**Spring, 2016**

András Kárpáti & Mór Kapronczay

Rajk László COllege for Advanced Studies Econometrics Course

Instructor: Balázs Muraközy

Pick me!

Analysis of NFL Quarterback statistics of the past decade

Abstract

In this paper the authors analyze data about quarterbacks drafted between 2005 and 2009. The aim of the authors was to explore the utility function of scouts regarding drafting quarterbacks, and the preference relations on the draftees. Then these results are used to estimate the relationship between the Draft and the future NFL career of a player. An important finding of this study is that the Draft is indeed a market with close to perfect information, where scouts’ preference relations concerning quarterbacks are close to predictable, but neither success in the Draft nor any other statistical data can trustworthily predict future success in the NFL

**Keywords: NFL, Draft, quarterback, econometrics, modeling**

**JEL-code: Z20**

Table of Contents

[I. Introduction 1](#_Toc451097620)

[II. The Data 4](#_Toc451097621)

[1. Datasets and Variables 4](#_Toc451097622)

[2. Evidence of strategic behavior in Combine data 6](#_Toc451097623)

[III. Hypotheses 7](#_Toc451097624)

[1. Draft 7](#_Toc451097625)

[2. NFL 8](#_Toc451097626)

[IV. Descriptive Analysis 9](#_Toc451097627)

[1. Athletic performance 9](#_Toc451097628)

[2. Performance in the NCAA 9](#_Toc451097629)

[V. Results 11](#_Toc451097630)

[1. Draft 11](#_Toc451097631)

[a. Utility function 11](#_Toc451097632)

[b. Predicting draft results from combine and NCAA data 14](#_Toc451097633)

[c. An original interpretation of these results 14](#_Toc451097634)

[2. NFL 15](#_Toc451097635)

[VI. Concluding remarks and further research 18](#_Toc451097636)

[VII. References 19](#_Toc451097637)

[List of Figures: 20](#_Toc451097638)

[List of Tables: 20](#_Toc451097639)

# Introduction

|  |
| --- |
| Source: ESPN.com; self-made chart |
| 1. Figure: The Draft is a multi-million dollar market (Quarterback median and total contract value) |

In the recent decade, statistical analysis of sports went through a revolution after *Moneyball: The art of winning an Unfair Game* (Lewis, 2003) was published. Lewis told the story of Billy Beane, a baseball team coach, who achieved an unlikely success with a very low budget. The key to the team’s success was Beane’s statistical evidence based draft strategy, called sabermetrics. Since then, the development of information technology and the abundance of feasible data opened up an interesting field of research.

The aim of this paper is to analyze the performance and career of NFL quarterbacks in the time period between 2004 and 2015. Our main goal is to understand the basic mechanics of the NFL Draft, and the relation of the Draft and other statistics known at the time of Draft, to future performance. The intuition of this analysis partly comes from Eckstein’s (2009) research done on NFL running backs. For our analysis, it is vital to understand the Draft and its underlying mechanisms, as well as the basics of the background game theory aspects.

Player selection in the NFL is a complicated process with many stages. Players attend college, where they play in the college’s football team, then they go to the Combine, where they take athletic and IQ tests, then they get drafted (or not) into NFL teams. First they are usually not starters in their team, but they gradually move from being a replacement to a starting player. This process can be accelerated or decelerated by injuries, transfers and more.

The NFL Draft in our view is a multi – million dollar market with almost perfect information. Because of the excessive statistical coverage of the NFL, statistical data, various analyses, advanced metrics and even qualitative analyses are available for long time periods. Moreover, the richness of data even goes to back to college years of the player, and on top of it, the Draft supplies us with comparable basic information on the players’ physical, and mental abilities. Economists, such as the authors of this paper are baffled by such a playground, since in most real-life business cases an empirical analysis would not be possible. However, it is still an open question, whether we can make predictions based on this data.

But this is a very important “almost” in the first sentence. Perfect information applies to players’ qualities, which teams seem to value the same way. However, there is almost no information concerning players later success in the NFL. So, what teams can do in case they need a quarterback, is to choose the “best” player based on Combine and college football data, the scouts’ instincts, stories and many other.

It is also important, that he NFL has numerous mechanisms to ensure equality of opportunity between teams. For instance, NFL franchises have a fixed budget every year, from which they must maintain their activities. The franchise has to pay the coaching staff, the players and operate their infrastructure from the same amount of money. (overthecap.com, s.a.)[[1]](#footnote-1) Therefore, an affluent owner cannot change much of an NFL franchise (by the way selling a franchise is extremely rare), and moreover efficient consumption of resources is inevitable to success.

