**Robbing Peter to Pay Paul George and His Teammates:**

**The Effect of an Increase in Salary Cap on Salary Distribution Among Teams in the NBA.**

James Caron

Southern Utah University

*In this paper I analyze the effect of increasing the salary cap in the NBA on salary distribution within teams. Using NBA player salary data from 1990 to 2018 and accounting for CBAs, and holdout years I use interaction terms in OLS regression to find the difference in difference to show where the salary is distributed when the salary cap increases by more than the average of 6%. I conclude that the distribution of salary is equal among players when the salary cap is increased, though these findings are inconclusive and further research may be required.*

In the NBA there is one thing that prevents the league from being pay-to-win: salary cap. Before the 1984-85 season, NBA teams could spend any amount of money they wanted on players. This created a system commonly known today as pay-to-win: whichever owner felt like spending the most money to get the best players, would probably win. Thus, in the year 1984 the salary cap was introduced to level the playing field among teams. The salary cap amount is determined year to year by the league’s collective bargaining agreement (CBA) and is based on the league’s revenue. Unlike many other leagues, the NBA has a soft salary cap which means there are many exceptions that permit teams to go above the salary cap. Because the league has increasingly become more about the individual players than the teams, these exceptions were introduced to incentivize players to stay on their respective teams. This is done to foster and cultivate fan support for each team. Since the salary cap was introduced, it has been steadily increasing by an average of 3.2 million dollars per year while having only one year where it decreased, and two years where it remained the same. The question I seek to answer is: “Where is salary distributed among players in the NBA when the salary cap is increased?”

In 2001 J. Richard Hill and Peter A. Groothuis conducted a study to test whether the newest CBA (at the time) in the NBA improves the distribution of salaries in the NBA rather than further skewing it. Using the median voter model- which is used as a standard to analyze union behavior- and salary and performance data on players- they concluded that the CBA did even the distribution of salaries among players in the NBA. In 2012 James R. Hill did another study, this time with Nicholas A. Jolly to find how changes made to the NBA CBAs between 1990 and 2008 altered the salary distribution and experience-earnings profile of players. Using yearly statistics for NBA players between 1989 and 2008, salary figures between 1990 and 2008, and biographical data they used three empirical techniques: Shorrocks R, kernel density estimates, and wage regressions. They found that the changes to the CBAs did modestly even the distribution among players in the NBA. Though these studies are very similar to the question I am trying to answer, I am attempting to look at the effect of the salary cap being raised rather than the effect of CBAs. In order to do this, I will need to control for the years where a new CBA was introduced.

In 2005 Lawrence M. Kahn and Malav Shah did a study on race and pay in the NBA for the 2001-2002 season using data that included contract details and player performance. Using multiple models in linear regression, they concluded that for players who were neither free agents nor on rookie scale contracts, there were statistically significant ceteris paribus nonwhite shortfalls in salary, total compensation, and contract duration. However, for players under the rookie salary scale (first-round draft picks) and free agents, race effects were small and insignificant. Though this study isn’t directly aligned, this shows evidence for a need to control for race, and whether a player was under their rookie contract or previously a free agent. Unfortunately, I was unable to find a database that includes the race of all players from 1990 to 2018. In order to find out why there was no race data, with some help I was able to get into contact with Omeed Selbe, an analytics consultant in Los Angeles who has worked with sports and racial data. Selbe's research indicates that while no one is prohibited from collecting or releasing racial data at the individual level, this data tends to be much more useful in aggregate. Due to sports analytics' focus on player performance and maximizing wins, racial data becomes less applicable to most analyses. However, I will control for the year of their contract, and I believe this will proxy for these variables appropriately.

**Data**

The data used for this analysis was player data that uses players each year as the unique identifiers, and included salary data, and controls for: the player’s year in the NBA (1st, 2nd etc), whether or not a new CBA was introduced that year, whether or not the season was a season that increased the salary cap by more than 6%, and whether there was a holdout that delayed the season. The reason I chose 6% is because it’s roughly the average amount the salary cap increased per year within the timeframe of my data, and it splits the very evenly at 50%. Though half the time the salary cap did increase by more than 6% and half the time it didn’t, this does not mean all my data is split 50/50 due to the variance in player count year to year. However, this shouldn’t have any significant effect on my study. The data includes all NBA players that were listed on each team each year from 1990 to 2018. The salary data was taken from <https://data.world/datadavis/nba-salaries> which is ultimately sourced from basketball-reference.com. It was created three years ago by user @datadavis and was most recently updated five months prior to this research. This data provided the player id, salary for that player, the season that the salary applied (start and end), and the team the player was on. I added another variable that listed the salary cap for season for each observation. This was done manually by filtering in excel and copying the associated salary cap found on basketball-reference.com. Doing this manually creates a higher risk for errors, but I am confident none were made in this process. I also created the rest of the control variables in Rstudio. Some players in the data were tied in salary, so when I created the variable that ranked players I rounded down. For example, if 2 players were tied for the top paid player on a team, they were ranked 1.5 and the top\_player variable would have a 1 for both of them. There is variance in the distribution of player’s salaries can be seen by the summary statistics and the histogram below:



