Evaluating the Performance of Offensive Lineman in the NFL

Nikhil Byanna, Abdolghani Ebrahimi, Diego Klabjan

***Abstract*—**In this paper we objectively measure the performance of an individual offensive lineman in the NFL. The existing literature proposes various measures that rely on subjective assessments of game film, but has yet to develop an objective methodology to evaluate performance. Using a variety of statistics related to an offensive lineman's performance, we develop a framework to objectively analyze the overall performance of an individual offensive lineman and determine specific linemen who are overvalued or undervalued relative to their salary. We identify eight players across the 2013-2014 and 2014-2015 NFL seasons that are considered to be overvalued or undervalued and corroborate the results with existing metrics that are based on subjective evaluation. To the best of our knowledge, the techniques set forth in this work have not been utilized in previous works to evaluate the performance of NFL players at any position, including offensive linemen.

***Keywords*—**Offensive lineman, player performance, NFL, machine learning.

I. Introduction

THE front office of teams in the National Football League (NFL) face critical decisions on a daily basis. From weekly game planning to filling coaching vacancies, teams must constantly evaluate their personnel to make short- and long-term decisions that are in the best interest of the franchise. One of the most important decisions NFL front office must make on a yearly, and even weekly, basis is determining which players should be on the 53-man roster.

There are three main avenues through which teams acquire players for their roster: the NFL draft, free agency, and trades. This paper will focus on the latter two. In both free agency and trading, teams must not only assess their personnel needs, but also consider the salary cap, or a league-wide budget constraint capping how much teams can spend on players' salaries each year. A problem that teams face in when negotiating contracts with players involves the notion of Illusory Superiority. A player has a cognitive bias that leads him to overestimate how good he is relative to other, e.g. linemen, in the league. Thus, the player will demand higher salaries in negotiations if the team cannot logically defend their salary offer. Therefore, teams are constantly trying to develop analytical methods to more accurately assess a player's true value, and thus better inform their free agency and trade decisions.

According to an interview with the Director of Pro Personnel for an NFL team, the evaluation of offensive linemen is 95% based on watching game film and 5% on using performance metrics and quantitative analysis. Once enough film has been watched for a given player and grades have been assigned to individual games and the lineman as a whole, the scouting team will then find "comparables," or other players who were similar to the lineman in question across various numerical dimensions. These dimensions are typically descriptive statistics such as height, weight, and number of Pro Bowl appearances. Once a team finds 4 to 5 comparables, they determine what would be a reasonable offer to make to the player.

The goal of this research is to develop a methodology that can be used to determine the monetary value of an offensive lineman. The framework is structured to help an NFL executive answer the following question when making salary decisions in free agency and trades: how much is a given offensive lineman worth? The methodology clusters players together based on statistics that are determined to be priced into the salaries of offensive linemen in the NFL labor market, effectively creating a pool of players who deserve to be compensated similarly based on their performance. If a player is found to have a salary that is anomalous as compared to the salaries of other players within his cluster, then his worth can be determined relative to his previous contract's APY (average per year) salary. If a player is found to be overvalued, then an NFL executive has reason to believe that the NFL labor market has inefficiently priced that player's salary and the player is worth less than the APY salary of his previous contract. Similar logic holds for a player that is found to be undervalued.

By creating a framework to analyze a player's performance based on objective statistics from the season in question, we reduce the subjectivity inherent in current metrics used to evaluate offensive linemen, such as the Pro Football Focus [4]-[6] metric described in Section II. The novelty of our approach is that it considers the actual outcomes that occur as a result of an offensive lineman's performance on a given play, as opposed to subjectively determining the impact that the given lineman had on the outcome of a play. This allows NFL executives to tie player performance to realized outcomes, which is directly related to the ultimate goal of teams: putting their team in a position to generate outcomes that lead to wins.

II. Related Work

The position in the NFL that has been the easiest to analyze is the quarterback. One of the primary advanced metrics used to analyze quarterbacks is the "Total Quarterback Rating (Total QBR)," which attempts to measure the situational performance,

Nikhil Byanna was with the Department of Industrial Engineering and Management Sciences at Northwestern University, USA (e-mail: nikhilbyanna @gmail.com).

Abdolghani Ebrahimi (PhD candidate) is with the Department of Industrial Engineering and Management Sciences at Northwestern University, USA (corresponding author, e-mail: ghani@u.northwestern.edu)

Diego Klabjan (Professor) is with the Department of Industrial Engineering and Management Sciences at Northwestern University, USA (e-mail: d-klabjan@northwestern.edu).

as opposed to the absolute performance, of the quarterback [1]. A large component of Total QBR is the concept of Win Probability Added (WPA), which measures the impact of a given play on the team's chances of winning the game [2]. The other important metric Total QBR considers is the Expected Points Added (EPA) of each play, which is similar to WPA but measures the impact of a given play on the team's expected total points for the game. By using a combination of these two metrics, Total QBR creates a robust metric to measure a quarterback's performance. If a universal metric like Total QBR could be created for other positions in the NFL, teams would have a much easier time comparing players based on their statistical performance.

The concept of WPA and EPA has been applied to the majority of positions in the NFL by Advanced Football Analytics, a website dedicated to the analysis of the NFL using mathematical and statistical models. The only position that has not had WPA and EPA applied to individuals is the offensive lineman: left and right guard, left and right tackle, and center. Advanced Football Analytics currently has a WPA and EPA calculated for the offensive line as a whole, but has no WPA or EPA calculation for an individual lineman. Given the lack of tangible statistics for offensive linemen, the majority of football analysts have had the same difficulty that Advanced Football Analytics has had in evaluating the performance of an individual offensive lineman.

The most widely cited and used metric for individual offensive lineman is published by Pro Football Focus (PFF), a website published by PFF Analysis Ltd. PFF grades every single snap on offense, defense, and special teams, assigning a grade to each player on the field. These grades are obtained by watching film from every game in an NFL season. The grades range from +2.0 to -2.0 in increments of 0.5, with a grade of "0" being viewed as the "expected" grade for an NFL player. The main factors considered for an offensive lineman are pass protection, run blocking, screen blocking, discipline, and procedure (see [3]). These grades are aggregated across the season and normalized across the position before a final rating is produced. This methodology appears to be sound and is the only notable metric that has been published thus far in relation to offensive linemen. However, one major aspect of the methodology is the subjectivity involved in grading each play. While some of the subjectivity can be controlled for by using consistent graders and validating across graders, the nature of the system is inherently subjective. Based on current research, there has yet to be an objective methodology created to evaluate the performance of an individual offensive lineman, which is what this research aims to develop.

III. Data

The primary dataset used for analysis was obtained from STATS LLC, a global statistics and sports information company that tracks, analyzes, and distributes play-by-play data from a variety of sports. The data contained game-by-game data for every offensive lineman that was recorded as having played at least 1 snap in a regular season or playoff game from the 2013-2014 and 2014-2015 seasons. The variables are detailed in Table I.

