An Imperfect Science of NBA Scouting

Josh LaFronz

February 1, 2026

Introduction

The NBA Draft is a structured crapshoot. Overlooked players break out, can't-miss prospects flame out, and others perform as expected. That doesn't even factor in the luck of the draft lottery. Franchises can change with just one decision: Either they pick the guy that will change their franchise trajectory, or they'll be subject to the despair of the what-ifs: "Look who team X passed on," "Player Y was drafted five picks later," "Player Y became a multi-time all-star while player Z is out of the league."

To guide this pivotal decision, scouts and team executives often take the prospect's body of work, collegiate statistics in this case, with a grain of salt. They're not trying to build an NCAA all-star team. They examine a prospect's college performance for traits and skills that will translate from an amateur game to a professional league. One of those traits is perimeter shooting.

In the modern and analytical NBA, shooting ability is a huge pivot point for a prospect's scouting report. Questions about a prospect's offensive polish can lead to a drop in the draft. It can be tricky to predict. Despite the uncertainty, there is a scouting staple that exists and is cited by draft analysts. College three-point shooting percentage is not truly the best metric for projecting that long term shooting capability. Instead, the convention is to look at a college prospect's free throw shooting. A high free throw percentage can be indicative of a solid foundation: good shooting mechanics and a feel for the game.

I've always accepted this staple. I have anecdotal evidence of players I was personally wrong about turning into good shooters: Jabari Smith, TyTy Washington, Ochai Agbaji. I want to analyze how much better college $FT\%$ is at projecting NBA shooting, as well as to highlight the other readily available college stats that are strong predictors. I'm going to break down my entire process from gathering data, pivoting in focus, model building, and evaluation. I found that college $FT\%$ is a superior predictor compared to college $3P\%$, and other contextualizing statistics (volume, age, and attempt rate) were also strong predictors.

Building a Dataset

Putting together a dataset was a surprisingly challenging task. Scraping the data was straightforward, but setting thresholds for inclusion was not very straightforward.

Initially, I thought it was best to include players with large NBA sample sizes; to include only very established NBA players, whether it be stars, multi-year veterans, or quality rotation players. My first thresholds were collecting all college players and draft prospects who debuted within the years 2000 to 2020 who played a minimum of 5000 career regular season minutes and averaged at least two three-point attempts per game. However, I did find a few issues with this set of thresholds. It's too restrictive; the goal of this project is to investigate the relationship between college free throw shooting and NBA three-point shooting, and these limits exclude players who perhaps flamed out or shot well, but didn't add much otherwise. I can't solely include successful NBA players in the dataset; I need to include those who didn't shoot well enough in an effort to carve out an NBA role. Also, two three-point attempts per game was a steep limit; many players either add a shot later in their career, or they are capable shooters who didn't attempt threes as often. When I actually performed modeling, I found that a model trained on relaxed thresholds was able to generalize and perform slightly better on unfamiliar data. I decided to proceed with career thresholds of at least 1500 minutes and one three-point attempt per game on NBA rookies and draft picks from 2000-20; 515 players in total.

Model Building

I had several different variables in consideration for model selection. The variables covered shooting, age, volume, and even physical traits.

- Accuracy
  - $nba3P\%$ - NBA Three-Point Percentage (**The parameter of interest**)
  - $c3P\%$ - College Three-Point Percentage
  - $cFT\%$ - College Free Throw Percentage
- Volume
  - cG - Total College Games Played
  - $cFGA$ - Total College Field Goal Attempts
  - $cFTA$ - Total College Free Throw Attempts
- Rate-Based Volume
  - $cFTr$ - College Free Throw Attempt Rate (FTA/FGA)
  - $c3PAr$ - College Three-Point Attempt Rate ($3PA/FGA$)
- Physical Characteristics
  - $draftAge$ - Age of a prospect on the day of the draft that followed his last college season
  - $height$ - Height in inches

When I first approached building models, I only considered college freshmen for my training and testing data. I collected more data from players from 2020-23 to perform testing on. However, I encountered a few problems that led me to pivot to using college players of any career length.