The Draft system is also one of these measures for equality of opportunity, and this is an important aspect of the Draft. The primary source of players for NFL teams is drafting college football players who played in the NCAA football league before.[[2]](#footnote-2) Teams have the opportunity to sign players apart from the draft, but this is a phenomenon hard to inspect. Teams select in a reverse order based on their success ranking in the past season – hereby rewarding the least successful team last year with the best opportunity to improve their squad with the best college football players. (NFL.com, s.a.)

Positions also make the picture a little more complicated. Teams have different needs in terms of position, which need is affected by roster depth (how much player playing this position a team has), team philosophy, coaches and numerous other things this paper does not aim to cover. Therefore, a team’s Draft selections are also effected by position needs. In addition, in terms of different positions, it is also important that different abilities or other statistics can be the main factor, because different traits make a good wide receiver, and a good defensive lineman To handle these issues, we decided to inspect one position, and rank of players playing in that same position.

We chose the **quarterback** position as a subject of our analysis. In any later time, when we refer to players, we refer to quarterbacks. On top of that, we recognized that the overall draft order is not relevant: the order in which the quarterbacks (or any players playing in the same position) are selected is the important thing. The overall draft order is affected by many other factors, but the differentiation between players playing in the same position is definitely based on how good a player is.

The reason we chose quarterback is twofold. On the one hand, the quarterback is practically the most important position in an NFL team’s offense. Every offensive game starts with the “center” throwing the ball back to the quarterback. Then, depending on the game plan, the quarterback can do different things, i.e. keep the ball and then throw it to a receiver, hand it to the runner, etc. A game plan can even include option type plays, when the quarterback has to decide based on the defense’s moves to keep the ball or to hand it to the running back.

On the other hand, quarterback position is the most easily assessable position based on data. American football is a team game, therefore every player has an effect on the players performance, but throwing statistics are much better proxies for quarterback worth, than number of tackles, or weight is for a linebacker, because these statistics are not the best proxies, and more intensely effected by teammate’s performance as well.

The goal of the authors with this paper is twofold. First, being a market with almost perfect information, the Draft’s results – in our view – should be at least partly predictable. We expect to at least partly give an explanation on Draft behavior of teams in terms of quarterback selection based on statistical data. Second, we think the NFL performance of a quarterback is unpredictable from combine, college statistics or any other metrics and is more likely a consequence of other, not measurable factors such as attitude. We use econometric analysis to support our hypotheses.

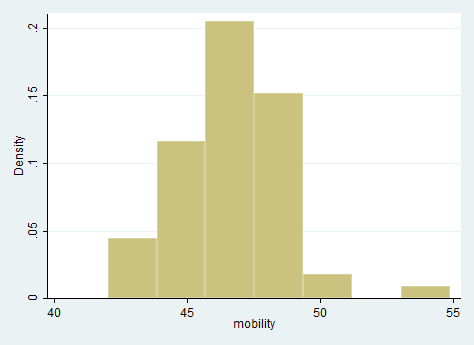
In terms of structure of this paper, we start by giving a brief summary of our dataset and variables, including the advanced metrics which we employed. Then, we develop the data, we are going to work with. After that, we explore our hypotheses, then present the descriptive statistics supporting these hypotheses. Later we present the results of further econometric analysis, first the Draft, then the NFL career, and finally we draw our conclusions.

# The Data

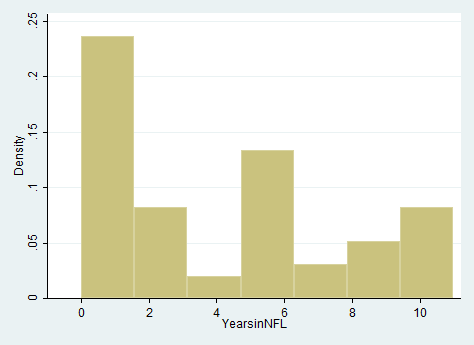
Our dataset consists of observations for every quarterback drafted for the NFL between 2004 and 2009. Our variables include college statistics, combine results, NFL statistics and salary for at least 6 years, and at most 11 years (those who got drafted earlier have more observations). The uniqueness of our study is related to the fact that it is impossible to construct a longer time series from publicly available data, since college statistics are not available prior to 2003.

