

Effect of homegrown players on professional sports teams

ISYE 2028

Rahul Patel 902949215

Problem Description:

Football is commonly referred to as America's favorite pastime. However, for thousands of people that work in the industry and the millions of fans who root for their teams every Thursday, Sunday, and Monday, winning seems to hold a degree of unparalleled importance in our hearts. Obviously, many things can affect whether a NFL team wins or loses a game, (e.g. coaches, injuries, weather, etc.) however one huge factor in determining a team's success is if a team has a talented roster. Often, teams have to make tough choices when constructing their rosters. For example, player contracts often have to be juggled due to the salary cap restrictions imposed on the league. Because of these salary cap restrictions, often, teams have to make decisions between keeping good players they have acquired in the draft, or picking up new, very high-value players who have recently become free agents. The opportunity of picking up very good free agents is that their contracts usually take up a large amount of cap space which prevents the team from signing other talented players. For my project I will be looking at the success of all 32 NFL teams from the past 4 seasons. I specifically wanted to look at the correlation between whether a team can experience a large amount of success while not signing large high-value free agents and relying more on their drafting process. This will provide insight on whether it's worth it for teams to spend large sums to attract free agents to their teams.

Data Source:

Information about each team's win/loss record and information about the team history of the roster is available on the Pro-Football-Reference.com. The data is available in a .csv format where each column represents the name of the player on the roster, their position, their age, their height/weight, and the team they were drafted by.

Method:

This study will analyze the correlation between the number of players on a team's roster that were drafted by that same organization (known from here on out as the "draft ratio") and each team's successes from the past 4 seasons. The criteria for success will be number of wins a

team had in a season. A team will play 16 regular season games and if they have a record good enough to qualify them for the playoffs they can potentially have 4 more wins (4 rounds in the playoffs) for a max of 20 wins possible in one season.

1. Compiling and organizing the Data

In order to figure out the draft ratios we want to parse through each of the 128 .csv files (32 teams for 4 years), I composed a Python script that would open each .csv file, determine which team the file is for, determine what year the file is for, and then count how many players have been drafted by the aforementioned team. The script in its final form looks like this:

```
import csv

teams=["was","oti","tam","ram","sea","sfo","sdg","pit","nor","phi","rai","nyj","nyg","nwe","min","mia",
,"kan","jax","clt","htx","gnb","det","den","dal","cle","cin","chi","car","buf","rav","crd","atl"]
teamsearch=["Redskins","Titans","Tampa","Rams","Seattle","Francisco","Chargers",
"Steelers","Saints","Eagles","Raiders","Jets","Giants","Patriots","Vikings","Dolphins",
"Chiefs","Jaguars","Colts","Texans","Packers","Lions","Broncos","Cowboys","Browns",
"Bengals","Bears","Panthers","Bills","Ravens","Cardinals","Falcons"]
years=["2011","2012","2013","2014"]

for team in teams:
    for year in years:
        filename="teams_" + team + "_" + year + "_roster_games_played_team.csv"
        ##All the files follow the above nomenclature using "team" and "year" as identifiers where
        "team" is an item in the list of teams and "year" is an item in the list of years.
        f=open(filename, "r")
        reader=csv.reader(f)
        a=teams.index(team)
        teamname=teamsearch[a]
        data=list(reader)
        roster=int(len(data))-3
        f.seek(0)
        stuff=f.read()
        drafted=stuff.count(teamname)
        print(team,year)
        print(drafted)
        print(roster)
        DraftPercent=drafted/roster
        print(DraftPercent)
```

The output for this script is the team abbreviation, the year of the team and the Draft Percentage. These values were then organized into a spreadsheet as seen below:

