Increasing the shot at a quality draft-decision –
A Bayesian approach to improve predicting three-point-accuracy translation in the NBA Draft

Tobias Berger | Friedrich-Schiller-Universität Jena | tobias.berger@gmx.net
Frank Daumann | Friedrich-Schiller-Universität Jena

1. Introduction

The NBA Draft is a mechanism to regulate competitive balance of a sports league. Its ambition - to give the worst franchises a chance to acquire young prospects, improve long-term and close the on-field talent gap to better teams - is clear and noble. In theory the policy strengthens the NBA’s product in many ways. It is supposed to distribute league-entering player capital equally, and by this dynamic improve the quality of play [14, 32] while increasing the competitive balance among all competitors by raising the uncertainty of outcome on a game and season-level. This has been deemed an important goal for sports leagues for a long time [63]. The potential benefits of the mechanism should lead to increasing attractiveness of the entertainment product the NBA offers.

The crux of these theoretical outcomes is the nature of the mechanism itself. For the described dynamics to work, all NBA managers need to constantly seize the opportunity the NBA Draft provides them with. Unfortunately, this is an incredibly complex expectation to have for the acting decision-makers. In the draft environment, choices are not made within a known variance, but among widely unknown outcomes. The policy can therefore be treated as a proxy for decisions under uncertainty [51]. Within this complex process decision-makers ultimately do not only have to judge who the best prospect is at the moment of the draft, but also who will develop into the most valuable player in the future, while additionally (among considering other external factors) forecasting the direction in which the sport itself is moving in the long-term. This provides much room for individual failures due to personal misjudgments or misconceptions by the managers, and leaves the door open for systematic errors on a leaguewide level caused by collective decision-making biases [64].

Over the past few decades many efforts have been made to improve draft-decision-making-quality and reach the goals the draft mechanism intends to fulfill. Still, the mechanism does not produce such positive outcomes constantly, even though it would be beneficial for the entire league. Research has shown that it does not pay off to solely build up a franchise through the draft [11, 54]. Meanwhile, the NBA historically has performed as the most
imbalanced of all the North American sports leagues [70]. At least to some extend this state of the association can be based on the rather poor draft-decision-making-quality next to salary cap-induced factors [74].

We identified the increase of managerial decision-making-quality within the dynamics as a crucial component to fulfill the aims of the fundamental policy proposal. Initial judgements in talent evaluation and future predictions of performance for the potential draftees and the environment they will play in need to be improved in their accuracy and reliability to reduce uncertainty inside the complex choice mechanism and thus, provoke better overall results.

This paper aspires to reach this goal by examining one of the key components of the sport more closely – shooting the basketball from distance. We will explore league-wide three-point scoring tendencies and investigate in which direction the sport has developed and will progress. Afterwards we strive for improving the prediction of shooting-translation from the college to the professional level by implementing a Bayesian approach to pre-draft metrics and including a play-by-play-based metric into modeling. Being able to project an important basketball skill more accurately should improve the individual managers player evaluations and, in the end, increase the draft-decision-making-quality on a leaguewide level. The results will provide a new baseline for the judgement of draft prospects within one of the most important performance facets of the sport.

2. Background and related literature

2.1. The NBA Draft policy and its shortcomings
The draft as a regulating sports mechanism has been an integral part of all North American major leagues and is investigated in many environments. Researchers mostly analyzed the structure as well as the decision-making within the process or look for inefficiencies from an economic standpoint. Papers on the draft, can be found for hockey [e.g. 24, 73], baseball [e.g. 20, 69], football [e.g. 34, 49] and basketball in the WNBA [e.g. 31, 33].

The draft policy of the NBA was firstly discussed from a strictly legal perspective. Over half a century ago, the mechanism was debated focusing on antitrust laws, labor rights and its general validity [e.g. 2, 18, 21]. It took a while until early structural criticism was raised, because it failed to deliver the intended league-wide outcomes concerning competitive balance through equal talent distribution. As Burkow and Slaughter (1981) pointed out team success was more dependent on the managerial decision-making qualities in other areas such as personnel hiring and player trading [19]. Drafting did not seem to level the playing field as much as it was supposed to, since most of the deciding organizations seem to lack proper talent evaluation abilities [12]. Stronger teams tend to profit more from the policy than bad ones, which is completely contrary to its intentions [11]. Consequently, research even advises against the draft policy as league-wide tool to strengthen competitive balance. Historical results show, there are not enough competent decision-makers to implement the
mechanism well enough to produce those favorable outcomes for the entire association constantly [55].

