ECON 490 Community Engaged Learning Project: Predicting Winger Salary Given Performance in the National Hockey League

Sutton Schiefner

36621143

The University of British Columbia

Abstract

Using player performance and salary data provided to me from CKM Sports Management, I estimate a multiple linear regression specification that utilizes statistics I believe to capture performance factors such as puck possession, passing, and aggressiveness in order to predict salary for the position of wingers, excluding rookie contracts for the initial estimation. I then utilize that function to predict the salaries for *all* wingers given those performance factors in order to see the degree of over and under payment at the contract, team, ice-time levels. I find that most rookies are likely to be over-payed than under, while standard and 35+ contract types are relatively even in terms of the number of over and under payments. Most teams, on average, overpay their players more so than underpay them, and further, when looking at ice-time I find that those on the lower end of the spectrum are predicted to be more over-payed than those with more ice-time and those at the top are, on average, under-payed more so than over-payed.

Introduction

Toronto Maple Leaf's center Auston Matthews was drafted first overall in 2016 at the age of nineteen years old. After playing three seasons under a rookie contract with a standard salary cap of $925,000, this February he signed a 5-year, $58.17 million dollar contract extension, with an annual cap of $11.63 million; over eleven times his previous salary. With such a drastic increase in salary, one would be justified in wondering if there was a similarly drastic change in his skillset over the first three years of his playing career. In reality, his total game points per season has actually gone down since he first started in the league. So what are the possible factors that lead to this stark increase in his salary cap?

This leads us to the question of salary determination. Sports statisticians and fans alike have been particularly interested in the concept of salary determination for some time now. The introduction of advanced statistical methods, has led to much research regarding this field in the other three professional sports leagues, but minimal research has been done in terms of salary determination in the National Hockey League (NHL).

Previous research has centered around using goals and total games played as main determinants of skill for forward positions, as well as identifying exogenous factors such as team and league effects such as revenue, location, and draft position among others (Levine, 2011) (Shirreffs & Sommers, 2006) (Idson & Kahane, 2000) (Vincent & Eastman, 2009) (Jones & Walsh, 1988) (Cebula, 2009). While these studies do return admirable results, the bulk of their data is focused in subject areas that do not inform us much on the actual skill of the player, and what attributes make for an effective winger. I stray away from focusing on *outcome-oriented* variables such as goals, and instead I look into metrics that relate to shot creation, puck possession, as well as a certain degrees of aggressiveness. Doing so, I can capture the effects of game points, creating a more descriptive definition of player skill.

I estimate a salary function using these metrics for the positions of left and right wingers for the 2017/2018 NHL season (excluding rookies), with the ultimate goal of using this function to predict a player's salary given their performance in the specified regions. Eventually, comparing wingers predicted worth salary-wise with their actual monetary value. I then analyze this prediction relating to rookies. Given their strict salary-caps they are most likely being grossly under or over-payed, and this function gives us a quantitative measure of that gap. This sort of knowledge could be of great use to player-agents seeking to attain the best renegotiations for their players, as well as team G.M.'s looking to allocate their salary budget effectively.

Thanks to a comprehensive dataset provided by CKM Sports Management in Vancouver B.C., I was given access to 166 different player statistics at the individual and team level across three lines of play: regular five-versus-five (5V5), powerplay (PP), and penalty-kill (PK). Additionally I was given access to a complete list of player salaries, including their contract type, and free agency among less relevant measures for this research.

I utilize a multiple linear regression model, focusing my research on the thirteen variables that I found to have the greatest effect on salary when regressed individually, and in varying combinations. I also include controls for player age, contract type, free-agency, team, as well as whether the player partakes in PP or PK lines, in order to capture other exogenous effects mentioned in the previous research but not included in the model.

My results provide some context to which measures are of most import for player skill. With the final regression including all controls, I was able to capture 80.8% of the variability in salary, with 65.4% of that variation due to the performance variables chosen. On the 5V5 line, increases in total shots taken ($icf60_{5v5i}$) and giveaways ($igva60_{5v5i}$), and decreases in hit against ($iha60_{5v5i}$) lead to increases in salary. While on the PP, decreases in total shots ($icf60_{PPi}$) and takeaways ($icf60_{PPi}$), and increases in giveaways ($igva60_{PPi}$) and hits against ($iha60_{PPi}$) lead to increases in player salary. Finally, on the PK, increases in unblocked shots ($iff60_{PKi}$), giveaways ($igva60_{PKi}$), and hits against ($iha60_{PKi}$), all lead to increases in salary. The time on ice variables (toi_{5v5i}, toi_{PPi}, toi_{PKi}) are positive for all three lines, but time on ice for the PP. However, for reasons I will discuss in greater detail later, I need to be wary of my interpretation of these variables.

Further, in applying this completed model to *all* wingers I find that very few players are actually getting payed their worth, with large over and under-payments across all contract types, teams, and ice-time levels. With entry-level contracts having the highest average overpayment, as well as the largest gap between over and under-payments, while the gap between the over/under of standard and 35+ contracts stayed relatively narrow. When the predicted salaries were analyzed at the team level, I find that 25 of the 31 teams in the NHL fall heavier on the side of overpayment than underpayment, with only 3 teams that have a gap that is less than six-figures. Finally, when dividing all players up based on ice time I find a large difference in the average predicted salaries with those on the lower end of ice-time having a large average overpayment, and those on the higher end having a small average underpayment. I believe this is indication that there is much room for improvement in terms of salary *determination* as well as salary *allocation*.

The rest of the paper is organized as such: Section A discusses the background. Section B presents the data and the construction of my sample. Section C presents the empirical model and the methodology of the research. Section D presents the results, as well as a discussion of the findings. Section E discusses my various checks of robustness. Finally, I make some concluding remarks and discuss limitations in section F.

