Analyzing NBA Player Statistics and Salary

ISYE 6783 Final Project Report

Richard Hu

May 3, 2021

1 Introduction

The National Basketball Association (NBA) is a North American-based professional basketball league which generates about $8 billion in yearly revenue and consists of 30 teams, each with an individual valuation of over $1 billion and an average team valuation of $2.2 billion (Forbes). Despite these already-considerable estimates, winning the NBA championship and consistently performing well can still cause valuations to drastically increase. The 2019 NBA Champion Toronto Raptors saw their valuation increase by 25% from the previous year (CNBC), and the Golden State Warriors, arguably the most successful NBA franchise in the last 10 years with 3 NBA championships and 5 NBA Finals appearances, have seen their valuation jump more than 900% from $450 million to $4.7 billion since 2011, rising from the 8th most valuable team to the 2nd most valuable team (Forbes). A 2009 study by Humphreys and Lee has also shown that franchise quality, including on-court performance, is the main driver in franchise valuation (Humphreys & Lee).

That being said, how does an NBA team increase its chances of winning? While there are numerous factors that go into assembling a championship-caliber team, the bottom line is that teams are given a certain limit of money to spend, the salary cap, and using this allocated money, the team needs to accurately identify and recruit players that will hopefully be worth their salary and possibly even outperform it. For example, from 2013 to 2017, Stephen Curry of the Golden State Warriors had a yearly salary ranging from being the 60th highest paid player to the 82nd highest paid player each year, while simultaneously winning two MVP (Most Valuable Player) awards, two All-NBA 1st Team awards, and two All-NBA 2nd Team awards (the latter two are essentially awarded respectively to the top 5 and top 10 players in the league). Our goal is to identify overvalued and undervalued players in the NBA using their salary and statistics. By recruiting correctly-valued or undervalued players and avoiding overvalued players, a team can maximize their chance of winning with a limited total salary, and thus increase their franchise valuation.

2 Data Collection

All data handling was done in Python. Player salary information was gathered from HoopsHype and player statistic information was collected from Basketball Reference. BeautifulSoup was used to scrape table information from the two sites and Pandas was used to format the data into a usable arrangement. Fortunately both above websites have URLs that can easily be changed to go from one year to another year's data. Then, using the BeautifulSoup HTML parser, we identified the relevant tables either using class names, or looping across data cell HTML and fetching the data at certain indices. Some mild data cleaning then took place; salaries were scrubbed of commas and dollar signs, and player names which had accented characters had them replaced with standard alphabet characters for consistency.

The data was then split into a training set and test set. As we are looking to find valuations of current player performance for the immediate future, the test set was naturally salaries and stats from the current 2020-2021 NBA season, while the training set would be salaries and stats from all prior seasons.

3 Data Preprocessing

When analyzing the NBA player salary information, we have to acknowledge the general trend of NBA player salaries increasing over time. The NBA salary cap, or the limit of total money a team can pay its players, is calculated as a percentage of the NBA's total revenue from the previous year. As NBA revenue has generally increased over the last 20 years, so has the NBA salary cap, and thus, NBA player salaries. This caused issues with initial regressions, as within the training set, we would always find that the players from earlier years with relatively smaller salaries but similar production would routinely be considered undervalued and more recent player stats and salaries would frequently be overvalued. To account for this, along with potential inflation changes, for the response variable, we will use player salary as a percentage of the yearly salary cap. The previously described trend can be seen below, where we have a histogram of overvalued players by year using salary and salary cap percentage, and we define an overvalued player to be in the highest 20% of actual salary minus predicted salary across all training data.