## Datasets and Variables

|  |
| --- |
|  |
| 3. Figure: Half of drafted quarterbacks do not play in the NFL |

College statistics include every passing statistic from the last two years of college. College statistics were somewhat problematic because of two reasons. First, a large fraction of players were injured or did not play for some other reason during at least one season during our college observation period. Second, college football in the US is organized in different leagues, and most databases miss at least one of these leagues. (sportrac.com, 2005 - 2009)

2. Figure: Distribution of Mobility

The Combine observations are sparse at best. The only variables available for every player were weight and the result of the 40-Yard dash. From these two variables we constructed a measure of **Mobility**, which is the quotient of the weight and the time of the sprint. We adopted the use of this metric from (Eckstein, 2009). This is based on the fact, that if 2 quarterbacks have the same weight, then the quicker is preferred, while if they run equally fast, the stronger and heavier is preferred. Therefore, the higher the quotient of these two values, the higher the mobility of the player.

Source: ESPN.com; self-made chart

Source: ESPN.com; self-made chart

The assumption, that this mobility measure is a better fit than the linear combination of Weight and time of the 40 Yard dash is tested later in the paper. (NFLCombineResults.com, 2005 - 2009)

Some players have available Wonderlic (a sort of intelligence test) scores, but the missing cases would make our estimations biased and less robust. The missing values must be considered endogenous, since most players usually choose which Combine tests to take.

NFL statistics include every kind of passing statistics for quarterbacks for every season up to 2015. Only 36 of our players have at least 4 seasons in the NFL, and of course we will investigate why some players get to have a career in the NFL, and some spend their career on the bench. The longest playing quarterbacks played for 11 seasons.

Our main sources for data was ESPN, and other football statistics websites (NFLCombineResults.com, 2005 - 2009); (Yahoo! Sports, s.a); (DraftHistory.com, s.a); (The Football Database, s.a.); (sports-reference.com, s.a.). Unfortunately, we had to do a lot of data carpenting, since every database misses a few players or years, so in most cases variables come from at least 2 datasets.

On the other hand, there were some pathological cases, which we excluded from our dataset. For example, Ingle Martin, a quarterback drafted in 2006 played in the Canadian League before he was drafted. He did not play in the NFL at all, he was a reserve throughout his employment. There was no data available from him on college, combine or NFL, therefore we excluded him from the dataset.

The temporal structure of our data is as follows: is the year of the draft for each individual player, negative time is college seasons, positive time means years after the draft. So instead of using absolute time, such as the year 2008, we use relative time “2nd year after his draft”, and make corrections using year dummies.

Our assumption is that the performance of the quarterbacks does not exists in a vacuum, but is rather heavily influenced by both teammates and opponents. Since we do not analyze match-to-match data, but season data, we control for the opponents (or should I say the changes of the whole NFL) by year dummies.

As a measure for **Team Efficiency** we used an advanced metric by Football Insiders called Offense DVOA (Defense-adjusted Value over Average). DVOA also breaks down every single play, assigning each play a value. This is also adjusted for the opponent, situation etc. as in Total Quarterback Rating plus it controls for season as well. It basically measures how well a team uses their resources to gain yards, therefore, it is a great measure to control for different NFL offenses. (FootballOutsiders, 2016)

The **Total Quarterback Rating** is a measure, which aims to contextualize *every play made by the quarterback*. To contextualize these plays is to control for the opponent, the current score, remaining time and so on. It also takes into account the value added of teammates taking part in the play. This means that the sheer performance and contribution of the quarterback is being measured. (ESPN, 2011)

The athletic qualities are represented by the variable **Mobility**, which is the ratio of **Weight** and **40 Yard Dash time** measured on the Combine. The higher the Mobility score, the better the quarterback.

And of course, our dataset also include the usual football statistics: Yards per Attempt, Yards thrown, Interceptions, Touchdowns, Fumbles, Attempts, Passer Rating and Completed Passes.

## Data Structure

We constructed two datasets: one for assessing draft performance and one for the future NFL performance. These datasets contain the variables described above, however their structure is different. The first is a cross-section dataset, while the second one is panel data. Both of them are subject to the temporal adjustment described above, while the first database contains 1 observation for every player (which means around 60 observations), while the other one contains 1 observation for each player and each year (only for the years they played), this means roughly 200-250 observations.

## Variables measuring draft performance

**Rank#, Contract Value (ln)** and **Pick#** are first used as outcome of the utility function of scouts. When we estimate NFL performance we employ them as proxies for quarterback ability. We find Quarterback Pick# and Contract Value to be more closely tied to quarterback quality then Pick#, since Pick# is heavily influenced by strategic behavior by the teams and the overall quality of the pool each year.

### Contract Value

Each picked player is awarded a contract at the draft by the picking team. Early picks and star players (the first 50 picks in any given Draft) usually get a very high value contract, while players picked after the 50th Pick usually get a sum roughly around the median. It is also noteworthy, that this variable is heavily influenced by strategic behaviour on both the players' and the teams' side. A player may refuse a low value contract, while the team is constrained by the Salary Cap, that limits the total monthly salary of a teams' players. Since contract value is measured in dollars, we use the natural logarithm of contract value for a better linear fit.

