Worth the Wage

A Comparative Analysis of NHL Player Contracts using Linear Weights

by

Mathieu Parisien

An essay submitted to the Department of Economics
in partial fulfillment of the requirements for
the degree of Master of Arts

Queen's University
Kingston, Ontario, Canada

August 1st 2019

Acknowledgement

First and foremost, I would like to thank my beautiful wife Amielle. Without her unconditional support throughout my entire journey in returning to school, I would not have reached the heights I have attained so far. From doing all the little things at home to maintain our every day life together while I was busy with school work, to even proofreading every single paper in my studies, everything you did was tremendously helpful and I will be eternally grateful for it. You've done too much to express on a single page. I mean, you even followed me to a different city so I can do my Masters! You deserve the Ford Raptor that I will get you someday!

I would like to thank my supervisor, Professor Charles Beach for allowing me to apply my passion for hockey and to actually use it for a course requirement. I appreciate the guidance you have provided me throughout this process.

I would also like to thank Professor Louis-Philippe Morin, who originally allowed me to apply econometrics to hockey in my undergrad. Thank you for your input and for putting up with my constant questions for this project, even though you weren't even officially a part of it.

I would like to thank Zied Haddad, my unofficial editor. You really helped a lot in allowing me to express my ideas in a more comprehensible way. I hope you at least took some enjoyment in the process when reading my thoughts and opinions related to sports.

Last but not least, I would like to thank my family for your constant support and constant interest in what I do. Thanks Mom and Dad for letting me live at home and feeding me for the first few years at university. And thank you Dad for originally introducing me and teaching me about hockey. It has led to a level of obsession high enough to enroll into university just so I can do a project like this. Thank you Manon and Wilf for constantly keeping me on my toes by asking me questions about my project, and of course for letting me marry your daughter.

Table of Contents

Section 1: Introduction

Following the 2004 National Hockey League (NHL) lockout, the game of hockey shifted into a new direction. The lockout resulted in the implementation of a salary cap for team spending. This salary cap set limits for both a maximum and minimum amount a team could spend on player contracts. The purpose of the salary cap was to reduce the spending disparity between large market teams and those in smaller markets. The salary cap, which limits team spending, has caused a great need for efficiency when signing players to contracts. Teams must now optimize their spending to obtain the best possible team under the restrictions of a salary cap. One can easily find news articles such as "The Worst Free Agent Signings in NHL history" (Burtch, 2017) which refers to players not preforming to the level their contract would suggest, signifying the existence of inefficient contracts. I believe these types of contracts exist because of improper valuation of player contributions. With my research, I aim to identify and monetarily quantify players' individual contributions towards winning. I believe that my model will be able to reasonably evaluate the importance of different aspects of the game and will be used to estimate fair valued contracts.

The basis of this project is that the classical method of measuring labour productivity would not be sufficient in this scenario, and that a different approach is required. The classical method would result in finding how teams currently value different aspects of a player's contribution, which I believe is being done inefficiently. If teams properly valued player contribution, then bad contracts would not exist. I believe teams currently over value certain player statistics and undervalue others. This project plans to uncover an objective way to value player contribution by conducting regression analysis to measure the importance of certain statistics towards team success. To accomplish this, the first step to take is to create a statistic to properly evaluate a player's contribution towards winning, which I have named the *Contribution Score*.

Baseball's Pythagorean theorem, developed by Bill James (Winston, 2012, p.3), shows a strong relationship between the *probability of winning* and *runs scored* and *runs allowed*. When this theorem is applied to hockey, we can identify the same strong relationship between goal scoring and winning. For this reason, I believe the best way to properly evaluate player performances is by their contributions towards scoring *goals for* their team and preventing goals from the opposition.

The approach taken will be to perform multiple linear regression analysis to identify how different statistics affect the amount of goals scored and allowed for each team. Two main regressions will be used, one with *goals for* as the dependent variable and the other with *goals against*. The independent variables will vary depending on which regression that will be analysed. Regression analysis will be performed on individual game data for each team as well as season aggregate data. This will allow for the identification of important contributors to team success. Following the regression analysis, in order to evaluate players, the regression results will be applied to individual player statistics and combined in order to determine how much they individually contributed to their team's success. The results of this application will be the player's *Contribution Score,* which will be a new statistic created as a result of this study*.* This analysis will allow for objective player evaluation by properly crediting players for their contributions utilising the available statistics that impact team success.

The second step of this project is to give a monetary value to a player's contributions in order to estimate fair contracts. This will be accomplished by assigning the *replacement player* the minimum allowed salary under the salary cap regulations and comparing other players accordingly. The *replacement players* will be identified as the bottom of the league's players, in terms of *Contribution Score*, who is outside the standard number of regularly active players. A similar method was used in the book *Mathletics* (2012) to identify the replacement players in baseball for a similar analysis; this method is what inspired me to apply it to hockey. The monetary value will be obtained by crediting the league minimum wage to the average *Contribution Score* of the *replacement players*. Once the equation is adjusted to the monetary value of the *replacement player*, each player in the league will be compared to the league minimum in order to determine their fair monetary value. This will be an estimate of a fair contract value for each player based off of their measured contributions for that season. These results will then be used to estimate each player's potential 2019 contract value.

There exist many different models for evaluating player performance across multiple sports. The approach to use linear regressions to evaluate team success and then apply it to evaluate player performances has only been used so far in baseball, shown in the book *Mathletics* (2012). This same approach, applied to hockey, will be used in this paper in order to determine player contributions. Most player analysis used in hockey is done using individual player data or frequency ratios, not often using regression analysis. Using player evaluation models to estimate the value of fair contracts in hockey is very seldom used as well. This project attempts to advance the literature by addressing these two current gaps. The most common analysis involving player salary in hockey is to simply obtain per dollar ratios of their statistics using existing contracts. I believe the major fault of current salary research is the use of existing contracts for their analysis, instead of setting fair salary values based off the objectively determined player evaluation results.

This project aims to target two components. The first will be to properly credit players for their contribution to goals by analysing certain statistics that might be overlooked or not properly valued in today’s NHL. The second accomplishment of this project will be to estimate the fair monetary valuation of each player. With the salary cap looming over every team, managers need to optimize every dollar they spend, which will be made easier with this analysis. Using a different approach by utilising team data and setting fair contracts, inspired by analysis in other sports, will advance the Hockey Analytics literature.

Section 2: Literature Review

There is a large amount of literature that exists on mathematical analysis of sports. The literature referred to for this project was gathered both from online sources and from printed books. Most of the analysis for hockey is posted online spread across various blogs. Some of the literature refers to analysis done on other sports, but their methods used will be applied to hockey in this project.

2.1 - Mathletics

The book *Mathletics* (2012), authored by Wayne Winston, thoroughly explains and demonstrates how mathematics can be applied to sports analysis. Winston attempts to guide the reader towards a better understanding of sports analytics through examples of Excel applications, and by breaking down mathematical equations for applications in sports. From player and team evaluations, to arguing the bias of sport officials, Winston applies statistical analysis to a variety of topics in his book. For the purpose of this paper, I will be focusing on the sections in the book which apply to player analysis.

2.1.1 - Baseball Player Evaluation

Much of the analytical methods used in sports today have their origins tied to baseball. The foundations of many of those methods come from the most famous of sport mathematicians, Bill James. One of James' most known contributions is baseball's Pythagorean Theorem (Winston, 2012, p.3), named this way because of its use of a summation between two squared variables, much like the traditional Pythagorean Theorem from classical geometry. Baseball's Pythagorean Theorem provides an estimation of a team's *winning percentage* with the formula:

estimation of percentage of games won = $\frac{(runs\ scored\ )^2}{(runs\ scored\ )^2 + (runs\ allowed\ )^2}$.

This formula demonstrates the impact that *runs scored* and *runs allowed* (opponent's runs) have on a team's probability of winning. As explained by Winston (2012), this formula has multiple desirable analytical properties (p.3), such as:

- the result will always be between 0 and 1
- its derivative with respect to *runs scored* will be positive, meaning an increase in *runs scored* will lead to an increase in the *winning percentage*
- its derivative with respect to *runs allowed* will be negative, meaning an increase in *runs allowed* will lead to a decrease in *winning percentage*.

When applied to 2006 Major League Baseball (MLB) data, Winston (2012) finds that the Pythagorean equation properly estimated a team's games won within 3 wins per team on average (p.5).

The Pythagorean equation has also been shown to work in other sports, as Winston (2012) demonstrates with the NBA and in the NFL. However, applying it to different sports may require some adjustments to the exponent in order to increase accuracy. The Pythagorean equation will be applied to hockey later in this paper, with the ideal exponent found to be a value of 2.

The Pythagorean equation mathematically proves an obvious fact, the more runs a team scores and the fewest runs they allow, the greater chance that team has of winning. Based on this, in 1979, Bill James created a statistic called Runs Created (Winston, 2012, p.11). This equation demonstrates how a player can be evaluated by how they contribute to scoring runs:

$$Runs\ Created = (hits + BB + HBP) * \frac{TB}{(AB + BB + HBP)}.$$

The variables are defined as follows:

- *hits* is the amount of hits a hitter has had, allowing him to reach a base
- *BB* is the amount of bases a hitter has reached due to being walked
- *AB* is the amount of plate appearances the hitter has
- *HBP* is the amount of bases a hitter has reached due to being hit by a pitch
- *TB* is the amount of total bases a hitter has reached on his hits (1 for a single, 2 for a double etc.).

When applied to the season statistics from 2000 to 2006, the predicted runs created is on average within 4% of the actual amount of runs scored (Winston, 2012, p.13). This equation can be applied to individual players by using their individual statistics as opposed to the team total statistics for the purpose of player evaluation. Further manipulation can be made to this statistic to increase accuracy, such as using per game ratios to equalize player comparisons. This is the first method presented to evaluate player performance from standard in game statistics.

A secondary approach to evaluate player performance explained in Winston's book, and more in line with the empirical work of this paper, is using linear weights to evaluate hitters (Winston, 2012, p.17). The method used to determine the linear weights is multiple linear regression analysis. Using team season statistics from 2000 to 2006, Winston analyses the following regression:

Runs scored = $\beta_0 + \beta_1$ (BB+HBP) + β_2 (singles) + β_3 (2B) + β_4 (3B) + β_5 (HR).

The variables are defined as follows:

- *BB+HBP* is the sum of bases reached by being walked and by being hit by a pitch
- *singles* is a hit that lead to a single base being reached by the hitter
- *2B* is a hit that leads to the hitter reaching two bases, referred to as *a double*
- *3B* is a hit that leads to the hitter reaching third base, referred to as *a triple*

- *HR* is a *home run,* which is equivalent to a hit that leads to the hitter reaching all four bases.

The result of this regression is an R^2 of 0.91 (Winston, 2012, p.21) with all variables being significant. When compared with the Runs Created method, the linear regression method is more accurate at predicting the amount of runs scored by a team in a given season. Therefore, Winston (2012) proclaims that the linear weights method is more accurate at predicting a team's total runs than the Runs Created method (p.21).

To apply the regression results for player evaluation, Winston (2012) estimates the amount of runs a player would create as if the entire team consisted of that player (p.23). To do this, he calculates a scaling value to convert the individual player's statistics to a team level. Once the statistics have been converted, they are then multiplied by the corresponding coefficients to obtain a prediction of the total amount of runs the observed player would score if he consisted of the entire team (Winston, 2012, p.24). The scaling of the statistics is important, as it is necessary to be able to properly compare the results between different players, especially between those who have had less *at bat* opportunities.

To complete the regression analysis, Winston (2012) calculates *Runs Created Above Average* for each player (p.26) using the regression results. He does this by first estimating the amount of runs scored by an average team by applying league average statistics to the regression results. Once that is determined, the statistics are adjusted to simulate how the player being evaluated would replace one of the average players on that team. The difference between the runs scored of the average team, and those scored by the average team plus that specific player, is how much that player would impact the average team. This method is used for a more relative comparison of a player's statistic.

The use of linear weights in baseball has a long history, which dates back to 1916 (Winston, 2012, p.22). Multiple models have been developed to determine the values of those weights. Winston (2012) believes that the best method to determine linear weights is by using Monte Carlo simulation (p.22). For the purpose of this paper, linear regression will be the method of choice for this analysis. However, the Monte Carlo simulation could be applied in a continuation of the project.

2.1.2 - Basketball Player Evaluation

The concept of using linear weights can be applied in the same way for multiple sports. In the book *Mathletics*, Wayne Winston (2012) continues demonstrating the uses of linear weights by applying these in the later sections to both Football and Basketball. Since Hockey and Basketball are similar in their fluidity of game play, for the focus of this paper, I will elaborate more on the basketball application of linear weights.

The first regression model presented is described as a four-factor model, originally created by statistician Dean Oliver (Winston, 2012, p.188). The four-factor model is used to identify how the team has been performing, and to identify the team's strengths and weaknesses (Winston, 2012, p.188). For this model, eight total variables are used, four describing offensive statistics and four describing defensive statistics. The four offensive statistics are the team's *Effective Field Goal Percentage (EFG)*, *Turnovers Committed per Possession (TPP)*, *Offensive Rebounding Percentage (ORP)* and *Free Throw Rate (FTR)*. The four defensive statistics used are the team's *Opponent's Effective Field Goal Percentage (OEFG)*, *Defensive Turnovers Caused Per Possession (DTPP)*, *Defensive Rebounding Percentage (DRP)* and the *Opponent's Free Throw Rate (OFTR)*. To formulate the regression equation, the difference is taken between the offensive and defensive statistics and used as the independent variables, with Games Won as the dependent variable.

Regression: Games Won = $\beta_0 + \beta_1$(EFG-OEFG) + β_2(TPP-DTPP) + β_3(ORP – DRP) + β_4(FTR-OFTR).

The results found that all of the variables are statistically significant at the 5% level, except for *Free Throw Dev* which is significant at the 10% level. The entire equation has an R^2 of 0.909 (Winston, 2012, p.193). Although using the differences in the offensive and defensive statistics does identify which aspects of the game of basketball are important, I believe this method overlooks key information. By taking the difference, one is assuming that offense and defense have the same impact on winning, which may not be a fair assumption to make. I believe that by evaluating statistics separately, the results will be more detailed in identifying which aspects are truly important to the sport being analysed.

In terms of player evaluations using linear weights, several different methods have been developed. Three such methods described by Winston (2012) are the *NBA Efficiency metric*, the *Game Score* and the *Win Score* (p.195).

The NBA Efficiency metric is a very simple linear equation accounting for various player statistics:

Efficiency per game = (points per game) + (rebounds per game) + (assists per game) + (steals per game) + blocked shots per game) – (turnovers per game) – (missed Field Goals per game) – (missed Free Throw per game).

This method, developed by Dave Heeren (Winston, 2012, p.195), is over-simplified because it values all statistics the same by assigning “weights” of +1 for good statistics and -1 for bad statistics. This linear weights method lacks the proper use of the weights to be an effective measure of performance.

John Hollinger builds off of the Efficiency metric by creating the Game Score statistic (Winston, 2012, p.196). The Game Score is an improvement on the Efficiency metric because it uses different weights for each statistic:

Game Score = 1.0(points) + 0.4(Field Goals Made) - 0.7(Field Goals Attempted) - 0.4((Free Throws Attempted – Free Throws Made)) + 0.7(Offensive Rebounds) + 0.3(Defensive Rebounds) + 0.1(Steals) + 0.7(Assists) + 0.7(Blocks) - 0.4(Personal Foul) -1.0(Turn Overs).

It is unclear how the weights are determined or justified. The Game Score is a simpler version of Hollinger's more complex Player Efficiency Rating (PER) (Winston, 2012, p.196). The main arguments made against the PER or the Game Score is that the results are offensively biased. This bias is common for many player evaluation formulas in sports, mainly due to the limited countable defensive statistics available. The final linear player evaluation method mentioned in *Mathletics* is the Win Score, created by Berri, Schmidt and Brooks (BSB) (Winston, 2012, p.199).

Player Win Score = points + rebounds + steals + 0.5(assists) + 0.5(blocked shots) – FG attempts – turnovers – 0.5(FT attempts) – 0.5(personal fouls).

BSB uses the Player Win Score to evaluate players by converting it into Wins Produced, and the summation of all the players' Wins Produced on a team will be approximately equal to that team's number of games won that season (Winston, 2012, p.199). The Win Score allows for a better evaluation of how much a player contributed to his team's success. The issue with this statistic, like many others, is that it doesn't account for much defensive contributions because box scores have more offensive statistics. Winston (2012) believes Win Score is the best system to evaluate players, but that it is limited by the box score statistics available (p.201).

2.1.3 - Salary Analysis

Once the methods for player evaluations are established, Winston (2012) then addresses the methods for applying these evaluations to determine player salary. The methods used by Winston (2012) to determine player salary in both baseball and basketball follow the same methodology. The first step taken is to identify what is known as a *replacement player*. A *replacement player* is a lower-tier player in the league, with the skill level of a player who would fill in for an injured player. Winston (2012) determines the replacement level in baseball by identifying the bottom 20% of the league in number of plate appearances (p.79). For basketball, Winston (2012) establishes the replacement player by identifying the bottom 10% of the league in the WINVAL statistic (p.233). WINVAL (winning value) is a player rating statistic developed by Jeff Sagarin of USA today (Winston, 2012, p.204), measuring a player's plus/minus accounting for teammates and score. The process to determine the fair value of a win varies between sports, as it is dependent on the statistics used to evaluate the players.

Winston (2012) chooses to use the SAGWIND statistic to evaluate baseball players for this exercise (p.80). SAGWIN is a type of point system quantifying each play a hitter or pitcher makes in correlation with how much that play impacts the probabilities of winning the game (2012, p.72). Using the SAGWIN points, Winston (2012) determines that a team of only replacement level players would have a season record of 44 wins and 118 loses, which represents -74,000 SAGWIN points split between the pitchers and the hitters (p.79). Winston (2012) determines that each plate appearance of a replacement level player is worth -5.97 SAGWIN points (p.80). It is assumed at this point that a team of replacement players has a team salary of zero (Winston, 2012, p.81). To properly evaluate a player's contribution above the replacement level, the VORPP (Value Over Replacement Player Points) statistic is used where a player's VORPP is simply 5.97

added to his SAGWIN. Using the average team payroll of $77 million and making the assumption that an average team would have a season record of 81 wins and 81 losses, Winston finds that every VORPP point above replacement is worth $1,040. This method quantifies a player's contribution and can be applied to determine a player's salary from the VORPP statistic above replacement level.

For the basketball evaluation, the WINVAL statistic is used to translate player performance into salary values (Winston, 2012, p.233). Winston (2012) found that -6 is the average WINVAL value for a replacement level player (p.233). Unlike the MLB, the NBA has a salary cap which implies a minimum player salary of 400,000 dollars (Winston, 2012, p.233). Therefore, Winston (2012) makes the association that -6 WINVAL points is worth 400,000 dollars (p.233). By applying the WINVAL statistics to the Pythagorean Theorem, it is determined that a replacement level starting lineup of five players is projected to win zero games, and costs two million dollars in payroll (Winston, 2012, p.234).

By making the same assumptions as earlier, assuming that a team with the average payroll of $66 million will win 50% of their games, it is determined that one win is worth just over $1.5 million (p.234). To apply this *per win* monetary value to players' WINVAL statistic Winston (2012) performs some manipulations with the WINVAL by including minutes played by each team. With this manipulation, Winston (2012) determines that one win is worth 2,880 points above replacement per minute (p.234). Using all of this information, Winston (2012) creates a *wins generated* by player statistic as follows (p.234):

$$Wins\ generated = (WINVAL + 6) * \frac{minutes\ played}{2{,}880}$$

With the *wins generated* statistic for each player, and the valuation of $1.5 million per win, Winston (2012) has demonstrated the process to determine an NBA player's fair salary.