TABLE I
List of Variables Contained in the Dataset

Descriptive Variables	Other
Unique game code	Base salary
Playoff vs Regular Game	Signing bonus
Game date	Incentives
Player name	Cap Value
Team	Snaps
Opponent	Holding penalties on rush attempts
Position	Holding penalties on pass attempts
Rookie year	Passing yards
Draft round	Dropbacks
Draft pick	Passing attempts
Birthday	Sacks to side/not to side
Rush attempts to side/not to side	Sack yards to side/not to side
Stuffs to side/not to side	Pressures to side/not to side
Rush yards to side/not to side	Hurries to side/not to side
Yards after contact to side/not to side	Knockdowns to side/not to side
Rush touchdowns to side/not to side	Quarterback release time/attempts
Successful rushes to side/not to side	Release time/attempts under pressure

As seen in Table I, the statistics cover a wide range of a player's attributes, including attributes related to the game, the player's salary, and the running or passing plays in which the player was involved. Each statistic was tracked on a play-by play basis such that it was only included for a given player if the player was on the field during the play. The play-by-play statistics were then aggregated on a game-by-game basis. It is worth noting the majority of statistics include a "to side" and "not to side" entry – these statistics were determined based on the location in which the play occurred. Based on the tracking methodology of STATS, the line is divided into five distinct splits –LS, L, M, R, RS— where LS indicates the far left side of the line, RS indicates the far right side of the line, and the remaining 3 categories are evenly divided between LS and RS. For a given position, the "to side" statistics were aggregated based on any play that was recorded as corresponding to a split in column 2 of Table II. For that same position, the "not to side" statistics were then aggregated based on any play that was recorder as going to any split not listed in column 2 of Table II.

The rationale behind this methodology is as follows. The reason that it has been difficult to objectively evaluate an offensive lineman's performance is because it is nearly impossible to track individual statistics for linemen. One way to approach this impediment is to track statistics that the offensive lineman is responsible for ("to side"), as well as the statistics that the offensive lineman is not responsible for ("not to side"). The "to side" statistics provide information on the direct contribution that the offensive lineman makes on a given play. The "not to side" statistics effectively serve as a control, showing how the statistics differ based on the other players on the field. For example, if a lineman on a given team has a large number of "to side" rushing yards but also has a large number of “not to side” rushing yards, this indicates that the numbers might be influenced by an exceptional running back or exceptional offensive lineman on the other side of the line. Based on this logic, “differential statistics” were created to highlight the differential between the “to side” and “not to side” statistics. In most cases, a positive differential indicates that the player outperforms the linemen on the other side, while a negative differential indicates that the linemen on the other side of the line outperform the player in question. A derived attribute was created for every statistic in Table I that includes a “to side” and “not to side” entry.

TABLE II
SPLITS OF THE LINE THAT ARE CATEGORIZED AS “TO SIDE” FOR A GIVEN POSITION

Position	Splits that are "to side"
LT	LS, L
LG	LS, L, M
C	L, M, R
RG	M, R, RS
RT	R, RS

Offensive linemen are often involved in run plays that are directed to the other side of the line – most notably in zone or power schemes. Ideally, a dataset used for this analysis would include a variable that indicates whether a specific lineman was involved in a block when the ball is run to the other side of the line, in the instance of a stretch, pull, or counter play. In this instance, our methodology could be adapted to calculate the differential between plays that involved the offensive lineman and plays that did not involve the offensive lineman in such plays, which would serve as a control for the ability of other players on the field. However, this type of data is currently not tracked, which leads to the methodology described above as the best proxy to control for the ability of other players on the field. A further discussion of the use of "differential statistics" can be found in Appendix A.

Aside from the data acquired from STATS, additional data needed to be extracted from the internet to capture all aspects of an offensive lineman’s performance. A key measure of an offensive lineman’s overall performance that could not be pulled from the STATS database is the player’s past performance, which can be represented by the number of selections the player has had to the Pro Bowl and All-Pro teams. There are 3 different All-Pro teams that are selected: Associated Press 1st Team, Associated Press 2nd Team, and Pro Football Writers 1st Team, in order of respective prestige. Thus, there were four additional variables created in the data set, one for the number of Pro Bowl selections and one for the number of selections to each of the three All-Pro teams.[1] The last data set used involves the Pro Football Focus grades described in Section II. The individual grades for each offensive lineman were extracted from 2007 – 2015. These grades are used for two purposes: to serve as potential independent variables in the regressions outlined in Section IV.A and to serve as a validation mechanism for the findings in our work.[2]

IV. METHODOLOGY

A. NFL Labor Market Pricing of Offensive Lineman

In order to appropriately identify overvalued and undervalued offensive linemen, it must first be determined how the NFL labor market prices the salaries of offensive linemen. The first analysis seeks to determine which player characteristics, both performance and experience based, are valued in the NFL labor market. Using a player’s average salary per year over the entirety of the contract (subsequently denoted as Cap Value), adjusted for cap inflation, as the dependent variable in a linear regression, we can determine which characteristics are priced into a player’s salary by examining the independent variables that are statistically significant in the model specification. Since we are using the APY salary of contracts signed before the 2013-2014 and 2014-2015 seasons, it is reasonable to assume that the APY salary of a contract is a function of a player's expected performance, and that the player's actual performance in the given season can be a proxy for their expected performance (i.e. executives have perfect foresight into projecting a player's future performance). A preliminary stepwise regression is run using the data to determine a linear model specification that accurately characterizes the data.

The initial set of potential variables includes 15 variables corresponding to the data described in Section III. This includes demographic characteristics (e.g. age, experience, and draft round), proxies for past performance (e.g. All Pro and Pro Bowl selections, average of past season PFF +/- ratings) and 5 "differential statistics" for the year in question believed to be important in describing a lineman's performance (e.g. stuff percentage differential, rushing yards per carry).[3] The player's PFF rating for the year in question was also included as a potential predictor to assess whether it's a stronger representation of the player's on field performance than the "differential statistics" used.[4]

It is worth noting that certain players were excluded from the dataset before the regression model was run. Any player in the dataset who was still under their rookie contract was eliminated from the dataset and excluded from the remainder of the dataset. The reason for this is because rookie contracts are determined based on a fixed scale and thus are not representative of a free market in which the true value of a player can be determined. Additionally, any players who were not unrestricted free agents at the time of signing their existing contract were excluded. This is due to the fact that the team they were with at the time of contract negotiation owned additional rights to the player that they otherwise would not have if the player were an unrestricted

[1] The data was manually pulled from [4-5] and merged with the existing STATS dataset.

[2] The data has been manually extracted from the [6] and merged with the existing STATS dataset.

[3] See Appendix B for a more detailed description of the differential statistics.

[4] It is worth noting that the player's PFF rating for the current year was not found to be significant in the regression, indicating that the differential statistics that are used have a stronger correlation to a player's salary than PFF ratings.

free agent.[5] Lastly, only players who signed contracts since 2011 (when the new CBA was signed) were included in the regression data.

While the initial regression model considers a variety of factors specific to the offensive lineman in question, a factor that it omits is the ability of the other offensive linemen on his line. Based on this omission, a control was created to account for the ability of the linemen on the same side of the line, to determine whether this is significant in a regression that models an offensive lineman's salary. To create this metric, the individual predictors from the initial regression model are categorized into “experience” predictors and “performance” predictors so we can separate the player's current performance from their past experience. “Performance” predictors are any predictors that directly relate to the player’s performance on the field during the season. All other individual predictors are categorized as “experience” predictors. We denote by E, P the set of all “experience, performance” predictors, respectively. To calculate these metrics, we seek to assign a relative weight to each predictor based on its effect on a player’s salary, which is represented by the coefficient of the predictor that results from the regression model. Thus, using the coefficients of each predictor, α_j , and the set that the predictor belongs to (E or P), normalized weights can be calculated for each of the predictors of the model and can subsequently be used to calculate the desired metrics.