A. Background

Previous research has focused on salary arbitration hearings and their possible effects *on* player performance (Levine, 2011) (Shirreffs & Sommers, 2006), as well as how player or team productivity, league effects, or discrimination effects can alter player salary (Idson & Kahane, 2000) (Vincent & Eastman, 2009) (Jones & Walsh, 1988) (Cebula, 2009). While these studies do return admirable results, the bulk of their metrics regarding player performance are limited to *outcome-oriented* variables. By outcome oriented I mean variables that focus on the results of players skill as opposed to what they do to achieve those results. By looking at metrics that relate to shot creation, puck possession, as well as a degrees of aggressiveness, we can capture the effects of game points, while looking at how these players are able to achieve those points. Crafting a more explanatory definition of player skill. I estimate a salary function using the aforementioned performance based metrics for the positions of left and right wingers. I seek to use this function to estimate player salary given those metrics, and evaluate how much all types of players are being over or under-payed, most interestingly how much rookies, given their strict salary caps are being payed. This information could be of great value to team owners and general managers looking to assess the worth of their players, and allocate player budgets efficiently and effectively. It could also be of great use to player agents looking to ensure their clients are receiving the salaries they are entitled to.

B. Data Source & Sample Construction:

The datasets came courtesy of CKM Sports Management in Vancouver B.C.. I was provided with player performance data for regular five-versus-five play, powerplay, and penalty-kill lines, as well as NHL salary data for all players. The 5V5 data contains 166 different performance statistics for n=801 players for the 2017/2018 NHL season, in addition to their team and position. Given the fact that not all players are played during powerplays and penalty kills, the amount of players in the PP and PK datasets are much smaller (n=353, and n=357 respectively), however, they observe the same 166 performance statistics. Using the 5V5 set as the master, the three performance statistics datasets were merged using player name, team, position, and season as the identifying variables. Since team, position and season are consistent for each player across the three sets, only the 5V5 versions of those variables have been kept as they encompass the highest amount of players. The final dataset contained salary data for n=804 NHL players during the 2017/2018 season; also including player age, contract type, and agency type. Merging with the salary data was only done by name, and any players that were not consistent among the two sets have been dropped. I have also dropped any salary outliers that may be causing an upward bias on our prediction. The cut off I chose was any salary above 7 million, as those were clearest outliers. Upon removal there is a more consistent upward trend, among the salary data. *Figure 1.* depicts the salary distribution before and after the removal of the outliers. If a player has played on multiple teams over the course of a season, they were dropped from the regression. If a player plays multiple positions, the first position listed was the one taken. The season variable was also dropped as all the results are from the same 2017/2018 season. Since rookies are on contracts that greatly restrict their salary cap, I have removed them from the regression until the point in which we will predict their salary given our final function,

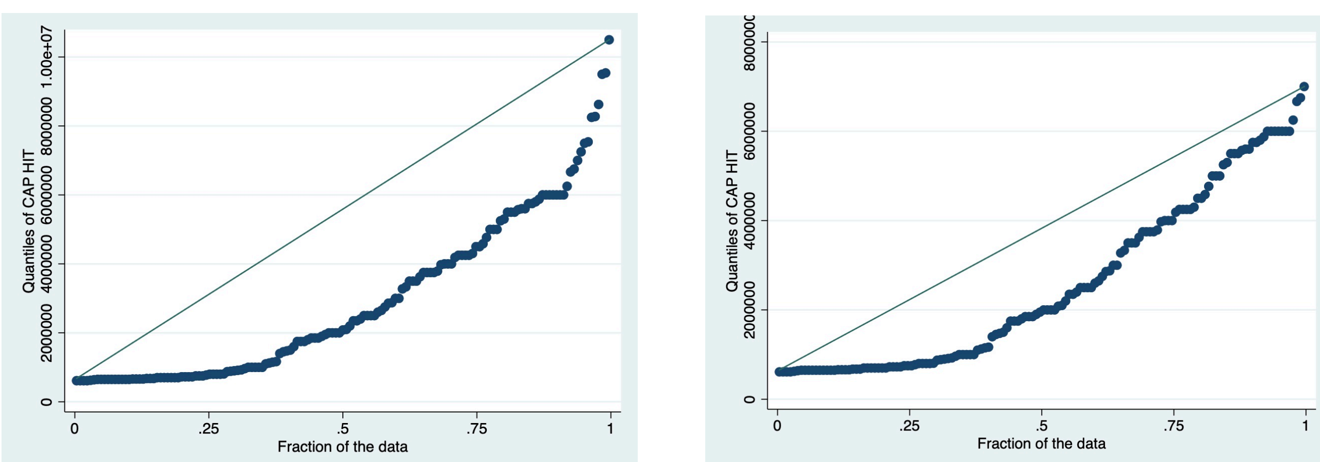

Figure 1. Salary Distribution Before Outlier Removal (Left) and After (Right)

to avoid a downward bias on the prediction. This leaves me with a sample size of n=144 different wingers for my regression. Lastly, since there some players who are not played on power plays or penalty kills, and thus do not have statistics for those variables, I generated two dummy variables: noPP and noPK, with a value of zero if the player has no data for those lines, and a value of 1 if they do. More on this and its importance in the robustness section.

C. Empirical Model and Methodology:

Given the predictive nature of the research, I will be utilizing a multiple linear regression specification for determining earnings with a selected set of the 166 variables provided to me. Based on the line of play (powerplay, penalty kill, or five-versus-five) the variables may vary in determining what skills are of higher emphasis for a winger on that line. The primary role of wingers in hockey depends on whether they are playing in defensive, or offensive zone positions. The variables chosen are all taken at the individual player level, and additionally I have chosen to use the variables at a per-sixty minutes basis to keep everything at a consistent and comparable metric.

i. Specification Construction

To determine which of these variables *individually* caused the most variation in salary ($caphit_i$), I utilized a bivariate regression analysis for each of the 33 variables I thought to be of the most importance to wingers, and regressed them against caphit. These variables are, for the 5V5, PP and PK levels: corsi-for ($icf60_i$), fenwick-for ($iff60_i$), shots-for ($isf60_i$), penalties taken ($ipent60_i$), penalties drawn ($ipend60_i$), giveaways ($igva60_i$), takeaways ($itka60_i$), hits for ($ihf60_i$), hits against ($iha60_i$), blocks ($iblk60_i$), and time on ice (toi_i). Since similar variables most likely capture similar effects, I only considered the variable with the highest R^2 value within a similar group (the shooting variables for example) for my initial regression specification. PP statistics have the highest effects on salary when looking at their individual R^2 levels. This is most likely due to the fact that better players are going to be played on powerplay lines more often than worse players. This effect on salary should be captured in the control variables for the dummies: noPK and noPP. After trying various regressions with differing combinations of these variables I narrowed the list down to only the thirteen with the highest prediction threshold, given a varying number of attempted combinations. To estimate the impact of player performance on salary I specify the following equation:

