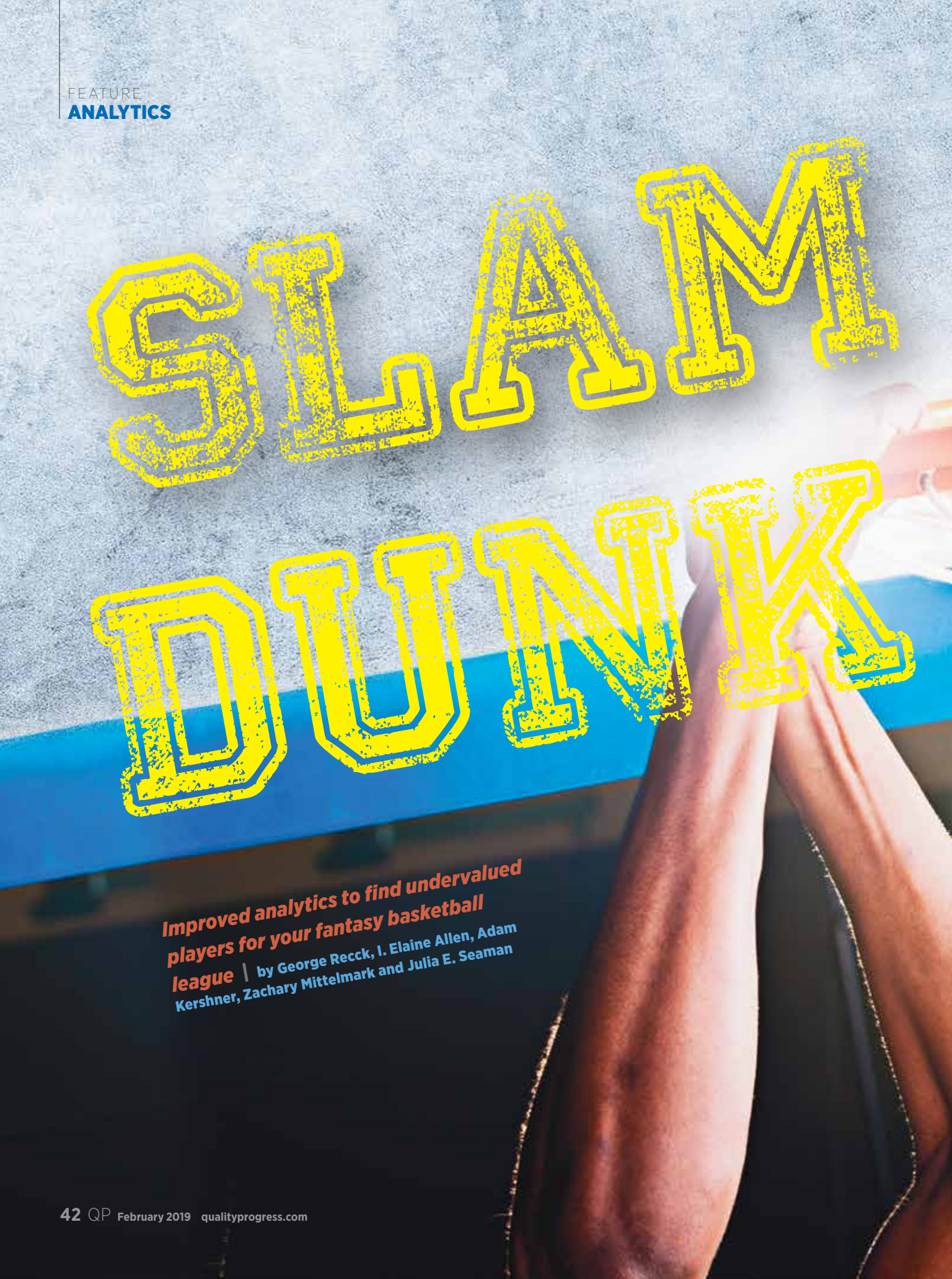
SLAM DUNK

Improved analytics to find undervalued players for your fantasy basketball league

by George Recck, I. Elaine Allen, Adam Kershner, Zachary Mittelmark and Julia E. Seaman



Just the Facts

Fantasy basketball leagues are becoming more popular, and rankings and analytics are an important part of managing a team.

