Evaluating the Performance of NFL Contracts

Ian Gatlin
MIT
iagatlin@mit.edu

ABSTRACT

This project creates a metric called avg_diff that is used to compare the performance of a franchise's player contracts to other teams. This metric is a relationship between the amount of money a teams are spending on players and those players' performances based on PFF grade. This metric is calculated by finding the difference between how much a team is paying a player and the average salary other teams are paying players of similar performance. With this metric, we can see that winning teams perform better in free agency since there is a correlation that they have a lower avg_diff scores. This metric also allows NFL front offices to evaluate their contracts and can help them determine for which positions they need to reevaluate their scouting tactics.

1 INTRODUCTION

My project is about evaluating NFL player contracts. I did this by comparing a player's salary to other players' salaries who are of similar skill level. By doing this on a large scale, we can determine if a team is efficiently choosing players that are helping them win by creating a metric that tell us if a team underpays more often for players than other teams. This is important because NFL teams operate under a salary cap. Teams have a limited amount of money they can spend on their roster's payroll, so if you overpay a player who isn't delivering, you may not have the cap space to sign another expensive player.

The difference between a player's salary and the average salary of similar players was the metric I created to compare teams' contract performance. I made this calculation for every player that had a starting role and who wasn't on a rookie contract for every year between 2012 and 2021. Then, after grouping this metric by team, I could determine how a team's contracts are performing.

From Figure 1, we can see that higher performing receivers are more likely to be paid more. This correlation allows us to expect the average salary of higher performing players to be greater than the average salary of lower performing players.

My goal was to create a metric called 'avg_diff' that conveys whether a team normally underpays or overpays for players. To evaluate my metric, I wanted to achieve a correlation coefficient of around 0.43 for the relationship between avg_diff and the win percentage of NFL teams from 2012 to 2016. I felt this was a good goal since a study from UC

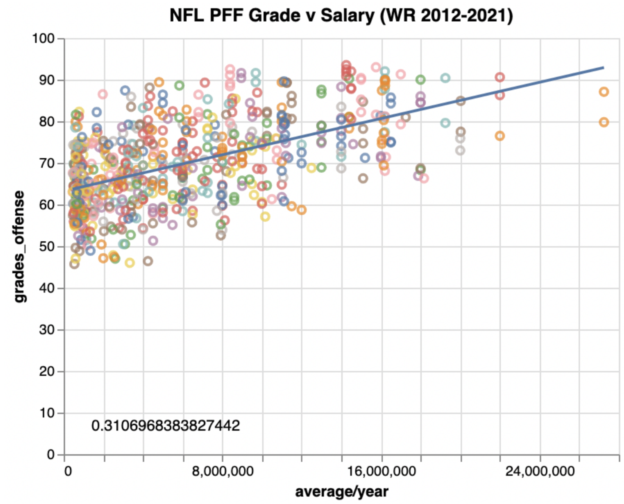

Figure 1: There is a positive correlation between PFF grade and salary for wide receivers. R^2 : 0.311

Berkeley developed a similar metric which had a correlation coefficient of 0.43 from 2011 to 2016[1].

The results of my project yielded a correlation coefficient of -0.091 from 2012 to 2016, and a correlation coefficient of -0.158 from 2012 to 2021.

Although I was not able to produce as convincing results similar to the study from Berkeley, we both came to the takeaway that for players of similar skill level, teams that win more pay less for players than teams that win less[1]. These trends are exciting because I was able to replicate them with a different data source. A strength of this project is that I was able to take a lot of data from a wide range of years and find similar trends for different year intervals. Also, the high quality of PFF data lets me evaluate performance at a very precise level.

A big weakness of this project is the reduced number data entries due to the omitting of rookie contracts. Since rookie contract values depend on draft position and do not relate to how teams structure contracts in free agency, I omitted them from the metric calculations. This greatly reduced the available data points when calculating avg_diff, and possibly skewed the metric towards teams with more rookie talent.

In the future, teams can use these metrics to reevaluate contract strategies, and see with which positions they are misevaluating talent.