2.2 - HockeyNomics

HockeyNomics (2009), written by Darcy Norman, is eight chapters long and covers topics from the History of Sports Statistics to analyzing Wayne Gretzky's greatest accomplishments. In the second chapter, which is the focus of this review, Norman describes his process for identifying the best free agent per dollar spent from the 2007 off-season.

Norman (2009) begins the analysis by identifying the best unsigned players from the 2006-2007 season based purely on goal count (p.45). The analysis begins with goals because, as per Norman (2009), "[m]any other factors come into play in the course of a game or season, but the fact remains that if a team doesn't score, it will win exactly zero games" (p.45). The clear importance of goals in an inspiring factor for this project. Norman emphasizes that statistics in their pure form are not enough for a proper analysis and that more context is required. Norman (2009) argues that a player's goal production is highly correlated with the amount of playing time they get (p.51). To add context, Norman (2009) incorporates each player's ice time to obtain their goals *per 60 minutes* played (p.53). Norman (2009) continues by admitting that there is more credit to be given for a goal than to only the player who scores it (p.56). Therefore, Norman (2009) adds the credit of assists by utilising the *Goals Created* statistic instead of pure goals (p.61). *Goals Created* (GC) is a statistic originally created by Alan Ryder (Norman, 2009, p.61). *Goals Created* distributes the credit of each goal to the player scoring the goal, and to the players who assisted on the goal (Norman, 2009, p.61). A simplified version of the formula is used in Norman's analysis, as follows:

$$GC = \left(Goals + (0.5\ Assists)\right) * \left(\frac{Team\ Goals}{Team\ Goals + (0.5\ Team\ Assists)}\right).$$

When a goal is scored in hockey, the last player to touch the puck on the offensive team gets credited with a goal, and up to the last two players on his team who touched the puck before him get credited with assists. Typically, one goal leads to the distribution of three possible points, where a player's points are measured as the sum of a player's *total goals* and *assists*. Instead of having each goal or assist measured as one point each, *Goals Created* counts each goal as one unit and distributes partial credit for that goal amongst the goal scorer and those who assisted on it (Norman, 2009, p.63). GC introduces the concept of distributing credit given for one goal to not only the goal scorer but to those who helped the scorer accomplish it.

Norman (2009) continues his player analysis using GC by manipulating it to account for a player's ice time in manpower situations (p.67). After analysing data from 2000 to 2006, Norman (2009) finds that power play goals are 2.5 times more frequent per minute compared to at even strength (p.70). Norman (2009) also finds that there are only 0.4 shorthanded goals (when a team is one player short) per minute compared to at even strength (p.70). To account for manpower situations, Norman (2009) simply multiplies the quantity of ice time by the frequency of the goals scored (Power Play ice time * 2.5, Penalty Kill ice time *0.4) to obtain the Adjusted GC *per 60 minutes* played (AdjGC/60) for each player (p.73). Although AdjGC/60 is a thorough tool in and of itself, Norman (2009) acknowledges that it does not account for defensive play (p.76). Using the AdjGC/60 on its own would not provide a fair assessment for players who play a more defensively responsible game. As much as goal scoring is important to the game, goal prevention is also a key factor.

To account for defense, Norman (2009) creates his own statistic to distribute blame amongst players for *goals against* their team while they are on the ice (p.82). Norman explains that when a goal is scored against a team, it is most likely a result of multiple events caused by multiple players. This means that all players on the ice are partly to blame for the *goal against* to some degree. The equation used to account for this is:

$$Goals\ Against = (Individual\ Player's\ Goals\ Against)\frac{Team\ Goals\ Against}{Sum\ of\ Individual\ Players'\ Goals\ Against}.$$

For his analysis, Norman (2009) simplifies this equation by creating *Kinda Goals Created Against* (KGCA), which is simply dividing the Individual Player's on-ice *goals against* by 6 (p.83). KGCA gives very close results to the *Goals Against* formula and is much easier to apply (p.84). Norman (2009) adjusts the KGCA statistic to account for manpower situations and ice time to obtain KGCA *per 60 minutes* (KGCA/60) (p.85). With KGCA/60, the defensive aspect of the game is now accounted for.

To obtain the final statistic to be used to determine the best player available, Norman (2009) uses *Net Goals Created per 60 minutes* (Net GC/60), which is the difference between AdjGC/60 and KGCA/60 (p.90). *Net Goals Created* credits both the offensive contributions of a player and his defensive contributions in goal prevention. Using *Net GC/60* Norman is able to compare and truly identify who the best overall player is among a list of unsigned players.

The final step of Norman's analysis is to account for monetary spending. The goal of Norman's entire exercise was not only to identify who the best player was, which was accomplished by the *Net GC/60* statistic, but to find the best player per dollar spent.

To accomplish this final task, Norman (2009) takes the players' season total Net GC/60, and divides it by the player's obtained contract from 2007 (p.95). This result finds how many *Goals Created per 60 minutes of play* over a season that player will produce per dollar spent.

With this analysis, Norman has demonstrated a reasonable and innovative method to analyse a player's statistics. He managed to add context to the data by accounting for a player's ice time and he presented new statistics to incorporate the aspects of the game which he found important. The methods used by Norman should be seen as the proper foundation for any type of player analysis in today's professional hockey.

2.3 - Introduction to Weighted Points Above Replacement

2.3.1 - Part 1

The most similar approach to what will be attempted in this paper from the existing literature is the Weighted Points Above Replacement (WPAR) approach. WPAR is presented by a blogger from the Hockey Graphs website who goes by the name of EvolvingWild (henceforth EW). The approach which was published in two parts in August 2017. The author's motivation for this player evaluation model is the use of an aggregate statistic to explain player performance. The belief here is that "combining multiple aspects of the game into one number can be insightful and extremely useful" (EW, 2017) when looking at player performance. The Weighted Points approach is the most relevant work for this paper because, unlike other works in hockey, linear regression is used to obtain the proper weights associated with the statistics used. I share a similar belief to EW in that there is more than only the counted statistics (goals and assists) that contribute to team success, and that statistics should not all be weighted the same. The author begins his player analysis by establishing a Weighted Points (WP) equation.

The data used in EW's work is individual player data from the 2010 to 2017 regular seasons. To account for the fact that a player's position has a large impact on his statistical output, the forwards and defencemen are separated into separate groups for the analysis. The regression method used here is referred to as the "split-half regression technique" (EW, 2017), which has been utilized in other Hockey Analytics literature. This technique will allow the result to show if a statistic is indeed relevant and a repeatable skill.

The author comes to the conclusion that the best dependent variable used for the evaluation of player performance is *on-ice goals*. *On-ice goals for* or *on-ice goals against* are credited to a player when a goal was scored while he was on the ice, regardless which player actually scored the goal. Using this as the dependent variable allows a player to be given partial credit for a goal scored, without necessarily having a direct impact on the goal itself. This is a similar evaluation strategy that I will be using in my study, however I will be utilising a goal at the team level for my regression analysis. To finalize the list of independent variables to be used, EW runs multiple regressions and retains only the most significant statistics. Some of the independent variables are *goals*, *primary assists* and *corsi* differential. *Corsi* is a proxy for *shot attempts* calculated by taking the sum of *shots on goal,* shots that missed the net (*missed shots)* and *blocked shots*. To expand on the equation, EW turns to past literature to obtain weights for statistics he was not able to test himself in the regression analysis, like the value of penalties and faceoffs. To incorporate these into his equation, EW simply adds these weights to his regression equation to finalize his Weighted Points(WP) equation. To obtain each player's total WP value, the equation is simply applied to each player's individual statistics.

The final step taken in this portion of the analysis is to obtain an *Above Average* value of the Weighted Points. This allows for easier interpretation, as a player's WP will be relative to the league average. This manipulation is made by taking the ratio of each player's WP value per minutes played that season, then obtaining the difference between the player's ratio and the ratio of the league's average *WP per minute* value. To finalize the manipulation, EW multiplies the per-minute difference by the player's total ice time to obtain the Weighted Points Above Average (WPAA) for each player. The *Above Average* manipulation is only used initially, as the data is then further manipulated to establish an *Above Replacement* level instead of *Above Average* in EW's follow up blog post.

2.3.2 - Part 2

As a continuation of the work done to obtain Weighted Points Above Average (WPAA), EW converts the Weighted Points Above Average statistic to Above Replacement. The Above Replacement baseline is quite common, as it is used in player evaluations across many sports like baseball and basketball. The article clarifies that, "Replacement level is the performance we would expect to see from a player a team could easily sign or call up to 'replace' or fill a vacancy. In theory it is the lowest tier NHL player" (EW, 2017). *Replacement level* players for basketball and baseball analysis are often seen as the players waiting on the bench, not those in the play. In hockey, there are no *replacement players* on the bench, as all players on the bench enter the game in shifts. Therefore, a different approach is needed.

In order to properly classify the *replacement players*, EW approaches this problem by analyzing player salary. The salary method adapted is as follows: players are identified as *replacement level players* if they are over the age of 24 and signed a new contract for the league minimum value. Following these restrictions, this allows the *replacement level* player to be at prime age and to be seen by the league as an expendable player.

Once the *replacement player* has been identified, their total WPAA is calculated for both forwards and defencemen. The results found show negative WPAA numbers for both the replacement forwards and defensemen. This result makes sense, as the league average is measured at a WPAA equal to zero, so these replacement players are worse than average.

The final step to the analysis is to convert all players' WPAA statistics into Weighted Points Above Replacement (WPAR) by incorporating the established *replacement level* into the WPAA equation. The author does this by calculating the difference between the WPAA of each player and the replacement level rate, and he also incorporates the individual player's total ice time. This calculation is done for each player in the league for individual seasons. To justify the use of the WPAR statistic, the author concludes the article by showing that a player's WPAR statistic is a repeatable skill and a good measure of player performance by showing the high correlation between a player's WPAR from one season to the next.

The purpose of WPAR is to obtain a clear way to evaluate individual player performance, one that can be used to easily compare players to one another. I attempt to follow the lead of this article by achieving that same goal and answering the same questions. I will equally be utilizing regression analysis to evaluate players, but will be taking a slightly different approach in terms of establishing the variable weights and relevant statistics. Although the WPAR statistic is quite informative, I believe my work will take it one step further by applying the results to establish fair salaries to players based on their contributions.

2.4 - Measuring Single Game Productivity: An Introduction to Game Score

Determining who is the best player in a game is a debate constantly had by hockey fans. OMGITSDOMI (henceforth OM) attempts to shed some light on this topic in his blog post *Measuring Single Game Productivity: An Introduction to Game Score* on the hockey-graphs website. OM explains his process of identifying a player's contribution in an individual game by applying a modified form of the Game Score developed for basketball by John Hollinger (OM, 2016) mentioned earlier in the section 2.1.1. The Game Score is calculated from a linear equation of an individual player's statistics with different weights assigned to them. This equation is an attempt to converge the different statistics used to evaluate players into one aggregate statistic and obtain a simple way to measure a player's productivity in a specific game (OM, 2016).

The first step used in this article to apply the Game Score concept to hockey was to determine which statistics to include in the equation. The statistics chosen were *goals, primary assists, secondary assists, shots on goal, blocked shots, penalty differential, faceoffs, 5-on-5 corsi differential, and 5-on-5 goal differential* (OM, 2016). It is not clear what process was used to determine the choice of statistic, but I believe he used personal preference.

OM's (2016) method to find the appropriate weights for each statistic was to weigh each statistic in relation to its frequency to goals. To make the Game Score statistic more interpretable and comparable to other standard statistics, OM (2016) scaled back most of the weights by 75 percent, but not all of them. I disagree with this method of determining weights. I don't believe that the frequency of a statistic is a proper measure of its importance. I also disagree with the uneven manipulation of the weights because if the statistics are not manipulated evenly throughout the equation, the equation could miss interpret the impact of each statistic and skew the results.

The final equation is:

$$
\begin{aligned}
&Player\ Game\ Score \\
&\qquad = 0.75(Goals) + 0.7(Primary\ Assist) + 0.55(Secondary\ Assist) \\
&\qquad + 0.075\ (Shot\ On\ Goal) + 0.05(Blocked\ Shots) \\
&\qquad + \ 0.15\ (Penalties\ Drawn - Penalties\ Taken) \\
&\qquad + 0.01(Faceoff\ Won - Faceoff\ Lost) + 0.05(Corsi\ For - Corsi\ Against) \\
&\qquad + 0.15(Goals\ for\ Differential)\ .
\end{aligned}
$$

OM (2016) does extend the application of the game score to goaltenders as well with a similar process only using *goals against* and *saves* in the equation:

$$Goalie\ Game\ Score = \ -0.75(Goals\ Against) + 0.1(Saves).$$

After applying his *Game Score statistic* to player data, OM shows it has some merit. Some of the greatest individual game performances in recent history obtained the highest Game Scores. OM continues his application by applying the Game Score to a season total level to evaluate a player's performance over an entire season. This results in many of the biggest names in the sport having the highest average game scores. OM also shows that a player's Game Score from one season is strongly correlated to his Game Score from the previous season, meaning that it is repeatable skill and a good estimation for a player's talent.

The author does admit that the statistic has some downfalls. The Game Score does not account for factors such as quality of competition and the impact of teammates. It also fails to account for special team situations (manpower situations), and, due to the statistics used, relies much more heavily on offense. This lacks defensive evaluation, which is a common issue faced in many sports. The Game Score, however, provides a good foundation for player evaluation statistics and provides many opportunities for future work.

Section 3: Data

The research in this paper consists of two parts: the first involves obtaining a Player Contribution model, and the second involves assigning a monetary value to those contributions. Different data sets and information sources are needed to accomplish both of these tasks. Appendices 1 and 2 are the summary statistics tables for the data used in the regression analysis.

3.1 - Player Contribution Data

To build a Player Contribution Model, both team and player data are required. Much of this data was obtained from the online database naturalstattrick.com (n.d). The first stage of the analysis consists of measuring how different aspects of the sport of hockey contribute to goal scoring and goal prevention at the team level. To accomplish this, individual game data for every NHL team from the 2014-15 season until the 2018-19 season will be used for initial regression analysis. From the 2014-2017 seasons, there were 30 active NHL teams, and from the 2017-2018 season, the league expanded to 31 teams. With 82 games played per season, for 5 seasons, and every individual game counting as a separate observation for each playing team, there is a total of 12,464 observations being analysed. The statistics used from this data set are g*oals for, goals against, shots for, shots against, blocks for, blocks against, missed shots for, missed shots against, and shooting percentage.* All the statistics used are available on the website or are obtained with minor mathematical manipulations. The manipulations needed were the decompositions of the reported *corsi* and *fenwick* statistics which are proxy statistics for puck possession. *Corsi* is a summation of *shots on net, missed shots and blocked shots,* and *fenwick* is the same summation with the exclusion of *blocked* *shots*. The decompositions of c*orsi* and *fenwick* to obtain *missed shots and blocked shots* were done to be able to measure the statistics in their most primitive form and to avoid double counting in the final model. Following the methodology used by Darcy Norman (2009) while developing the *Goals Created* statistic, *per 60 minutes* ratios will be used for each statistic in order to equate the data value across teams.

Certain statistics such as *hits, primary assists, secondary assists, takeaways, giveaways and faceoff percentages* were not available in the game level data. In order to account for these statistics, team data at the season aggregate level was used. These statistics were obtained by manually summing individual player data for every team and every season individually. *Goals for* and *goals against* at the aggregate season level was also used for this portion of the analysis. Over the 5 years observed, this consisted of 152 observations.

Team data will be used, both at the season level and individual game level, to run regression analysis to obtain the necessary weights for the statistics being analysed. Following the analysis, the regression results will be applied to individual player data. To accomplish this, the use of individual player data at the season level from the 2016-17 season to the 2018-19 season is also needed and obtained from the online database naturalstattrick.com (n.d). The same statistics used at the team level in the regression analysis will be used at the individual level for this application. As described later in the methodology section, certain statistics need to be manipulated to obtain a player's individual value.

One aspect that makes hockey a difficult sport to analyse empirically is the multiple man-advantage situations that could take place within a game. Darcy Norman (2009) explains this issue best when he states that not all goals should be viewed as equal (p.67). When a team is on a power-play and has additional players on the ice compared to the opposition, they are at an advantage and it is easier for them to score with an extra player compared to when both teams are at even strength. To make the statistics as comparable as possible across players, only 5-on-5 data will be used. Approximately 75% of NHL games are played at even strength (Duroux, 2017), which consists of the majority of the action, and even some statistics such as *corsi* are only measured at even strength to have a fair measurement. Measuring the full game statistics as being equal, as EW (2017) did in his Weighted Points models and OM (2016) did in his Game Score model, could lead to skewed evaluation results, as EW mentions in his conclusion. Using the full game data would be ideal for the best results; however, segregating each possible scenario to properly measure them is a complex task as there are several different man-advantage situations that could occur. To add to the difficulty, not every statistic can be used in every situation. As previously mentioned, *corsi* is only measured at even strength, which would limit the use of *missed shots and blocked shots* in man power situations as some of these statistics are not individually available in the data set. Using 5-on-5 data allows for all the statistics to be measured equally, eliminates advantages some players might have by playing large amounts of their ice-time on the power-play and eliminates any complication of needing to account for different man-advantage situations.

3.2 - Monetary Data

For the player valuation portion of this paper, I will be using data on player contracts from the online database spotrac.com (n.d). This website lists all the details of current player contracts, which include the total dollar value of the contract, the length of the contract, as well as the *annual average value* (AAV). From this website, I also have access to total spending amounts for each team. This information will allow me to compare my results of estimated fair contracts with existing contracts to identify potential spending inefficiencies. I will also be using information on the salary cap, obtained from the online database puckreport.com (Puck Report, 2017), to identify the current maximum and minimum spending limits for both individual contracts and team spending.

Section 4: Methodology

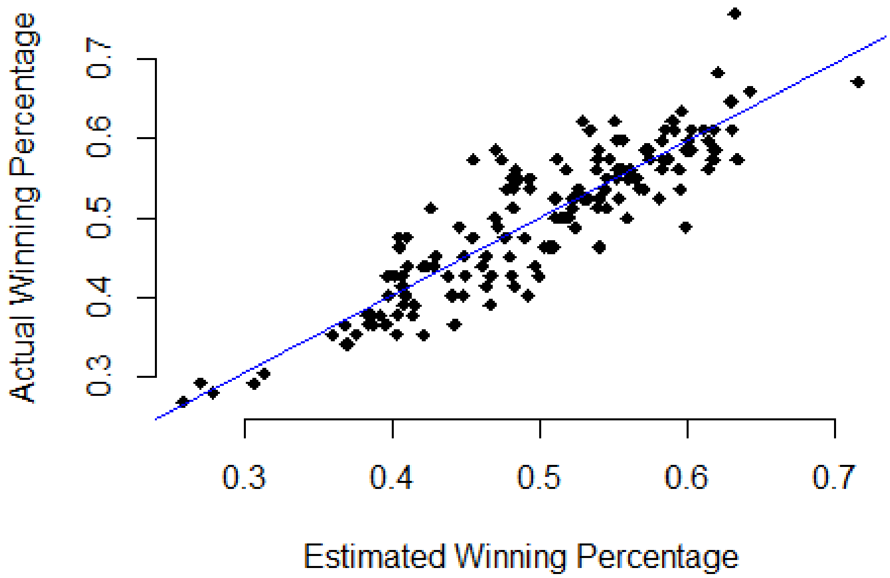
The work required in order to monetarily quantify a player's contributions towards winning must be done in two stages. The first stage is to construct a Player Contribution Model using regression analysis in order to properly evaluate individual player performance. The second stage will be to apply the model to individual player statistics and assign a monetary value to those contributions.