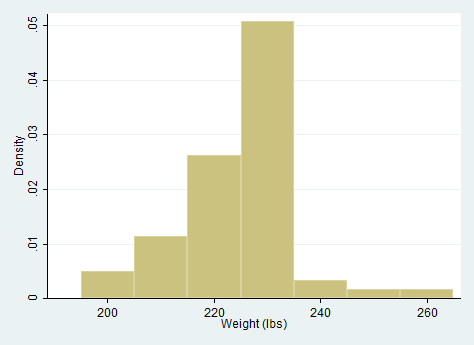
### Pick

Pick measures how many players did the drafting teams pick before the player. It ranges from 1 to over 200 in our dataset (see Figure 8 for example).

### Rank (Quarterback Pick#)

This is basically an ordering of the quarterbacks. We use this as a dependent variable, because we are interested in how well do these quarterbacks do against each other, rather than how well do they do against the complete Draft pool. So the quarterback picked first has Rank 1, while he may be the 10th Pick.

## Evidence of strategic behavior in Combine data

The fact, that weight is measured at the Combine affects the weight of players. As in many other aspects of Draft, when we observe the distribution of Weight, we can observe evidence of strategic behavior. We can see that most players are under 235 lbs., and it is well known that athletes adjust their weight before measurements, we can suspect strategic behavior behind the Weight data. This phenomenon is not further assessed in this paper, but in our view it is interesting enough to take note of.

|  |
| --- |
| Source: NFLCombineResults.com; self-made chart |
| 4. Figure: Distribution of weight through the whole sample. |

# Hypotheses

Being economics students (at least one of us) who love the NFL (at least one of us), before starting the analysis, we already had our hypotheses. Firstly, our prediction was that there is no statistic, metric, score, or any other sort or combination of data that can trustworthily predict a quarterback’s performance in his upcoming NFL years. We implicitly assume, that there are a lot of other factors, primarily attitude towards work that determines later success. However, it is important to stress that this does not contradict our hypothesis, that scouts know their job, and their valuation of the players reflects true potential, and is much better than randomly picking players.

Secondly, we believe that the Draft is a “playground” for exploring scout strategies. As we mentioned before, the NFL Draft is a market with almost perfect information and totally rational actors who seek to maximize their utility (which combines success in terms of NFL results and in economic sense). In terms of the total Draft procedure scouts act strategically. Take the example, when player A is a better (in any meaning of the word) player than player B, and the team’s first two preferences are these two players, respectively. If the drafter expects A to be present in the next round of the draft, but is uncertain about B, he will choose B in the first round, and A in the second. This is the sort of noise in Pick# we mentioned earlier.

This is the reason why the order of selecting quarterbacks in the Draft (Quarterback Pick#, or quarterback rank) is better explanatory variable for NFL performance, and a better explained variable in Draft results than the overall order of selecting any player (Pick#).

## Draft

When we only consider the drafted quarterbacks, there should be a set of preference-relations, which describe the quality of quarterbacks in relation to each other. In other words, we should be able to determine the better quarterback from any two choices.

Our aim is to explore this preference-relation, determine which factors increase the quality of a quarterback, and to reproduce the quality-order of the players for possibly predictive purposes.

Our hypothesis regarding the matter is that scouts look for quarterbacks who are **fast and strong** regarding athletic qualities, and have a **good passing ability**. The passing ability is measured in 2 ways: on one hand, we can include the **Passer Rating** as an explanatory variable, or we can include the building blocks of Passer Rating, such as Percentage of Completed Passes, Interceptions, Thrown Yards, etc.

## NFL

Our hypothesis regarding the NFL is twofold: first we want to investigate which measures available at the Draft are significant in future NFL performance. Second, we hypothesize that scouts know how to do their job, meaning that quarterback quality metrics (Contract Value and Quarterback Pick#) are well correlated with future NFL performance.

|  |
| --- |
| **Pick# Pct Mobility**  **Passer Rating**  **Yards / Attempt**  **Did not have a career Did have a career**  Source: NFLCombineResults.com; ESPN.com self-made chart |
| 5. Figure: Box chart of successful and unsuccessful players. LHS: played less than 3 years; RHS: played at least 3 years All variables are standardized, therefore the vertical axis is interpreted as standard deviation |

These hypotheses are based on the descriptive statistics on Figure 4. Pick#, Pct (percentage of successful passes), Mobility, Passer Rating, Yards/Attempt (playing data is from College Years, athletic is from Combine) are better. Moreover, for players who did not have a successful NFL career all the metrics have a smaller quantile ranges. Having a career in the NFL means that the player had at least one game where he threw at least one pass. All variables are standardized for comparability.

