

Grinding the Bayes

A Hierarchical Modeling Approach to Predicting the NFL Draft

Benjamin Robinson, Grinding the Mocks LLC

October 19, 2020

Abstract

Using the 2018 NFL Draft as a case study, this paper extends my initial work (Robinson 2020) on the efficacy of using mock draft data to forecast player-level draft outcomes. By using the same data and applying a more rigorous analytical approach (Bayesian hierarchical modeling with Markov Chain Monte Carlo simulations), methods are developed that allow for NFL decision makers to more accurately gauge when a prospect is likely to be selected in the draft. This log-adjusted measure of Expected Draft Position (EDP) explains about 85 percent of variance in actual log-adjusted draft outcomes. Finally, I discuss the implications of using these metrics to inform draft strategy and compare how EDP relates to on-field production.

1 Introduction

As the prominence of the NFL and its annual player draft has grown, predicting the draft through the creation of “mock drafts” has grown into a cottage industry. In this paper, I treat mock drafts developed by members of the draft community (media, experts, and fans), so called “Draftniks”, as data, collect them, and develop measures to quantify and predict the range of the Expected Draft Positions of 2018 draft-eligible collegiate football players.

In “Grinding the Mocks: Using Mock Draft Data to Predict the NFL Draft” (Robinson 2020), I found that mock draft data paired with relatively simple methods have strong associations with actual draft selection on a logarithmic basis. However, these approaches also have downsides: imprecise confidence/prediction intervals, a failure to account for the nested nature of mock draft data, and have features that are not a good match for the underlying complexity of the draft.

Interest in the NFL Draft is high and rightly so for NFL teams and their fans. Research has shown that from 2003 to 2014 NFL, a team’s record from the previous season was not as highly correlated with their record in the following season as it was in other sports (Lopez 2016). This research largely attributes this phenomenon to the NFL’s relatively shorter sixteen game schedule because a difference in record of a few games from the year prior is a larger change in winning percentage than in a sport like baseball where teams play one hundred and sixty two games. Additionally, the worst teams get higher draft selecting, theoretically giving poor teams a great chance at draft higher quality players and improving the talent level of those teams.

Teams can gain a competitive edge in this area by optimizing their draft strategy by selecting the players they deem to be the “best” inside those players’ range of expected draft outcomes. Within the context of the draft, data informed decision making can give teams a better understanding of how the draft is expected to shake out. From there, front office decision makers can use these insights to craft effective strategies that take into account the inherent uncertainty of the draft with the ultimate goal of maximizing the efficiency of scarce draft capital.

2 Literature Review

2.1 The NFL Draft

There has been quite a bit of research published about the NFL Draft. This research has largely focused on viewing the draft through the lense of market economics. Using this economic lens, researchers have analyzed the draft to understand how efficiently NFL teams value players and ultimately how that relates to NFL success (see (Berri and Simmons 2009) and (Mulholland and Jensen 2014)). Of particular note is Massey and Thaler’s landmark research exploring the lack of value gained by teams that traded up for higher draft selections versus those that traded down to accrue more picks (Massey and Thaler 2013).

However, there has been little research overall about predicting draft outcomes. In 2014, Brian Burke, currently at ESPN, developed what he called the “Bayesian Draft Analysis Tool” (Burke 2014). Burke used expert draft analyst prospect rankings and employed a Bayesian framework using the past performance of those analysts to make adjustments to where they ranked players. This early work greatly influenced ESPN’s “NFL Draft Predictor” product, which they unveiled for the 2020 NFL Draft (Walder 2020). Burke initially intended his tool to provide a “reasonable first guess” and to act as a decision support tool that would equip teams and fans with a better sense of how likely players were to be selected at each pick in the draft.

In addition to Burke’s work, one study in particular investigated the use of mock drafts to forecast draft outcomes. The research looked at individual mock draft accuracy using data from The Huddle Report, a website that hosts an annual mock draft contest for qualifying Draftniks and analysts (Caudill and Mixon 2013). The authors critique the mock draft industry’s lack of statistical rigor in evaluating which analysts produce the most accurate predictions and also propose alternative measures of mock draft accuracy: Euclidean distance, Spearman rank-correlation, and salary misallocation.