4.1 - Player Contribution Model

4.1.1 - Team Analysis

The approach for the Player Contribution Model will be to perform multiple linear regression analysis in order to identify how different statistics affect the number of *goals scored* and *goals allowed* for each team. As previously mentioned, the Pythagorean Theorem of baseball developed by Bill James (Winston, 2012, p.3), which is an estimation of *winning percentage*, can also be applied to hockey. After the Pythagorean Theorem is applied to the season statistics from 2014 to 2019, the results in Figure 1 show a clear relationship between goal scoring and winning, which is why the variables *goals scored* and *goals allowed* will be used as the dependent variables in the regression analysis. As shown in Wayne Winston's work in his book *Mathletics* (2012), and in EW's Weighted Points model (2017), regression analysis is a justifiable approach to measure player performance. The Player Contribution Model will be more along the lines of Winston's linear weights regression used in his baseball player evaluation (Winston, 2012, p.20). The approach will be to regress different statistics on g*oals* and then apply the regression results to individual players' statistics.

FIGURE 1: Pythagorean Hockey Equation



$$Estimated\ Winning\ Percentage\ = \frac{GF^2}{GF^2 + GA^2}$$

This approach differs from the Weighted Points Model by using team statistics in the regression analysis as opposed to taking individual player statistics. I believe this approach is optimal because the results will show the true effects of how certain statistics help the entirety of the team. Winston has applied this method to evaluate players in both baseball and basketball. It has yet to be reproduced for hockey.

The independent variables will vary depending on the regressions being analysed and their significance levels. Due to the data restrictions mentioned in the data section, separate sets of regressions will be conducted with one set using individual game statistics and the other set using season total statistics. The econometric models will take the following forms:

Set 1 - Season total regressions:

$$(1.1)\ GF.60_{i,t} = \beta_1 + \beta_2\ A1.60_{i,t} + \beta_3\ A2.60_{i,t} + \beta_4\ Giveaways.60_{i,t} + \beta_5\ Takeaways.60_{i,t} + \beta_6\ HitsF.60_{i,t} + \beta_7\ HitsT.60_{i,t} + \beta_8 Face.Perc_{\,i,t} + \beta_9 Team_i + B_{10} Year_t + \varepsilon_{i,t}$$

$$(1.2)\ GF.60_{i,t} = \beta_1 + \beta_2\ Ast.60_{i,t} + \beta_3\ Giveaways.60_{i,t} + \beta_4\ Takeaways.60_{i,t} + \beta_5\ HitsF.60_{i,t} + \beta_6\ HitsT.60_{i,t} + \beta_7 Face.Perc_{i,t} + \beta_8 Team_i + B_9 Year_t + \varepsilon_{i,t}$$

$$(2)\ GA.60_{i,t} = \beta_1 + \beta_2\ Giveaways.60_{i,t} + \beta_3\ Takeaways.60_{i,t} + \beta_4\ HitsF.60_{i,t} + \beta_5\ HitsT.60_{i,t} + \beta_6\ Face.Perc_{\,i,t} + \beta_7 Team_i + B_8 Year_t + \varepsilon_{i,t}$$

Set 2 - Individual game regressions:

$$(3)\ GF.60_{i,m,t} = \beta_1 + \beta_2\ SF.60_{i,m,t} + \beta_3\ BlockA.60_{i,m,t} + \beta_4\ Missed.F.60_{i,m,t} + \beta_5 SH.Perc_{i,m,t} + \beta_6\ Team_i + B_7 Year_t + \varepsilon_{i,m,t}$$

$$(4)\ GA.60_{i,m,t} = \beta_1 + \beta_2\ SA.60_{i,m,t} + \beta_3\ BlockF.60_{i,m,t} + \beta_4\ Missed.A.60_{i,m,t} + \beta_5\ Team_i + B_6 Year_t + \varepsilon_{i,m,t}$$

Regression equation (1.1) will measure the effects of offensively oriented statistics on a team's goal scoring ability at the season aggregate level. For this regression, the dependent variable is *goals for per 60 minutes played (GF.60)* for team *i*, in season *t*, which are the goals scored by team *i*. For the independent variables we have *primary assists per 60 minutes played (A1.60)* which is awarded to the last player to pass the puck to the goal scorer*, secondary assists per 60 minutes played (A2.60)* which is awarded to the last player to pass the puck to the primary assister, *giveaways per 60 minutes played (Giveaways.60)* which is when a player from team *i* gives up possession of the puck to the opposing team, *takeaways per 60 minutes played (Takeaways.60)* which is when a player from team *i* takes possession of the puck away from the opposition, *hits for per 60 minutes played (HitsF.60)* which is when a player from team *i* makes physical contact with the opposition leading to a loss of possession, *hits taken per 60 minutes played (HitsT.60)* which is when the opposition makes physical contact with a player on team *i* leading to a loss of possession, and *faceoff percentage (Face.Perc)* which is the percentage of faceoffs a team wins in a season. All the variables in the regression are for team *i*, at season *t*. The two final independent variables in the equation will be control variables. The first is a *Team* control variable which will control for consistent differences across teams, with the Anaheim Ducks as the reference group. The second is a Year control variable which will control for scoring trends across different seasons, with 2014 as the reference year.

Regression equation (1.2) is identical to regression (1.1) with the exception of using *total assists (Ast.60)* as an independent variable as oppose to *primary assists (A1.60)* and *secondary assists (A2.60)* separately.

Regression equation (2) will measure the effects of defensively oriented statistics on how the number of goals a team allows at the season aggregate level. For this regression the dependent variable is *goals against per 60 minutes played (GA.60)* for team *i*, in season *t*, which are the goals scored against team *i*. For the independent variables we have *giveaways per 60 minutes played (Giveaways.60)* which is when a player from team *i* gives up possession of the puck to the opposing team, t*akeaways per 60 minutes played (Takeaways.60)* which is when a player from team *i* takes possession of the puck away from the opposition, *hits for per 60 minutes played (HitsF.60)* which is when a player from team *i* makes physical contact with the opposition leading to a loss of possession, *hits taken per 60 minutes played (HitsT.60)* which is when the opposition makes physical contact with a player on team *i* leading to a loss of possession, and *faceoff percentage (Face.Perc)* which is the percentage of faceoffs a team wins in a season. All the variables in the regression are for team *i at* season *t*. The two final independent variables in the equation will be control variables. The first is a *Team* control variable which will control for consistent differences across teams, with the Anaheim Ducks again as the reference group. The second is a Year control variable which will control for scoring trends across different seasons, again with 2014 as the reference year.

Regression equation (3) will measure the effects of offensively oriented statistics on a team's goal scoring ability at the individual game level. For this regression, the dependent variable is *goals for per 60 minutes played* (*GF.60*) for team *i*, in game *m*, and season *t*, which are the goals scored by team *i*. For the independent variables we have *shots for per 60 minutes played* (*SF.60*) which is a *shot on goal* for team *i,* b*locked shots against per 60 minutes played (BlockA.60)* which is a shot attempt for team *i* that was blocked from reaching the goal by the opposition, *missed shots for per 60 minutes played* (*Missed.F.60*) which is a *shot attempt* by team *i* which was not blocked but missed the net, and *shooting percentage (SH.Perc)* which is the percentage of shots that counted as a goal over the total amount of shots on net. All the variables in the regression are for team *i*, in game *m*, and season *t*. The two final independent variables in the equation will again be Team and Year control variables.

Regression equation (4) will measure the effects of defensive oriented statistics on the number of goals a team allows at the individual game level. For this regression the dependent variable is *goals against per 60 minutes played (GA.60)* for team *i*, in game *m*, and season *t*, which are the goals scored against team *i*. The independent variables are, *shots against per 60 minutes played (SA.60)* which is a *shot on goal* for *i's* opposing team, *blocked shots for per 60 minutes played (BlockF.60)* which is *a shot attempt* for team *i's* opponent that was blocked from reaching the goal by team *i,* and m*issed shots against per 60 minutes played (Missed.A.60)* which is a *shot attemp*t by team *i's* opponent which was not blocked but missed the net. All the variables in the regression are for team *i*, in game *m*, and season *t*. The two final independent variables in the equation will again be control variables for Team and Year.

As explained in the data section, the majority of the statistics used are in *per 60 minute* ratios in order to properly compare their values across different games. All statistics used in these regressions are at the team level. This portion of the analysis allows us to identify the importance of each statistical event on goal scoring and goal prevention at the team level.

4.1.2 - Player Application

To finalize the Player Contribution Model, the coefficients found in the regression analysis for each variable will be used as weights and applied to players' individual statistics, in order to then determine each player's contribution. This same method is used by Winston in his Linear Weights Model. This application will allow for the proper evaluation of a player's contribution to winning and to give fair credit to players for their overall play on the ice.

The application is quite intuitive as individual player statistics are multiplied by the corresponding weights found in the regression results for a specific season. This will result in numbers representing each player's contribution to their team's *goals for* and *against per 60 minutes* of play. These results will then be combined to obtain a player's *Contribution Score*.

For the measure of a player's offensive contributions, the results from regressions (3) and either (1.1) or (1.2) will be used, depending on the results. Once the regression results are analysed and the appropriate variables are identified, their corresponding weights will be multiplied by the individual player statistics. This will provide us with two separate results, one for each offensive regression. In order to fully obtain a player's contribution towards goal scoring, much like in EW's Weight Points Model, the resulting numbers are combined by simply being summed together. This summation will allow for the full measure of a player's offensive contributions to account for all of the independent variables analysed.

The same process will also apply to the measure of a player's defensive contribution, applying regressions (2) and (4) to the individual statistics and summing the results. For the defensive contributions, however, there requires additional manipulation. Many defensive statistics are not measurable individually the same way offensive statistics are. As an example, an individual *shot for* is awarded to the individual player who took the shot, but a *shot against* is awarded to every player who was on the ice when the opponent took a *shot on goal*. Therefore, blame needs to be distributed, but as explained in Darcy Norman's *HockeyNomics* (2009), distributing blame is quite difficult to do. To remedy this issue, the same adjustments made in *HockeyNomics* will be applied. The adjustment is simplified to dividing the corresponding defensive statistic by 6, accounting for each player on the ice, including the goalie for team i, because of the use of 5-on-5 data in this analysis. This adjustment will distribute the blame equally to all players involved.

Once the application of the weights is complete for both offensive and defensive contributions, following the method used by Norman (2009) in his *Net Goals Created* statistic, the difference between a player's offensive and defensive contributions will be taken. This difference is representative of a player's overall contributions to his team for both the offensive and defensive sides of the game. As a result of this stage of the analysis, we will have a player contribution statistic for each player in the league for each season played, which will be known as a player's *Contribution Score*. To be able to fairly compare players' contribution results, forwards and defensemen will be separated into their own respective groups for the remainder of the analysis.

4.2 - Monetary Valuation of Player Performance

What many existing player evaluation models in hockey lack is a translation to monetary value based off their findings. This topic has been applied in other sports by Winston in *Mathletics* (2012), and will be applied to hockey in this second stage of the analysis. The first step of this stage, much like in past work, is to identify the replacement level player. As previously mentioned, the replacement level is the lowest ranked NHL player who would be called upon in case of injury.

Each team must normally have 12 forwards and 6 defensemen dressed in order to play a game. Using the results from the *Contribution Scores*, the highest ranked 372 forwards (350 before 2017) and 186 defensemen (180 before 2017) per season are the players recommended to be dressed for every game. All players ranked lower than these will be identified as a replacement level player. To eliminate distorted results due to small sample sizes, only players with at least 60 minutes of total playing time in a given season will be used for this classification process. This will result in approximately 30% of the league's players being identified as *replacement players*.

To make the *Contribution Score* statistic comparable between the recommended players and replacement players, a R*eplacement Contribution Score* (RCS) will be calculated. The *RCS* will be calculated by taking the average *Contribution Score* of the replacement level players, much like how Winston obtains his replacement statistic. There will be one *RCS* for forwards and one for defencemen to maintain their separate analysis. The *RCS* will allow for easier comparison between players by comparing their contributions to an average replacement player.

To assign a monetary value to a player's contributions, we turn to the salary requirements under the rules of the league's Salary Cap. The Salary Cap is the maximum amount a team can spend on their player contracts, which fluctuates based on league revenue (Puck Report, 2017). In addition to the Salary Cap, under the guidelines of the league, there is also a minimum and maximum limit to the Average Annual Value (AAV) of a player's contracts. A player's salary can vary from year to year, as long as the average salary of the entire contract falls within these limits, as the AAV is what is used to calculate a team's salary structure. The Salary Cap and the wage limits are determined before the beginning of each new season. The maximum AAV contract is limited to 20% of the Salary Cap amount, while the player contract minimum is a predetermined fixed amount (Puck Report, 2017).

The final step is to provide a monetary value for each point of the contribution statistic. The method proposed will be to appoint the league minimum salary for each season to the *RCS*. The method of applying the league minimum wage to the *replacement player* is similar to the method used by Winston (2012) for his basketball player analysis. Once the *RCS* is established for each season, the league minimum salary will be divided by the *RCS* controlling for position, which will determine the monetary value for each point of *Contribution Score*. To determine the estimated fair contract for each player in the league, the monetary value for each *Contribution Score* will be then multiplied by each player's total *Contribution Score,* which will provide a fair estimate of a player's AAV of his contract for that season.

To properly evaluate a player's performance, it is more appropriate to consider multiple seasons as opposed to only one. A player's season can be affected by injuries, personal events or extremely good luck, which does not illustrate his true abilities. The method used in this paper will be to measure a player's weighted average estimated contract for the past three season. The weights used per season will be double the value of the previous season. This approach is used by Rob Vollman, author of multiple hockey analytics books such as Hockey Abstract (2014) and Stat Shot (2016), when evaluating players. For example, if the projected contract year will be the 2019-20 season, the estimated contract of each player for the 2018-19 season will be weighted at the value of 4, the 2017-18 estimated contract will be weighted at 2 and the 2016-17 estimated contract will be weighted at 1. The total will then be divided by the sum of the weights, 7. This will result in a weighted estimated contract. This method will eliminate over evaluation of a player who luckily outperformed his abilities for one season, as well as under evaluating a player who experienced an off year.

4.3 – Identifying Discrepancies

The work so far has focused on establishing fair salary based on player contributions. To identify the NHL's efficiency at determining player salaries, the existing contracts and the projected contracts determined in this paper will be compared. Using player contract information from the spotrac.com database (n.d), the differences between the current contracts and the projected contracts will be determined.

Section 5: Results

5.1 - Anticipated Results

The following Tables 1-4 summarize our expectations for the signs of the regressions to be analysed in this study.

Table 1

Season Regression Goals For		
Variables	Sign	Reasoning
Assists	+	Only way to get an assist is when a goal is scored, so they are highly correlated
Giveaways	-	As giveaways increase, the team has less possession of the puck, a goal cannot be scored without possession
Takeaways	+	As takeaways increase, a team is increasing its possession time
Hits For	+	As hits increase, there is a higher chance of causing turnovers
Hits Taken	-	As hits taken increase, there is a higher chance of losing the puck
Faceoff Percentage	+	As faceoff percentage increases, a team increases its possession time which would lead to scoring chances

Table 2

Season Regression Goals Against		
Variables	Sign	Reasoning
Giveaways	+	As giveaways increase, the opposing team has possession which means they have more scoring opportunities
Takeaways	-	As takeaways increase puck possession, chances against should decrease
Hits For	-	As hits increase, there is a higher chance of causing turnovers and obtain possession from the opposition
Hits Taken	+	As hits taken increase, there is a higher chance of losing the puck to the opposition and providing them with scoring chances
Faceoff Percentage	-	As faceoff percentage increases, a team increases its possession time which would lead to less scoring chances against them

Table 3

Game Regression Goals For		
Variables	Sign	Reasoning
Shots For	+	A shot for is a scoring opportunity, the more chances you have the more likely you are at scoring
Blocks Against	-	Opponent blocks the scoring chance, and stops it from getting to the goal
Misses For	-	Shot is not on goal, so it is impossible to go in
Shooting Percentage	+	For it to increase a goal must be scored

Table 4

Game Regression Goals Against		
Variables	Sign	Reasoning
Shots Against	+	More shots a team faces, the more likely they are at surrendering a goal
Block For	-	More shots are prevented from reaching the net, the less scoring chances the opponent has
Misses Against	-	The more shots that don't reach the net, the less likely they are at scoring

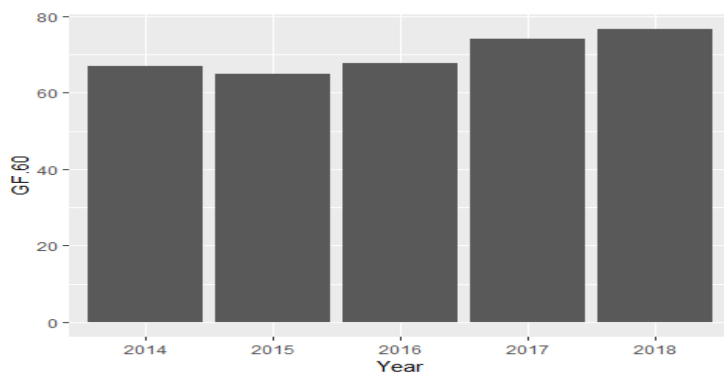
5.2 - OLS Regression Results

The regression Tables 5-8 quantify the effects of different team statistics on goal scoring on a *per 60 minutes* played ratio. The first column of the regression results in each table shows the effects on goals scoring when only the team statistics are used, the second column adds Team fixed effects to capture the differences amongst the teams, and the third column adds Year fixed effects to capture the yearly trends in goal scoring. The analysis of the regression results will focus on the fully controlled model.

The Team controls are included in the models in order to control for potential differences between teams. The intent is to account for the different coaching and ownership strategies implemented by each team. The addition of this set of controls is justified, because the controls add significant information to the majority of the models. The coefficients of the separate team control variables are generally not very large and not statistically significant. It is important, however, to control for different team playing styles as it does impact goal scoring.

The Year controls are included in the models to account for possible scoring trends from one season to the next. As can be seen from Figure 2, the coefficients are consistent with the average yearly scoring trends. To have a better understanding of the impact of the statistics used in the models, the goal scoring trends from one season to the next needed to be accounted for.

Figure 2 - Season Goal Totals



5.2.1 - Season Level Regressions

The season regressions focus more on statistics that dictate the flow of the game and less on direct scoring chances by teams. Having the data at the season aggregate level could lead to small sample issues and many of these statistics, like *hits, giveaways* and *takeaways*, are subjectively accumulated by an individual watching the game.

5.2.1.1 - Goals For (GF) Regressions

Regression Table 5 quantifies the effects of different team statistics on a team's goal scoring *per 60 minutes* played at the season level. The first three columns show the results for the first version of the regression. This regression uses the team total *primary assists*, *secondary assists*, *giveaways*, *takeaways*, *hit*s and *faceoff percentages* as independent variables. The three last columns show the results for the second version of the regression. This version uses the same variables as the first with the exception of replacing *primary* and *secondary assists* with the single variable *total assists*. Commentary will focus on equations (1.3) and (2.3) with the full set of controls.

