The dilemma of choosing talent: Michael Jordans are hard to find

Peter A. Groothuis\(^a\), James Richard Hill\(^b\,*\) and Timothy Perri\(^c\)

\(^a\)3093 Raley Hall, Appalachian State University, Boone, NC 28608, USA
\(^b\)317 Sloan Hall, Central Michigan University, Mt. Pleasant, MI 48859, USA
\(^c\)3092 Raley Hall, Appalachian State University, Boone, NC 28608, USA

This article explores the dilemma of choosing talent using NBA data from 1987 to 2003. We find there is much uncertainty in selecting talent. If superstars are found, they are usually identified early. However, more false positives exist than correct decisions with high draft picks. Our results suggest the dilemma of choosing talent is not so much a winner’s curse but more like a purchase of a lottery ticket. Most times you lose, but, if you are going to win, you must buy a ticket.

I. Introduction

Economics has a long history of situations where agents have \textit{ex post} regrets from decisions made under uncertainty. In the now classic case of the winner’s curse, agents who have differing beliefs about an amenity value will find, in an auction, the winner of the auction will be the bidder who overvalued that amenity. Capen \textit{et al.} (1971) and Pepall and Richards (2001) suggest a winner’s curse emerges in competitive bidding environments; Cassing and Douglas (1980) provide an example of the winner’s curse in baseball free agency. More recently, Lazear (2004) identifies the Peter Principle as a situation where individuals who are promoted may have been lucky in a stochastic sense and been promoted above their performance level.

Nowhere is the problem more pronounced than in the pursuit of superstar talent. Rosen (1981) outlined the theoretical constructs of the market for superstars and recognized the pervasiveness of the search. Sports teams are in pursuit of the next Michael Jordan, movie studios pursue the next Titanic and music producers seek the next Beatles. Yet player after player, movie after movie and singer after singer fail to meet expectations. In the pursuit of superstars, there are many false positives. We identify this problem as the dilemma of choosing talent.

In Section I, we model the dilemma of choosing talent when the distribution of talent is known to be from the upper portion of a talent distribution. In Section II, we test the theory using a panel study of players in the NBA from 1987 to 2003. We conclude with a discussion of the dilemma of choosing talent and how it relates to the economics of superstars.


*Corresponding author. E-mail: hill1jr@cmich.edu
data to study the winner’s curse in hiring international basketball players. Other studies have analysed the draft mechanism in choosing talent. Hendricks et al. (2003) analyse uncertainty, option value and statistical discrimination in the NFL draft. Groothuis et al. (2007) analyse early entry in the NBA draft, while Lavoie (2003) focuses on discrimination in the NHL draft. Our study, following the same structure, provides insights into the NBA as well as to the labour market in general.

II. The Model

To formally model the problem of choosing talent, consider what happens to the probability of finding high quality talent when the lower bound for high quality increases. Assume:

- \( x = \) talent, \( x_L \leq x \leq x_H \);
- \( x \sim \) continuously with a p.d.f of \( f(x) \) and a c.d.f of \( F(x) \);
- \( x^* \) is the minimum level for high quality talent;
- A potential employer observes a binary signal which is either favourable or unfavourable and
- \( P = \text{prob}(x > x^*)|\text{favourable}) \).

Thus, from Bayes theorem we have:

\[
P = \frac{\text{prob}(\text{favourable}|x > x^*)\text{prob}(x > x^*)}{\text{prob}(\text{favourable}|x > x^*)\text{prob}(x > x^*) + \text{prob}(\text{favourable}|x < x^*)\text{prob}(x < x^*)}
\]

Note \( \text{prob}(x > x^*) = 1 - F(x^*) \) and \( \text{prob}(x < x^*) = F(x^*) \).

Now suppose the probability of a favourable signals increases linearly in \( x \): \( \text{prob}(\text{favourable}) = x/x_H \). This means those with \( x = x_H \) have a probability of one of receiving a favourable signal; others have a smaller probability of a favourable signal.

