Optimizing the Allocation of Capital Among Offensive Positions in the NFL

by

Paul G. Calvetti, Jr.

B.S. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2022

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Paul G. Calvetti, Jr.
Department of Electrical Engineering and Computer Science
May 12, 2023

Certified by: Christina Chase
Lecturer, Department of Mechanical Engineering
Managing Director, MIT Sports Lab
Thesis Supervisor

Accepted by: Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Optimizing the Allocation of Capital Among Offensive Positions in the NFL

by

Paul G. Calvetti, Jr.

Submitted to the Department of Electrical Engineering and Computer Science
on May 12, 2023, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Building a successful National Football League (NFL) team is a challenging task, requiring front offices to balance player selection and compensation while operating under a salary cap constraint. The salary cap represents the maximum amount a team can spend on player salaries in a given season. Effective team construction entails strategic allocation of resources across different positions to maximize performance within this budget. This paper focuses on the critical aspect of allocating salary cap resources among offensive positions to maximize team success. We introduce a novel model that considers the interplay between players at different offensive positions, as well as the variations in salaries and performance levels observed between players under rookie and veteran contracts. By framing the allocation challenge as a constrained optimization problem, we aim to help teams maximize their points per game while staying within the salary cap limit. Our model's predictions enable us to identify the optimal distribution of resources across offensive positions, providing valuable insights for NFL front offices as they seek to allocate their salary cap to achieve maximum offensive performance and increase their chances of success on the field.

Thesis Supervisor: Christina Chase
Title: Lecturer, Department of Mechanical Engineering
Managing Director, MIT Sports Lab

Acknowledgments

First and foremost, I would like to thank my family, especially my parents, for their support of all of my endeavours ranging from athletic to academic. From a young age they always encouraged me to passionately explore my interests. They set an example for me and my siblings of how to work hard. I certainly would not be in the position to write this thesis without their love and support.

I would also like to thank my advisor, Christina Chase, not only for all her help with writing my thesis, but also for all of her help as my academic advisor throughout my undergrad years at MIT. Her support and advice was instrumental throughout my undergrad and over the past year while working on this thesis.

I would also like to thank everyone in the MIT sports lab for their help over the past year, especially Professor Anette (Peko) Hosoi. Peko's technical advice was highly valuable throughout my research process.

I would also like to thank my girlfriend, Rachel, for all of her love and support. She has a been a great influence on me and improved my quality of life. She helped broaden the horizon of my interests immensely, with our latest endeavour being yoga.

I would like to thank Dan West not only for his delicious meals, but also for his friendship throughout my time in college. His conversation kept me sane, and his meals fueled me as both a student and an athlete for 4 years.

I would like to thank all of my coaches and teammates on the MIT football team over the past 5 years. Football has taught me so much about myself, and being able to continue to play at MIT was an amazing experience.

I would like to thank all of my friends at MIT. I have been lucky enough to meet a great group of friends while at MIT and am thankful for all of the fun times we had together. Collaborating with many of them while taking courses helped me learn through my time in college. I was especially lucky to meet friends in my fraternity, DKE, with whom I spent my junior year during COVID taking classes remotely while relaxing on the beach in Plymouth and skiing in Winter Park, Colorado.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Each season National Football League (NFL) front offices are tasked with filling a roster with the best possible 53 players, with even more on the practice squad. Filling a roster with top level players often leads to teams success which can boost revenues for a team. However, since 1994 when the salary cap was implemented, teams must work to achieve this goal of fielding the best possible team with the constraint of the salary cap.The salary cap places a limit on the total amount of money each season that a team is able to pay its players. One strategy teams may try to implement is identifying players that they feel are undervalued. Players who possess similar skill levels to their peer at a lower cap hit are some of the most valuable players on a team. Paying these players lower salaries enables teams to use the saved capital to sign more players that help strengthen their roster while still saying below the salary cap. Another strategy teams employ is to identify the best draft prospects, yet there is minimal control in this process when filling a roster since draft picks are a valuable commodity that teams do not give away without high compensation. As a result, the draft receives considerable attention each year.

This research aims to address the question of how NFL teams should distribute their salary cap among offensive positions to maximize their offensive performance, as measured by their scoring per game. The allocation of capital to each position is a complex decision as it involves various factors such as the interplay between different positions and wide variations in pay scales. To address this question, this paper proposes a framework to model how a team's salary cap allocation among offensive positions affects their offensive performance. By analyzing this framework, teams can gain valuable insights into how they can optimize their salary cap allocation to maximize their scoring potential.

Chapter 2

Background and Related Work

NFL teams attempt to solve the problem of filling their roster with players that will lead to a successful team under the financial constraint of the salary cap. The salary cap enforces a maximum number of dollars that a team is able to pay its players each season. In the end, all cash a team pays to its players in both salary and bonus eventually must count against their salary cap. Although, there are tactics teams may use such as offering players a signing bonus from which the cap hit is amortized over the length of the players contract. This paper focuses on how a team chooses to allocate this salary cap among the various position groups.

One aspect of salary cap allocation is the selection of players in the draft. The draft is an opportunity for teams to add rookies to their roster who will be on relatively inexpensive contracts for four years. Interestingly, Past research done by Cade and Thaler demonstrated that teams overvalue early draft picks [8].

One approach that has been used to investigate the optimal salary cap allocation uses a linear regression model where the amount of money paid at each position group is used to predict a team's wins [6]. The study suggested that Tight Ends and Kickers are the positions where the marginal investment of salary leads to the largest expected increase in wins. They hypothesize that these positions are identified as having the most benefit because they experience injuries at lower rates than most other positions[6].

Other work incorporated Pro Football Reference's Approximate Value (AV) which is meant to be a measure of player value which is comparable between different positions and seasons [7]. This method utilized salary cap hit as a predictor of AV. Then the cumulative AV at each position is used to predict a team's number of wins. The study concludes that it is best to pay for top players at guard, defensive tackle, and free safety [9]. Both of these studies assume a linear relationship between salary, or talent, at a certain position and the win contribution from that position. However, various synergistic and antagonistic effects exist in football that make this assumption questionable. For example, an elite quarterback might elevate the players around him, or an elite wide receiver might lead to a team running the ball less often and, therefore, lower the contribution of a running back.

Utilizing a different approach, Froelich uses a conditional logit model to determine the impact the amount of salary cap allocated to the offensive line positions, especially left tackle, on game outcomes [3]. The study identified the wide range of 5% to 15% as an optimal range for salary cap spending on left tackle. Ness used a similar model to determine the impact of the percentage of the salary cap allocated to defensive positions, and concluded that only under certain conditions does increasing the percentage of salary cap allocated to defensive players improve a team's chance of winning [10].

Other studies have focused on the concentration of salary cap and the impact of having a more even distribution of salary cap versus having the salary cap concentrated on a few key players. A study by Zimmer used the Herfindahl–Hirschman index (HHI) as a measure of salary cap concentration along with various other predictors including the percentage of cap allocated to quarterback and running back in order to predict a team's success. [16]. They found that the best allocations were those that were most even and those the were most uneven.

This paper uses some of the same ideas found in the literature including the use of Pro Football Reference's Approximate Value (AV). However, this paper places an emphasis on understanding the interactions between players and how this should inform the allocation of capital among the offensive positions.

Chapter 3

Methodology

I aim to develop an approach that understands for the interactions between players on the same team. These interactions depend on how a team is constructed in terms of talent allocation. For example, adding a star wide receiver has a different impact on a team which already has a great running back and offensive line than a team which does not have great player at these positions. In this section, we discuss how our model is designed to capture these relationships, and offer quantitative, data driven insights into capital allocation among positions.

Additionally, our model accounts for differences between the compensation of players on rookie and veteran contracts. Players on rookie contracts are often paid below what the free agency market would pay for a player of their skill level due to their salary being constrained based on their draft position. In this paper, we understand the impact of this difference in compensation through the creation of metric we call *effective salary*.

We develop a model to determine the optimal position to allocate capital in order maximize points per game, which may differ for each team. We choose to focus on points per game since it enables us to limit our analysis to just the offensive positions. Although a teams defense may directly score points or put the offense in advantageous point scoring situations, to simplify our analysis we assume these effects are small and points per game can be explained by the performance of offensive players alone.

3.1 Data

I obtained salary cap data from Spotrac.com [2]. I obtained the rest of the data discussed in this paper from Pro-Football-Reference.com [1]. This includes Approximate Value (AV), wins, and points per game data. All data is obtained for NFL seasons from 2011-2021. This range is chosen because it the the range in which Spotrac had salary cap data available.

3.1.1 Position Information

We use abbreviations, such as QB for Quarterback, to refer to the various football positions. In this section we define all these abbreviations as well as include a short description of that position's role. Additionally, different positions have a different number of players both on the roster and on the field at one time. We note that for the number of players on the team this number includes any player that was on the teams payroll in this season including practice squad players, or players who only played for part of the season. These differences are also be outlined below. Lastly, we outline differences in the average total cap hit of different positions.

3.2 Player Performance Model

To develop a model that understands the interactions between players at different positions, I began by developing a model for player performance. This model uses salary to predict player performance. The inherent assumption in this model is that teams assign players' salaries in an efficient manner so that players receive a salary commensurate to their skill level. Although this may not always the case, with enough data, we hope to be able to capture the average performance a team can expect for players at a certain position with a given salary. We achieve this through the use of a measure we term *effective salary* which accounts for players on rookie contracts receiving lower salaried relative to players on veteran contracts with a similar skill level.

Position	Abbreviation	Typical # players on field	Avg. # players on team	Avg Cap %
Quarterback	QB	1	1.9	8.2
Running Back	RB	1 to 2	4.9	4.5
Wide Receiver	WR	1 to 4	6.2	8.6
Tight End	TE	0 to 2	3.6	3.9
Right Tackle	RT	1	2.3	2.5
Left Tackle	LT	1	1.6	3.7
Guard	G	2	3.6	4.9
Center	C	1	1.9	2.7
Kicker	K	0 to 1	1.1	1.1

Table 3.1: Position Information. The table shows the abbreviation for each position, which are be used throughout the paper. Additionally, it shows the typical number of players on the field at each position which can vary depending on the offensive personnel package, formation, and type of play. The table also shows the average number of players on a team at each position. This number includes any player that was on the teams payroll in this season including practice squad players, or players who only played for part of the season. Lastly, the table include the average cumulative cap percentage allocated to players at each position which varies widely with quarterbacks receiving 8.2% of the cap on average while kickers only receiving 1.1% of the cap on average.