5.2.1.1.1 - Version 1 Results

The results show that *primary assists* have a large positive and statistically significant effect on a team's ability to score goals. As a team's *primary assists* increase by 1 assist *per 60 minutes* of play in a season, their amount of goals scored in a season increases by 1.061 goals *per 60 minutes* of play, holding everything else constant in equation (1.3). The coefficient is statistically significant at the 0.1 percent level of significance. This is logical as *primary assists* can only be awarded when a goal is scored.

The effects of *secondary assists* on goal scoring, however, is quite small and the sign is counterintuitive. *Secondary assists* appear to have a very low impact on a team's ability to score, and the coefficient is not statistically different from zero.

Giveaways have a negative effect on goal scoring, as expected, and is quite small. As the amount of a team's *giveaways* increase by 1 *giveaway per 60 minutes* of play, their amount of goals scored decreases by 0.010 goals *per 60 minutes* of play, holding everything else constant in equation (1.3). The coefficient is statistically significant at the 5 percent level of significance. This result follows the logic that *giveaways* surrender puck possession to the opposing team, and a team cannot score goals without puck possession.

Takeaways have a small positive effect on goal scoring. As a team's amount of *takeaways* increase by 1 *takeaway per 60 minutes* of play, their amount of goals scored increase by 0.015 goals *per 60 minutes* of play, holding everything else constant. The coefficient is statistically significant at the 5 percent level of significance. The magnitude of the effect of *takeaways* is similar to that of *giveaways* with the exception of the opposite sign. It is reasonable that *giveaways* and *takeaways* have a similar impact as they are their near opposite. Increasing *takeaways*, increases puck possession which results in greater scoring for the observed team.

The results of physical play, as seen in the variables *hits for* and *hits taken* coefficients, require a more in-depth interpretation. As for *hits for*, the sign changes as we add the sets of controls, and in the final model has a very small and positive impact on a team's scoring ability. As the amount of *hits* delivered by a team increases by 1 *hit per 60 minutes* of play, their amount of goals scored increases by 0.001 goals *per 60 minutes* of play, holding everything else constant. *Hits taken,* however, has a negative impact on goal scoring in column (1.1) and (1.2), but changes to positive in column (1.3) once the year is accounted for. Neither coefficient for *hits* is statistically significant. This can perhaps be interpreted as saying that a team can only be *hit* when they have puck possession. Said differently, the more *hits* a team receives, the more often they have the puck. The positive effect of *hits* for can be interpreted as a team causing turnovers with their physical play, increasing scoring chances for their team.

The effect of a team's *faceoff percentage* on a team's ability to score goals is positive. As a team's *faceoff percentage* increases by 1 percent, their amount of goals scored increases by 0.077 goals *per 60 minutes* of play, holding everything else constant. The coefficient is not statistically different than zero. This is logical because as a team wins more faceoffs, they increase their puck possession and can create scoring opportunities.

F-statistics tests were run to obtain the joint significance level of the Team and Year controls (see section 5.4 below). It is determined that for this model, both sets of control variables fail to be statistically significant and therefore do not add any significant information to the model.

Overall, the only coefficients with statistical significance are *primary assists*, *giveaways* and *takeaways*. This argues that physical play does not have a significant impact on a team's offensive success. Surprisingly, faceoff percentage appears to not have a significant impact either, which contradicts expectation. However, this may simply reflect the relatively small number of observations and degrees of freedom available at the season level of the analysis.

5.2.1.1.2 - Version 2 Results

When *total assists* are used, the results show that assists have a positive effect on a team's ability to score goals. As a team's total assists increase by 1 assist *per 60 minutes* of play, their amount of goals scored increase by 0.54 goals *per 60 minutes* of play, holding other factors constant in equation (2.3). This result is significant at the 0.1 percent level of significance. This result follows the logic that assists can only be awarded when a goal is scored. The result that the coefficient is worth half a goal reflects the fact that two assists can potentially be awarded for each goal scored, therefore rendering it worth nearly half as much.

In this second regression, the effect of *giveaways* remains negative but increases in magnitude. As the amount of a team's *giveaways* increase by 1 *giveaway per 60 minutes* of play, their amount of goals scored decrease by 0.013 goals *per 60 minutes* of play, holding other factors constant. This result is statistically significant at the 5 percent level of significance.

Similar to the results of *giveaways*, the effects of *takeaways* in this second regression is larger in magnitude and remains the same sign compared to the first regression. As the number of *takeaways* for a team increases by 1 *takeaway per 60 minutes* of play, their amount of goals scored increases by 0.026 goals *per 60 minutes* of play, holding other factors constant. This result is significant at the 1 percent level of significance.

The effect of physical play on a team's ability to score goals does change in this second regression compared to the first. Both *hit* coefficients are not at all statistically different from zero. Although insignificant, these results could be an argument for teams to focus more on skilled players over physically dominating players. These results could also simply be showing how today's game has become less physical.

The biggest difference in effects between the two *goals for* regressions at the season level is the *faceoff percentage* coefficient. In the second regression, the effect of *faceoff percentage* changes signs to have a negative effect on a team's ability to score goals. This result is also not at all statistically significant.

After running F-statistics tests to obtain the significance level of the Team and Year controls, it is determined that both controls are indeed significant and add relevant information to the model.

Overall this second regression model for a team's goal scoring ability at the season level does give interesting results. The measured effects of *total assists* do make more sense with a value of 0.5 goals compared to the effects of primary and *secondary assists*, especially since the *secondary assists'* coefficient is negative. Generally, the other variables do not change very much, other than increased statistical significance and slightly greater magnitude. This second regression, however, does add confusion for the *faceoff percentage* variable as now it is determined to have a counterintuitive negative effect.

5.2.1.2 - Goals Against Regression

The regressions in Table 6 quantify the effects of different team statistics on a team's goals allowed *per 60 minutes* played at the season level. Commentary will focus on the results of equation (3) which include all the Team and Year controls.

As the results show, the effects of *giveaways* on goals allowed is small and positive. As a team's amount of *giveaways* increase by 1 *giveaway per 60 minutes* of play, their amount of goals allowed increases by 0.0022 Goals *per 60 minutes* of play, holding everything else constant. The coefficient is not statistically different from zero. This can be interpreted as, when a team gives up possession of the puck they allow more scoring opportunities to the opposing team.

Unlike *giveaways*, the results of *takeaways* are counterintuitive. The results show that a *takeaway* also has a positive effect on a team's goals allowed, but the coefficient is not statistically different from zero. This is counterintuitive because as a team increases *takeaways*, it steals possession of the puck from the opposition which should prevent the opposing team from scoring goals. A possible explanation for the positive results is that, to have a *takeaway* the opposing team must first have possession of the puck meaning they have more overall puck possession, but this is not a strong explanation.

Much like in the *goals for* regression from Table 5, the effect of physical play on *goals allowed* is difficult to comprehend. *Hits for* has a small positive effect on a team's goals allowed, meaning the more a team plays physical the more goals are scored against them. The coefficient is again not statistically different than zero. Although this interpretation goes against the definition of a *hit*, where a *hit* is physical contact that leads a loss of possession of the puck, the positive effect does make sense. For a team to make a *hit*, the opposing team needs to have puck possession, meaning the opposing team with possession has more opportunities to score.

The *hits taken* variable has a small, significant and negative effect on a team's goals allowed. As the amount of *hits* received by a team increases by 1 *hit per 60 minutes* of play, their amount of goals allowed decreases by 0.0319 goals *per 60 minutes* of play, holding everything else constant. The coefficient is statistically significant at the 1 percent level of significance. Similar to the interpretation for *hits for*, for a *hit* to be delivered, the receiving team must have possession of the puck which means they can produce more scoring chances.

The effect of *faceoff percentage* on *goals against* is counter intuitive, especially after analysing its effects on goals from Table 5 column (1.3), but follows that of column (2.3). The regression results show that a team's *faceoff percentage* has a positive effect on the amount of goals they allow. Said differently, as a team's *faceoff percentage* improves, the team allows more goals. The coefficient, however, is not at all statistically different than zero.

After running F-statistics tests to obtain the significance level of the Team and Year controls, it is determined that both sets of controls are significant and add relevant information to the model.

Following the footsteps of the past literature, these results show that defensive play is difficult to measure. With an R^2 of 0.48 in the fully controlled model in column (3), there is much to be desired. The only statistically significant variable in this model is *hits* taken, which carries with it a questionable result. Like in the *goals for* model from Table 5, the effects of *faceoff percentage* are unusual. In this case, *faceoff percentage* has a positive affect on goals allowed which contradicts general belief. This could be caused by the small sample size and that *faceoff percentage* was taken in aggregate at the season level. Perhaps a clearer understanding of the effects could be obtained using individual game data for *faceoff percentages*.

5.2.2 - Game Level Regressions

The data used for the game level regressions are statistics that directly measure scoring chances. These regressions have large samples and should provide more accurate insight as these involve more reliable statistics.

5.2.2.1 - Goals For (GF) Regressions

The regressions in Table 7 quantify the effects of different team statistics on a Team's goal scoring *per 60 minutes* played at the individual game level. There are thus over twelve thousand observations used for each regression.

The results show that *shots taken* by a team have a positive effect on a team's goal scoring abilities. As a team's *shots for* increase by 1 *shot per 60 minutes* of play, their amount of goals scored increases by 0.076 goals *per 60 minutes* of play in equation (3), holding other factors constant. This coefficient is statistically significant at the 0.1 percent level of significance. This certainly does make sense as a goal cannot be scored without a shot directed at the goal.

Shot attempts taken by the observed team that are blocked by the opposing team's players is also shown to have a positive effect on that team's goal scoring abilities, however it is a small effect and is not statistically different from zero. A *block against* can be interpreted as a possible scoring chance for the attacking team, and thus a positive coefficient does have some reasoning because a partially blocked shot does still have a chance of scoring.

A *shot attempt* taken by team *i* that is unblocked but misses the net generates a negative effect on that team's scoring ability. As the amount of *missed shots* increase by 1 additional *missed shot per 60 minutes* of play, the team's amount of goals scored decreases by 0.0025 goals *per 60 minutes* of play, holding everything else constant, and the coefficient is significant at the 1 percent level of significance. It is impossible for a *shot attempt* that misses the net completely to be counted as a goal, therefore a negative sign is indeed reasonable.

A team's *shooting percentage* has a positive and large effect on the team's scoring ability. As a team's *shooting percentage* increases by 1 percent, their amount of goals scored increases by 0.274 Goals *per 60 minutes* of play, holding everything else constant and the coefficient is also significant at the 0.1 percent level of significance. This result again makes sense as their *shooting percentage* represents a team's scoring success from their *shots taken*.

F-statistics tests were run to obtain the significance level of the sets of Team and Year controls. It is determined that for this model, both sets of control variables fail to be statistically significant and so do not add any significant information to the model.

Overall the recorded statistics at the individual game level have a lot of explanatory power for a team's offensive ability. All three versions of the model have R^2s of 0.94, and the majority of the non-control variables used are very statistically significant. This adds to the concept that the current statistics being tracked do well at measuring teams' offensive abilities.

5.2.2.2 - Goals Against (GA) Regressions

The regressions in Table 8 quantify the effects of different team statistics on a team's goals allowed *per 60 minutes* played at the individual game level.

The results show that *shots against* a team has a positive effect on the amount of goals they allow. As the number of *shots* a team faces increases by 1 additional *shot per 60 minutes* played, their amount of goals allowed increases by 0.045 Goals *per 60 minutes* of play in equation (3), holding everything else constant. The coefficient is also statistically significant at the 0.1 percent level of significance. As a team allows a higher volume of *shots*, it is expected that they will also allow a higher volume of goals, which is demonstrated here.

An opponent's *shot attempt* that is blocked by a player of the observed team still has a positive effect on the number of goals allowed. As the number of blocks by the team increases by 1 additional *block per 60 minutes* played, their amount of goals allowed increases by 0.03 goals *per 60 minutes* of play, holding everything else constant. The coefficient is again statistically significant at the 0.1 percent level of significance. Although the initial belief is that blocking shots would prevent goals, a *blocked shot* is still considered a *shot attempt* by the opposition, meaning they are generating offense. A *blocked shot* does, however, have a smaller coefficient than a *shot against.* This difference signifies that blocking shots does lower the quality of the scoring chance compared to an unblocked shot on goal.

An opponent's *shot attempt* that misses the net has a negative effect on the number of goals allowed by the team observed. As the number of *missed shots against* increases by 1 additional *missed shot per 60 minutes* played, the amount of goals allowed by the observed team decreases by 0.036 goals *per 60 minutes* of play, holding everything else constant. The coefficient is also statistically significant at the 0.1 percent level of significance. This can be interpreted as the defending team preventing the opponent from having good scoring opportunities and thus playing well defensively.

After running F-statistics tests to obtain the joint significance level of the sets of Team and Year controls, it is determined that both sets of controls are significant and add relevant information to the model.

With small values for the R^2 measures, the GA models follow along with the largest issue in hockey analytics. It is difficult to quantify defensive ability with the data available. Although the overall measure of explanatory power is small in these models, all of the non-control variables used are all strongly individually statistically significant.

5.3 - Other Regression Results

To gain further understanding of the robustness of the above results, some additional regressions were ran. Tables 9 and 10 show the results using Seemingly Unrelated Regression equations (Greene, 2017, p.328), and Table 11 shows the results of the OLS regressions using natural logs. The results do not provide much additional insight to the OLS regressions already discussed so no further comment is required here.

5.4 - F-Test Results

The following Tables 12-13 show the results of the F-statistic hypothesis tests conducted for the OLS regression models. The hypothesis tested here is if the set of control variables are equal to zero. The equation used to find the F-statistic is:

$$F\ Statistic = \frac{\frac{SSR1 - SSR2}{m}}{\left(\frac{SSR2}{(n-k-1)}\right)}.$$

The SSR1 corresponds to the Sum of Squared Residuals (SSR) for the regression without the control being tested in the model. SSR2 corresponds to the SSR for the regression that included the control being tested, m corresponds to the number of variables in the control being tested, n corresponds to the number of observations in that specific model, and k corresponds to the number of variables in the model including the control variables. For the hypothesis to be proven false, i.e. the control variable as a group are statistically different from zero, the estimated F-statistic must be greater than the corresponding F-critical value found in the F-statistic table, at the 5 percent level of significance.

Table 5: Goals For Regression at the Season Level

	(1.1)	(1.2)	(1.3)	(2.1)	(2.2)	(2.3)
(Intercept)	0.004618	0.1014	-0.01572	0.2445	0.3476	0.118
	(0.105)	(0.149)	(0.155)	(0.168)	(0.216)	(0.218)
A1.60	1.143 ***	1.086 ***	1.061 ***			
	(0.04)	(0.049)	(0.050)			
A2.60	-0.1464 **	-0.06533	-0.02569			
	(0.046)	(0.052)	(0.054)			
Ast.60				0.5478 ***	0.5328 ***	0.5406 ***
				(0.013)	(0.015)	(0.015)
Giveaways. 60	-0.003317	-0.008695 *	-0.0103 *	-0.006682	-0.007942	-0.013345 *
	(0.002)	(0.004)	(0.004)	(0.003)	(0.006)	(0.006)
Takeaways. 60	0.005875	0.01165	0.01499 *	0.009922 *	0.02458 **	0.02588 **
	(0.003)	(0.006)	(0.006)	(0.005)	(0.009)	(0.009)
HitsF.60	0.001063	-0.00003724	0.0007229	0.0001274	-0.003939	-0.002028
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
HitsT.60	-0.00003884	-0.001675	0.0001343	0.0001123	-0.001192	0.003258
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002716)	(0.003)
Face.Perc.	0.07793	0.06825	0.07737	-0.1575	-0.04961	-0.01339
	(0.194)	(0.236)	(0.231)	(0.3123)	(0.3456)	(0.327)
Team Arizona Coyotes		-0.01383	-0.01826		-0.03056	-0.04608
		(0.031)	(0.031)		(0.045)	(0.043)
Team Boston Bruins		-0.01425	-0.02616		-0.03057	-0.05021
		(0.031)	(0.031)		(0.045)	(0.043)
Team Buffalo Sabres		-0.02948	-0.02070		-0.06881	-0.05141

	(0.033)	(0.033)	(0.048)	(0.046)
Team Calgary Flames	-0.08525 *	-0.08358 *	-0.2091 ***	-0.1696 **
	(0.036)	(0.037)	(0.051)	(0.051)
Team Carolina Hurricanes	-0.06835	-0.08171 *	-0.1665 **	-0.1671 **
	(0.041)	(0.040)	(0.058)	(0.056)
Team Chicago Blackhawks	-0.02988	-0.03573	-0.09120	-0.09503
	(0.037)	(0.037)	(0.054)	(0.051)
Team Colorado Avalanche	-0.08062 *	-0.07913 *	-0.1095 *	-0.1010 *
	(0.036)	(0.035)	(0.052)	(0.050)
Team Columbus Blue Jackets	-0.02504	-0.02366	0.01207	0.01245
	(0.035)	(0.035)	(0.051)	(0.049)
Team Dallas Stars	-0.01511	-0.01087	-0.04594	-0.02451
	(0.029)	(0.029)	(0.043)	(0.041)
Team Detroit Red Wings	-0.02321	-0.01420	-0.04975	-0.03003
	(0.030)	(0.029)	(0.043)	(0.042)
Team Edmonton Oilers	-0.03582	-0.03717	-0.04445	-0.03514
	(0.031)	(0.030)	(0.045)	(0.043)
Team Florida Panthers	-0.01675	-0.020324	-0.06063	-0.05182
	(0.031)	(0.031)	(0.046)	(0.044)
Team Los Angeles Kings	0.02017	0.01816	0.03562	0.02229
	(0.028)	(0.028)	(0.041)	(0.039)
Team Minnesota Wild	-0.04113	-0.03383	-0.05448	-0.03801
	(0.038)	(0.038)	(0.056)	(0.054)
Team Montreal Canadiens	-0.009781	-0.01778	-0.02462	-0.03641
	(0.030)	(0.030)	(0.044)	(0.042)
	-0.02560	-0.02092	-0.0002982	0.01234

Team Nashville Predators	(0.032)	(0.032)	(0.047)	(0.045)
Team New Jersey Devils	-0.04570	-0.04084	-0.03633	-0.03037
	(0.034)	(0.034)	(0.050)	(0.048)
Team New York Islanders	-0.007715	-0.01409	-0.003886	-0.005684
	(0.030)	(0.030)	(0.044)	(0.042)
Team New York Rangers	0.01033	0.008026	-0.02351	-0.01033
	(0.032)	(0.032)	(0.046)	(0.045)
Team Ottawa Senators	-0.0066	-0.01252	-0.01535	-0.01972
	(0.029)	(0.029)	(0.043)	(0.041)
Team Philadelphia Flyers	-0.04758	-0.05530 *	-0.08686 *	-0.1026 **
	(0.028)	(0.028)	(0.041)	(0.039)
Team Pittsburgh Penguins	-0.01733	-0.03586	-0.00003987	-0.04144
	(0.032)	(0.032)	(0.046)	(0.045)
Team San Jose Sharks	-0.05605	-0.06416	-0.1760 ***	-0.1643 ***
	(0.034)	(0.033)	(0.047)	(0.046)
Team St Louis Blues	-0.07778 *	-0.08553 *	-0.09984	-0.1144 *
	(0.036)	(0.036)	(0.053)	(0.051)
Team Tampa Bay Lightning	-0.04797	-0.05170	-0.01368	-0.02191
	(0.032)	(0.031)	(0.046)	(0.044)
Team Toronto Maple Leafs	0.03149	0.01680	0.004403	-0.01096
	(0.034)	(0.034)	(0.049)	(0.048)
Team Vancouver Canucks	-0.006007	0.0001527	-0.05056	-0.03498
	(0.034)	(0.033)	(0.049)	(0.047)
Team Vegas Golden Knights	-0.01131	-0.03157	-0.006922	-0.03641
	(0.049)	(0.048)	(0.071)	(0.068)