2.2 Social Science Research

Led by the University of Pennsylvania’s Philip Tetlock, the technique of aggregating the predictions of experts to estimate the likelihood of an event has become popular among social scientists. Tetlock’s influential work on the border of political science and social psychology used the “wisdom of crowds” to assess how well expert forecaterers were at predicting the likelihood of a range of international political events. In his 2005 book, “Expert Political Judgment: How Good Is It? How Can We Know?”, Tetlock explored the accuracy of predictions made by experts in various fields (Tetlock 2008). Further investigation found that not only were expert forecaterers often just as accurate as others but that these prognosticators were neither held accountable for their inaccurate predictions nor open to criticism when faced with the news of their inaccurate forecasts.

Tetlock followed up this work with “Superforecasting: The Art and Science of Prediction”. This book focused on the findings of “The Good Judgement Project”, an initiative to determine if predictions derived from the “wisdom of crowds” could outperform expert predictions in a multi-year government sponsored global political events prediction tournament (Tetlock and Gardner 2015). The authors employed Brier scores, developed by meteorologists to rate the accuracy of weather forecasts, to determine which predictions were most accurate (Brier 1950). Using this data, Tetlock investigated the qualities of seemingly neophyte forecaterers whose forecasts outperformed subject matter experts and how the qualities of these “superforecasters” could be understood and harnessed to improve forecasting and prediction making.

3 Data and Methodology

3.1 Data

The main novelty in my approach to analyzing and predicting the NFL Draft is in the data collection. For each mock draft, which were gathered through a combination of web scraping and manual data entry, I collected: identifying information about the mock draft, identifying information about the draft eligible prospects, and descriptive information about the pick. This allows for the prospect’s “draft stock” to be tracked throughout the 2018 draft season. Additionally, a linear weight was created based on the inverse difference between the date of the mock daft and the day after the start of the NFL Draft. This index is used

to downweight the importance of mock drafts created before the draft and upweight the importance of mock drafts created closer to the draft that have more complete information about the players’ themselves and the needs of the teams selecting them.

Table 1: Summary Statistics of NFL Mock Draft Data. Source: Grinding the Mocks, 2020.

Draft Year	Number of Mock Drafts	Number of Draft-Eligible Players	Number of EDP-Eligible Players*
2018	424	414	178

* EDP-Eligible Players must have 10 or more Mock Drafts.

3.2 Methodology

In developing this Expected Draft Position (EDP) metric, I focus on implementing a Bayesian hierarchical modeling framework. In “Grinding the Mocks” (Robinson 2020), I employed a frequentist approach to estimating EDP because the main goal of the original work was to prove the concept. With the concept proven, it is fair to critique the shortcomings of that approach. Predicting where a player is likely to be drafted is a highly non-trivial problem of path dependent decision making in the presence of uncertainty.

At its core, the draft is a stochastic optimization problem whose solution requires an appropriately probabilistic model of the drafting process as a basis for running Markov Chain Monte Carlo (MCMC) simulations to find the range of players’ draft outcomes. Simple unconditional statistical measures are just not well suited to untangle the underlying complexity of the draft. Additionally, it is necessary to employ a hierarchical approach because of the nested nature of the mock draft data. With that knowledge in hand, a hierarchical model is best suited because it takes advantage of the relationships within the dataset that make observations dependent on each other, an important assumption of linear models that makes them less suited to answering this question.

The hierarchical approach also benefits from the use of a Gamma distribution (with a logarithmic link function) to provide a more realistic and flexible distribution for modeling mock draft selections. The Gamma distribution is widely used to model variables that are always positive and is often used for modeling the amount of time until an event occurs in a Poisson process (Mun 2005). Another benefit to this approach is that the logarithm of the Gamma distribution is approximately normal. This helps quite a bit with the interpretability of the model and the model’s output. An important feature of Bayesian methods are their ability to include information about the prior distribution of the data in their modelling approach. This application of Bayesian models does not include any information about the prior distribution of the data because I am only using one year of data in this paper. Future research will include developing appropriate priors and optimal specifications of this model.

