2017

Salary Inequality in the NBA: Changing Returns to Skill or Wider Skill Distributions?

Jonah F. Breslow
Claremont McKenna College

Claremont McKenna College

Salary Inequality in the NBA: Changing Returns to Skill or Wider Skill Distributions?

submitted to

Professor Ricardo Fernholz

by

Jonah F. Breslow

for

Senior Thesis

Spring 2017

April 24, 2017

Abstract

In this paper, I examine trends in salary inequality from the 1985-86 NBA season to the 2015-16 NBA season. Income and wealth inequality have been extremely important issues recently, which motivated me to analyze inequality in the NBA. I investigated if salary inequality trends in the NBA can be explained by either returns to skill or widening skill distributions. I used Pareto exponents to measure inequality levels and tested to see if the levels changed over the sample. Then, I estimated league-wide returns to skill. I found that returns to skill have not significantly changed, but variance in skill has increased. This result explained some of the variation in salary distributions. This could potentially influence future Collective Bargaining Agreements insofar as it provides an explanation for widening NBA salary distributions as opposed to a judgement whether greater levels of inequality is either good or bad for the NBA.

Acknowledgements

This thesis is dedicated to my parents, Patrice and Adam Breslow, who have always selflessly supported me in every possible way.

I would also like to thank Professor Ricardo Fernholz for his dedication to his students and for making the thesis process both enjoyable and enriching. I thank Professor Heather Antecol for all her help in this process.

All my professors and coach who made my experience at CMC so invaluable: Professor Eric Helland, Professor Cameron Shelton, Professor Ananda Ganguly, Professor Lisa Meulbroek, Professor Manfred Keil, Professor Winston Ou, Professor Gary Gilbert, Professor James Kreines, and Coach Charlie Griffiths.

I would like to shout out the suite: Phil, John, Mac, Alex, Ari, Graham, and Matt.

I have to thank everyone in Poppa who helped me learn how to use LaTeX or else this thesis would never have been completed.

Finally, I want to acknowledge both my biological brother, Ted Breslow, and my non-biological brothers, Austin Aleman and Micah Ornelas, for our unfaltering friendships.

Contents

List of Figures

List of Tables

1 Introduction

Income and wealth inequality have become topical issues in economics recently – not only politically but academically as well – as seen in Piketty and Saez (2003) and Atkinson et al. (2011). Piketty (2014) becoming a New York Times and Amazon best seller indicates that this issue extends beyond academics into peoples' lives. Many political movements have been centered around reducing the trend of increasing inequality, of which Bernie Sanders' 2016 presidential campaign was the most recent and influential.

In this paper, I examine long-term trends of NBA salary distributions. I formally test whether salary inequality has been statistically increasing since the 1985-86 season. Further, I examine if the NBA has experienced changes in returns to skill. While I do not intend to answer the moral question of whether inequality is either good or bad, I propose a method to determine whether inequality can be explained by skill. Moral judgements about inequality are subjective I have no intention of arguing one way or another – I simply provide an explanation for salary distributions in the NBA.

I find that salary inequality in the NBA increased from the 1985-86 season to the 2000-01 season. Beyond the 2000-01 season, salary inequality has not statistically changed. Further, I provide evidence that this phenomenon can be explained by a widening skill distribution. I propose that there are three main mechanisms that potentially explain NBA salary distributions. The three mechanisms are:

- Higher returns to skill — if returns to skill increase and can explain more variation in salaries, then the best athletes will earn more while the less skilled athletes will likely earn less.

- Wider skill distributions — if skill distributions widen, *ceteris paribus*, then the more skilled athletes will earn more and the less skilled athletes will earn less.

- Secular trends — there may be natural trends towards greater inequality or other factors that leads to wider salary distributions.

2 Literature Review

2.1 Income Inequality and Salary Caps in the NBA

Atkinson et al. (2011) highlight the rising levels of income inequality in most countries, including the United States. Many sports leagues, in line with this general trend, have also experienced an increase in salary inequality. Salary distributions have been a topic of contention in many sports leagues including the National Basketball Association (NBA). While there were lockouts in 1995 and 1996, the 1998, 2005, and 2011 lockouts were significant because they all concluded in the genesis of new CBAs. The 1998 CBA was widely considered, at the time, to favor team owners and mid-level athletes because it cut costs in various ways, most notably by placing limits on individual player salaries and on the share of revenue that could go to the athletes. This was the first time the NBA placed limits on athlete salaries. While this agreement had certain exemptions that allowed for exceeding salary limits, these were not the main facets of the CBA. Furthermore, if teams did exceed the salary cap, they were forced to pay a "luxury tax" for violating the limit.[1]

Hill and Groothuis (2001) examined the effects of the 1998 CBA on salary inequality in the NBA. Prior to the 1998 CBA, Hill and Groothuis (2001) found consistently increasing inequality. Without any limit on maximum salaries, the top athletes in the NBA commanded much higher salaries. Hill and Groothuis (2001) found that the 1998 CBA effectively decreased salary inequality in the NBA in the following 1999-2000 season.

In the 1998 CBA, the percent of revenue that could go to the athletes was capped at 48%. However, the 2005 CBA increased this share to 51%. Along with other provisions of the agreement regarding contract length and specific exemptions, the new 2005 salary cap helped teams retain all-star players without being penalized by the luxury tax. Provisions in the 2011 CBA did not change individual salary limits. Slight alterations were made regarding players transferring to and from the NBA Development League and all-star rookie salary caps, but rules that

[1]This paragraph summarized what was presented in Beck (2011).

would affect a large contingent of the NBA were not drastically altered. Essentially, the 1998 CBA shifted the NBA towards a de facto policy of egalitarianism and neither the 2005 nor the 2011 CBA significantly changed this.[2] However, the question of whether this shift towards egalitarianism was made because superstars were overcompensated and average athletes under-compensated or because some general sentiment towards egalitarianism seemed more attractive is still left to be answered.

As labor markets in professional sports have become liberalized and as veteran free-agency has become common, athletes have begun earning more. Scully (2004) examines the effects that veteran free-agency has on athlete compensation in the MLB, NBA, NFL, and NHL. He finds that, with free-agency, athletes not only earn more, but they command a higher proportion of franchise earnings as well. In other words, profits move from the owners to the athletes.

2.2 Racial Inequality

The question of race-based salary inequality in the NBA has been a topic of research since the late 1970s. Mogull (1977) examined the salaries of 28 players, equally black and white, and found no statistically significant difference in earned salaries between races controlling for individual performance adjusted by minutes per game. One of the main issues of this study was its extremely small sample size. Kahn and Sherer (1988) reexamined the race-wage question by looking at salaries for the 1985-86 NBA season. Contrary to Mogull (1977), they found black players earned around 20% less than white players. Hill (2004) examined salary data from the 1990s and found that white players earned more than their black counterparts because, on average, white athletes were 2 inches taller than black athletes. The coefficient on race variables was insignificant because it was correlated with the height variable, which was significant itself.