(1)

$$\begin{aligned} caphit_i = {} & \beta_0 + \beta_1 icf60_{5v5i} + \beta_2 igva60_{5v5i} + \beta_3 iha60_{5v5i} + \beta_4 toi_{5v5i} \\ & + \beta_5 icf60_{PPi} + \beta_6 igva60_{PPi} + \beta_7 itka60_{PPi} + \beta_8 iha60_{PPi} + \beta_9 toi_{PPi} \\ & + \beta_{10} iff60_{PKi} + \beta_{11} igva60_{PKi} + \beta_{12} iha60_{PKi} + \beta_{13}\, toi_{PKi} + \beta_{14} noPP \\ & + \beta_{15} noPK + \beta_{16} age_i + \boldsymbol{\eta_i} + \boldsymbol{\delta_i} + \boldsymbol{\theta_i} + \varepsilon_i \end{aligned}$$

Where $caphit_i$ represents player salary at the individual level. The $icf60_i$ variables represent the total shots taken by an individual player per 60 minutes of game time, while the $iff60_i$ variable represents the total unblocked shots by an individual player per 60 minutes. These variables indicate the level of shot creation each player contributes on their respective lines of play, and allows us to get a general idea as to their aggressiveness towards the net. The $igva60_i$ variables represent the total giveaways by an individual player per 60 minutes, meaning the amount of times the player had the puck, then lost it to the opposing team, and the $itka60_i$

variable represents the opposite; the amount of the times that player took the puck away from the opposing team. I will have more on what these variables represent for the overall equation in the results section later on. The $iha60_i$variables represent the total amount of hits against the player, a possible metric for puck possession or proximity, but more on this later as well. The toi_i variables represent the total amount ice time for each respective line of play. $noPP$ and $noPK$ take the value of 1 if the player has statistics for powerplay or penalty kill lines respectively, and zero if they do not. The $age60_i$ variable is a control for player age, to estimate any effects on salary based on the age of the player. $\boldsymbol{\eta_i}$ is a set of dummies controlling for the agency of the individual player, meaning if they are a restricted or unrestricted free agent. $\boldsymbol{\delta_i}$ denotes a set of dummies relating to the type of contract a player has, in this case whether the player is playing on a standard contract or a 35+ contract. A 35+ contract is a special case in which if a player who is 35 years or older signs a multiyear contract, their cap hit is deducted from the total team cap hit. It was designed to prevent the demotion of veteran players to minor leagues, and could count as a further metric for individual player legacy. $\boldsymbol{\theta_i}$ is a set of dummies controlling for the team each individual player is on, taking the value of one for whichever team they are signed to, and a zero for all others. This controls for any team effects that may have been omitted in my analysis. ε_i is the error term.

D. Results and Discussion

Part I: Salary Regression

i. Baseline Regression:

The table shown in *Figure 2.* represents my results for the varying stages of the regression. The initial baseline regression is represented in column (1), without any of the controls for age, team, free-agency, or type, only the necessary controls for penalty kill and powerplay lines are included. We receive a R^2 value of 0.6540, significant at the 99% confidence level. We receive some interesting results in this first regression. Many of the coefficients do not go in the expected direction based on intuitive hockey knowledge. Corsi-for variables predict drops in salary for an increase in shot attempts, which does not seem intuitive given the context of the sport, the effects of these shot attempts could perhaps be captured in another variable such as giveaways (since a missed shot may count as a giveaway) however, I am inclined to believe there is some unknown effect that is captured in the controls that may be causing a downward bias on these coefficients. The giveaways variables for 5V5 and powerplay lines have strong significant positive coefficients, which again, seems unintuitive given the context of hockey. However, giveaways could be capturing the effects of a player with a high amount of passing or shooting, where a missed pass, shot, or interception could lead to that play being counted as a giveaway. Given that interpretation, positive coefficients on the giveaway variables are rather quite intuitive as a proxy for puck aggressiveness. The negative coefficient for the penalty-kill line makes sense given the context, as a loss of the puck in that scenario greatly increases the opponents chance of scoring, much more so than that powerplay and 5V5. Hits against on 5V5

		Regression Output			
	(1)	(2)	(3)	(4)	(5)
	reg1	reg2	reg3	reg4	reg5
VARIABLES	CAP HIT	CAP HIT	CAP HIT	CAP HIT	CAP HIT
CAP HIT					
5V5					
iCF/60	-13,082	-1,818	4,749	13,463	52,974
	(40,140)	(38,538)	(37,574)	(36,212)	(40,271)
	0.745	0.962	0.900	0.711	0.192
iGVA/60	316,354**	297,112**	253,716*	305,147**	472,956***
	(152,734)	(146,247)	(143,136)	(138,437)	(169,170)
	0.0403	0.0443	0.0787	0.0293	0.00627
iHA/60	-116,607*	-51,685	-61,494	-52,268	-60,468
	(66,263)	(65,953)	(64,275)	(61,847)	(68,080)
	0.0808	0.435	0.341	0.400	0.377
TOI	159.5	370.8	629.8	892.5*	749.3
	(568.8)	(547.5)	(540.5)	(525.4)	(567.7)
	0.780	0.500	0.246	0.0919	0.190
PP					
iCF/60	-71,381***	-61,073***	-62,110***	-49,610**	-53,636**
	(23,362)	(22,540)	(21,938)	(21,412)	(21,895)
	0.00274	0.00767	0.00540	0.0221	0.0161
iGVA/60	240,503*	266,303**	278,786**	253,272**	187,581
	(123,744)	(118,628)	(115,527)	(111,311)	(125,475)
	0.0541	0.0265	0.0173	0.0246	0.138
iTKA/60	-560,653**	-507,794**	-476,609**	-439,223**	-349,328
	(243,061)	(233,050)	(227,059)	(218,549)	(233,533)
	0.0227	0.0312	0.0378	0.0466	0.138
iHA/60	89,405	92,719	68,545	60,437	77,864
	(109,488)	(104,771)	(102,312)	(98,380)	(104,688)
	0.416	0.378	0.504	0.540	0.459
TOI	18,058***	17,414***	15,392***	13,233***	13,163***
	(2,809)	(2,694)	(2,716)	(2,688)	(2,810)
	2.32e-09	1.98e-09	9.35e-08	2.64e-06	9.36e-06
PK					
iFF/60	104,651	118,148	104,255	99,083	60,712
	(74,987)	(71,853)	(70,094)	(67,398)	(74,916)
	0.165	0.103	0.139	0.144	0.420
iGVA/60	-93,343	-128,629	-72,984	-47,118	3,762
	(238,600)	(228,525)	(223,248)	(214,741)	(239,037)
	0.696	0.575	0.744	0.827	0.987
iHA/60	59,733	81,468	25,875	41,116	259,258
	(174,663)	(167,243)	(163,921)	(157,639)	(167,516)
	0.733	0.627	0.875	0.795	0.125
TOI	6,059	5,734	4,100	1,735	3,852
	(4,685)	(4,484)	(4,401)	(4,289)	(5,043)
	0.198	0.203	0.353	0.687	0.447
noPP	-209,112	-76,694	-274,467	-315,054	-345,449
	(667,216)	(639,520)	(626,220)	(602,092)	(619,543)
	0.754	0.905	0.662	0.602	0.578
noPK	957,116	963,136	731,806	626,144	883,824
	(637,207)	(609,736)	(598,907)	(576,570)	(620,944)
	0.136	0.117	0.224	0.280	0.158
AGE		90,510***	52,944*	119,196***	81,350**
		(25,303)	(27,938)	(33,283)	(37,609)
		0.000493	0.0604	0.000488	0.0330
(Dummy Controls)					
Expiry	No	No	Yes	Yes	Yes
Type	No	No	No	Yes	Yes
Team	No	No	No	No	Yes
Constant	988,140	-2.343e+06*	-1.511e+06	-5.203e+06***	-5.451e+06***
	(933,757)	(1.291e+06)	(1.290e+06)	(1.654e+06)	(1.941e+06)
	0.292	0.0718	0.244	0.00208	0.00604
Observations	144	144	144	144	144
R-squared	0.654	0.686	0.705	0.729	0.808