Now \( \text{prob}(\text{favourable} > x^*) = \int_{x_H}^{x_H} \frac{x}{x_H} f(x) dx / [1 - F(x^*)] \), and \( \text{prob}(\text{favourable} < x^*) = \int_{x_L}^{x^*} \frac{x}{x_H} f(x) dx / F(x^*) \). We can then simplify Equation 1:

\[
P = \frac{\int_{x_H}^{x_H} x f(x) dx}{\int_{x_H}^{x_H} x f(x) dx}
\]

The denominator of (1') is the population mean of \( x \). Clearly \( \partial P/\partial x \) is negative: the higher the level of talent desired (\( d(x^*) > 0 \)), the smaller the probability someone with a favourable signal exceeds the cut-off for high talent (\( x^* \)). Also \( \partial P/\partial \overline{x} \) is negative: the more talented the population, on average, the smaller the probability someone with a favourable signal exceeds the cut-off for high talent.

Note: these results do not depend on a ‘thin tail’ at the upper end of the ability distribution; all we have specified is that the distribution is continuous.

For further insight, suppose \( x \sim \) uniformly on \([\overline{x} - \Delta, \overline{x} + \Delta] \). We have:

\[
P = \frac{(x_H)^2 - (x^*)^2}{4 \Delta \overline{x}} \quad (1')
\]

Suppose \( \overline{x} = 6 \) and \( \Delta = 5 \). A firm that desired an above-average worker (\( x^* = 6 \)) would, choosing at random, obtain such a worker with a 50% probability. Using (1’), the signal would correctly identify such an individual 71% of the time. If the firm desired someone with \( x > 10 \), choosing at random, it would obtain such an individual 10% of the time. Using the signal, it would obtain such an individual 17.5% of the time.

III. Empirical Results

To empirically test the model of the dilemma of choosing talent, we focus on NBA data for performance from the 1987–1988 season to the 2003–2004 season. We use a measure of player performance called the efficiency formula to develop a distribution of talent. As reported by NBA.com, this index is calculated per game as: (points + rebounds + assists + steals + blocks) – ([field goals attempted – field goals made] + [free throws attempted – free throws made] + [turnovers]). This formula provides a measure of quality that is based upon performance in all aspects of the game. In Table 1, we report the mean, median, SD and highest level of the efficiency rating. We find in all cases the mean is higher than the median, suggesting a right-skewed distribution of talent. We also find that the highest value is always over 3 SDs from the mean. In Fig. 1, we plot a distribution of efficiency ratios for the 2001–2002 season. The distribution is skewed right with only a few players in the top tail of the distribution.

In Table 2, we focus on players whose efficiency rating is 2 SDs from the mean. We find from 12 to 22 players a season have efficiency ratings over 2 SDs from the mean. During this time period, we find only two players, who were in this elite category, undrafted, Ben Wallace in 2001–2002 season and Brad Miller in the 2003–2004 season. Many were on the list a multiple of times, some as many as 9 years.
During this time, we find many of the number one picks and lottery picks are in the elite category. Some number one picks, however, never show up on the list. Still others only make the list one time in their career.

In Table 3, we look at only the top five players in efficiency ratings. We find, in our 17-year panel, only 19 players fill the 85 spots in this time period. Most were on the list a multiple of times. The lowest rank in the draft on this list was the 13th pick – two players, Karl Malone in 1985 and Kobe Bryant in 1996. Many of the top players were number one draft picks. Many number one picks, however, did not make the top five players in the NBA. In fact, many of the top picks did not make it to two deviations above the mean. There are many false positives.

In Table 4 the mean, SD, minimum value, maximum value and number of observations for efficiency are reported by draft number. The figures in this table reveal some interesting results. First, the drop-off in efficiency between the first pick in the draft and the second pick is statistically significant. The decrease in mean efficiency is also statistically significant between the fifth and sixth picks. There is a general negative relationship between mean efficiency and draft number; exceptions to this trend occur when lower picked players overachieve (e.g. both Karl Malone and Kobe Bryant were thirteenth picks in the draft). Overall, the draft appears to represent either an efficient judge of talent or a self-fulfilling prophesy (teams may give number one picks more minutes and more opportunities to be a superstar). Hoang and Staw (1995) support the latter view; they find teams grant more playing time to their most highly drafted players even after controlling for performance, position and injury.