2.2. The NBA Draft dynamic as a proxy for decisions under uncertainty
The reasons for these policy shortcomings of the NBA Draft are manifold but can be mostly be attributed to the 'human judgement-factor' in many parts of this particular decision-making equation. The decision problem of the entire mechanism is clear. Every franchise strives to optimize the opportunity its current draft position provides them with by selecting the best talent available. Talent in this environment can be seen as a mixture of mostly on-court but also off-court benefits the prospect will generate for his new employer. This value allows the drafting organization to maximize financial and sportive utility following classical economic theory [25, 41]. The process towards solving this problem is immensely complex due to the hurdles the decision-makers must master. Within this straight-forward setup a draft choice requires NBA front offices to give their judgement in two main areas which both present huge levels of uncertainty, as probabilities of decision outcomes in this framework can only be roughly estimated and are mostly unknowable [78].

One the one hand, a manager needs to evaluate the talent level of all the players at the moment of the draft, while factoring their potential development in the future. On the other hand, the direction of the entire sports, as an underlying environment the athlete needs to perform in, must be correctly anticipated as well. The NBA has come a long way over the past two decades by installing more sophisticated processes around these interwoven dynamics. Especially with the recent emergence of sports analytics huge steps have been taken towards a more scientific approach of the sport to better understand the entire discipline and reduce uncertainties [1, 46]. The evaluation of player talent along the triangulation of eyes (e.g. scouting, personal workouts), ears (e.g. personal interviews, medical records exploration, further background research) and numbers (e.g. anthropometric measurements, high school and college performance metrics) has become much more sophisticated [8], while league-wide playing trends can be discovered and followed more closely than ever before ([68].

However, judging individuals is extremely difficult because basketball as a sport has not been ‘solved’ yet. It is hard to isolate the personal impact of one player in a dynamic team discipline [48]. And even if managers knew what exact data to look at, how to weigh all the information correctly to anticipate career arcs of players perfectly, there would probably still be discussions about career values of prospects arguing about peak versus average performance-level, relative value or longevity as potentially driving utility dimensions [e.g. 7, 72].

At some point the analysis of those factors simply comes down to taste, preferences, and the decision-maker's philosophy on how the sports of basketball should be played. These clear aspects of judgement, which are defined as ‘the set of evaluative and inferential processes
that people have at their disposal and can draw on in the process of making decisions’ [43, p.XV] open up the entire process to the realm of classical decision-making-quality-lowering fallacies and biases. Many of those have been investigated in the NBA Draft environment to either reduce uncertainty or uncover hidden systematic errors. Basketball draft research has covered nationality bias [54] as well as recency and availability bias [12, 17, 38]. Additionally, it was shown that age and favorable positional length as proxies for potential [4, 13, 29] get misjudged constantly, while certain boxscore statistics do not translate as smoothly from the amateur to the professional level as one would expect [e.g. 22, 31, 65].

2.3. The NBA three-point revolution
Arguably the most influential rule change in the history of basketball was the implementation of the three-point line. After much experimentation in other basketball environments the three-point line was introduced to the NBA in 1979. This rule change was a clear reaction to the dominance of the bigmen in early decades of the sport. It was intentionally designed to give smaller players better opportunities to score, while making the game more enjoyable to watch for the fans [30].

At first the league was slow to react to this change. Only a few players possessed the shooting skills to use this new offensive weapon constantly, while managers and coaches had not fully grasped the advantages the additional way of scoring does provide. But as more athletes were brought up with the three-point line being present for their entire basketball upbringing shooting-competence and the acceptance of the three-pointer rose in the NBA [27]. In a trickle-down fashion this change affected college basketball as well, which introduced the line in 1986.

![Figure 1. Leaguewide Average of 3PArG over time in NBA and NCAA.](image-url)
With the general skill level of the players catching up, the era of sports analytics led to shooting from distance becoming even more fashionable. Building on the ‘Moneyball’-ideas early introduced by the Oakland As in baseball [45], which center around the use of data to find exploitable inefficiencies within the sport [e.g. 1, 46], the three-pointer as a strategic option was finally recognized as an excellent driver for efficient basketball offenses [57]. Backed up by numbers, managers and players not only started to realize the huge benefit the potential extra-point the shot grants. They also recognized the positive effect the threat of such attempts alone produced, forcing opposing defenses to space out, opening up room for other basketball actions as well. The entire geometry of the sport changed [e.g. 27, 67].
The strategic and stylistic innovation the extensive use of the three-pointer has caused within sport is enormous. The increased application of sports analytics only fuels this development to this day. Currently, some of the arguably most influential players of the sport make more three-pointers over a season than entire teams did over a year just three decades ago [6]. This trend should last, since it has led to a vast increase in offensive efficiency, which is one key factor in winning basketball games [57]. With the 2010s Golden State Warriors and Cleveland Cavaliers disproving the old NBA sentiment, that teams relying on the jump shot will never win championships [76] and no diminishing rule-changes on the horizon, the three-point-movement is here to stay. Thus, the league-wide demand for competent three-point shooters is increasing.