Team Washington Capitals		-0.07710 *	-0.08318 **		-0.08357	-0.08323
		(0.030)	(0.030)		(0.045)	(0.043)
Team Winnipeg Jets		-0.02337	-0.02643		-0.05520	-0.04856
		(0.031)	(0.030)		(0.045)	(0.043)
Year 2015			0.03566 **			0.05581 **
			(0.012)			(0.017)
Year 2016			0.03282 *			0.07239 ***
			(0.014)			(0.019)
Year 2017			0.02597			0.06186 **
			(0.014)			(0.020)
Year 2018			0.02574			0.06638 **
			(0.015)			(0.021)
N	152	152	152	152	152	152
R2	0.98	0.98	0.99	0.94	0.96	0.97
SSR	0.302761	0.210826	0.194467	0.800749	0.456558	0.395100

Standard errors in parentheses

*** p < 0. 001; ** p< 0.01; * p < 0.05.

Table 6: Goals Against Regression at the Season Level

	(1)	(2)	(3)
(Intercept)	2.738 ***	1.575	2.038 *
	(0.576)	(0.804)	(0.822)
Giveaways.60	0.02034	0.02804	0.002226
	(0.012)	(0.022)	(0.023)
Takeaways.60	0.04382 *	0.07279 *	0.03619
	(0.017)	(0.032)	(0.033)
HitsF.60	0.004992	0.01177	0.009882
	(0.006)	(0.008)	(0.008)
HitsT.60	-0.01852 **	-0.03251 **	-0.03192 **
	(0.007)	(0.01)	(0.011)
Face.Perc	-1.134	0.7484	0.8037
	(1.095)	(1.322)	(1.287)
Team Arizona Coyotes		0.4136 *	0.3418 *
		(0.171)	(0.169)
Team Boston Bruins		0.02303	0.05292
		(0.172)	(0.171)
Team Buffalo Sabres		0.4042 *	0.3311
		(0.180)	(0.178)
Team Calgary Flames		-0.1459	-0.01498
		(0.195)	(0.202)
Team Carolina Hurricanes		0.07624	0.2134
		(0.218)	(0.217)
Team Chicago Blackhawks		0.3510	0.3477
		(0.207)	(0.202)
Team Colorado Avalanche		0.2455	0.2081
		(0.200)	(0.197)
Team Columbus Blue Jackets		0.1466	0.09534
		(0.195)	(0.191)
Team Dallas Stars		0.01942	0.07172
		(0.163)	(0.162)
Team Detroit Red Wings		0.2648	0.2252
		(0.165)	(0.163)
Team Edmonton Oilers		0.2823	0.3907 *
		(0.169)	(0.169)
Team Florida Panthers		0.1091	0.1886
		(0.175)	(0.173)

Team Los Angeles Kings	0.07271	0.01956
	(0.157)	(0.155)
Team Minnesota Wild	0.05826	-0.02370
	(0.213)	(0.212)
Team Montreal Canadiens	0.1516	0.2052
	(0.169)	(0.166)
Team Nashville Predators	-0.1572	-0.1580
	(0.179)	(0.176)
Team New Jersey Devils	0.2281	0.1920
	(0.188)	(0.184)
Team New York Islanders	0.07184	0.1564
	(0.168)	(0.165)
Team New York Rangers	0.003125	0.1082
	(0.1762)	(0.176)
Team Ottawa Senators	0.3397 *	0.3940 *
	(0.165)	(0.162)
Team Philadelphia Flyers	0.2860	0.2446
	(0.156)	(0.153)
Team Pittsburgh Penguins	0.2969	0.2859
	(0.175)	(0.174)
Team San Jose Sharks	0.07058	0.1867
	(0.179)	(0.179)
Team St Louis Blues	-0.02558	-0.08973
	(0.203)	(0.200)
Team Tampa Bay Lightning	0.1125	0.08583
	(0.172)	(0.168)
Team Toronto Maple Leafs	0.2659	0.3757 *
	(0.188)	(0.187)
Team Vancouver Canucks	0.3485	0.2988
	(0.186)	(0.183)
Team Vegas Golden Knights	-0.03715	0.04995
	(0.272)	(0.269)
Team Washington Capitals	-0.1394	-0.07326
	(0.169)	(0.166)
Team Winnipeg Jets	-0.003189	0.03505
	(0.171)	(0.168)
Year 2015		-0.09036
		(0.066)
Year 2016		-0.05859

			(0.074)
Year 2017			0.05727
			(0.078)
Year 2018			0.1364
			(0.082)
N	152	152	152
R2	0.16	0.43	0.48
SSR	9.897810	6.749412	6.174038

Standard errors in parentheses

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 7: Goals For Regression at the Game Level

	(1)	(2)	(3)
(Intercept)	-2.150 ***	-2.144 ***	-2.142 ***
	(0.023)	(0.031)	(0.031)
SF.60	0.07657 ***	0.07656 ***	0.07644 ***
	(0.001)	(0.001)	(0.001)
Block.A.60	0.0012454	0.001126	0.001064
	(0.001)	(0.001)	(0.001)
Missed.F.60	-0.002192 **	-0.002391 **	-0.002453 **
	(0.001)	(0.001)	(0.001)
SH.Perc	0.2744 ***	0.2743 ***	0.2743 ***
	(0.001)	(0.001)	(0.001)
Team Arizona Coyotes		-0.02902	-0.02916
		(0.029)	(0.029)
Team Boston Bruins		-0.03902	-0.03885
		(0.029)	(0.029)
Team Buffalo Sabres		-0.03648	-0.03679
		(0.029)	(0.029)
Team Calgary Flames		0.01794	0.01804
		(0.029)	(0.029)
Team Carolina Hurricanes		-0.02611	-0.02599
		(0.029)	(0.029)
Team Chicago Blackhawks		-0.0003691	-0.0001947
		(0.029)	(0.029)
Team Colorado Avalanche		0.009983	0.009886
		(0.029)	(0.029)
Team Columbus Blue Jackets		-0.01242	-0.01228
		(0.029)	(0.029)
Team Dallas Stars		-0.01127	-0.01102
		(0.029)	(0.029)
Team Detroit Red Wings		0.006416	0.006111
		(0.029)	(0.029)
Team Edmonton Oilers		-0.01260	-0.01260
		(0.029)	(0.029)
Team Florida Panthers		-0.01904	-0.01903
		(0.029)	(0.029)
Team Los Angeles Kings		0.02005	0.02010
		(0.029)	(0.029)

Team Minnesota Wild	-0.008147	-0.008081
	(0.029)	(0.029)
Team Montreal Canadiens	0.0006193	0.0008460
	(0.029)	(0.029)
Team Nashville Predators	0.01211	0.01238
	(0.029)	(0.029)
Team New Jersey Devils	0.001708	0.001231
	(0.029)	(0.029)
Team New York Islanders	0.04040	0.04059
	(0.029)	(0.029)
Team New York Rangers	-0.03036	-0.03038
	(0.029)	(0.029)
Team Ottawa Senators	0.02348	0.02348
	(0.029)	(0.029)
Team Philadelphia Flyers	0.02261	0.02263
	(0.029)	(0.029)
Team Pittsburgh Penguins	-0.01114	-0.01075
	(0.029)	(0.029)
Team San Jose Sharks	-0.01032	-0.01003
	(0.029)	(0.029)
Team St Louis Blues	0.005952	0.006028
	(0.029)	(0.029)
Team Tampa Bay Lightning	0.02099	0.02116
	(0.029)	(0.029)
Team Toronto Maple Leafs	0.03699	0.03717
	(0.029)	(0.029)
Team Vancouver Canucks	-0.01530	-0.01554
	(0.029)	(0.029)
Team Vegas Golden Knights	0.007704	-0.00001588
	(0.038)	(0.039)
Team Washington Capitals	-0.005784	-0.005680
	(0.029)	(0.029)
Team Winnipeg Jets	-0.007080	-0.007044
	(0.029)	(0.029)
Year 2015		-0.008514
		(0.012)
Year 2016		0.003394
		(0.012)
Year 2017		0.0009429

			(0.012)
Year 2018			0.02277
			(0.012)
N	12464	12464	12464
R2	0.94	0.94	0.94
SSR	2154.96	2149.98	2148.65

Standard errors in parentheses

*** p < 0.001; ** p < 0.01; * p < 0.05

Table 8: Goals Against Regression at the Game Level

	(1)	(2)	(3)
(Intercept)	0.8695 ***	0.8161 ***	0.8327 ***
	(0.086)	(0.118)	(0.121)
SA.60	0.04763 ***	0.04624 ***	0.04497 ***
	(0.002)	(0.002)	(0.002)
Block.F.60	0.029002 ***	0.03076 ***	0.029999 ***
	(0.003)	(0.003)	(0.003)
Missed.A.60	-0.03399 ***	-0.035802 ***	-0.03648 ***
	(0.003)	(0.003)	(0.003)
Team Arizona Coyotes		0.2109	0.2151
		(0.115)	(0.115)
Team Boston Bruins		-0.02253	-0.02421
		(0.115)	(0.115)
Team Buffalo Sabres		0.3043 **	0.3067 **
		(0.116)	(0.115)
Team Calgary Flames		0.1911	0.1916
		(0.115)	(0.115)
Team Carolina Hurricanes		0.2677 *	0.2673 *
		(0.115)	(0.115)
Team Chicago Blackhawks		-0.08074	-0.07749
		(0.116)	(0.116)
Team Colorado Avalanche		0.1783	0.1815
		(0.115)	(0.115)
Team Columbus Blue Jackets		0.1248	0.1253
		(0.115)	(0.115)
Team Dallas Stars		0.05506	0.05641
		(0.115)	(0.115)
Team Detroit Red Wings		0.1280	0.1273
		(0.115)	(0.115)
Team Edmonton Oilers		0.3300 **	0.3316 **
		(0.115)	(0.115)
Team Florida Panthers		0.1867	0.1864
		(0.115)	(0.115)
Team Los Angeles Kings		-0.07428	-0.07593
		(0.115)	(0.115)
Team Minnesota Wild		0.06663	0.06523
		(0.115)	(0.115)

Team Montreal Canadiens	-0.1117	-0.1101
	(0.116)	(0.116)
Team Nashville Predators	-0.1557	-0.1559
	(0.115)	(0.115)
Team New Jersey Devils	0.1682	0.1664
	(0.116)	(0.115)
Team New York Islanders	0.1024	0.1046
	(0.115)	(0.115)
Team New York Rangers	0.1043	0.1065
	(0.115)	(0.115)
Team Ottawa Senators	0.2595 *	0.2639 *
	(0.115)	(0.115)
Team Philadelphia Flyers	0.04458	0.04576
	(0.115)	(0.115)
Team Pittsburgh Penguins	-0.01586	-0.01562
	(0.115)	(0.115)
Team San Jose Sharks	0.1471	0.1465
	(0.116)	(0.115)
Team St Louis Blues	-0.08038	-0.08315
	(0.115)	(0.115)
Team Tampa Bay Lightning	-0.02078	-0.02114
	(0.115)	(0.115)
Team Toronto Maple Leafs	0.1844	0.1900
	(0.116)	(0.115)
Team Vancouver Canucks	0.2100	0.2114
	(0.115)	(0.115)
Team Vegas Golden Knights	0.1225	0.03831
	(0.153)	(0.154)
Team Washington Capitals	-0.06105	-0.06065
	(0.115)	(0.115)
Team Winnipeg Jets	0.06414	0.06417
	(0.115)	(0.115)
Year 2015		-0.06087
		(0.047)
Year 2016		0.006740
		(0.047)
Year 2017		0.08019
		(0.047)
Year 2018		0.1691 ***

			(0.047)
N	12464	12464	12464
R2	0.05	0.05	0.06
SSR	34043.54	33842.6	33766.88

Standard errors in parentheses

*** p < 0.001; ** p < 0.01; * p < 0.05

Table 9: Season Level SUR Regression Results

VARIABLES	(1) GF.SUR V1	(2) GA.SUR V1	(3) GF.SUR V2	(4) GA.SUR V2
A160	1.053***			
	(0.0416)			
A260	-0.0170			
	(0.0451)			
Giveaways60	-0.0103***	0.00223	-0.0133***	0.00223
	(0.00359)	(0.0201)	(0.00511)	(0.0201)
Takeaways60	0.0151***	0.0362	0.0259***	0.0362
	(0.00520)	(0.0286)	(0.00731)	(0.0286)
HitsF60	0.000682	0.00988	-0.00203	0.00988
	(0.00131)	(0.00708)	(0.00184)	(0.00708)
HitsT60	0.000182	-0.0319***	0.00326	-0.0319***
	(0.00174)	(0.00969)	(0.00245)	(0.00969)
FacePerc	0.000760	0.00804	-0.000133	0.00804
	(0.00196)	(0.0110)	(0.00280)	(0.0110)
Ast60			0.541***	
			(0.0126)	
Constant	-0.0138	2.038***	0.118	2.038***
	(0.131)	(0.705)	(0.187)	(0.705)
Observations	152	152	152	152
R-squared	0.985	0.475	0.970	0.475

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Regressions Contain Team and Year controls

Table 10: Game Level SUR Regression Equations

VARIABLES	(1) SUR	(2) SUR
SF60	0.0764***	
	(0.000552)	
BlockA60	0.00127*	
	(0.000722)	
MissedF60	-0.00253***	
	(0.000859)	
SHPerc	0.274***	
	(0.000642)	
SA60		0.0449***
		(0.00218)
BlockF60		0.0303***
		(0.00287)
MissedA60		-0.0365***
		(0.00337)
Constant	-2.143***	0.829***
	(0.0314)	(0.121)
Observations	12,464	12,464
R-squared	0.940	0.055

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Regressions Contain Team and Year controls

Table 11: Log Season Regression Results

	log Season GF equation (1.1)	log Season GF equation (1.2)	log Season GA equation (2)
(Intercept)	0.01734	-0.5386 ***	1.324 *
	(0.104)	(0.15)	(0.552)
ln.A1.60	0.9923 ***		
	(0.047)		
ln.A2.60	-0.01644		
	(0.04)		
ln.Ast.60		0.912 ***	
		(0.026)	
ln.Giveaways.60	-0.03544 *	-0.06015 *	0.004775
	(0.017)	(0.024)	(0.091)
ln.Takeaways.60	0.05049 *	0.09828 ***	0.1445
	(0.02)	(0.027)	(0.102)
ln.HitsF.60	0.008967	-0.01958	0.1315
	(0.018)	(0.025)	(0.093)
ln.HitsT.60	0.007009	0.04313	-0.3638 **
	(0.023)	(0.032)	(0.122)
ln.Face.Perc	0.01200	-0.005328	0.1323
	(0.051)	(0.072)	(0.271)
Team Arizona Coyotes	-0.007123	-0.02720	0.1421
	(0.014)	(0.020)	(0.075)
Team Boston Bruins	-0.01206	-0.02802	0.009248
	(0.014)	(0.019)	(0.073)
Team Buffalo Sabres	-0.008549	-0.02643	0.1361
	(0.015)	(0.021)	(0.078)
Team Calgary Flames	-0.3585 *	-0.07611 **	-0.02384
	(0.017)	(0.023)	(0.087)
Team Carolina Hurricanes	-0.03733 *	-0.08183 ***	0.08497
	(0.017)	(0.024)	(0.089)
Team Chicago Blackhawks	-0.01704	-0.04852 *	0.1425
	(0.017)	(0.023)	(0.088)
Team Colorado Avalanche	-0.03642 *	-0.05450 *	0.07228
	(0.017)	(0.023)	(0.088)
Team Columbus Blue Jackets	-0.01241	-0.007707	0.02801
	(0.016)	(0.023)	(0.086)
Team Dallas Stars	-0.006932	-0.01537	0.02433

	(0.013)	(0.018)	(0.070)
Team Detroit Red Wings	-0.005022	-0.01298	0.1012
	(0.013)	(0.019)	(0.070)
Team Edmonton Oilers	-0.02111	-0.02383	0.1488 *
	(0.014)	(0.019)	(0.072)
Team Florida Panthers	-0.01285	-0.03196	0.0618
	(0.014)	(0.020)	(0.075)
Team Los Angeles Kings	0.01192	0.01737	0.008737
	(0.013)	(0.018)	(0.067)
Team Minnesota Wild	-0.0167	-0.02866	-0.01203
	(0.018)	(0.026)	(0.098)
Team Montreal Canadiens	-0.01051	-0.01730	0.07385
	(0.013)	(0.018)	(0.069)
Team Nashville Predators	-0.01023	-0.002894	-0.07909
	(0.014)	(0.020)	(0.076)
Team New Jersey Devils	-0.01991	-0.01995	0.07002
	(0.015)	(0.022)	(0.080)
Team New York Islanders	-0.009206	-0.01060	0.05010
	(0.013)	(0.019)	(0.071)
Team New York Rangers	-0.0009514	-0.01327	0.02858
	(0.014)	(0.020)	(0.076)
Team Ottawa Senators	-0.006927	-0.01556	0.1425 *
	(0.013)	(0.018)	(0.070)
Team Philadelphia Flyers	-0.0248 *	-0.04756 **	0.1002
	(0.012)	(0.017)	(0.065)
Team Pittsburgh Penguins	-0.01846	-0.02774	0.1042
	(0.014)	(0.020)	(0.074)
Team San Jose Sharks	-0.032*	-0.07757 ***	0.06998
	(0.015)	(0.020)	(0.077)
Team St Louis Blues	-0.03977 *	-0.06508 **	-0.04995
	(0.018)	(0.025)	(0.093)
Team Tampa Bay Lightning	-0.02308	-0.01643	0.03167
	(0.014)	(0.020)	(0.072)
Team Toronto Maple Leafs	0.00189	-0.01939	0.14925
	(0.015)	(0.021)	(0.080)
Team Vancouver Canucks	0.00189	-0.01744	0.1264
	(0.015)	(0.021)	(0.080)
Team Vegas Golden Knights	-0.01469	-0.02785	0.01412
	(0.021)	(0.029)	(0.109)
Team Washington Capitals	-0.0362 **	-0.04175 *	-0.05236

	(0.014)	(0.019)	(0.072)
Team Winnipeg Jets	-0.01272	-0.02784	0.001656
	(0.014)	(0.019)	(0.072)
Year 2015	0.01628 **	0.02754 ***	-0.03503
	(0.005)	(0.007)	(0.028)
Year 2016	0.01455 *	0.03476 ***	-0.02201
	(0.006)	(0.008)	(0.032)
Year 2017	0.01176	0.03034 ***	0.02641
	(0.006)	(0.009)	(0.033)
Year 2018	0.01092	0.03181 ***	0.05664
	(0.007)	(0.009)	(0.035)
N	152	152	152
R2	0.98	0.97	0.48
SSR	0.0383591	0.0770658	1.1125118

Standard errors in parentheses

*** p< 0.001; ** p < 0. 01; * p < 0.05.

.