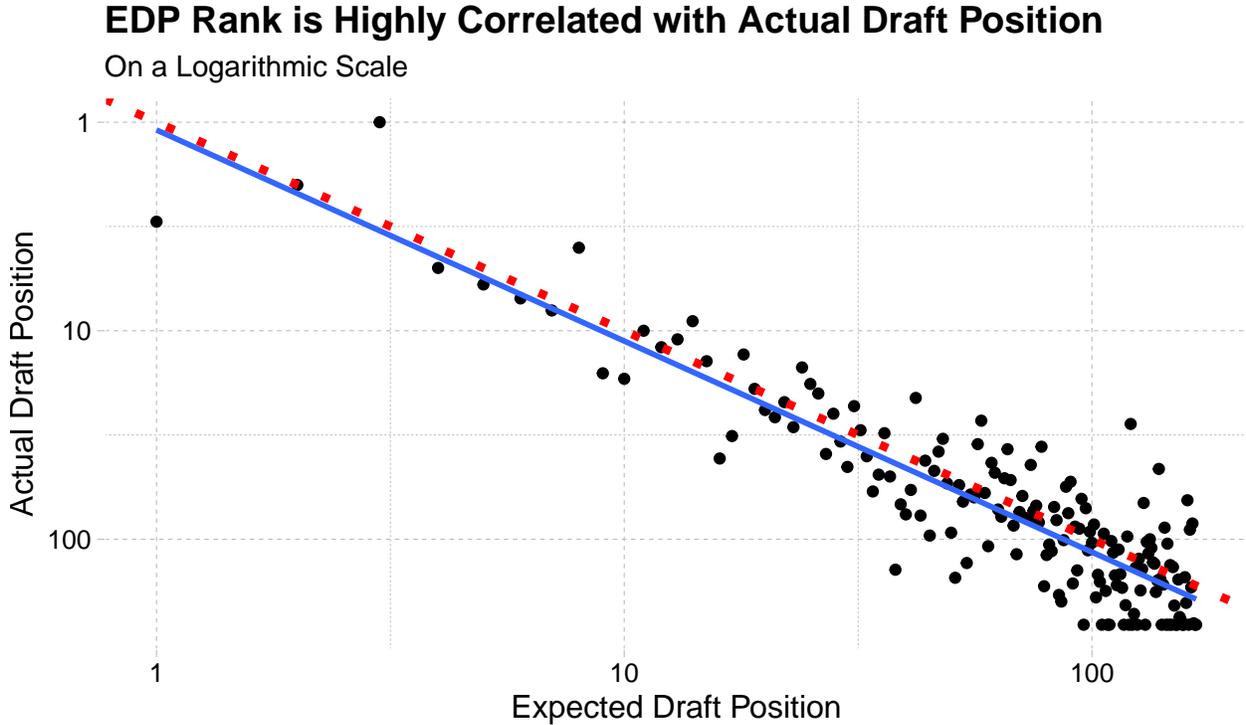
Despite its frequentist approach, there are multiple lessons to be taken from my prior research (Robinson 2020). One is to include Expected Draft Rank (the ranking of players by their EDP) as the metric of choice and that the performance of the model predictions improve drastically after implementing logarithmic transformations and weighting players’ EDP based on their share of mock drafts. A benefit of the logarithmic transformation is that it constricts the numerical space of the distribution but more importantly also implicitly values earlier picks in the draft higher than later ones. This makes the difference between later picks that might be of the same magnitude linearly much different logarithmically. For example, the difference between 4 and 1 and 104 and 101 is the same: 3. However, the difference between the logarithm of 4 and 1 is larger (about .6) than the difference between the logarithm of 104 and 101 (about .01) by a factor of 47, accounting for the high value of early picks.

The Bayesian model is run using the mock draft pick as the outcome variable with a random effects specification for players and their draft weight index values. This allows individual players to vary randomly in terms of their intercept and the effect of time before the draft (which is known as a “random slope”). I chose not to include any fixed effects in the model because I believe that quite a lot of the issues that fixed effects would be used for are already accounted for in the mock drafts themselves. Including them in the model would

Table 2: Results of Linear Regression Models of Log Transformed EDP. Source: 'Grinding the Mocks', 2020.

Metric	Transformation	R-Squared	Adjusted R-Squared	AIC	BIC
EDP Rank	Logarithmic	0.8622545	0.8614197	50.2533	59.60728