2 BACKGROUND

The NFL has become a hotbed for data science applications. The two main applications of data science in the NFL are in-game decision making (eg. making the decision to go for it on fourth down) and game planning (eg. finding tendencies of opposing teams). Even though there are other problems data could solve in the NFL, these two fields have boomed in popularity because of their simplicity and effectiveness.

Outside of in-game decisions and game planning, I find talent acquisition to be the problem with the most potential. If a team could quickly and effectively evaluate talent for scheme, coaching, and positional fit, NFL front offices could take the guess work out of the draft and free agency and greatly improve their teams change for success.

Even though optimizing the draft and free agency with data science would be greatly beneficial, it is very hard to do in practice. Previous studies have found it to be very hard to use data from the draft combine and college football to predict career success before teams draft rookies[2, 3]. In free agency, teams currently spend hundreds of hours watching game film, interviewing players, and determining scheme fits for potential signees. It would be great to optimize this process, but the amount of variables in this process is very hard to model.

Therefore, I decided that evaluating how teams' contracts perform retroactively would be an attainable but still interesting project.

3 DATA USED

For this exploration, I needed a data source with NFL contract information, and a NFL performance database of some sort of metric. I also needed each team's win percentage for each year. I manually created my own table for win percentage, I used Spotrac.com to get contract data, and I chose Pro Football Focus (PFF) as the performance metric.

3.1 Spotrac

Spotrac is a sports contract website for all professional sports leagues. I scraped data from Spotrac's 'Contracts By Position' page, where I was able to get contract data from each position group by year in which the contracts were signed.

I took the player's name, signed age, year signed, contract duration, contract value, and the contract's average annual value from Spotrac.

Although Spotrac had all of the fields I needed, there were some problems with the data. I was able to overcome the problem of one player having multiple contracts on the same year by overwriting the least significant contract with the more valuable contract. Since I was scraping data year by year and contract extensions are independent of previous contract duration, I had to make sure I populated the tables in a way that would keep the most most recent contract's values. This problem was common in the data, and was most likely exacerbated by the NFL's franchise tag, a one year contract that teams can use so they have more time to negotiate a more lengthy extension with a player.

A second problem that I had to deal with was missing contract data. There were some years where players simply did not have contract data. I believe this problem originated from some contracts that were signed a year before they went into effect. There were some entries that claimed that a player had signed a 3 year contract, but who did not sign an extension until 4 years after that date. After attempting to cross check a handful of these entries with further internet research, I decided to omit players who were missing contract data for that year. I still kept entries of that player for years where there was contract data. This may have skewed some of the data, since this led to lucrative contracts being ignored if they were missing from Spotrac. However, this option was better than omitting a player from all the tables if a year of contract data was missing, as that choice would have significantly limited the data pool.

3.2 Pro Football Focus

Pro Football Focus is a football analytics company that grades players in the NFL and NCAA. PFF grades a player's game by grading every play on a scale of -2 to +2 and then scaling the overall performance to a 0-100 scale. Players are also assigned letter grades similar to the American high school grading scale. These grades are different from regular statistics since they are do not only depend on a numerical outcome (eg. yards gained), but also take into account the player's performance independent of how other player's may unfairly affect them. For example, a quarterback may throw a terrible pass that could have easily been intercepted, but the defender may have dropped the ball. This is an incompletion, which statistically is the same as throwing the pass away or having an offensive receiver drop the pass. But with PFF, this poor throw will count more negatively towards the quarterback's grade than a regular incompletion. Each position is graded on a different rubric and each play is graded by humans who watch every play from multiple angles.

I thought this data source would be the best to determine a player's performance because of its precision and lack of bias. The study from Berkeley used Madden ratings[1]. I found this decision poor because Madden ratings do not reflect a player's immediate impact on the field. A high Madden rating is accumulated over years of play at a high level, while PFF grades immediately evaluate player's contributions on the field. It was for this reason I chose PFF.

I was able to download CSV files straight from the PFF website. I took player name, position, team, snap count, and offensive grade or defensive grade.