This development becomes especially obvious by looking at functional demands for different positions. The tallest of players of the NBA used to have the simple job of converting scoring opportunities close to the basket. Since the start of the analytics era, most teams want all their players to be able to shoot the three, making the ability to hit open triples one of the most valuable individual basketball traits no matter the height or position [23]. As we can see everybody is supposed to shoot more. Curiously, NBA front offices seem to be slow to react to this trend. They arguably still draft more non-shooting bigmen in higher spots than they should [58], which potentially illustrates rigid underlying management structures and a potential status-quo-bias.

2.4. Three-point shooting as an NBA Draft trait
Due to the current league environment and general playstyle, we have identified three-point shooting as a valuable skillset-element for every NBA player. Experts forecast this association-wide shooting-trend to last since an increasing number of athletes and teams is
adapting [e.g. 16, 27, ] (e.g. Nasru, 2017). Consequently, this trait becomes an increasingly interesting pre-draft ability to evaluate. Especially at the moment, it is often deemed the swing skill for players that can make or break successful careers [35]. Identified and correctly projected, the accurate judgement of a prospects three-point shooting skills can result in better draft decision-making. On the one hand, drafting underrated elite long-range shooters can setup a team for a decade of success [e.g. 52, 53]. On the other hand, simply avoiding players who fail to meet their shooting projection and turn out to be less of a talent as anticipated for that matter indirectly helps a franchise as well [e.g. 44, 47]. Either way, the projection of this ability is key to be able to seize the opportunities the NBA draft provides for the franchises.

However, predicting the translation of pre-draft three-point shooting towards NBA performance in this category is rather complex and difficult because many factors need to be considered. From a player’s internal perspective, managers need to judge the shooting technique [e.g. 37, 42] while considering the prospect’s mental and physical capabilities [e.g. 3, 59]. External variables like adaption to the greater NBA shooting distance1, differing level of play around the player, varying role within the team system, change of player position as well as potentially shifting degrees of difficulty of the types of three-point attempts a prospect is taking, complicate a clean projection. Due to differing circumstances and the uncertainty such external factors produce, even smart forecasts for NBA performance based on pre-draft indicators can never be fully accurate [56].

Additionally, within these pre-draft components several biases can cloud judgement. Recency, availability, or small sample size-biases can negatively influence evaluators, if a prospect has a hot shooting streak in the march madness tournament [12]. Overconfidence might also lead to inaccurate projections on three-point ability, as managers falsely believe that they will teach a non-shooting player how to hit more threes, or on the contrary, irrationally disregard a solid shooter because of their unorthodox throwing motion [64]. To avoid some of these biases, analytics would suggest blocking out as much subjective noise as possible by projecting shooting and its potential development using a model based on tangible, historic data.

3. Methodology

3.1. Data

The website ‘Hoop-Math’ was used as a primary data resource [36]. Their data base allowed us to capture the shooting statistics of every NCAA athlete and their respective schools from 2012 to 2020 (N = 47408). We isolated the drafted players from this sample and were able

---

1 The NBA line (7.24m) is farther back than the one in college and FIBA basketball (both 6.75m) [77].
to match further, missing shooting data from the stats-portal ‘Barttorvik’ as well as scraping and adding the NBA performance data of all draftees from ‘Basketball-Reference’ [5, 6].

Consequently, the constructed data set includes pre-draft information (e.g. name, college, position, age, years of pre-draft-experience), detailed pre-draft shooting data (e.g. shot attempts and makes from several distances) based on play-by-play information and detailed post-draft performance data (e.g. shot attempts and makes from several distances) of all college players who were drafted between 2012 and 2020 (N = 386). 2012 became the natural starting point for our records since college basketball play-by-play data and resulting deeper statistics for full draft classes were only available to us from this point on.

The ongoing nature of the dataset is an important limitation we want to point out. While the players pre-draft statistics cannot change anymore, their post-draft numbers are still developing as most of the players are currently active in the league. The data was scraped in June 2020 and does not include the most recent continuation of the NBA season after an intermission of the 2019/20 season, induced by the COVID-19 pandemic.

We used the ‘R’-software to prepare and analyze the data. For empirical Bayes binomial estimations, we relied on the “ebbr”-package within this program [62].