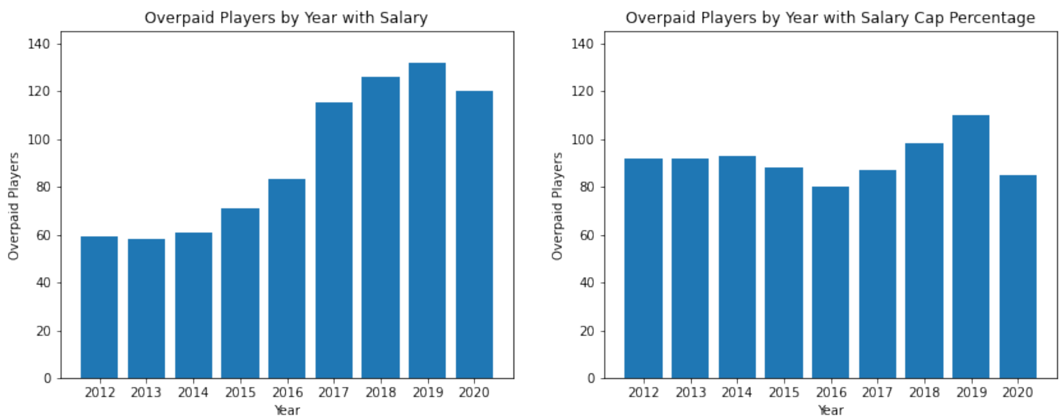

Figure 1: Overvalued players analysis from basic training data regression using salary and salary cap percentage as response variables. Note the jump in later years for the salary-based regression compared to the evenness for the salary cap percentage-based regression.

Deciding what NBA player statistic information to use contains multiple caveats. We will start with a smaller detail that will affect our future feature selection: pace of play. Players who play more minutes and teams who are faster-paced and take more shots will have generate higher stat counts per game than players who play less and teams who are slower-paced. We begin with initial regressions using per-game stats and per-100-possessions stats, the latter being normalized to a player on the court for 100 of his team's possessions. The results are shown in Appendix in Figure 11. The two models have very similar sets of p-values despite their changes in measurement. While we will address the issue of pace later, at the moment and more importantly, we begin to see issues with multicollinearity. Rebounds, one of the main basketball counting stats, are not significant in any of its three values: ORB, DRB, TRB (a full glossary of basketball abbreviations can be found in Reference 6). This is an easier case to handle, as their relationship takes the form $ORB + DRB = TRB$. If we build a regression model using only ORB and DRB or only TRB (shown in Appendix Figure 12), we can now see significance in the DRB or TRB variables. The more difficult stats to handle are points and shot-attempt related stats. These can be given as either of the following

$$PTS = 3 * 3PM + 2 * 2PM + FTM$$
$$PTS = 3 * 3P\% * 3PA + 2 * 2P\% * 2PA + FT\% * FTA$$

For now, we will assume that PTS on its own, is able to explain all the previously given variables in the equation.

4 Initial Regression Model

Note, the following regression model does not use TO; with the lessened features, it was still not significant at $\alpha = 0.05$. We additionally, chose not to use the “Minutes" predictor. Not only was it insignificant in the previous regressions, but as can be seen in the correlation matrix below, we believe it had multicollinearity with several other predictors. This makes sense, as the more minutes a player plays, the more opportunities they have to accumulate statistics.

	Minutes	TRB	AST	BLK	STL	PF	PTS
Minutes	1.000000	0.668110	0.678469	0.389820	0.760180	0.723578	0.882886
TRB	0.668110	1.000000	0.281699	0.701756	0.430948	0.695597	0.622656
AST	0.678469	0.281699	1.000000	0.036286	0.696563	0.382122	0.688107
BLK	0.389820	0.701756	0.036286	1.000000	0.224691	0.543467	0.345380
STL	0.760180	0.430948	0.696563	0.224691	1.000000	0.544777	0.674458
PF	0.723578	0.695597	0.382122	0.543467	0.544777	1.000000	0.591221
PTS	0.882886	0.622656	0.688107	0.345380	0.674458	0.591221	1.000000

Figure 2: Correlation matrix.

Thus, for our first regression model, we have

	coef	std err	t	P>|t|	[0.025	0.975]
Age	0.0008	7.9e-05	9.653	0.000	0.001	0.001
TRB	0.0092	0.001	14.414	0.000	0.008	0.010
AST	0.0086	0.001	10.937	0.000	0.007	0.010
STL	-0.0110	0.003	-3.458	0.001	-0.017	-0.005
BLK	0.0106	0.003	3.598	0.000	0.005	0.016
PF	-0.0240	0.002	-13.896	0.000	-0.027	-0.021
PTS	0.0055	0.000	20.961	0.000	0.005	0.006

Omnibus:	767.383	**Durbin-Watson:**	0.783
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1845.460
Skew:	1.036	**Prob(JB):**	0.00
Kurtosis:	5.539	**Cond. No.**	106.