To identify the best players to select for your fantasy team, a new rating system has been developed.

This player valuation model examines how much an individual player contributes to a team's overall field goal percentage and free throw percentage.

In the past 30 years, we have entered the world of big data. The magnitude and accessibility of data—and the ability to analyze those data—have advanced dramatically. This advancement of data processing power can be seen in diverse fields, including investing, web analytics and sports metrics.

Following this trend, fantasy sports have significantly developed, both in terms of sophistication and popularity. Rankings and analytics in sports leagues are everywhere and becoming more important. One of the easiest places to see this is in fantasy sports because the data are public, and there are online business and betting opportunities. The difficulty in summarizing this data for choosing players for fantasy sports teams lies in identifying which metrics are best at discriminating between the highest quality players and the rest of those in the league.

This article focuses on fantasy basketball, introducing a new method to estimate the overall quality of a given basketball player, contrasted in tandem with current procedures used by sports websites to estimate player value. This approach has led a few of this article's authors to three championships and two second-place finishes in the last eight years in a mature fantasy basketball league.

TABLE 1

Fantasy league points awarded for PTS and AST: 2016-2017

Team	PTS	Points
SGFF	**8,752**	**9**
SalesDog	**8,719**	**8**
Mr. C	**8,352**	**7**
The Illini	**8,235**	**6**
Crew	**8,156**	**5**
Viper	**8,119**	**4**
Ghost	**8,012**	**3**
Cobra	**7,658**	**2**
FengD	**7,284**	**1**

Team	AST	Points
Mr. C	**2,083**	**9**
SalesDog	**2,044**	**8**
The Illini	**1,954**	**7**
FengD	**1,919**	**6**
Crew	**1,917**	**5**
Viper	**1,877**	**4**
SGFF	**1,792**	**3**
Cobra	**1,627**	**2**
Ghost	**1,207**	**1**

***AST** = assists* ***PTS** = points scored*

Fantasy basketball overview

Previously overshadowed by baseball and football, fantasy basketball has been growing. In fact, fantasy basketball revenue surpassed that of fantasy football in 2016 at FanDuel, and it was projected to do so in 2017 at DraftKings. These two sites represent the largest fantasy gambling outlets.[1] In addition, a recent ruling by the U.S. Supreme Court gave states the ability to authorize sports betting, which all but nine states have now allowed.[2]

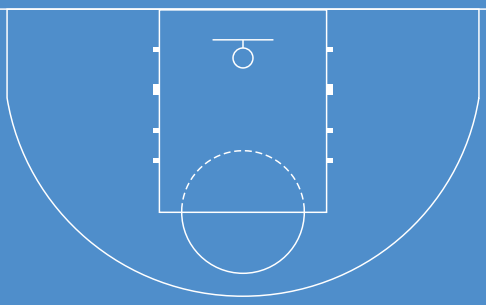
Many believe the rapid growth of fantasy basketball is due to sports fans' desire to remain engaged with sports and fantasy leagues during Major League Baseball's off-season. Like baseball, the National Basketball Association (NBA) has many games played daily, many traditional as well as new statistical scoring measurements, and the advent of digital tracking to provide endless amounts of data per game or player. In addition, data analysis opportunities are numerous: Most data in basketball are larger in magnitude than other sports, and much are continuous or pseudo-continuous in comparison to baseball.