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Summary Statistics for Salary Variables** | | | | | |
| Variable | Obs | Mean | Std. Dev. | Min | Max |
| Player Salary Adjusted | 12,929 | 4,277,603 | 5,031,235 | 4,078 | 51,800,000 |
| Salary Cap Adjusted | 12,929 | 57,600,000 | 19,800,000 | 22,800,000 | 101,000,000 |
| Team Total Salary Adjusted | 12,929 | 69,300,000 | 26,200,000 | 14,500,000 | 163,000,000 |

These salary numbers are adjusted for 2018 dollars using the average CPI listed from the [U.S. Department of Labor Bureau of Labor Statistics.](http://www.bls.gov/)

**Model & Results**

In testing my question of “Where is salary distributed within teams when the salary cap increases year to year?” I started with this simple model:

Where ln(salary) is the natural log of players salary adjusted to 2018 dollars. Where is the average log of salary for a not top player in a year when the salary cap doesn’t increase. Where *cap\_increase* is a 1 if the salary cap raised by more than 6% that year and a 0 if it wasn’t. will tell us what the average percentage increase in player salary when the salary cap is increased by more than 6% compared to when it isn’t for not top players. Where *top\_player* is a 1 if the player was a top paid player and a 0 if they weren’t. will tell us the average percentage increase in player salary for a top player compared to other players when there isn’t a salary cap increase above 6%. And where is my interaction term that will tell us the difference in difference in salary for a top player in a year where the salary cap was increased by more than 6%, compared to the salary of a top player in a year where the salary cap wasn’t increased by more than 6%, compared to the salary of a not top player in a year where the salary cap was increased by more than 6%, compared to the salary of a not top player in a year where the salary cap wasn’t increased by more than 6%. These results can be seen below in table 1:

|  |  |
| --- | --- |
| **Table 1** | |
| OLS Regression Estimate of the Distribution of Player Salary when the Salary Cap Increase by More Than 6% | |
| Dependent Variable: Log of Player Salary Adjusted for Inflation | |
| Independent Variables |  |
| 6% Increase (1=yes) | -0.202\*\*\* |
| [0.0218] |
| Top Player (1=yes) | 1.901\*\*\* |
| [0.0604] |
| Top Player Interaction Term | 0.0344 |
| [0.0860] |
| Constant | 14.58\*\*\* |
| [0.0155] |
|  |  |
| Observations | 12,929 |
| R-squared | 0.139 |
| Note: Standard errors are shown in brackets. P-values are indicated as follows: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | |

The most surprising finding in these results is that this model tells us, with significance at the 99% level, that the average player salary decreases by 20% when the salary cap goes up by more than 6%. In addition, the model also tells us that a top player makes an average of 190% more than a not top player. Unfortunately, the interaction term isn’t significant, and therefore can be relatively ignored for now as I am going to run a more complex model. It can be inferred that these results are due to the nature of averages. It makes sense that a top player makes 190% more on average than a not top player, and my 6% increase variable supports the hypothesis that the distribution gets worse when the salary cap is increased. It does not conclude it. It is important to consider that this model only explains 14% of the variation in salary, and my interaction term isn’t significant, but the fact that these variables are significant should not to be ignored.

For my second model I added variables for different groups of player salary ranks and control variables.