# Methodology

Our goal is to explore what makes one quarterback better than another, therefore we are interested in the preference relations of the scouts. In consequence, we are interested in the ranking of the quarterbacks with respect to each other.

To test our models, we compare the ranking produced by our model and the actual result of the Draft. There are two ways to do that: the first metric we use is Spearman’s Rank Correlation, and the other one is called our Preference Congruence metric.

## Spearman’s Rank Correlation

Spearman’s rank correlation coefficient is best understood as “how well can we explain the relationship between two set of variables using monotonic functions”. In our case, 78,9% means that the relationship between Pick# and our explanatory variables is explained in 80% using a monotonic function.

## Preference Congruence metric

However, if we further think about what preference-relations are, then we realize that Spearman’s is not what we are primarily interested in. A set of preference relations is a function which maps ordered pairs of quarterbacks to relations:

What we are really interested in is how much of these do we get right compared to a random ordering. For this purpose we construct a matrix whose entries are 1, if (scouts prefer to ), and -1 otherwise. Of course for all . Our model gives us , and our Preference Congruence metric will be

The interpretation of this metric is “if we pick two quarterbacks at random, then what is the probability that we can determine the one that is preferred by the scouts”.

If the corresponding elements in the matrix are the same, , if they are different, then . So is the number of decisions we got wrong. There are entries in our matrices, but there are elements in the diagonals, which are 0. So our metric indeed gives us the aforementioned probability.

For randomly generated lists, .

# Descriptive Analysis

Supporting our hypotheses, we are presenting the following descriptive statistics.

## Athletic performance

One can observe on the scatterplot that the highest paid quarterbacks usually have average running performance. This supports the hypothesis that running faster alone does not make you a better quarterback, as it can be seen in Figure 7.

On the other hand, the scatterplot of Mobility and the natural log of Contract Value indicate a positive relationship between the two variables, see Figure 6.

Source: NFLCombineResults.com; Yahoo! Sports self-made chart

|  |  |
| --- | --- |
| C:\Users\KapronczayMor\Documents\Rajk\Öko2015osz_2016tavasz\2felev\beszamolo\c_mobility.png | C:\Users\KapronczayMor\Documents\Rajk\Öko2015osz_2016tavasz\2felev\beszamolo\c_40yard.png |
| 6. Figure: The relation of mobility with the logarithm of contract value. | 7. Figure: The relation of running performance with the logarithm of contract value. |

If we take a look at the outliers on these plots, we can observe the best quarterbacks: those who have a contract value significantly above the median. It is worth noting that there is variance in mobility and speed among the top players as well, which means that there is no ultimate quarterback: some stand out from the crowd with their mobility, some with their passing skills and so on.

## Performance in the NCAA

Performance in College football is not only determined by the individual, but is rather heavily influenced by the team. With this assumption, it is still a surprising finding, that there is no clear direct relationship between college statistics and Draft rank. Figure 9 shows the effect of Interceptions per Attempt on Draft rank. It shows no clear effect, except for the fact that crossing the assumed thresholds of 6% is seemingly heavily penalized by scouts. The same is true for Passer Rating, see Figure 8. All in all, the most valuable quarterbacks turn out to have average passing statistics.

|  |  |
| --- | --- |
| C:\Users\KapronczayMor\Documents\Rajk\Öko2015osz_2016tavasz\2felev\beszamolo\RAT2_player.png | C:\Users\KapronczayMor\Documents\Rajk\Öko2015osz_2016tavasz\2felev\beszamolo\int_pl.png |
| 8. Figure: The effect of Passer Rating on Draft rank. | 9. Figure: The effect of interceptions on Draft rank. |

|  |  |
| --- | --- |
| 1. Table: Draftees and Careers | |
| **Number of Draftees** | **Number of Colleges** |
| 1 | 47 |
| 2 | 11 |
| 3 | 2 |
| 4 | 1 |

In our view, the explanation of this phenomena is twofold. First, teams draft from many different colleges, and there is no available statistical data on the effect of team efficiency, like in the NFL. On Table 1 it is demonstrated that in the observed time period the majority of colleges contributed just one quarterback to the Draft. There are some “quarterback star maker” colleges, which contribute more than one quarterback during the observed 5 years. We assume, that the scouts have firsthand information about the coaches and schemes of NCAA teams, therefore they can weight college statistics the adequate way.

Second, NFL teams are purchasing an asset for their playing scheme, not an already made quarterback with his own style. Therefore, the performance of a certain player in a potentially different game environment does not determine performance in a future NFL environment.