These weights allow an overall “Experience” and “Performance” metric to be created for each player i in the dataset. For each player in the data set, the Experience metric for the player that played on the left and right side of the player in the given season are averaged to get a “Team Experience Metric.” Likewise, the Performance metric for the player that played on the left and right side of the player in the given season are averaged to get a “Team Performance Metric.” It is worth noting that players who play Left and Right Tackle, which means that they are at the end of the line, only have one player playing next to them, and thus had the team metrics take on the value of the metric for the one player that played next to them. These new statistics are now included as potential independent variables in a new stepwise regression model, along with all of the original independent variables, to create a final linear regression model that is used in the remainder of the analysis. Fig. 1 shows the results of the final linear regression model, while Fig. 2 provides a description of the differential statistics used in the regression. Please see Appendix B for a description of why the Experience and Performance metrics were not found to be significant in the final regression.

Variable	Estimate	t value	Pr(>|t|)
(Intercept)	5399261	9.427	2.12E-15
Avg, PFF Rating Prior to Contract	56697	3.081	0.00268
Experience	-199134	-2.694	0.00831
Draft Round	-264405	-4.224	5.38E-05
Pro Bowl Selections	624910	3.715	0.000338
Stuff % Differential	-82247	-2.551	0.012284
Yds per Attempt Differential	382197	2.035	0.044516
Sack %	-2516	-2.31	0.022994
Adjusted R-squared			**0.50**

Fig. 1 Final regression results

B. Clustering

The linear regression model, formulated after performing the stepwise regression from the previous section with m independent variables, determines which player characteristics are valued by the NFL labor market, and thus priced into the salaries of offensive linemen. We next seek to group similar players together based on a comparison of the characteristics specified in the model. This is done via a k-means cluster analysis, which seeks to create k distinct clusters of players from the overall dataset of n players [7], who each containing m standardized attributes.

To determine the optimal number of clusters, the k-means clustering algorithm is run for values of $k \in \{1, \dots, 20\}$ and the Krzanowski-Lai statistic [8] is computed for each iteration of k, as defined by:

$$C_k = \left|\frac{DIFF(k)}{DIFF(k+1)}\right| \ where\ DIFF(k) = (k-1)^{\frac{2}{m}} \cdot Within\ SS_{k-1} - k^{\frac{2}{m}} \cdot Within\ SS_k \qquad (1)$$

Here “Within SS” is the sum of square distances within all clusters. Once C_k is determined for $k \in \{1, \dots, 20\}$, it can then be plotted as a function of k to determine the optimal number of clusters. To determine k^*, we identify all k values that correspond to local maxima of C_k as potential candidates and then further examine the data to choose from these candidates. After k^* is determined, the k-means clustering algorithm is run with k^* clusters.[6]

C. Characterization of Clusters

The goal of forming the k^* clusters created in the previous section is to group players of similar ability into the same cluster, thereby providing a basis for player comparison. Based on the objective function of the clustering algorithm, a player should be placed in a cluster with other players who are similar to him across the m dimensions in the data. The end goal of the analysis is to identify players within the clusters that are worth more or less than the APY salary of their previous contract based on their performance. However, the clusters must first be inspected in an attempt to provide a characterization based on the attributes of the players within the cluster.

[5] Players with two observations (one for the 2013-2014 season, one for the 2014-2015 season) that were under the same contract had only one data point included due to heteroscedasticity concerns. Their statistics from the 2 seasons were averaged to create a single data point. After the model was specified, these two observations were then disaggregated for the clustering analysis.

[6] Refer to Appendix C for a detailed analysis of the k-means application.

Differential Statistic	Description
Stuff % differential	% of carries to lineman's side minus % of carries not to lineman's side that were stopped
Yds/Attempt differential	Yards per carry to lineman's side minus yards per carry not to lineman's side
Successful Run % differential	% of carries to lineman's side minus % of carries not to lineman's side that resulted in a
Sack %	Sacks allowed by lineman divided by sacks allowed by other linemen on field

Note: A successul play occurs on 1st down when a play achieves greater than 40% of the yardage to gain, on 3nd down when it achieves greater than 50% of yardage to gain, on 3rd or 4th down when play results in a 1st down. Touchdowns are always successful plays; turnovers are always unsuccessful plays

Fig. 2 Description of "differential statistics" used in the regression analysis

To create a consistent method of inspection across clusters, t-tests are performed using the sample mean of each predictor within a given cluster and the sample mean of the overall dataset. The t-test is carried out under the null hypothesis that the population mean of the cluster is equal to the overall population mean, with the alternative hypothesis being that the population mean of the cluster is not equal to the overall population mean.

This hypothesis test is carried out for each predictor j within each cluster w, with p-values calculated for each hypothesis test. Using a significance level of 1%, predictors within each cluster that had p-values that are less than .01 are deemed to be statistically significant. Thus, there is evidence that suggests that the population mean of cluster w is different from the population mean of the overall population for any predictor j that has a corresponding p-value of less than .01. Once these predictors are identified for each cluster w, a qualitative assessment of the p-values is conducted to determine whether subsequent hypothesis testing (described in Section IV.D and Section IV.E) should be one or two sided. If the predictors with significant p-values all suggest that a player should be either compensated with a high or low salary, then a one-sided test should be conducted. For example, if a cluster is found to have less experience and below average run blocking performance in the given year, then the cluster exhibits below average performance. Thus, we want to find players who are paid a significantly higher salary than the rest of the players in the cluster even though they are not worthy of a higher salary. If a cluster does not contain clear indications such as the preceding example, then a two-sided test should be performed on the cluster.

D. Distribution Fitting

As mentioned in the previous section, the final component of the analysis is to test for a players worth within the clusters that were created. This test is based on the underlying assumption that the empirical salary data is randomly drawn from some parametric distribution for each cluster w. Therefore, these empirical salary distributions must be determined before proceeding to test for the worth of a given player. To estimate the parameters for the empirical salary distribution of each cluster, statistical software is used to fit a distribution to each cluster based on the salary values associated with each player within the cluster. The empirical distribution of salaries for any cluster is bounded below by the minimum base salary of an NFL contract, which has been historically increasing on a yearly basis. Thus, for the purposes of distribution fitting, the lower bound of a potential distribution should be the minimum base salary across all years included in the data set.

Based on this information, a select group of empirical distributions are plausible to model the salary of a given cluster. Furthermore, the Lognormal, Gamma, Beta, Pareto, and Weibull distributions have been empirically found to be descriptive models for the distribution of income [9]. Thus, the distributions used for consideration should be restricted to these 5 families of distributions insofar as they can provide a reasonable fit based on the criteria described in the remainder of this section. The first criteria on which to evaluate the fit of a distribution are the Chi-Squared statistic and the Akaike information criteria, which are both statistical measures of goodness of fit. Evaluating these values for a given distribution, in relation to the other candidate distributions, provides a proxy for the relative fit of the distribution. Once the list of candidate distributions is narrowed down based on these statistics, an examination of the P-P (probability-probability) and Q-Q (quantile-quantile) plots will help to inform which distribution to choose. Based on these two criteria, a distribution is chosen for each of the clusters created in Section IV.C. For clusters of size $n \leq 15$, it is recommended that distributions are not fit to the data given the small sample size, and thus such small clusters are neglected from further analysis. A sample distribution fit, and the corresponding P-P and Q-Q plots can be found for Cluster 1 in Figs. 3-5.