Figure 3: Initial multiple regression model.

These predictors represent the 5 standard counting stats, along with player age and fouls per game. All of them are significant at $\alpha = 0.05$, and most of the coefficients agree with intuition. The coefficient for player age should be positive, as older players have higher guaranteed minimum salaries, and the “max" contract a player can sign has jumps in price for players who have been in the NBA for at least 7 years and 10 years. Fouling is a negative action in basketball, so it has a negative coefficient. TRB, AST, BLK, PTS are all beneficial stats for players to acquire, and they have positive coefficients. The strange coefficient is STL being negative, as it's another beneficial stat.

4.1 Model Metrics

We will evaluate our multiple regression models using three very popular metrics: R^2, Mean Absolute Error, and Root Mean Squared Error. To try to increase our initial model strength, we build two additional models, one using standardization on the predictors and the other using normalization on the predictors. As we can see in the figure below, the multiple regression model using scaled and normalized predictors both perform worse in all three metrics than the initial model. The scaled model, while its error measurements are of a similar scale to the other two, has notably poor R^2 predictive power.

Model	R^2	MAE	RMSE
Initial	0.562862	0.042490	0.056019
Scaled	0.052969	0.067930	0.082453
Normalized	0.426121	0.048216	0.064185

Figure 4: Regression model metrics for initial model along with scaled and normalized models.

5 Regression Model Refinement

5.1 Feature Selection

We take a more refined look at the model feature selection now. We build a stepwise-regression feature selection method in Python that uses a standard forward-backward stepwise algorithm based on predictor p-values. In this algorithm, we find the best predictor to add to the model (forward) if it has p-value under 0.04; then we check the current model with all added predictors to see if any have p-value over 0.05 and potentially remove them. We iterate until we have selected a model whose features are all p-value under 0.05. This gives us the following predictors: $FTM, DRB, AST, FGM, PF, Age, BLK, STL$. The exact model summary can be found in the Appendix. The last feature selection we will implement is using Lasso regression. After optimizing the alpha parameter, we generate three models based on the original data, standardized data, and normalized data. The coefficients for each will be in the Appendix. The metric evaluation of each, along with the stepwise-regression selection can be seen below.

Model	R^2	MAE	RMSE
Initial	0.562862	0.042490	0.056019
Stepwise	0.566595	0.041838	0.055779
Lasso	0.636126	0.037504	0.051109
Lasso Scaled	0.627890	0.037268	0.051684
Lasso Normalized	0.416608	0.047089	0.064715

Figure 5: Regression model metrics for various models.

Though the stepwise-regression feature selection model performs slightly better than the original model across all 3 metrics, the lasso regression model with original data and scaled data both significantly outperform the other models.

5.2 Data Improvements

As previously discussed, we anticipated issues with the per-game statistics since they are team and year-dependent. We previously tried standardizing and normalizing the data but found it to be consistently performing more poorly than the original data. However, there are a variety of statistics in the NBA analytics world called “advanced" statistics. These attempt to standardize and normalize player performance and produce unbiased measures of player performance across all teams and all years. The list of advanced statistics can be seen in Figure 15 in the Appendix. Their descriptions and how they are calculated are all explained in the glossary of terms found at Reference 6. Next, we will look at regression models using these already-adjusted advanced statistics.

Regression metrics are shown below in the figure. We first found that on their own, with feature selection, the advanced statistics were not as strong as the per-game statistic model. However, we believed that some of the predictive power of the advanced statistics could be combined with the per-game statistics, so we then built two additional models using feature-selection techniques on both sets of predictors (per-game and advanced statistics). The joint model using stepwise regression feature-selection ended up outperforming the previous best model, the per-game statistics with lasso regression. The selected predictors were: $Age, TS\%, OBPM, DRB, PER, FTM, 3PAr, FGM, PF, WS/48, TO, BLK, AST\%, USG\%, BPM, STL$.

