, and this difference can be completely blown out of proportion when we do not normalize salary numbers relative to the league cap. Therefore, a conclusion can be reached that there exists gender bias when it comes to how professional basketball athletes are compensated in their respective leagues.

Perhaps some of these differences can be justified by the economics behind the leagues. The WNBA simply does not have anywhere near the resources that the NBA does, with a revenue of $60 million while the NBA generates $7.4 billion per year, leading to an average salary of $71,635 compared to $6.4 million (Jope, 2019). Factors that contribute to the NBA's market include an international presence, national broadcasting, ticket sales, higher ratings and much more, all which significantly favor the NBA. Hopefully this illustrates just how large a landslide advantage the NBA has when it comes to marketing. What may be confusing is that this gap is only widening and no real solutions have been attempted to bridge this gap in gender inequality between the leagues. One of the most important elements when it comes to boosting WNBA player recognition is media coverage, but clearly not much effort has been put forth in this aspect as we have already learned that only 3.2% of U.S media coverage is devoted to female sports (Cooky, 2015). Even though this observational study masked some of these issues by representing player salaries as a percentage of the cap, it does not help matters when NBA players are entitled to 50% of the league's revenue while WNBA players are limited to only 20% of the revenue (Jope, 2019). Overall, despite some WNBA athletes playing well enough on the court to justify earning the same percentage of the cap as NBA players, gender bias and external league circumstances appear to prevent this from being the case.

Weaknesses

The first thing to note is that during the propensity score estimation, the covariates used to predict the outcome binary treatment variable was not optimized so it may not be the ideal prediction of whether an athlete should be classified as an NBA or WNBA player. The same problem also exists when applying the multiple regression model to find the best predictors for the salary response variable. It is very difficult to list every single covariate that is related to both the treatment as well as the overall outcome variable, as nearly every statistic would have to be considered in that case in which the model grows far too complex. However, this does not mean that there may not be potentially more optimal covariate predictors that may show stronger correlation with the treatment or outcome variable, as the predictors selected in this study were based on the criteria that the modern basketball game was more tailored to favor strong offensive stats. Thus, experimenting with combinations of other variables as predictors is an interesting alternative option to consider.

One of the limitations of this study is that there are unaccounted cases for which player performance is affected from external circumstances. The most prevalent component of this is when players are traded mid-season, in which teams value players differently and players will have different roles which affects their performance throughout the season. However, this is a safe assumption to ignore for the context of this report as it is impossible to determine how each team evaluates players. This is also a fairly uncommon scenario relative to the total number of players so the overall effect on the analysis is minimal. In addition, another issue that comes up with younger athletes who are playing on rookie-scaled or entry-level contracts in the NBA and WNBA, who are capped to earning a smaller percentage of the salary cap. This leads to a common scenario that arises, with these young

players often outperforming their contracts as they are among the lower-tiers of earnings for basketball athletes in the league, leading to potential outlier data points. Despite this problem, there is no clear solution as player statistics cannot be normalized simply due to high-performing young players being locked in to their entry-level contracts. For more context, this is not a sport-specific problem, as other top leagues across the globe all have rookie athletes who are restricted to lower salaries to start their careers.

Finally, there is the issue of how to evaluate other potential impact of female versus male basketball athletes outside of pure salary discrepancies. In this study, the scope was limited to analyzing how female and male basketball players differed in how they were paid according to their performance on the court. While this study partly succeeded in identifying that there does appear to be some type of gender bias, there are other factors that go beyond just an athlete's salary such as historical influence, popularity, and impact on the sport. There is no doubt that these are all important elements to consider when searching more into the gender bias problem in sports, but it is hard to incorporate all these aspects into a single model for analysis and so a decision was made in this study to solely focus on how much basketball athlete salaries.

Next Steps

As a continuation to this last problem, an intriguing future possibility to take is to perform a set of analysis with the outcome variable being focused on the popularity (number of fans) of top male and female athletes worldwide. Potential relevant predictor variables for this type of study can include historical achievements, championships, level of competition, and so on. This seems to be a good outcome of interest for analysis because it gives us a different perspective on how male and female athletes compare with respect to overall influence, as the number of fans an athlete has is often indicative of their historical impact in their respective sport. If the findings of this study provides us with the expected results that the top male athletes are dominant in this outcome, it provides strong additional support that gender bias is indeed problematic in the world of sports and steps are needed to be taken to address the issue.

References

2020-2021 Hollinger NBA Player Statistics - All Players. (n.d.). Retrieved from http://insider.espn.com/nba/hollinger/statistics

Adams, L. (2019, June 29). NBA Minimum Salaries For 2019/20. Retrieved from https://www.hoopsrumors.com/2019/06/nba-minimum-salaries-for-201920.html

Calculating PER. (n.d.). Retrieved from https://www.basketball-reference.com/about/per.html

Cooky, C., Musto, M., & Messner, M. (2015). A Quarter Century of Excluding Women's Sports in Televised News and Highlight Shows. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/2167479515588761

Jope, C. (2019, July 18). NBA vs WNBA: Revenue, Salaries, Attendance, Ratings. Retrieved from https://www.wsn.com/nba/nba-vs-wnba

Lyons Jr, R., Jackson Jr, N., & Livingston, A. (2018, July 13). Determinants of NBA Player Salaries. Retrieved from https://thesportjournal.org/article/determinants-of-nba-player-salaries/

NBA Player Earnings. (n.d.). Retrieved from https://www.spotrac.com/nba/rankings/2019-20/cap-hit/

NBA - Salary Cap History. (n.d.). Retrieved from https://basketball.realgm.com/nba/info/salary_cap

NBA Team Salary Cap Tracker. (n.d.). Retrieved from https://www.spotrac.com/nba/cap/

Pinak, P. (2020, July 09). The NBA's Longest Game Took 6 Overtimes and Lots Of Standing Around. Retrieved from https://fanbuzz.com/nba/how-long-is-a-basketball-game/

R Tutorial 8: Propensity Score Matching. (n.d.). Retrieved from https://sejdemyr.github.io/r-tutorials/statistics/tutorial8.html

Strupp, J. (2020, November 05). WNBA and NBA player comparisons; league salaries. Retrieved from https://www.kaggle.com/therealpumpkin/wnba-and-nba-player-comparisons-league-salaries

WNBA Salary Rankings. (n.d.). Retrieved from https://www.spotrac.com/wnba/rankings/2019/cap-hit/

WNBA Team Salary Cap Tracker. (n.d.). Retrieved from https://www.spotrac.com/wnba/cap/