Source:; TheFootballDatabase, ESPN.com self-made table

# Results

## Draft

### Preference ordering of scouts

Our first task was to explore the preference ordering of drafters. It is important to stress, that at this stage we do not seek to objectively evaluate quarterback performance (that would be based on later NFL data), but we want to understand how teams pick players: which are the most important factors for a good quarterback in the eyes of a scout.

We were particularly interested in the role of mobility, which is supposed to be a better explanatory variable for the draft than the linear combination of Weight and the time of the 40 Yard Dash.

|  |
| --- |
| C:\Users\KapronczayMor\Documents\Rajk\Öko2015osz_2016tavasz\2felev\beszamolo\NoLNCV.png  Source: NFLCombineResults.com; Yahoo! Sports self-made chart |
| 10. Figure: The relationship of Contract Value and Pick |

Table 2 summarizes the regressions we ran on the data. Every estimation is an OLS regression on the quarterbacks. In this table it is also tested, that Mobility is a better measure, than linear combination of Weight and time of the 40 Yard dash. (See Column 3 for example)

The explained variable is either the rank in which the quarterback is picked in the draft (in terms of quarterback picks), or the natural log of the players’ Contract Value. In the first case, qualities which are valued by scouts should have a negative coefficient, while in the second case the same coefficients should be positive. Our model is summarized by the following equation:

Before the regression results, it is also important to stress the relationship of the two success proxies. Figure 10 presents the observation on Contract Value and Quarterback Pick# (No. stands for quarterback rank in our database). It is clear from the chart, that early picks earn multiple times more than the average quarterback. But the relationship is not straight forward, signs of other aspects of contracts can also be seen from the figure.

Regarding Table 2, our main regression is in column 3, and the others act as robustness checks. First, we only include Mobility as explanatory variable, then we involve college statistics and finally year dummies which control for the changes of the pool. The first 5 columns the explained variable is the quarterback rank, while the last 5 columns try to explain the variation in Contract Value. These regressions are included as robustness checks for functional form. This is possible, since the relationship of the two explained variables is non-linear.

Our initial hypothesis, that mobility is an important factor for scouts, is confirmed by the robustness checks. Our second finding is that including Interceptions, Completion and Yards per Attempt is a better strategy than including Passer Rating. This is based on the observation that the first set of controls achieve a higher adjusted R-squared in the model.

Our third observation is that Interceptions are heavily penalized in quarterback rank, but is not significant when we try to explain Contract Value.

Altogether our best model explains about 40% of the variation in Pick and 32% in Contract Value. This concludes our exploration of the preference ordering of scouts.

2. Table: Relation of Draft and performance



### Predicting draft results from combine and NCAA data

To obtain an estimation on draft rank, we used our best regression from Table 2, the regression marked with a 3. We identificated from this model the Quarterback Pick# variable, and the natural logarithm of Contract Value. Then between the rank of the actual variables, and the predicted ones Spearman’s ρ can be calculated. Table 3 shows the results on different predicted variables in different years. Columns marked with either “Regression” or “Poisson” signs if the predicted value is based on a simple OLS, or a Poisson regression.

First of all, models based on salary are only slightly better than random guessing. However, our Preference Congruence metric indicates that our best models can explain 3 out of 4 preference relations on average.

3. Table: Predictive performance of our model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Preference Predictions** | **No.** | | | | **LNCV** | |
| **PC** | **Spearman** | **PC** | **Spearman** | **PC** | **Spearman** |
| Regression | | Poisson | |
| **2005** | 79% | 79% | 80% | 79% | 76% | 75% |
| **2006** | 73% | 67% | 70% | 60% | 62% | 34% |
| **2007** | 73% | 60% | 69% | 52% | 56% | 20% |
| **2008** | 66% | 40% | 63% | 31% | 59% | 27% |
| **2009** | 76% | 67% | 74% | 66% | 65% | 41% |

Where does the rest of the variation come from? First, the “gut-feeling” of scouts. Evidence of this can be found on scouting webpages, where besides Combine results one can often find qualitative comments such as “Threw nice in practice”. Second, college results are heavily influenced by respective teams and opponents, and there is no reliable statistical data available on that matter.

## NFL

Our first estimation on NFL performance of drafted quarterbacks concerns whether they get to play in the league after the draft. For this, we constructed a dummy variable, which is 1 if a drafted player played at least 3 seasons in the NFL and 0 otherwise.

We used a Linear Probability model, a probit and a logit regression to estimate the effects of various factors on the probability of having an NFL career. Out is these factors, the only significant one is the quality of the team they got drafted into. The mechanism is simple: **if a good team drafts you, you have a lesser chance to be a starter**.