While it varies by group and hosting site, anywhere from four to 20 people can form a fantasy basketball league during the NBA season. The participants act as general managers (GM) for their respective teams. These GMs meet to draft a roster of players from the pool of NBA players. After a player is drafted, no other GM can draft that player for his or her team. More recently, drafts are typically held online. Drafts may be held all at once, over several days or even over several weeks. Many leagues place dollar values on players and have a corresponding salary cap, or have position or play time requirements, to restrict and control team makeups. Most leagues even allow trading to occur among GMs.

During the season, after the draft is completed, GMs set their starting lineups for the starting period. This can be done on a daily, weekly or monthly basis, or just once for the whole NBA season, depending on the fantasy league preferences. Thereafter, statistics for the players in a GM's starting lineup accumulate and are automatically ranked. The overall goal is to earn the most points in the league to win the season.

Despite the potential to win cash, the return on time invested is typically lower than minimum wage. Thus, participating in a fantasy league is done purely out of enthusiasm for the game and bragging rights. Many leagues even have intermediate prizes (monthly or daily), which can keep a team out of the race for first place engaged in the league. These are the typical eight categories involved in calculating an overall fantasy NBA team ranking:

+ Assists (AST).
+ Blocks (BK).
+ Field goal percentage (FGP).
+ Free throw percentage (FTP).
+ Points scored (PTS).
+ Three-pointers made (3PT)
+ Total rebounds (TRB).
+ Steals (ST).



GENERAL CALCULATION OF ADDED VALUE

Tables 4 and 5 show how to calculate this added value to FGP or FTP for an individual player. Below are the equations to do so for any player where AvgFG and AvgFT represent the team's average number of field goals and free throws multiplied by four to represent the other four players. AvgFGA and AvgFTA represent average attempts, while FB, FGA, FT and FTA represent those made by the individual player being evaluated.

$$FG\ Add = \frac{(4*AvgFG)+FG}{(4*AvgFGA)+FGA} - \frac{AvgFG}{AvgFGA}$$

$$FT\ Add = \frac{(4*Avg\ FT)+FT}{(4*Avg\ FTA)+FTA} - \frac{Avg\ FT}{Avg\ FTA}$$

After applying this calculation to the top 50 players, we can build the following regression models:

- **Predict FG Add based on FG made, FG missed and FG attempted**
- **Predict FT Add based on FT made, FT missed and FT attempted.**

Calculating league standings

Table 1 illustrates how league points were awarded for PTS and AST categories based on the statistics accumulated over the 2016-2017 season of a mature fantasy NBA league.

The table shows that regardless of whether a player wins a category by one or 1,000, only nine points are assigned to the category—thus, you need to win a category by just a single unit to accumulate nine points. You might think of this in the same way as winning a bid on a contract: A business bidding for services would want to offer just enough to beat its competitors, without sacrificing too much of its own resources.

Table 2 (p. 46) shows how the points earned in each ranked statistical category throughout the season are totaled to determine the league's ending standings. It is not necessary for a GM to win the most points in every category, but rather have the highest average value per ranked category. Ideally, the GM should aim to score a seven across the categories for the best chance to win the league.

Valuing players

A player's value for fantasy leagues is based on his statistics in the ranking categories. In today's modern setting, actual NBA data are readily available online from sources such as CBS Sports, ESPN and Yahoo Sports, among others. Most of these sites provide rankings for NBA players based on their proprietary formulas that they promote for use in fantasy leagues. However, while these sites provide a generalized overview of the league, not all fantasy leagues measure the same categories, so rankings may not be precisely tailored for a specific league.

We were interested in deciphering these rankings for two reasons:

1. To determine their relevance to the eight fantasy scoring categories.
2. To evaluate the quality of current ranking schemes and identify higher quality metrics for choosing players.

For sake of brevity, all further analyses in this article will use the CBS Sports rankings.