Figure 2. Regression Stages 1 Through 5

return a loss in salary for increases in hits against, indicating a possible loss of the puck in the 5V5 context. For the powerplay and penalty-kill lines, however, hits against returns a positive coefficient most likely indicating puck possession in these contexts, which is outweighing the negative effects of losing the puck; especially in a PK context where the opposing team has greater chances of handling the puck, puck possession and by proxy hits against could also be an indicator of skill. Similarly, the negative coefficient on takeaways for powerplays could just as likely indicate the opposite. If a given player is taking away the puck from the other team many times, that indicates that the opposing team has puck possession on the other teams powerplay, a possible negative determinant of skill. Time on ice returns positive coefficients, intuitively, with PP time on ice of a much higher value at an $18,058.32 increase in salary per minute of powerplay time compared to $159.51 and $6059.42 for 5V5 and PK respectively. However it is important to note, that there could be some reverse causality in terms of time on ice. Players could easily be receiving more ice time *because* they are of a higher skill set, instead of the reverse.

In the further stages of this regression I will incrementally be adding the rest of the controls for age [column (2)], free agency [column (3)], contract type [column (4)], and team [column (5)], each contributing to incremental increases in the R^2 value, while remaining significant at the 99% confidence level.

ii. Stage II: Age

Adding the control for age [column (2)] increases the amount of variance explained by the regression to an R^2 squared of 0.6857. The direction of the coefficients for each variable do not change, but the magnitudes slightly do, but no enough to alter my previous interpretation of the questionable directions of the previous variables. The 5V5 corsi-for variable seems to be moving in the positive direction, perhaps indicating that some of the omitted variables that could have been causing a negative direction are now being accounted for. The control for age has significant effects on salary, with an increase in one year in age leading to an increase in salary of around $90,000. Signifying that perhaps the length of a players career has some effect on their salary regardless of skill. It could also indicate effects of a players skills progressing over their career leading to higher salary.

iii. Stage III: Free agency

Adding the dummy for the categorical variable of free-agency in the third stage [column (3)] also further increases the explanatory value of this regression with an R^2 of 0.7047. Once again, the coefficient directions don't change with the exception of 5V5 corsi-for, which is now in the intuitive direction indicating that a one unit increase of shot attempts in 5V5 play lead to an increase in salary of around $4,800 per year. Again, the effects which were causing this negative direction for corsi-for are unknown, but whatever effect that was causing the negative bias has been captured by this new control. The variable for restricted free agency has been omitted by Stata, but for unrestricted free agents it seems they have an higher salary by around $766,000 per year. A rather large effect on the salary of a player. Since unrestricted free agency requires a certain level of experience in the league, it is safe to assume that this is also a potential indicator for career length in the league, and in turn potentially skill.

iv. Stage IV: Contract type

In the next stage I add another dummy for a categorical variable: contract type. The R^2 increases once again to a value of 0.7293. Since we are exclusively basing this model off of non-rookie contracts, this is solely for standard and 35+ contracts. The remaining variables' coefficients have not changed in direction, and remain roughly the same value. Stata omitted the variable for 35+ contracts, leaving us with an interpretation for the coefficient for standard contracts. Having a standard contract as a player, on average, increases their yearly by roughly $1.9 million. Once again, standard contracts could capture the effect of gained experience in the league exiting a rookie contract.

v. Stage V: Team

In the final stage of this regression I add a set of dummies for the categorical variables relating to each team in the NHL. Based on the previous research, we know that there are various team effects that can greatly increase a player's salary. These effects include but are not limited to: team revenue, team legacy, location, state or province laws, coaching staff, teammates and so on. By controlling for the team each player is on, we can to an extent, assume they capture the variation in salary caused by these effects. Adding these dummies to the regression greatly increases the salary variability explained to an R^2 of 0.8082. Once again, there are no changes in the directions of the coefficients, but the coefficient for corsi-for in 5V5 play is much more intelligible when compared to its' initial value. Now a one unit increase of total shots taken in 60 minutes leads to an increase in salary of roughly $52,000. However, the corsi-for for powerplays still seems unintuitive, but could be capturing the effects of potential giveaways on the powerplay, while the giveaways variable may be capturing puck aggressiveness which is considered valuable in powerplay. All teams have positive coefficients with the exception of the Detroit Red Wings. This is somewhat intuitive, as the Red Wings placed 5th last in the league for the 2017/2018 season, however, given that logic I would expect the last four teams to also return negative coefficients and yet they do not. There could be some other effects of playing in Detroit that are causing this downward pressure on salary that may require some further research. The top three teams in terms of positive salary effects are the Boston Bruins (4th in the league overall) The Tampa Bay Lightning (3rd in the league overall) and the New York Islanders (22nd in the league overall). The first two are intelligible given their legacy and place in the league, however the large positive effect of the islanders is rather puzzling given their place in the league, however, effects relating to the city of New York could be assumed to have large positive outcomes on a team's value and ability to pay higher salaries.

