Optimal Decision-Making in the NBA Free Agency Market

Pravek Karwe, Rishi Khera, Narain Krishnamurthy, Sarah Tracy

December 15, 2016

1 Abstract

The National Basketball Association is the most competitive basketball league in the world, with the competition period comprising of 82 games spanned over the course of 24 weeks. The success of a team on the court has a direct effect on the revenue and popularity of a team. While each NBA player reacts uniquely to the stress of practice and competition, there remains a clear need to use sports performance models to inform starting points for player preparation. One of the most common methods of doing this is using game-related statistics to evaluate technical and tactical behavior. Game-related statistics not only measure perception and action from the players, but also can provide a measure of co-adaptation since each player is ultimately part of a larger unit (the team). NBA teams have two options when deciding how to improve the players on their team; either they try to acquire talent through the NBA Draft, or they sign more experienced players through NBA free agency. Often, teams are hesitant to build talent through the draft due to the amount of time it takes for young players to develop and the potential for a given player to bust. With free agency, however, teams have the ability to sign players in their athletic primes that can immediately make an impact. For this reason, free agency offers the most potential to improve a team dramatically from one year to the next. In this paper, we analyze the optimal choices for a team in free agency to ensure the most improvement from any given season to the next.

In this paper, we will explore how to optimize player acquisitions in the free agency market. We will first cover an overview of NBA Free Agency and our considerations in Section 2 and 3. In Sections 4 and 5 we will discuss how we modeled the on-court value of a player and used that to model free agency selection as a knapsack problem. Finally in Section 6 we explore the results and implications of our model, and discuss future steps in Section 7.

2 Introduction to the Free Agency Problem

At the end of every NBA season there are a number of players with expired contracts who become unrestricted free agents. These players have the right to sign with any team of their choice. The NBA institutes a salary cap restricting the total amount that teams can spend on salaries. Since all the details of existing contracts is public information, all participants know the amount of money each team has to sign free agents. The NBA does not have a hard salary cap, but rather teams have the ability to raise the cap by taking advantage of salary cap exceptions. However, accounting for these exceptions is complicated, so in our model we will assume the salary cap is a hard upper bound.

In addition, there are constraints on the amount of players on a team. There must be between 13 and 15 players on a roster. Finally, five players are on a court at the same time, and these players generally fill five standard basketball positions: Point Guard, Shooting Guard, Small Forward, Power Forward and Center. Thus it would be beneficial to a teams success to have decent depth on the roster for each position.

3 Considerations in Model Selection

In order to maximize the performance of an NBA team, we wanted to value players such that all offensive statistics for the team are maximized. A team with the most points, rebounds, assists, etc possible is more likely to be a winning team. However, we also wanted to take into account the efficiency of players in order to capture how much they can produce in a given amount of time. In order to capture the total value of players, factors beyond the box score must be taken into account. A player can have an extremely large impact on the game without putting up any points, simply due to impact on the floor. There are a variety of intangible factors that cause players to be better than their expected value. Defensive impact is often hard to keep in the box score defensive statistics currently indicate (steals, blocks, defensive rebounds). Thus we wanted to formulate a model such that intangible factors to account for the intangibles, but ultimately decided to use intangible factors as a separate entity to maximize. For this reason, we wanted to start with a baseline statistic that measures tangible stats so that we could ultimately develop an intangible addition to that statistic.

4 Modeling Player Value

Our starting point is a well-established basketball statistic, the Player Efficiency Rating. Player Efficiency Rating (PER) is an all-in-one stat developed by John Hollinger of ESPN. It combines most of a player's attributes into one number, and sufficiently communicates the performance of a player. It is also standardized across the league, so PER is an easily comparable statistic between players of the same team, different teams, and even different years. The formula to calculate PER is shown below in:

$$\begin{split} uPER = &\frac{1}{\mathrm{MP}} * \\ & \left[3P \\ & +\frac{2}{3} * AST \\ & + \left(2 - factor * \left(\frac{\mathrm{team}_\mathrm{AST}}{\mathrm{team}_\mathrm{FG}} \right) \right) * FG \\ & + \left(FT * 0.5 * \left(1 + \left(1 - \left(\frac{\mathrm{team}_\mathrm{AST}}{\mathrm{team}_\mathrm{FG}} \right) \right) \right) + \frac{2}{3} * \left(\frac{\mathrm{team}_\mathrm{AST}}{\mathrm{team}_\mathrm{FG}} \right) \right) \right) \\ & -VOP * TOV \\ & -VOP * DRB\% * (FGA - FG) \\ & -VOP * 0.44 * (0.44 + (0.56 * DRB\%)) * (FTA - FT) \\ & +VOP * (1 - DRB\%) * (TRB - ORB) \\ & +VOP * DRB\% * ORB \\ & +VOP * STL \\ & +VOP * DRB\% * BLK \\ & \end{bmatrix} \end{split}$$

$$factor = \frac{2}{3} - \left(0.5 * \left(\frac{\text{lg}_\text{AST}}{\text{lg}_\text{FG}} / \left(2 * \frac{\text{lg}_\text{FG}}{\text{lg}_\text{FT}}\right)\right)\right)$$
$$VOP = \frac{\text{lg}_\text{PTS}}{\text{lg}_\text{FGA} - \text{lg}_\text{ORB} + \text{lg}_\text{TOV} + 0.44 * \text{lg}_\text{FTA}}$$
$$DRB\% = \frac{\text{lg}_\text{TRB} - \text{lg}_\text{ORB}}{\text{lg}_\text{TRB}}$$

 $aPER = uPER * \frac{\text{lg_PACE}}{\text{team_PACE}}$ $PER = aPer * (15/\text{lg}_aPer)$

Note: Any statistic with lg in front corresponds to the league average for that statistic.

PER is standardized such that the league average is 15. The all-time leader in PER in a season Wilt Chamberlain, with a PER of 31.82 for the 1962-63 season. The all-time leader in PER over a career is Michael Jordan, with a PER of 27.91. The interpretation of the value of PER is as follows:

All-time great season	35+
Hands-down MVP	30-35
Strong MVP candidate	27.5-30
Long-shot MVP candidate	25-27.5
Definite All-Star	22.5-25
Borderline All-Star	20-22.5
Second offensive option	18-20
Third offensive option	16.5-18
Slightly above-average player	15-16.5
Rotation player	13-15
Non-rotation player	11-13
Fringe roster player	9-11
Player who won't stick in the league	5-9

Table 1: Interpretation of PER values [8]

While PER is a very useful tool, and the go-to stat for NBA player analysis, it receives some criticism for being an offensive-heavy statistic [8]. It only includes two defensive statistics–steals and blocks–and doesn't incorporate any hustle statistics that the NBA have recently started to track.

We would like to use PER plus some additional statistics which incorporate defensive hustle statistics as the coefficients of the objective function in our model. That is to say, we we want to select the free agent with maximum PER subject to our constraints. However, we do want to update the model to more accurately reflect a player's defensive and hustle skills and abilities. In addition to PER, we will add an extra value to a player by incorporating five defensive and hustle statistics: screen assists, ball deflections, loose balls recovered, charges drawn and contested shots. These statistics have only started to be tracked recently. We incorporate these statistics by adding one stat, which we will label INT, that will incorporate these statistics which are missing from PER, to the base PER value of a player.

INT will incorporate the five defensive and hustle statistics mentioned by comparing a player statistics in these areas to the league averages. INT is a sum of the percentage comparison to the league average in each of these statistics. The mathematical formulation of INT is the value seen below:

$$INT = \frac{SA - \lg_SA}{\lg_SA} + \frac{DEF - \lg_DEF}{\lg_DEF} + \frac{LBR - \lg_LBR}{\lg_LBR} + \frac{CHR - \lg_CHR}{\lg_CHR} + \frac{CS - \lg_CS}{\lg_CS}$$

We structured INT in this manner such that each term in INT is standardized to 0, and so the league average for this is 0. We chose to standardize INT to have a league average of 0 so that our PER comparison metrics (e.g. an all-time great season is 35+) still applies to PER + INT. We chose the five terms mentioned because we believed that they effectively communicate a player's non-offensive abilities without overfitting with too many terms. Including more statistics in our INT value might not make our model better, only more complicated. INT does include defensive statistics (such as contested shots and charges drawn), as well as formerly "intangible" statistics

(such as loose balls recovered, or screen assists).