	Draft % 2014/20	Draft % 2013/20	Draft % 2012/20	Draft % 2011/2012		Win Count 2014/	Win Count 2013/	Win Count 2012/	Win Count 2011/2012
Arizona Cardinals	0.3870967742	0.3833333333	0.4516129032	0.3965517241	Arizona Cardinals	11	10	5	8
Atlanta Falcons	0.4590163934	0.593220339	0.5789473684	0.5614035088	Atlanta Falcons	6	4	15	10
Baltimore Ravens	0.4714285714	0.5423728814	0.55	0.5	Baltimore Ravens	11	8	14	13
Buffalo Bills	0.4237288136	0.4	0.4	0.3692307692	Buffalo Bills	9	6	6	6
Carolina Panthers	0.3181818182	0.435483871	0.4393939394	0.4626865672	Carolina Panthers	8	13	7	6
Chicago Bears	0.3055555556	0.3548387097	0.393442623	0.3650793651	Chicago Bears	5	8	10	8
Cincinnati Bengals	0.55	0.4769230769	0.4603174603	0.4406779661	Cincinnati Bengals	10	11	10	9
Cleveland Browns	0.3088235294	0.375	0.4918032787	0.3492063492	Cleveland Browns	7	4	5	4
Dallas Cowboys	0.393442623	0.3188405797	0.3731343284	0.4761904762	Dallas Cowboys	13	8	8	8
Denver Broncos	0.4210526316	0.3965517241	0.4237288136	0.3965517241	Denver Broncos	13	16	14	9
Detroit Lions	0.4126984127	0.3846153846	0.4117647059	0.3387096774	Detroit Lions	11	7	4	10
Green Bay Packers	0.625	0.5538461538	0.5	0.6181818182	Green Bay Packers	14	8	12	16
Houston Texans	0.4426229508	0.3768115942	0.4833333333	0.4615384615	Houston Texans	9	2	13	11
Indianapolis Colts	0.253968254	0.3150684932	0.3235294118	0.4305555556	Indianapolis Colts	13	12	11	2
Jacksonville Jaguars	0.3636363636	0.3188405797	0.3421052632	0.3289473684	Jacksonville Jaguars	3	4	2	5
Kansas City Chiefs	0.344222951	0.4677419355	0.4761904762	0.4642857143	Kansas City Chiefs	9	11	2	7
Miami Dolphins	0.4117647059	0.4545454545	0.3692307692	0.3968253968	Miami Dolphins	8	8	7	6
Minnesota Vikings	0.5245901639	0.5254237288	0.5660377358	0.4406779661	Minnesota Vikings	7	5	10	3
New England Patriots	0.4696969697	0.4307692308	0.4626865672	0.4225352113	New England Patriots	16	14	14	16
New Orleans Saints	0.5087719298	0.3333333333	0.3166666667	0.4210526316	New Orleans Saints	7	12	7	14
New York Giants	0.347826087	0.46875	0.5	0.4166666667	New York Giants	6	7	9	7
New York Jets	0.3968253968	0.3064516129	0.3768115942	0.3220338983	New York Jets	4	8	6	8
Oakland Raiders	0.3939393939	0.3650793651	0.4032258065	0.4375	Oakland Raiders	3	4	4	8
Philadelphia Eagles	0.4655172414	0.4655172414	0.4848484848	0.5087719298	Philadelphia Eagles	10	10	4	8
Pittsburgh Steelers	0.5357142857	0.5238095238	0.5573770492	0.6	Pittsburgh Steelers	11	8	8	12
San Diego Chargers	0.3538461538	0.3870967742	0.4153846154	0.453125	San Diego Chargers	9	10	7	8
San Francisco 49ers	0.4696969697	0.45	0.5357142857	0.6111111111	San Francisco 49ers	8	14	14	15
Seattle Seahawks	0.3714285714	0.3692307692	0.4745762712	0.3714285714	Seattle Seahawks	15	17	12	7
St. Louis Rams	0.4406779661	0.3965517241	0.3389830508	0.2816901408	St. Louis Rams	6	7	7	2
Tampa Bay Buccaneers	0.2857142857	0.359375	0.4328358209	0.4393939394	Tampa Bay Buccaneers	2	4	7	4
Tennessee Titans	0.4	0.5081967213	0.5384615385	0.5740740741	Tennessee Titans	2	7	6	9
Washington Redskins	0.3661971831	0.4098360656	0.3787878788	0.3492063492	Washington Redskins	4	3	10	5

The values on the right side of the table are the calculated win counts of the teams in each season. This information was also available on Pro-Football-Reference.com.