Table 12: Team Control F-Test Results

Model Analysed	SSR1	SSR2	m	n	k	F statistic	F Critical Value	Conclusion
Season Level Goals For Regression Version 1	0.30276	0.210826			37	1.29		Team Controls are not statistically significant
Season Level Goals For Regression Version 2	0.800749	0.456558		152	36	2.89	1.57	Team Controls are statistically significant
Season Level Goals Against Regression	9.89781	6.749412	30		35	1.804		Team Controls are statistically significant
Game Level Goals For Regression	2154.96	2149.98		12464	34	0.96	1.46	Team Controls are not statistically significant
Game Level Goals Against Regression	34043.54	33842.6			33	2.46		Team Controls are statistically significant

Table 13: Year Control F-Test Results

<table>
<tr><th>Model Analysed</th><th>SSR1</th><th>SSR2</th><th>m</th><th>n</th><th>k</th><th>F statistic</th><th>F Critical Value</th><th>Conclusion</th></tr>
<tr><td>Season Level Goals For Regression Version 1</td><td>0.210826</td><td>0.194467</td><td rowspan="5">4</td><td rowspan="3">152</td><td>41</td><td>2.31</td><td rowspan="3">2.46</td><td>Year controls are not statistically significant</td></tr>
<tr><td>Season Level Goals For Regression Version 2</td><td>0.456558</td><td>0.3921</td><td>40</td><td>4.32</td><td>Year controls are statistically significant</td></tr>
<tr><td>Season Level Goals Against Regression</td><td>6.749412</td><td>6.174038</td><td>39</td><td>2.61</td><td>Year controls are statistically significant</td></tr>
<tr><td>Game Level Goals For Regression</td><td>2149.98</td><td>214865</td><td rowspan="2">12464</td><td>38</td><td>1.92</td><td rowspan="2">2.37</td><td>Year controls are not statistically significant</td></tr>
<tr><td>Game Level Goals Against Regression</td><td>33842.6</td><td>33766.88</td><td>37</td><td>6.966</td><td>Year controls are statistically significant</td></tr>
</table>

Section 6: Player Application of the Regression Results

6.1 - Regression Adjustments

To account for irregular and insignificant regression results, the regression equations will be adjusted before applying them to the individual player data.

The season level regressions presented in Tables 5 and 6, where far fewer observations are available, will be slightly adjusted. From the *goals for* regressions shown in Table 5, regression equation (2.3) will be used in the continuation of the analysis instead of equation (1.3). As the results on *secondary assists* are counterintuitive, using equation (2.3) which includes *total assists* is more reasonable. The *faceoff percentage* variable will be removed in both the *goals for* and *goals against* equations as it was not statistically significant and its estimated effect is counterintuitive. In addition, in equation (3) for season *goals against* regression from Table 6*,* the *takeaways* variable will similarly be removed. *Takeaways*, like *faceoff percentage* is not statistically significant and is counterintuitive. These new regression results are formally presented in Table 14.

The full model regression equations from Table 7 and 8 (both presented in their respective column (3)) applied to the individual game data will remain as is. All of the variable coefficients were either statistically significant or were not counterintuitive, therefore will remain in the player application. These regression results are formally presented in Table 15.

As for the control variables, both the *Team* and *Year* variables will remain in all four equations. The controls are statistically significant for the majority of the models and will be kept to be consistent across the equations.

Table 14: Player Application Season Level Regressions

	Equation (1): Goals For	Equation (2): Goals Against
(Intercept)	0.1107	2.547 ***
	(0.127)	(0.444)
Ast.60	0.5406***	
	(0.015)	
Giveaways.60	-0.01335 *	0.01232
	(0.006)	(0.022)
Takeaways.60	0.02586 **	
	(0.009)	
HitsF.60	-0.002009	0.008498
	(0.002)	(0.008)
HitsT.60	0.003246	-0.02961 **
	(0.003)	(0.011)
Team Arizona Coyotes	-0.04578	0.3624 *
	(0.043)	(0.163)
Team Boston Bruins	-0.05004	0.1153
	(0.043)	(0.156)
Team Buffalo Sabres	-0.05085	0.3269
	(0.044)	(0.168)
Team Calgary Flames	-0.1692 **	0.07932
	(0.05014)	(0.167)
Team Carolina Hurricanes	-0.1670 **	0.3740 *
	(0.056)	(0.156)
Team Chicago Blackhawks	-0.09441	0.3699 *
	(0.049)	(0.185)
Team Colorado Avalanche	-0.1005 *	0.2690
	(0.049)	(0.175)
Team Columbus Blue Jackets	0.01281	0.1515
	(0.048)	(0.174)
Team Dallas Stars	-0.02436	0.1104
	(0.041)	(0.155)
Team Detroit Red Wings	-0.02973	0.2118
	(0.041)	(0.160)
Team Edmonton Oilers	-0.03459	0.4170 **
	(0.041)	(0.151)
Team Florida Panthers	-0.05137	0.2432
	(0.042)	(0.150)

Team Los Angeles Kings	0.02249	-0.02576
	(0.039)	(0.150)
Team Minnesota Wild	-0.03767	0.03460
	(0.053)	(0.197)
Team Montreal Canadiens	-0.03602	0.1904
	(0.041)	(0.161)
Team Nashville Predators	0.01252	-0.09868
	(0.044)	(0.162)
Team New Jersey Devils	-0.02968	0.2098
	(0.045)	(0.163)
Team New York Islanders	-0.005278	0.1957
	(0.041)	(0.149)
Team New York Rangers	-0.009768	0.1540
	(0.042)	(0.151)
Team Ottawa Senators	-0.01949	0.4458 **
	(0.041)	(0.148)
Team Philadelphia Flyers	-0.1026 **	0.2497
	(0.039)	(0.153)
Team Pittsburgh Penguins	-0.04102	0.3026
	(0.044)	(0.164)
Team San Jose Sharks	-0.1639 ***	0.2435
	(0.044)	(0.160)
Team St Louis Blues	-0.1142 *	-0.006975
	(0.050)	(0.1795)
Team Tampa Bay Lightning	-0.02152	0.1068
	(0.043)	(0.157)
Team Toronto Maple Leafs	-0.01070	0.4702 **
	(0.047)	(0.1557)
Team Vancouver Canucks	-0.03439	0.3110
	(0.045)	(0.169)
Team Vegas Golden Knights	-0.03604	0.2186
	(0.067)	(0.202)
Team Washington Capitals	-0.08284 *	-0.03163
	(0.042)	(0.149)
Team Winnipeg Jets	-0.04829	0.08675
	(0.042)	(0.153)
Year 2015	0.05582 **	-0.1065
	(0.017)	(0.064)
Year 2016	0.07241 ***	-0.06361

	(0.019)	(0.074)
Year 2017	0.06188 **	0.06264
	(0.020)	(0.078)
Year 2018	0.06640 **	0.1487
	(0.021)	(0.081)
N	152	152
R2	0.97	0.47

Standard errors in parentheses

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 15: Player Application Game level Regressions

	Equation (3): Goals For	Equation (4): Goals Against
(Intercept)	-2.142 ***	0.8327 ***
	(0.031)	(0.121)
SF.60	0.07644 ***	
	(0.001)	
SA.60		0.04497 ***
		(0.002)
Block.A.60	0.001064	
	(0.001)	
Block.F.60		0.029999 ***
		(0.003)
Missed.F.60	-0.002453 **	
	(0.001)	
Missed.A.60		-0.03648 ***
		(0.003)
SH.Perc	0.2743 ***	
	(0.001)	
Team Arizona Coyotes	-0.02916	0.2151
	(0.029)	(0.115)
Team Boston Bruins	-0.03885	-0.02421
	(0.029)	(0.115)
Team Buffalo Sabres	-0.03679	0.3067 **
	(0.029)	(0.115)
Team Calgary Flames	0.01804	0.1916
	(0.029)	(0.115)
Team Carolina Hurricanes	-0.02599	0.2673 *
	(0.029)	(0.115)
Team Chicago Blackhawks	-0.0001947	-0.07749
	(0.029)	(0.116)
Team Colorado Avalanche	0.009886	0.1815
	(0.029)	(0.115)
Team Columbus Blue Jackets	-0.01228	0.1253
	(0.029)	(0.115)
Team Dallas Stars	-0.01102	0.05641
	(0.029)	(0.115)
Team Detroit Red Wings	0.006111	0.1273

	(0.029)	(0.115)
Team Edmonton Oilers	-0.01260	0.3316 **
	(0.029)	(0.115)
Team Florida Panthers	-0.01903	0.1864
	(0.029)	(0.115)
Team Los Angeles Kings	0.02010	-0.07593
	(0.029)	(0.115)
Team Minnesota Wild	-0.008081	0.06523
	(0.029)	(0.115)
Team Montreal Canadiens	0.0008460	-0.1101
	(0.029)	(0.116)
Team Nashville Predators	0.01238	-0.1559
	(0.029)	(0.115)
Team New Jersey Devils	0.001231	0.1664
	(0.029)	(0.115)
Team New York Islanders	0.04059	0.1046
	(0.029)	(0.115)
Team New York Rangers	-0.03038	0.1065
	(0.029)	(0.115)
Team Ottawa Senators	0.02348	0.2639 *
	(0.029)	(0.115)
Team Philadelphia Flyers	0.02263	0.04576
	(0.029)	(0.115)
Team Pittsburgh Penguins	-0.01075	-0.01562
	(0.029)	(0.115)
Team San Jose Sharks	-0.01003	0.1465
	(0.029)	(0.115)
Team St Louis Blues	0.006028	-0.08315
	(0.029)	(0.115)
Team Tampa Bay Lightning	0.02116	-0.02114
	(0.029)	(0.115)
Team Toronto Maple Leafs	0.03717	0.1900
	(0.029)	(0.115)
Team Vancouver Canucks	-0.01554	0.2114
	(0.029)	(0.115)
Team Vegas Golden Knights	-0.00001588	0.03831
	(0.039)	(0.154)
Team Washington Capitals	-0.005680	-0.06065
	(0.029)	(0.115)
Team Winnipeg Jets	-0.007044	0.06417

	(0.029)	(0.115)
Year 2015	-0.008514	-0.06087
	(0.012)	(0.047)
Year 2016	0.003394	0.006740
	(0.012)	(0.047)
Year 2017	0.0009429	0.08019
	(0.012)	(0.047)
Year 2018	0.02277	0.1691 ***
	(0.012)	(0.047)
N	12464	12464
R2	0.94	0.06
SSR	2148.65	33766.88

Standard errors in parentheses

*** p < 0.001; ** p < 0.01; * p < 0.05

To maintain the focus of the results of this application to the player's contributions, the intercepts have been ignored when the equations are applied to the player statistics. The reasoning for excluding the intercept is because the application of the equations is simply for comparative analysis and not used for predictions. Because the same equations are applied to all players, excluding the intercept will not impact the marginal differences between players. The intercept was also ignored in EW's Weighted Points Model (2017) mentioned in the literature review.

6.2 - Contribution Score

The equations that will be applied to the individual players will then be:

(5) Season level Goals For contribution = (0.5406)*Assists – (0.01335)*Giveaways + (0.02586)*Takeaways – (0.002009)*Hits For + (0.003246)*Hits Taken + $Team_i$ + $Year_t$

(6) Game level Goals For contribution = (0.07644)*Shots For + (0.001064)*Blocked Shot Attempts – (0.002453)*Missed shots + (0.2743)*Shooting Percentage + $Team_i$ + $Year_t$

(7) Season level Goals Against contribution = (0.01232)*Giveaways + (0.008498)*Hits For - (0.02961)*Hits Taken + $Team_i$ + $Year_t$

(8) Game level Goals Against contribution = (0.04497)*iShots Against + (0.029999)*Blocked Shots – (0.03648)*iMissed Shots + $Team_i$ + $Year_t$.

The *Team* and *Year* variables are the set of controls for the 31 different teams, and the 5 separate years analysed. With the Anaheim Ducks as the team of reference, the *Team* variable will take the value of the corresponding coefficient that the observed player belongs to. This will signify how different teams are compared to Anaheim, other things being equal. With the reference year being 2014, the *Year* variable will take the value of the corresponding season being analysed. This will signify how different the scoring trends are compared to the 2014 season, other things being equal.

As explained in the Methodology section, equations (5) to (8) will be applied to a player's individual statistics. Equations (5) and (6) will capture the player's offensive contributions, and equations (7) and (8) will capture the player's defensive contributions. Two equations are needed to measure each player's offensive and defensive contributions because of limited data. All four equations consist of important variables needed to properly evaluate a team's success. Not all variables were available at the individual game level, therefore separate season level regressions were used to measure their importance. All of the regressions used had either *goals for per 60 minutes* or goals against *per 60 minutes* as the dependent variables. The use of the same ratios for all of the data has made it possible to include all the variables to make one aggregate measure of offensive contribution by combining the results of equation (5) and (6). The same is done with equations (7) and (8) to measure defensive contribution.

To account for the individual player statistics, *shots against*, *missed shots* and *blocked shots* will be adjusted as previously explained using Darcy Norman's method from *HockeyNomics (2009)*, by dividing the statistic by 6. These statistics are measured as on-ice events, meaning the count is the total amount that occurred when a player is on the ice. For example, an opponent takes a *shot on goal*, that shot is awarded to each player on the ice for team *i* as a *shot against*. To properly divide the blame to every player evenly, that shot will be divided by the total number of players on the ice for team *i* including the goalie, which is 6. Individual *blocked shots* is a statistic that is measured and is available in the data; however, for this application I will use *on-ice blocked shots* and dividing it by 6, instead of *individual blocked shots*. *On-ice blocked shots* are the amount of shots blocked by the observed player's team while he is on the ice. Using *on-ice blocked shots* will remain consistent with the other defensive statistics used, even though a player doesn't directly block a shot, his defensive play could still be partly responsible for the shot being blocked by a teammate.

An apparent exclusion in equations (5) and (6) is a player's individual goal production. This variable could not be accounted for in the regression analysis because GF.60 was used as the dependent variable, and therefore could not also be an independent variable. Although the actual count of a player's individual goal production is not in the equation, it is still implicitly accounted for through the inclusion of *shooting percentage* (SH.Perc) and *shots for* (SF.60). The SH.Perc can only increase when a player scores a goal and they can only do so by taking a shot. Therefore, by having both in the equation we are indirectly accounting for a player's goals. Similarly, a player's amount of *goals against* while on the ice is also not directly included in equations (7) and (8) simply because it is implicitly estimated by these equations as their contribution to goals against.

Table 16 illustrates the application of the above equations (5) to (8) to a player's statistics to obtain the player's *Contribution Score*. The player used in the example is the Edmonton Oilers' Connor McDavid from the 2018 season. As illustrated in the table, each individual statistic is multiplied by its respective coefficient and then summed together to obtain a result for each equation. The results of equations (5) and (6) correspond to the player's offensive contributions towards goal scoring, and equations (7) and (8) are the player's contribution towards *goals against* his team. To be able to easily compare players and to be able to calculate their estimated contracts, an aggregate contribution number for each player is required, which will be their *Contribution Score.* To obtain each player's *Contribution Score,* each player's *total* *offensive contribution* will be reduced by his *total contribution towards goals against*. To obtain the player's *total offensive contribution* the results from equation (5) and (6), their game level and season level offensive contributions are summed together. Equally, to obtain the player's *total contribution towards goals against* his team, the results of equations (7) and (8) are summed together. Finally, to obtain the player's *Contribution Score* we take the difference between the player's *total offensive contribution* and *total contribution towards goals against.* These simple manipulations are made possible because all the statistics are in the same ratio of *per 60 minutes* of play.

The *Contribution Score* equation is as follows:

$$\begin{aligned} Contribution\ Score &= Total\ Offensive\ Contribution - Total\ Goals\ Against\ Contribution \\ &= [equation\ (5) + equation\ (6)] - [equation\ (7) + equation\ (8)] \end{aligned}$$

Table 16: Contribution Score Calculation

Year: 2018	Player: Connor McDavid		Team: Edmonton Oilers		Position: Forward			
Equations	(5)		(6)		(7)		(8)	
	Ast.60	coefficient	SF.60	coefficient	Giveaways.60	coefficient	iSA.60	coefficient
	1.76	0.540601	6.91	0.07644	2.99	0.012322	5.658092687	0.044972
	Giveaways.60	coefficient	Block.A.60	coefficient	HitsF.60	coefficient	iBlock.60	coefficient
	2.99	-0.01335	2.9	0.001064	1.36	0.008498	2.208931127	0.029999
	Takeaways.60	coefficient	Missed.F.60	coefficient	HitsT.60	coefficient	iMissed.A.60	coefficient
	3.92	0.025862	2.56	-0.00245	3.92	-0.02961	1.959417312	-0.03648
	HitsF.60	coefficient	SH.Perc	coefficient	Team	coefficient	Team	coefficient
	1.36	-0.00201	15.29	0.2743	EDM	0.417048	EDM	0.331564
	HitsT.60	coefficient	Team	coefficient	Year	coefficient	Year	coefficient
	3.92	0.003246	EDM	-0.0126	2018	0.148709	2018	0.169106
	Team	coefficient	Year	coefficient				
	EDM	-0.03459	2018	0.02277				
	Year	coefficient						
	2018	0.66402						
Results	Season Level Goals For Contribution = 1.652		Game Level Goals For Contribution = 4.729		Season Level Goals Against Contribution = 0.498		Game Level Goals Against Contribution = 0.7499	
	Total Goals For Contribution = 6.381				Total Goals Against Contribution = 1.248			
	Contribution Score = 5.1336							

6.3 - Contract Estimation

Once each player’s *Contribution Score* has been calculated for each season independently, it is now possible to assign a monetary value to the *Contribution Score* for each player in each individual season. As mentioned in the Methodology section, the forwards and defensemen are separated for the remainder of the analysis. Also, as previously explained, to avoid small sample errors, only players with a total ice time over 60 minutes of play for the season are considered for this analysis.

A player is categorized as a replacement player when he is amongst the lowest tier in terms of *Contribution Score*. For this application, the replacement players will be identified as the players outside the regular lineup of 12 forwards and 6 defencemen per team for that season. The average *Contribution Score* for the replacement players was taken to obtain the *Replacement* Contribution Score. To obtain the monetary value of each *Contribution Score* point, the minimum wage for that season was divided by the *Replacement Contribution Score*.

Finally, all the players will have two separate measures for their financial worth. The first is their Estimated Value, and their second will be their Estimated Contract. A player’s Estimated Value will simply be his *Contribution Score* multiplied by the monetary value of each *Contribution Score* point. Because a negative *Contribution Score* is possible, this may result in a negative Estimated Value. In the real world, a player cannot be paid a negative salary. Therefore, in the case where any player’s Estimated Value is negative, or less than the league minimum wage, their Estimated Contract will be assigned the value of the league minimum wage. For any player where their Estimated Value is greater than the league minimum, their Estimated Contract will be assigned the same value as their Estimated Value. In the data set, players that have played for multiple teams within the same season are counted as multiple observations, one per team. To account for this, these players' *Contribution Scores* have been combined using weighted averages based on the amount of games played for each team, so that each player accounts for a single observation.

The results for the *Replacement Contribution Scores* and the dollar value of a *Contribution Score* point are displayed in Table 17. In Table 18, an example of the estimations is illustrated using Connor McDavid once again, as well as Brad Malone who has a negative *Contribution Score* for the 2018 season. The estimated results of the players' Estimated Values and Estimated Contracts per season are displayed in Table 19.

The results in Table 17 show fluctuations in the dollar value per *Contribution Score* point, in the last column, from year to year. A possible reasoning for this is the change in minimum wage and NHL expansion that occurred in 2017. Every two years the league minimum wage increases, which will affect the estimated dollar value for a score point from one year to the next. In 2017, the NHL expanded the league from 30 teams to 31, by introducing a team to Las Vegas. This expansion allowed players that would normally be replacement players to become regular players in the league to fill in the open positions. This caused the average replacement player level to drop, and thus increase the *Dollar Value per Contribution Score point*. In 2018, the league adjusted itself from the expansion and was able to filter out the weaker replacement players that had been relied upon in the expansion season.