E. Player Identification

The final component of the analysis can now be carried out to determine how much a given player is worth, using the distributions determined in the previous section. The goal of this analysis, in the case of overvalued players, is to provide executives with a statistical result that shows that the player in question is not worth as much as they were paid in their previous contract. This result can be formalized by utilizing the distributions found in Section IV.D.

Consider a highly compensated player i in cluster w with a cumulative distribution function f. Suppose an NFL executive chooses a player from cluster w and believes that his performance in the following year will be similar to his past year's performance. Based on this supposition, player i's performance from the following season would be expected to be placed in cluster w. The question that an NFL executive must ask is if the player is worth as much as his previous contract's salary S. The probability that a player that is placed in cluster w is paid greater than S dollars is equal to $P(x \geq S) = 1 - f(S)$. If $P(x \geq S) \leq .05$, then S is significant at the 5% level, and an

NFL executive has reason to believe that the player should have a compensation lower than S in the coming year. Similar logic is applicable to lower paid players.

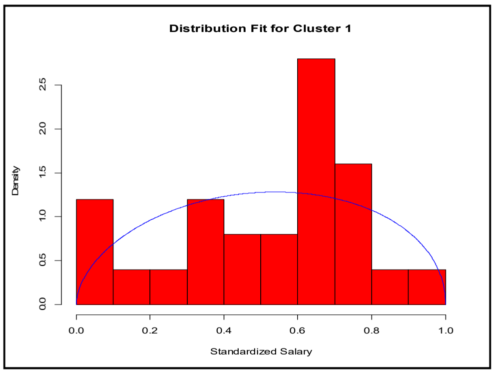

Fig. 3 Beta distribution fit for Cluster 1

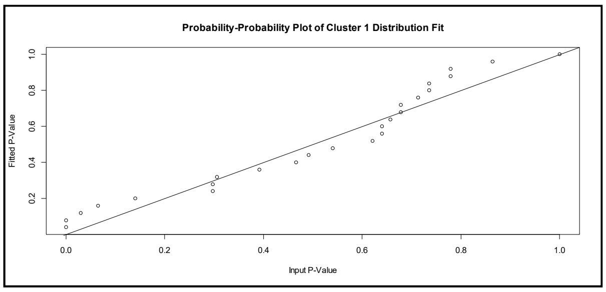

Fig. 4 Probability-Probability (P-P) plot for the beta distribution fit of Cluster 1

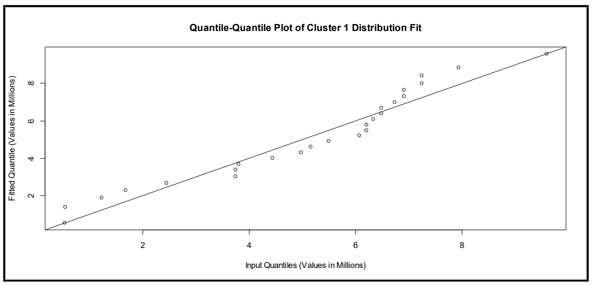

Fig. 5 Quantile-Quantile (Q-Q) plot for the beta distribution fit of Cluster 1

Based on the intuition outlined above, the analysis can be carried out for each individual cluster while performing one sided or two sided tests based on the characterizations from Section IV.D. From these tests, a list of overvalued and undervalued players is created as the output of the analysis. However, one additional assumption must be verified to create the final list of players that are determined to be overvalued or undervalued. An underlying assumption that is needed to apply the methodology's logic is that the characteristics of the players within a cluster are similar enough that one can consider any two players within a cluster to be similar. This is an assumption that, when violated, weakens the argument set forth in the remainder of the analysis. A way to formalize the concept of cluster homogeneity is through the silhouette value for a given point within a cluster analysis, which characterizes how well a player fits into his given cluster [10]. Let s(i) be the resulting silhouette value of player i. If s(i) is close to 1, player i is very well clustered. If s(i) is close to 0, player i is close to being placed in a neighboring cluster. If s(i) is close to -1, player i would be more appropriately clustered if placed in its neighboring cluster.

Given the construction of s(i), we seek an average s(i) value in the sample that is positive and significantly larger than 0 to be consistent with the assumption of cluster homogeneity. Given the small sample used for this analysis, the average s(i) value was lower than desired, taking on a value of slightly greater than 0.16. Theoretically, as more seasons of data are added to the sample, the clustering algorithm will be able to create more homogenous clusters and the s(i) value will increase to a value that is sufficiently large to be consistent with the assumption of cluster homogeneity (specific thresholds have not been empirically defined, but [11] suggest a value at least above .25 and ideally above .5). To account for the low average s(i) value in the sample, player i has to satisfy the criteria outlined previously in this section, as well as the criterion outlined in (7), to be considered overvalued or undervalued in the analysis:

$$s(i) \geq \frac{\sum_{j=1}^{n} s(j)}{n} \tag{2}$$

The criterion in (2) stipulates that the silhouette value of player i must be greater than the average silhouette value of the entire sample. It is worth noting that this is a simple heuristic that was used for the purpose of the analysis, and that other heuristics can be tested to develop a stronger condition that must be satisfied.

V. Results

A. Player Clusters

The analysis was conducted using the dataset outlined in Section III and the methodology set forth in Section IV. The clustering algorithm was run using 7 clusters (see Appendix C for why we chose 7 clusters), with the descriptive statistics of each of the 7 clusters, as well as the entire sample, shown in Fig. 6.

Cluster	# Players	Mean	Standard Deviation	Median	Minimum	Maximum
1	25	$ 5,118,847.06	$ 2,386,079.12	$ 6,067,019.30	$ 528,701.62	$ 9,588,292.68
2	17	$ 2,967,692.92	$ 1,834,427.10	$ 2,984,758.11	$ 541,463.41	$ 5,714,855.64
3	18	$ 3,838,068.63	$ 1,817,446.69	$ 4,004,880.79	$ 795,634.99	$ 6,731,317.82
4	23	$ 2,206,081.26	$ 1,236,261.33	$ 1,759,045.86	$ 666,587.32	$ 5,082,092.17
5	24	$ 2,528,372.25	$ 1,288,775.69	$ 2,523,431.23	$ 590,243.90	$ 5,283,149.56
6	8	$ 6,656,008.67	$ 1,215,143.94	$ 6,258,929.77	$ 5,719,508.25	$ 9,253,297.36
7	18	$ 3,098,081.01	$ 1,925,532.67	$ 2,559,429.46	$ 528,701.62	$ 6,889,518.41
Sample	133	$ 3,518,357.31	$ 2,146,048.26	$ 3,085,988.93	$ 528,701.62	$ 9,588,292.68

Fig. 6 Descriptive statistics of salary for each of the 7 clusters and the entire sample

Based on the sizes of the clusters, it was determined that Cluster 6 was not sufficiently large to proceed with the remainder of the analysis. Thus, Clusters 1, 2, 3, 4, 5, and 7 were further analyzed in order to provide a qualitative characterization and determine whether to test the cluster for undervalued players, overvalued players, or both. Fig. 7 depicts the characterization of each cluster, based on the analysis of the p-values that resulted from the t-test methodology described in Section IV.C.