3.2. Modeling three-point accuracy translation by using different shooting measures

Predicting the translation of three-point accuracy from the college to the NBA level is difficult. There are many context variables (e.g. pre-draft team tactical system, teammates, competition) that would need to be included into a more holistic approach. We chose to disregard these factors completely to simplify our approach as much as possible and reduce the amount of potential statistical noise around the two introduced metrics. Consequently, we approached shooting-ability simply as a measure of how many of the three-pointers a player attempted (3PA) he converted (3PM).

Introducing shooting as this basic measure of volume and efficiency already poses some difficulties. Research has shown that three-point percentage (3P%) as a metric stabilizes at around 750 attempts. Hence, it needs a certain kind of volume until this statistic becomes a reliable measure for a player’s shooting competence [15]. This creates a draft projection dilemma, because almost no potential draftee plays enough official games to be able to offer such a dependable pre-draft sample-size. Especially since drafted players have been getting a lot younger over the past few decades [79].

Current publicly available models try to account for this complication by factoring in other potentially useful dimensions that contribute to measuring shooting ability. Regular three-point metrics in these models are usually considered with raw percentages and a shooting volume estimate based on three-point makes or attempts on a ‘per-40-minutes-basis’ to be able to compare talents with different playing times. Pre-draft free-throw-percentage (FT%) is a popular addition in such designs as well. It has been shown that including free-throw percentage in models significantly improves projections of NBA three-point accuracy, even
though the shots seem to differ a lot in terms of distance or game circumstance [e.g. 28, 39, 40, 71]. Despite the differences, it seems logical, that free-throw percentage contributes to the projection of three-point skills. In its core it is only capturing another throwing-activity on the court and is therefore just an additional, interesting measure of ‘shooting touch’, while increasing the data basis the system can gather meaningful information from.

In our later modeling we will add two-point-jumper accuracy (2P)\% as another shooting metric to our prediction attempt with the same logic. We assume two-point jumpers require a draftee to apply many of the same biomechanical capabilities that are needed to hit a three-pointer or a free-throw. Even though the game context differs a lot in the given situations, all scoring attempts are usually executed with a similar shooting technique or style and require a level of ‘shooting touch’ to successfully put the ball into the basket [80]. Hence, two-point-jump shooting could also be a reliable measure of overall shooting-ability and should therefore contribute to the prediction of three-point competence in the NBA. This hypothesis will be tested.

### 3.3. Adding a Bayesian approach to identify true pre-draft shooting competence more accurately

To ‘apply a Bayesian approach’ usually means to use statistical techniques relying on ‘Bayes theorem’, which assigns a rather subjective treatment to probabilities and handles unknown information probabilistically [10]. Taking a Bayesian perspective when dealing with statistical problems has become popular in many fields because it provides several advantages. For the discipline of sports science and analytics particularly - facing a data influx rapidly growing in volume and complexity - it proves to be beneficial especially because it allows to:“

- incorporate expert information or prior believes, […]
- use Bayesian learning where the current posterior distribution becomes the prior for future data, […]
- model complex problems, […]
- deal more effectively with small dataset using prior information to improve the parameter estimates, […]
- make predictions taking into consideration uncertainty” [66, p. 2]

among other listed benefits. Various applications of Bayesian methods for basketball shooting problems and modeling can be found [9, 26, 60, 75]. In our opinion, such an approach should be favorable for pre-draft three-point shooting as well because it allows to add more context and potentially assign more meaning to this measure by using an empirical Bayesian shrinkage towards a Beta prior based on historical data. This technique has been applied successfully to produce a more well-informed assessment of baseball’s batting average, an estimate of a player’s hitting ability [61]. Statistically speaking, the logic of batting average is very similar to three-point-shooting in many ways. This approach has therefore already been applied to pre-draft three-point percentage and provided interesting
results in the basketball realm as well [50]. With our analysis, we will build on this first modeling attempt and hope for even more accurate results by adding more informed data, factoring in two-point-jumper accuracy.

4. Analysis

4.1. Empirical Bayes estimation of pre-draft three-point percentage

Three-point-percentage is a straight-forward statistic. It is calculated by dividing the number of three-point-makes by the number of attempts:

$$3P\% = \frac{3PM}{3PA}$$

As the term shows, three-point-percentage measures shooting accuracy and will always be a value somewhere between 0 and 1. Carelessly just going by 3P% a manager looking for a good shooting prospect would have to draft a talent that went 1/2 over a player who converted 40/100 of his triples as .5 is a lot greater than .4. But when evaluating and projecting shooting ability and not only accuracy this approach seems silly. We intuitively know that the number of attempts a player needed for his results should be factored in since even though they present the same accuracy on the surface neither 1/2 and 50/100 nor 0/1 and 0/100 should be treated equally.

