An Examination of the National Football League Combine Statistics and Draft Stock - Draft

Trevor R. Jones*

May 9, 2020

Abstract

This paper will examine the impact of measurements taken at the NFL combine on individual athletes' potential draft stock, specifically focusing on the round and overall pick based on each position subgroup. To do this, two approaches are applied. The first is a robust linear regression that predicts draft stock with a player's age, weight, height, 40-yard dash time, 3-cone drill time, 20-yard shuttle time, vertical jump height, and broad jump distance. In the second model, I created a tree model with the regression listed above with the goal of predicting an athlete's draft stock. Results in Progress. . .

*Department of Economics, University of Oklahoma. E-mail address: trevorjones@ou.edu

1 Introduction

Following their final college football season, concluding sometime between December and January, prospective professional players trade their shoulder pads and a helmet for shorts and a t-shirt to train for the National Football League combine, with the goal of outperforming other prospects at their position. If players receive an invite to the NFL Combine, they spend several days undergoing a series of physical tests, drills, and measurements in pair with interviews with the staff of interested football clubs.

These physical tests and measurements receive a plethora of media attention, as many players climb up or fall down the draft boards of teams and sports analysts alike based on their performance. Some players choose to opt out of individual tests, but the vast majority of athletes participate, hoping to increase their stock and signing bonus come draft day. The physical tests include two speed-based tests in the 40-yard dash and a 20-yard shuttle, an agility-focused test in the 3-cone drill, two "explosive" measurements in the vertical and broad jump, and one strength-based test in the number of repetitions on the bench press with 225 pounds. Further, teams also receive official measurements of the player's height, weight, wingspan, and hand size. Athletes also undergo a series of position-specific drills, but these are more qualitatively measured than quantitatively, so they are left out of this analysis.

2 Literature Review

1. The National Football League (NFL) Combine: Does Normalized Data Better Predict Performance in the NFL Draft (2010)

- Used data between 2005-2009 for 8 combined physical tests, only from players drafted
- Major finding: draft performance has little to do with draft success, as only straight sprints and jumping ability have any major success
- Suggests the NFL alter the combine toward measurements that are more important to teams and incorporate mental and technical skills into the combine

2. The National Football League Combine: A Reliable Predictor of Draft Status? (2003)

- Data from the 2000 NFL Combine for 9 performance tests
- Findings: combine can be used to accurately predict draft status of RB, WR, and DBs.

3. The National Football League Combine: Performance Differences Between Drafted and Non-drafted - Players Entering the 2004 and 2005 Drafts (2008)

- Data from 2004 and 2005 combines, divided players into skill players, big skill players, and linemen
- Skill position Drafted players outperformed undrafted players in the 40-yard dash, vertical jump, 3-cone,
- Linemen: drafted outperformed undrafted in the 40-yd dash, bench press, and 3-cone drill

4. Does the NFL Combine Really Matter (2006)

- Looks at data for QB, WR, and RB from 2000 to 2010
- Finds outstanding performance at the NFL Combine is not sufficient to guarantee one's success as a professional player

5. The NFL Combine as a Predictor of ON-Field Success (2017)

- A number of combined measures were statistically significantly associated with improved on-field NFL performance

- quarterbacks and running backs with faster forty-yard dash speeds were associated with better NFL rushing statistics

3 Data

The data in this project comes from official NFL combine statistics and the official NFL draft results between 2002 and 2022. The data for this project comes from the nflreadr package, specifically from its combine and draft picks library. The following data gives a base summary statistics from the data, will revise and add a definition table as well.