In total, our model will view PER + INT as the on-court value of a player. In formulating our model, we will aim to maximize this quantity from the players chosen from the free agency pool.

5 Modeling Free Agency

5.1 Knapsack Formulation

As mentioned, our goal is to maximize the value of the players on our team, subject to positional needs and the league-instituted salary constraint.

As a starting point, we chose to view the free agency problem as a knapsack linear program. A knapsack model is an obvious choice since each player has a clear cost (salary), value (PER + INT), and players can either be wholly in or out of the team (the knapsack in our case). Each player would represent a binary value representing whether that player would be part of the final team formulation for the next year. The goal would be to maximize value of the players in the knapsack, in this case the value of their Player Efficiency Rating plus our additional intangibles calculation, subject to their cost - salary - under the league constraint of the salary cap.

Different from a standard knapsack problem, we will subject the optimization to additional constraints. The main one to first abide by pertains to the positional needs of a team. There can be a minimum of thirteen players on a roster, and a maximum of fifteen. Each of these players is trained for one out of five standard basketball positions. In order to properly play games throughout the season, and to have backups in case starting players become injured, the roster should have a mostly even distribution of positions.

Given these constraints and problem formulation, we can run the model as a linear program, seen below:

Variables:

- P is an individual player
- FA is the set of all free agents
- SC is the team's salary cap
- C_P is the salary (cost) of player P
- V_P is the value of player P
- Pos_P is the static indicator variable as to whether player P plays position Pos
- X_P is the indicator variable as to whether the team chooses player P

LP Problem:

Maximize: $\sum V_P X_P$ Subject to:

- $13 \leq \sum_{P \in A} X_P \leq 15$
- $.9 * SC \leq \sum_{P \in A} C_P X_P \leq SC$
- For each position $Pos \in S: 2 \leq \sum X_P Pos_P \leq 4$
- For each player P:

$$-X_P \in 0, 1$$

 $- Pos_P \in 0, 1$ for each $Pos \in S$

$$-\sum_{Pos\in S} Pos_P = 1$$

5.2 Coding the Model

To produce solutions solution to this linear program, we wrote a set of Python classes, which are publicly viewable at https://github.com/sarahjtracy/OR2.

Our first aim was to gather the data for all players and teams in the NBA. Fortunately this data is available publicly at stats.nba.com [7]. In order to scrape this data from the website, we ran calls through the open-sources nba_py package [9]. This was initially a complication as nba_py is a poorly-document package still in production. As we also previously mentioned, hustle stats have only been collected recently. We realized later on that the NBA hustle stats could only be collected from 2016 [4]. However, we wished to run our model on data from 2013-14 so that we would have salary data available to use (we assessed that determining the market value of a player by ourselves was an entirely different problem). Still desiring to see the effects of intangible statistics on the value of a player, we opted to use the 2016 hustle statistics data, assuming the playing style of an NBA player would not change drastically over two years, and since it was the best data available. If a player had moved out of the league by 2016, we would use the league average. In the future, our model will be able to run on current statistics as those become more readily tracked.

As mentioned, we also needed salary data and the list of current free agents. This data was available from ESPN, although, lacking a script to pull the data from ESPN's website, we had to scrape the data ourselves [6] [5]. In addition, the ESPN salary data sometimes did not match up with the free agents data. In this case, problems were manually searched since the complications were of a low percentage from the around 150 free agents.

Within our python package, we calculated the value for each player. The mathematical calculations for PER and INT are located back in Section 4, and we converted those same mathematical formulas into Python code for each player.

Having the positional data, value data, salary data, and current free agents list, we could then formulate the linear program in Python. We used the PuLP Python package to run the linear program maximization, running through the free agents list and creating variables and constraint formulas for each constraint mentioned in the problem formulation in section 5.1 [10].