Multivariable modeling

Multivariable linear regression models are used to evaluate the current ranking metrics using a separate model by year for nine NBA seasons. Each model calculates a coefficient for each of the eight metrics as predictors of the overall CBS rank. Using this method, each predictor's coefficient controls for the effect of the other seven metrics, and its statistical significance in the overall model is evaluated. Both the size of the coefficient and its consistency over multiple years are important.

This model uses the last nine NBA seasons (from 2008-2009 to 2016-2017) based on the top 100 players as ranked by CBS Sports. Table 3 (p. 47) offers this nine-year history of regression coefficients for each statistical category and their relative strength.

The strength of a category can be measured by the size (weight) of the regression coefficients.

TABLE 2

Final fantasy league standings: 2016-2017

Rank	Team	3PT	AST	BK	FGP	FTP	PTS	ST	TRB	Total
1st	Illini	3	8	7	9	3	6	6	8	50
2nd	SalesDog	9	3	8	5	2	8	5	3	43
3rd	Viper	8	6	4	3	9	4	3	5	42
3rd	Cobra	7	5	2	8	8	2	8	2	42
5th	Mr. C	4	9	9	6	1	7	1	4	41
5th	SGFF	6	4	3	2	6	9	2	9	41
5th	Ghost	5	7	1	1	7	3	9	1	34
7th	FengD	1	2	6	7	5	1	7	6	35
9th	Crew	5	1	8	1	4	1	2	3	25

AST *= assists* ***BK*** *= blocks* ***FGP*** *= field goal percentage* ***FTP*** *= free throw percentage*
PTS *= points scored* ***3PT*** *= three-pointers made* ***TRB*** *= total rebounds* ***ST*** *= steals*

The change in importance in a particular category can be seen in the colors of the heatmap (with red indicating the most important categories and white identifying the least important categories).

Note that some categories are highly significant ($p < 0.001$) but do not add much weight to the overall ranking because of the small size of the coefficient. In the heatmap, coefficients of 0.4 or greater are colored in red, coefficients between 0.3 and 0.39 are colored in dark orange, coefficients between 0.2 and 0.29 are colored in light orange, and the remaining coefficients have no coloring.

Throughout the years, it's easy to note the consistently strong (low) p-values for the first six categories (3PT, TRB, AST, ST, BK and PTS) throughout the nine-year history. However, FGP and FTP emerge as statistically insignificant most of the time. In the three seasons that FTP is significant, its significance is borderline enough that it is still not a strong predictor of rank, especially in comparison to the six significant variables.

The high p-values in the six years that FTP is statistically insignificant also indicate that its occasionally low p-value is likely more a fluke than anything statistically meaningful. There are also variabilities in the strength of the associations, as noted in the heatmap colors in Table 3. 3PT, ST and BK are consistently stronger than the other three statistics that show consistent significance.

These results indicate that FGP and FTP are not strong predictors of rank, even though they are an important aspect of the fantasy scoring component. This is further illustrated by comparing an initial model that includes FGP and FTP to a new model without FGP and FTP. Here, the R-squared value in the initial model was 81.5%, whereas the R-squared value in the new model that excludes FGP and FTP is 80.7%. Such a small decrease in this value indicates that FGP and FTP, while statistically significant variables, contribute a very small amount to explaining variation in rankings.

Looking only at the individual regression coefficients and p-values for the covariates for the 2016-2017 season (the last column), we can use these coefficients to form the ranking model that follows.

The regression equation for 2016-2017 to predict CBS rank is:

CBS Sports rank = 212.6 + 0.22(3PT) + 0.03(TRB) + 0.04(AST) + 0.29(ST) + 0.35(BK) + 0.01(PTS) + 0.57(FGP) + 0.34(FTP)

With the regression analysis results in Table 3, we have shown that sports websites, such as CBS Sports, do not consider FGP and FTP as high-quality metrics in developing their player rankings. This is a major flaw given the importance of these statistics as fantasy point categories. Our subsequent analysis introduces a new method to more accurately value players of the highest quality for fantasy leagues and further analyze CBS Sports' method to rank players.