Before I used these in my regression, I tested to see that each of these grouping of player salaries were statistically significant. From table 2 (presented below) we can see that each of them are:

|  |  |  |
| --- | --- | --- |
| **Table 2** | | |
| Player Salary Rank Variables | | |
| Ranks | Observations | Statistically Different Salaries? |
| Total | 12929 | -- |
| Top Player | 6% | Yes\*\*\* |
| Top 2 Players | 13% | Yes\*\*\* |
| Top 3 Players | 19% | Yes\*\*\* |
| Starting 5 | 32% | Yes\*\*\* |
| Note: P-values are indicated as follows: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | | |

The interpretability of our model changes so that is the average log salary for non-starting players in a year where there wasn’t a holdout, and under the first CBA. The interpretation of remain the same as from the prior model, except for instead of comparing to not top players, it compares to not starting players. The interpretation of are the same as except instead of comparing the top player to not starting players, it compares top 2 players, top 3 players, and starting players to not starting players. The interpretation of are all interaction terms that are the same as except instead of interacting *top\_player* with *cap\_increase*, they interact variables *top2\_players*, *top3\_players*, & *starting\_players*. The results of this model can be seen in Table 3 below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 3** | | | | | | | |
| OLS Regression Estimate of the Distribution of Player Salary when the Salary Cap Increase by More Than Average | | | | | | | |
| Dependent Variable: Log of Player Salary Adjusted for Inflation | | | | | | | |
| Independent Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 6% Increase (1=yes) | -0.202\*\*\* | -0.198\*\*\* | -0.197\*\*\* | -0.165\*\*\* | -0.0936\*\*\* | -0.270\*\*\* | -0.268\*\*\* |
|  | [0.0218] | [0.0201] | [0.0201] | [0.0190] | [0.0194] | [0.0576] | [0.0574] |
| Top Player (1=yes) | 1.901\*\*\* | 0.300\*\*\* | 0.300\*\*\* | 0.220\*\*\* | 0.233\*\*\* | 0.236\*\*\* | 0.238\*\*\* |
|  | [0.0604] | [0.0655] | [0.0655] | [0.0619] | [0.0609] | [0.0605] | [0.0602] |
| Top Player Interaction Term | 0.0344 | 0.0282 | 0.0290 | 0.0621 | 0.0531 | 0.0485 | 0.0481 |
| [0.0860] | [0.0930] | [0.0930] | [0.0878] | [0.0863] | [0.0858] | [0.0853] |
| Top 2 Players (1=yes) |  | 0.244\*\*\* | 0.244\*\*\* | 0.231\*\*\* | 0.232\*\*\* | 0.232\*\*\* | 0.229\*\*\* |
|  |  | [0.0662] | [0.0661] | [0.0625] | [0.0614] | [0.0610] | [0.0607] |
| Top 2 Players Interaction Term |  | -0.00210 | -0.00285 | 0.00858 | 0.0124 | 0.0147 | 0.0147 |
|  | [0.0936] | [0.0936] | [0.0885] | [0.0869] | [0.0864] | [0.0860] |
| Top 3 Players (1=yes) |  | 0.363\*\*\* | 0.363\*\*\* | 0.327\*\*\* | 0.335\*\*\* | 0.336\*\*\* | 0.337\*\*\* |
|  |  | [0.0570] | [0.0570] | [0.0539] | [0.0530] | [0.0526] | [0.0524] |
| Top 3 Players Interaction Term |  | -0.0344 | -0.0343 | -0.0314 | -0.0367 | -0.0350 | -0.0349 |
|  | [0.0810] | [0.0810] | [0.0765] | [0.0752] | [0.0747] | [0.0743] |
| Top 5 Players (1=yes) |  | 1.457\*\*\* | 1.457\*\*\* | 1.258\*\*\* | 1.302\*\*\* | 1.307\*\*\* | 1.305\*\*\* |
|  |  | [0.0357] | [0.0357] | [0.0341] | [0.0336] | [0.0334] | [0.0332] |
| Top 5 Players Interaction Term |  | 0.0384 | 0.0380 | 0.0704 | 0.0809\* | 0.0703 | 0.0698 |
|  | [0.0507] | [0.0507] | [0.0479] | [0.0471] | [0.0468] | [0.0466] |
| Holdout Year (1=yes) |  |  | 0.0873\*\*\* | 0.0503 | 0.119\*\*\* | -0.161\*\*\* | -0.168\*\*\* |
|  |  |  | [0.0326] | [0.0308] | [0.0336] | [0.0588] | [0.0588] |
| Player's Year |  |  |  | 0.0846\*\*\* | 0.0691\*\*\* | 0.0671\*\*\* | 0.0675\*\*\* |
|  |  |  |  | [0.00215] | [0.00225] | [0.00226] | [0.00226] |
| Fixed Effect (CBA) | No | No | No | No | Yes | Yes | Yes |
| Fixed Effect (Year) | No | No | No | No | No | Yes | Yes |
| Fixed Effect (Team) | No | No | No | No | No | No | Yes |
| Constant | 14.58\*\*\* | 14.12\*\*\* | 14.11\*\*\* | 13.73\*\*\* | 13.42\*\*\* | 13.49\*\*\* | 13.48\*\*\* |
|  | [0.0155] | [0.0142] | [0.0144] | [0.0168] | [0.0227] | [0.0737] | [0.0734] |
|  |  |  |  |  |  |  |  |
| Observations | 12,929 | 12,929 | 12,929 | 12,929 | 12,929 | 12,929 | 12,929 |
| R-squared | 0.139 | 0.470 | 0.470 | 0.527 | 0.543 | 0.550 | 0.555 |
| Note: Standard errors are shown in brackets. P-values are indicated as follows: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. This Table displays the results of seven OLS regression models. Each model controls for additional factors that could contribute to variance in salary distribution among players. | | | | | | | |