It was important to take the snap count of each player because PFF grades independently of playing time. Players that only played 5 snaps in a season can have the same PFF grade as a regular starter although their contributions on the field as a whole are very different. I sorted and filtered by snap count during data cleaning to make sure that I was only taking grades from players who had played a substantial amount of time on the field.

4 MOTIVATION

In the NFL there is a strict cap space rule. A team's payroll must stay below a set monetary figure, or the team will face punishments from the league's front office. For reference the cap will be $\$208.2M$ in 2022. It is in the team's best interest to stay under the cap while signing players.

This project is important for optimizing the limited amount of money teams have. If a team signs a player for a lot of money and that player doesn't perform to expectations, the team is now left with less cap space with no output on the field. Although this project only compares teams' performances against each other, this work has the potential to allow teams to determine if they have a trend of misevaluating talent in certain positions so they can reevaluate their scouting techniques and optimize their cap.

Bad contracts can become very problematic for NFL teams who are struggling to sign new players. Since larger contracts in the NFL usually are around 4 years, teams can be stuck with dead cap for years after making a poor signing. This is why this project is so important.

5 DESIGN

The design of this project was mostly concerned with accurately joining the tables from PFF and Spotrac together, and then accurately grouping and aggregating the data.

5.1 Data Cleaning and Merging

First, I worked on cleaning the PFF data. I started with 50 tables separated by year and position (I used 7 positions in this project: OL, RB, WR, TE, DL, LB, DB. I omitted quarterbacks from the project because of the relatively small amount of big name free agent signings from the quarterback position. Also, 4 of the 7 position groups I used shared tables, leaving me with 5 tables for each of 10 years).

For each individual table, I searched for duplicate names. If there were duplicate names, I decided to keep the player PFF ranked highest. Then I took the top n players from each position group from each team. I did this by ranking a position group by snap count, grouping by team, and taking the top n players from that grouping. I took an arbitrary n because in the NFL a team normally will have a set group of starting players, but a larger array of players may get significant playing time. In this project, I took the amount of players I personally determined to normally receive a significant amount of playing time. Then I ranked all of the players in each position group with an integer based off of the PFF value for that year. I also had to clean variables, such as team names for franchises that have relocated in the past 10 years.

After cleaning the individual tables, I concatenated all of the PFF tables together by position. I now had 7 PFF tables separated by position.

Then I started working on the salary tables. I dropped duplicate names in each table here as well, and then I concatenated them by position. I had to clean some values to cast them as integers, and then I filtered out rookie contracts by removing all entries for contracts signed when a player was 23 years old or younger.

I also had to clean the names on the salary tables to make sure that all suffixes like 'Jr.' and 'II' matched on the salary and PFF tables.

Then I had solve one of the biggest challenges I faced doing this project. I had to convert the salary tables so each contract had an individual row for each year the contract was valid. I did this by creating a new table row by row from the previous table and looping over rows for the amount of years the contract was valid while incrementing the year column by 1. This was highly inefficient, but I could not find a way to use an existing pandas function.

It was then time to do an inner merge between the salary and PFF tables. At this point I had 7 tables in total.

Then I created another new table for each position and iterated over the rows again. This time, I added a new column that stored the average value of players with similar contracts. Since the tables were separated by position and they were ranked by PFF grade, I could simply use $.iloc[x : y, z].mean()$ to find the average salary of a range of players that have similar PFF grades.

I took the difference between a players actual salary, and the average salary of similarly skilled players. This difference it the metric 'diff' and once I found the average of a teams 'diff' value for a year, the metric becomes the 'avg_diff'. Because I subtracted the average salary from the actual salary, more negative values are better than more positive ones.

5.2 Data Aggregation

For each of the 7 tables for each position group, I grouped all of the 'diff' metrics by team and year. Now, I knew the average amount of money a team over or underpayed for each position for every year between 2012 and 2021. To find the overall amount a team over or underpaid, I would need to combine these 7 tables.

Because some positions get paid more on average than others, I standardized every positions avg_diff on a scale between -1 and 1 using the z-score method. I chose this method because I wanted the data to be centered around 0 so I could utilize negative values after standardizing.