3.2.1 Measuring Player Performance: Approximate Value

To develop a model that understands the interactions between players at different positions, I began by developing a model for player performance. I use Pro Football Reference's Approximate Value (AV) as a measure of player performance. AV is a metric which Pro Football Reference publishes for every player for every season since 1960. It is meant to be a measure of player value which is comparable between different positions and seasons [7]. Although they admit the metric is a "very approximate" metric, we still can use it as our measure of player performance since it exists for all players in our analysis.

3.2.2 Predicting AV: Breakdown by Rookie Contract Status

We model AV based on salary on the team level and not the player level. The reason for this decision is that we want to remove a level of complexity from our model. If we predict AV at the player level, in our model design we implicitly must answer the question of whether we prefer paying a few players a large amount of money or many players a small amount of money. Instead, we assume teams make this decision in some approximately optimal fashion.

Now we create a model from salary to AV. We break down this regression by position, and type of contract. We consider the following position groups for our model of AV: [C, G, LT, RT, TE, WR, QB, RB, K]. The reason we use a separate regression for each position is that pay level varies greatly between position, so we must consider a player's salary in the context of their position. We consider two types of contracts which are rookie and veteran contracts. We include undrafted free agent (UDFA) contracts in the rookie contracts. This decision is mostly due to it being convenient given the our dataset. Additionally, the salary given to a late round draft pick and a undrafted free agent is comparable since both are paid approximatley the league minimum. We define a rookie contract as any contract that is the player's first NFL contract, while we define a veteran contract to be any contract after the player's first contract. The purpose of this distinction is that players are often be paid below

Contract Type	Average AV	Average Cap %	Percentage Of Players
Veteran	1.91	3.74	60
Rookie	0.63	3.09	40

Table 3.2: Rookie vs Veteran Contract Player Cap Hit and AV Comparison. The table shows the average AV and cap hit for players on veteran and rookie contracts. It is apparent that players on veteran contracts are on average paid a higher percentage of the salary cap and also have a higher average AV. Additionally, the table shows that 60% of players in our data set and on veteran contracts while the other 40% of players are on rookie contracts.

their market value when they on rookie contracts since their pay is constrained to a predefined amount based on their draft position. Similarly, a UDFA is an unproven player in the NFL and thus is often be paid the league minimum. On the contrary, veteran players are usually proven player who have made it through a rookie contract and performed well enough to have earned another contract. Thus, we expect the to have higher average salaries and performance, which is reflected in the data as shown in table 3.2.

Each regression takes the following form for each team to predict rookie AVs.

$$AV_{i,r} \sim \alpha_0^{i,r} + \alpha_1^{i,r} \log\left(1 + S_{i,r}\right) \tag{3.1}$$

Where:

- $AV_{i,r}$ = Sum of the AVs of rookies at position i on a team.
- $S_{i,r}$ = Salary cap hit as a percentage of the total cap for rookies at position i on a team.

Similarly, each regression takes the following form for each team to predict veteran AVs.

$$AV_{i,v} \sim \alpha_0^{i,v} + \alpha_1^{i,v} \log\left(1 + S_{i,v}\right) \tag{3.2}$$

Where:

- $AV_{i,v}$ = Sum of the AVs of veterans at position i on a team.

- $S_{i,v}$ = Salary cap hit as a percentage of the total cap for veterans at position i on a team.

We include salary cap hit as a predictor since we believe free agency markets should be fairly efficient and teams pay players commensurate to their skill level. Including salary cap hit as a predictor allows us to understand how the distribution of the salary cap impacts team performance. We note that we use salary cap hit as a percentage of total cap in order to normalize for salaries across different seasons. The salary cap has increased from $120,375,000 (in 2011) to 208,200,000 (in 2022) over the course of time period we have data. Thus, it is not fair to compare the dollar amount of salary for a player in 2011 to one in 2022, but we can compare the salary as a percentage of the salary cap.

3.2.3 Computing *Effective Salary*

In order to account for the differences in salary between rookies and veterans we created the notion of *effective salary*. *Effective salary* is the salary that the team would be paying for the level (AV) of players they have, if all the players were on veteran contracts. *Effective salary* enables us to model AV in way that it makes sense to answer the question of how to allocate capital to maximize points per game. More specifically, when modelling player performance based on salary, it prevents our model from being biased by whether players are under rookie or veteran contracts. In order to compute the *effective salary* for a team at position i we solve the following expression.

$$AV_{i,v} = AV_{i,r}$$

Which based on equations 3.1 and 3.2 we can solve for $S_{i,v}$ in terms of $S_{i,r}$. This yields the following expression for the veteran salary at position i we would pay for a player of equal AV.

$$S_{i,v} = \exp\left(\frac{\alpha_0^{i,r} - \alpha_0^{i,v}}{\alpha_1^{i,v}}\right) \cdot (1 + S_{i,r})^{\alpha_1^{i,r}/\alpha_1^{i,v}} \tag{3.3}$$

This expression tells us how much we expect a player at a certain position on a rookie contract would be paid if they were a veteran, assuming they perform as expected based on their salary. Of course, there is a lot of noise in this measure since contracts are not perfectly efficient, but we should expect with enough data we can estimate this relationship with a satisfactory level of accuracy. Thus, we can define a function $f_{i,r}$ for each position i to convert rookie salaries to *effective salaries*. We note that we ensure that this function returns 0 if the rookie salary at position i, $S_{i,r} = 0$ since if a team has no rookies they should not add to the *effective salaries*.

$$f_{i,r}(S_{i,r}) = \begin{cases} \exp\left(\frac{\alpha_0^{i,r} - \alpha_0^{i,v}}{\alpha_1^{i,v}}\right) \cdot (1 + S_{i,r})^{\alpha_1^{i,r}/\alpha_1^{i,v}}, & \text{for } S_{i,r} > 0 \\ 0, & \text{for } S_{i,r} = 0 \end{cases} \tag{3.4}$$

We can similarly define a function $f_{i,v}$ for each position i to convert veteran salaries to *effective salaries*. By the definition of *effective salary* we can notice that $f_{i,v}$ is simply be the identity function for all positions. It is the identity function since we assume free agency markets are efficient, and therefore, players on veteran contracts are paid a fair market price for their skill level.

$$f_{i,v}(S_{i,v}) = S_{i,v} \tag{3.5}$$

Now for each position i we can define our *effective salary*, $S_{i,e}$, to be the rookie contract salary, $S_{i,r}$, converted to an *effective salary* plus the veteran contract salary, $S_{i,v}$, converted to an *effective salary* according to functions 3.4 and 3.5, respectively.

$$S_{i,e} = f_{i,v}(S_{i,v}) + f_{i,r}(S_{i,r}) = S_{i,v} + \exp\left(\frac{\alpha_0^{i,r} - \alpha_0^{i,v}}{\alpha_1^{i,v}}\right) \cdot (1 + S_{i,r})^{\alpha_1^{i,r}/\alpha_1^{i,v}} \tag{3.6}$$

This one number, *effective salary*, can now be used to predict a team's AV at each position.

3.2.4 Digression on Use of *Effective Salary* to Evaluate Which Position to Draft

Additionally, *effective salary* provides a framework for considering which positions are most valuable to draft. Positions which have much higher *effective salaries* for rookies for the same level of pay may be the most valuable to draft. In the NFL draft there is a set salary for players based on their draft position. For a given pick a team knows the salary they must pay that player, S. With team needs set aside, a team may choose to adopt the strategy of drafting a player at position x such that:

$$x = \arg\max_i f_{i,r}(S)$$

The reason that we position x may be the best to draft is that it is the position that we expect the team would need to pay the most money to sign a veteran player with the same expected performance (AV). This follows from the definition of *effective salary* as the salary the team would need to pay a player if they were on a veteran contract. Thus, the team saves cap space by signing a player at this position to a rookie contract, and signing players at other positions to veteran contracts. This analysis is incomplete since it ignores many key variables including individual team needs and the win (or points per game) contribution of each position. Nonetheless, it provides an interesting starting point for further research evaluating the value of drafting different positions at different points in the draft. This approach is not implemented in this paper, but it is briefly discussed in the future work section 5.1.

3.2.5 *Effective Salary* to Predict AV

We now use this new measure of *effective salary* to predict AV. More specifically, the model assumes a logarithmic relationship between the *effective salary* paid to players at position i, $S_{i,e}$, and the AV at position i, AV_i. The model is of the following form.

$$AV_i \sim \alpha_0^{i,e} + \alpha_1^{i,e} \log(1 + S_{i,e}) \tag{3.7}$$

We solve for the parameters $\alpha_0^{i,e}$ and $\alpha_1^{i,e}$ using an ordinary least squares regression.

3.3 Offensive Performance Model

In this section we explain our model to predict a team's offensive performance. We use points per game (PPG) as the metric for offensive performance we want to predict. We already have a model to predict player performance, outlined in section 3.2. In this section we model PPG as a function of player performance at each offensive position.

To remove a level of complexity from our model, we make the assumption that each team utilizes players at a certain position in a way that if they have the same skill level they will have an equal contribution to wins. This is a simplification since different teams employ different offensive strategies with some more heavily focused on running and other focused more heavily on passing. In a running focused offense, the quarterback or wide receivers may have a less important role than running backs or offensive lineman. This assumption enables us to develop one model for points per game that can be applied to each team.

We use the sum of the players' AV at each position to predict PPG. We use AV_x to denote the sum of the AV of players at position x. We consider the sum since considering each player individually complicates the problem since we must differentiate between at team with two running backs earning an AV of 5 each (for a cumulative AV of 10) and a team with one running back earning an AV of 10. Taking this level of complexity out of our model help us avoid over-fitting. We note that we include cross terms of the form $\sqrt{AV_x \cdot AV_y}$ to capture the interactions between player performance and the contribution to PPG at positions x and y. We include the square root to keep the units of our regression consistent and to control for the correlation between AV at various positions. These cross terms are an important aspect of our model that enable us to quantify the interactions between positions and develop a better understanding of how adding a player at a certain position impacts teams differently. A model without these terms would suggest that adding a star wide receiver has the same impact on any offense. Our model can understand that a team having a great quarterback changes the PPG contribution of that receiver.