Our equation for the model is:

4. Table: Probability of having a career

|  |  |  |  |
| --- | --- | --- | --- |
|  | **(1)** | **(2)** | **(3)** |
| **VARIABLES** | **Probit** | **Linear Probability** | **Logit** |
|  |  |  |  |
| **Passer Rating** | -0.00591 | -0.00191 | -0.00919 |
|  | (0.0130) | (0.00430) | (0.0219) |
| **Contract Value (ln)** | **1.442\*** | **0.137\*\*** | **2.595\*** |
|  | (0.845) | (0.0641) | (1.510) |
| **Draft rank** | 0.106 | 0.00585 | 0.186 |
|  | (0.106) | (0.0268) | (0.183) |
| **Mobility** | 0.202 | 0.0324 | 0.331 |
|  | (0.133) | (0.0299) | (0.229) |
| **Drafting team** | **-3.738\*\*** | **-0.831\*** | **-6.334\*\*** |
|  | (1.897) | (0.448) | (3.198) |
| **Constant** | -29.61\*\* | -2.606 | -52.02\* |
|  | (15.08) | (2.091) | (27.34) |
|  |  |  |  |
| **Observations** | 60 | 60 | 60 |
| **R-squared** |  | 0.347 |  |
| Standard errors in parentheses | |  |  |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | |  |  |

In our second estimation, we tried to capture Total Quarterback Rating in terms of data available at the time of the Draft. The most important explanatory variables are the natural log of Contract value, Quarterback Pick# and Passer Rating in college. This strategy is employed to test the effectiveness of the Draft as a market. In other words, the significance of Contract Value or Quarterback Pick# would indicate that the draft is an effective market, and their insignificance would mean that scouts are not able to tell the difference between future superstars and future benchwarmers.

Our control variables include year dummies and Team Offensive Efficiency. Since Total Quarterback Rating is already adjusted for the team, we expect the latter variable to be statistically insignificant. The model is described by the equation:

Results of the estimations shed light on some interesting facts. First, quality proxies determined at the Draft by scouts are statistically significant throughout our estimations. This confirms our initial hypothesis that scouts assess player quality well. Second, college Passer Rating is not a significant explanatory of Total Quarterback Rating in the NFL, and that confirms our hypothesis that college performance is not carried over to the NFL.

5. Table: NFL Performance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VARIABLES** | (1) | (2) | (3) | (4) |
| **QBRAT with Pick#** | **QBRAT with Contract Value (ln)** | **QBRAT with Quarterback pick#** | **QBRAT with All** |
|  |  |  |  |  |
| **Pick#** | -0.0987\*\*\* |  |  |  |
|  | (0.0242) |  |  |  |
| **Experience** | 5.759\*\* | 4.857\* | 5.663\*\* | 4.943\* |
|  | (2.802) | (2.796) | (2.818) | (2.828) |
| **Experience2** | -0.527\* | -0.437 | -0.518\* | -0.447 |
|  | (0.286) | (0.280) | (0.291) | (0.286) |
| **Passer Rating** | -0.107 | -0.0917 | -0.135 | -0.113 |
|  | (0.124) | (0.117) | (0.129) | (0.118) |
| **Team Efficiency** | 14.01 | 11.47 | 13.90 | 12.04 |
|  | (17.59) | (17.94) | (17.53) | (18.22) |
| **Contract Value (ln)** |  | 6.312\*\*\* |  | 5.235\*\* |
|  |  | (1.378) |  | (2.320) |
| **Quarterback pick#** |  |  | -2.024\*\*\* | -0.513 |
|  |  |  | (0.514) | (0.807) |
| **Constant** | 89.37\*\*\* | -17.33 | 96.36\*\*\* | 5.442 |
|  | (21.02) | (19.07) | (22.42) | (44.34) |
| **Observations** | 235 | 235 | 235 | 235 |
| **Number of id** | 44 | 44 | 44 | 44 |
| **Standard errors in parentheses** | |  |  |  |
| **\*\*\* p<0.01, \*\* p<0.05, \* p<0.1** | |  |  |  |

# Concluding remarks and further research

We analyzed a dataset of every drafted quarterback between 2005 and 2009. Our research supports our initial assumption that the Draft is a highly efficient market, where players’ publicly available data serves as a basis for scout choice. Out best models could predict 75% of preference relations. We performed a number of robustness checks, and this result was firm.

Regarding our questions about the relation of the Draft and NFL performance, we found that a number of other factors define NFL appearance, such as the team which drafts the player, and probably other non measureable factors, and not the quality / ability of the player. We implicitly assume, that the main factors which determine later NFL success are the players intrinsic properties, such as discipline, mental toughness and attitude, i.e. towards work.[[3]](#footnote-3) Further research concerning this relation can be an interesting topic.