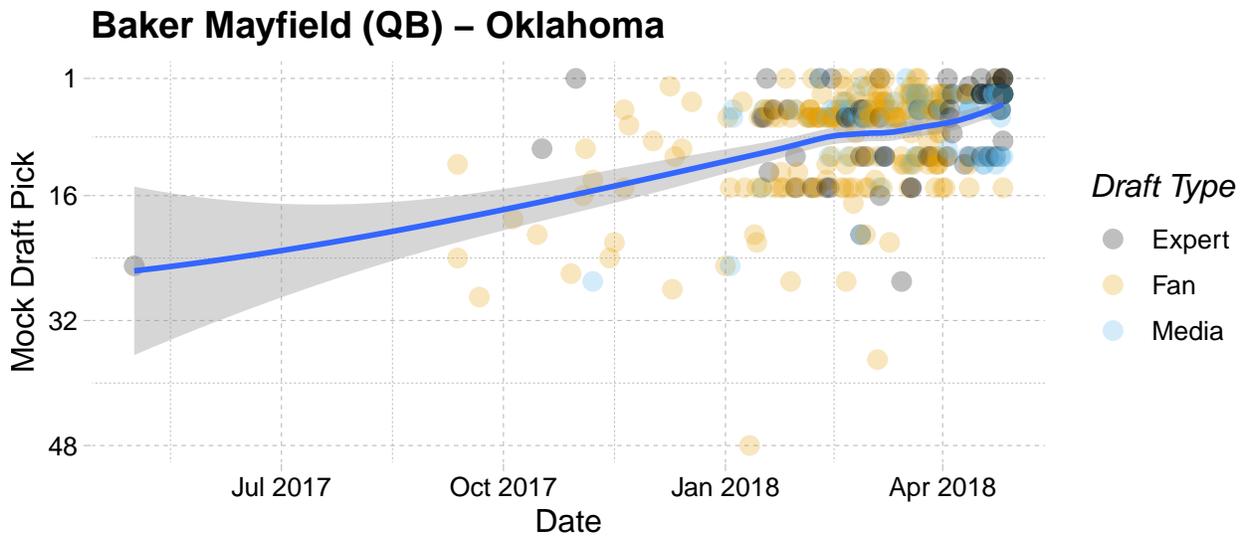
introduce needless complexity and ultimately distort the model's predictions. This means I end up with levels for each player and random effects estimates on the draft weight index as well as random slopes. From there, the model is fit using a Markov Chain Monte Carlo simulation with 4 chains and 5000 iterations to produce each prospect's posterior distributions.



Source: Grinding the Mocks, 2020.

4 Analysis

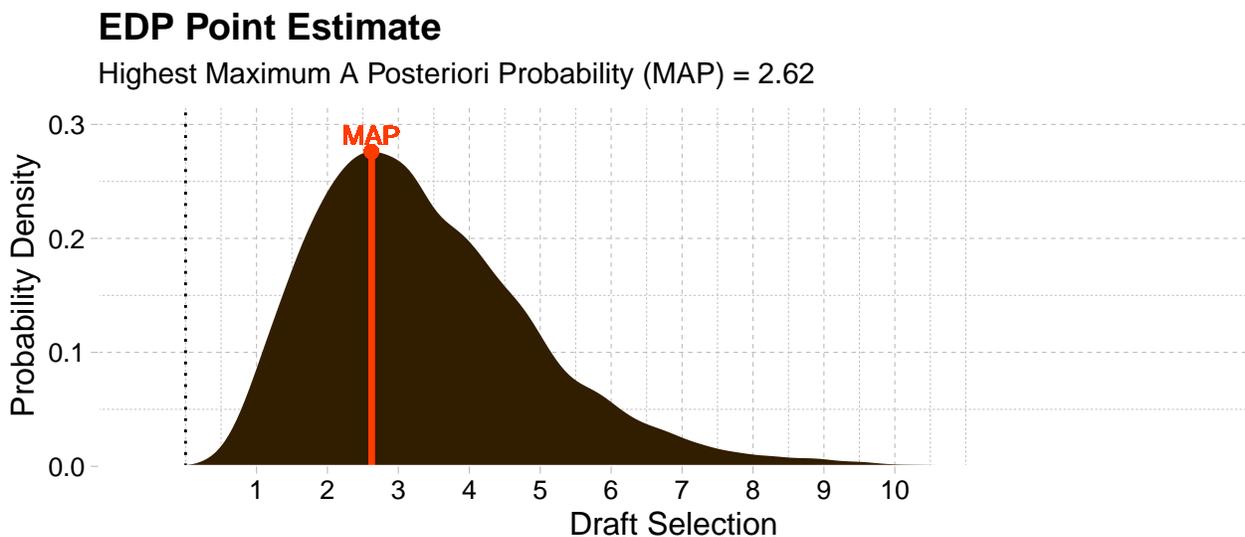
Throughout this paper, I will be using Baker Mayfield, who played quarterback at the University of Oklahoma and was the first pick in the 2018 NFL Draft, as a case study for examining how the Bayesian EDP model treats individual player level outcomes. Mayfield's draft story is an interesting one. While he was selected with the first overall pick, he was not expected to be drafted at that pick by mock drafts. That title belongs to University of Southern California quarterback Sam Darnold. How likely was Mayfield to be drafted first overall? And what were his range of expected outcomes? Let's take a look at his mock draft data to see how his stock changed over time.