To combine these standardized scores, I weighed each of the positions' avg_diff score depending on how many players I determined received a significant amount of playing time in games. I weighed them by the same n value when I took the top n players from each position for each team per year. Then I joined all of the position tables, and took the sum of their weighted standardizations. This yielded the average difference compared to other teams that each franchise was spending on contracts for players of similar skill between 2012 and 2021. Now, I can take aggregations and averages of different year ranges, and compare how teams' contracts have performed over time.

5.3 Visualizations

This data was now ready for visualization. There were also some opportunities to visualize some of the intermediate tables from the design process, but none of them heavily related to the problem I aimed to solve. Some of those tables yielded interesting data that could be used in future applications though.

I simply had to plot avg_diff against the win percentage for each team. Then I found the correlation coefficients and r^2 values for different year ranges using Altair.

6 EVALUATION

For this project I wanted to evaluate my results by comparing them to the FAR study from Berkeley. I wanted to see I if I could replicate a moderately strong or better correlation between avg_diff and win percentage.

The Berkeley students found a moderately strong correlation coefficient of 0.43 when graphing FAR against win percentage from 2011-2016[Figure 2][1]. When I did this process with avg_diff from 2012-2016, I found a very weak correlation coefficient of -0.091[Figure 3]. However, from 2012 to 2021, I found a slightly less weak correlation coefficient of -.158[Figure 4]. This is most likely the result of many weaknesses in the design that could be improved upon with further iteration.

Although these results are not the greatest, I still found a negative correlation between the two variables along three different year ranges. This negative correlation supports the conclusion made by the FAR study that winning is a factor in free agency performance[1].

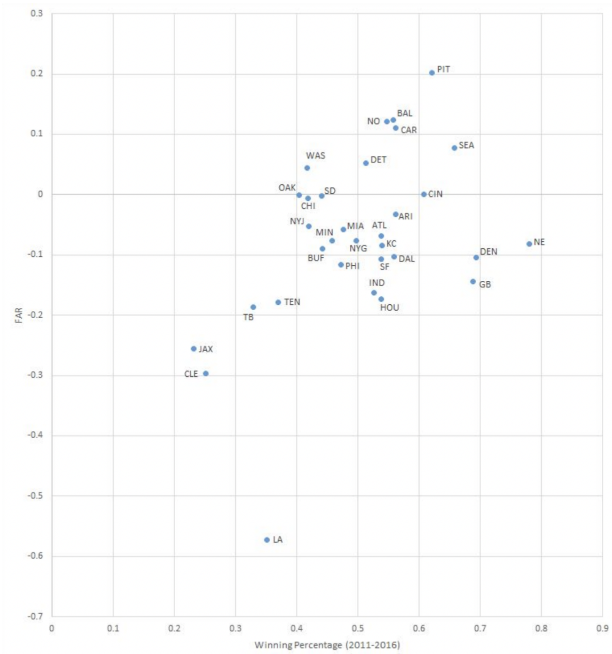

Figure 2: Graph plotting FAR and win percentage from 2011 to 2016. R : 0.43 **[1]**

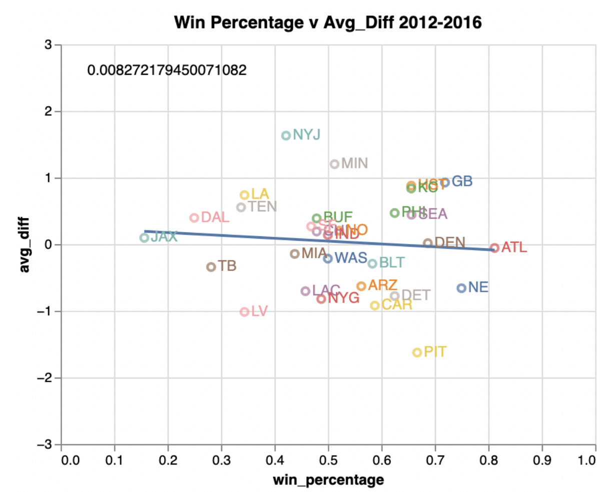

Figure 3: Graph plotting avg_diff and win percentage from 2012 to 2016. $R : -0.091$

7 RELATED WORK: FAR

I was inspired to do this project by a study from UC Berkeley where four students in the sports analytics group developed their own Free Agency Rating (FAR) metric. Instead of using PFF grades, they used Madden grades as their performance metric, but the general idea and design process was the same.