First, the combined train and test datasets were very small. There were 124 players in the training set and 35 in the test set. Along with the small test set, given that the maximum NBA career length was six years, it's noisy, and many players within it are not finished products given that their careers are so short. The train set contains players with short careers, but the test set is entirely composed of short careers. The performance results were very scattered. So I shifted from that split back to the previous dataset of 515 college prospects of all career lengths from 2000-20.

In addition to expanding the scope, I also altered my choice of resampling method. Initially, I opted for a typical 80/20 train-test split, but I was disturbed at what happened when I simply altered the random seed that controlled how the train-test split was divided. Here's an excerpt of what happened when I changed the seed and nothing else: Same model, same larger dataset.

$$nba3P\% \sim c3PAr + cFT\% + cFTr + draftAge + cG + cFTA$$

Seed	Test MSE	Spearman's ρ
0	9e-04	.52
1	0.053	.32

Table 1: I'm going to be sick

Suffice it to say that my dataset had some outliers. In the next section, I'll expand on how some of those outliers came to be. To address this issue of seed misfortune, I shifted from a single train-test split to a 5-fold CV approach. From there, I could calculate the mean and spread of several evaluation metrics to determine the model quality.

Additionally, I had doubts about my relatively low threshold of 1500 minutes. Such a relaxed barrier of entry led to the inclusion of names like Brandon Goodwin and Manny Harris, a pair of undrafted guards who both played about 1600 career NBA minutes. This is not meant to disrespect either of them. I was personally in attendance for Manny Harris's best NBA game: a 27-point explosion against the Phoenix Suns. My qualm was that it felt wrong to include such players with such brief NBA stays, despite the intuition of including such players to expose the model to players who didn't pan out. However, that anxiety was eased due to the pair of models that I utilized:

1. **Binomial GLM**: The response variable is $nba3P\%$ with weights corresponding to the volume of both makes and misses.

2. **Weighted Least Squares**: The response variable is $nba3P\%$ with weights corresponding to the total number of NBA minutes played.

Those weighted models address the issue of treating player A and player B the same, when player A played 1500 minutes, while player B logged 15000 career minutes.

After all these fires were put out, I was able to proceed with my analysis.

Model Evaluation

In building models, I used the following goodness-of-fit tests as metrics:

- Deviance and RSS
- AIC
- BIC

- Mallow's C_p
- R^2_{adj} and pseudo R^2
- Test MSE
- Spearman's ρ
- Kendall's τ

The last two metrics were especially important in my evaluation process. After some thought and having read a fantastic article that covers this same topic, I realized that the goal is not necessarily to predict a point estimate of a player's $3P\%$. Rather, the goal is to assert that "player A will be a better three-point shooter than player B," and in that lies the scouting decision.

Spearman's ρ and Kendall's τ are both non-parametric ordinal tests for association between two sets of ranks. In this case, the test was to show an association between actual $3P\%$ rank and predicted rank. A small enough p-value indicates that the association between the actual/predicted ranks can not reasonably be explained by chance, and it also indicates model strength. Both Spearman and Kendall tests are structured like this. Without loss of generality,

$$Ho : \tau = 0$$

$$Ha : \tau \neq 0$$

The farther the reported τ or ρ value is from zero, the stronger the association between the two data vectors. This was a huge philosophical breakthrough because I was initially getting discouraged by the models' inability to accurately predict extremes. When a player actually shot 29% in their career, the models would say 32-34%; the same goes for a 43% shooter. Both the binomial GLM and weighted LM struggled to extend predictions beyond the 32-39% range. That's where the factor of ordering comes in; those shooters with predicted $3P\%$ around the 32% mark were scraping the bottom of the projections' range, and that was where the truth shined! Ordinal metrics added a layer of validity to strong models that was easier to discern beyond minute changes in Test MSE. We are dealing with decimals here.

After many iterations of training, I decided to proceed with the weighted least squares approach instead of the binomial regression model. The weighted LMs on average produced better values of Test MSE and ordinal estimates. They were also often in accordance as to which models were stronger; the same model selection decision typically had the same result in both model types. With both factors in mind, that's why I decided to continue with just the weighted linear model, which I am going to leverage to argue that $cFT\%$ is a better predictor of $nba3P\%$ than $c3P\%$.