Mayfield started the draft process as a potential late first round pick in mock drafts and saw his stock rise throughout his senior season, where he was the winner of the Heisman Trophy as the most outstanding player in college football and led Oklahoma to the College Football Playoffs. What did the Bayesian Model think of his stock on draft night? It gave him an EDP of about 3, which ranked him as the third highest prospect in the 2018 NFL Draft. However, the 95% confidence interval of his EDP put his draft stock range between 1 and 7 so maybe we shouldn't have been too surprised that Mayfield was selected first overall. This prediction passes the sniff test but it doesn't tell us quite enough about the probability of Mayfield being selected at each pick.

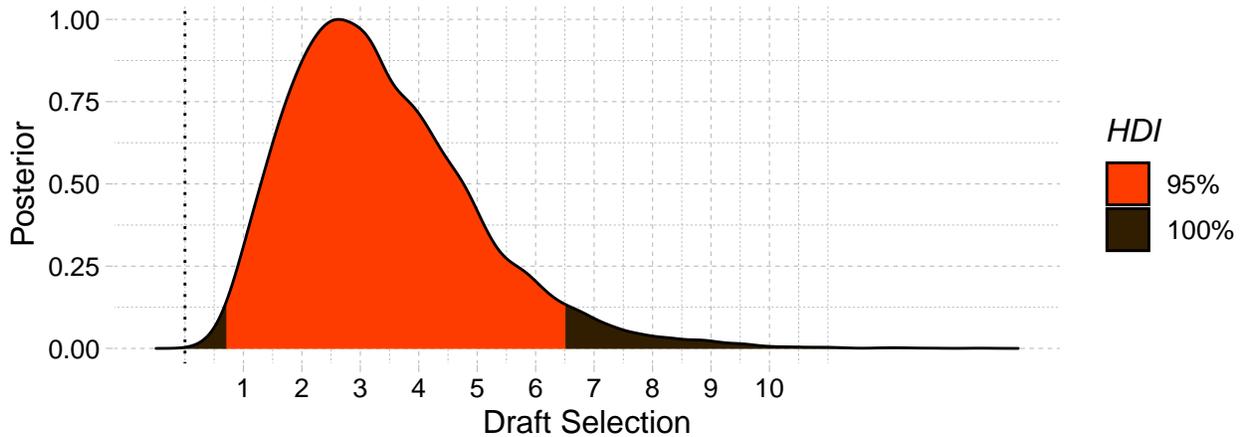
Table 3: Results of Bayesian Prediction for Baker Mayfield's EDP. Source: Grinding the Mocks, 2020.

Name	Position	School	Estimate	Est.Error	Q2.5	Q97.5
Baker Mayfield	QB	Oklahoma	3.390166	1.609972	1.032777	7.191859



EDP Credible Interval

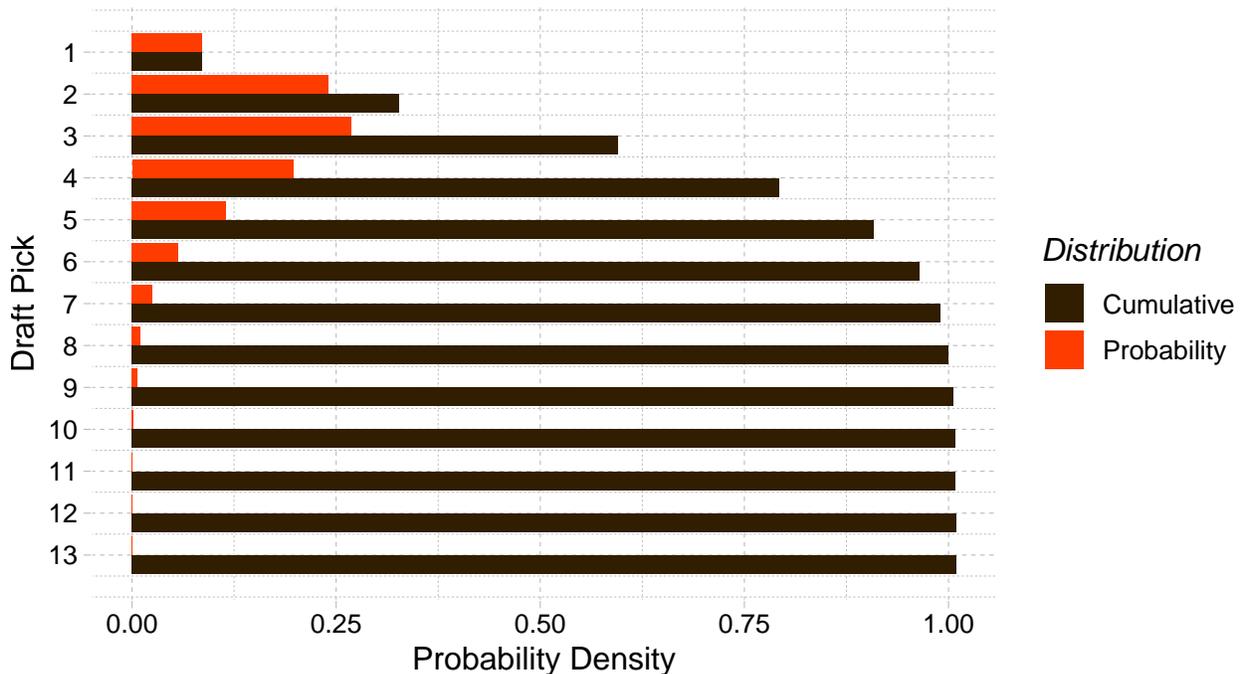
Highest Density Interval (HDI) = 0.7 – 6.51