6 Results

6.1 Analyzing the use of INT

We ran our model to simulate free agency on team rosters after the 2013-14 season, using the free agents available in 2014. During this time the league salary cap was about \$63 million. Here we will jump straight into some results:

The San Antonio Spurs, who won the championship in 2014, had ten players remaining with contracts extending past 2014. Below we can see the output of our model, listing first the players, along with their salaries and positions, who remained in the roster, followed by the selection of free agents that optimized our model. The selection of free agents includes a comparison to the selections that would have been made without our additional of intangible statistics:

```
_____
CURRENT ROSTER
_____
Kawhi Leonard 2894059 SMALL FORWARD
Marco Belinelli 2873750 SMALL FORWARD
Danny Green 4025000 SHOOTING GUARD
Cory Joseph 2023261 POINT GUARD
Tony Parker 12500000 POINT GUARD
Jeff Ayres 1828750 POWER FORWARD
Manu Ginobili 7000000 SHOOTING GUARD
Tim Duncan 10361446 POWER FORWARD
Tiago Splitter 9250000 POWER FORWARD
Austin Daye 53838 POWER FORWARD
Calculation including Intangibles
                                        Calculation with just PER
_____
                                        _____
FREE AGENTS
                                        FREE AGENTS
_____
                                        _____
Jameer Nelson 2732000 POINT GUARD
                                        Isaiah Thomas 7238606 POINT GUARD
Greg Monroe 5479933 CENTER
                                        James Jones 915243 SHOOTING GUARD
Cole Aldrich 915243 CENTER
                                        Cole Aldrich 915243 CENTER
Jimmer Fredette 915243 POINT GUARD
                                       Miroslav Raduljica 48028 CENTER
Michael Beasley 53838 SMALL FORWARD
                                       Michael Beasley 53838 SMALL FORWARD
_____
                                        _____
Total Salary = 62906361
                                        Total Salary = 61981062
Total PER = 254.825201157
                                        Total PER = 258.008390608
Total Intangibles = 2.08097385795
                                        Total Intangibles = -6.2908394686
```

Figure 1: Model output for the San Antonio Spurs, 2014

The model opted to fill out the roster with 15 players and a total salary of \$61,981,062. As can

be seen, the Spurs ended all of their contracts with centers in 2014, and consequently our model picked two centers from the free agency pool to add to the team. On the other hand, the remaining roster contained four power forwards, maxing out the need for them, and thus no power forwards were added from the free agency pool.

As noted above, we can view the performance of the model before and after we added the intangibles calculation. The high negative value for intangibles of our team output from a model with just PER as the value of player indicates that this team could be undervaluing defense. Interestingly, our model updated with intangibles calculations opts to still go for free agents Cole Aldrich and Michael Beasley, but swaps out the other three.

While it is impossible for us to say with total certainty that our team output calculated with intangibles is better, we can use our existing understanding of basketball to analyze and predict the success of this team. The addition Greg Monroe gives the Spurs some added defense around the rim (and pairing him up with Tim Duncan creates a formidable shot-blocking, impenetrable duo in the paint). Further, the addition Jameer Nelson adds some speed and hustle to the team. We can expect him to inject this team with energy through his fast hands, which will lead to ball deflections on defense, and some steals which can turn into points on the other end. The Spurs team resulting from the inclusion of intangibles is a much better one on paper.

Let us observe the model output for another team, the Cleveland Cavaliers:

CURRENT ROSTER

```
Jarrett Jack 6300000 POINT GUARD
Kyrie Irving 7070730 POINT GUARD
Dion Waiters 4062000 SHOOTING GUARD
Matthew Dellavedova 816482 SHOOTING GUARD
Sergey Karasev 1533840 SHOOTING GUARD
Tristan Thompson 5138430 CENTER
Anthony Bennett 5563920 POWER FORWARD
Anderson Varejao 9704545 POWER FORWARD
Carrick Felix 510000 SMALL GUARD
Alonzo Gee 915243 SMALL FORWARD
Tyler Zeller 1703760 CENTER
Calculation including Intangibles
                                        Calculation with just PER
_____
                                         -----
FREE AGENTS
                                         FREE AGENTS
_____
                                         _____
Marcin Gortat 10434782 CENTER
                                        Isaiah Thomas 7238606 POINT GUARD
                                        Dirk Nowitzki 7974482 POWER FORWARD
Dirk Nowitzki 7974482 POWER FORWARD
                                        Cole Aldrich 915243 CENTER
Cole Aldrich 915243 CENTER
Michael Beasley 53838 SMALL FORWARD
                                        Michael Beasley 53838 SMALL FORWARD
_____
                                         _____
Total Salary = 62697295
                                        Total Salary = 59501119
Total PER = 216.34633677
                                        Total PER = 219.491695307
Total Intangibles = 12.9492344668
                                        Total Intangibles = 4.729227044
```

Figure 2: Model output for the Cleveland Cavaliers, 2014

The Cavaliers' roster was also filled out in our model. In contrast to the San Antonio Roster, there was only one free agency choice that was changed with the addition of intangibles. Possibly, since intangibles were already a positive value in the model calculated entirely on PER, adding intangible values is not as much of a necessity.