Needle moving for a new valuation model

"Needle moving" represents a new approach to valuing success in the FGP and FTP categories, raising their value in player evaluation. Essentially, we asked ourselves, "How much does player X improve a team's overall FGP or FTP?" In baseball, a similar term—wins above replacement—has also become a popular fan statistic.

To answer this question, we calculated the added (or lost) value each player contributes to the team's overall FGP or FTP as follows:

1. We calculated the average field goal (FG) made and free throw (FT) data for the top 50 players to create a baseline of the average player. We chose the top 50 players as opposed to the top 100 to include those players likely to play most of the games. Thus, their average is the profile of the average player they are likely to replace.

TABLE 3

Heatmap and significance of regression coefficients 2008-2017 for each category

Category	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17
3PT	**0.34*****	**0.36*****	**0.40*****	**0.44*****	**0.30*****	**0.31*****	**0.25*****	**0.21*****	**0.22*****
TRB	**0.07*****	**0.07*****	**0.08*****	**0.10*****	**0.04*****	**0.05*****	**0.06*****	**0.05*****	**0.03*****
AST	**0.08*****	**0.08*****	**0.10*****	**0.12*****	**0.09*****	**0.08*****	**0.05*****	**0.05*****	**0.04*****
ST	**0.23*****	**0.32*****	**0.31*****	**0.43*****	**0.27*****	**0.33*****	**0.27*****	**0.25*****	**0.29*****
BK	**0.29*****	**0.33*****	**0.43*****	**0.54*****	**0.41*****	**0.41*****	**0.36*****	**0.33*****	**0.35*****
PTS	**0.02*****	**0.02*****	**0.01*****	**0.02*****	**0.01***	**0.02**	**0.02****	**0.01****	**0.01*****
FGP	**0.07**	**0.05**	**0.21**	**0.26**	**0.30**	**0.07**	**0.26**	**0.52**	**0.57**
FTP	**0.25**	**0.21**	**0.30***	**0.18***	**0.29**	**0.06**	**0.35**	**0.20**	**0.34***

*** < 0.05 ** < 0.01 *** < 0.001**

AST *= assists* ***BK*** *= blocks* ***FGP*** *= field goal percentage* ***FTP*** *= free throw percentage*
PTS *= points scored* ***3PT*** *= three-pointers made* ***TRB*** *= total rebounds* ***ST*** *= steals*

2. We incorporated the FG and FT data for an individual player of interest.
3. The change in FGP or FTP from the four-player team to the five-player team that includes player X represents how much does player X moves the needle. We call this the added (or subtracted) value that player contributes to (or costs) the team.

The scope of the analysis that follows focuses on the 2016-2017 season. Table 4 (p. 48) shows a calculation for NBA player Kevin Durant's added value to a four-player team's FGP. We define a player's added (or subtracted) value calculated in step three as FG Add (when calculating the change in FGP) or FT Add (when calculating the change in FTP).

Table 5 (p. 48) gives the same calculation using FT added value rather than for FGs. Again, Durant represents player X. As in Table 4, the values for players one through four represent the averages for the top 50 players. See the sidebar "General Calculation of Added Value" for more explanation.

Between the two years for which we performed the earlier calculations, there was a small increase in the added value, although this was not significant. We use the coefficients from these models to calculate our new player valuation model.

Using the data from Tables 4 and 5, we get the following regression equations:

FG Add = -0.02 + 0.018(FG) - 0.0087(FGA)

FT Add= 0.097 + 0.074(FT) - 0.06(FTA)

Using these regression coefficients, we can now create a new statistic for the percentage categories: field goal score (FGS) and free throw score (FTS).

FGS= (1.8*FG) - (0.87*FGA)

Each field goal made receives 1.8 points, while a penalty of 0.87 points is assigned to each field goal attempted.