2. Graphing the data and observing a correlation

I'll be graphing the variables using the draft ratio as my x and the win count as my y. Hopefully, I'd like to show there is a correlation between the two variables.

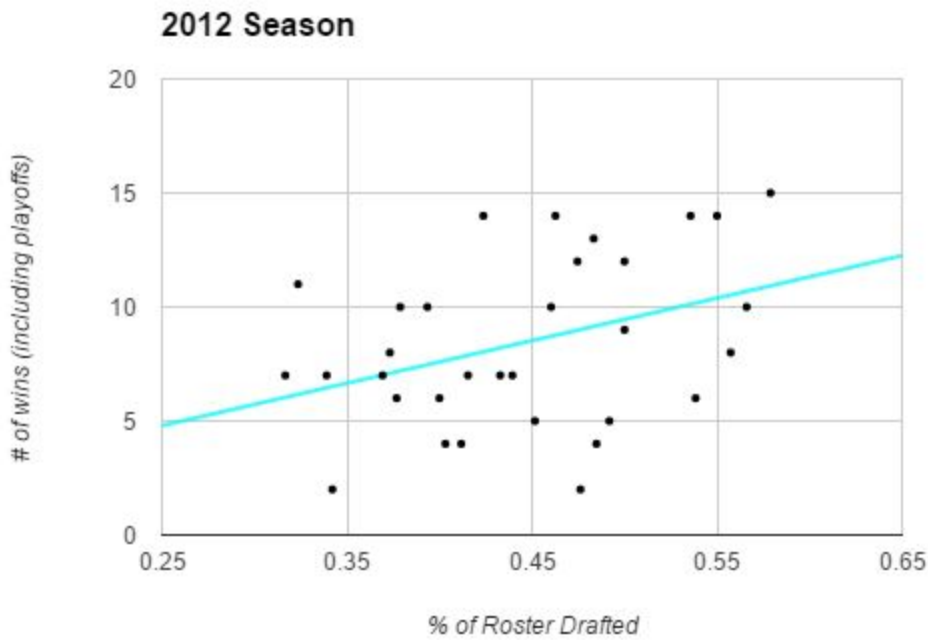
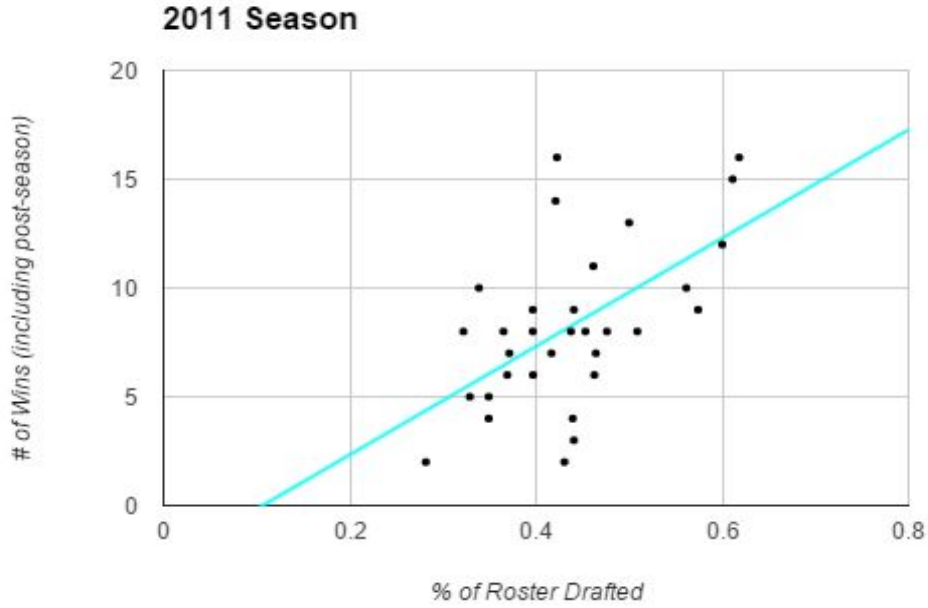
The regression line can be expressed in terms of $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}$$

where $\bar{y} = (1/n) \sum_{i=1}^n y_i$ and $\bar{x} = (1/n) \sum_{i=1}^n x_i$.