Argument I: Correlation

Given my dataset of college prospects from 2000-20, here's the correlation matrix of the three aforementioned variables.

```
            nba3Ppct     c3Ppct    cFTpct
nba3Ppct 1.0000000 0.2717003 0.4112595
c3Ppct   0.2717003 1.0000000 0.4165641
cFTpct   0.4112595 0.4165641 1.0000000
```

Figure 1: Correlation Matrix of NBA $3P\%$ and the two competing predictors

When examining the first column (or row) of the matrix, we can see that college $FT\%$ has a greater correlation to NBA $3P\%$ than college three-point shooting; it's over 50% greater. For further proof, I performed bootstrapping with 2000 iterations, and the sentiment is echoed:

$$\theta = cor(cFT\%, nba3P\%) - cor(c3P\%, nba3P\%)$$

$$Ho : \theta = 0 \ ; \ Ha : \theta > 0$$

Lower Bound	Upper Bound	Verdict
0.04545	0.2311	**Reject** Ho

Table 2: Bootstrapping Results for Correlation

To go a step further, in the bootstrapping interval, 99.85% of iterations resulted in $\theta^B > 0$; a staggering display of support.

Argument II: Simple Models

For this exercise, I created two very crude models with the purpose of comparing goodness-of-fit metrics. Both models are weighted by the career NBA minutes played.

$$\text{Model 1: } nba3P\% \sim c3P\%$$

$$\text{Model 2: } nba3P\% \sim cFT\%$$

Using 5-fold cross validation, I calculated the average of each goodness-of-fit metrics from the list that began this section. The stronger reported value is bolded:

Model	RSS	AIC	BIC	Mallow's C_p	Adj. R^2	Test MSE	Spearman's ρ	Kendall's τ
Model 1	4258	-1628	-1616	-407.9	0.08172	0.001067	0.2998	0.2049
Model 2	**3773**	**-1678**	**-1666**	**-408.0**	**0.1864**	**9.560e-04**	**0.3971**	**0.2733**

Table 3: Crude Model Comparison

A clean sweep for Model 2! However, given there were only five folds, for the sake of both rigor and transparency, I conducted a similar series of bootstrapping tests for the eight metrics above. Some of the results were equally adamant; others, not quite.

$$\theta = S_2 - S_1, \text{ where } S_i \text{ is a statistic from model } i$$

$$Ho : \theta = 0 \ ; \ Ha : \theta \neq 0$$

$$\text{Confidence Interval: Reject } Ho \text{ if } 0 \notin [\hat{\theta}^{(.025*B)}, \hat{\theta}^{(.975*B)}]$$

(Less specific since half of the statistics are stronger if they are lesser.)

Stat	Lower Bound	Upper Bound	Verdict
RSS	-663.4	-299.0	**Reject** Ho
AIC	-68.58	-31.74	**Reject** Ho
BIC	-68.58	-31.74	**Reject** Ho
Mallow's C_p	-0.007369	-0.003597	**Reject** Ho
Adjusted R^2	0.06552	0.1413	**Reject** Ho
Test MSE	-2.483e-04	2.139e-05	Fail to Reject Ho
Spearman's ρ	-0.08641	0.2685	Fail to Reject Ho
Kendall's τ	-0.05803	0.1915	Fail to Reject Ho

Table 4: Goodness-of-Fit Bootstrap Results

Difficult stretch at the end there. This is particularly significant because these three statistics are performed on unfamiliar data points, which calls into question the significance of the upgrade from Model 1 to Model 2. Not to sound like a sore loser, but I will argue why Model 2 is still superior, practically speaking. Unfortunately, yes, the last three tests did not reach the threshold for a $\alpha = .05$ significance level. However, there is a less rigorous argument in favor of them.

$$\text{Let } Z = \left(\sum_{i=1}^{B} \Psi[\hat{\theta}(i) > 0] \right) / B$$

In plain English, Z represents the percentage of stats within the bootstrap interval that satisfy Ha[1]. It's to show that while these three tests failed, it was quite close:

Stat	Z
Test MSE	0.9545
Spearman's ρ	0.86
Kendall's τ	0.865

Table 5: Coping

Z demonstrates that the difference in ordinal ability and test set accuracy can't and shouldn't be ignored. It's not as powerful as the hypothesis test, but the high values of Z combined with the huge AND significant differences in the other goodness-of-fit metrics demonstrate why the crude model with college $3P\%$ performed better.