Overall, the variables chosen appear to account for around 65% of the variation in salary in this model, with the controls accounting for a further 16%, leaving 19% left unexplained. Of the controls, it seems that effects on skill, and in turn salary caused by mere exposure and experience in the league account for 8% of the variation, and team effects account for a further 8% in the variability. While there are some puzzling coefficients initially, the effects I explained could account for the seemingly counter-intuitive directions of the coefficients. However, there could be other omitted effects causing this downward pressure on the coefficients that I am unaware of.

Part II: Predicted Salary vs. Actual

Using the aforementioned regression specification (1) I will now be estimating the salary of all players including rookies, to see given their performance and this model, what they hypothetically *should* be getting paid.

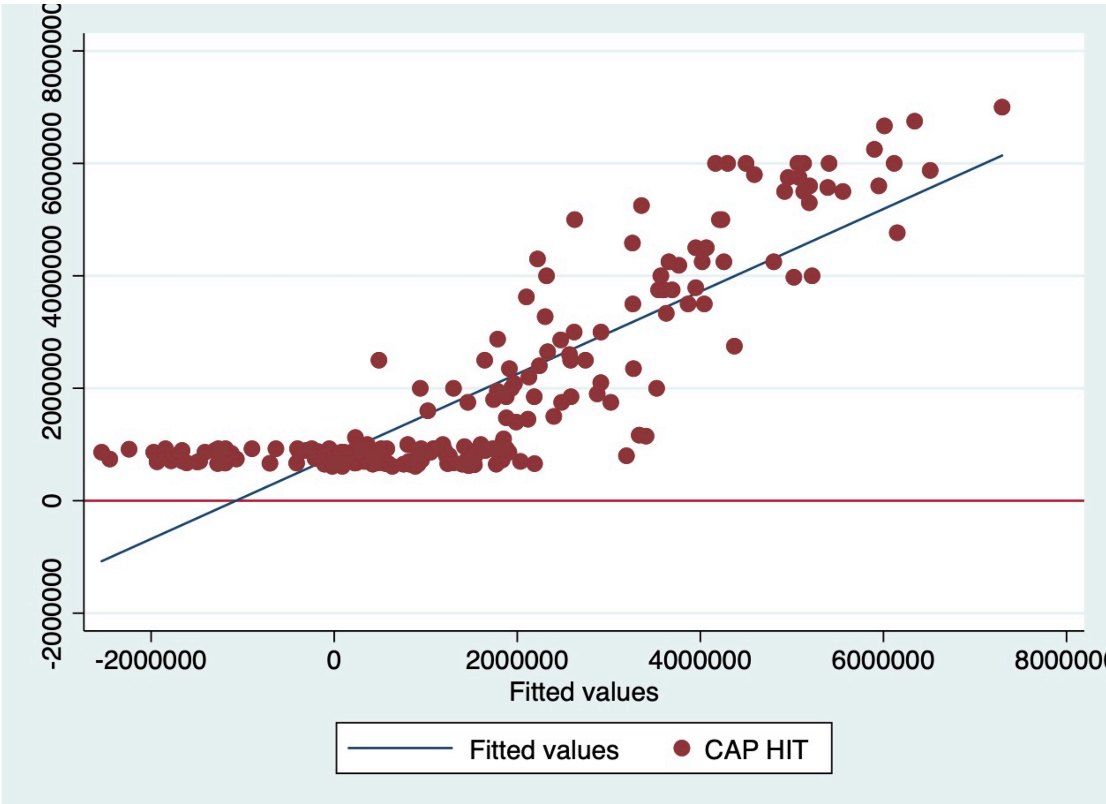

Figure 3. Actual Salary Values vs. Predicted Values

From *Figure 3.* you can see that rookies (the horizontal trend towards the bottom of the graph) clearly would've biased the initial specification downwards as they all have similar set salary caps that do not vary by a significant amount based on performance. I then took the initial caphit values, and subtracted the predicted values form them, generating a new difference variable "diff". A negative value for diff indicates that a player is being under-payed when compared to that of their predicted value, and a positive value indicates that a player is being over-payed. From my initial interpretation, the model predicts negative *salary* values for some players, but mostly for rookies. This could be due to a lack of the necessary positively biased statistics as some of these players are quite new, combined with poor experience effects and possible poor team effects. Regardless, it seems a large amount of rookies according to this specification are overplayed but I will touch on this further later.

i. Contract

For rookies: 18 are under-payed and 47 are predicted to be over-payed with an average underpayment of around $480,000 and an average overpayment of around $1.7 million. It is important to remember, for the overpayment, my model predicated negative salaries for a large amount of the rookies so according to this model they are not even worth a baseline salary, which could be troubling.

For 35+ contracts, 4 were predicted under-payed, and 3 over-payed with an average underpayment of around $656,000 and an average overpayment of around $875,000

For standard contracts, 65 were predicted to be under-payed and 72 were predicted to be over-payed, with an average underpayment of around $682,000 and an average overpayment of around $615,000. For the standard contract, there were some predicted negative values but not as many as with the rookies.

i. Team

As seen in the table depicted in *Figure 4.* below, I have calculated the average amount each team is predicted to be over or under-paying their players. 25 of the 31 teams are overpaying more on average than they are underpaying, indicating much potential for more efficient salary allocation according to this model. Of the 31 teams, the 3 teams that had the narrowest gap between their average over payments and average underpayments are the Minnesota Wild, L.A. Kings, and the Vegas Golden Knights. Meaning they, on average, overpay some players roughly the same amount that they underpay others. These teams are the only ones