As seen in the tables, the value of a contribution point is very different between defensemen and forwards. Looking at the *replacement Contribution Score* in Table 17, it appears that the replacement level defencemen contributes far less to team success than does a replacement level forward. This suggests that there is a larger marginal difference between defensemen than there is between forwards.

This result could be signifying that teams would be better off offering contracts to high level defensemen and settling with lower tier forwards due to the smaller marginal differences between forwards. A possible explanation for this result is the way forwards and defensemen are used differently. First, there is the simple difference in the number of players at each position. On a typical team there are 12 forwards and 6 defensemen, meaning the number of defensemen in the league is much smaller than the number of forwards. Taking the average replacement defensemen of such a smaller group could be skewed by the smaller sample. The second difference between forwards and defensemen is their style of play. A defenseman is not typically relied upon for offensive contribution the same way forwards are, especially a replacement defenseman filling in a spot in the lineup. With the regression results being much more offensively oriented and the fact that replacement defensemen would not be relied on as much in a game, these might be the reason replacement defensemen do not contribute as much as replacement forwards.

Table 17: Replacement Player Value

Season	Position	Number of Active Players	Number of Regular Players	Number of Replacement Players	Replacement Contribution Score	Minimum Wage	Dollar Value per Score point
2016-17	Defence	262	180	82	0.287216231	$ 575,000.00	$ 2,001,975.99
	Forwards	505	350	155	1.387620198	$ 575,000.00	$ 414,378.52
2017-18	Defence	278	186	92	0.152712743	$ 650,000.00	$ 4,256,357.32
	Forwards	513	372	141	1.309907084	$ 650,000.00	$ 496,218.40
2018-19	Defence	282	186	96	0.565464725	$ 650,000.00	$ 1,149,496.99
	Forwards	517	372	145	1.689219042	$ 650,000.00	$ 384,793.20

Table 18: Estimated Contract Example

Year	Player	Contribution Score	Dollar Value per Score point	Estimated Value	Estimated contract
2018	Brad Malone	-0.238526428	$ 384,793.20	$ -91,783.35	$ 650,000.00
2018	Connor McDavid	5.133568104	$ 384,793.20	$ 1,975,362.10	$ 1,975,362.10

Table 19: 2019 Estimated Values and Estimated Contracts

			Mean	Standard Error	Standard Deviation	Minimum	Maximum	Count
2018-19	Forwards	*Contribution Score*	3.41032	0.0635	1.4432	-0.2385	9.6521	
		Monetary Value ($)	1.312 M	24 K	555 K	-92 K	3.714 M	517
		Estimated Contract ($)	1,348 M	22 K	495 K	650 K	3.714 M	
	Defensemen	*Contribution Score*	1.5268	0.0546	0.9176	-0.5181	5.4461	
		Monetary Value ($)	1.755 M	62 K	1.055 M	-596 K	6.260 M	282
		Estimated Contract ($)	1.814 M	58 K	971 K	650 K	6.260 M	
2017-18	Forwards	*Contribution Score*	2.8977	0.0598	1.3545	-0.6812	9.4181	
		Monetary Value ($)	1.438 M	30 K	672 K	-338 K	4.673 M	513
		Estimated Contract ($)	1.475 M	27 K	608 K	650 K	4.673 M	
	Defensemen	*Contribution Score*	1.1630	0.0636	1.0598	-0.6539	9.0975	
		Monetary Value ($)	4.950 M	271 K	4.511 M	-2.783 M	38.722 M	278
		Estimated Contract ($)	5.164 M	254 K	4.241 M	650 K	38.722 M	
2016-17	Forwards	*Contribution Score*	3.0097	0.0640	1.4391	-0.5373	8.2565	
		Monetary Value ($)	1.247 M	27 K	596 K	-223 K	3.421 M	505
		Estimated Contract ($)	1.288 M	24 K	528 K	575 K	3.421 M	
	Defensemen	*Contribution Score*	1.3040	0.0635	1.0286	-0.5198	8.7792	
		Monetary Value ($)	2.611 M	127 K	2,059	-1.041 M	17.576 M	262
		Estimated Contract ($)	2.699 M	120 K	1,948	575 K	17.576 M	

Note: monetary results are expressed in thousands of dollars (K) and millions of dollars (M)

6.4 - Identifying Discrepancies

The goal of this paper is to objectively evaluate player contributions and to identify current inefficiencies in player salaries; i.e., differences between the estimated values of their contributions, and what they are actually paid, are henceforth referred to as inefficiencies. Using the estimated values for the 2016,2017 and 2018 seasons, an estimated contract for 2019 will be calculated using weighted averages and compared to the current active contracts as of the time of writing (July 18, 2019). The method used to assign the weights for each season will follow the same method used by Rob Vollman in the Hockey Abstract (2014). This method is simply assigning each year a weight that is double the weight used for the previous year. The weights used were 4 for 2018, 2 for 2017 and 1 for 2016, then divide the total by 7.

The estimated 2019 player contracts were calculated for all players that were active from the 2016-17 season to the 2018-19 season. An example using Connor McDavid is illustrated in Table 20 and the overall results are displayed in Table 21. Some players with gap years were still given an estimated contract; however, their results could be skewed due to missing data. For the contract comparison portion of the analysis, only players with data for all three seasons will be included in the analysis.

To identify inefficiencies, each player's 2019 Estimated Value and their 2019 Estimated Contracts are compared with the currently active contracts in the league. Using the online database from spotrac.com (n.d), a list of currently active contracts has been obtained. Cross-referencing this list with the players from the data set to properly match each player to their respective contract, the difference was taken between the player's actual contract and the calculated Estimated Value and Estimated Contract. Some players were unable to be cross-referenced and therefore were dropped for this portion of the analysis. The results of the 2019 Estimated Contracts are shown in Table 21.

Table 20: McDavid 2019 Estimation Example

Player: Connor McDavid			
	Value/ Contract*		Calculation
2018 Estimation	$	1,975,362.10	4* 2018 Estimation
2017 Estimation	$	940,965.66	2*2017 Estimation
2016 Estimation	$	1,654,767.03	2016 Estimation
2019 Estimation	$	1,634,020.96	summation /7
Actual 2019 Contract	$	12,500,000.00	
Difference	$	10,865,979.04	Actual contract - Estimated 2019
* Connor McDavid has an Estimated Value over the minimum wage; therefore, his Estimated Value and Estimated Contract are all equal			

In Table 21, the Estimated 2019 Player Value is the monetary estimation of a player's contract using the payers' Estimated Value of past seasons. The Estimated 2019 Contract is the monetary estimation of a player's contract using each player's past Estimated Contracts. The Actual Contract is the player's current annual average salary. The Difference in Value and the Difference in Contract are the measured differences between a player's actual contract and their Estimated 2019 Value and Estimated 2019 Contracts, respectively. A positive difference means the actual contract is larger than the estimated amount signifying the player is overpaid based on this study's methodology. A negative difference means the actual contract is lower than the estimated value, signifying the player is underpaid based on this study's methodology.

The results for forwards in Table 21 show that, while using this method to estimate player contracts, forwards are on average being overpaid by approximately $2.4 million per year. The average actual contract for a Forward is just under $3.9 million, but according to this method should be closer to $1.4 million. The highest

Estimated Contract is approximately $2.4 million, when the highest actual contract is $12.5 Million. The largest underpayment is only by $1.4 million, but the largest overpayment is over $10 million.

The results for defensemen in the lower panel of Table 21 show that the estimated contracts on average are approximately $3 million, quite close to the $3.7 million average figure of the actual contracts. The largest actual contract for a defenseman is $11.5 million, but the method in this study estimates that it should be just over $8.5 million. The largest underpayment for defensemen is estimated to be over $7.8 million, and the largest overpayment is just under $9 million. On average, it is estimated that defensemen are overpaid by less than one million dollars.

From the results in Table 21, we can assume that the largest discrepancies in player contracts lie in the evaluation and quantification of contributions from forwards. It appears the current method to monetarily evaluate contracts overcompensates the marginal differences between forwards. Overall, these findings suggest that players are being overpaid for their level of production when compared to replacement level players.

Tables 22 and 23 illustrate correlation matrices between the player's estimated contracts and their actual contracts. As displayed in Tables 22 and 23, there is a surprisingly weak correlation between the players' actual contracts and their estimated salaries. Figures 3 and 4 are scatter plot graphs displaying the same information. This result displays evidence that the methodology presented in this paper is evidently a different approach than that currently used by teams. A possible reasoning for the large differences in results is that perhaps the methodology presented does not account for factors that occur outside of the playing surface that affect player salaries. Such factors as fan preferences could be a driving factor for the differences found. Fans are the people who buy the teams' tickets and merchandise. The fans are what truly drive a team's revenue. If fans prefer to see flashy players over efficient players, perhaps that is one reason why teams overpay so much for forwards based on this paper's findings. Perhaps the methodology presented here does not account for fan utility in a team's playing style. The findings are based on a player's contribution in on-ice play; outside factors may be a possible reason for the large salary differences found, especially for forwards.

Appendices 3 to 6 display the top 10 players with the biggest discrepancy between their actual contracts and estimated contracts using the *Contribution Score* method for evaluation.

Table 21: Estimated 2019 Player Value and Contract

Position		Mean	Standard Error	Standard Deviation	Minimum	Maximum	Count
Forwards	*Estimated 2019 Player Value*	$ 1,412,026.40	$ 18,605.24	$ 340,023.10	$ 448,528.84	$ 2,429,500.65	334
	Estimated 2019 Contract	$ 1,430,591.77	$ 17,699.55	$ 323,471.01	$ 769,346.50	$ 2,429,500.65	
	Actual Contract	$ 3,855,078.22	$ 141,533.79	$ 2,586,624.33	$ 625,000.00	$ 12,500,000.00	
	Difference in Value	$ 2,443,051.82	$ 137,467.87	$ 2,512,316.98	$ -1,408,995.54	$ 10,865,979.04	
	Difference in Contract	$ 2,424,486.46	$ 137,608.03	$ 2,514,878.60	$ -1,408,995.54	$ 10,865,979.04	
Defensemen	*Estimated 2019 Player Value*	$ 2,930,152.38	$ 97,514.52	$ 1,311,923.74	$ -759,697.49	$ 8,662,868.33	181
	Estimated 2019 Contract	$ 3,009,315.47	$ 91,399.23	$ 1,229,650.84	$ 639,285.71	$ 8,662,868.33	
	Actual Contract	$ 3,737,413.59	$ 162,434.98	$ 2,185,339.17	$ 547,500.00	$ 11,500,000.00	
	Difference in Value	$ 807,261.20	$ 174,491.83	$ 2,347,547.46	$ -7,837,868.33	$ 8,982,850.77	
	Difference in Contract	$ 728,098.12	$ 173,204.13	$ 2,330,223.26	$ -7,837,868.33	$ 8,982,850.77	

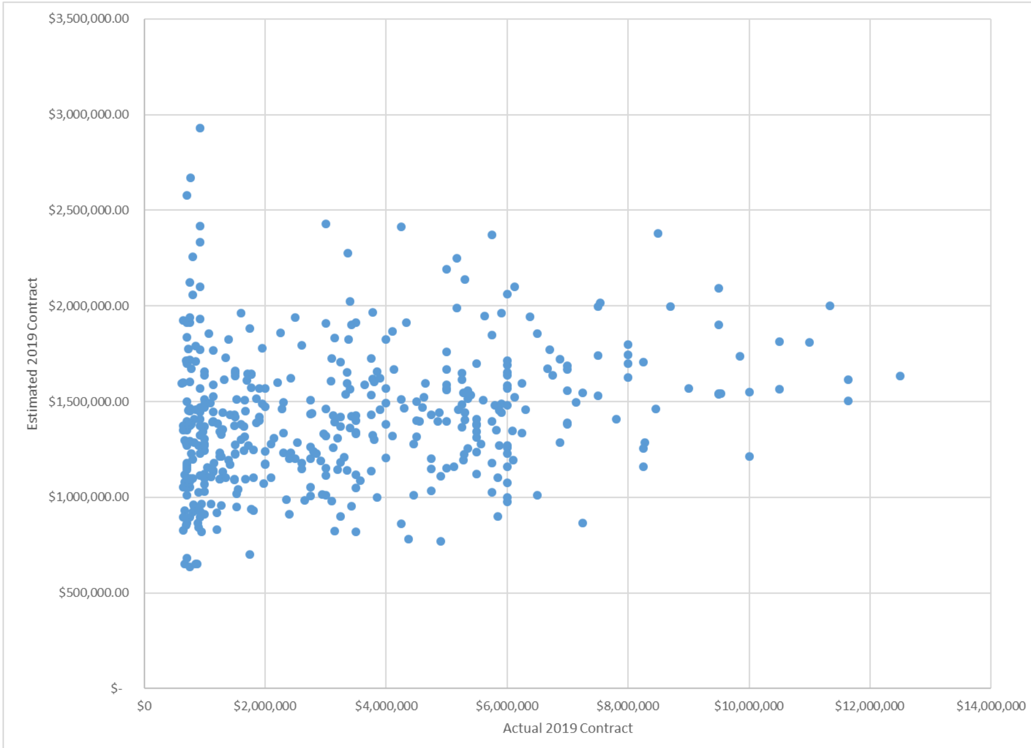
Table 22: Forward Correlation Matrix

	Estimated 2019 Value	*Estimated 2019 Contract*	*Actual 2019 Contract*
Estimated 2019 Value	1		
Estimated 2019 Contract	0.9755	1	
Actual 2019 Contract	0.2399	0.2366	1

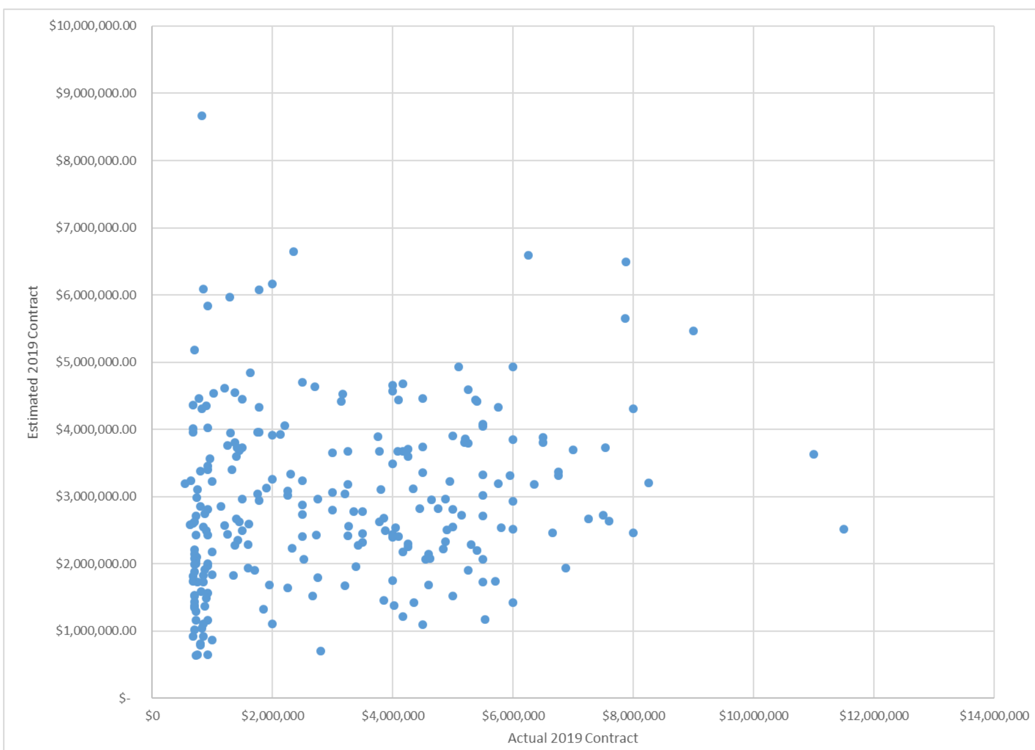
Table 23: Defensemen Correlation Matrix

	Estimated 2019 Value	*Estimated 2019 Contract*	*Actual 2019 Contract*
Estimated 2019 Value	1		
Estimated 2019 Contract	0.989905666	1	
Actual 2019 Contract	0.189463863	0.189405708	1

<u>Figure 3 : Contract Correlation For Forwards</u>



<u>Figure 4: Contract Correlation for Defensemen</u>



Section 7: Possible Extensions and Alterations

7.1 - Extension

This type of project presents many different avenues for further experimentation. While remaining with the concept of using linear weights to measure player performance, methods other than OLS regression analysis could be used. A possible extension to the project would be to use Monte Carlo simulations to estimate the linear weights, or even perhaps certain Machine Learning techniques could be useful. Advanced predictive analytics could also be used to attempt to estimate player salaries.

Other possibilities would be to use regression analysis in the more classical Labour Economics approach, to estimate player's productivity. By classical method I am referring to regressing a player's salary on the player's individual statistics. This extension could then be compared with the findings of this paper to find the biggest differences in how teams value their players.

7.2 - Possible Improvements

Although this project was successful in achieving its goal of estimating player contracts based off of the measured importance of certain statistics for team success, improvements could still be made. One improvement would be to obtain all of the measured statistics at the individual game level as opposed to using some at the season aggregate level. This adjustment would likely improve the results as all the statistics would be measured within the same regressions, as opposed to needing several different regressions, and would allow the use of a greater sample size. If all the variables were included in the same regression as opposed to separated throughout multiple regressions, the coefficients would likely be more reliable and the errors would potentially be smaller as all the major factors would be accounted for. This could result in more variables being statistically significant, which would lead to more robust findings. Having all the variables in a single regression would also lead to consistently using the same larger sample size for the analysis.

The findings in this research fell victim to the same issues found in past literature in sports analysis. The findings of this paper were more offensively focused and lacked defensive measurement. The original belief was that, by running separate regressions for *goals against*, the results would show a better way of measuring defensive ability. Unfortunately, the regression results measuring the impact on *goals against* have very low R^2 values. Possibly the use of different statistics would lead to improvements on these results, but for the time being, it seems that being limited to using box score statistics limits the ability to properly measure defensive ability.

Section 8: Conclusion

This project had two main objectives. The first was to identify the importance of different statistics and how they impact team success. The second was to apply this measure of importance to the statistics of individual players, to ultimately measure how much they helped their team and how much they should be financially compensated for it. The results of the second objectives were then used to identify current inefficiencies in how teams value player contributions.

The findings from the regression analysis showed that the statistics available have strong explanatory power for measuring a team's offensive abilities, but very little at measuring defensive abilities. Surprisingly, some statistics such as *faceoff percentage* and physical play had much smaller impacts on team success than previously anticipated. These results still provided enough information to obtain useful estimations for a player's contributions towards team success. Once applied to individual player statistics, these results offer a different perspective on how to evaluate player performance. The results showed that, compared to their respective replacement level player, defensemen have a larger impact on team success than do forwards. The largest estimated fair salary for forwards was just under $2.5 million, where the largest fair salary for a defenseman was over $8.5 million. This suggest that teams should invest into stable defensemen as opposed to relying on replacement level defensemen. Additionally, based on the estimated player values, the largest contract discrepancies were found to be for forwards, with an average overpayment of approximately $2.5 million a year. The average contract discrepancy for defensemen was only approximately $800,000. These estimates suggest that the current method used by teams to evaluate players overestimates the marginal contributions made by forwards.