Final Model

After about twenty different iterations of model selection, I was left with two models that I believed to be the best choices.

WLS Model 12.1: $nba3P\% \sim c3PAr + cFT\% + cFTA + cFGA + draftAge + height + cG$

WLS Model 12.2: $nba3P\% \sim c3PAr + cFT\% + draftAge + height + cG$

These two models have the lowest or second lowest of nearly all of the goodness-of-fit metrics. This result challenged some assertions I had:

[1]Note: it's actually $\hat{\theta}(i) < 0$ for Z when the statistic is Test MSE.

1. **Adding degree to draft age** - In the NBA draft, three- and four-year players fall in favor of raw prospects. Even if a college player had an incredible career or an unforgettable March Madness run, if they lack enticing physical traits, then they will drop. Miles McBride, Jalen Brunson, and Tre Jones are examples of second-round steals. However, for every Brunson, there are three that didn't quite work: Admiral Schofield, Justin Jackson, Cassius Winston, etc. Given the arguably disproportionate emphasis placed on age, I thought that $draftAge^2$ would be a prudent predictor. Models with this higher-degree predictor performed well, but the best models did not include it. Adding a degree to $draftAge$ resulted in slight decreases in all metrics except RSS.

2. **Favoring Rate Over Raw Volume** - My first instinct was to include $cFTr$, because that statistic demonstrates how frequently a player shoots free throws per-FGA. I figured that rates would be a better predictor than raw totals. One, totals will have multicollinearity concerns. $cFTA$, $cFGA$, and cG all correlate with more time spent in college, allowing more time to accumulate stats. However, the models with $cFTr$ instead of $cFGA$ and $cFTA$ performed a little worse. That multicollinearity is still an issue, which I'll get into soon.

I included both models because they are both compelling, but if I had to choose just one, I would go with Model 12.2. It's stronger in many key areas:

		RSS		AIC		BIC		Mallow's C_p		Adj R^2	
Model	df	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Model 12.1	8	**3360**	95.55	**-1714**	8.541	-1678	8.541	-396.0	0.001253	**0.2648**	0.01154
Model 12.2	6	3399	112.0	-1713	9.874	**-1685**	9.874	**-400.0**	0.001244	0.2600	0.01487

Table 6: Goodness-of-Fit Stats

	Test MSE		Spearman's ρ		Kendall's τ	
Model	Mean	SD	Mean	SD	Mean	SD
Model 12.1	**9.255e-04**	6.367e-05	0.4301	0.09450	0.2984	0.06432
Model 12.2	9.293e-04	7.153e-05	**0.4357**	0.08436	**0.2988**	0.05760

Table 7: Test Data Results

There are arguments for both:

- In favor of Model 12.1
  - Lower Test MSE with less spread
  - Higher Adjusted R^2
  - Lower RSS
- In favor of Model 12.2
  - **Lower df**, which means lower Mallow's C_p
  - Lower Ordinal Estimates and less spread, too
  - Lower BIC
  - Nearly equal AIC

– No multicollinearity issues ($cFGA$)

We're really splitting hairs here, but the one bolded point is the most important. Both models have nearly equal metrics everywhere, but Model 12.2 performs as well with two less predictors. Trading for two less degrees of freedom for similar performance convinced me. Also, as I mentioned earlier, Model 12.2 experiences multicollinearity issues. Using the 'vif' function from R's 'car' package reveals the extent of the issue:

c3PAr	cFTpct	draftAge	height	cG	CFTA	cFGA
1.971090	1.365809	3.387600	1.255272	5.373331	3.919661	6.013732