FTS= (7.4*FT) - (6*FTA)

Each free throw made receives 7.4 points, while a penalty of six points is assigned to each free throw attempted. We use these scores in our construction of a new model: the player valuation model.

New player valuation model

For a team to have the best chance of winning the league, the team should try to achieve a "seven," or close to it, in each category. Given that there are typically eight statistical categories, we can assume that a score of 56 (eight categories x seven) points should win the league. Of course, the number of points it takes to win the league will vary slightly by season, but 56 points is a good goal to set.

To build the player valuation model, we begin with PTS as a baseline value. From Table 1, in 2016-2017, to "beat the seven" in points, 8,352 points were needed. In addition, to "beat the seven" in assists, 1,954 points were needed. The ratio of these two-point totals shows that each assist is worth 4.3 points.

With this ratio, the individual category weights for each statistical category are determined in "points." These weights

TABLE 4

Needle moving—FG 2016-2017

Player #	FG made	FG missed	FG attempts	FGP
Player 1	**492**	**586**	**1,078**	**45.6%**
Player 2	**492**	**586**	**1,078**	**45.6%**
Player 3	**492**	**586**	**1,078**	**45.6%**
Player 4	**492**	**586**	**1,078**	**45.6%**
Team total	**1,968**	**2,344**	**4,312**	**45.6%**
Durant "Player X"	**698**	**683**	**1,381**	**50.5%**
				New FGP
New team total	**2,666**	**3,027**	**5,693**	**46.8%**
		Kevin Durant's added value to FGP		**1.2%**

***FG** = field goals* ***FGP** = field goal percentage*

TABLE 5

Needle moving—FT 2016-2017

Player #	FT made	FT missed	FT attempts	FTP
Player 1	**266**	**72**	**338**	**78.7%**
Player 2	**266**	**72**	**338**	**78.7%**
Player 3	**266**	**72**	**338**	**78.7%**
Player 4	**266**	**72**	**338**	**78.7%**
Team total	**1,064**	**288**	**1,352**	**78.7%**
Durant "Player X"	**447**	**51**	**498**	**90.0%**
				New FTP
New team total	**1,511**	**339**	**1,850**	**81.7%**
		Kevin Durant's added value to FTP		**3%**

***FT** = free throws* ***FTP** = free throw percentage*

are used to calculate a rating for each player. Note that these category weights vary slightly from year to year based on that season's rankings, similar to the regression coefficients. However, as with regression coefficients—except in the case of a significant trend—this variation is small and therefore can be used to determine next season's points thresholds.

We call this newly calculated rating the new valuation rating (NVR), which can be calculated as shown below. This rating incorporates the new approach to FGS and FTS introduced in the needle moving section.

NVR= (1*PTS) + 11.7(3PT) + (2.8*TRB) + (4.3*AST) + 17*ST) + (21.5*BK) + [(2.5*FG) - (1.1*FGA)] + [(7.4*FT) - (6*FTA)]

This new rating system helps identify undervalued—or falsely overvalued—players from traditional ranking systems by using points as a baseline value. In addition, players are rated on a per-game basis, which accounts for how frequently they are playing. This can often drastically increase (or decrease) a player's value, causing an inconsistency between the rating and the CBS ranking. In addition, the new FGS and FTS metrics incorporate the added (or lost) value each player brings to the team.

The player valuation model using the rating will help a fantasy league participant analyze the value of each individual player to construct a well-rounded team capable of winning in many scenarios. Because it is nearly impossible to attempt to win all eight categories, this model values players to "beat the seven," which, if achieved in all statistical categories, nearly guarantees winning the league overall.

Figure 1 is a histogram that shows the number of players plotted against the NVR for the 2016-2017 season. Therefore, after getting past the elite players, this model can be used to find undervalued players to acquire or overvalued players to avoid. The histogram shows most players in the 90-150 range with two elite players, James Harden and Russell Westbrook, who are well above 220.