Cluster	Characterization	Value to test for
Cluster 1	Players who were early draft selections and have above average run blocking abilities	Undervalued
Cluster 2	Players who had above average PFF ratings prior to their contract being signed	Undervalued
Cluster 3	Players who were early draft selections, have less experience, and fewer pro bowl selections	Both
Cluster 4	Players who were late draft selections, have below average PFF ratings prior to contract being signed	Overvalued
Cluster 5	Players who were late draft selections, have fewer Pro Bowl selections and below average run blocking abilities	Overvalued
Cluster 7	Players who had below average PFF ratings prior to contract being signed	Overvalued

Fig. 7 Characterization of Clusters 1, 2, 3, 4, 5, and 7

The resulting cluster that each player was placed in can be found in Fig. 8. It is worth noting that whenever a player is listed twice in Fig. 8, the first entry corresponds to the player's performance in 2013 and the second entry corresponds to the player's performance in 2014.

B. Player Identification Through Clusters

Employing the methodology described in Section IV.E to clusters 1, 2, 3, 4, 5, and 7, four players were identified as being overvalued or undervalued in the 2013-2014 and 2014-2015 seasons combined. It is worth noting that the original analysis identified eleven overvalued and undervalued players, but the additional criterion in (7) eliminated seven of the eleven players. The two undervalued players can be found in Fig. 9, along with the given year, the team they were on, the position they played, and their cap value.

Player	Cluster	Player	Cluster	Player	Cluster
Alex Boone	4	Evan Mathis	2	Mike Brisiel	5
Andre Smith	1	Evan Mathis	2	Mike McGlynn	3
Andrew Gardner	4	Fernando Velasco	5	Mike McGlynn	3
Andy Levitre	1	Gosder Cherilus	1	Nate Chandler	4
Andy Levitre	3	Gosder Cherilus	1	Nick Hardwick	2
Anthony Collins	3	Harvey Dahl	5	Paul Fanaika	4
Austin Howard	4	J.D. Walton	7	Paul McQuistan	3
Austin Howard	5	Jake Long	6	Phil Loadholt	1
Ben Grubbs	1	Jared Veldheer	3	Phil Loadholt	3
Ben Grubbs	7	Jeremy Zuttah	7	Ramon Foster	4
Brad Meester	2	Jermon Bushrod	1	Ramon Foster	4
Branden Albert	1	Jermon Bushrod	1	Roberto Garza	2
Branden Albert	1	Jeromey Clary	5	Roberto Garza	2
Breno Giacomini	4	Joe Barksdale	1	Rodger Saffold	3
Breno Giacomini	4	Joe Barksdale	7	Ryan Clady	6
Brian de la Puente	4	Joe Berger	2	Ryan Kalil	6
Brian de la Puente	5	John Jerry	3	Ryan Kalil	6
Bryant McKinnie	1	Jon Asamoah	1	Ryan Wendell	4
Byron Bell	4	Jonathan Goodwin	2	Ryan Wendell	5
Chad Rinehart	3	Josh Sitton	2	Samson Satele	3
Chad Rinehart	7	Justin Blalock	3	Samson Satele	7
Charlie Johnson	5	Justin Blalock	7	Scott Wells	5
Charlie Johnson	5	Kevin Boothe	5	Scott Wells	5
Chris Chester	1	Khalif Barnes	1	Sebastian Vollmer	1
Chris Chester	1	Khalif Barnes	1	Shawn Lauvao	3
Chris Clark	4	King Dunlap	4	T.J. Lang	3
Chris Myers	2	King Dunlap	4	T.J. Lang	3
Chris Myers	2	Kory Lichtensteiger	7	Ted Larsen	4
Chris Williams	7	Kory Lichtensteiger	7	Todd Herremans	2
Dan Connolly	5	Kraig Urbik	7	Todd Herremans	2
Dan Connolly	5	Kraig Urbik	7	Tony Pashos	2
Daryn Colledge	3	Kyle Cook	5	Travelle Wharton	2
Daryn Colledge	3	Logan Mankins	6	Tyler Polumbus	4
Davin Joseph	7	Logan Mankins	6	Tyson Clabo	4
Davin Joseph	7	Louis Vasquez	1	Will Beatty	1
Donald Penn	4	Louis Vasquez	7	Will Beatty	1
Doug Free	1	Lyle Sendlein	5	Will Montgomery	5
Doug Free	1	Lyle Sendlein	5	Will Montgomery	5
Doug Legursky	5	Mackenzy Bernadeau	5	Willie Colon	2
Eric Winston	1	Manny Ramirez	7	Willie Colon	2
Erik Pears	4	Manny Ramirez	7	Zach Strief	4
Erik Pears	5	Marshal Yanda	6	Zach Strief	5
Eugene Monroe	1	Marshal Yanda	6	Zane Beadles	7
Evan Dietrich-Smith	4	Matt Slauson	5		
Evan Dietrich-Smith	4	Michael Oher	3		

Figure 8 Cluster placement of players in dataset

Player	Year	Team	Position	Cap Value
John Jerry	2014	New York Giants	G	$ 795,635
Mike McGlynn	2014	Kansas City Chiefs	G	$ 1,037,594

Fig. 9 Players found to be undervalued based on the clustering analysis

These findings suggest that both of these players, based on their 2013 or 2014 characteristics, were similar to players in the dataset who were paid a significantly higher salary. Based on this conclusion, executives should be willing to pay the player the same salary, or a potentially higher salary, as their previous contract. The analysis also identified two overvalued players, which can be found in Fig. 10. For a detailed explanation of how the methodology outlined in Section IV.A – Section IV.E was applied to the dataset and produced the findings outlined in this section, please refer to Appendices B and C.

Player	Year	Team	Position	Cap Value
Scott Wells	2013, 2014	St. Louis Rams	C	$ 5,283,150
Davin Joseph	2013	Tampa Bay Buccaneers	G	$ 6,889,518

Fig. 10 Players found to be overvalued based on the clustering analysis

When inspecting the statistics of the 2 overvalued players identified in Fig. 10, the player's performance in the season(s) identified was significantly worse than the average performance of all players in the sample. Davin Joseph and Scott Wells both had significantly lower yards per attempt differentials than the league average, as well as significantly worse stuff percentage differentials. Davin Joseph also had a significantly lower average PFF rating for his time in the league prior to the contract being signed.

These statistics help strengthen the findings, showing that the three players identified as being overvalued did indeed perform significantly worse during the season than other linemen in the league. Furthermore, the actions taken by these players' teams following the season help to further corroborate the findings. Davin Joseph was released by the Buccaneers following the 2013 season, and Scott Wells was released by the Rams following the 2014 season.

