Luck Vs Skill in the NFL DRAFT

In this article, I'm going attempt to test for skill when it comes to the NFL Draft. I know there are few others who've done this, including the economist Cade Massey, but I thought it would be interesting to test it using a different set of methods. The methods I use are all directly adapted from a working paper by Professors Eugene Fama and Kenneth French, titled, "Luck versus Skill in the Cross section Mutual Fund Alpha estimates." I'll be referencing this paper throughout the article.

On the surface, testing for skill appears straightforward. If we start by assuming that everything is based on 50-50 chance, then a successful performance over a long period of time would suggest skill. The outliers, those guys and girls who consistently perform well over a very long time, would seem to be the most compelling evidence for skill.

However, thinking of it in this way is potentially misleading. To get a sense of why, imagine if you asked 1000 people to toss a coin 10 times. Most would get something around 5 heads and 5 tails, but not everyone would. Some would get 3 heads and 7 tails and few might even get 1 tail and 9 heads. In fact, if we increase the number of people we ask from 1000 to 100,000, we would expect 97 people to get all of one and none of the other. If we didn't know these were coin tosses, we might actually think those 97 people were really skilled.

The point? Its not enough to identify the outliers and proclaim them skillful. The reason being, if we run something enough times, we're going to find some extreme results even just by chance alone.

Taking this to the NFL, we know there are front offices that find a lot of great players in the draft and many who don't. It's not enough to identify the one's who have great success over time and proclaim them skillful. The challenge instead becomes trying to distinguish if they were skilled or if they were just lucky. This is especially hard given the fact that our sample, much like that of mutual funds, consists of a lot of GMs who weren't hired for very long and few that survived for a very long time. This compositional mix makes the problem all the more difficult.

This is where the genius behind the paper comes in. Rather than model the results directly, we can instead infer if the results we see are similar to the one's we'd see if everything was done just by chance. In essence, imagine two worlds, the first is what we observed and the other is what we create if everything occurs by chance alone. And by formally comparing the two worlds, we can see how similar they are and infer whether or not we are seeing luck or skill.

In order to do this, we need to define three things: Draft Value, Draft Regimes, and a Skill measurement.

Draft Value is a metric that quantitatively measures a player's career success, like a stock's return is a measure of a stock's value. A popular Draft Value metric is profootballreference.com's Approximate Value(AV). However, AV has some issues that make it problematic. Take a guy like Peyton Manning, who is the leader in career AV. Even if you believe Bill Polian deserves complete credit for drafting Peyton Manning, the amount of AV he gets from him is going to enormous, just by virtue of Manning

being Manning. Manning is also QB and by the nature of the position, is going to accrue much more AV than say a standard hall of fame running back. This causes scoring biases across all the positions.

Instead, I created my own draft value metric, based off a set of criteria like starts, probowls, and all pros, but it is structured so that all positions are graded equally and on a similar scale. Thus, a QB like Manning(or Brady) will not be scored higher than a hall of fame RB/LT/G.

Draft Regime represents the front office or primary decision maker for the team. A Draft Regime is defined as a period tracking the longest tenured stay of either the GM or the head coach. For example, the current Ravens regime runs from when Ozzie Newsome took the job, despite having two different head coaches. The current Patriots regime runs since Belichick became the head coach, despite having a few different GMs. For our sample, only regimes that made made a minimum of 24 picks were included(roughly 6 picks per year over 4 years). Special teams players, specialists, and fullbacks were not included as picks in the sample. I can address why in the comments section.

Finally, we come to our measurement for *skill*. To see this, we take a regression:

$$\text{Draft Value} = \alpha + \beta * \text{Pick} + \varepsilon$$

The actual regression is a bit more complicated, but the intuition is the same. For those unfamiliar with regression analysis, don't worry too much about all the symbols, just focus on the intuition. The right hand variable, **Pick**, represents where player happened to be selected in the draft. The aformentioned Manning, for example, was selected with the first overall pick, so his pick would be 1.

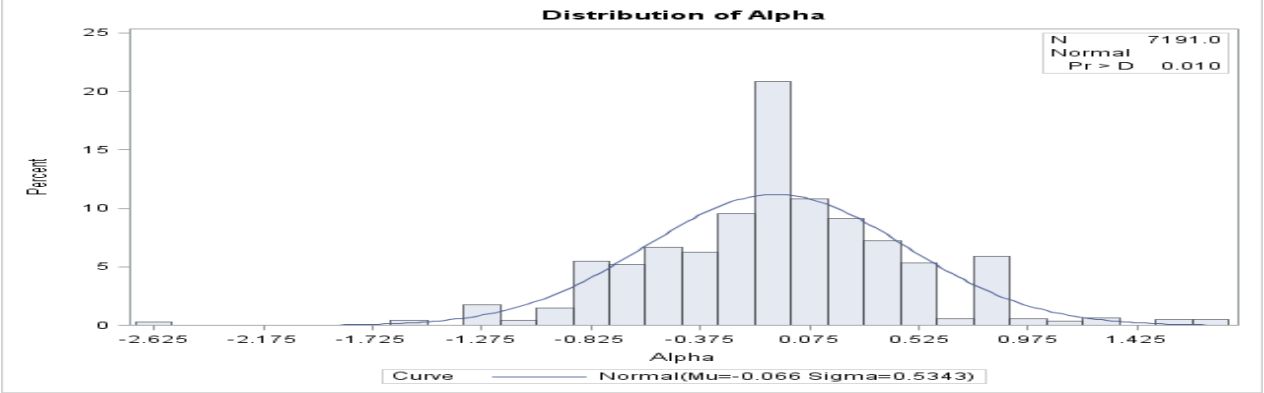
In principle, it is assumed that high draft picks are spent on players with better abilities. By contrast, we can assume lower round picks are not as talented as players selected in the first round. In other words, where a player is selected is going to have a big say about his Draft Value; because the best players are often selected early. John Elways of the world go 1st overall for a reason and likely have a hall of fame career no matter who drafts them.