Model	R^2	MAE	RMSE
Adv Stepwise	0.504847	0.039562	0.059646
Adv Lasso	0.510477	0.040643	0.059306
Joint Stepwise	0.646862	0.036442	0.050371
Joint Lasso	0.617115	0.037833	0.052450

Figure 6: Regression metrics using feature selection techniques with per-game data and advanced statistics.

6 Salary Analysis & Conclusions

Player	Salary %	Pred	Diff
Luka Doncic	0.073753	0.244216	-0.170463
Trae Young	0.060214	0.206298	-0.146083
Bam Adebayo	0.046871	0.188153	-0.141282
Donovan Mitchell	0.047604	0.179027	-0.131423
Shai Gilgeous-Alexander	0.037945	0.166276	-0.128331
Jayson Tatum	0.090683	0.204920	-0.114238
Carmelo Anthony	0.023500	0.120444	-0.096944
Mike James	0.000907	0.095281	-0.094374
Jarrett Allen	0.035825	0.128590	-0.092765
Zion Williamson	0.093875	0.184864	-0.090990

Player	Salary %	Pred	Diff
Blake Griffin	0.310612	0.118641	0.191972
John Wall	0.378000	0.197340	0.180660
Otto Porter	0.261034	0.081029	0.180005
Chris Paul	0.378952	0.201727	0.177225
Steven Adams	0.271144	0.096333	0.174811
Mike Conley	0.316127	0.158349	0.157778
Andrew Wiggins	0.270680	0.117801	0.152879
D'Angelo Russell	0.262500	0.109847	0.152653
Kemba Walker	0.315000	0.165299	0.149701
Kevin Love	0.286405	0.138727	0.147678

Figure 7: Regression model built with stepwise regression feature selection using per-game and advanced statistics. Left figure is underpaid players; right figure is overpaid players.

Using our new joint model, we can now analyze player salaries. Using the regression model to predict salary percentage, and subtracting from the actual salary, we look at this 2020-2021 NBA season's most underpaid and overpaid players. It is worth nothing that 9 of these players are on their rookie contracts, and have already become some of the top players in the league. It is unlikely that their teams would be willing to trade them or release them barring internal conflict. It is important to note that this means that keeping them should be a high priority for winning basketball. We do see one role player, Mike James, show up as well. To optimize their salary, the Brooklyn Nets should prioritize keeping him on a cheap contract. On the other hand, the names on the overpaid list are some of the most frequently cited examples of "bad contracts" in NBA discourse. Future teams looking to sign new players might want to think about avoiding them at a similar cost. Below in Figure 8, we see an additional visualization of salary vs predicted salary. We can see some of the underpaid players as small outliers in the top left of the data, while the overpaid players stand out being to the bottom right.

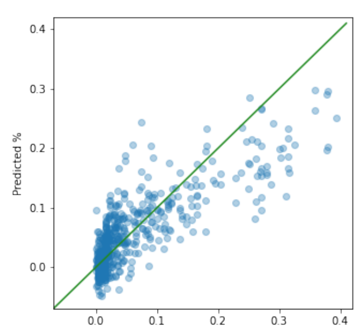

Figure 8: Salary % vs predicted. The green line is $y = x$. Players above are underpaid, and below are overpaid

In Figure 9 below, we see the summed salary differences for each team. Very few teams have highly underpaid players, but most of them are much younger teams with young players with high potential.

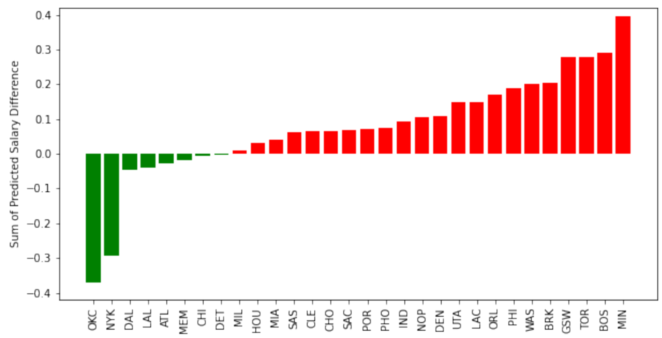

Figure 9: Summed predicted salary differences for each NBA team.