Jenkins (1996) took a different approach; he examined only negotiated free-agent salaries. He used free-agent salary data from 1983 to 1994 — approximately 370 athletes — and found that the return to increased performance was the same

[2] All the information presented in this paragraph is derived from Coon (2011).

for both races. In other words, the free-agent market treats both black and white athletes equally. He finds, however, that players are evaluated differently based on race; this is not salary discrimination in the traditional sense, but does indicate racial discrimination in the NBA during the 1990s. When regressing log salary on different performance measures with a binary variables for race, Jenkins (1996) finds that athletes' salaries are statistically influenced by different factors depending on their race. For example, black athletes' salaries are statistically influenced by their total career time on the court while white athletes' salaries are not.

2.3 This Paper

In this paper, I aim to add to the existing literature by 1) documenting and testing salary distributions over the long-term 2) determining the returns to skill over the sample period and 3) examining if returns to skill affect salary distributions. This approach will append Hill and Groothuis (2001), as I intend to explain why salary inequality occurs, not the effects that specific contracts have on inequality.

3 Data and Methods

I use individual player data collected from Basketball-Reference.com in five-year intervals from the 1985-86 season to the 2015-16 season.[3] From the data, I use each athlete's PER, age, and salary. In each season, there are athletes for which Basketball-Reference.com does not have salary data. Furthermore, in each season, there are athletes who are paid under the league minimum for a variety of reasons. All of these athletes, for whom I lack salary data or who earned less than the league minimum, are omitted from the sample. In the 1985-86, 1990-91, 1995-96, 2000-01, 2005-06, 2010-11, and 2015-16 seasons 37, 53, 48, 34, 29, 34, and 46 observations are omitted, respectively, for the aforementioned reasons. In percentage terms, approximately 13.7% of the 1990-91 season data is omitted, which was the season that saw the most observations nullified by these criteria.

[3] I used Basketball-Reference (2017) to get all of the data used in this paper. It will take exploration into each athlete in each season to find his salary data, but the other variables are easily found by clicking the desired season, hovering over the 'Player Stats' option, and then clicking 'Advanced.'

Furthermore, I omit athletes who had fewer than 15 total minutes of play time because measuring their performance proved to be quite inaccurate. In the 1985-86, 1990-91, 1995-96, 2000-01, 2005-06, 2010-11, and 2015-16 seasons 9, 5, 7, 8, 13, 7, and 4 observations are omitted, respectively, because the athlete played fewer than 15 minutes. Overall, there are 2,639 observations. The 1985-86, 1990-91, 1995-96, 2000-01, 2005-06, 2010-11, and 2015-16 seasons have 279, 331, 376, 397, 420, 410, and 426 observations, respectively.

3.1 Salary and Rank

In this study, all of the regressions are run using $ln(Salary)$ as the dependent variable. I ranked each NBA athlete by his salary such that the highest paid athlete was rank 1 and the lowest paid player was rank n, where n is the number of athletes in the NBA in the given season.

There are many instances where multiple athletes earn the same salary. In this case each player is not given the same rank, the ranking between these players is arbitrary. I do this because when I regress $ln(Salary)$ on $ln(Rank)$, it is accurate to account for the fact that there are groups of athletes who earn the same salary. Consider if the data had five groups of twenty athletes and in each group the athletes earn equal salaries, but each 'better' group's salary is 10% greater than the next 'worse' group's. If each group is considered the same rank, it is misleading to say that as an athlete moves from rank 2 to rank 1 his salary increases by 10% because the athlete would actually need to move from, for example, rank 26 to rank 20 to earn that 10% salary increase.

3.2 Player Efficiency Rating (PER)

Before explaining how I cleaned the data based on the Player Efficiency Rating (PER), I will explain what the PER is. The PER is a standardized rating of player skill; every season, the league average PER is set to be 15.00. This statistic was created by John Hollinger, who currently is the Vice President of Basketball Operations for the Memphis Grizzlies, to be an all-encompassing rating of player skill and efficiency. In Hollinger's own words, "the PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance."[4] While understanding the specific formula for the PER is interesting, the purpose of this paper will not be aided by a detailed explanation of the PER.[5]

The PER adds a player's positive achievements — field goals, free throws, total rebounds, offensive rebounds, defensive rebounds, assists, steals, and blocks — subtracts negative achievements — missed field goals, missed free throws, and turnovers — and weighs each achievement differently. The calculation of the weights is beyond the scope of this paper. To get a per minute efficiency rating, the sum of all the positive and negative achievements are then weighted by minutes played. This yields the unadjusted PER (uPER). The final adjustment that must be made to the uPER is for team pace. Pace is defined to be the number of possessions per game. If a team has a higher pace then the uPER will be adjusted downward and vice versa. This equalizes the each player's per minutes opportunities for both positive and negative achievements. Some teams, such as the Brooklyn Nets, move the ball much faster and have more average possessions per game while other teams, like the Utah Jazz, move the ball slowly and have fewer possessions per game.[6]

A major criticism of the PER is that it largely favors offensive players. Since many defensive skills are not as easy to quantify as offensive skills, the PER is biased towards offensive players. Hollinger even admits, "Bear in mind that PER is not the final, once-and-for-all evaluation of a player's accomplishments during the season. This is especially true for defensive specialists – such as Quinton Ross and Jason Collins — who don't get many blocks or steals."[7] However, the PER, perhaps more accurately than any other statistics, incorporates many aspects of an athlete's game and successfully indexes them such that comparing any two athletes — across any season — can be done quickly and accurately, without

[4] Basketball-Reference

[5] For a more comprehensive look at how the PER is calculated, see Basketball-Reference

[6] See ESPN (2017) for NBA team pace statistics.

[7] In Hollinger (2011), Hollinger explains what the PER is and the downsides to using it as a performance metric. However, the reason why it is considered perhaps the best perfomance metric is becuase it "allows us to unify the disparate data on each player we try to track in our heads"

having to weigh certain statistics against others. Since the PER is the variable I am using for skill, from here onwards I will refer to PER distributions as skill distributions.

3.3 Age

Ideally, I would have access to data on years in the NBA for each player within each season of interest. However, this data was not made available by Basketball-Reference.com. Instead, I use player age in place of years in the league. Using age as a control may address the issue in which veteran players have higher salary minimums, regardless of skill. So, if two players with the same PERs have different veteran statuses, they will have different minimum salaries. Indeed, it appears that years in the league will have a relationship with salaries paid. The assumption I make is that age is closely correlated with years in the league. After I ran the regressions, I did find that age explained some of the variation in $ln(Salary)$.

3.4 Data Summary

In Table 1 below, the summary statistics for Salary, Log Salary, PER, and Age are shown for all seasons.