Source: Grinding the Mocks, 2020.

Using Baker Mayfield’s draft night posterior distribution generated from the Bayesian EDP model, we can more closely investigate his EDP and the probabilities surrounding his draft selection. Let’s first take a look at the probability of him being selected at each pick and calculate his highest Maximum A Posteriori Probability, also known as MAP, estimate. MAP is basically the mode of the distribution of EDP probability estimates. In Mayfield’s case, his MAP was equal to about 2.59, with the probability density being between .25 and .3. Another metric we can use to evaluate Mayfield’s EDP probability estimates is called the Highest Density Interval, or HDI. The HDI is Bayesian statistic’s equivalent to a confidence interval in frequentist statistics and communicates the range of selections between which Mayfield has a higher probability of being selected than not based on the posterior distribution of his Bayesian EDP estimates. In Mayfield’s case, his HDI ranges from .7 to 6.51, which aligns fairly closely with his initial prediction interval.

Baker Mayfield's Pick by Pick Probability of Draft Selection



Source: Grinding the Mocks, 2020.

The final metric we will explore is Baker Mayfield’s (cumulative) probability density distribution. This metric will give us an estimate of his probability of being selected at each integer pick the model has an estimate of. Instead of getting the smoothed out probability density chart where we saw Mayfield’s MAP, this chart provides a more discrete measure of the probability of being selected at each pick ranging from pick one to pick thirteen, which is the last pick where the model can provide a probability density estimate. Looking at his cumulative probability distribution, we can see that the marginal likelihood of Mayfield being drafted increased much slower going from pick three to pick four. In the 2018 NFL Draft, the third pick belonged to the New York Jets and mock drafts expected them to select Baker Mayfield because they expected for the Cleveland Browns to draft Sam Darnold first overall. This would have allowed Mayfield to “fall” to the Jets at pick three. The Browns had multiple picks in the top five so Draftniks assumed (correctly) that they would use their first pick on a quarterback and draft a player at a different position at pick 4.

5 Considerations for an Optimal Draft Strategy

5.1 Considerations

So what does this all mean for teams trying to make the best decisions they can in the draft? This paper assumes that teams are focused on getting the most out of their scarce draft capital and thus would not want to select a player outside of his range of expected draft outcomes. This is especially true given the random nature of post-draft outcomes in the NFL. However, teams make decisions without regard for this efficiency principle quite often for any number of reasons. Each team has its own process for player evaluation that leads them to draw different conclusions about the relative merits of players. Thus it would also make sense that teams would have similarly disparate practices for how they think about when predicting when players will be selected in the draft. Some teams may lean on the industry connections of their player personnel and college scouting groups to get a sense of where individual players are valued. Some might employ more quantitative methods like I have outlined in this paper. And other teams may not care much at all about either of these considerations in their decision making process during the draft.

An important note to make about EDP is that it does not explain 100% of draft outcomes and only has access to information the general public has. Therefore, teams using EDP’s insights will have to learn to understand what type of risk they are willing to accept when making decisions in the draft. How comfortable are decision makers with giving up the opportunity of drafting a player in hopes that the player will be available at a later pick? Given this variation in decision making and in risk profiles, what is an optimal use of Expected Draft Position as part of a team’s overall draft process? A story from the 2020 NFL Draft illustrates these concepts well.