C. Player Identification Predicted Salary

To identify additional players who may be overvalued or undervalued based on the analysis, the linear regression model from Section IV.A is used to compare a player's predicted salary to their actual salary in a given year. If a player's predicted salary is greater than 2 standard deviations away from their actual salary, they are identified as a candidate for being overvalued or undervalued. If these players were indeed overvalued or undervalued, this would suggest that the differential statistics used to predict future salary can predict future PFF rating better than a player's actual salary. This would show that the differential statistics are able to provide information about players that GMs are not picking up as they price offensive lineman's salaries. The four additional players identified can be found in Fig. 11, along with the given year, the team they were on, the position they played, their actual salary, and their predicted salary.

Player	Year	Team	Position	Actual Salary	Predicted Salary	Overvalued/Undervalued?
Branden Albert	2013	Kansas City Chiefs	T	$9,588,292.68	$5,820,659.08	Overvalued
Doug Free	2013,2014	Dallas Cowboys	T	$7,235,190.04	$3,521,875.32	Overvalued
Eric Winston	2013	Arizona Cardinals	T	$1,219,512.20	$4,666,476.67	Undervalued
Joe Barksdale	2013	St. Louis Rams	T	$528,701.62	$3,873,952.16	Undervalued

Fig. 11 Players found to be overvalued or undervalued based on their predicted salary

Similar to the overvalued players identified in the clustering methodology, the actions taken by the Chiefs and Cowboys corroborate the findings in the analysis. Albert was released by the Chiefs following the 2013 season, and Free was resigned at a lower APY salary contract value by the Cowboys in 2015. The results from Section V.B and V.C were combined and evaluated in the subsequent section.

D. Comparison to Pro Football Focus Metrics

Given the lack of published metrics that evaluate individual offensive lineman, it is difficult to use multiple methods to validate the findings in an empirical manner. As mentioned in Section II, the most widely cited metric used to evaluate individual offensive lineman is the Pro Football Focus rankings that are published on a weekly basis. A way to judge whether a player over performed or underperformed relative to their salary is to compare their relative performance ranking, as judged by Pro Football Focus, with their relative salary ranking based on each player’s cap value from the given season. If a player’s relative performance ranking is lower than their relative salary ranking, then they over performed relative to their salary.[7] Conversely, if a player’s relative performance ranking is higher than their relative salary ranking, then they underperformed relative to their salary. This intuition provides a basis for evaluation of whether a player is overvalued or undervalued solely based on their performance from the given season, which will be used to validate the analysis. Fig. 12 highlights the relative performance and salary ranking of each of the eight players found to be overvalued or undervalued, as compared to every other player in the sample that plays the same position as the player in question.[8]

As seen in Fig. 12, the comparison of relative performance rank to relative salary rank corroborates the findings from this paper for six of the eight players identified. In the case of two of the four players found to be undervalued, their relative performance rank was lower than their relative salary rank, suggesting that they over performed relative to their salary. In the case of all four players found to be overvalued, their relative performance rank was significantly higher than their relative salary rank, suggesting that they underperformed relative to their salary. It is worth noting that further application of the methodology to larger datasets would need to be conducted to validate these findings. Furthermore, the findings suggest that this analytical approach should not be used as the only method of player evaluation, but rather in conjunction with other evaluation methods to help defend salary decisions.

Player	Year	Position	Relative Performance Rank	Relative Salary Rank	Under/Overvalued
John Jerry	2014	G	23	28	Undervalued
Mike McGlynn	2014	G	31	27	Undervalued
Eric Winston	2013	T	21	19	Undervalued
Joe Barksdale	2013	T	9	22	Undervalued
Scott Wells	2013, 2014	C	12	3	Overvalued
Davin Joseph	2013	G	30	1	Overvalued
Branden Albert	2013	T	11	1	Overvalued
Doug Free	2013, 2014	T	7	3	Overvalued

Fig. 12 Comparison of relative performance rank and relative salary rank for identified players

Why do NFL teams not just use a comparison of relative performance rank and relative salary rank on a yearly basis to determine players that are overvalued and undervalued? The main reason we believe this is not a sound approach for evaluation is the subjectivity inherent in the formulation of the performance ranks. The performance rank is based on judgment, whereas the analysis in this work is data driven. Furthermore, we argue that having a relative performance rank that is significantly higher or lower than your relative salary rank is a necessary but not sufficient condition to be considered overvalued or undervalued. If a player is identified as having a significantly higher or lower performance ranking than his relative salary ranking, this is solely indicative that the player’s performance in the given year is not worthy of the salary that he has been paid. However, this assessment does not consider the past performance of the player, which is captured in statistics such as experience, age, and awards that the player has received. A player with multiple All-Pro 1st-team selections that underperforms in a given year relative to his salary may still be worth his salary if the season was an anomaly. The analysis accounts for these past performance indicators and determines whether they are strong enough to outweigh a poor performance in a given season by a player. Based on this intuition, the analysis outlined in this paper provides a result that is stronger evidence to determine if a player is overvalued or undervalued.

VI. Discussion

A. Consideration for NFL Executives

Sections V.B and V.C present an objective result that informs NFL executives of players that can be considered to be overvalued or undervalued relative to their performance.

[7] The relative performance and salary ranking are both computed in ascending order. Thus, the player with the best performance ranking has a relative performance ranking of 1, while the player that is paid the highest salary has a relative salary ranking of 1.

[8] For the purposes of this validation, the three positions considered were center, guard, and tackle. Thus, right guards and left guards were grouped together, as well as right tackles and left tackles.

However, this does not suggest a team should with absolute certainty give a player a more lucrative contract if that player was found to be undervalued or with absolute certainty release or trade away a player that was found to be overvalued. Consider a player who is found to be undervalued based on this methodology. Although this informs an executive that the player's overall season performance makes his worth at least as high as his previous contract, it does not provide information regarding other underlying factors not described by the analysis. Thus, once players are identified via the methodology presented in the paper, NFL executives are encouraged to further evaluate exogenous factors such as the player's health and the technical development of the player before making decisions.

While this analysis provides a conclusive recommendation as to whether a player should be offered more or less money, it does not provide executives with a player's true value, or in other words the exact salary that a player deserves.2 Thus, it is imperative for an executive to consider the marginal increase or decrease in performance that would result from signing or releasing an identified player, while also considering the marginal increase or decrease in cost that would result. There exists no universal equilibrium when considering this trade off given the different states that a team may be in regarding their salary cap space and the utility that they derive from having a higher performing team. Therefore, this trade off must be evaluated on a situational basis.

This analysis is most helpful for teams that have minimal salary cap space and must make financial decisions that will allow them to be comfortably under the cap. In the case of undervalued players, the team can potentially exploit the market inefficiency and sign a player for less than he is worth, while still acquiring a player who performed relatively well in the prior season. In the case of overvalued players, the team can attempt to trade the player to free up salary cap space that they can use to sign a different lineman or use to address other positional needs.

B. Limitations

There are various shortcomings and novelties of the data that are unable to be accounted for in the analysis. The most significant shortcoming is the fact that the data only contains statistics from the 2013-2014 and 2014-2015 NFL seasons, which manifests itself in the two ways described below.