This study provided a hypothetical estimation of player salaries based off the measured importance of statistics to a team's success using linear regression analysis. We were able to identify inefficient contracts from these results and ultimately conclude that NHL teams may not be maximizing their contract allocations. Although not a perfect estimation, these findings could suggest an alternative approach on how to efficiently build a hockey team to optimize team spending.

References

Adnan. (2011, October 03). Introduction to Advanced Hockey Statistics: Corsi. Retrieved from https://www.silversevensens.com/2011/10/3/2461198/introduction-to-advanced-hockey-statistics-corsi-ottawa-senators

Bell, J. (2019, July 01). Free Agent Busts: The Worst Free Agent Signings in NHL History. Retrieved from https://thehockeywriters.com/the-worst-free-agent-signings-in-nhl-history/

Burtch, S. (2012, July 23). Intro To Advanced Hockey Statistics - Corsi. Retrieved from https://www.pensionplanpuppets.com/2012/7/23/3173579/what-is-corsi-how-do-you-use-corsi-who-is-corsi-don-cherry-hates-corsi

Calculating Point Shares. (n.d.). Retrieved from https://www.hockey-reference.com/about/point_shares.html

Cane, M. (2014, December 10). Score Adjusted Weighted Shots. Retrieved from https://puckplusplus.com/2014/12/10/score-adjusted-weighted-shots/

Cane, M. (2015, February 24). From the Ottawa Hockey Analytics Conference: Looking at Weighted Shots. Retrieved from http://www.hockeyprospectus.com/from-the-ottawa-hockey-analytics-conference-looking-at-weighted-shots/

Duroux, A. (2017, December 31). Demystifying Hockey Analytics, Part 1: Intro and The Basics. Retrieved from https://bsndenver.com/demystifying-hockey-analytics-part-1-intro-and-the-basics/

Duroux, A. (2017, December 31). Demystifying Hockey Analytics, Part 2: Stats Sites. Retrieved from https://bsndenver.com/demystifying-hockey-analytics-part-2-stats-sites/

Duroux, A. (2017, December 31). Demystifying Analytics, Part 3: Stats Frontier. Retrieved from https://bsndenver.com/demystifying-hockey-analytics-part-3-stats-frontier/

Eric T. (2013, December 12). Bringing Shot Quality into the Mix. Retrieved from https://www.sbnation.com/nhl/2013/12/12/5114366/nhl-stats-shot-quality-player-evaluation

EvolvingWild. (2017, August 01). Introducing Weighted Points Above Replacement – Part 1. Retrieved from https://hockey-graphs.com/2017/08/01/introducing-weighted-points-above-replacement-part-1/

EvolvingWild. (2017, August 04). Introducing Weighted Points Above Replacement – Part 2. Retrieved from https://hockey-graphs.com/2017/08/02/introducing-weighted-points-above-replacement-part-2/

Garret9. (2013, October 23). The Importance and Misconceptions of Advanced Hockey Analytics. Retrieved from https://www.arcticicehockey.com/2013/10/23/4862840/the-importance-and-misconceptions-of-advance-hockey-analytics

Greene, W. H. (2017). *Econometric Analysis* (8th ed.). New York, NY: Pearson.

Hawerchuk. (2010, November 25). Retro NHL and Anger at Corsi. Retrieved from https://www.arcticicehockey.com/2010/11/25/1835371/retro-nhl-and-anger-at-corsi

Hohl, G. (2017, December 01). Behind the Numbers: What Makes a Stat Good. Retrieved from https://hockey-graphs.com/2017/12/01/behind-the-numbers-what-makes-a-stat-good/

Hollinger, J. (n.d.). Calculating PER. Retrieved from https://www.basketball-reference.com/about/per.html

How to Calculate Wins Produced. (n.d.). Retrieved from http://wagesofwins.com/how-to-calculate-wins-produced/

JLikens. (2011, February 16). Shots, Fenwick and Corsi. Retrieved from http://objectivenhl.blogspot.com/2011/02/shots-fenwick-and-corsi.html

Madeupcall. (2016, March 29). 1-2-3 Let's Have Fun: Getting Started With Hockey Analytics. Retrieved from http://www.corsica.hockey/blog/2016/03/29/1-2-3-lets-have-fun-getting-started-with-hockey-analytics/

Mbmccurdy. (2014, November 13). Adjusted Possession Measures. Retrieved from https://hockey-graphs.com/2014/11/13/adjusted-possession-measures/

Myers, D. (n.d.). About Box Plus/Minus (BPM). Retrieved from https://www.basketball-reference.com/about/bpm.html

Natural Stat Trick. (n.d.). Retrieved from http://www.naturalstattrick.com/

NBA Win Shares. (n.d.). Retrieved from https://www.basketball-reference.com/about/ws.html

NHL Contracts. (n.d.). Retrieved from https://www.spotrac.com/nhl/contracts/

Norman, D. (2009). *HockeyNomics: What the Stats Really Reveal*. Montréal: OverTime Books.

Novet, A. (2019, June 06). How to Get Started in Hockey Analytics. Retrieved from https://hockey-graphs.com/2018/11/27/how-to-get-started-in-hockey-analytics/

Omgitsdomi. (2016, July 13). Measuring Single Game Productivity: An Introduction To Game Score. Retrieved from https://hockey-graphs.com/2016/07/13/measuring-single-game-productivity-an-introduction-to-game-score/

Paine, N. (2016, June 17). LeBron James Is Still In His Prime. Retrieved from https://fivethirtyeight.com/features/lebron-james-is-still-in-his-prime/

Pothier, J. (2017, September 02). Advanced Stats For VGK Dummies: Goals Created. Retrieved from http://sinbin.vegas/advanced-stats-for-vgk-dummies-goals-created/

Puck Report. (2009, July 3). Rethinking NHL Player Salary Structure. Retrieved from http://www.puckreport.com/2009/07/rethinking-nhl-player-salary-structure.html

Puck Report. (2017, June 22). NHL Minimum Wage & Maximum Wage By Year. Retrieved from http://www.puckreport.com/2009/07/nhl-minimum-wage-maximum-wage-by-year.html

Ryder, A. (2010, September 28). Goals Created. Retrieved from http://hockeyanalytics.com/2006/10/goals-created/

Ryder, A. (2013, December 15). Expected Goals. Retrieved from http://hockeyanalytics.com/2012/03/expected-goals/

Ryder, A. (2015, January 08). The Ten Laws of Hockey Analytics. Retrieved from http://hockeyanalytics.com/2008/01/the-ten-laws-of-hockey-analytics/

Staples, D. (2015, August 18). Why it's problematic to use on-ice stats like Corsi to rate individual NHL players. Retrieved from https://edmontonjournal.com/sports/hockey/nhl/cult-of-hockey/why-its-problematic-to-use-on-ice-stats-to-rate-individual-nhl-players

Tango, T. (2014, November 30). Introducing Weighted Shots Differential (aka Tango). Retrieved from http://tangotiger.com/index.php/site/article/introducing-weighted-shots-differential-aka-tango

Vollman, R., Awad, T., & Fyffe, I. (2014). *Rob Vollman's Hockey Abstract*. Place of publication not identified: Createspace.

Vollman, R., Awad, T., & Fyffe, I. (2016). *Stat Shot: The Ultimate Guide to Hockey Analytics*. Toronto, ON: ECW Press.

Wendorf, B. (2014, October 08). THN Analytics: An Introduction. Retrieved from https://thehockeynews.com/news/article/thn-analytics-an-introduction

Wilson, K. (2011, July 23). The Theory and Nature of Current Advanced Hockey Analysis. Retrieved from https://flamesnation.ca/2011/07/23/the-theory-and-nature-of-contemporary-hockey-analysis/

Winston, W. L. (2012). *Mathletics: How Gamblers, Managers, and Sports Enthusiasts use Mathematics in Baseball, Basketball, and Football*. New Jersey: Princeton University Press.

Appendix

Appendix 1: Team Game Data Summary Statistics

Variables	Definition	Mean	Standard Error	Standard Deviation	Minimum	Maximum
TOI	Time on Ice	48.252	0.037	4.169	32.617	60.000
SH%	Shooting Percentage (GF/Shots For)	7.968	0.053	5.948	0.000	40.000
SF	Shots on Net For	23.942	0.053	5.970	5.000	56.000
SF/60	Shots on Net For Per 60 minutes of play	29.793	0.063	7.085	6.560	65.630
SA	Shots on Net Against	23.942	0.053	5.970	5.000	56.000
SA/60	Shots on Net Against Per 60 minutes of play	29.793	0.063	7.085	6.560	65.630
GF	Goals Scored For	1.855	0.012	1.362	0.000	10.000
GF/60	Goals Scored For Per 60 minutes of play	2.310	0.015	1.693	0.000	12.890
GA	Goals Scored Against	1.855	0.012	1.362	0.000	10.000
GA/60	Goals Scored Against Per 60 minutes of play	2.310	0.015	1.693	0.000	12.890
Block.F	Opposing Shots Blocked For	11.494	0.040	4.421	0.000	37.000
Block.F/60	Opposing Shots Blocked For Per 60 minutes of play	14.287	0.048	5.348	0.000	46.790
Block.A	Shots Blocked by Opposition	11.494	0.040	4.421	0.000	37.000
Block.A/60	Shots Blocked by Opposition Per 60 Minutes of play	14.287	0.048	5.348	0.000	46.790
Saves.A	Shots Saved by Opposing Goalie	22.086	0.052	5.860	4.000	53.000

Saves.A/60	Shots Saved by Opposing Goalie Per 60 Minutes of play	27.483	0.063	6.993	5.250	62.110
Saves.F	Opposing Shots Saved by Goalie	22.086	0.052	5.860	4.000	53.000
Saves.F/60	Opposing Shots Saved by Goalie Per 60 Minutes of play	27.483	0.063	6.993	5.250	62.110
Missed.F	Shots that Missed the net	9.288	0.033	3.703	0.000	28.000
Missed.F/60	Shots that Missed the net Per 60 Minutes of play	11.551	0.040	4.510	0.000	33.820
Missed.A	Opposing Shots that Missed the net	9.288	0.033	3.703	0.000	28.000
Missed.A/60	Opposing Shots that Missed the net Per 60 Minutes of play	11.551	0.040	4.510	0.000	33.820
Observations	12464					

Appendix 2: Team Season Data Summary Statistics

Variables	Definition	Mean	Standard Error	Standard Deviation	Minimum	Maximum
TOI	Time on Ice	3956.630	5.123	63.162	3772.900	4086.850
TOI/GP	Average Time on Ice per Game	48.252	0.062	0.770	46.011	49.840
A1	Primary Assists	142.645	1.533	18.895	98.000	193.000
A1.60	Primary Assists per 60 minutes of play	2.163	0.023	0.282	1.484	2.964
A2	Secondary Assists	112.730	1.313	16.190	74.000	159.000
A2.60	Secondary Assists per 60 minutes of play	1.709	0.020	0.243	1.146	2.411
Ast	Total number of Assists	255.375	2.804	34.574	175	350
Ast.60	Total number of Assists per 60 minutes of play	3.872	0.042	0.517	2.710	5.374
Giveaways	Giveaways	571.921	10.926	134.708	303.000	970.000
Giveaways.60	Giveaways per 60 minutes of play	8.663	0.162	1.992	4.616	14.665
Takeaways	Takeaways	471.803	7.919	97.629	252.000	763.000
Takeaways.60	Takeaways per 60 minutes of play	7.149	0.117	1.441	3.803	11.401
HitsF	Hits for	1726.803	22.285	274.753	1096.000	2517.000
HitsF.60	Hits for per 60 minutes of play	26.199	0.343	4.228	16.181	37.720

HitsT	Hits Taken	1726.855	19.675	242.566	1179.000	2315.000
HitsT.60	Hits Taken per 60 minutes of play	26.190	0.299	3.680	17.963	34.518
Face.Perc	Faceoff Percentage	49.989	0.161	1.981	44.006	54.207
GF	Goals Scored For	152.138	1.597	19.690	106.000	206.000
GF.60	Goals Scored For Per 60 minutes of play	2.306	0.024	0.293	1.600	3.160
GA	Goals Scored Against	152.138	1.561	19.248	112.000	219.000
GA.60	Goals Scored Against Per 60 minutes of play	2.306	0.023	0.279	1.720	3.240
Observations	152					

Appendix 3: Estimated Top 10 Overpaid Forwards

Player	POS	TEAM	Actual Annual Contract	Estimated 2019 Value	Value Difference	Estimated 2019 Contract	Contract Difference
Connor McDavid	C	EDM	$12,500,000.00	$ 1,634,020.96	$ 10,865,979.04	$ 1,634,020.96	$ 10,865,979.04
Auston Matthews	C	TOR	$11,634,000.00	$ 1,492,394.24	$ 10,141,605.76	$ 1,502,081.56	$ 10,131,918.44
Artemi Panarin	LW	NYR	$11,642,857.00	$ 1,591,705.11	$ 10,051,151.89	$ 1,614,349.96	$ 10,028,507.04
Leon Draisaitl	C	EDM	$11,333,333.00	$ 1,999,729.17	$ 9,333,603.83	$ 1,999,729.17	$ 9,333,603.83
John Tavares	C	TOR	$11,000,000.00	$ 1,810,533.31	$ 9,189,466.69	$ 1,810,533.31	$ 9,189,466.69
Jonathan Toews	C	CHI	$10,500,000.00	$ 1,564,652.55	$ 8,935,347.45	$ 1,564,652.55	$ 8,935,347.45
Jack Eichel	C	BUF	$10,000,000.00	$ 1,213,608.12	$ 8,786,391.88	$ 1,213,608.12	$ 8,786,391.88
Patrick Kane	RW	CHI	$10,500,000.00	$ 1,814,599.00	$ 8,685,401.00	$ 1,814,599.00	$ 8,685,401.00
Anze Kopitar	C	LAK	$10,000,000.00	$ 1,517,105.05	$ 8,482,894.95	$ 1,548,085.50	$ 8,451,914.50
Tyler Seguin	C	DAL	$ 9,850,000.00	$ 1,738,154.22	$ 8,111,845.78	$ 1,738,154.22	$ 8,111,845.78

Appendix 4: Estimated Top 10 Underpaid Forwards

Player	POS	TEAM	Actual Annual Contract	Estimated 2019 Value	Value Difference	Estimated 2019 Contract	Contract Difference
Ivan Barbashev	C	STL	$925,000.00	$ 2,333,995.54	$ -1,408,995.54	$ 2,333,995.54	$ -1,408,995.54
Nic Dowd	C	WAS	$750,000.00	$ 2,124,122.63	$ -1,374,122.63	$ 2,124,122.63	$ -1,374,122.63
Tyler Ennis	C	OTT	$800,000.00	$ 2,058,460.52	$ -1,258,460.52	$ 2,058,460.52	$ -1,258,460.52
Greg McKegg	C	NYR	$750,000.00	$ 1,913,502.31	$ -1,163,502.31	$ 1,913,502.31	$ -1,163,502.31
Alan Quine	C	CGY	$735,000.00	$ 1,556,377.49	$ -821,377.49	$ 1,776,011.80	$ -1,041,011.80
Brayden Point	C	TBL	$692,500.00	$ 1,712,892.64	$ -1,020,392.64	$ 1,712,892.64	$ -1,020,392.64
Oskar Sundqvist	C	STL	$700,000.00	$ 1,629,758.98	$ -929,758.98	$ 1,700,168.21	$ -1,000,168.21
William Carrier	LW	VGK	$725,000.00	$ 1,714,677.97	$ -989,677.97	$ 1,714,677.97	$ -989,677.97
Colton Sissons	C	NSH	$625,000.00	$ 1,595,523.14	$ -970,523.14	$ 1,595,523.14	$ -970,523.14
Pontus Aberg	LW	MIN	$650,000.00	$ 1,598,367.86	$ -948,367.86	$ 1,598,367.86	$ -948,367.86

Appendix 5: Estimated Top 10 Overpaid Defensemen

Player	POS	TEAM	Actual Annual Contract	Estimated 2019 Value	Value Difference	Estimated 2019 Contract	Contract Difference
Erik Karlsson	D	SJS	$11,500,000	$ 2,517,149.23	$8,982,850.77	$ 2,517,149.23	$8,982,850.77
Drew Doughty	D	LAK	$11,000,000	$ 3,625,563.31	$7,374,436.69	$ 3,625,563.31	$7,374,436.69
Brent Burns	D	SJS	$8,000,000	$ 2,391,136.28	$5,608,863.72	$ 2,464,514.40	$5,535,485.60
Oliver Ekman-Larsson	D	AZ	$8,250,000	$ 3,204,754.37	$5,045,245.63	$ 3,204,754.37	$5,045,245.63
Dustin Byfuglien	D	WPG	$7,600,000	$ 2,639,206.76	$4,960,793.24	$ 2,639,206.76	$4,960,793.24
Brent Seabrook	D	CHI	$6,875,000	$ 1,863,130.80	$5,011,869.20	$ 1,936,245.19	$4,938,754.81
Aaron Ekblad	D	FLA	$7,500,000	$ 2,015,746.61	$5,484,253.39	$ 2,727,492.20	$4,772,507.80
Kris Letang	D	PIT	$7,250,000	$ 2,666,934.13	$4,583,065.87	$ 2,666,934.13	$4,583,065.87
Alexander Edler	D	VAN	$6,000,000	$ 1,073,153.55	$4,926,846.45	$ 1,419,027.95	$4,580,972.05
Duncan Keith	D	CHI	$5,538,462	$ 1,176,207.06	$4,362,254.94	$ 1,178,315.08	$4,360,146.92

Appendix 6: Estimated Top 10 Underpaid Defensemen

Player	POS	TEAM	Actual Annual Contract	Estimated 2019 Value	Value Difference	Estimated 2019 Contract	Contract Difference
Paul LaDue	D	LAK	$ 825,000.00	$ 8,662,868.33	$ -7,837,868.33	$ 8,662,868.33	$ -7,837,868.33
Slater Koekkoek	D	CHI	$ 925,000.00	$ 5,839,735.79	$ -4,914,735.79	$ 5,839,735.79	$ -4,914,735.79
Zach Werenski	D	CBJ	$1,775,000.00	$ 6,073,130.11	$ -4,298,130.11	$ 6,073,130.11	$ -4,298,130.11
Zdeno Chara	D	BOS	$2,000,000.00	$ 6,165,246.03	$ -4,165,246.03	$ 6,165,246.03	$ -4,165,246.03
Yannick Weber	D	NSH	$ 675,000.00	$ 4,363,076.39	$ -3,688,076.39	$ 4,363,076.39	$ -3,688,076.39
Matt Irwin	D	NSH	$ 675,000.00	$ 3,959,623.34	$ -3,284,623.34	$ 3,959,623.34	$ -3,284,623.34
Scott Harrington	D	CBJ	$1,633,333.00	$ 4,843,854.73	$ -3,210,521.73	$ 4,843,854.73	$ -3,210,521.73
Kevin Connauton	D	COL	$1,375,000.00	$ 4,397,080.78	$ -3,022,080.78	$ 4,553,650.77	$ -3,178,650.77
Luca Sbisa	D	NYI	$1,500,000.00	$ 4,451,195.41	$ -2,951,195.41	$ 4,451,195.41	$ -2,951,195.41
Alex Goligoski	D	AZ	$ 547,500.00	$ 2,783,776.80	$ -2,236,276.80	$ 3,190,428.50	$ -2,642,928.50