First, note how mean Salary and mean Log Salary have been increasing steadily in every season in the sample. While this indicates nothing regarding the distributions of salaries, it does show us that NBA athletes, on average, earn more today than they did earlier in the league's history.

Recall, according to Hollinger, the PER is supposed to have an average of 15.00 every season. In every season I examine in this study, the average PER is under 15.00. This means that many of the athletes I omit from the data set had PERs that were in the right tail of the PER distribution. However, a PER of 15.00 is within one standard deviation of every season's mean PER.

Table 1: Summary Statistics by Season

Season		Observations	Salary ($)	Log Salary	PER	Age
	Mean		472,007	12.9	14.7	27.0
1985-86	Median	279	265,000	12.5	13.6	26.0
	SD		352,296	0.6	4.0	3.3
	Mean		876,790	13.4	13.7	26.9
1990-91	Median	331	700,000	13.5	13.3	26.0
	SD		656,271	0.8	4.2	3.7
	Mean		1,745,097	14.0	13.2	27.3
1995-96	Median	376	1,319,000	14.1	13.1	27.0
	SD		1,683,824	1.0	4.4	3.9
	Mean		3,479,283	14.6	13.3	27.8
2000-01	Median	397	2,250,000	14.6	12.9	27.0
	SD		3,543,836	1.0	4.6	4.6
	Mean		4,121,180	14.7	13.1	26.9
2005-06	Median	415	2,586,164	14.8	12.6	26.0
	SD		4,096,187	1.0	4.8	4.3
	Mean		4,694,811	14.9	13.2	26.6
2010-11	Median	415	3,000,000	14.9	13.0	26.0
	SD		4,614,811	1.0	4.7	4.2
	Mean		5,050,944	14.9	13.7	26.6
2015-16	Median	426	2,880,600	14.9	13.5	26.
	SD		5,228,593	1.0	4.7	4.4

Note that the mean and median of Log Salary are essentially equal for each season. This indicates the the distribution of salaries is log normal, which means that the distribution of salaries is skewed right. In other words, the right tail of the salary distribution is stretched and there are more observations farther away from the mean salary in the right tail than in the left tail.

4 Empirical Methods and Results

In section 4.1 I discuss trends in the NBA salary distribution and utilize the Pareto exponent to determine levels of salary inequality. I then test each season's Pareto exponent against each other season's Pareto exponent to test for growing inequality.

In section 4.2 I discuss the trends in PER distributions. In section 4.3 I use regression analysis to determine each season's return to skill and test for changes in return to skill over the sample period.

4.1 Growing Salary Inequality

4.1.1 Salary Distributions

Since the 1985-86 NBA season, there has been a notable increase in salary inequality. Ranked by their salaries, the top 1% and the top 10% of NBA athletes have seen only a modest increase in their share of total salaries paid by the NBA. For example, the top 1% of athletes in the 1985-86 season was paid approximately 6% of the total salaries paid while the top 1% of athletes in the 2015-16 season was paid approximately 5% of total salaries paid, a single percentage point decrease. The top 10% of athletes saw a four percentage point increase in share of salaries paid moving from approximately 31% to 35%. However, this increase is not of the same nature as what was found in Atkinson et al. (2011) as the top 1% did not see a large increase in their share of salaries paid. This is illustrated in Figure 1 below.

The groups that saw larger increases in their share of total salaries paid were the top 25% and top 50% of NBA athletes. The top 25% commanded approximately 56% of total salaries paid in the 1985-86 season while in the 2015-16 season, they earned approximately 64% of the total salaries paid, an eight percentage point increase. The top 50% of athletes earned a similar increase in their share of total salaries paid. In the 1985-86 season, they earned approximately 80% of the total salaries paid and in the 2015-16 season they earned approximately 86% of total salaries paid, a six percentage point increase. This is illustrated in Figure 2 below.

Figure 1: Share of Salaries Paid to Top 1% and 10% of NBA Athletes

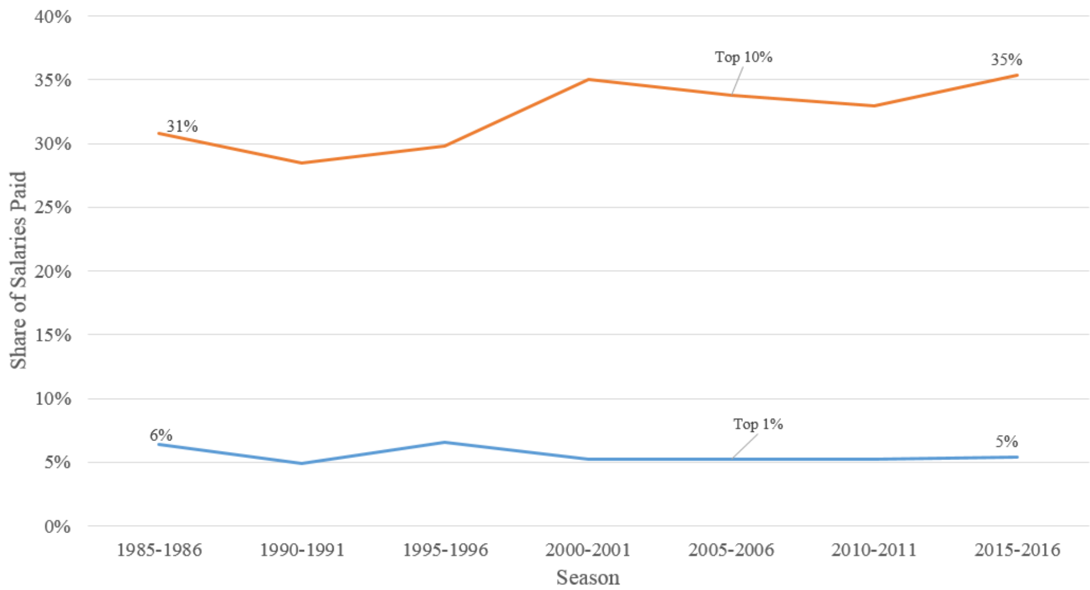
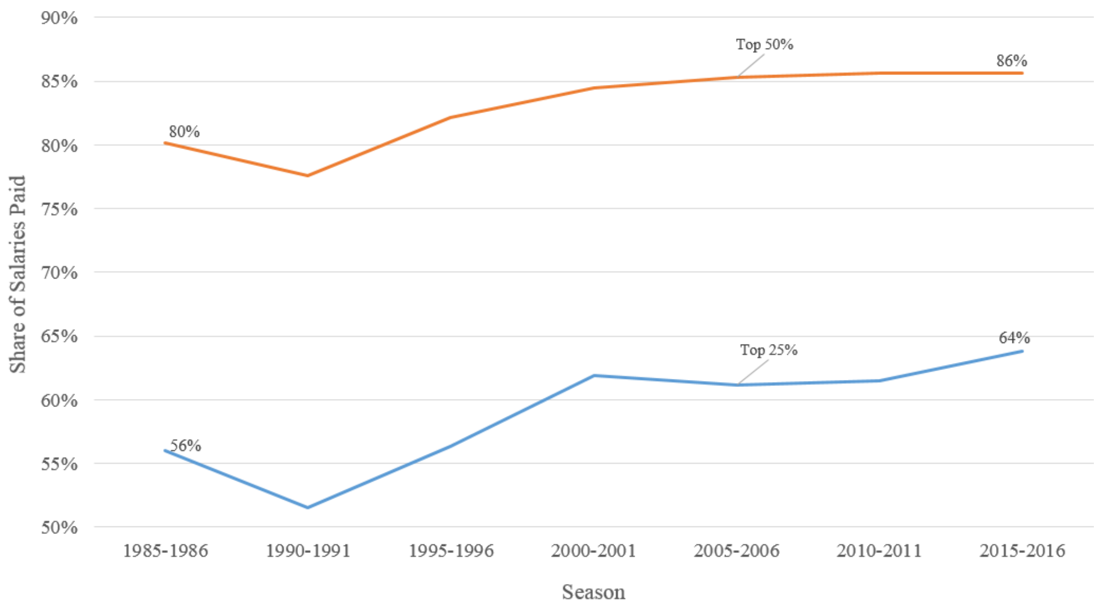