From these results I can first see that in the end did not end up changing much at all, which is again surprising. However, our interpretation is slightly different this time. This time shows us that non-starting players’ salary goes down by an average of 27% in years where there wasn’t a holdout and the first CBA was in place. Though theoretically they would be on their 0th year of playing, which disrupts the interpretability a little bit. The second thing to notice is that though all the player rank variables were significant, none of the interaction terms were. For the player rank variables this means that player’s salaries are on average 24%, 23%, 34%, & 130% for top, top two, top three, and starting players respectively compared to non-starting players in years where the salary cap didn’t increase by more than 6%. Though the 130% seems as though there should an error, this is because each variable is comparing to the average of all the other players. In fact, these variables aren’t really that interpretable because of this. Another interesting note about these results is that the control variables had relatively little effect on the overall R-squared. This makes sense because it is much more valuable to know if a player ranks (according to pay) in the top 5 players and if so which rank in trying to predict how the % of salary they have on the team. The year, team, and current CBA is much less useful intuitively. We can also see that according to this model, for each additional year a player plays, their salary increases by 6.75%. Unfortunately though, this isn’t important in regards to the goal of this analysis, and the variables that are important, are all insignificant.

**Conclusion**

In conclusion our findings show that when we control for the CBA, year, team, the player’s year in the NBA, and whether it was holdout year, the distribution of salary is not uneven among the top 5 players compared within each other, and compared to those who don’t start (not top 5 players). It is possible that the players distribution of revenue isn’t even among the players who don’t start, but I didn’t research that because it’s value of importance isn’t as nearly as high as the starting players.

Due to the complexity of NBA contracts and pay, there are many implications in these findings, and many places where further research may be needed to come to anything conclusive. We did find that certain player ranks do get paid more than others on average, but the interpretability of this limited due to the way our model is set up. This could be fixed, but the assumptions of OLS would have to be reevaluated. I understand that the actual amount of money a player earns is much more complicated than just their salary with the numerous amounts/types of bonuses. For the purpose of my analysis I chose to ignore this because it becomes much more complicated. The data on bonuses is much harder to find, and how I would track/control for it would have to be evaluated as well. I assumed that salary would be a close proxy for any bonuses received, and therefore thought it safe to exclude bonuses from my analysis. The next steps I would take in continuing this study would be to control for bonuses and see if that makes any of my interaction terms significant. One of the biggest errors in this analysis that was not accounted for was that these models were predicting player salary of the same year when the salary cap was increased by more than 6%. When in fact, a player’s salary isn’t affected until their contract expires. In order to accurately do this analysis I would need to find contract data that includes when their contract expired, and when their new one was negotiated, and do a lot of data manipulation to get it to merge nicely with the data I currently have.

**Work Cited**

Hill, J. R., & Groothuis, P. A. (2001). The new NBA collective bargaining agreement, the median voter model, and a Robin Hood rent redistribution. *Journal of Sports Economics*, *2*(2), 131-144.

Hill, J. R., & Jolly, N. A. (2012). Salary distribution and collective bargaining agreements: A case study of the NBA. *Industrial Relations: A Journal of Economy and Society*, *51*(2), 342-363.

Kahn, L. M., & Shah, M. (2005). Race, compensation and contract length in the NBA: 2001–2002. *Industrial Relations: A Journal of Economy and Society*, *44*(3), 444-462.