Our model is of the form:

$$PPG \sim \beta_0 + \sum_{i \in \mathrm{P}} \beta_i AV_i + \sum_{i,j \in G} \beta_{i,j} \sqrt{AV_i AV_j} \tag{3.8}$$

Where:

- P = List of all offensive positions.
- G = All pairs of positions we include cross terms for. The positions included here are discussed in the results section.

We mainly use our priors about football to determine the position pairs in G. We include positions that we expect to depend on each other's performance the most, while not including too many terms to avoid over-fitting.

One note is that we only include the offensive positions in P in regression 3.8 since we want to model and optimize points per game which we assume is determined by offensive player performance. We include kicker as an offensive positions since they score points by kicking field goals and extra points. Having a better kicker increases to a team's average points per game since they are able to make kicks more consistently and from further away. The rest of the positions we include in P are the traditional positions on a football offense.

3.4 Finding Optimal Cap Allocation

In the NFL, a front office has many players already under contracts, so they must decide where to allocate capital, given the other players they have on the roster. In this section we describe two methods to determine how to allocate capital in order to maximize a team's points per game. The first approach is greedy approach which finds the position where the next marginal dollar provides the greatest expected increase in points per game given the other players already on the team. The second approach finds the optimal allocation for a team starting from scratch (i.e. with no players already under contract) by solving a non-linear optimization problem. This approach could be extended to solve where a team should allocate the remainder of their capital given some partial allocation of players already under contract but is not discussed in this paper. It is important to note that these computations focus only on offensive positions since we focus on maximizing points per game.

3.4.1 Finding the Optimal Position to Allocate Next Dollar: A Greedy Approach

We first took a greedy approach to capital allocation. A greedy algorithm is one which make the best decision at the current state without regard for whether this will yield an optimal solution later on. The approach we describe in this section is greedy since we find the position where adding a dollar of salary yields the largest expected improvement in points per game without regard for how this will impact the efficacy all future dollars.

We can use the results of the regression from *effective salary* to AV (Equation 3.7) to determine the player AV we can expect for a given salary. Then, we can use the results of the regression to predict points per game using player AV (equation 3.8) to determine the point per game we add by increasing AV at each position. We focus on veteran salaries at each position x ($AV_{x,v}$) since since we assume we must sign players through free agency and not the draft. Mathematically we want to solve for the position x that maximizes the following expression.

$$\arg\max_{x} \frac{\partial PPG}{\partial S_{x,v}} \tag{3.9}$$

If we want to find the place where if we allocate capital, we can expect the largest improvement in PPG we can use calculus. We find the position where the partial derivative of expected PPG is the largest w.r.t the AV at that position. More specifically, we want to find the position x that maximizes $\frac{\partial PPG}{\partial S_{x,v}}$ where $S_{x,v}$ is the salary of veterans at position x. Again, we notice that we focus on the acquisition of veteran players since players must be acquired through free agency.

$$\arg\max_x \frac{\partial PPG}{\partial S_{x,v}} = \arg\max_x \frac{\partial PPG}{\partial AV_x} \cdot \frac{\partial AV_x}{\partial S_{x,e}} \cdot \frac{\partial S_{x,e}}{\partial S_{x,v}} \tag{3.10}$$

We can first note the following based on equation 3.6.

$$\frac{\partial S_{x,e}}{\partial S_{x,v}} = 1$$

We can now compute $\frac{\partial PPG}{\partial AV_x}$ where AV_x is the predicted AV at position x.

$$\begin{aligned}\frac{\partial PPG}{\partial AV_x} &= \arg\max_x \frac{\partial}{\partial AV_x}\left[\beta_0 + \sum_{i \in \mathrm{P}} \beta_i AV_i + \sum_{i,j \in G} \beta_{i,j}\sqrt{AV_i AV_j}\right] \\ &= \beta_x + \frac{1}{2}\sum_{y:x,y \in G} \beta_{x,y}\sqrt{\frac{AV_y}{AV_x}}\end{aligned}$$

Now we can compute $\frac{\partial AV_x}{\partial S_{x,v}}$ utilizing regression equation 3.2. We also recall that the salary paid to position i, $S_i = S_{i,r} + S_{i,v}$.

$$\begin{aligned}\frac{\partial AV_x}{\partial S_{x,e}} &= \frac{\partial}{\partial S_{x,e}}[AV_x] \\ &= \frac{\partial}{\partial S_{x,e}}[\alpha_0^{x,e} + \alpha_1^{x,e}\log(1 + S_{x,e})] \\ &= \frac{\alpha_1^{x,e}}{1 + S_{x,e}}\end{aligned}$$

Finally we can write an explicit expression for $\frac{\partial PPG}{\partial S_{x,v}}$ by substituting the computed quantities in equation 3.10.

$$\frac{\partial PPG}{\partial S_{x,v}} = \frac{\alpha_1^{x,e}}{1+S_{x,e}} \cdot \left(\beta_x + \frac{1}{2} \sum_{y:x,y \in G} \beta_{x,y} \sqrt{\frac{AV_y}{AV_x}} \right) \tag{3.11}$$

We can recall that we model the AV at position i, AV_i, can be modeled using the *effective salary* at position i, $S_{i,e}$, according to regression equation 3.7.

$$\frac{\partial PPG}{\partial S_{x,v}} = \frac{\alpha_1^{x,e}}{1+S_{x,e}} \cdot \left(\beta_x + \frac{1}{2} \sum_{y:x,y \in G} \beta_{x,y} \sqrt{\frac{\alpha_0^{y,e} + \alpha_1^{y,e} \log\left(1+S_{y,e}\right)}{\alpha_0^{x,e} + \alpha_1^{x,e} \log\left(1+S_{x,e}\right)}} \right) \tag{3.12}$$

Now we have an explicit expression for $\frac{\partial PPG}{\partial S_{x,v}}$ in terms of *effective salaries.* We can recall by equation 3.6 that the *effective salary* at position x is a function of both $S_{x,r}$ and $S_{x,v}$ the salary given to rookies and veterans, respectively at position x. It also depends on the salaries given to players at positions we consider related since we sum over all y such that $x, y \in G$. We can recall that G is the set of all of our position pairing which is first introduced in equation 3.8. These positions are the ones that we believe interact the most with position x, and thus have the largest effect on position x's contribution to points per game.

This expression is useful because it can solve the problem of where a team should allocate their next dollar of salary cap in order to maximize their points per game (PPG). As stated in equation 3.9 we can find the position x that maximizes $\frac{\partial PPG}{\partial S_{x,v}}$ for which we now have an explicit expression.

3.4.2 Finding the Optimal Allocation: A Non-linear Optimization

The greedy approach described in the previous section finds locally optimal decisions, but not necessarily globally optimal solutions. This is especially important if a team has a large amount of capital to allocate. To find the optimal solution for where a team should allocate it capital, whether it be from scratch or from some partial allocation, we must solve a non-linear optimization. In this section we outline the approach to solve this optimization from scratch.

We solve an optimization where our objective is points per game, and we are constrained by the total *effective salary* cap we can spend among the offensive positions. We also ensure we maintain the relationship between *effective salary* and AV at each positions and the relationship between AV at each position and points per game. The optimization below can be used to find the cap allocation that maximizes the expected points per game based on our model.

$$\max_{S_e} \quad \beta_0 + \sum_{i \in \mathrm{P}} \beta_i AV_i + \sum_{i,j \in G} \beta_{i,j} \sqrt{AV_i AV_j} \tag{3.13a}$$

$$\text{s.t.} \quad AV_i = \alpha_0^{i,e} + \alpha_1^{i,e} \log\left(1 + S_{i,e}\right) \tag{3.13b}$$

$$S_e^T \vec{1} \leq \text{MAX\_TOTAL\_CAP\_PERCENTAGE} \tag{3.13c}$$

$$S \geq \vec{1} \cdot \text{MIN\_POSITION\_CAP\_PERCENTAGE} \tag{3.13d}$$

$$AV_{OL} = AV_{LT} + AV_{RT} + AV_G + AV_C \tag{3.13e}$$

The variables used in the above optimization are defined below.

- S_e is an array of the *effective salary* given to each position as a percentage of the total cap. $S_{i,e}$ is the *effective salary* given to position i. The is a vector with a length equal to the number of offensive positions which is 9 in our analysis.

- AV_i is the AV we expect at position i based on the *effective salary*.

- MAX_TOTAL_CAP_PERCENTAGE is the maximum total cap percentage (in units of *effective salary*) that we are able to spend on the offense. We should expect this number to be close to 50, since teams tend to spend about half their cap space on the offense. It could even be greater than 50 since we are dealing with *effective salary*, and rookies can have an *effective salary* higher than their actual cap hit.

- MIN_POSITION_CAP_PERCENTAGE is the minimum *effective salary* cap hit we can assign to a position. We can assign this value as the minimum of the veteran minimum contract and rookie salary minimum converted to an *effective salary*. We let this number be 0.5 in our analysis.

- The parameters $\beta_0, \beta_i, \beta_{i,j}$ are learned through regression 3.8 for all values of i and relevant values of j.
- The parameters $\alpha_0^{i,e}, \alpha_1^{i,e}$ are learned through regression 3.7 for all values of i and relevant values of j.

Now we explain the term we maximize in equation 3.13a, which is equivalent to maximizing our points per game. This term comes from the regression from AV as each position to points per game (PPG) in equation 3.8. We optimize our points per game based on the AV contributed by players at each position. We maximize this equation with respect to the *effective salary* paid to players at each position, since AV is constrained to be a function of the *effective salary*.

Now we explain each of the constraints in the optimization.

- Constraint 3.13b ensures that we maintain the relationship between the AV at position i, AV_i, and the *effective salary* at position i, $S_{i,e}$, as governed by regression equation 3.7.
- Constraint 3.13c ensures we keep our total effective cap spending below, or equal to, the limit we set with MAX_TOTAL_CAP_PERCENTAGE.
- Constraint 3.13d ensures we keep out cap spending at each position above, or equal to, the lower bound we set with MIN_POSITION_CAP_PERCENTAGE.
- Constraint 3.13e ensures the offensive line (OL) AV is the sum of the AV at each of the OL positions. This is an artifact of how we compute offensive line AV, as the sum of the AVs at each of the offensive line positions rather than predict it based on *effective salary*.