An issue with the modeling can be seen below. While NBA player salaries can go up to almost 40% of the cap, the predicted salary percentage maxes out at less than 30%. This makes it very difficult to evaluate players with max contracts. It leads to ask about future research regarding whether max players actually contribute that much towards a team's performance. The model does have fairly strong predictive power for performance; aside from one or two players, a lot of the players on the below list would be in current NBA top 10 lists. All except one made the all star game, and several are frontrunners for the current season MVP award.

Player	Salary %	Pred	Diff
LeBron James	0.359351	0.298623	0.060728
Russell Westbrook	0.378952	0.295067	0.083885
James Harden	0.378000	0.290224	0.087776
Giannis Antetokounmpo	0.252227	0.285585	-0.033358
Nikola Jokic	0.270680	0.266694	0.003986
Joel Embiid	0.270680	0.264796	0.005884
Kevin Durant	0.357879	0.264155	0.093724
Stephen Curry	0.394048	0.250188	0.143859
Luka Doncic	0.073753	0.244216	-0.170463
Damian Lillard	0.289783	0.242503	0.047280

Figure 10: Highest predicted salary percent players in the NBA.

Expanded work could include additional predictors. Height could be interesting, as the average NBA player height has been going up over recent years. However, the number of high-level big men is much smaller than wing players who are slightly shorter, possibly indicating a quadratic fit. There could also be a better version of data standardization or normalization that we have not considered which could improve model efficacy. There are also other regression models that could be considered for the problem such as Ridge regression and SVR.

Overall, we have built a multiple regression NBA player salary estimator using forward-backward stepwise feature selection. The model uses the last 9 years of player salaries and stats as training data for the current year test data. The R^2 of 0.65 shows solid predictive power, and our mean absolute error and root mean squared error are fairly low, both around 0.04. The current visualizations only focus on the extrema, identifying the most underpaid and overpaid players, some of which, popular consensus already deem to be underpaid or overpaid. But with some future adjustments to filtering, teams can identify strong players deserving of strong salaries and potential breakout players on low salaries, all while avoiding overpaid players that will limit team budget flexibility in the future.

7 References

1. https://www.cnbc.com/2020/02/11/the-latest-forbes-nba-valuations-show-how-much-winning-can-pay-off.html
2. https://www.forbes.com/sites/kurtbadenhausen/2021/02/10/nba-team-values-2021-knicks-keep-top-spot-at-5-billion-warriors-bump-lakers-for-second-place
3. https://college.holycross.edu/RePEc/spe/HumphreysLee_FranchiseValues.pdf
4. https://www.basketball-reference.com/leagues/NBA_2021_per_game.html
5. https://hoopshype.com/salaries/players/
6. https://www.basketball-reference.com/about/glossary.html