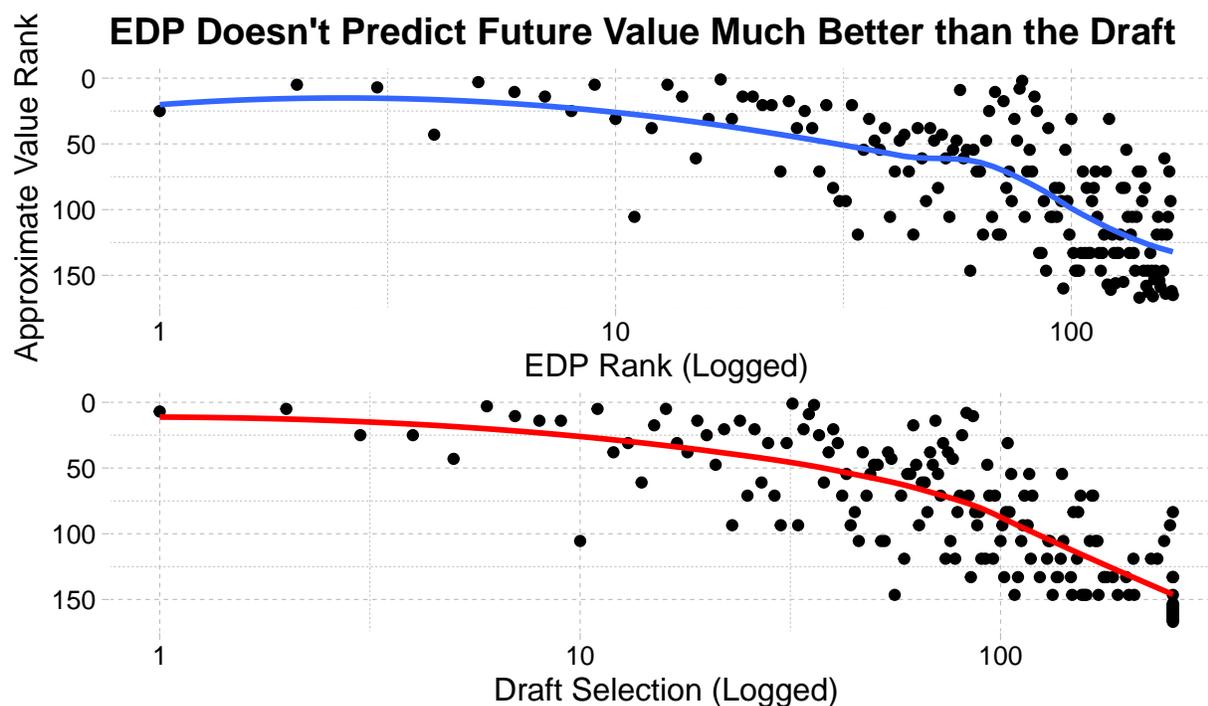
First year general manager of the Cleveland Browns Andrew Berry has a decision to make: the Indianapolis Colts are on the phone asking if the Browns would be willing to trade their 41st overall selection (a second round pick) in exchange for the Colts 44th (also a second round pick) and 160th (a fifth round pick) overall picks. Berry has multiple considerations in front of him: he can stay where he is and select Louisiana State University defensive back Grant Delpit (the player he’s been targeting all along) or he can accept the trade, move down three spots, and hope that Delpit is available at that point. He understands that there is a risk in losing out on drafting Delpit but he’s comfortable with it. Why is that? First, Berry consults with his pro scouting staff to get a sense of what the needs are of the team’s drafting in front of the Browns. Then he listens to his research and development staff to get a sense of what the odds are of Delpit coming off the board at each pick. And he ultimately makes an educated decision. Berry decides to accept the Colts offer, which the Colts use to draft Wisconsin running back Jonathan Taylor, and three picks later Delpit is still there and he selects him: Hook, line, and sinker (Zegura 2020).

5.2 Future Value

One last question with significance to NFL teams is: which metric is more predictive of a player’s future value in the NFL; his EDP or his draft selection? To answer this question, I used data from Pro Football Reference on each player in the 2018 NFL Draft class’s Approximate Value to examine if, all things being equal, EDP outperformed draft selection in terms of future value.