We solve this optimization using Gurobi, an optimization software [4].

Chapter 4

Results

In this section we discuss the results of implementing the methodology outlined in the previous section.

4.1 Player Performance Model

As discussed in section 3.2 we develop a model to predict player performance based on salary. In this section we outline these results.

4.1.1 Predicting AV: Breakdown by Rookie Contract Status

We begin by breaking down our prediction of AV by rookie contracts and non-rookie contracts according to regression equations 3.1 and 3.2, respectively. The results of each of these regressions are shown in tables A.1 and A.2, respectively. We can notice that for all positions, i, the coefficient term $\alpha_1^{i,v}$ is larger than $\alpha_1^{i,r}$. This relationship indicates that for all positions adding cap percentage to players on rookie contracts results in a larger expected increase in AV contribution than if that cap percentage were given to players on veteran contracts. This advantage of signing rookies is confirms teams' value placed on draft picks since the picks allow them the opportunities to sign players that will on average play at a higher level than a veteran with the same salary.

4.1.2 Computing *Effective Salary*

We can use the results of the regressions to predict AV by rookie contract status to compute *effective salary*, as discussed in section 3.2.3. We can recall that *effective salary* is defined as the salary a team would pay players if they were on a veteran contract. For veteran players this is just their cap hit since they are already on veteran contracts while for rookies it is the amount we expect they would be paid if they were on a veteran contract. Plots of rookie contract cap hit ($S_{i,r}$) vs rookie contract *effective salaries* ($f_{i,r}(S_{i,r})$) are shown in figure B-1 where rookie contract *effective salary* is the amount the rookie player would be paid if they were on a veteran contract. These plots show how the function $f_{i,r}(\cdot)$ differs for various positions. We can notice that these lines show that $f_{i,r}(S_{i,r}) > S_{i,r}$ which confirms that rookies are paid below their market value if they were veterans.

For most positions it appears that veterans are paid at approximately twice the level of rookies meaning $f_{i,r}(S_i, r) \approx 2S_{i,r}$. Additionally, the curve seems to be well approximated by a linear function meaning this relationship does not change depending on the value of the rookie contract. One position where this is not the case is center (C) where the curve appears to have a more exponential than linear shape. Additionally, if we look at a center on a rookie contract reviving 1% of the cap we can tell by looking at the curve (or substituting into equation 3.6) that we expect that player would receive over 3% of the cap if they were on a veteran contract. This suggests that center may be one of the more valuable positions to draft based on our model. It is important to note that this could also simply be due to our limited (11 seasons) data set containing many centers who perform exceptionally well on rookie contracts which inflates the coefficient $\alpha_1^{C,r}$ in regression 3.1.

4.1.3 *Effective Salary* to Predict AV

Effective salary can be used to model AV, as described in section 3.2.5. According to regression equation 3.7 we fit parameters $\alpha_0^{i,e}$ and $\alpha_1^{i,e}$ for each offensive position, i. The results of the regression for each offensive position are shown in table 4.1. The

Position	$\alpha_0^{i,e}$	$\alpha_1^{i,e}$	R-Squared
QB	7.42	2.1	0.06
RB	9.12	2.58	0.04
WR	6.67	5.89	0.13
TE	2.17	2.57	0.09
RT	1.18	5.19	0.39
LT	1.45	4.15	0.44
G	3.01	7.17	0.19
C	4.07	3.33	0.18
K	1.88	1.34	0.12

Table 4.1: *Effective Salary* to AV Regressions for Each Offensive Position. Predict cumulative AV using *effective salary*. Parameters fit according to regression equation 3.7. The column labeled $\alpha_1^{i,e}$ tells us, for each position i, the added AV we expect for a one unit increase in the log of the *effective salary* cap hit at that position, denoted by $\log(S_{i,e})$.

relationship between effective cap hit and AV for each position with our modeled (log) relationship between effective cap hit and AV is shown in figure 4-1. Additionally figure B-2 shows the relationship between our predicted AV and the realized AV. Based on these figures we can notice that the model accuracy differs for each position's model. Our best model, the right tackle (LT) model, has an R-squared value of 0.44 while our worst model, the running back (RB), model has an R-squared values of 0.04. These low R-squared values are unsurprising since player salary is just one variable that can be used to explain player performance. One important aspect of player performance is the influence of other players which is not included in this model. This may explain why running back, a position that relies heavily on the offensive line, yields the model with lowest R-squared value. The correlation of player AV between 2 consecutive seasons is 0.65, so even if one knew the AV of a player in the last year, there is still a lot of noise in predicting the AV in this season. It is important to keep in mind that the purpose of these models is to understand how player performance (AV) scales with salary cap hit on average rather than to make accurate predictions.

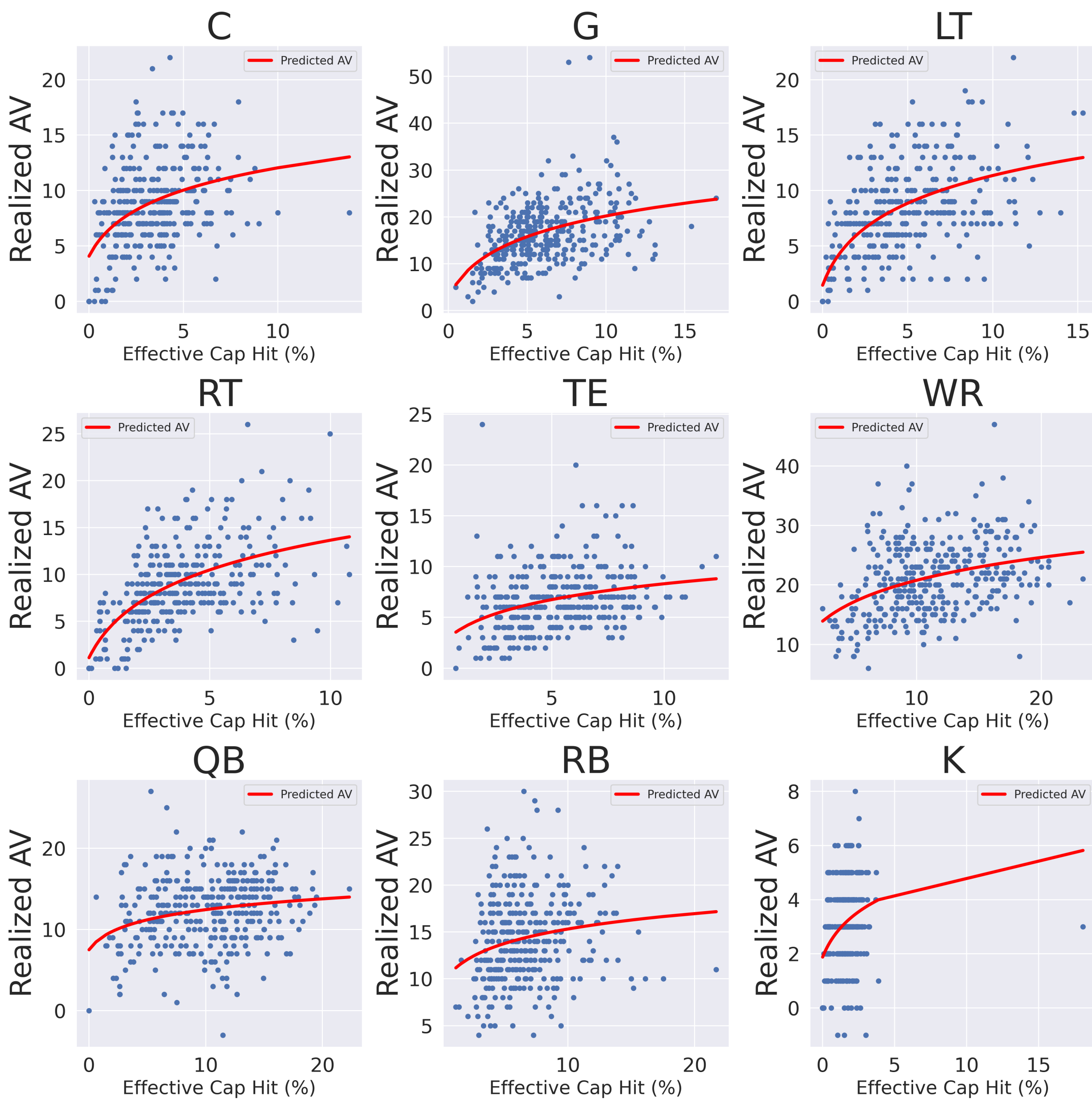

Figure 4-1: Effective Cap Hit (as a percentage of total salary cap) vs Realized AV for each position where each data point is one position, on one team, in one season. The red line is predicted AV based on the regression equation 3.7. This plot is used to justify the use of the log of effective cap hit to predict AV. The plots show that the relationship between effective cap hit and AV is sub-linear due to diminishing marginal returns of allocating capital to a certain position. It is also apparent that for certain positions, such as right tackle (RT) and center (C), the relationship between effective cap hit and AV is stronger than for other positions, such as running back (RB) and kicker (K). This difference in the strength of the relationship is reflected in the difference in r-squared values in table 4.1.

4.2 Offensive Performance Model

In this section we discuss the results of our model to predict offensive performance based on the player performance at each offensive position. We use points per game (PPG) as our measure of offensive performance and cumulative AV by position as our measure for player performance.

A vital aspect of this model is the selection of the positions which are included G, the pairs of positions for which cross terms are included. We can notice that G is included in our regression from positional AV to points per game shown in equation 3.8. We include positions in G that impact each others performance the most based on our prior beliefs and understanding about how players interact on a football team. Since this decision is somewhat arbitrary we include results for multiple assignment of G which we denote with $G_1, G_2, \ldots$. The different values of G are as follows.

- $G_0 = \{\} = \emptyset$
- $G_1 = \{(QB, WR), (RB, OL)\}$
- $G_2 = \{(QB, WR), (RB, OL), (QB, TE), (LT, OL)\}$

One note is that we define offensive line (OL) AV to be the sum of the AV at each of the offensive line positions which include left tackle (LT), right tackle (RT), guard (G), and center (C).