An underlying assumption that is required to employ a 2 sample t-test is that the data from both populations are normally distributed. Given the analysis is limited to only 2 seasons worth of data, there may be a concern that the sample size is not sufficiently large to make this assumption. However, as the dataset grows and more seasons are added, it is expected that the clusters will be sufficiently large, provided that the optimal number of clusters k^* does not increase proportionally to the increase in size of the dataset.

Another concern with the small sample size involves the distribution fitting process as described in Section IV.D. When fitting a distribution to historical data, the best fit is determined mainly based on its goodness of fit relative to other candidate distributions, with the exception of the inspection of the P-P and Q-Q plots. The P-P and Q-Q plots suggest that the majority of distributions that were fit are relatively good fits. However, some clusters had too small of samples to fit an appropriate distribution and thus were discarded from the analysis. This is not to say that the players from these clusters are valued correctly, but there is not sufficient evidence to say that they are overvalued or undervalued.

The last limitation involves the use of differential statistics in the analysis. An argument can be made that a lineman with "bad" linemen on the other side of the line are given an unfavorable advantage and vice-versa for a lineman with "good" linemen on the other side. While this is a potential concern, we believe that the differential statistics are a valid measure given the findings in Appendix A and other considerations mentioned throughout the paper. Furthermore, we can reason that defenses would place less of an emphasis on stacking the side of the line with "bad" offensive linemen to stop the run, making it more difficult for the lineman on the "good" side to generate successful runs.

C. Future Research

The analysis conducted in this work revolved primarily around using past performance to determine whether a player warranted the salary he was paid. While this is one way to approach the task of identifying overvalued and undervalued players, another approach is to try to map past performance to future performance based on other machine learning techniques classified as "semi-supervised" and "supervised" learning techniques [12]. Using these techniques, the dataset would be used as training data, in which each observation would include an input object and a desired output value. The input object would be the input vector of characteristics, while the output value would be the player's salary in either the current season or season that follows. An inferred function would then be determined, which could be used to map new input vectors to a given salary, providing teams with an estimate of a player's salary given his characteristics for a given season. It is worth noting that similar limitations may arise regarding the small sample size or the ability of the algorithm to accurately map input vectors to salaries.

Another alternative to the methodology described above is to further explore the salary regression model described in Section 4.1. If an accurate explanatory or predictive salary model could be developed using regression techniques, it would be able to inform NFL executives of how certain characteristics are precisely valued in the labor market, as well as provide them with a model that can precisely determine the salary a player deserves based on his performance. The most difficult aspect of developing an accurate model is being able to control for the

[2] Using the regression methodology outlined in Section IV.A, it is possible to create a linear model with salary as the dependent variable that can be used to predict a player's salary, or true value. However, given the limitations of the small sample, the predictive power of the model was not considered to be strong enough to include in the analysis.

various factors that impact an offensive lineman's performance. This work attempts to create the control through the use of “differential statistics". Some limitations of the use of "differential statistics" are described in Section III, and it is encouraged to use more descriptive data if enhanced player data tracking techniques are employed in the future by sports analytics companies.

Furthermore, it may be the case that the outcomes used to derive these statistics are a function of strategic decisions made by teams. For example, a defensive team may purposely place their best defensive lineman on the side of the line as the WORST offensive lineman on the opposing team, thus exacerbating that offensive lineman's poor statistics. Furthermore, if the best offensive lineman is on the opposing side of the line in this situation, then his statistics are artificially improved since the lineman on the other side is playing so poorly. Thus, additional controls may be needed to develop an accurate explanatory of predictive salary model. Potential controls include the ability of the defensive lineman on the field and the in-season performance statistics of the defense that is being played.

VII. Conclusion

Evaluating the performance of individual offensive linemen is a task that has been difficult to accomplish without the extensive use of rating players by watching game film. The most widely cited metric that is used to quantify the performance of offensive linemen is the statistic created by Pro Football Focus, which is calculated based off of individual grades given to offensive linemen based on an assessment of game film. This work aims to create an evaluative approach that is based off of objective statistics gathered relating to each player, with the goal of emphasizing the actual outcome of a play rather than subjectively assessing how the player's actions contributed to the outcome.

Through a multi-step methodology that groups similar players into clusters and subsequently evaluates the salary distribution of the clusters, certain players are identified as being overvalued or undervalued based on the salary that they were paid in the given year. Using the dataset obtained for this paper, 4 players were found to be overvalued and 4 players were found to be undervalued in the 2013-2014 and 2014-2015 NFL seasons. The Pro Football Focus metric was then used as a proxy for player performance, and it was found that 6 of the 8 players significantly over performed or underperformed relative to their salary in the same direction as identified by the analysis. Given the small sample, we believe these results are promising albeit not conclusive. As more seasons worth of data become available, the hope is to conduct a more robust analysis that can be used by NFL executives across the league to make more informed decisions regarding the acquisition and release of offensive linemen in the National Football League.

Appendix

A. Differential Statistics Methodology

To further investigate the efficacy of utilizing the "not to side" statistics in the differential statistics, as opposed to only using the "to side" statistics, regressions were run to determine the value of including the "not to side" statistics in the analysis. Using the methodology outlined in Appendix B, the stepwise regression was run using "to side" statistics instead of differential statistics. The results of this regression are presented in Fig. 13. The Adjusted R-squared from this regression is 0.47.

Variable	Estimate	t value	Pr(>|t|)
(Intercept)	2945845	2.22	2.87E-02
Avg, PFF Rating Prior to Contract	58801	3.143	0.002206
Yards per Attempt to Side	664269	2.215	0.029027
Successful Run % to Side	-7186	-2.872	4.99E-03
Experience	-216627	-2.918	0.004359
Draft Round	-275462	-4.333	0.0000353
Pro Bowl Selections	595882	3.457	0.000808
Adjusted R-squared			**0.47**

Fig. 13 Results of regression model with "To Side" statistics

To determine if "Not to Side" statistics are important in explaining a player's salary, the "Not to Side" statistics for "Yards per Attempt" and "Successful Run %" were added to the regression to examine the effect on the model. The results of this subsequent regression with an Adjusted R-squared of 0.5 are in Fig. 14.

Variable	Estimate	t value	Pr(>|t|)
(Intercept)	3823923	2.434	0.016753
Avg, PFF Rating Prior to Contract	57296	3.043	0.003014
Yards per Attempt to Side	802234	2.446	0.016235
Yards per Attempt Not to Side	-369593	-1.152	0.252264
Successful Run % to Side	-39231	-1.124	0.263812
Successful Run % Not to Side	33745	0.951	0.344178
Experience	-204426	-2.709	0.007981
Draft Round	-276854	-4.314	3.86E-05
Pro Bowl Selections	601169	3.475	0.000766
Adjusted R-squared			**0.50**

Fig 14 Results of regression model including "Not to Side" statistics

As seen in Fig. 14, including the "Not to Side" statistics impacts the model in a non trivial manner. Both of the "Not to Side" statistics have a coefficient that is opposite in sign to the "To Side" statistic, implying that the "Not to Side" statistics have the opposite effect on a player's salary than the "To Side" statistics – exactly what would be expected if the "Not to Side" statistics served as an adequate control for the ability of other players on the field (e.g. running back, linemen on the other side).