For example, DeMar DeRozan was ranked 72nd by CBS, but had an NVR of 150 that ranked 23rd overall. This suggests DeRozan may be a better fantasy draft pick than CBS expected. This is from the fact that his FG and FT scores may have been undervalued by CBS.

There are many more examples of differences between the CBS rank and the NVR, and these players may represent the under or overvalued draft picks for a fantasy team. While optimizing single player value is important, complementing across the combined team statistics is also important to score in every ranked category. Draymond Green, ranked by CBS as seventh overall, had an NVR of 149, which ranks 25th because he had negative FG and FT scores.

Isaiah Thomas, Kawhi Leonard, Damian Lillard or Jimmy Butler would all serve as better picks over Green, given that they all had an NVR of 175 or better, with strong FG and FT scores. Note that all of these players played between 74 and 76 games,

which is important because CBS creates its ranking based on aggregate totals rather than per-game totals, so there is a dampening effect for players who played many games versus players with a few good games.

New predictor model for CBS rank

Using the newly calculated score values (FGS and FTS) for the top 100 players, we return to predicting CBS rank using a regression analysis. The purpose of this second attempt at predicting CBS rank is to analyze the p-values for FGS and FTS, to evaluate whether websites such as CBS Sports are incorporating FG and FT data in any way to calculate rankings. We already established that the raw percentages are not significant, but we now analyze whether the weighted scores we created would be meaningful impactors in predicting CBS rank.

Table 6 (p. 50) shows the regression output for predicting CBS rank based on the same eight categories, except with scores (FGS and FTS) rather than points (FGP and FTP).

Here, as with our initial model discussed in Tables 2 and 3, the first six categories (3PT, TRB, AST, ST, BK and PTS) are highly significant in predicting CBS rank, whereas the remaining two (in this case, FGS and FTS) are not significant predictors. This result confirms that CBS Sports is not incorporating FG and FT data into its player rankings, indicating a possible flaw in its rankings.

While the FGS and FTS metrics we created are not statistically significant predictors of CBS rank, they help identify the added (or subtracted) value an individual player brings to a given team and increase the quality of player rank by giving more weight to consistency of play and added value to a team.

Player valuation model and imbalances

Despite the plethora of data now available to analysts, popular fantasy sports websites such as CBS Sports, ESPN and Yahoo Sports are at a loss for how to use some of that data to rank players. Through numerous regression models over a nine-year span, we saw that the percentage categories (FGP and FTP) are not statistically significant predictors of player rank.

We created a new rating system—rather than a ranking system—to identify hidden value in players (or even sometimes, overvalued players). To do this, we began by what we call "needle moving," which asks how much an individual player contributes to a team's overall FGP or FTP. This player's contribution may either increase or decrease the team's percentage.

Not only does this help to identify higher-quality players, but by building a regression model to predict this added value ("FG Add" or "FT Add"), we can use those regression coefficients