Hence, to produce a more useful or even ‘true’ estimate of shooting abilities, more context needs to be added to the simple percentage. Often basic filtering is sufficient. As described earlier, after crossing a certain threshold in sample-size plain percentages become reliable statistics. For three-point shooting this point has proven to be around 750 attempts [15], meaning we can use simple 3P% as a tool to compare shooting capabilities of talents somewhat confidently if all of the evaluated players have hit this mark.

However, this excludes a lot of prospects. As we want to be able to project all potential NBA players in their shooting translation, we chose empirical Bayes estimation as our method to work with pre-draft three-point-percentage - a good example for a binominal distribution. With this technique, context is added by fitting a beta distribution to all available data, gaining a prior that adds information based on the investigated environment. Later, this prior is individually updated with the evidence each observation gave and converted to a posterior distribution [61].

To produce our prior, we first need to fit a regular beta distribution to the histogram of all the three-point-percentages of our data base, using this model:

$$X \sim Beta(\alpha_0, \beta_0)$$
This estimation gives us $\alpha \approx 114$ and $\beta \approx 201$ for all players who attempted at least one three-pointer, using the most data points from our data set possible.

![Figure 5. Density plot of Pre-Draft-3P% with beta distribution fit curve.](image)

The density plot of the beta distribution fit with maximum likelihood over a histogram of three-point percentage seems to match our data satisfyingly well. Hence, we can apply Bayesian inference, adapt the found values as our priors and update each case based on this foundation using the following formula:

$$Beta (\alpha_0 + makes, \beta_0 + misses)$$

This means that the expected mean value of the beta distribution is

$$\frac{\alpha_0}{\alpha_0 + \beta_0} \approx 0.362$$

The found prior is the head start basically every player gets assigned with before evaluating their actual performance. Depending on the three-point makes and misses a talent showed the found prior is updated. The greater the sample size a prospect produced, the greater the chance to shift the prior in one direction or the other. Simply applying the examples from earlier, we can see that this approach would estimate that an athlete hitting 40/100 is more likely to be a better shooter than a competitor making 1/2 despite having a lower three-point percentage:
\[
\frac{1 + \alpha_0}{2 + \alpha_0 + \beta_0} = \frac{1 + 114}{2 + 114 + 201} \approx 0.363 \\
\frac{40 + \alpha_0}{100 + \alpha_0 + \beta_0} = \frac{40 + 114}{100 + 114 + 201} \approx 0.371
\]

Applying this method, we updated each observation using our found prior. Or as this figure shows, we basically shrunk each value towards the expected mean value. The smaller the sample-size a player put up the more reason the model has to move the three-point-percentage towards the general expectation of average based on the investigated environment.

![Figure 6. Raw Pre-Draft-3P\% versus Shrunken-3P\%](image)

In our first step, we updated all three-point-percentages of our database this way and got a little closer to a more insightful metric to use for our shooting translation projection. Extreme cases of either extremely high or low accuracy got corrected if they were based on little evidence.

### 4.2. Beta binomial regression for pre-draft three-point percentage

Shrinking three-point percentages towards an expected mean seems to have improved the metric already. All players getting lucky by hitting most attempts of a very low number of tries from behind the three-point line are not listed as the players potentially possessing the most shooting accuracy anymore. But our method did not account for one important dynamic within the sport of basketball: Weak shooters generally tend to shoot less than better long-range threats due to e.g. a lack of self-confidence or coaching staffs restricting prospects from taking triples. On the flipside strong shooters usually get the green light in their team’s offense, even if they hit a cold-streak that could already influence a small college-
sample immensely, because they showed their talents in practices. As we can see, we did not account for this complication in our first estimation. Players shooting very little tend to get overrated by only using a one-fit-all-prior.

![Figure 7. Comparison Raw 3P% versus Shrunken 3P% by CBB 3PA.](image)

Factoring in this complication, forces us to adjust our found estimates again, since we can see that our Empirical Bayes estimate towards our prior (indicated by the red dotted line) systematically overvalues shooters that took very little shots. The blue line, showing the median average for every attempt range, bears out that sentiment.