Furthermore, including the "Successful Run % Not to Side" statistic causes the "Successful Run % To Side" statistic to no longer be significant in the regression. This is further evidence that "Not to Side" statistics are important to consider in the model. If the "Not to Side" statistics were not important, than they would not affect the significance of the "To Side" statistics, and would potentially have the same sign as well. Although not all variables are significant in the regression, the evidence discussed above suggests that incorporating "Not to Side" statistics is important to develop a stronger regression model. Based on this information, and the fact that the regression model with the "differential statistics" has a higher adjusted R-squared value, we believe that the use of "differential statistics" serves as a reasonable control given the limitations of the data.

B. Labor Market Pricing of Offensive Lineman

This Appendix details the application of the methodology outlined in Section IV.A to the dataset used for the analysis.

The initial set of potential variables includes 15 variables corresponding to the data described in Section III. This includes demographic characteristics (e.g. age, experience, draft round and pick), proxies for past performance (e.g. All Pro and Pro Bowl selections, past PFF +/- ratings) and 5 "differential statistics" for the year in question believed to be important in describing a lineman's performance (e.g. stuff percentage differential, rushing yards per carry, successful run percentage, pressures allowed percentage, sacks allowed percentage). The player's PFF rating for the year in question was also included as a potential predictor, to assess whether it is a stronger representation of the player's on field performance in a given year than the "differential statistics" used. The summary statistics for the dataset and the potential variables can be found in Fig. 15.

The model specified by the stepwise regression included 8 independent variables, having an adjusted R-squared value of .50. This will be referred to as “initial salary model.” Using the initial salary model, the Performance and Experience metrics were created for each player and the “Team Experience Metric” and “Team Performance Metric” was calculated for each player, as described in Section IV.A. The categorization of variables into performance and experience variables can be found in Fig. 16**.**

Variable	Mean	St. Dev
Salary (APY)	$3,270,186.97	$2,141,891.57
Overall PFF Rating Before Contract	5.69	9.51
Overall PFF Rating in Current Season	1.05	13.60
Log(Age)	1.47	0.03
Experience	6.31	2.14
Draft Round	4.48	2.54
Log(Draft Pick)	1.98	0.45
1st Team All Pro Selections	0.08	0.30
2nd Team All Pro Selections	0.13	0.60
PWF All Pro Selections	0.09	0.35
Pro Bowl Selections	0.43	1.06
Stuff% differential	1.47	4.63
Yds/Attempt differential	-0.25	0.83
Successful Run % Differential	-0.41	6.32
Pressure %	27.47	139.58
Sack %	25.79	139.83
Sample Size (n)		**106**

Fig. 15 Summary Statistics of data used in regression

A new stepwise linear regression was then run that included these two metrics and their squared terms as independent variables. However, the final model determined by this stepwise selection process was deemed to fit the data worse than the initial salary model. Thus, the controls for the players on the line to both sides of a given player were not found to be significant and were not included in the final regression model. Instead, the initial salary model was used for the remainder of the analysis.

Experience	Performance
Age	Stuff % differential
Experience (Years in League)	Yds/Attempt differential
Draft_Pick	YAC/Attempt Differential
1st Team All Pro	YBC/Attempt Differential
2nd Team All Pro	Successful Run % Differential
PWF All Pro	Pressure Allowed Differential
Pro Bowl	Pressure %
PFF rating from past seasons	Sack %
	Rush Tds/Attempts Differential
	Att/Dropback
	PFF rating from curent season

Fig. 16 Categorization of independent variables into experience vs performance variables

There are a few possible reasons that the controls were not found to be significant in the new regression model. A given player's salary at the time a contract is signed is a function of their past performance, as well as their future expected performance. The initial salary model attempts to control for the future expected performance of players on the other side of the line (as well as the running back and other players on the field) through the use of differential statistics, using actual performance as a proxy. The "Team Performance" metric attempts to control for the future expected performance of players on the same side of the line. However, players on the same side of the line may have similar enough statistics from the current season to the lineman in question that it does not provide additional explanatory power in a regression model. The "Team Experience" metric explores whether the other linemen's past performance has an impact on the given lineman's salary. In theory, an NFL team may try to sign players who they believe to be worth less money if their other linemen are strong and the team has stronger needs at other positions. Based on this, the past performance of the other linemen may help explain the salary of a given lineman. However, based on the regression, this is not the case. It is not surprising that the "Team Experience" metric was not found to be significant, but we wanted to explore the possibility to conduct as thorough of an analysis as possible.

C. Clustering Analysis

This Appendix details the application of the methodology outlined in Section IV.B to the dataset used for the analysis.

Using the characteristics from the “Salary Model,” input vectors were created for each of the players in the dataset as outlined in Section IV.B. The input vectors were then used to run the k-means clustering algorithm for $k \in \{1,..,20\}$. The Krzanowski-Lai statistic was then calculated for each value of k, with the results shown in Fig.17.

Based on Fig. 17, it was determined that $k^* = 7$ since there exists a local maxima at $k = 7$ and the marginal decrease in the Krzanowski-Lai statistic from $k = 7$ to subsequent local maxima is not as large. Thus, the k-means algorithm was run with 7 clusters to group players together for the remainder of the analysis.

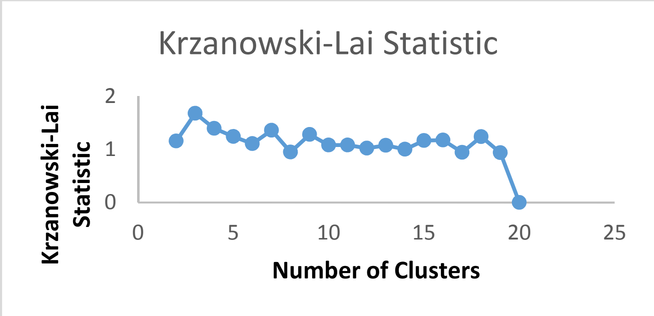

Fig. 17 Krzanowski-Lai statistic for various values of *k*

ACKNOWLEDGMENTS

We would like to acknowledge Ryan Warkins, Matt Scott and STATS, LLC for generously providing the dataset used in the analysis.

REFERENCES

[1] Oliver, Dean. (2011). Guide to the Total Quarterback Rating. Retrieved from http://espn.go.com/nfl/story/_/id/6833215/explaining-statistics-total-quarterback-rating.

[2] Burke, Brian. (2010). Win Probability Added (WPA) Explained. Retrieved from http://archive.advancedfootballanalytics.com/2010/01/win-probability-added-wpa-explained.html.

[3] Grading. (n.d.) Retrieved from https://www.profootballfocus.com/about/grading/

[4] Pro Football Reference. (2014). 2014 NFL All-Pros. Retrieved from http://www.pro-football-reference.com/years/2014/allpro.htm

[5] Pro Football Reference. (2014). 2014 NFL Pro Bowlers. Retrieved from http://www.pro-football-reference.com/years/2014/probowl.htm

[6] Pro Football Focus. (2015). *Cumulative Guard Summary*. Retrieved from https://www.profootballfocus.com/data/by_position.php?tab=by_position

[7] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, 100-108.

[8] Krzanowski, W. J., & Lai, Y. T. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 23-34.

[9] McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica: Journal of the Econometric Society*, 647-663.

[10] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

[11] Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. New York: Wiley Print

[12] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]." IEEE Transactions on Neural Networks 20.3 (2009): 542-542.