The assignment G_0 is the case where we ignore all interactions between players at different positions.

The assignment G_1 includes the relationships between QB/WR and RB/OL. The reason these relationships are included is because they rely heavily on each-others performance. Quarterbacks depend on wide receivers to catch their passes while wide receivers depend on quarterbacks to pass them the ball. Similarly, running backs rely on offensive lineman to block defenders to they can run the ball effectively while offensive lineman rely on running backs to read their blocks and run through the correct openings in the defensive front. We include the relationship between offensive line and running back rather than the relationships between running back and each of the offensive line positions since this limits the number of predictors we use which helps avoid over-fitting.

The assignment $G = G_2$ includes the relationships between QB/WR, RB/OL, QB/TE, and LT/OL. The relationships between QB/WR and RB/OL was already included and discussed for G_1. We also include the relationship between QB/TE and LT/OL in this regression for G_2. In a similar way that quarterbacks and wide receivers rely on each other, tight ends rely on quarterbacks to pass them the ball while quarterback rely on tight ends to catch their passes. We include the relationship between left tackle and rest of the offensive line since we believe there may be differing strategies of either building up the entire offensive line or focusing on hiring an elite left tackle. The left tackle is traditionally believed to be the most important offensive lineman, so that is why we chose to include this term.

We create separate train and test sets in order to determine if we over-fit our model. We randomly select 20% of our data set to be in our test set where each data point s one team in one season. The results of each of these regressions including the learned coefficients and the train and test r-squared value are shown in table 4.2. Additionally, scatter plots of our predicted points per game (PPG) versus realized points per game are shown in figures B-3, B-4, and B-5 for G_0, G_1, and G_2, respectively.

We can determine that it seems that we slightly over fit our model in each assignment G_0, G_1, and G_2. This over fitting however was marginal since the r-squared values are relatively close (within 3%) for all of G_0, G_1, and G_2. Additionally it seems adding in more cross terms (terms of the form $\sqrt{AV_i AV_j}$) did not increase the amount of over fitting since the gap between the train and testing r-squared remained relatively constant between G_0, G_1, and G_2. Based on the r-squared values of each regression the G_2 model, which included the most cross terms, is the best model since it had the highest train and test r-squared value. This model being the best tells us that understand the relationship and interactions between position is important to accurately predicting a team's points per game. Prior literature had not addressed these interactions, so this is an exciting insight.

The scatter plots shown in figures B-3, B-4, and B-5 demonstrate the accuracy

	G_0		G_1		G_2	
Predictor	$\boldsymbol{\beta}$	**Significance**	$\boldsymbol{\beta}$	**Significance**	$\boldsymbol{\beta}$	**Significance**
Constant	5.56	0.00	5.24	0.00	5.18	0.00
AV_{QB}	0.23	0.00	0.48	0.00	0.54	0.00
AV_{RB}	0.18	0.00	0.85	0.00	0.84	0.00
AV_{WR}	0.23	0.00	0.36	0.00	0.33	0.00
AV_{TE}	0.15	0.00	0.18	0.00	0.40	0.08
AV_{RT}	0.14	0.00	0.36	0.00	0.37	0.00
AV_{LT}	0.11	0.00	0.33	0.00	0.44	0.00
AV_G	0.16	0.00	0.38	0.00	0.39	0.00
AV_C	0.14	0.00	0.36	0.00	0.36	0.00
AV_K	0.07	0.34	0.07	0.33	0.07	0.34
$\sqrt{AV_{QB}AV_{WR}}$	-	-	-0.35	0.08	-0.29	0.17
$\sqrt{AV_{RB}AV_{OL}}$	-	-	-0.78	0.01	-0.77	0.01
$\sqrt{AV_{QB}AV_{TE}}$	-	-	-	-	-0.31	0.33
$\sqrt{AV_{LT}AV_{OL}}$	-	-	-	-	-0.06	0.23
Train R-Squared	-	0.80	-	0.81	-	0.81
Test R-Squared	-	0.76	-	0.77	-	0.78

Table 4.2: AV to Points Per Game (PPG) Regression Results. Predict PPG using offensive player AV according to regression equation 3.8 for $G \in \{G_0, G_1, G_2\}$. The table shows that the models using G_0, G_1, and G_2 are increasingly accurate, without over-fitting, since both the train and test r-squared values are increasing.

of the model. This accuracy is the result of the strong role player performance (AV) plays in team offensive success (PPG). Based on the r-squared values, the G_2 model is the best, followed by G_1, and then G_0 is the worst, but this is difficult to tell based on the scatter plots. The r-squared values are also quite close. One reason why we still want to include the cross terms in our model despite it only adding a small amount of predictive accuracy is that it can enable us to better understand what positions are best for each team, rather than to make general conclusions for all teams. For example, including the cross term between quarterback and wide receiver can enables us to understand how a certain wide receiver has have a different impact on each team depending on that team's quarterback(s).

The main takeaways from the G_0 regression is that it seems that adding AV at quarterback is the most valuable, followed by wide receiver and guard. This conclusion follows from those positions having the largest coefficients (β_i terms) in the regression. This confirms a common belief that, in the NFL, the quarterback is the most important offensive position. Wide receivers are also a highly valued position so that result is unsurprising. The result that guard is one of the most valuable positions is somewhat unexpected. One reason for this result could be that since there are two guards on the offense (left guard and right guard), the sum of their performance is less noisy and a better representation of the offensive performance than all the other offensive line positions which only have one player on the offense at those positions. The data set we have access to does not differentiate between left and right guard, but with access to that data it would be possible to determine if the grouping of these positions is impacting the results.

In the regression using G_1 we can notice that the coefficients for each of the positions (β_i) is positive, but the coefficients are negative for the cross terms ($\beta_{i,j}$). This is an interesting insight that tells us about the value of adding talent at different positions based on the level of talent at other positions. The coefficient $\beta_{QB,WR} < 0$ which means our model believes that adding AV to wide receivers is less valuable if you have a better quarterback, and vice versa. Similarly, since the coefficient $\beta_{RB,OL} < 0$, adding AV to running backs is less valuable if you already have a better offensive line.

The G_2 regression has a similar result to the G_1 regression where the coefficients for each of the individual positions is positive while the coefficients are negative for the cross terms. The same reasoning of diminishing marginal returns applies here. An intuitive argument why this is the case is that if you already have incredible wide receivers you do not need a great quarterback to throw them the ball. A similar argument can apply for the other position parings.

4.3 Finding the Optimal Allocation

This section outlines two methods of finding the optimal capital allocation. The first is a greedy approach that finds where the next marginal dollar adds the most points per game given the players already on a roster. The second utilizes a non-linear optimization to find an optimal capital allocation starting from an empty roster. Both

Position	G_0	G_1	G_2
QB	2	13	9
RB	1	23	22
WR	15	240	242
TE	4	76	79
RT	107	0	0
LT	52	0	0
G	132	0	0
C	39	0	0
K	0	0	0
Count	352	352	352

Table 4.3: Count for Each Position of Number of Times it is the Optimal Position to Allocate Capital Towards for G_1, G_2, and G_3. For each team in each season we find the position that allocating the next marginal dollar towards maximizes their points per game based on the salary they are paying rookies and veterans at each positions. This corresponds with maximizing the expression $\frac{\partial PPG}{\partial S_{x,v}}$ where $S_{x,v}$ is the salary given to veterans at position x, as described in equation 3.12.

of these approaches for finding the optimal capital allocation utilize the results from the player performance model and offensive performance model.

4.3.1 Finding the Optimal Position to Allocate the Next Dollar: A Greedy Approach

In order to find the optimal position to allocate the next dollar towards we utilize calculus. As discussed in section 3.4.1, we find the position x that maximizes the quantity $\frac{\partial PPG}{\partial S_{x,v}}$. We can utilize the expression we derive in equation 3.12 to compute this quantity for every offensive position on every team in every season. This expression depends on the position parings in G, so we repeat this calculation for each value of G which include G_0, G_1, G_2. We then select the position x for which $\frac{\partial PPG}{\partial S_{x,v}}$ for each team in each season. We can solve this problem in two ways. The count of how many times each position was the best position to allocate the next dollar to is shown in table 4.3.

Table 4.3 shows that the position that maximizes the expected improvement in points per game depends heavily on which position pairing we consider from

G_0, G_1, G_2. The table shows that for G_0 the best positions were guard 132 times, right tackle 107 times, left tackle 52 times, and center 39 times. The model believes one of the offensive line positions was the most valuable to allocate capital towards in 330 out of 352 samples of different teams in different seasons. Since G_0 does not consider interactions between positions it appears our model believes the offensive line is the most valuable to add capital towards when ignoring interactions between position groups.

The results of the best position to allocate capital towards are very similar for the position groupings G_1 and G_2. In contrast to the results for G_0 where we include no position pairings, for G_1 and G_2 the model never believes that it is optimal to allocate capital to an offensive line position. Alternatively, it finds that wide receiver is the best position to allocate capital towards 240 times and 242 times for G_1 and G_2, respectively. The model also finds that tight end is the best position to allocate capital towards 76 times and 79 times for G_1 and G_2, respectively. Next, the model finds that running back is the best position to allocate capital towards 23 times and 22 times for G_1 and G_2, respectively. Lastly, the model finds that quarterback is the best position to allocate capital towards 23 times and 22 times for G_1 and G_2, respectively. Thus, it is apparent that when the model uses position pairing G_1 and G_2 it suggests allocating the next dollar towards skill positions, including wide receiver, running back, quarterback, and tight end, is more valuable than the offensive line.

It is clear the the selection of the position pairings G included when training the AV to points per game model is vital to the overall results of our model. It would be interesting to explore the positions to include in a more systematic way.