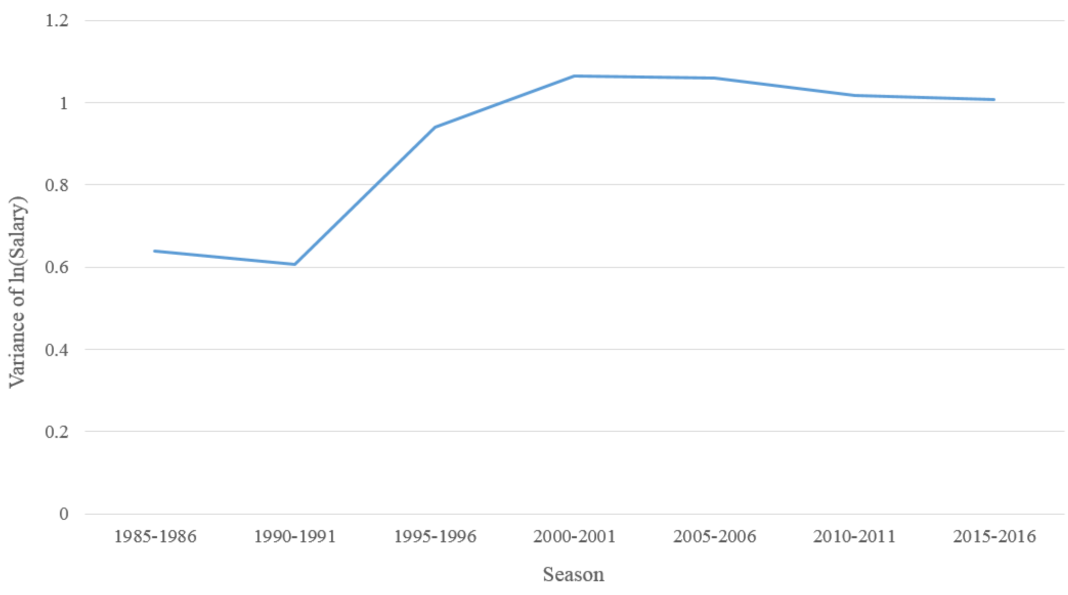
Figure 2: Share of Salaries Paid to Top 25% and 50% of NBA Athletes



From Figures 1 and 2, it becomes apparent that the largest gains in share of salaries paid went to athletes between the top 10% and top 25%. Athletes between the top 25% and 50% experience the next largest gain. Athletes between the top 1% and top 10% saw the third largest gain. The top 1% was the only group that experienced a loss in their share of total salaries paid. In other words, this shows that the top 10-50% of athletes saw gains in their salaries, not the top 1%. This is likely because many of the highest ranked athletes cannot have salaries that exceed salary limits as discussed in the literature review.

Furthermore, it will be useful to examine trends in log salary. In Figure 3 below, I have plotted variance of log salaries over time. It becomes apparent that variance of the log salary distribution increased over the period between the 1985-86 and 2000-01 seasons. After the 2000-01 season, the variance of log salaries does not appear to have changed significantly; if anything, it appears to have shrunk.

Figure 3: Variance of Log Salary: 1985-86 Season to the 2015-16 Season



This suggests that after the 2000-01 season, each season's salary distribution actually tightened implying that salary inequality likely did not grow from the 2000-01 season onward.

4.1.2 Power Laws and the Pareto Exponent

One way to quantify inequality is by looking at the relationship between $ln(Salary)$ and $ln(Rank)$:

$$ln(Salary)_{Sn} = \alpha_{Sn} + \zeta ln(Rank_{Sn}) + \epsilon_{Sn} \quad (1)$$

By regressing $ln(Salary)$ on $ln(Rank)$, one gets a power law relationship. This power law illustrates that given a relative change in rank, there is a commensurate relative change in salary scaled by ζ. What determines the relationship between the relative changes is the ζ coefficient on $ln(rank)$. This coefficient, also known as the Pareto exponent, can be interpreted as follows: for a one percentage change in rank, the associated percentage change in salary is ζ.[8]

Since I have ranked the data such that the highest paid athlete is rank 1, ζ will necessarily be negative because as rank decreases, salary increases. It is important to note that more negative $\zeta's$ indicate a greater level of inequality. For example, if ζ were -1, then a 1% decrease in rank would indicate a 1% increase in salary. However, if ζ were -2, then a 1% decrease in rank would indicate a 2% increase in salary. In other words, the more negative the Pareto exponent is, the more the top athletes earn compared to athletes ranked lower than them.

Using power laws and the Pareto exponent, I can statistically test if salary inequality is increasing in the NBA. In Table 2, I have provided a table of ζ coefficients and in Table 3 I present a matrix of statistical test for if every subsequent season's ζ coefficient, $(\zeta_{Season+5}, \zeta_{Season+10}, \ldots, \zeta_{2015})$, is greater than ζ_{Season}.[9]

In order to estimate the Pareto exponent for each season, I ran the following regression:

$$ln(Salary)_{Sn} = \alpha + \zeta_{1985} ln(Rank_{1985}) + \zeta_t ln(Rank_{1990}^{2015} \times Season) + \gamma(Season) + \epsilon_{Sn} \quad (2)$$

[8]Gabiax (2016) provided the inspiration to measure salary inequality by using power laws and Pareto exponents. This method is a simple and accurate way to evaluate salary inequality and changes in salary inequality.