Using the data acquired we can visualize these results in the form of a fitted regression line over a scatter plot of our data from the past 4 seasons.



Using the regression analysis equations above the fitted regression lines

$$2011: \hat{y} = 24.9x - 2.649$$

$$2012: \hat{y} = 18.647x + 0.133$$

$$2013: \hat{y} = -2.878x + 9.647$$

$$2014: \hat{y} = 14.124x + 2.601$$

For all of the seasons bar 2013 there is a surprisingly high correlation between the draft ratio and the # of wins. While this is a relatively small sample size (only 4 seasons), we can notice the slope for the regression lines 3 of the 4 instances are positive. Initially, this would indicate that the higher draft ratios would mean that teams had a greater chance of success but let's analyze this further.

3. Further Regression Analysis.

We want to analyze the strength of the regression using the sample variance and mean standard error to make inferences about our correlation. We calculate this using an unbiased estimator for population variance as so:

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \text{ where } s^2 \text{ is the sample variance.}$$

The sample variance tells us how spread out our data is in our model.

Now we calculate an unbiased estimator of the variance ($\hat{\sigma}^2$) of our regression model also known as the mean squared error. The equation for this is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \text{ where } y_i - \hat{y}_i \text{ is the error of our regression model for point } i.$$

These values have been calculated for all 4 seasons below.

2011 season:

Draft % 2011/2012	Win Count 2011/2012	y(hat)	(Y-Y(average))^2	Sample Variance	y(average)	(Y-Y(hat))^2	Mean Squared Error
0.3965517241	8	7.225137931	0.0625	14	8.25	0.6004112259	9.694998201
0.5614035088	10	11.32994737	3.0625			1.768760003	
0.5	13	9.801	22.5625			10.233601	
0.3692307692	6	6.544846154	5.0625			0.2968573314	
0.4626865672	6	8.871895522	5.0625			8.247783892	
0.3650793651	8	6.44147619	0.0625			2.428996465	
0.4406779661	9	8.323881356	0.5625			0.4571364209	
0.3492063492	4	6.046238095	18.0625			4.187090342	
0.4761904762	8	9.208142857	0.0625			1.459609163	
0.3965517241	9	7.225137931	0.5625			3.150135364	
0.3387096774	10	5.784870968	3.0625			17.76731276	
0.6181818182	16	12.74372727	60.0625			10.60331207	
0.4615384615	11	8.843307692	7.5625			4.65132171	
0.4305555556	2	8.071833333	39.0625			36.86716003	
0.3289473684	5	5.541789474	10.5625			0.2935358338	
0.4642857143	7	8.911714286	1.5625			3.65465151	
0.3968253968	6	7.231952381	5.0625			1.517706669	
0.4406779661	3	8.323881356	27.5625			28.34371269	
0.4225352113	16	7.872126761	60.0625			66.0623234	
0.4210526316	14	7.835210526	33.0625			38.00462925	
0.4166666667	7	7.726	1.5625			0.527076	
0.3220338983	8	5.369644068	0.0625			6.91877233	
0.4375	8	8.24475	0.0625			0.0599025625	
0.5087719298	8	10.01942105	0.0625			4.078061388	
0.6	12	12.291	14.0625			0.084681	
0.453125	8	8.6338125	0.0625			0.4017182852	
0.6111111111	15	12.56766667	45.5625			5.916245444	
0.3714285714	7	6.599571429	1.5625			0.1603430408	
0.2816901408	2	4.365084507	39.0625			5.593624725	
0.4393939394	4	8.291909091	18.0625			18.42048364	
0.5740740741	9	11.64544444	0.5625			6.998376309	
0.3492063492	5	6.046238095	10.5625			1.094614152	

2012 season:

Draft % 2012/2013	Win Count 2012/2013	y(hat)	(Y-Y(average))^2	Sample Variance	y(average)	(Y-Y(hat))^2	Mean Squared Error
0.4516129032	5	9.751225806	11.81640625	13.73790323	8.4375	22.57414666	13.76506189
0.5789473684	15	12.12563158	43.06640625			8.26199382	
0.55	14	11.58585	30.94140625			5.828120223	
0.4	6	8.7888	5.94140625			7.77740544	
0.4393939394	7	9.523378788	2.06640625			6.367440507	
0.393442623	10	8.66652459	2.44140625			1.778156669	
0.4603174603	10	9.913539683	2.44140625			0.007475386495	
0.4918032787	5	10.50065574	11.81640625			30.25721354	
0.3731343284	8	8.287835821	0.19140625			0.08284945979	
0.4237288136	14	9.231271186	30.94140625			22.7407745	
0.4117647059	4	9.008176471	19.69140625			25.08183156	
0.5	12	10.6535	12.69140625			1.81306225	
0.4833333333	13	10.34271667	20.81640625			7.061154714	
0.3235294118	11	7.362852941	6.56640625			13.22883873	
0.3421052632	2	7.709236842	41.44140625			32.59538532	
0.4761904762	2	10.20952381	41.44140625			67.39628118	
0.3692307692	7	8.215046154	2.06640625			1.476337156	
0.5660377358	10	11.88490566	2.44140625			3.552869349	
0.4626865672	14	9.957716418	30.94140625			16.34005656	
0.3166666667	7	7.234883333	2.06640625			0.05517018028	
0.5	9	10.6535	0.31640625			2.73406225	
0.3768115942	6	8.356405797	5.94140625			5.552648281	
0.4032258065	4	8.848951613	19.69140625			23.51233174	
0.4848484848	4	10.3709697	19.69140625			40.58925488	
0.5573770492	8	11.72340984	0.19140625			13.86378081	
0.4153846154	7	9.075676923	2.06640625			4.308434689	
0.5357142857	14	11.31946429	30.94140625			7.185271716	
0.4745762712	12	10.17942373	12.69140625			3.314497959	
0.3389830508	7	7.651016949	2.06640625			0.4238230681	
0.4328358209	7	9.401089552	2.06640625			5.765231038	
0.5384615385	6	11.37069231	5.94140625			28.84433586	
0.3787878788	10	8.393257576	2.44140625			2.581621218	

2013 season:

Draft % 2013/2014	Win Count 2013/2014	y(hat)	(Y-Y(average))^2	Sample Variance	y(average)	(Y-Y(hat))^2	Mean Squared Error
0.3833333333		10	8.543766667	14.70564516	8.4375	2.120615521	15.14626177
0.593220339		4	7.939711864			15.52132957	
0.5423728814		8	8.086050847			0.007404748348	
0.4		6	8.4958			6.22901764	
0.435483871		13	8.393677419			21.21820772	
0.3548387097		8	8.625774194			0.3915933413	
0.4769230769		11	8.274415385			7.428811496	
0.375		4	8.56775			20.86434006	
0.3188405797		8	8.729376812			0.5319905333	
0.3965517241		16	8.505724138			56.1641707	
0.3846153846		7	8.540076923			2.371836929	
0.5538461538		8	8.053030769			0.002812262485	
0.3768115942		2	8.562536232			43.06688179	
0.3150684932		12	8.740232877			10.6260817	
0.3188405797		4	8.729376812			22.36700503	
0.4677419355		11	8.30083871			7.285471671	
0.4545454545		8	8.338818182			0.1147977603	
0.5254237288		5	8.134830508			9.827162317	
0.4307692308		14	8.407246154			31.27889558	
0.3333333333		12	8.687666667			10.97155211	
0.46875		7	8.2979375			1.684641754	
0.3064516129		8	8.765032258			0.5852743559	
0.3650793651		4	8.596301587			21.12598828	
0.4655172414		10	8.307241379			2.865431748	
0.5238095238		8	8.13947619			0.01945360771	
0.3870967742		10	8.532935484			2.152278294	
0.45		14	8.3519			31.90103361	
0.3692307692		17	8.584353846			70.82310019	
0.3965517241		7	8.505724138			2.26720518	
0.359375		4	8.61271875			21.27717427	
0.5081967213		7	8.184409836			1.40282666	
0.4098360656		3	8.467491803			29.89346662	