8 Appendix

OLS Regression Results

Dep. Variable:	Salary %	**R-squared (uncentered):**	0.709
Model:	OLS	**Adj. R-squared (uncentered):**	0.708
Method:	Least Squares	**F-statistic:**	527.3
Date:	Wed, 28 Apr 2021	**Prob (F-statistic):**	0.00
Time:	11:18:04	**Log-Likelihood:**	6106.8
No. Observations:	4124	**AIC:**	-1.218e+04
Df Residuals:	4105	**BIC:**	-1.206e+04
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Age	0.0008	8.74e-05	9.305	0.000	0.001	0.001
Minutes	0.0002	0.000	0.643	0.520	-0.000	0.001
FGM	0.0201	0.028	0.714	0.475	-0.035	0.075
FGA	0.0322	0.019	1.715	0.086	-0.005	0.069
3PM	-0.0218	0.025	-0.874	0.382	-0.071	0.027
3PA	-0.0389	0.019	-2.071	0.038	-0.076	-0.002
2PM	-0.0318	0.020	-1.571	0.116	-0.071	0.008
2PA	-0.0302	0.019	-1.606	0.108	-0.067	0.007
FTM	0.0055	0.012	0.450	0.653	-0.018	0.029
FTA	-0.0033	0.003	-0.987	0.324	-0.010	0.003
ORB	-0.0259	0.018	-1.450	0.147	-0.061	0.009
DRB	-0.0130	0.018	-0.730	0.465	-0.048	0.022
TRB	0.0255	0.018	1.431	0.152	-0.009	0.060
AST	0.0079	0.001	7.354	0.000	0.006	0.010
STL	-0.0100	0.003	-2.952	0.003	-0.017	-0.003
BLK	0.0099	0.003	3.263	0.001	0.004	0.016
TO	-0.0025	0.003	-0.811	0.417	-0.009	0.004
PF	-0.0219	0.002	-11.146	0.000	-0.026	-0.018
PTS	0.0087	0.012	0.737	0.461	-0.014	0.032

Omnibus:	792.622	**Durbin-Watson:**	0.809
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1948.036
Skew:	1.060	**Prob(JB):**	0.00
Kurtosis:	5.616	**Cond. No.**	1.47e+03

OLS Regression Results

Dep. Variable:	Salary %	**R-squared (uncentered):**	0.658
Model:	OLS	**Adj. R-squared (uncentered):**	0.656
Method:	Least Squares	**F-statistic:**	393.9
Date:	Wed, 28 Apr 2021	**Prob (F-statistic):**	0.00
Time:	11:22:57	**Log-Likelihood:**	5752.4
No. Observations:	4114	**AIC:**	-1.146e+04
Df Residuals:	4094	**BIC:**	-1.134e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Age	0.0059	0.000	26.807	0.000	0.005	0.006
FGM	-0.0249	0.030	-0.838	0.402	-0.083	0.033
FGA	0.0405	0.020	2.061	0.039	0.002	0.079
3PM	-0.0083	0.027	-0.311	0.756	-0.061	0.044
3PA	-0.0426	0.020	-2.170	0.030	-0.081	-0.004
2PM	0.0011	0.021	0.053	0.958	-0.041	0.043
2PA	-0.0402	0.020	-2.048	0.041	-0.079	-0.002
FTM	-0.0064	0.013	-0.507	0.612	-0.031	0.018
FTA	0.0004	0.001	0.404	0.686	-0.002	0.002
ORB	0.0056	0.019	0.296	0.767	-0.031	0.042
DRB	0.0103	0.019	0.545	0.586	-0.027	0.047
TRB	-0.0066	0.019	-0.351	0.725	-0.043	0.030
AST	0.0036	0.000	7.786	0.000	0.003	0.004
STL	-0.0019	0.001	-1.718	0.086	-0.004	0.000
BLK	0.0021	0.001	1.947	0.052	-1.47e-05	0.004
TO	-0.0011	0.001	-1.109	0.267	-0.003	0.001
PF	-0.0068	0.001	-13.360	0.000	-0.008	-0.006
PTS	0.0151	0.013	1.200	0.230	-0.010	0.040
ORtg	-0.0003	0.000	-2.385	0.017	-0.001	-5.13e-05
DRtg	-0.0013	0.000	-10.584	0.000	-0.001	-0.001

Omnibus:	457.948	**Durbin-Watson:**	0.766
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	774.282
Skew:	0.769	**Prob(JB):**	7.36e-169
Kurtosis:	4.467	**Cond. No.**	6.10e+03

Figure 11: Regression with per-game stats on the left and per-possession stats on the right.