[9]For the full regression output, see data appendix 6.1

By interacting Rank and Season, I was able to run a single regression. The ζ_{1985} is the Pareto exponent for the 1985-86 season. Each coefficient from ζ_{1990} to ζ_{2015}, represented by the ζ_t coefficient on the interacted $Rank_{1990}^{2015} \times Season$ vector, must be added to ζ_{1985} to yield the respective season's Pareto exponent. In Table 2, you will find all the ζ coefficients.

Table 2: Pareto Exponents

Season	1985	1990	1995	2000	2005	2010	2015
ζ_{Season}	-1.76***	0.15	-0.20	-0.41***	-0.25	-0.36**	-0.40***
	(0.08)	(0.12)	(0.13)	(0.15)	(0.18)	(0.15)	(0.14)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

The Pareto exponent for the 1985-86 season is -1.76. The Pareto exponents for the 1990-91 and 1995-96 seasons are not statistically less than the 1985-86 season. The 2000-01, 2010-11 and 2015-16 Pareto exponents are all statistically less than the 1985-86 season at the 1% level. The 2005-06 Pareto exponent is statistically less than the 1985-86 season at the 10% level. The R^2 of this regression is 0.8611, which means that 86.11% of the variation in $ln(Salary)$ can be explained by this model. In other words, this model is relatively accurate in explaining $ln(Salary)$.

To extend the testing beyond the 1985-86 season, I ran F-tests between the Pareto exponents from the 1990-91 season to the 2015-16 season. I converted the F-statistics to t-statistics by taking the square root and then assigning the correct sign based on the regression coefficients.[10] In Table 3, you can find the results from all of these tests.

Table 3: Left-Tailed Testing of Pareto Exponents

	1990	1995	2000	2005	2010	2015
1985	1.17	-1.50*	-2.78***	-1.33*	-2.42***	-2.73***
1990		-2.51***	-3.64***	-2.08**	-3.28***	-3.60***
1995			-1.36*	-0.27	-1.04	-1.31*
2000				0.80	0.27	0.00
2005					-0.57	-0.76
2010						-0.22

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 3 reiterates what was stated above: barring the 1990-91 season, every

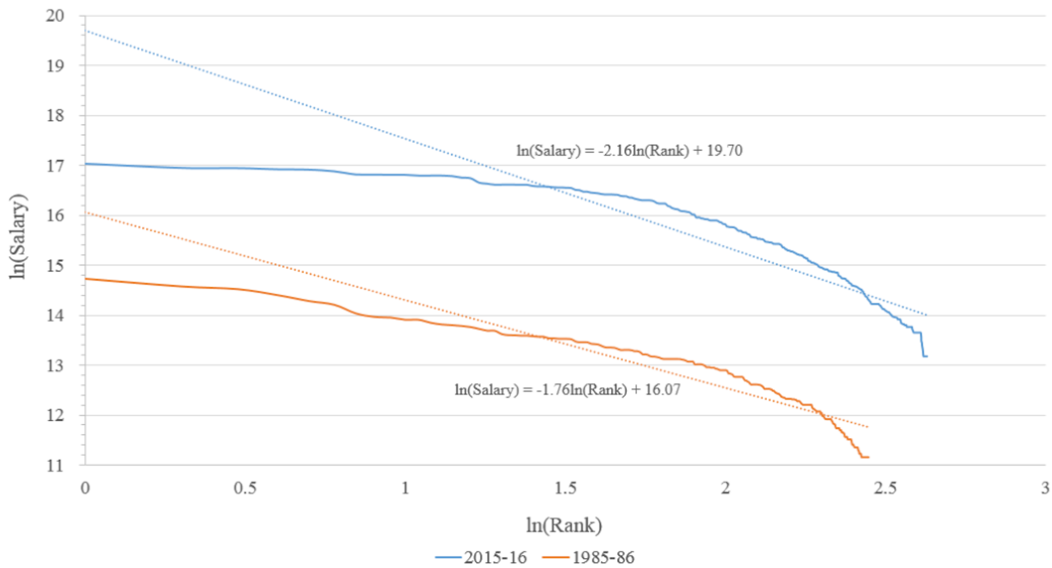
[10] I was able to convert the F-statistics to t-statistics because each F-test had only 1 degree of freedom.

season after the 1985-86 season had a statistically significantly smaller Pareto exponent than the 1985-86 season. Similarly, the 1995-96, 2000-01, 2010-11, and 2015-16 seasons had significantly smaller Pareto exponents than the 1990-91 season at the 1% level. The 2005-06 Pareto exponent was significantly smaller than the 1990-91 Pareto exponent at the 5% level. The 2000-01 and 2015-16 seasons had significantly smaller Pareto exponents at the 10% levels. From the 2000-01 season onwards, there was no statistically significant changes in the Pareto exponent.

These results indicate that over the period from the 1985-86 season to the 2016-17 season, inequality increased significantly. However, from the 2000-01 to the 2016-17 season inequality has not changed significantly. In other words, salary inequality widened from the 1985-86 season until the 2000-01 season but from the 2000-01 season onward, inequality has not significantly changed.

In Figure 4 below, I present the Pareto distributions for each season of the study. This graph visually presents the relationship between $ln(Salary)$ and $ln(Rank)$. While it is not extremely visually striking, it is apparent that the slope of the 2015-16 OLS, ζ_{2015}, is more negative than that of the 1985-86 season, ζ_{1985}. This indicates that there is more salary inequality in the 2016-17 season than in the 1985-86 season.

Figure 4: Pareto Distributions: 1985-86 and 2015-16 Seasons



4.1.3 Conclusions: Salary Distributions

These results show there exists a statistically significant trend in which inequality is increasing in the sample. I have demonstrated that the Pareto exponents have been decreasing since the 1985-86 season but have not continued a statistically significant decline since the 2000-01 season. Perhaps the conclusion that inequality has not changed since the 2000-01 season corroborates Hill and Groothuis (2001), who find that the 1998 CBA had the effect of curbing increasing inequality in NBA salaries. Notwithstanding the trends from the 2000-01 season onward, there still exists a statistically significant increase in salary inequality from the 1985-86 season to the 2015-16 season.

4.2 Growing Skill Disparity

In this section, I will examine if there were increases in the variance of each season's PER distribution from the 1985-86 season to the 2015-16 season.

4.2.1 Increasing PER Variance

As was previously discussed in the section on PER, John Hollinger designed the PER such that the mean PER is equal to 15.00 every season. However, this is the only fixed moment of the PER distribution. This raises the question of if other moments of the PER distribution have changed over time, particularly variance. An increase in the variance of the PER over time would indicate that there is a divergence in skill — the top athletes are getting better while the bottom athletes are getting worse. In Figure 5, I plotted the relative PER distributions. *Prima facie*, the distributions do not appear to have significantly changed over time. But, in Figure 6 I plot the variance of the PER over time and find that there is an increase in the sample.