TEAM	AVERAGE OVER-PAYMENT	# OF OBVS (OVER)	AVERAGE UNDER-PAYMENT	# OF OBVS (UNDER)	PLUS/MINUS
ANA	$ 934,005.60	3	$ (160,232.00)	1	$ 773,773.60
ARI	$ 786,397.86	6	$ (172,780.38)	1	$ 613,617.49
BOS	$ 1,095,230.58	4	$ (750,441.00)	3	$ 344,789.58
BUF	$ 1,168,829.61	4	$ (926,569.21)	3	$ 242,260.40
CAR	$ 1,036,672.83	3	$ (240,222.66)	4	$ 796,450.18
CBJ	$ 408,261.88	4	$ (299,594.03)	4	$ 108,667.84
CGY	$ 1,269,353.02	4	$ (574,708.69)	5	$ 694,644.33
CHI	$ 976,285.47	2	$ (425,529.00)	2	$ 550,756.47
COL	$ 934,580.23	6	$ (730,970.79)	3	$ 203,609.44
DAL	$ 807,216.59	4	$ (1,334,996.38)	1	**$ (527,779.78)**
DET	$ 1,606,243.73	4	$ (425,118.08)	3	$ 1,181,125.65
EDM	$ 1,037,803.20	6	$ (613,558.38)	2	$ 424,244.83
FLA	$ 2,376,333.00	4	$ (670,470.08)	3	$ 1,705,862.92
LA	$ 612,023.75	3	$ (635,425.50)	1	**$ (23,401.75)**
MIN	$ 1,552,212.45	4	$ (1,525,612.38)	2	$ 26,600.08
MTL	$ 1,008,254.71	9	$ (870,378.48)	3	$ 137,876.23
NJ	$ 784,937.16	6	$ (1,271,920.72)	2	**$ (486,983.56)**
NSH	$ 1,066,870.78	2	$ (794,669.94)	4	$ 272,200.84
NYI	$ 895,735.75	2	$ (393,788.38)	3	$ 501,947.38
NYR	$ 29,196.25	1	$ (294,024.17)	3	**$ (264,827.92)**
OTT	$ 1,048,879.06	2	$ (237,637.63)	1	$ 811,241.44
PHI	$ 1,062,445.35	4	$ (808,851.17)	3	$ 253,594.18
PIT	$ 1,464,353.67	3	$ (441,510.44)	4	$ 1,022,843.23
SJ	$ 691,928.67	3	$ (401,954.05)	4	$ 289,974.62
STL	$ 977,154.80	4	$ (811,952.94)	2	$ 165,201.86
TB	$ 608,053.78	4	$ (1,384,210.50)	1	**$ (776,156.72)**
TOR	$ 1,080,100.19	4	$ (542,981.81)	2	$ 537,118.38
VAN	$ 897,326.21	3	$ (597,425.99)	5	$ 299,900.22
VGK	$ 1,053,584.75	7	$ (1,144,267.75)	2	**$ (90,683.00)**
WPG	$ 1,300,207.82	3	$ (750,669.43)	5	$ 549,538.40
WSH	$ 867,913.16	4	$ (742,676.42)	3	$ 125,236.74

Figure 4. Average Team Over/Under Payment

in the league with a plus/minus that is less than six figures. The remaining 6 teams that underpay more on average than they overpay are the Dallas Stars, L.A. Kings, New Jersey Devils, New York Rangers, Tampa Bay Lightning, and the Vegas Golden Knights.

i. Time On Ice

For this analysis, I broke up all 209 players whose salary I have predicted into four sets based on their time on ice for 5V5 play. Time on ice varies from a minimum of 52 minutes to its’ maximum for this dataset of 1,284 minutes. **Group 1** contains any players with a time on ice from 0-321 minutes (n=49), **Group 2** from 322-642 minutes (n=38), **Group 3** from 643-963 (n=71), and **Group 4** from 944-1284 minutes (n=51). Using the *diff* variable generated, I averaged each players net predicted salary across the four time groups. **Group 1** easily had the largest, with a general average overpayment of roughly $1,115,000. **Groups 2** and **3** also both returned average overpayments for their corresponding ranges at $356,646.05 and $72,938.86 There is an apparent trend here between time on ice and salary. Those with little ice time, on average, are predicted to have a much lower salary than the salary cap they received at this time. This over payment gradually lowers, until we reach the players with much higher time on ice statistics, and the average overpayment transitions into an average underpayment (see *Figure 5.*). Again, we need to keep in mind the reverse casualty of the time on ice variables, as we are still

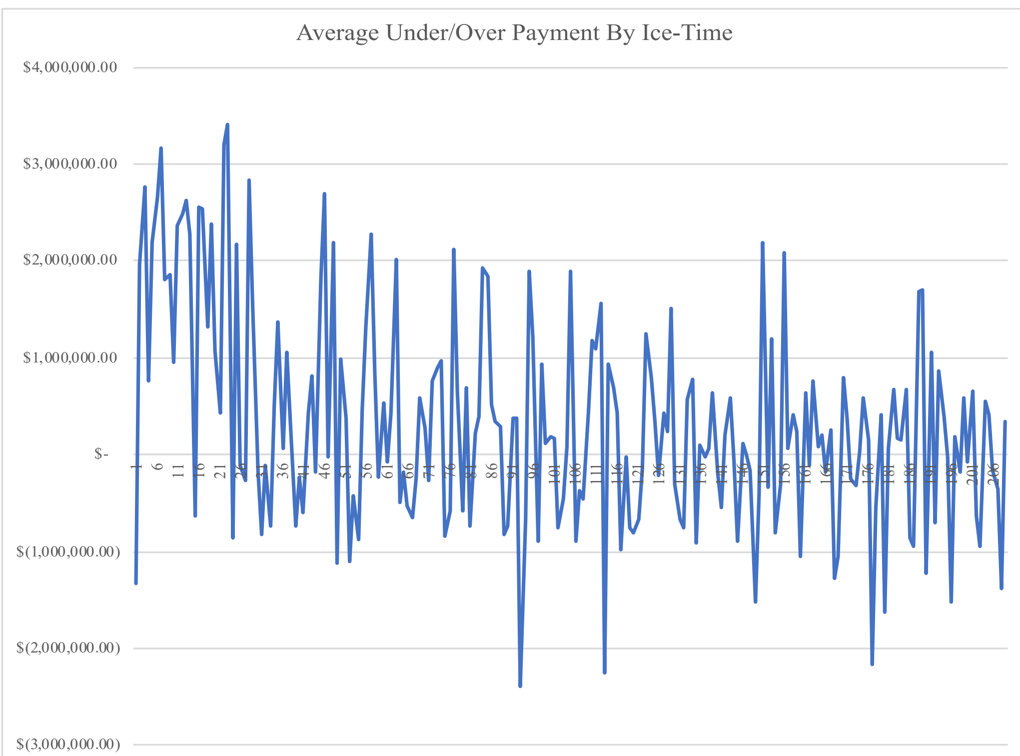

Figure 5. Average Over/Under Payment By Ice-Time

unsure as to which direction this relationship goes at this point.