On the other hand, we confirmed that scouts do a good job: their evaluation of the quarterbacks at the draft is highly significant when explaining future performance of players who got a chance to play in the league. The utility function of scouts can almost perfectly be replicated using metrics concerning abilities of players (40 yard dash, Wonderlic test, etc.), and interestingly college football statistics are almost irrelevant. This supports our view, that teams look for assets to build in their playing schemes, therefore abilities matter the most and the past performance of these “assets” in different environments does not count.

# References

*DraftHistory.com.* (s.a). Forrás: Quarterbacks: http://www.drafthistory.com/index.php/positions/qb

Eckstein, N. (2009). *http://economics.uoregon.edu/.* Forrás: A Statistical Analysis of American Football Running Backs' Draft Position and Subsequent Earnings: http://economics.uoregon.edu/wp-content/uploads/sites/4/2014/07/football.pdf

ESPN. (2011. 8 4). *Guide to Total Quarterback Rating*. Forrás: espn.go.com: http://espn.go.com/nfl/story/\_/id/6833215/explaining-statistics-total-quarterback-rating

FootballOutsiders. (2016. 4 1). *Team Efficiency Ratings*. Forrás: Footballoutsiders: http://www.footballoutsiders.com/stats/teameff

Lewis, M. (2003). *Moneyball: The Art of Wininng an Unfair Game.* New York: W.W. Norton.

*NFL.com*. (s.a.). Forrás: The Rules of the Draft: http://operations.nfl.com/the-players/the-nfl-draft/the-rules-of-the-draft/

NFLCombineResults.com. (2005 - 2009). *NFL Combine Results*. Forrás: Historical NFL Scouting Combine Data: http://nflcombineresults.com/nflcombinedata.php?year=2010&pos=QB&college=

*overthecap.com*. (s.a.). Forrás: Salary Cap Space: http://overthecap.com/salary-cap-space/

sportrac.com. (2005 - 2009). *NFL Draft Tracker*. Forrás: Sportrac: http://www.spotrac.com/nfl/draft/2005/quarterback/

*sports-reference.com.* (s.a.). Forrás: Sports-Reference: http://www.sports-reference.com/cfb/schools

*The Football Database.* (s.a.). Forrás: NFL Database: http://www.footballdb.com/teams/nfl/

*Yahoo! Sports.* (s.a). Forrás: NFL Stats: https://sports.yahoo.com/nfl/

*Every statistical database was accessed between 2016.05.25 -30 multiple times.*

# List of Figures:

[1. Figure: The Draft is a multi-million dollar market (Quarterback median and total contract value) 1](#_Toc454296155)

[3. Figure: Half of drafted quarterbacks donot play in the NFL 4](file:///C:\Users\karpatika\Documents\RAJK\ECONOMETRICS\BESZÁMOLÓ\Ökonometria16tavasz_KapronczayKárpáti_ujrairt_21_17.docx#_Toc454296156)

[2. Figure: Distribution of Mobility 4](file:///C:\Users\karpatika\Documents\RAJK\ECONOMETRICS\BESZÁMOLÓ\Ökonometria16tavasz_KapronczayKárpáti_ujrairt_21_17.docx#_Toc454296157)

[4. Figure: Distribution of weight through the whole sample. 7](#_Toc454296158)

[5. Figure: Box chart of successful and unsuccessful players. LHS: played less than 3 years; RHS: played at least 3 years 9](#_Toc454296159)

[6. Figure: The relation of mobility with the logarithm of contract value. 11](#_Toc454296160)

[7. Figure: The relation of running performance with the logarithm of contract value. 11](#_Toc454296161)

[8. Figure: The effect of Passer Rating on Draft rank. 12](#_Toc454296162)

[9. Figure: The effect of interceptions on Draft rank. 12](#_Toc454296163)

[10. Figure: The relationship of Contract Value and Pick 13](#_Toc454296164)

# List of Tables:

[1. Table: Draftees and Careers 10](#_Toc453697300)

[2. Table: Relation of Draft and performance 13](#_Toc453697301)

[3. Table: Predictive performance of our model 15](#_Toc453697302)

[4. Table: Probability of having a career 16](#_Toc453697303)

[5. Table: NFL Performance 17](#_Toc453697304)

1. Teams do not have exactly the same amount of money, every team’s salary cap is base salary cap after corrections. [↑](#footnote-ref-1)
2. NCAA stands for National College Athletic Association. [↑](#footnote-ref-2)
3. One of the most successful NFL head coaches, Bill Belichick of New England Patriots has a famous advise for players: „Do your job!” [↑](#footnote-ref-3)