Since there aren't many publicly available metrics for football that cover every player on the field and attempt to measure player value on the same scale (such as Wins Above Replacement in baseball), Approximate Value is the best option public analysts have. A plus of AV is that it covers every single position, which means we don't have to deal with biased "counting" metrics that don't do a good job assessing play at positions such as offensive line, where players don't really provide those types of statistics.

Ultimately, the answer is close. Despite the limited amount of players and time for players to accrue Approximate Value and even when accounting for the value of higher picks (using the same logarithmic transformation applied previously) there doesn't seem much of a meaningful difference between EDP and draft slot when it comes to predicting future value in the NFL. This result highlights that player evaluation can and should still play a big role in determining how teams set up their draft boards and that EDP can guide decision makers in using their limited draft resources most efficiently to capture as much of a class's potential future value as possible.



Source: Grinding the Mocks & Pro Football Reference, 2020.

6 Conclusion

This paper introduces a Bayesian framework for the analysis of mock draft data to predict the player-level outcomes of the 2018 NFL Draft. The problem is important because NFL draft prospects have different draft probability distributions and decision makers should take this into account when managing their drafts.

In an article he penned for *The Athletic*, Michael Lombardi (Lombardi 2019), the long-time NFL front office executive, writes: "Building an algorithm of all the mocks. . . allows teams to know and understand which players are considered in the top 100, top 200 and top 300. But it's not foolproof." Lombardi is right that aggregating mock drafts is not foolproof. However, in light of this evidence, it would be unwise to ignore this data as a tool to inform decision-making as a compliment to the hard work of scouts and team personnel. Teams have limited selections in the draft and it is in their best interest to gain the greatest possible efficiency from these choices. As analytics and data science continue to make progress in the NFL, teams that can utilize these tools in varied and innovative ways are more likely to find "edges" to give their organizations a stronger chance to succeed on Draft Day and perhaps eventually on the field.

7 References

- Berri, David J., and Rob Simmons. 2009. "Catching a Draft: On the Process of Selecting Quarterbacks in the National Football League Amateur Draft." *Journal of Productivity Analysis* 35: 37–49.
- Brier, W. G. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78: 1–3.
- Burke, Brian. 2014. "Bayesian Draft Prediction Model." Edited by Advanced Football Analytics. <http://archive.advancedfootballanalytics.com/2014/04/bayesian-draft-prediction-model.html>.
- Caudill, Mixon Jr., Steven B., and C. Paul Mixon. 2013. "NFL Draftnikology: Euclidean Metrics and Other Approaches to Scoring Ranking Predictions." *Communications in Statistics - Simulation and Computation* 43: 237–48.
- Lombardi, Michael. 2019. "'A control center of information': What goes on inside a team's NFL draft room." Edited by The Athletic. <https://theathletic.com/926108/2019/04/16/inside-an-nfl-draft-war-room/>.
- Lopez, Michael. 2016. "Exploring Consistency in Professional Sports: How the NFL's Parity is Somewhat of a Hoax." Edited by Sloan Sports Analytics Conference. http://www.sloansportsconference.com/mit_news/exploring-consistency-in-professional-sports-how-the-nfls-parity-is-somewhat-of-a-hoax/.
- Massey, Cade, and Richard Thaler. 2013. "Loser's Curse: Overconfidence Vs. Market Efficiency in the Nfl Draft." *Management Science* 59: 1479–95.
- Mulholland, Jason, and Shane T. Jensen. 2014. "Predicting the Draft and Career Success of Tight Ends in the National Football League." *Journal of Quantitative Analysis in Sports* 10: 381–96.
- Mun, Johnathan. 2005. *Advanced Analytical Models: Over 800 Models and 300 Applications from the Basel II Accord to Wall Street and Beyond*. Wiley.
- Robinson, Benjamin. 2020. "The Value of Mock Drafts." Edited by Football Outsiders. <https://www.footballoutsiders.com/stat-analysis/2020/value-mock-drafts>.
- Tetlock, Philip. 2008. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.
- Tetlock, Philip, and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction?* Crown Publishers.
- Walder, Seth. 2020. "ESPN NFL Draft Predictor: Chances Tua Tagovailoa be available at No. 5 and more." Edited by ESPN. https://www.espn.com/nfl/insider/story/_/id/29035553/espn-nfl-draft-predictor-chances-tua-tagovailoa-available-no-5-more.
- Zegura, Nathan. 2020. "Andrew Berry recaps 2020 NFL Draft with Nathan Zegura." Edited by Cleveland Browns. <https://www.clevelandbrowns.com/video/andrew-berry-recaps-2020-nfl-draft-with-nathan-zegura>.