An interesting point to notice is, this is the year LeBron James returned to Cleveland, but he was not chosen in our model from the free agency pool. This note reminds us that there is intangible monetary value from off-court value that cannot be assessed in our model, or, in addition, our model is too linear to give an extra value to hiring star players.

A final observation to note is that both the Cavaliers and the Spurs have models that output they should aim to sign Cole Aldrich and Michael Beasley. While our model outputs the most optimal choices out of the entire free agents pool, it is not guaranteed that a team will be able to acquire all of the top players it desires during a transfer window. Cole Aldrich, for example, stayed with the same team his current contract had ended with. Our model outputs the top choices out of the entire pool of free agents for the year - it would need to be run again as free agents gradually become unavailable as they are acquired by competing teams.

6.2 Further Results - Modeling Team Mindset

Our original model optimized all of the on-court statistics within the PER and Intangibles calculations. To modify the model from there, we wished to take into account the mindset of the team. Based on their win percentage in the previous season, a team can be realistic about its goals for the following season, deciding whether it should aim to perform well in playoffs or whether the current team has a low chance of making playoffs, and thus should work on rebuilding the team for the future.

To account for this, we recognized teams as needing to rebuild if they won less than half of their regular season games in the previous season. If a team fell into this category, they may wish to prioritize future growth over acquiring the best team on the court for the next season. This corresponds to an incentive to hire younger players, hoping they will continue to grow into better performers and thus build a better team over several years.

To quantify this desire, we calculated the average age of players on the court over the season. Then we added an age factor into the value of a player by using the difference between the league average and the age of the player, using the positive difference if a player is younger and negative if the player is older. Multiplying by a vector reduces the variance to a scale more fitting with overall PER. The formulation is seen below:

$AF = 0.5 * (lg_AGE - AGE)$

The age factor will modify the value of the player by a factor comparing the player age to the league average. This modification gave a boost to younger players, and penalized adding older players. By testing with higher weights, we determined a multiple of 0.5 as other data too heavily changed the model.

Since the San Antonio Spurs won the championship in 2014, the were not placed in the rebuilding category, and thus this addition was not added to the model. However, the Cleveland Cavaliers had a win percentage of just .293, and we can see the effect adding the age factor has on our model results:

```
FREE AGENTS

Isaiah Thomas 7238606 POINT GUARD

Dirk Nowitzki 7974482 POWER FORWARD

Cole Aldrich 915243 CENTER

Michael Beasley 53838 SMALL FORWARD

-----

Total Salary = 59501119

Total PER = 219.491695307

Total Intangibles = 4.72922704442
```

Figure 3: Updated Rebuilding Model output for the Cleveland Cavaliers, 2014

Where the previous model had Marcin Gortat, age 30 at the time, the age factor swapped him out for Isaiah Thomas, age 25. Interestingly, Isaiah Thomas was also seen as an output in the model accounting for just PER alone. We can also see that as a tradeoff for prioritizing age, the intangible value receives an 8 point loss, although this allows PER to actually increase slightly.

7 Suggestions for Further Steps

As our analysis stands, we incorporate the traditional five-position system in our free agency selection process. We select players based on their position on paper, whether they are a Point Guard, Shooting Guard, Small Forward, Power Forward or Center. However, not all players of the same position play the same way. For example, Carmelo Anthony and Kawhi Leonard are both Small Forwards, but the impact they have on their respective teams are very different. Carmelo Anthony is an offensive-minded mid-range shooter, while Kawhi Leonard is a defensive specialist and a 3point shooter.