2014 season:

Draft % 2014/2015	Win Count 2014/2015	y(hat)	(Y-Y(average))^2	Sample Variance	y(average)	(Y-Y(hat))^2	Mean Squared Error
0.3870967742	11	8.067967742	6.56640625	14.31854839	8.4375	8.596813162	13.39604413
0.4590163934	6	9.083688525	5.94140625			9.509134917	
0.4714285714	11	9.258985714	6.56640625			3.031130743	
0.4237288136	9	8.585322034	0.31640625			0.1719578156	
0.3181818182	8	7.094681818	0.19140625			0.8196010103	
0.3055555556	5	6.916361111	11.81640625			3.672439908	
0.55	10	10.36865	2.44140625			0.1359028225	
0.3088235294	7	6.962514706	2.06640625			0.001405147275	
0.393442623	13	8.157590164	20.81640625			23.44893302	
0.4210526316	13	8.547526316	20.81640625			19.82452191	
0.4126984127	11	8.429539683	6.56640625			6.607266244	
0.625	14	11.427875	30.94140625			6.615827016	
0.4426229508	9	8.852163934	0.31640625			0.02185550228	
0.253968254	13	6.187793651	20.81640625			46.40615534	
0.3636363636	3	7.736636364	29.56640625			22.43572404	
0.3442622951	9	7.463016393	0.31640625			2.362318607	
0.4117647059	8	8.416352941	0.19140625			0.1733497716	
0.5245901639	7	10.00978689	2.06640625			9.058817095	
0.4696969697	16	9.234530303	57.19140625			45.77158022	
0.5087719298	7	9.786385965	2.06640625			7.763946745	
0.347826087	6	7.513347826	5.94140625			2.290221643	
0.3968253968	4	8.205365079	19.69140625			17.68509545	
0.3939393939	3	8.164606061	29.56640625			26.67315576	
0.4655172414	10	9.1755	2.44140625			0.67980025	
0.5357142857	11	10.16689286	6.56640625			0.6940675115	
0.3538461538	9	7.598369231	0.31640625			1.964568813	
0.4696969697	8	9.234530303	0.19140625			1.524065069	
0.3714285714	15	7.846685714	43.06640625			51.16990527	
0.4406779661	6	8.824694915	5.94140625			7.978901364	
0.2857142857	2	6.636142857	41.44140625			21.49382059	
0.4	2	8.2502	41.44140625			39.06500004	
0.3661971831	4	7.772802817	19.69140625			14.2340411	

Analyzing the Results:

For 3 of the 4 seasons, regression analysis demonstrated a positive correlation between draft ratio and success, agreeing with the initial hypothesis that higher draft ratios correlated with higher success. While the relationship between MSE and Variance wasn't as clear cut in all the seasons, for the 2011 season it was clear that the MSE was a great deal smaller than the Variance meaning it draft ratio was a good indicator for success that season.

There are a few reasons why I think the data was not as consistent as I would have liked.

- 1) Sample size: The NFL season is relatively short compared to most other team sports (soccer, hockey, basketball, baseball, etc.) and the NFL typically sees more upsets than most other leagues. Maybe, the hypothesized trends would be more pronounced with an analysis of an 82 game NBA season versus a 17 game NFL season. Unfortunately it was difficult to acquire data from other leagues so I was restricted to using NFL data.
- 2) Injuries: NFL teams value certain players more than others (a Quarterback is typically more valuable to a team than a punter), so if a key player becomes injured for a team, then, a team that might have been predicted to have a high amount of success could

feel the impact of the injury more so than a team with a lower draft ratio but fewer injuries.

- 3) Coaching: NFL coaching staffs have a huge impact on the way a team plays, and often times coaching changes can force teams to make adaptations that might affect their performance during the season. Conversely, good NFL coaching staffs are typically able to do less with more, so to speak. An example of this would be the New England Patriots, who typically do not draft well but acquire undrafted free agents for low fees, but experience high degrees of success because of a great coaching staff (the Patriots had the best record in the NFL each of the 4 years I took data from but never had a draft ratio higher than .47 which is 0.05 above the average of .42).