E. Robustness

i. Secondary Variable Choice

Initially as I stated, when choosing my variables I went with only the ones that had the highest r-squared values given their bivariate regressions against salary. Those results were admirable, however not as explanatory as my final regression as noted from the table in *Figure 6.* below.

ii. Log-linear Model

I considered, given the long positive tail of the salary distribution given here, to try and utilize a log-linear specification in an attempt to make the distribution more normal, and perhaps fix some of my negative predictive values explained in the previous section. However as you can see in *Figure 7.* & *Figure 8.* below, after taking the log the normality didn't increase by much, and neither did the fit of my predictions. Further, after running the same specification with a log-transformed dependent variable, the R^2 value actually lowered (see *Figure 9.*), giving me no suitable reason to use this model. Indicating that the linear model I used is more explanatory given these variables than the log-linear model.

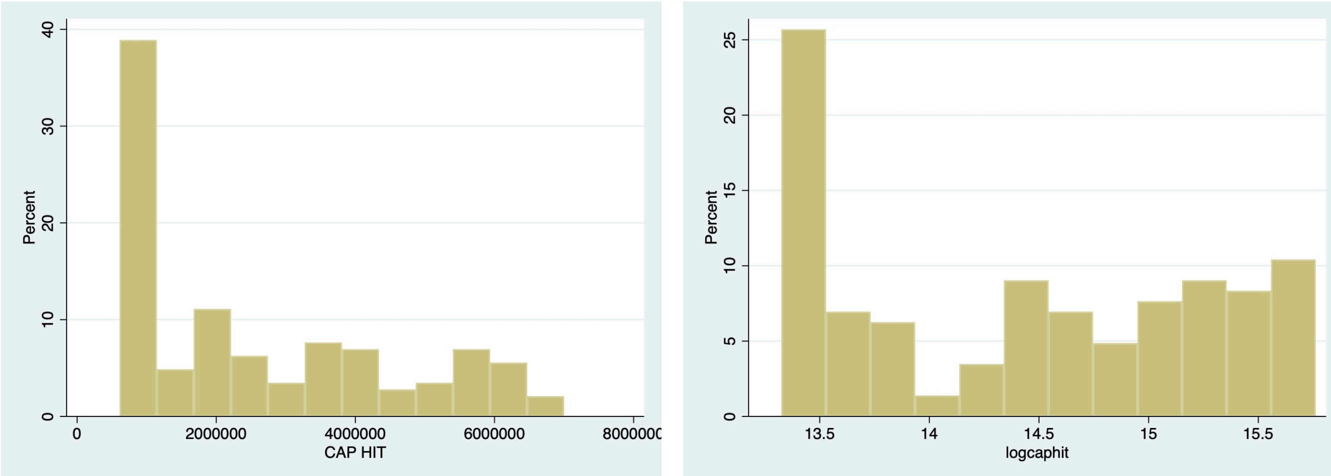

Figure 7. Salary Distribution (Left) vs. Log Salary Distribution (Right)

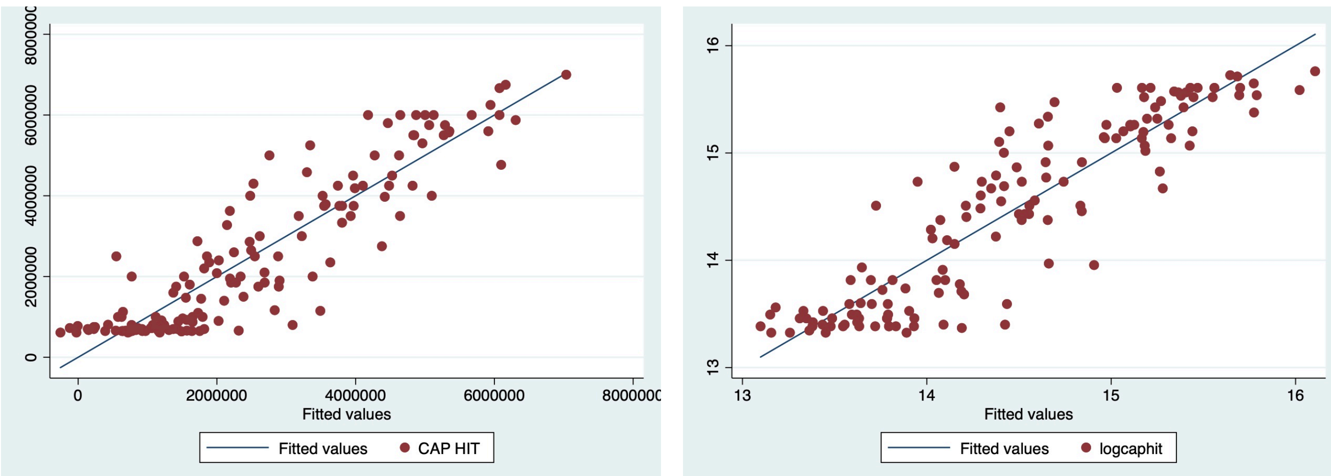

Figure 8. Fitted Values (Left) vs. Log Fitted Values (Right)

iii. noPP noPK Dummies

The inclusion of these two dummies was in order to reduce the amount of total variables in the regression. Otherwise, with the inclusion of a dummy value for those with powerplay or penalty-kill statistics, the interaction of these dummies with each variable in the regression would greatly increase the amount of overall variables causing the specification itself to be of an immense size. Instead, by adding these two dummies to the regression alone, we assume that it averages that effect across all players, and that the return to these statistics for both lines is the same for the players who are played on both. It may not be an overly accurate assumption, but it the filler I can provide without causing my regression to become too large. Regardless, without these controls as well my regression returns a lower R^2, as can be seen in *Figure 10.*.