Removing this bias within our estimates, we need to adjust our model. We want our priors to be influenced by the number of threes a prospect has taken/was allowed to take. Robinson (2017) suggests using beta-binomial regression and update the model in a generative process to the following, with $\pi$ being defined as the true shooting percentage of player $i$ and letting $3PA_i$ be known and fixed per player [61]:

$$p_i \sim Beta (\alpha_0, \beta_0)$$

$$3PM_i \sim Binom (3PA_i, p_i)$$

The "ebbr"-package in R lets us calculate individual priors for our shooting estimation, based on the number of attempts. Their distributions look like this:
Our newfound priors still state a fairly great amount of uncertainty. Attempting 750 threes in college still leaves open a broad corridor of a ‘true’ shooting-percentage of .31 to .46 for a prospect as the range of probable results. But with the range depending on 3PA, it can be assumed that a talent that only attempted 10 threes is almost certainly a worse shooter than a player taking 750 triples, assumed they had equal playing opportunities.
With the new priors most of the percentage estimates for players with little attempts drop quite a bit and correct for our former systematic overvaluation of shooting.

![Figure 10. Overview of Raw 3P%, Shrunken 3P% and 3P% estimate with regression by CBB 3PA.](image)

Investigating the effects of our last step, we reached our intended results. Our new basic shooting-estimates look more context driven and therefore more accurate than our raw percentages. We eliminated unreasonably high or low percentages produced by (bad) shooting luck in small sample-size circumstances, using a data educated prior. Afterwards we corrected our shrinkage values by introducing individual priors for every player based on their number of three-point attempts as basketball logic suggests that this factor should play a role in the estimation of shooting ability.

The newfound statistic is still problematic as various external factors (e.g. tactical system, injury, suspensions, competition) can influence the number of three-point attempts a player is taking despite his assumed shooting ability as our simple approach puts to the forefront here. However, our values now better line up with the median average a certain attempt threshold dictates, which is promising. Additionally, the found estimates are only one piece of the three-point-translation-model presented later.

**4.3. Modeling three-point-percentage translation adding two-point-jumper accuracy**

Finally, we can put together all the prepared components and combine them into a modeling approach. Based on our theoretical work we assume the pre-draft factors free-throw percentage, two-point-jumper percentage and our estimation of three-point-percentage (after an empirical Bayes estimation while according for three-point-attempts in our priors) to be a fruitful foundation for the projected translation of NBA three-point-shooting based
on pre-draft-data. Hence, the formula of our multiple regression model (M1) with post-shooting-accuracy as our dependent variable shall be:

\[ M1: \text{PostDraft}3P\% = \beta_0 + \beta_1 \text{PreDraft}FT\% + \beta_2 \text{PreDraft}2PJ\% + \beta_3 \text{PreDraft}3P\% \text{Estimate} \]

We will compare the results of this approach with two other popular methods we found [39, 40], using pre-draft three-point-accuracy-data without a shrinkage concept. While the second model (M2) adds simple pre-draft three-point-percentage and three-point attempt-rate per 40 minutes, the third model (M3) factors in a combination of the two variables by calling on three-pointers made per 40 minutes:

\[ M2: \text{PostDraft}3P\% = \beta_0 + \beta_1 \text{PreDraft}FT\% + \beta_2 \text{PreDraft}3P\% + \beta_3 \text{PreDraft}3P\% \text{Ap40} \]

\[ M3: \text{PostDraft}3P\% = \beta_0 + \beta_1 \text{PreDraft}FT\% + \beta_2 \text{PreDraft}3P\% \text{Mp40} \]

We tested all modeling ideas on our data set for all athletes who attempted at least one three in college (to be able to calculate a pre-draft three-point-percentage estimate) and played at least 50 games in the NBA already (N = 255). We chose this threshold to be able to include as many players as possible, while working with a reasonable sample-size of games. Even rookies from the 2019/20 season had the chance to qualify if they played enough until the COVID-19 intermission.

M1 proved to be a highly significant general model with an overall explanatory power of an adjusted R² of 0.231. We can reject the H0 of these factors not having an influence on the dependent variable. All individual components were highly significant as well and having an effect in the expected direction. Good pre-draft performance translates into a projection of higher post-draft skills:

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>-.134</td>
<td>.052</td>
<td>-2.576</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-FT%</td>
<td>.153</td>
<td>.062</td>
<td>.169</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-3P%-Estimate</td>
<td>.756</td>
<td>.146</td>
<td>.343</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Post-Draft-3P%

M2 also presents a highly significant model with an adjusted R² of 0.213. Contrary to M1, it does not only consist of significant parts. Curiously, Pre-Draft-3P\% does not improve the prediction of the design. Though, removing the variable only improves the explained variance by 0.002.
M3 is similar to M2 in its results. Presenting a highly significant modeling concept again, M3 has an adjusted $R^2$ of 0.216. Its individual variables are all highly significant and have an effect in the expected direction:

Comparing the explained variance of all the designs, M1 fares the best. Its prediction of post-draft three-point-percentage based on pre-draft-statistics is superior to the explored publicly available methods. Examining the standardized beta coefficients, the Pre-Draft-3P%-Estimate has the most influence on the projection of the dependent variable. But taking two-point-jump shoot-accuracy into account also contributes to the overall model performance. Without posing a huge increase in explained variance, our design still delivers a slight improvement over the known technique.