Figure 5: Relative PER Frequency by Season

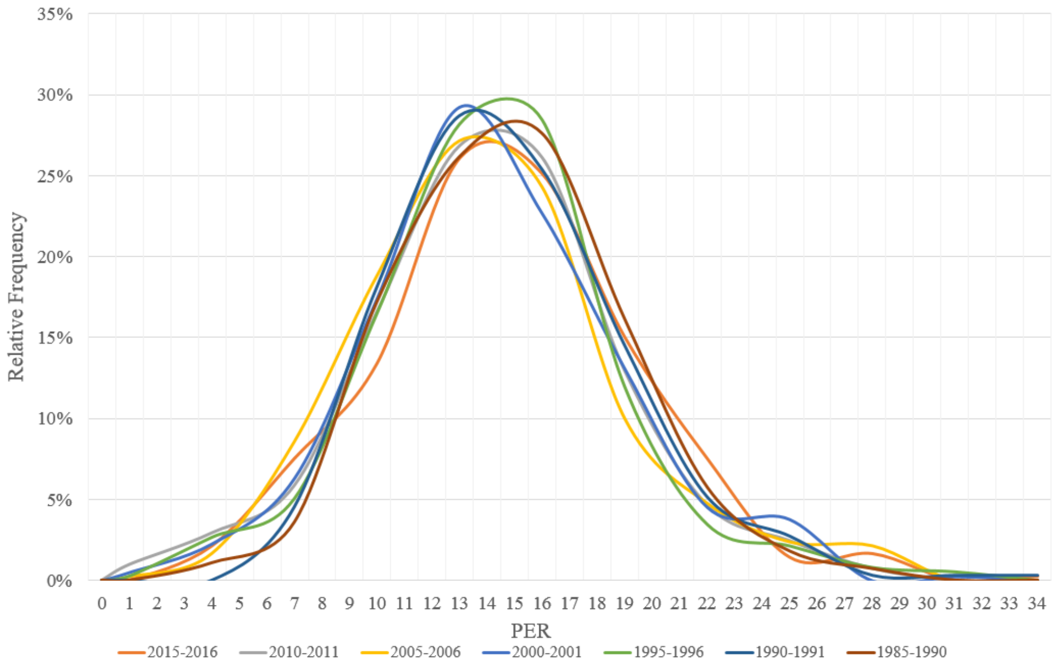
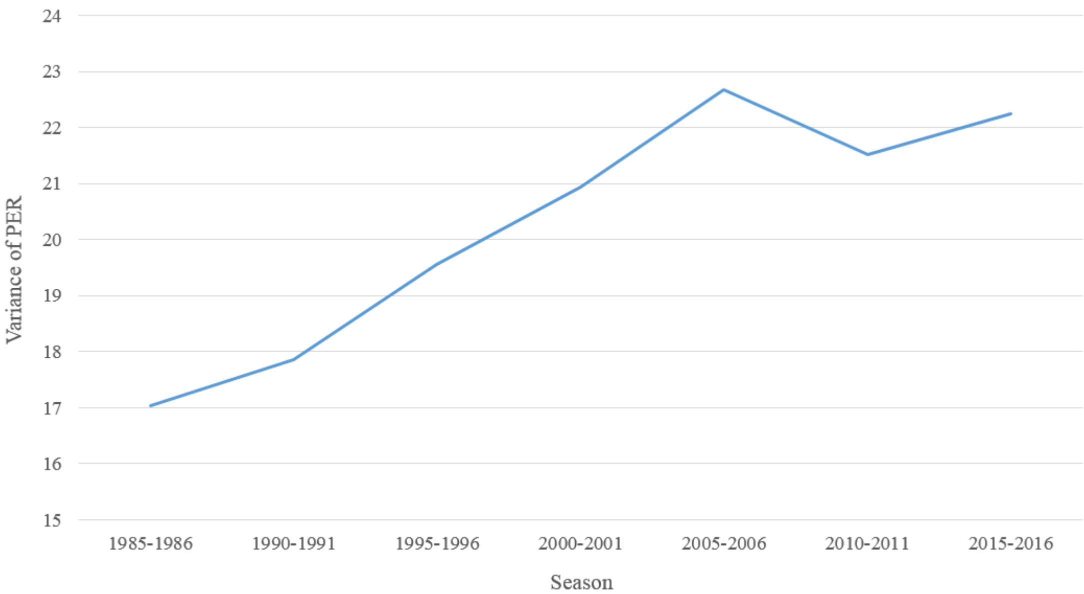


Figure 6: Variance of PER Distributions: 1985-86 Season to 2015-16 Season



With such an increase in the variance of the PER distribution over time, my hypothesis is that the increase in the variance of the skill distribution is a variable that is driving the increasing salary inequality in the league. However, we must first check if returns to skill have changed over time. To determine if this is the case, I used the regression that follows to determine the returns to skill for every season in the sample.

4.3 Regression Analysis

In this section, I will 1) estimate returns to skill from the 1985-86 season to the 2015-16 season and 2) examine whether the returns to skill have changed over the sample period.

4.3.1 Returns to Skill

I ran the following regression:

$$ln(Salary)_{Sn,p} = \alpha + \beta_{1985}(PER_{1985,p}) + \beta_2(PER^{2015}_{1990,p} \times Season) + \gamma_{1985}(Age_{1985,p}) + \gamma_2(Age^{2015}_{1990,p} \times Season) + \eta(Season_p) + \epsilon_{p,Sn} \quad (3)$$

In this regression, I examined the relationship between $ln(Salary)$ and PER. The subscripts p and Sn index each observation across all players and seasons, respectively. The β_{1985} coefficient represents the return to skill for the 1985-86 season. For example, an athlete in the 1985-86 season with a PER one point higher than another athlete will, on average, have a salary $(1+\beta_1)$ times higher. The β_2 represents each subsequent season's addition to the 1985-86 season's β_{1985}. If, for example, β_{1985} is 0.10 and β_2 for the 1990-91 season is -0.02, then the return to skill in the 1990-91 season would be 0.08, or 8%, which is 2% lower than in the 1985-86 season.

The Age variables work the same way; the γ_{1985} represents the return to age for the 1985-86 season and the γ_2 represents each subsequent season's change from the 1985-86 return to age. Once again, we included Age as a proxy for veteran status to control for the minimum salary for each class of veterans. Furthermore, I created a dummy variable for each season that allows the constant to move depending on which season the regression is evaluating, controlling for season-fixed effects.