The crux of the equation above essentially asks, how much of a Draft Regime's success is purely driven just by where he picks? By controlling for this, we can thus ask - can a Draft Regime find successful players above and beyond where they happened to be picked? Ie- can a skilled Draft Regime find Andrew Luck without needing the first overall pick? Can they find and develop valuable players in the lower rounds? That becomes our definition of skill.

Skill is represented here by the alpha coefficient: α. It represents the *difference* between the Draft Value of the player and the Draft Value explained by ***where*** the player was selected. In other words, a large positive α means a Draft Regime is really hitting on their picks above and beyond what we'd expect.

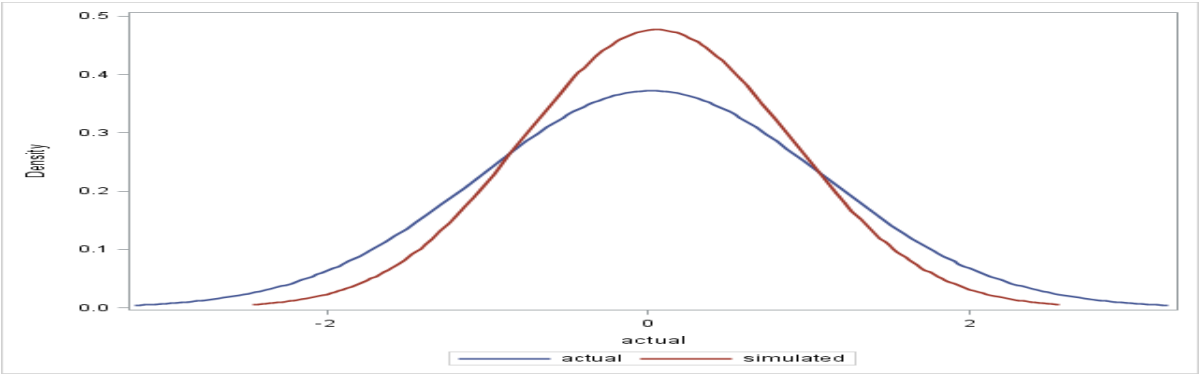
Conversely, a large negative α means a Draft Regime is consistently missing despite getting to pick early in the draft. Finally, an α close to 0 implies most Draft Regimes are getting pretty much the player we'd expect given the pick. In other words, no real skill.

Running the regression and then collecting all of the alphas for all of the Draft Regimes in the sample, I plot the distribution below:



Most regimes are clustered around 0, suggesting most demonstrate very little skill, but a few are stretched far out there. These are the Draft Regimes who are far outpicking(or failing to pick) their draft slots and ostensibly, demonstrating skill(or lack off). Of course, we can't stop here, since these results still could have happened by chance alone.

The next step is to simulate a world in which there is no skill and then compare the two. This is done by setting setting α to 0 and simulating this 10,000 times. We then generate a new set of of alphas, this one based strictly on chance. Plotting the two distributions, we get:



At first glance, they appear similar. We can formally test their similarity, using a Kolmogorov-Smirnoff Test. After running the test, we get a test statistic of 0.798, or not very significant. In other words, it is there's a good chance those outlier Draft Regimes were just lucky/unlucky.

We could stop here, but notice that the two distributions aren't EXACTLY the same. In the actual distribution, we see more extremes (demonstrations of skill) than our luck based simulation. This made me wonder, what could be done to make the two exactly similar? In other words, if instead of simulating no skill, I added a bit of skill to the simulation. How much skill would it take to make the two distributions look EXACTLY the same?

This is the final thing the paper did, adding a small bit of skill to the simulation. I mentioned the paper again because their results essentially matched my own. They found that a 0.5% level of skill pretty much produces a close match. You can't tell if 0.5% is a lot or a little unless you're in the data yourself, but it is a fairly small figure and as a result, a somewhat weak bit of evidence for skill. In fact, they remark, “This is not much performance, and perhaps as a result, the evidence against the no performance alternative (x = 0.0%) is rather weak." Translation: All it took was a very small addition of skill to make the two look exact, so maybe there isn't any skill there at all.

So what to make of all of it? I suppose that will come down to whether you're a cynic or an optimist. Here's my take: While its small, I think that 0.5% is evidence of some skill. It's not going to mean finding a Richard Sherman or Tom Brady, but it could mean finding a marginally better first rounder or a useful reserve in the 5th round.

Still, franchise success and GM reputations are rarely built on marginal value or solid 5th rounders. They get made on the star players, "diamonds in the rough", and the hall of famers that transcend their eras. And that I think is the main takeaway here. When it comes to the draft, whether you believe in skill or not, most of the success and failure is where you pick and what kind of player happens to be there.