In Table 5, we summarize the dilemma of choosing talent by calculating the percentage of players who obtain superstar status by draft number. Column one calculates the percentage of players who have at least one season of performance 2 SDs above the mean. We find that 80% of number one draft picks have at least one superstar season where their performance is 2 SDs above the mean. This percentage falls off quickly with only 40% of number two draft picks and 30% of number three draft picks having a superstar season. Column two reports the percentage of players by draft pick who make the top five players in the league. Here we find the dilemma of choosing talent is great; only 35% of number one draft picks perform at this level, and this falls off even more quickly. Finding superstars is a rare event indeed.

To further test the dilemma of choosing talent, we use a random effects panel model to estimate player’s efficiency ratings. A simple equation to represent the model is:

$$Eff_{it} = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i-1} + \epsilon_{it}$$

where $i$ refers to the individual player, $Eff_{it}$ represents the efficiency of the player in year $t$, $X_1$ is a vector of time-invariant player characteristics, $X_{2i-1}$ is a vector of experience measures and $\epsilon_{it}$ is vector of disturbances. The only time-variant player characteristics included in the model are experience and experience squared; no performance statistics are used since efficiency is computed from these statistics. Time-invariant personal characteristics used to explain efficiency are player height (measured in inches), years of college and a dummy variable equal to one for white players.

---

1 The value of the test statistic is 6.5239. This is greater than the critical value at the .005 level of significance given the degrees of freedom.
<table>
<thead>
<tr>
<th>Season</th>
<th>Player names, draft year and draft number</th>
</tr>
</thead>
</table>
Two options for estimating this model are the fixed effects approach and the random effects approach. In the fixed effects formulation of the model, differences across individuals are captured in differences in the constant term; thus any time-invariant personal characteristics are dropped from the regression. In this formulation of the model, it is impossible to determine if differences exist between players in terms of efficiency due to draft number or other time-invariant variables. Therefore the fixed effects model will not be used.

In the random effects formulation, the differences between individuals are modelled as parametric shifts of the regression function. This technique of estimating panel data allows for estimates of all of the time-invariant personal characteristics as well as the experience statistics. Breusch and Pagan (1980) developed a Lagrange multiplier test (LM-test) for the appropriateness of the random effects model compared to the OLS format.2 The LM-test statistic is 9109.99, which greatly exceeds the 95% Chi-squared with one degree of freedom, 3.84. Thus the simple OLS regression model with a single constant term is inappropriate.

In Table 6, we report the results of the random effects model run using data from the 1987–2002 seasons.3 In Regression I, draft number, experience, experience squared, years of college and race are all statistically significant determinants of efficiency; height is not. As expected, efficiency declines as draft number rises. Efficiency initially rises with experience then declines. Efficiency declines as years of college rise; this reflects the early entry of outstanding college or high school players. Regression II is run minus the draft number variable. The coefficient of height is now positive and significant. Obviously there is collinearity between draft number and height.

---


3 The last season used in this regression analysis, 2002–03 season was selected to avoid any selectivity bias that might have occurred from too many young high school players jumping into the league prior to the imposition of the 19-year-old rule and individual salary cap negotiated into the latest NBA agreement.
The negative coefficient for white players is interesting. *A priori* we would expect this coefficient to equal zero. The results suggest that white players may be drafted higher than their future performance would indicate. Lavoie (2003) has studied the NHL draft and concluded there is entry discrimination against French Canadian hockey players. Exit discrimination in the NBA has been the focus of recent articles by Hoang and Rascher (1999) and Groothuis and Hill (2004). Perhaps future research on entry discrimination is warranted.

The *R*-squared of the models is around 16–17% overall. It is somewhat higher in explaining variation in efficiency between players, approximately 21%, and between years for the same players, 26%. In general, the results suggest a great deal of unexplained variation in player efficiency from season to season. The weakness of the explanatory power of the model may be somewhat surprising given the plethora of data available to NBA executives prior to making draft decisions. In addition to college and/or high school performance statistics available for all players in the draft, the NBA holds camps in which the top players play against one another. Obviously there are characteristics and attributes that are not easily seen or measured that affect player performance.

**IV. Conclusions**

The dilemma of choosing talent suggests, when employers seek to find the very best of a pool of applicants, more false positive signals exist than correct decisions. Using NBA data, we find there is much uncertainty in selecting talent. However, stars and superstars are generally correctly identified in the draft. Our results suggest the dilemma of choosing talent is not so much a winner’s curse, but more like a purchase of a lottery ticket. Most times you lose, but, if you are going to win, you must buy a ticket.

**References**