Finally, in the data appendix 6.2 you can find the regression in which team was included as an explanatory variable but not fully interacted. I do not report the fully interacted model. Neither of these models were employed for two reasons. First, I am more concerned with a league-wide return to skill than any individual team's return — the fully interacted model produces the latter. Second, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are both notably higher than the other, more parsimonious regression. The fully interacted model has an AIC of 6,801.4 and a BIC of 10,416.4 while the model including only one team variable has an AIC of 6,308.7 and a BIC of 6,608.4. The model which excludes team completely has an AIC of 6,271.3 and a BIC of 6,394.7. By all information criteria, the model that excludes team as an explanatory variable is more parsimonious than the other two models.[11] The R^2 of the most parsimonious regression was 0.598, meaning that 59.8% of the variation in log salaries can be explained by the model.

4.3.2 Returns to Skill Over Time

Based on the most parsimonious model, I ran F-tests to determine whether each β_2 coefficient was equal to every other β_2 coefficient. However, to make these F-statistics interpretable, I converted each F-statistic into a t-statistic by taking the square root of each F-statistic.[12] This allowed me to apply the appropriate sign to each t-statistic. From here, I was able to do right-tailed t-tests to check if each subsequent PER coefficient is greater than the coefficient of interest.[13] In Table 4 below, you will find the estimates for the return to skill for each season in the sample and in Table 5, you will find the matrix of t-statistics.

[11]See data appendix 6.2 for the regression omitting team-fixed effects and the regression including non-interacted team-fixed effects. The regression with fully-interacted team-fixed effects is omitted because of length, but it is clear from the AIC and BIC that it is unnecessary.

[12]Once again, I was able to convert the F-statistics to t-statistics because each F-test had only 1 degree of freedom.

[13]For full regression output, see data appendix 6.2

Table 4: Returns to Skill

Season	1985	1990	1995	2000	2005	2010	2015
β_{Season}	0.10*** (0.01)	-0.03** (0.01)	0.01 (0.01)	-0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)

***$p < 0.01$, **$p < 0.05$, *$p < 0.10$

Table 5: Right-Tailed Testing of Returns to Skill

	1990	1995	2000	2005	2010	2015
1985	-2.35	0.93	-0.62	0.31	0.42	0.07
1990		3.30***	1.40*	2.65***	2.76***	2.35***
1995			-1.44	-0.62	-0.51	-0.83
2000				0.89	0.98	0.67
2005					0.11	-0.22
2010						-0.33

***$p < 0.01$, **$p < 0.05$, *$p < 0.10$

Essentially every season, barring the 1990-91, had statistically equal returns to skill. The 1990-91 season saw a statistically significant decrease in the returns to skill at the 1% level compared to every other season except the 2000-01 season, where the difference was significant at the 10% level. It is interesting to note that the 1990-91 season had the greatest Pareto exponent at -1.61, an indication that it was the season with the least inequality. Thus, for the 1990-91 season, as returns to skill declined, inequality did as well. As soon as returns to skill began regressed to its normal level, salary inequality normalized as well.

5 Discussion and Conclusion

I have shown that since the 1985-86 season, salaries in the NBA have become more widely distributed. Athletes in the top 10%-25% and 25%-50% have seen their salaries rise. However, from the 2000-01 season onward, inequality has not significantly risen. This was demonstrated by calculating the Pareto exponents for each season of the sample and doing t-tests on each exponent.

When I examined the distribution of skill via the distribution of PERs, I observed that the variance of PERs has increased over almost every season in the sample. Naturally, this raised the question of whether the increase in variance of the salary distribution was driven by some variable related to skill, either the variance of PERs or returns to skill. Thus, for each season, I determined the return to skill by regressing log salaries on PER controlling for age (as a proxy for veteran status). Indeed, I found that returns to skill have not increased since the 1985-86 season. However, the 1990-91 season saw a statistically significant decline in return to skill, the only season which saw this phenomenon.

These results indicate that the return to skill is not the factor that is driving the change in salary inequality from the 1985-86 season onward. Perhaps changes in the return to skill is partially responsible for the change in salary inequality from the 1990-91 season onward, yet increases in salary inequality was observed when comparing the 1985-86 season to all the seasons from the 1995-96 season onward. These seasons all had statistically equal returns to skill. This indicates that the widening skill distributions drove the increase in salary inequality, at least in part. Since returns to skill have not changed, a widening of the skill distribution necessarily implies that the top athletes earn more and the bottom athletes earn less, increasing salary inequality. I have not provided evidence against secular trends towards greater inequality, but I have provided evidence that increasing PER variances explain some of the increase in salary inequality.

To determine if there is a secular trend towards greater salary inequality in the NBA, future research could examine more data. In particular, examining the interim seasons between each season I used in this paper could answer the question of if there is a secular trend towards greater salary inequality or if salary distributions can be completely explained by changing skill distributions. Once again, I provided statistical evidence that changes in returns to skill do not affect salary inequality. Further, I provided observational evidence that widening skill distributions explain at least some of the growth in inequality. To formally check whether there exists a secular trend will require more data.

6 Data Appendix

6.1 Pareto Exponent Regression

In the following regression, I estimated the Pareto exponent for each season. Each variable is defined:

- logrnk: This is the 1985-86 Pareto exponent.
- logrnksn*: a specific addition or subtraction to the 1985-86 Pareto exponent that determines a subsequent season's Pareto exponent. For example, logrnksn2 is the addition for the 1990-91 season, logrnksn3 is the addition for the 1995-96 season, etc.
- sn* and _cons: This controls for season-fixed effects. This variable allows the constant to move depending on which season is of interest. For example, sn2 corresponds to the 1990-91 season, sn3 corresponds to the 1995-96 season, etc._ is the constant for the 1985-86 season.

Pareto Exponent Regression

	(1)
	logsalary
logrnk	-1.764***
	(0.0834)
logrnksn2	0.146
	(0.124)
logrnksn3	-0.196
	(0.130)
logrnksn4	-0.409**
	(0.147)
logrnksn5	-0.246
	(0.184)

logrnksn6	-0.363*
	(0.150)
logrnksn7	-0.402**
	(0.147)
logrnksn8	-0.509**
	(0.158)
sn2	0.723**
	(0.261)
sn3	2.091***
	(0.279)
sn4	3.216***
	(0.321)
sn5	3.080***
	(0.411)
sn6	3.470***
	(0.331)
sn7	3.636***
	(0.326)
sn8	4.097***
	(0.353)
_cons	16.07***
	(0.173)
n	2,639
R^2	0.861
adj. R^2	0.860

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.2 Returns to Skill Regressions

In the following regression, I estimated the returns to skill for each season. Each variable is defined:

- per: This is the return to skill for the 1985-86 season.
- persn*: This is a specific addition to the 1985-86 return to skill that determines a subsequent season's return to skill. For example, persn2 is the addition for the 1990-91 season, logrnksn3 is the addition for the 1995-96 season, etc.
- age: This variable is the return to age for the 1985-86 season. It is used as a control for veteran status.
- agesn*: This is a specific addition to the 1985-86 return to age. It is interpreted the same as persn*, but with age instead.
- sn*: This is the same variable from the previous regression.
- tm*: Each of these variables represents a specific team in the NBA.
  - tm1: Atlanta Hawks
  - tm2: Boston Celtics
  - tm3: Brooklyn Nets
  - tm4: Charlotte Hornets
  - tm5: Chicago Bulls
  - tm6: Cleavland Cavaliers
  - tm7: Dallas Mavericks
  - tm8: Denver Nuggets
  - tm9: Detroit Pistons

- tm10: Golden State Warriors
- tm11: Houston Rockets
- tm12: Indiana Pacers
- tm13: Los Angeles Clippers
- tm14: Los Angeles Lakers
- tm15: Memphis Grizzlies
- tm16: Miami Heat
- tm17: Milwaukee Bucks
- tm18: Minnesota Timberwolves
- tm19: New Orleans Pelicans
- tm20: New York Knicks
- tm21: Oklahoma City Thunder
- tm22: Orlando Magic
- tm23: Philadelphia 76ers
- tm24: Phoenix Suns
- tm25: Portland Trail Blazers
- tm26: Sacramento Kings
- tm27: San Antonio Spurs
- tm28: Toronto Raptors
- tm29: TOT (Undefined by Basketball-Reference.com)
- tm30: Utah Jazz
- tm31: Washington Wizards

• _cons: This is the regression constant.

Returns to Skill Regressions

	(1)	(2)
	logsalary	logsalary
per	0.105***	0.103***
	(0.00846)	(0.00829)
persn2	-0.0289*	-0.0269*
	(0.0116)	(0.0114)
persn3	0.00904	0.0109
	(0.0119)	(0.0117)
persn4	-0.00880	-0.00834
	(0.0135)	(0.0134)
persn5	0.00300	0.00361
	(0.0118)	(0.0118)
persn6	0.00423	0.00493
	(0.0119)	(0.0118)
persn7	0.000974	0.000890
	(0.0121)	(0.0121)
age	0.0800***	0.0800***
	(0.0126)	(0.0125)
agesn2	-0.0264	-0.0250
	(0.0171)	(0.0168)
agesn3	-0.0221	-0.0214
	(0.0171)	(0.0168)
agesn4	-0.00574	-0.00648
	(0.0157)	(0.0155)

agesn5	0.0244	0.0221
	(0.0164)	(0.0163)
agesn6	0.00969	0.00768
	(0.0164)	(0.0162)
agesn7	-0.0140	-0.0161
	(0.0153)	(0.0152)
sn2	1.981***	1.911***
	(0.454)	(0.444)
sn3	1.925***	1.876***
	(0.452)	(0.441)
sn4	2.287***	2.296***
	(0.443)	(0.435)
sn5	1.604***	1.654***
	(0.436)	(0.429)
sn6	2.118***	2.157***
	(0.437)	(0.427)
sn7	2.785***	2.838***
	(0.404)	(0.399)
tm1	0.0524	
	(0.136)	
tm2	0.0953	
	(0.130)	
tm3	0.0952	
	(0.139)	
tm4	0.0964	

	(0.129)
tm5	0.0758
	(0.127)
tm6	0.118
	(0.132)
tm7	0.0721
	(0.134)
tm8	0.0823
	(0.128)
tm9	0.0866
	(0.125)
tm10	0.159
	(0.137)
tm11	-0.0339
	(0.132)
tm12	0.140
	(0.126)
tm13	0.0449
	(0.133)
tm14	0.0808
	(0.133)
tm15	0.0732
	(0.150)
tm16	-0.0518

	(0.150)
tm17	0.140
	(0.133)
tm18	0.154
	(0.135)
tm19	0.183
	(0.154)
tm20	0.315*
	(0.139)
tm21	0.138
	(0.133)
tm22	0.0920
	(0.134)
tm23	0.0794
	(0.144)
tm24	0.114
	(0.135)
tm25	0.0934
	(0.137)
tm26	0.118
	(0.133)
tm27	-0.126
	(0.136)
tm28	0

	(.)	
tm29	0.110	
	(0.114)	
tm30	0.0243	
	(0.126)	
tm31	0.0233	
	(0.130)	
_cons	8.865***	8.980***
	(0.334)	(0.311)
n	2639	2639
R^2	0.601	0.598
adj. R^2	0.593	0.595
AIC	6308.7	6271.3
BIC	6608.4	6394.7

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7 References

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71. Retrieved from http://www.jstor.org/stable/29779750.

Basketball-Reference. Calculating per. Retrieved from http://www.basketball-reference.com/about/per.html.

Basketball-Reference (2017). Nba and aba league index. Retrieved from http://www.basketball-reference.com/leagues/.

Beck, H. (2011). For n.b.a. owners, 1999 labor deal was win with regrets. Retrieved from http://www.nytimes.com/2011/07/03/sports/basketball/nba-owners-threaten-lockout-to-revamp-labor-deal.html.

Coon, L. (2011). Larry coon's nba salary cap faq. retrieved from http://www.cbafaq.com/salarycap05.htm.

ESPN (2017). 2016-2017 hollinger team statistics. Retrieved from https://espn.go.com/nba/hollinger/teamstats/_/sort/paceFactor.

Gabiax, X. (2016). Power laws in economics: An introduction. *The Journal of Economic Perspectives*, 30(1):185–205. Retrieved from http://www.jstor.org/stable/43710016.

Hill, J. R. (2004). Pay discrimination in the nba revisited. *Quarterly Journal of Business and Economics*, 43(1/2):81–92. Retrieved from http://www.jstor.org/stable/40473375.

Hill, J. R. and 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):131–44. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/152700250100200203.

Hollinger, J. (2011). What is per? Retrieved from http://www.espn.com.au/nba/columns/story?id=2850240&columnist=hollinger_john.

Jenkins, J. A. (1996). A reexamination of salary discrimination in professional basketball. *Social Science Quarterly*, 77(3):594–608. Retrieved from http://www.jstor.org/stable/42863504.

Kahn, L. M. and Sherer, P. D. (1988). Racial differences in professional basketball players' compensation. *Journal of Labor Economics*, 6(1):40–61. Retrieved from http://www.jstor.org/stable/2534867.

Mogull, R. G. (1977). A note on racial discrimination in professional basketball: A reevaluation of the evidence. *The American Economist*, 21(2):71–75. Retrieved from http://www.jstor.org/stable/25603127.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T. and Saez, E. (2003). Income inequality in the united states. *The Quarterly Journal of Economics*, 118(1):1–39. Retrieved from http://www.jstor.org/stable/25053897.

Scully, G. W. (2004). Player salary share and the distribution of player earnings. *Managerial and Decision Economics*, 25(2):77–86. Retrieved from http://www.jstor.org/stable/25151278.