4.3.2 Finding the Optimal Allocation: A Non-linear Optimization

The results of using a non-linear optimization to find the optimal salary cap allocation are discussed in this section. We solve the optimization described in equation 3.13. We let MIN_POSITION_CAP_PERCENTAGE = 0.5 since this approximately cor-

Position	G_0	G_1	G_2	G_0, G_1, G_2 Average	League Average
QB (%)	4.1	4.0	2.4	3.5	9.7
RB (%)	4.1	0.5	0.5	1.7	6.4
WR (%)	13.3	12.1	12.0	12.5	10.8
TE (%)	4.1	2.9	4.35	3.8	4.9
RT (%)	6.6	8.6	8.2	7.8	3.5
LT (%)	3.9	5.1	6.3	5.1	4.3
G (%)	11.5	13.21	12.4	12.4	5.8
C (%)	3.88	5.12	4.4	4.5	3.2
K (%)	0.5	0.5	0.5	0.5	1.4
Total Cap %	52	52	52	52	52

Table 4.4: Optimal Allocation of *Effective Salary* Cap Among Offensive Positions. For each of G_1, G_2, and G_3 we find the allocation that maximizes their points per game based on optimization equation 3.13. The table shows that on average our models allocate above the league average effective cap percentage to all the offensive line positions (C, G, RT, and LT) and wide receiver (WR). For the other positions, including quarterback (QB), running back (RB), tight end (TE), and kicker (K), our models tend to allocate less than the league average effective cap percentage.

responds to the percentage of the salary cap that is the veteran minimum, so it corresponds with the minimum a team would be able to spend a position and still have a player at the position. We let the variable MAX_TOTAL_CAP_PERCENTAGE = 52. Recall that MAX_TOTAL_CAP_PERCENTAGE is in terms of *effective salary* so teams can have an *effective salary* cap hit that sums up to more than 100% if they have many rookies on their roster. Over the course of our data set which includes 11 seasons, the average *effective salary* cap allocated to offensive players was 52% of the salary cap. Thus, we choose to only allow the optimization to allocate 52% of the salary cap to the offensive positions. We note that doing this optimization in terms of *effective salary* is the same as if a team could only sign veteran players since in the absence of rookies we have that *effective salary* equals veteran salary ($S_{i,e} = S_{i,v}$).The results of the optimization for G_0, G_1, and G_2 are shown in table 4.4 along with the league average allocation.

The table shows the the optimal effective (veteran) cap allocation to quarterback ranges from 4.1% for G_0, 3.96% for G_1, to 2.42% G_2. This result is interesting since quarterback is often thought to be the most valuable position on a team, with teams paying an average of 9.7% of the *effective salary* cap percentage to quarterbacks. One explanation for our model allocating such a low percentage of the cap to quarterback is that it does not understand the causal nature of the interactions between players on a football team. For example, our model cannot understand whether the quarterback is making the other offensive players perform well, or vice versa. Since quarterback is an expensive position to obtain top players, it is unsurprising our model chooses to allocate capital to cheaper positions since it does not understand that quarterback is likely the position that influences the other positions more than they influence the quarterback and not the reverse.

Similarly our model allocates less cap percentage to running back than the league average. For G_0 our model allocates 4.1% of the salary cap to running backs, and it allocates 0.5% to running backs for G_1 and G_2. This is likely due to the coefficient associated with the $\sqrt{AV_{RB}AV_{OL}}$ term ($\beta_{RB,OL}$), which is included in G_1 and G_2, being negative. Since this coefficient is negative our model believes the return (added points per game) for adding AV at running back is inversely related to the skill level at offensive line. This relationship represents diminishing marginal returns where our model believes that there is less value in having a good running back if you already have an elite offensive line, and vice versa. Thus, since our model allocates a large portion of the cap to the offensive line positions, it chooses to allocate a small amount of capital to the running back.

Our model allocates between 13.3% and 12.05% of the cap towards wide receivers which is above the league average of 10.76%. The explanation for why we allocate more than the league average to wide receivers is similar to why we allocate more to offensive line rather than running back. The coefficient $\beta_{QB,WR}$ is negative so our model believes that there is diminishing marginal returns to adding value at quarterback if a team already has good wide receivers, and vice versa. Thus our model find a solution where we allocate above the league average to wide receivers and below the league average to quarterbacks. This is because it requires less effective cap percentage to add AV at wide receiver than quarterback. This is evident based on the results of the *effective salary* to AV regression described in regression equation

3.7 and table 4.1. This regression finds the coefficient for the receiver regression ($\alpha_1^{WR,e} \approx 5.9$) was greater than the coefficient for the quarterback regression ($\alpha_1^{QB,e} \approx 2.1$). This relationship tells us that it takes less capital to add AV at wide receiver than quarterback which makes sense because quarterback is the highest average paid position in the NFL.

Both center and left tackle are positions that our model allocates about the league average towards. Although our model tends to allocate heavily to offensive line it does not allocate as much to left tackle as guard or right tackle since left tackle is generally the highest paid offensive line position. Unsurprisingly, the results of the *effective salary* to AV regression described in regression equation 3.7 and table 4.1 show that the coefficients for center ($\alpha_1^{C,e}$) and left tackle ($\alpha_1^{LT,e}$) are the smallest two coefficients on the offensive line. Thus, we would conclude that these two positions yield the smallest AV improvement for a marginal addition of salary assuming all positions are at the same level salary.

Guard and right tackle are positions that our model chooses to allocate about double the league average percentage of the cap towards each. Our model allocates on average 12% (across G_0, G_1, G_2) of the cap to guard while the league average is 5.8% of the effective cap. Similarly, for right tackle, our model allocates an average of 7.8% of the effective cap while the league average is 3.5% of the effective cap. The reason for this is that these are the cheapest offensive line positions to add AV at based on their coefficients in the *effective salary* to AV regression shown in table 4.1. We have that the largest 3 coefficients $\alpha_1^{i,e}$ are $\alpha_1^{i,RT}, \alpha_1^{i,WR}$, and $\alpha_1^{i,G}$. Unsurprisingly, our model allocates an effective cap percentage above the league average to these positions. The coefficient for guard ($\alpha_1^{i,G}$) is the largest coefficeint, which contributes to guard being the position we exceed the league average allocation by the largest margin with our model allocating on average 12% (across G_0, G_1, G_2) of the effective cap to guard while the league average is 5.8% of the effective cap.

Our model is slightly below the league average in terms of tight end cap allocation. The league average for tight ends is 4.93% while our models allocations range from 2.91% for G_1 to 4.35% for G_2. Our model allocates the minimum of 0.5% to kickers which is below the league average of 1.4%.

Chapter 5

Conclusion

5.1 Future Work

One potential area for future work is to explore different methods for determining the position pairings, G, that we consider interactions between, as described in equation 3.8. It would be interesting to investigate systematic approaches to determine the optimal G assignments, rather than relying on priors about the function of a football offense as in this paper.

In addition, future work could incorporate the concept of *effective salary* cap to inform draft decision making. As discussed in section 3.2.4, a team could adopt a drafting policy that prioritizes selecting players at positions where they would need to pay a veteran player the largest salary to achieve the same expected performance as the drafted rookie. *Effective salary* can be used to determine this amount for each position.

Another potential area for future research is to modify the non-linear optimization described in equation 3.13 to solve for the optimal allocation given a partial allocation and the remaining salary cap available. This problem is closer to the actual problem faced by NFL teams, as they often have some players already under contract and some remaining capital they can use to sign other players. By utilizing such an algorithm, a team could allocate their remaining capital in a way that maximizes their expected points per game according to our model.

We acknowledge that considering a player's salary cap hit as a complete statistic of their compensation is an oversimplification of the highly complex and nuanced structure of NFL player contracts. Future work could focus on developing a more comprehensive approach that takes into account the intricacies of players' contracts, such as signing bonuses, incentives, and guarantees. This would provide a more accurate representation of a player's true compensation and could potentially impact the optimal allocation of salary cap resources among different positions.

5.2 Key Takeaways

This paper proposes a modelling structure that that helps teams determine the optimal allocation of capital among offensive positions. Our findings suggest that the choice of position parings, represented in G, and considered in equation 3.8, have an impact on the optimal allocation among offensive positions. According to our model, guard and right tackle are the positions where teams should allocate the most *effective salary* capital relative to the league average. This is followed by wide receiver, left tackle, and center, where our model allocates more than the league average, but by a smaller margin. Tight ends and kickers receive slightly less capital than the league average, while running backs and quarterbacks receive far below the league average. We believe that the low allocation given to quarterbacks is due to the lack of causality built into our modeling framework. Our model does not account for the fact that quarterbacks often have a significant influence on the performance of other positions, rather than the other way around. Therefore, the model allocates capital to positions where it takes less capital to add approximate value (AV), as quarterbacks are the highest-paid position in the NFL. Our model does not consider the causal nature of interactions in football, resulting in a preference to allocate capital towards positions where adding talent is cheaper. This reasoning explains why guard and right tackle receive the most capital relative to the league average since adding marginal talent to these positions costs less, as discussed in section 3.4.

One novel aspect of this research is the creation of the measure we call *effective* *salary*. We believe that *effective salary* is useful for a variety of purposes. In this research we utilize it to account for the differences between rookie and veteran contracts when modelling player performance based on salary. As discussed in the future work section 5.1 and the digression in section 3.2.4, it presents a framework for informing draft decision making. Rookie contracts are a valuable asset for teams, and *effective salary* presents a framework for understanding the how the value of rookie contracts change across different positions and pay levels.

This research proposes a model that understands the interactions between the offensive positions of a football team and determines how these relationships impact each position's contribution to offensive success, measured by points per game. Due to the complexity of our model, we need to solve a non-linear optimization problem to compute the optimal allocation of capital among offensive positions. This optimal allocation differs from past results since it is constructed using a model that accounts for the relationships between positions. We believe this model is valuable since it more accurately reflects the complex interactions between players within an NFL offense.

Appendix A

Tables

Position	$\alpha_0^{i,v}$	$\alpha_1^{i,v}$	**R-Squared**
QB	1.23	6.32	0.3
RB	0.53	9.17	0.35
WR	0.49	9.1	0.41
TE	-0.24	4.95	0.32
RT	1.34	7.19	0.36
LT	2.46	4.35	0.29
G	1.86	7.29	0.33
C	0.89	8.15	0.32
K	1.62	4.02	0.08

Table A.1: Salary to AV Rookie Regression for Each Offensive Position. Predict cumulative AV contributed by rookies using rookie cap hit. Parameters fit according to regression equation 3.1. The column labeled $\alpha_1^{i,r}$ tells us, for each position i, the added AV we expect for a one unit increase in the log of the rookie salary cap hit at that position, denoted by $\log(S_{i,r})$.