F. Concluding Remarks & Limitations

Ultimately, my salary function was able to predict an admirable level of variation in the salary for wingers on standard and 35+ level contracts. However, some of the coefficients are still unintuitive, and may require further research beyond the scope of my study at this moment. I have my best assumptions as to the reasons for these directions, but they are simply assumptions and thus we are limited by that. The results of the controls for team, and long term experience effects are consistent with that of the previous research, in that they do have a significant effect on the salary of an individual player regardless of skill level. We are limited in that aspect by only having one season of data to work with, as I cannot look at how these skills have progressed over the course of these players careers, perhaps to get a better understanding of what skills improve along with salary. However, given the season specific data, we are able to conclude with a degree of certainty that these variables as possible proxies for puck possession, passing, and aggressiveness are of import to salary determination, along with the other necessary controls. Additionally I was able to get an worthy estimate of all wingers salaries with the function, although there are some negative values in the salary predictions, further research may be beneficial to eliminate or diminish the effects that are causing the negative predictions of salary for these players. This may involve further restricting the sample to account for players who have very little time on ice. Additionally further research regarding the direction of the relationship between time on ice, salary could be beneficial to this study and others of its' kind.

Regression Output

	VARIABLES	(1) reg1 CAP HIT
	CAP HIT	
5V5	iFF/60	66,653 (53,661) 0.217
	iGVA/60	455,133** (179,137) 0.0127
	iTKA/60	55,239 (206,020) 0.789
	iHF/60	-861.0 (27,122) 0.975
	iHA/60	-51,206 (71,529) 0.476
	TOI	735.4 (600.8) 0.224
PP	iFF/60	-56,364* (29,325) 0.0577
	iGVA/60	79,522 (123,036) 0.520
	iHA/60	117,624 (106,852) 0.274
	TOI	13,409*** (2,944) 1.59e-05
PK	iFF/60	38,034 (78,418) 0.629
	iPEND/60	185,126 (235,839) 0.434
	iGVA/60	42,575 (247,532) 0.864
	iHA/60	271,275 (173,850) 0.122
	TOI	3,373 (5,335) 0.529
	noPP	-89,045 (641,399) 0.890
	noPK	853,606 (659,933) 0.199
	AGE	90,604** (38,977) 0.0223
	(Dummy Controls)	
	Expiry	Yes
	Type	Yes
	Type	Yes
	Observations	144
	R-squared	0.802

Figure 6. Alternate Variable

Regression Output

	VARIABLES	(1) reg1 logcaphit
	logcaphit	
5V	iCF/60	0.0158 (0.0184) 0.392
	iGVA/60	0.197** (0.0772) 0.0125
	iHA/60	-0.0283 (0.0311) 0.365
	TOI	0.000678** (0.000259) 0.0103
P	iCF/60	-0.0138 (0.00999) 0.170
	iGVA/60	0.0650 (0.0572) 0.259
	iTKA/60	-0.0757 (0.107) 0.479
	iHA/60	0.0723 (0.0478) 0.134
	TOI	0.00334** (0.00128) 0.0107
P	iFF/60	0.0422 (0.0342) 0.220
	iGVA/60	-0.0102 (0.109) 0.926
	iHA/60	0.109 (0.0764) 0.158
	TOI	0.00168 (0.00230)
	AGE	0.0412** (0.0172) 0.0182
	noPP	-0.0990 (0.283) 0.727
	noPK	0.492* (0.283) 0.0857
	(Dummy Controls)	
	Type	Yes
	Expiry	Yes
	Team	Yes
	Constant	10.68*** (0.886) 0
	Observations	144
	R-squared	0.785

Figure 9. Log-Linear Regression

Regression Output

	VARIABLES	(1) reg1 CAP HIT
	CAP HIT	
5V	iCF/60	59,648 (39,790) 0.137
	iGVA/60	462,110*** (168,193) 0.00716
	iHA/60	-77,028 (66,087) 0.247
	TOI	765.6 (567.9) 0.181
P	iCF/60	-45,878** (18,554) 0.0152
	iGVA/60	200,996 (125,109) 0.111
	iTKA/60	-283,820 (222,341) 0.205
	iHA/60	115,470 (84,537) 0.175
	TOI	13,205*** (2,612) 2.03e-06
P	iFF/60	2,076 (63,168) 0.974
	iGVA/60	81,581 (232,822) 0.727
	iHA/60	204,061 (162,079) 0.211
	TOI	-826.6 (3,823) 0.829
	AGE	86,277** (37,350) 0.0230
	(Dummy Controls)	
	Type	Yes
	Expiry	Yes
	Team	Yes
	Constant	-5.200e+06*** (1.814e+06) 0.00508
	Observations	144
	R-squared	0.804

Figure 10. Regression Without noPP & noPK Dummies

Works Cited

Cebula, R. J. (2009). Teaching how private enterprise works using professional sports: A brief note on the case of individual NHL players' salaries. *Journal of Private Enterprise, 24*(2), 165-174. Retrieved from http://ezproxy.library.ubc.ca/login?url=https://search.proquest.com/docview/215100968?accountid=14656

Idson, T., & Kahane, L. (2000). Team effects on compensation: An application to salary determination in the National Hockey League. Economic Inquiry, 38(2), 345-357. doi:10.1111/j.1465-7295.2000.tb00022.x

Jones, J., & Walsh, W. (1988). Salary Determination in the National Hockey League: The Effects of Skills, Franchise Characteristics, and Discrimination. *Industrial and Labor Relations Review, 41*(4), 592-604. doi:10.2307/2523593

Levine, T. (2011). Two Worlds Collide: Salary Arbitration for NHL Players in the Salary Cap Era. *Ohio State Journal On Dispute Resolution*, *26*(4), 729.

Shirreffs, B., & Sommers, P. (2006). The Effect of Salary Arbitration on NHL Player Performance. *International Advances In Economic Research*, *12*(1), 142-142. doi: 10.1007/s11294-006-6151-x

Vincent, C., & Eastman, B. (2009). Determinants of Pay in the NHL. *Journal of Sports Economics,10*(3), 256-277. doi:10.1177/1527002508327519