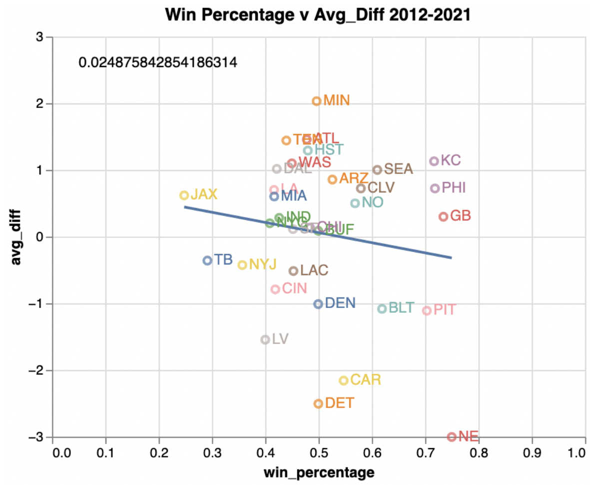

Figure 4: Graph plotting avg_diff and win percentage from 2012 to 2021. $R : -0.158$

Another big difference is that they also only took player performance metrics for years in which they signed the contract to calculate FAR, unlike avg_diff where I took performance metrics for every year of the contract. The study concluded that franchises that have a winning culture are more likely to succeed in free agency[1].

The study also explored relationships between FAR and other variables that may affect a team in free agency such as market size and weather, but found no significant correlations between the two other variables and FAR.

8 CONCLUSION

The main takeaway from this project is although the correlation was weak, teams who win more perform better in free agency.

The use of PFF and the range of years that was used in this project were highlights of my project's strengths. My project also can very quickly and easily calculate the avg_diff value for many teams over many different date ranges very quickly. However, there were many weaknesses in the project.

The first problem arose when filtering out rookies. Since I was evaluating contracts signed during free agency, it was important to remove young players. But since players on a rookie contract can make up a significant amount of a team's snaps, I lost a lot of rows when summing the weighted scores from each position group because a team didn't have enough non-rookie contract players in a certain position to yield a avg_diff value. I attempted mean value imputation on these null values by taking the mean of the avg_diff metric for other years where that position had a avg_diff metric, but the results of this did not yield reasonable results.

Another weakness stems from the arbitrary choice of the amount of n players I considered from each position group to contribute to the avg_diff. Different teams use different schemes that use different amounts of players from each position groups. Selecting a more relevant number of players from each position group for each team would greatly increase the accuracy of this project.

8.1 Future Lines of Work

The results from this project can be used by NFL front offices to evaluate how their signings are performing compared to other teams. It would be interesting to do individual case studies on line items from each team's free agency and comparing it to avg_diff.

The intermediary tables outlined during the design process also have potential for future work. It would be interesting to do further work on are the tables that have the avg_diff values for each individual position. This metric is a good way to see if a team is under or overspending on certain positions.

9 REPOSITORY LINK

Here is the GitHub link to my project's code, data, and slides: https://github.com/ijga/6.s079-project

REFERENCES

[1] Sampath Dudde, William Wu, Macdonald Austin, and Rohan Konnur. [n. d.]. Developing a Metric to Evaluate the Performance of NFL Franchises in Free Agency. Sports Analytics Group at Berkeley. ([n. d.]).

[2] Andrew J Meil. 2018. *Predicting Success Using the NFL Scouting Combine.* Master's thesis. University State University, Fullerton.

[3] Jason Mulholland and Shane Jensen. 2014. Predicting the Draft and Career Success of Tight Ends in the National Football League. *Journal of Quantitative Analysis in Sports* 10, 4 (2014), 381–396.