Position	$\alpha_0^{i,v}$	$\alpha_1^{i,v}$	**R-Squared**
QB	-1.01	5.3	0.48
RB	0.08	6.06	0.45
WR	-0.41	6.84	0.38
TE	-0.24	3.7	0.32
RT	1.17	4.81	0.4
LT	3.47	2.18	0.30
G	7.00	0.94	0.33
C	3.02	3.53	0.24
K	2.23	1.06	0.08

Table A.2: Salary to AV Veteran Regressions for Each Offensive Position. Predict cumulative AV contributed by rookies using veteran cap hit. Parameters fit according to regression equation 3.2. The column labeled $\alpha_1^{i,v}$ tells us, for each position i, the added AV we expect for a one unit increase in the log of the veteran salary cap hit at that position, denoted by $\log(S_{i,v})$.

Appendix B

Figures

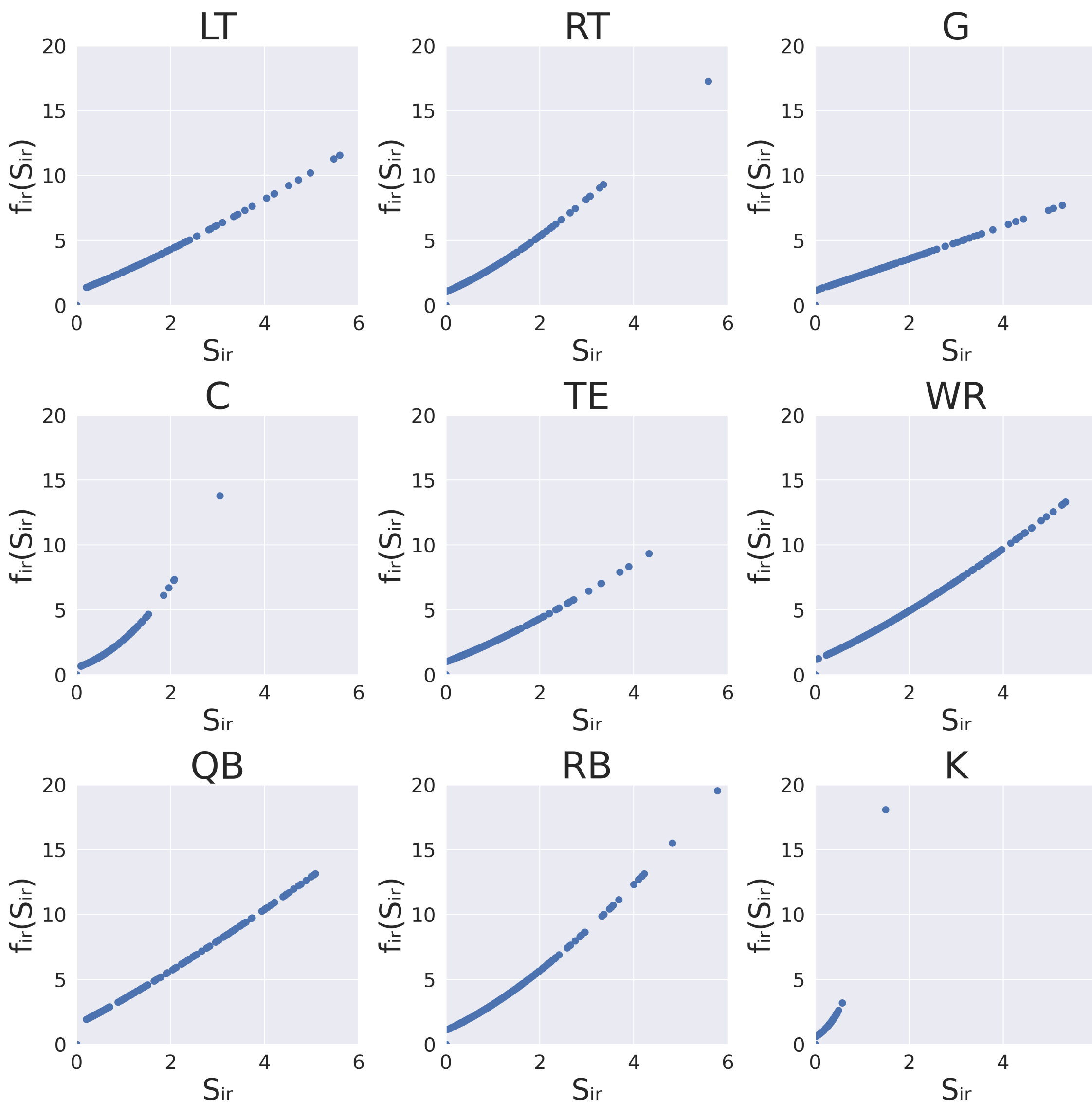

Figure B-1: Rookie salary cap percentage ($S_{i,r}$) vs rookie *effective salary* cap percentage ($f_{i,r}(S_{i,r})$) for each position i. Rookie *effective salary* cap percentage is computed based on equation 3.4 and tells us the amount we expect a player on a rookie contract would get paid if they were on a veteran contract based on their cap hit. This amount is different for each position and each plot shows this amount for a different position. All the plots use the same axis scales so it is easy to compare among different positions. One note about these plots it that they reflect that certain positions, such as kicker and center, are rarely picked early in the draft as reflected by the lack of data points with large rookie salaries (which would be to the right on the x-axis). We can conclude this because players picked early in the draft are paid a larger percentage of the salary cap on their rookie contracts.

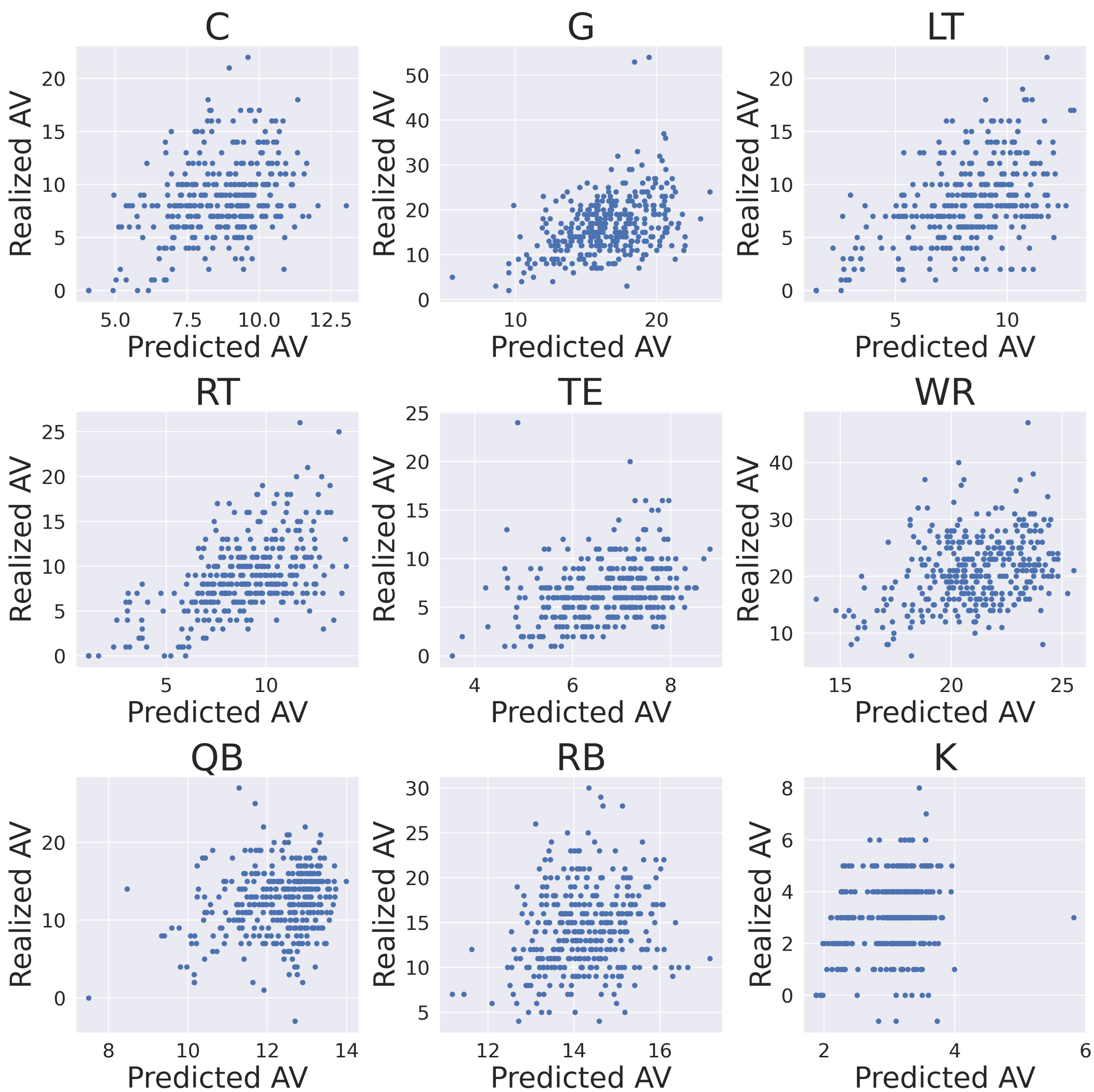

Figure B-2: Regression 3.7 results. Predicted AV vs realized AV where AV is predicted using *effective salary* for each team in each season at each position. Each data point is one team in one season's predicted AV vs realized AV at a specific position. It is apparent that our model is far more accurate at predicting AV for certain positions, such as left tackle (LT) and right tackle (RT), than other positions, such as quarterback (QB) and running-back (RB). As discussed in the results section, we hypothesize this may have to do with quarterback and running back being positions that rely on many other players performance in order to execute their jobs effectively.