With this in mind, we would like to more deeply analyze player style beyond the traditional five-position system. We would like to cluster every player in the NBA with like players based on play style, rather than position. Furthermore, we can apply this cluster analysis to existing teams to understand which type or types of players appear more on winning teams. This allows us to better understand the needs of the team, and therefore select a free agent that is a better fit for the team.

Additionally, as previously mentioned, we explored slightly the effect of different team mindsets. Our model currently only flips a switch, as to say, if a team is unlikely to be a playoff contender. If a team is of this type, we call it a rebuilding year and incentivize hiring younger players. However, there are many variations of team mindsets. A team that is not only close to a playoff spot, but is likely a championship contender, may want to exploring taking advantage of salary cap exceptions by overspending in an attempt to win the championship. This is just one example of an additional team mindset that could be explored. With extra time we would like to further explore the weights of certain values in our model. Since our model is linear, adding several bench players each having a small increase in value would be treated as equal to signing one star player. But since the star player would have a great time on the court, it is possible they should be much more desired. This was explored slightly by increasing the value of free agents who are higher valued than the best player in their position on the team that is looking to acquire them. Unfortunately, while this first added some higher-value players, since all players are selected at the same time, the data skewed towards adding too many of those positions which were originally undervalued. In order to solve this, we would need to update the model to add players one at a time and update the maximum values for each position in the team accordingly.

Another reason to update the model to add players one at a time is to better simulate how the market acts in the real world. As seen in our examples, sometimes multiple teams desire the same player. The player can only sign for one team, and this could be modeled by adding players to teams one at a time and then updating the available free agent pool. Updating the model to act on single players requires a large update in the code database.

Conclusion

Because the actual free agents that a team signs are dependent on many other variables, it is difficult to compare our results to the free agent transfers that occurred. Our model outputs the player choices that would optimize our value of the player, given by Player Efficiency Rating combing with our Intangibles piece, but whether that player would actually sign for the given team depends on the specific desires of that player, and the other teams in the league that also wish to acquire that player.

So forth, the interesting results of our model are how team selection can vary based on how one assesses the value of a player. As we hypothesized, adding an intangibles calculation to the model gives different results and more defensive-minded teams than assessing players strictly on PER.

We took several assumptions on how teams value the attributes maximized by PER, those maximized by hustle statistics, and the value of an underaged team. Our model could be additionally changed with different weights to these attributes depending on the mindset of the team. We explored one mindset, rebuilding a team with younger players, with some interesting results, but this mindset could greatly vary across the league and could definitely be explored further.

References

- "Calculating PER." Blog post. Basketball-Reference.com. Sports Reference LLC, 5 Apr. 2007. Web. 14 Dec. 2016. http://www.basketball-reference.com/about/per.html.
- [2] "A Cluster Analysis of NBA Players." Sloan Sports Conference (n.d.): n. pag. Web. http://www.sloansportsconference.com/wp-content/uploads/2012/02/44-Lutz_cluster _analysis_NBA.pdf.
- [3] "Hollinger NBA Player Statistics." ESPN.com. ESPN Internet Ventures, 2016. Web. 14 Dec. 2016. http://insider.espn.com/nba/hollinger/statistics/_/year/2014.

- [4] "Hustle Stats." NBA.com/Stats. NBA Media Ventures, LLC, 2016. Web. 14 Dec. 2016. http://stats.nba.com/hustle/.
- [5] "NBA Free Agents Tracker." Spotrac.com. Spotrac, 2016. Web. 14 Dec. 2016. http://www.spotrac.com/nba/free-agents/2014/.
- [6] "NBA Player Salaries." ESPN.com. ESPN Internet Ventures, 2016. Web. 14 Dec. 2016. http://www.espn.com/nba/salaries/_/year/2015.
- [7] "NBA Stats NBA.com." NBA.com/Stats. NBA Media Ventures, LLC, 2016. Web. 14 Dec. 2016. http://stats.nba.com/.
- [8] "Player Efficiency Rating." Wikipedia. Wikimedia Foundation, 14 Dec. 2016. Web. 14 Dec. 2016. https://en.wikipedia.org/wiki/Player_efficiency_rating.
- [9] nba_py. Github Repository. 2016. https://github.com/seemethere/nba_py/tree/master/nba_py
- [10] PuLP. Version 1.6.1. Python Software Foundation. 2016. https://pypi.python.org/pypi /PuLP