Figure 2: Model 12.1 VIF

$cFGA$ and cG have values greater than five, which is alarming. The model performs well enough, but such a strong correlation between these two predictors further increases the motivation to opt for the model that doesn't perform well with two less variables and doesn't include a problematic one.

c3PAr	cFTpct	draftAge	height	cG
1.466968	1.267379	3.361354	1.208420	3.353412

Figure 3: Model 12.2 VIF

As some final housekeeping for my model, I included three tests of the assumption of linearty and constant variances:

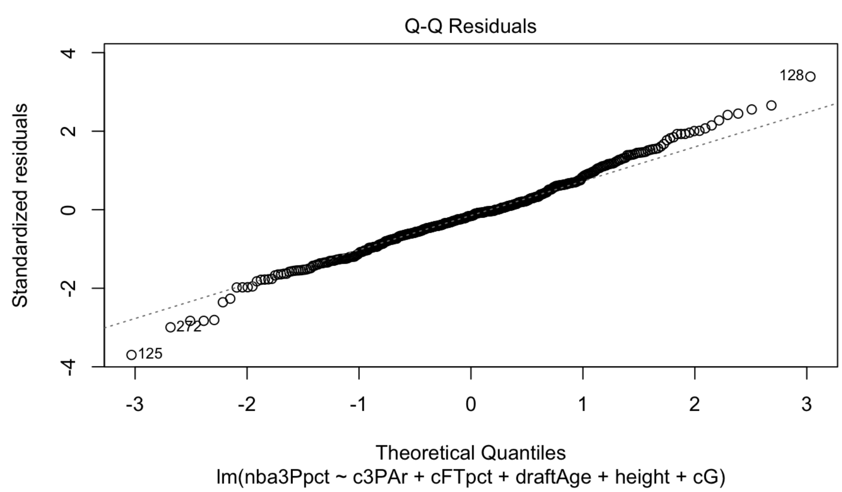

Figure 4: Model 12.2 QQ-Plot

```
F Test for Heteroskedasticity
-----------------------------
Ho: Variance is homogenous
Ha: Variance is not homogenous

Variables: fitted values of nba3Ppct

     Test Summary
-------------------------
Num DF     =    1
Den DF     =    410
F          =    0.2503802
Prob > F   =    0.617076
```

Figure 5: F-Test of Equal Variances from 'olsrr' package

```
        Rainbow test

data:  boot_mod
Rain = 0.99177, df1 = 206, df2 = 200, p-value = 0.5237
```

Figure 6: Rainbow Test from 'lmtest' package

Final Model

$nba3Ppct \sim c3PAr + cFTpct + draftAge + height + cG$

One more thing.

Argument III: Adding $c3P\%$ to the Final Model

For one last exercise, I took the final model and created two sub-versions of it:

- Model $base$: The untouched final model
- Model rft: $c3P\%$ replaces $cFT\%$
  - $nba3P\% \sim c3PAr + c3P\% + draftAge + height + cG$
- Model $w3p$: $c3P\%$ added as a new predictor
  - $nba3P\% \sim c3PAr + cFT\% + draftAge + height + cG + c3P\%$

I once again used bootstrapping to compare the three models' metrics. I have two separate goals. First, I want to show that replacing $cFT\%$ with $c3P\%$ creates a worse model. Second, I want to show that adding $c3P\%$ to the final model does not cause any metrics to improve. This means that both goals will have similar alternative hypotheses and I want to reject Ho for case 1, but not for case 2:

$$\theta_1 : S_{base} - S_{rft}$$

$$\theta_2 : S_{base} - S_{w3p}$$

$$Ho : \theta_i = 0;\ Ha : \theta_i \neq 0;\ i = 1, 2$$

Stat	Parameter	Ha	Lower Bound	Upper Bound	Verdict
RSS	θ_1	$\theta_1 < 0$	-486.8	-251.4	**Reject Ho**
RSS	θ_2	$\theta_2 > 0$	0.04140	54.2941	Reject Ho
AIC	θ_1	$\theta_1 < 0$	-56.49	-29.16	**Reject Ho**
AIC	θ_2	$\theta_2 > 0$	-1.995	4.808	**Fail to Reject Ho**
BIC	θ_1	$\theta_1 < 0$	-56.49	-29.16	**Reject Ho**
BIC	θ_2	$\theta_2 > 0$	-6.016	0.7874	**Fail to Reject Ho**
Adj R^2	θ_1	$\theta_1 > 0$	0.05549	0.1053	**Reject Ho**
Adj R^2	θ_2	$\theta_2 < 0$	-0.01014	0.001836	**Fail to Reject Ho**
Test MSE	θ_1	$\theta_1 < 0$	-1.859e-04	-6.876e-06	**Reject Ho**
Test MSE	θ_2	$\theta_2 > 0$	-3.108e-05	1.091e-05	**Fail to Reject Ho**
Spearman's ρ	θ_1	$\theta_1 > 0$	-0.002177	0.1908	Fail to Reject Ho
Spearman's ρ	θ_2	$\theta_2 < 0$	-0.01330	0.02314	**Fail to Reject Ho**
Kendall's τ	θ_1	$\theta_1 > 0$	-0.003436	0.1367	Fail to Reject Ho
Kendall's τ	θ_2	$\theta_2 < 0$	-0.01104	0.01562	**Fail to Reject Ho**

Table 8: Final Model Comparison Metric Bootstrap Results[2]

Once again, I'll be using Z to argue my case about the Spearman and Kendall results for θ_1. Just look at those intervals. They are hanging onto Ho by a hair. Quite literally:

Stat	Z
Spearman's ρ	0.97
Kendall's τ	0.9695

Table 9: Coping Part 2

Concluding Remarks, Notes & Improvements

In all, I proposed a model to relatively predict a college prospect's NBA perimeter shooting using readily available college stats. More importantly, I demonstrated the stronger influence that college free throw shooting possesses when compared to college three point shooting. Models with $cFT\%$ perform significantly better and models that include $c3P\%$ do not receive a significant upgrade. This aligns with the scouting staple that has been repeated for so long.

If I were to repeat the task of college prospect modeling, there are many notable updates and improvements that I would implement.

- Include other physical traits as variables: Wingspan, Vertical, Strength/Agility Drill Results at the Combine
- I would use more sophisticated modeling packages, such as Python's sklearn package. I believe that using random forests or bagging trees would have produced a strong model.
- Continuing from the last step, a model that uses Bayesian/Shrinkage methodology would work perfectly for this scenario.

[2] I didn't include Mallow's C_p because the intervals were so cramped, and I don't believe the results added to my point. Yes, Model $w3p$ has an additional variable, so it's Mallow's C_p will be "worse."

That is to say, this topic is incredibly dense. Scope creep was a very real issue, and I constantly tangled with feeling overwhelmed by it all: From building the dataset, to deliberating which models to use, to being cornered by thousands of numbers with small differences. Organization, structure, and crafting something coherent became exponentially more challenging with all these cooks in the kitchen. My initial goal was to simply answer "Does free throw percentage actually make for a better predictor?" and a larger endeavor took its place.

In the end, I was able to accomplish the goal and verify the long-standing scouting staple. I had little doubt. With how often it's reiterated, if it wasn't true, surely someone would have publicly challenged it. However, the process behind answering the question was fascinating. It's not that $c3P\%$ had zero predicting power, but it was interesting how unhelpful it was. The final model comparison exercise really highlights how $c3P\%$ being part of an already strong model was akin to earning an A on a group project without actual contribution.

References

[1] Comprehensive R Archive Network, "Heteroscedasticity," R, URL.

[2] Martin, Damien, "Shrinkage and Empirical Bayes to improve inference," *Statology*, 26 Dec. 2018, URL.

[3] Wieland, Ben, "NBA Draft: Prediction of NBA Three-Point Shooting using College Shooting Data," Hoops and Hoos, *SubStack*, 10 Jan. 2022, URL.

[4] Statistics are taken from *Basketball-Reference*.

[5] Model building, evaluation, reporting, and visualization was done using Python (BeautifulSoup, Pandas, NumPy), R (tidyverse, car, stats, olsrr, lmtest), and Excel.