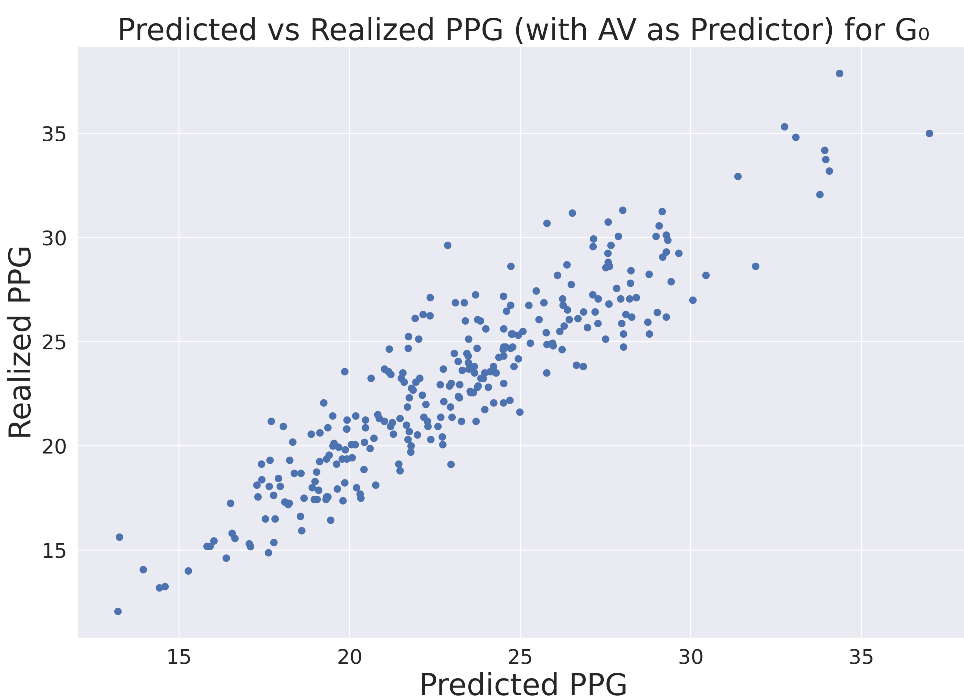

Figure B-3: Predicted PPG vs Realized PPG (Known AV) for G_0 Model. Points per game (PPG) predicted by regression 3.8 with position pairings G_0. Each data point is one team in one season. We use the players realized AV in a season to predict the points per game the team scored in that season. Visually, it is apparent the model is quite accurate, but in order to compare to other positions parings G it is useful compare the r-squared values in table 4.2. For this plot, using position pairings G_0, we achieve an r-squared value on our test set of 0.76 which is lowest out of all the models.

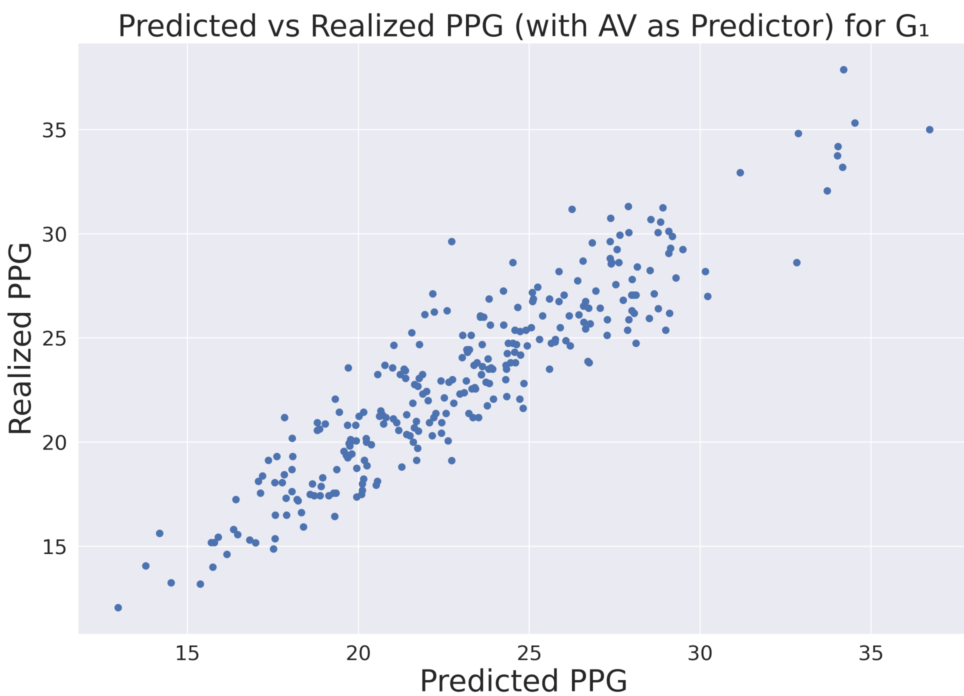

Figure B-4: Predicted PPG vs Realized PPG (Known AV) for G_1 Model. Points per game (PPG) predicted by regression 3.8 with position pairings G_1. Each data point is one team in one season. We use the players realized AV in a season to predict the points per game the team scored in that season. Visually, it is apparent the model is quite accurate, but in order to compare to other positions parings G it is useful compare the r-squared values in table 4.2. For this plot, using position pairings G_1, we achieve an r-squared value on our test set of 0.77 which is better than the model using G_0 but worse than the model using position pairings G_2.

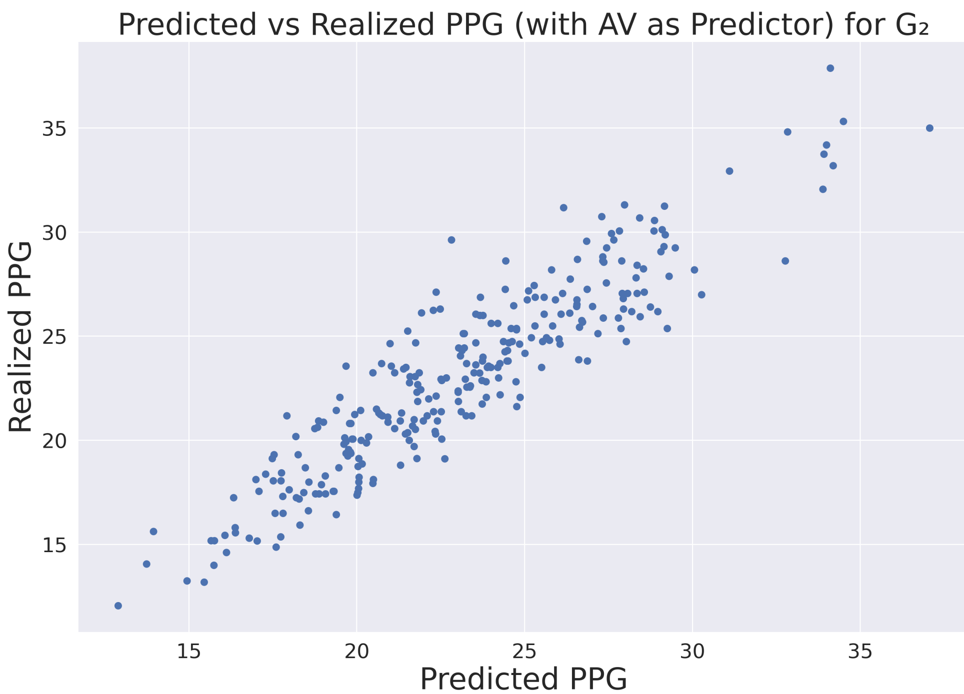

Figure B-5: Predicted PPG vs Realized PPG (Known AV) for G_2 Model. Points per game (PPG) predicted by regression 3.8 with position pairings G_2. Each data point is one team in one season. We use the players realized AV in a season to predict the points per game the team scored in that season. Visually, it is apparent the model is quite accurate, but in order to compare to other positions parings G it is useful compare the r-squared values in table 4.2. For this plot, using position pairings G_2, we achieve an r-squared value on our test set of 0.78 which is best out of all the models.

Appendix C

Software Packages

I utilized Python to perform this research. The list below is not intensive but outlines the main software packages used to perform the analysis presented in this paper.

- NumPy: Fundamental package for array computing in Python [5]
- Pandas: Powerful data structures for data analysis, time series, and statistics [11]
- Seaborn: Python visualization library [15]
- Matplotlib: Python visualization library [15]
- Statsmodels: Statistical computations and models for Python [14]
- BeautifulSoup: Screen scraping library [13]
- Jupyter: Interactive Python notebook interface [12]
- Gurobi: Optimization software [4]
- gurobipy: Python interface to Gurobi [4]

Additionally, I used ChatGPT exclusively for proofreading purposes once I had written a full draft of my paper.

Bibliography

[1] Pro football reference. https://www.pro-football-reference.com/. Accessed: 2023-03-01.

[2] Spotrack. https://www.spotrac.com/. Accessed: 2023-03-01.

[3] Kelly Ann Froelich. Is the blind side tackle worth it?: An analysis of the salary allocation of the nfl offensive line. *Duke University*, 2013.

[4] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2023.

[5] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[6] John Haugen. Team success and personnel allocation under the national football league salary cap. *The Park Place Economist*, 14(1), 2006.

[7] Sports Reference LLC. Football glossary and football statistics glossary, 2000.

[8] Cade Massey and Richard H. Thaler. The loser's curse: Decision making and market efficiency in the national football league draft. *Management Science*, 59(7):1479, 2013.

[9] Jason Mulholland. Optimizing the allocation of funds of an nfl team under the salary cap, while considering player talent. *University of Pennsylvania ScholarlyCommons*, 2016.

[10] Eric Ness. Is defense decisive? an examination of the national football league salary structure and game outcomes. *Duke University*, 2010.

[11] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[12] Fernando Pérez and Brian E. Granger. IPython: a system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

[13] Leonard Richardson. Beautiful soup documentation. *April*, 2007.

[14] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[15] Michael Waskom, Olga Botvinnik, Drew O'Kane, Paul Hobson, Saulius Lukauskas, David C Gemperline, Tom Augspurger, Yaroslav Halchenko, John B. Cole, Jordi Warmenhoven, Julian de Ruiter, Cameron Pye, Stephan Hoyer, Jake Vanderplas, Santi Villalba, Gero Kunter, Eric Quintero, Pete Bachant, Marcel Martin, Kyle Meyer, Alistair Miles, Yoav Ram, Tal Yarkoni, Mike Lee Williams, Constantine Evans, Clark Fitzgerald, Brian, Chris Fonnesbeck, Antony Lee, and Adel Qalieh. mwaskom/seaborn: v0.8.1 (september 2017), September 2017.

[16] Tim Zimmer. The impact of nfl salary cap concentration on team success. *CHOREGIA*, 12:53–66, 05 2016.