FIGURE 1

Histogram of new valuation ratings 2016-2017

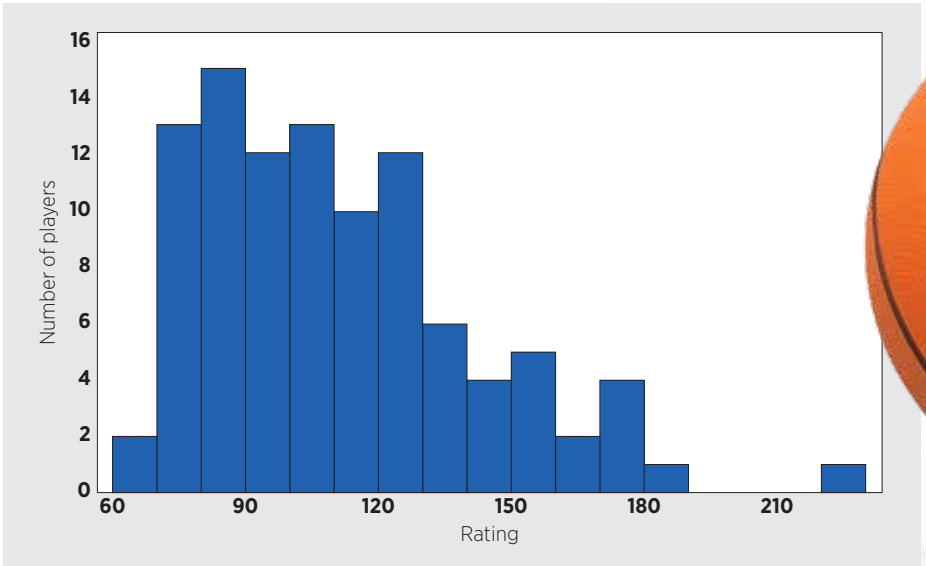


TABLE 6

Revised model coefficients and p-values: 2016-2017

Predictor	Coefficient	P-value
3PT	0.23	0.00
TRB	0.03	0.01
AST	0.04	0.00
ST	0.28	0.00
BK	0.35	0.00
PTS	0.02	0.01
FGS	0.04	0.06
FTS	0.01	0.36

***AST** = assists*
***BK** = blocks*
***FGP** = field goal percentage*
***FTP** = free throw percentage*
***PTS** = points scored*
***3PT** = three-pointers made*
***TRB** = total rebounds*
***ST** = steals*

to create a new metric for the field goal and free throw categories: FGS and FTS.

Thus, we created the player valuation model that seeks to rate each player based on per-game performance. Seeking to "beat the seven" in each statistical category (excluding FGs and FTs), we calculated the ratio of points to a given category, using the values that achieved "the seven." This ratio represents that statistical category's weight. By combining these weights with the FGS and FTS metrics we created, we arrive at the new valuation rating.

For future research, the player valuation model can be adjusted to identify trends before the regression analysis. For example, the nine-year history of regression coefficients shows an increase in value and a devaluing of three pointers as more and more players become high-quality three-point shooters. Our category weights in the player valuation model agree with this, and also show the same trend. However, this model would be more useful if it could identify this trend first, among several other potential trends.

You also might use the player valuation model in conjunction with scheduling imbalances, where NBA teams win several consecutive games, by taking advantage of the imbalances in a week-to-week format. This could greatly enhance your fantasy team's performance and competitive advantage over other players in the league who are only using the website rankings to select players for their roster. QP

REFERENCES

1. Daniel Roberts, "The Growth in Fantasy Sports Will Not Come From Football," Yahoo! Finance, Sept. 28, 2017, https://tinyurl.com/yahoo-fantasty-basket.
2. The Supreme Court of the United States, "Murphy, Governor of New Jersey, et al. v. National Collegiate Athletic Assn. et al." opinion, May 24, 2018, www.supremecourt.gov/opinions/17pdf/16-476_dbfi.pdf.

George Recck is director of the math resource center at Babson College in Wellesley, MA, and founder of Total Information Inc., an information consulting businesses providing service to small businesses. He holds an MBA from Babson College.

I. Elaine Allen is professor of biostatistics at the University of California, San Francisco, and emeritus professor of statistics at Babson College. She is also director of the Babson Survey Research Group. She earned a doctorate in statistics from Cornell University in Ithaca, NY. Allen is a member of ASQ.

Adam Kershner is a senior at Babson College with a concentration in business analytics. After graduation, he will begin a career in consulting at Ernst & Young in transaction advisory services.

Zachary Mittelmark is a senior associate at PNC Riverarch Capital in Pittsburgh. He has a bachelor's degree in business administration from Babson College.

Julia E. Seaman is research director of the Quahog Research Group and a statistical consultant for the Babson Survey Research Group at Babson College. She earned her doctorate in pharmaceutical chemistry and pharmacogenomics from the University of California, San Francisco.