OLS Regression Results

Dep. Variable:	Salary %	**R-squared (uncentered):**	0.708
Model:	OLS	**Adj. R-squared (uncentered):**	0.707
Method:	Least Squares	**F-statistic:**	622.1
Date:	Wed, 28 Apr 2021	**Prob (F-statistic):**	0.00
Time:	12:03:11	**Log-Likelihood:**	6096.1
No. Observations:	4124	**AIC:**	-1.216e+04
Df Residuals:	4108	**BIC:**	-1.206e+04
Df Model:	16		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Age	0.0008	8.23e-05	10.050	0.000	0.001	0.001
FGM	0.0189	0.028	0.670	0.503	-0.036	0.074
FGA	0.0307	0.019	1.629	0.103	-0.006	0.068
3PM	-0.0226	0.025	-0.902	0.367	-0.072	0.026
3PA	-0.0357	0.019	-1.896	0.058	-0.073	0.001
2PM	-0.0321	0.020	-1.584	0.113	-0.072	0.008
2PA	-0.0284	0.019	-1.506	0.132	-0.065	0.009
FTM	0.0062	0.012	0.505	0.614	-0.018	0.030
FTA	-0.0043	0.003	-1.281	0.200	-0.011	0.002
TRB	0.0091	0.001	12.694	0.000	0.008	0.011
AST	0.0087	0.001	8.307	0.000	0.007	0.011
STL	-0.0093	0.003	-2.885	0.004	-0.016	-0.003
BLK	0.0101	0.003	3.300	0.001	0.004	0.016
TO	-0.0022	0.003	-0.705	0.481	-0.008	0.004
PF	-0.0227	0.002	-12.597	0.000	-0.026	-0.019
PTS	0.0092	0.012	0.780	0.436	-0.014	0.032

Omnibus:	791.692	**Durbin-Watson:**	0.799
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1944.748
Skew:	1.059	**Prob(JB):**	0.00
Kurtosis:	5.614	**Cond. No.**	1.20e+03

Figure 12: Regression with per-game stats and one rebounding value.

OLS Regression Results

Dep. Variable:	Salary %	**R-squared (uncentered):**	0.708
Model:	OLS	**Adj. R-squared (uncentered):**	0.707
Method:	Least Squares	**F-statistic:**	1248.
Date:	Wed, 28 Apr 2021	**Prob (F-statistic):**	0.00
Time:	17:17:31	**Log-Likelihood:**	6097.2
No. Observations:	4124	**AIC:**	-1.218e+04
Df Residuals:	4116	**BIC:**	-1.213e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
FTM	0.0097	0.001	7.844	0.000	0.007	0.012
DRB	0.0121	0.001	13.603	0.000	0.010	0.014
AST	0.0074	0.001	9.435	0.000	0.006	0.009
FGM	0.0096	0.001	10.155	0.000	0.008	0.011
PF	-0.0220	0.002	-12.944	0.000	-0.025	-0.019
Age	0.0008	7.93e-05	10.182	0.000	0.001	0.001
BLK	0.0100	0.003	3.494	0.000	0.004	0.016
STL	-0.0101	0.003	-3.191	0.001	-0.016	-0.004

Omnibus:	780.098	**Durbin-Watson:**	0.799
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1907.177
Skew:	1.046	**Prob(JB):**	0.00
Kurtosis:	5.593	**Cond. No.**	101.

Figure 13: Regression where features are selected by stepwise regression algorithm.

Feature	original	standardized	normalized
Age	0.005414	0.020863	-0.141876
FGM	0.000000	0.000000	0.000000
FGA	0.000000	0.000470	-0.000000
3PM	-0.000000	0.000000	-0.000000
3PA	-0.000000	0.000000	-0.067545
2PM	0.000000	0.000000	0.000000
2PA	0.003071	0.011835	0.000000
FTM	0.000000	0.005414	0.076282
FTA	0.004333	0.004301	0.000745
ORB	-0.000000	0.000000	-0.000000
DRB	0.004776	0.015284	0.239264
TRB	0.003772	0.000000	0.003471
AST	0.004232	0.006328	0.110505
STL	-0.000000	0.000000	-0.000000
BLK	0.000000	0.000645	0.000000
TO	0.000000	0.000000	0.000000
PF	-0.009501	-0.002623	-0.428564
PTS	0.003318	0.013664	0.128133

Figure 14: Lasso regression coefficients.

PER	TS%	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%		OWS	DWS	WS	WS/48		OBPM	DBPM	BPM	VORP

Figure 15: Advanced statistics.