To further evaluate the models, we calculated the mean absolute deviation for all designs. M1 performed best again with a general error size of 0.0447 compared to M2 (0.0466) and M3 (0.0463). Once more, the differences are marginal, but still relevant.
Finally, we used absolute error terms to evaluate M1 individually. With an increasing number of NBA games and three-point attempts the size of the errors made decreases rapidly and seems to stabilize. Curiously, we can also show that predicting the shooting
accuracy of bigmen offers the most problems to our design and can clearly be picked out as an area of improvement, even though position variables did not prove to be significant in our initial modeling.

4.4. Predicting post-draft three-point-attempt rate
Shooting ability in the NBA is surely about how accurate players are at hitting the basket when taking shots from distance. Therefore, we investigated Post-Draft-3P% as a metric, managers should be eager to project accurately in a draftee. However, prospects also need to be willing to take these kinds of shots at a reasonable rate to make enough use of their own competence and keep defenses honest. Only a combination of accuracy and volume represents shooting-ability correctly. That is why we incorporated three-point-attempts into our estimations of ‘true’ pre-draft-three-point-percentage.

Consequently, in our quest of predicting shooting-translation from the NCAA to the NBA, we also need to look at the three-point-attempt-rate at the professional level and suggest a model that allows to project this metric as accurate as possible to aid draft decision-making even further.

Thus, we chose a rather exploratory approach since we could not derive a regression design solely from theory. We selected NBA three-point-attempt rate per 40 minutes as our dependent variable. For the independent components of the model, we took all six available factors from the three already explored ideas for three-point-percentage as all of these can be argued as measures for shooting ability (all pre-draft measures: 3P%, 3PAp40, 3PMp40, FT%, 2PJ%, 3P%-estimate). Afterwards, we applied a stepwise multiple regression method to check if a significant approach was there and which combination of the individual parts would offer the most explanatory power. Again, we used the entire data set, filtering for one college three attempt and 50 played NBA games.

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Estimate</th>
<th>R Square Change</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.750a</td>
<td>.562</td>
<td>.560</td>
<td>1.531</td>
<td>.562</td>
<td>324.692</td>
<td>1</td>
<td>253</td>
<td>.000</td>
</tr>
<tr>
<td>2</td>
<td>.759b</td>
<td>.575</td>
<td>.572</td>
<td>1.511</td>
<td>.013</td>
<td>7.962</td>
<td>1</td>
<td>252</td>
<td>.005</td>
</tr>
<tr>
<td>3</td>
<td>.767c</td>
<td>.588</td>
<td>.583</td>
<td>1.491</td>
<td>.012</td>
<td>7.547</td>
<td>1</td>
<td>251</td>
<td>.006</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Pre-Draft-3Pmp40
b. Predictors: (Constant), Pre-Draft-3Pmp40, Pre-Draft-2P%
c. Predictors: (Constant), Pre-Draft-3Pmp40, Pre-Draft-2P%, Pre-Draft-3P%-Estimate
d. Dependent Variable: Post-Draft-3PAp40
Table 5  
Coefficient Overview Stepwise Model

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>1.696</td>
<td>.181</td>
<td>9.396</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-3PMp40</td>
<td>.685</td>
<td>.038</td>
<td>.750</td>
<td>18.019</td>
</tr>
<tr>
<td>2</td>
<td>(Constant)</td>
<td>-.203</td>
<td>.696</td>
<td>-.292</td>
<td>.770</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-3PMp40</td>
<td>.687</td>
<td>.038</td>
<td>.752</td>
<td>18.320</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-2PJ%</td>
<td>4.977</td>
<td>1.764</td>
<td>.116</td>
<td>2.822</td>
</tr>
<tr>
<td>3</td>
<td>(Constant)</td>
<td>-3.830</td>
<td>1.488</td>
<td>-2.573</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-3PMp40</td>
<td>.563</td>
<td>.058</td>
<td>.616</td>
<td>9.642</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-2PJ%</td>
<td>4.990</td>
<td>1.741</td>
<td>.116</td>
<td>2.865</td>
</tr>
<tr>
<td></td>
<td>Pre-Draft-3P% Estimate</td>
<td>11.955</td>
<td>4.352</td>
<td>.176</td>
<td>2.747</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Post-Draft-3PAp40

The best model appears to include pre-draft three-pointer-make rate per 40 minutes, pre-draft-two-point-jumper-percentage, and our newly introduced estimate of pre-draft-three-point-percentage. The model itself is highly significant as well as all its individual components. The adjusted R² of 0.588 is fairly strong, while all component coefficients have an effect in the anticipated direction.