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
year	21	0	2012.0	6.1	2002.0	2012.0	2022.0
round	7	0	4.2	2.0	1.0	4.0	7.0
pick	262	0	127.9	73.9	1.0	128.0	262.0
age	10	4	22.5	0.9	20.0	23.0	28.0
allpro	8	0	0.1	0.4	0.0	0.0	7.0
seasons_started	18	0	1.8	2.8	0.0	0.0	17.0
w_av	117	9	16.2	20.1	−4.0	8.0	163.0
dr_av	110	15	12.8	17.0	−4.0	6.0	163.0
games	232	9	58.2	47.4	0.0	46.0	297.0
wt	198	18	245.3	45.5	155.0	236.0	384.0
forty	141	22	4.7	0.3	4.2	4.7	5.7
bench	46	43	21.5	6.4	2.0	21.0	49.0
vertical	53	35	33.3	4.2	20.5	33.5	46.0
broad_jump	58	36	115.6	9.3	83.0	117.0	147.0
cone	197	48	7.2	0.4	6.4	7.2	9.0
shuttle	140	47	4.4	0.3	3.8	4.3	5.2

4 Empirical Methods

While this paper covers both linear regression and tree models, the primary regression equations can be depicted in the following equation (this is subject to revision).

The following models regress some the athlete's age, position group (dummy variable created for each position), weight, 40-yard dash time, vertical jump height, broad jump distance, 3-cone time, and shuttle time on a selected dependent variable. Using a linear regression model where Y is the dependent variable and X is all independent variables, the following models are obtained:

$$Pick = \beta_0 + \beta_1 Age + \beta_2 Position + \beta_3 Weight + \beta_4 Forty + \beta_5 Vertical + \beta_6 Broad + \beta_7 Cone + \beta_8 Shuttle + \varepsilon \tag{1}$$

Where Pick is a discrete variable for the number overall an athlete is selected.

$$Round = \beta_0 + \beta_1 Age + \beta_2 Position + \beta_3 Weight + \beta_4 Forty + \beta_5 Vertical + \beta_6 Broad + \beta_7 Cone + \beta_8 Shuttle + \varepsilon \tag{2}$$

Where Round is a discrete variable for the round of the draft an athlete is selected.

$$Drafted = \beta_0 + \beta_1 Age + \beta_2 Position + \beta_3 Weight + \beta_4 Forty + \beta_5 Vertical + \beta_6 Broad + \beta_7 Cone + \beta_8 Shuttle + \varepsilon \tag{3}$$

Where Drafted is a dummy variable that equals 1 if the athlete was drafted and 0 if the athlete was not drafted.

5 Research Findings

Here are some of the initial regression results for the first two equations. The tree model equations are still in progress and I am currently working on creating the full model.

6 Conclusion

	Model 1	Model 2
(Intercept)	−665.372***	−665.372***
	(100.889)	(100.889)
age	16.480***	16.480***
	(1.653)	(1.653)
posDE	−1.748	−1.748
	(14.584)	(14.584)
posT	−0.681	−0.681
	(12.758)	(12.758)
posDB	29.603**	29.603**
	(10.121)	(10.121)
posDT	73.763***	73.763***
	(21.057)	(21.057)
posWR	24.267*	24.267*
	(9.673)	(9.673)
posTE	54.429**	54.429**
	(16.573)	(16.573)
posRB	7.695	7.695
	(9.761)	(9.761)
posLB	35.988+	35.988+
	(20.253)	(20.253)
posC	19.619+	19.619+
	(10.922)	(10.922)
posFB	15.275	15.275
	(25.417)	(25.417)
posP	39.569**	39.569**
	(13.039)	(13.039)
posNT	30.140+	30.140+
	(15.918)	(15.918)
posOLB	28.227	28.227
	(32.845)	(32.845)
posCB	31.053*	31.053*
	(12.371)	(12.371)
posS	8.338	8.338
	(16.982)	(16.982)
posLS	1.550	1.550
	(9.869)	(9.869)
posSAF	30.702**	30.702**
	(11.119)	(11.119)
posOT	16.286	16.286
	(12.865)	(12.865)
wt	−1.297***	−1.297***
	(0.137)	(0.137)
ht5-11	1.842	1.842
	(7.699)	(7.699)
ht5-5	18.887	18.887
	(64.348)	(64.348)
ht5-6	−26.478	−26.478
	(64.127)	(64.127)
ht5-7	−20.242	−20.242
	(22.163)	(22.163)
ht5-8	10.308	10.308
	[illegible]	[illegible]