This gives us reason to believe, that the found model (M4)

\[ M4: PostDraft3PAp40 = \beta_0 + \beta_1 PreDraft3PMp40 + \beta_2 PreDraft2PJ% + \beta_3 PreDraft3P%Estimate \]

should be useful in the projection of shooting competence translation, while our newly introduced components contribute to this prediction.

Again, we evaluated the model by investigating its error terms. The mean absolute deviation of the design is 1.2, which seems reasonable, when projecting 3PAp40.
Again, it can be observed that the model’s projections become more accurate with increasing NBA games played as well as threes taken, and then stabilize. The distinction between...
bigmen and non-bigs appears to be important again, even though bigmen-attempt-rates seem to be more predictable by the variables used than three-point accuracy.

5. Conclusion and Outlook

In this paper we investigated whether an empirical Bayesian derived ‘true’ pre-draft three-point percentage and pre-draft two-point-jumper-percentage could contribute positively to the quest of projecting the shooting translation of NBA draft prospects. Our analysis showed that in addition to pre-draft free-throw-shooting both statistics provide value as a basis for predictive modeling. The newfound three-point-percentage estimate based on data-introduced Bayesian priors, factoring in the number of attempts as an indicator for basic shooting-quality lends more context to the former simple percentages. The two-point jumper-percentage also behaves as expected. Our notion of this metric being another indicator of general ‘shooting touch’, just like free-throw percentage, turned out to be right and should be considered for shooting translation ideas from now on.

Interestingly, this statistic also contributes positively to the projection of three-point-attempt-rate, which should draw some attention. The showed connection indicates that the popular basketball narrative, implying that hitting midrange-shots effectively could be an indicator for a player having shooting-range which might be expandable to behind the three-point line, could be true.

The general model we found, including the components in question, presents an improvement to other publicly available designs of shooting-translation. We were able to show that compared to systems based on simpler metrics, our model was able to project three-point accuracy with less errors than former approaches. Additionally, we were pleased to see that our introduced variables offer an even more promising basis for the projection of post-draft three-point shooting-volume, adding even more information to the evaluation of a prospects shooting-competence. Such contributions could ultimately improve draft decision-making league-wide and therefore move the results of the policy towards the intended outcomes.

However, the simple modeling presented here, cannot be used reliably on its own. As we can see with the shares of variance that both models for shooting translation fail to explain, many important dynamics are not captured by the simple metrics including the Bayesian informed three-point percentage and two-point jump-shooting. The design has difficulties picking up on e.g. conservative college schemes suppressing shooting aspirations of talented bigmen or possible non-linear future development of players due to improved coaching or shooting mechanic change. Hence, it should rather be used as a valuable decision-informing tool during the draft process, triggering conversation and possibly closer investigations of prospects, rather than being the main argument for a choice.
To produce a model powerful enough to be used in this way, more research, time, and data is needed. Our approach regarding two-point-jumpers was based on pre-draft play-by-play data that has been available for NCAA prospects for a decade only. Such a sample-size is not satisfying, but will automatically expand in the coming years, maybe even with the addition of information on international prospects, to allow more accurate results.

Besides this natural progression, more sophisticated future approaches should consider capturing more of the external factors that potentially go into the projection of shooting progression. Having data on system-based shooting suppression/enhancement of a former team or a more detailed overview of the types of threes a player took, would further inform estimations of a ‘true’ pre-draft three-point percentage. Analysis of biomechanical or psychological attributes could educate estimates of development curves as well, by e.g. showing an obvious, but easy to fix flaw in the shooting motion, a lack of throwing consistency due to weak conditioning or a lack of confidence in the own abilities and then indicating a high likelihood of improvement in these areas by showing a good work-ethic and high coachability with psychological profiling. Picking up on such signals, which should inform shooting translation even further, would allow to improve draft decision-making-quality even more.

References

https://doi.org/10.7312/alam16292
https://doi.org/10.3389/fphys.2018.00075
https://barttorvik.com/
https://doi.org/10.1111/test.12220

Berri, D. J., & Schmidt, M. B. (2010). Stumbling on wins: Two economists expose the pitfalls on the road to victory in professional sports. FT Press.


Volz, K. G., & Gigerenzer, G. (2012). Cognitive processes in decisions under risk are not the same as in decisions under uncertainty. Frontiers in Neuroscience, 6, Article 105, 1–6.