THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

EVALUATING INEFFICIENCIES IN THE NATIONAL HOCKEY LEAGUE LABOR MARKET

ADAM RASTATTER
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Economics and Accounting
with honors in Economics

Reviewed and approved* by the following:

Colin Knapp
Senior Lecturer of Economics
Honors Adviser

Russell Chuderewicz
Senior Lecturer of Economics
Thesis Supervisor

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The National Hockey League has evaluated players using the same metrics for decades. Although several new statistics have been proposed, evidence suggests that managers still use outdated and imprecise measures to drive labor market decisions. Since the National Hockey League capped team salary expenditures at $73 million per year in 2016, every contract is crucial. Bill James started a revolution in Major League Baseball, using more advanced and exact measures to evaluate player performance. Do advanced baseball statistics such as on-base percentage have a hockey equivalent? If so, which of these statistics have a high correlation to player success? Finally, which of these statistics are currently underutilized in the National Hockey League labor market? This paper aims to explore these questions

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

-To my parents, who have pushed me to achieve my ambitious goals.

-To Professor Knapp for helping me throughout the entire process.

-To my sisters, who helped proofread and edit on multiple occasions.

-To Isai Martinez, for assisting me with my regression work.

Introduction

Moneyball

Before the 21st century most professional sports leagues used qualitative measures to evaluate players. This evaluation was often referred to as the "eye test." In 2003, a book called *Moneyball: The Art of Winning an Unfair Game* changed conventional thinking as it documented the introduction of alternative evaluative measures for managers. The book centered around the Oakland Athletics and their improbable journey to four consecutive American League Divisional Series appearances from 2000 to 2003. The Athletics were considered a small market team, spending less on player salaries than 26 other Major League Baseball clubs. In 2000, the Athletics nearly beat the Yankees in the American League Divisional Series despite spending $80,672,044 less on players. Much of the league was perplexed by the Oakland Athletics. How could they be in yearly contention while spending a fraction on payroll (2000 Baseball Payrolls)?

Applying SABR Metrics

How did a league dominated by big market clubs get turned upside down by the small market Athletics? The answer lies in the numbers. Oakland's general manager Billy Beane began hiring statisticians and economists to use Society of American Baseball Research or Sabremetrics to evaluate players. Sabremetrics focus on eliminating conventional biases in player evaluation and assessing players based on the single most important baseball measure – the creation of runs. These analysts used regression analysis to determine what characteristics were highly correlated to run creation but were being undervalued in the labor market. For example, instead of strictly assessing hitters on batting average (hits/total at bats) as was the norm, they began using on-base percentage ((hit+walk+hit by pitch)/total at bats). Billy Beane and his analysts recognized that a walk or hit by pitch created the same result as a single. Either way, a player ended up on base and provided an opportunity to create runs. This enabled the Oakland Athletics' management to sign bargain contracts with players that were exceptional at reaching base in nonconventional. By digging into the numbers and recognizing inefficiencies in the evaluation of players, the Athletics were able to prosper as a small market team in an inequitable league (James).

The Batting Average of Hockey

Much like other professional sports, there is a need for quantitative evaluation in hockey. For a number of years, management and the media have used a statistic called plus/minus as a means for assessing individual players. Like batting average, plus/minus is a flawed statistic. Plus/minus measures a player's contribution by netting his team's goals for and against while he is on the ice. The issue is that goals are an increasingly rare occurrence in the National Hockey League. To take a statistic with a low occurrence and a high variability and place reliance on it for player evaluation is problematic. To then have teams base labor decisions on such a statistic will create inefficiencies in the market. The National Hockey League caps total team salaries at $73 million. Because this amount is spread across 23 salaries, each contract is imperative. Consequently, there is an increasing interest to create statistics similar to on-base percentage that will capture characteristics that ultimately contribute to the most important result in hockey – scoring goals.

Chapter 1: Literature Review of Advanced Statistics in Hockey

Statistics in Hockey

Statistics have existed in hockey since the advent of the game. Goals and assists have been tracked since the inception of the league. In the 1960s, the Montreal Canadiens realized that these simple statistics did not entirely account for a player's impact while on the ice. The Canadiens created plus/minus as a way to capture more information about player effectiveness. Unfortunately, it has many flaws that limit its ability to fairly evaluate players (Corsi Plus-Minus).

Plus/Minus

Plus/minus is used today by coaches, general managers, and the media as an authoritative statistic for player evaluation. The premise is simple. When a player is on the ice and a goal is scored by his team, that player earns a plus. When a player is on the ice for a goal scored against his team, he earns a minus. A player's pluses and minuses are aggregated through their season or career to arrive at a final valuation. A player with a plus/minus greater than zero is considered proficient, while a value below zero is deficient. The issue many hockey statisticians have with the statistic is that "goals are rare, and with any rare statistical event it becomes easy for the sample to become skewed by outliers and not be indicative of the "true-talent level" or population (Hohl)." For example, Alexander Ovechkin is considered by many to be the best pure goal scorer of this generation. In 2013, Ovechkin led the league with 51 goals – the next closest being 43 – while also finishing third worst in the league among forwards with a -36 plus/minus. Ovechkin is the fourth highest paid player in the league, making $9.5 million per year (NHL Salary Rankings). Does his plus/minus make him undeserving of his $9.5 million contract? Or, is it accurate to conclude that Ovechkin is the best goal scorer in the league?

One flaw plus/minus has in its current state is the way it is measured. The National hockey league limits plus/minus to even strength and shorthanded situations. Stated differently, goals scored while on the power play are not factored into a player's plus/minus. Therefore, players are not rewarded for participating in situations with a high probability of scoring. But, the statistic develops inconsistencies because goals scored shorthanded are included in a player's plus/minus. So, when a player participates on the power play his plus/minus either remains unchanged (regardless if his team scores) or is negatively affected (if a shorthanded goal is scored against). Conversely, a player that only participates on the penalty kill, and not the power play, can only improve his plus/minus (Go Figure).

Another flaw inherent in the measurement of the statistic is that goals scored with a goalie pulled are included in plus/minus. Pulling a goalie is a move of desperation that leads to high probability scoring chances. By arbitrarily including some goals, plus/minus creates biases for certain player types. Defensively oriented players are rewarded while offensively gifted players are punished. Figure A further illustrates this point (Hohl). The graph is an accumulation of all players since 2007 and their subsequent ice time and plus/minus. Time on ice is indicative of player ability as better players will generally play more. If plus/minus is indicative of player ability, it should have a correlation to time on ice. As we can see, there is no trend between a player's plus/minus and the amount of minutes he plays.

Besides inherent flaws from the composition of the statistic, there are a number of other issues with plus/minus. One issue is that a player's goalie directly affects a skater's plus/minus differential. Suppose, for example, that you have two players: A and B. Assume both players are identical in every way. Each strictly plays even-strength hockey. Each takes and allows 15 shots per night and scores on 10% of the shots they take. Now assume that Player A has a goalie that saves 94% of shots faced while Player B's goalie only saves 88%. Each player's plus/minus differential over the course of the season would be calculated as follows:

Player A Game Plus/Minus Differential = [15 (Shots per game) x 10% (shooting %)] - [15 (Shots allowed per game) x (1-94%) (goalie save %)] = 0.60+/- Per Game

Player A Season Plus/Minus Differential = 0.60 x 82 (Games/Season) = 49.2

Player B Game Plus/Minus Differential = [15 (Shots per game) x 10% (shooting %)] - [15 (Shots allowed per game) x (1-88%) (goalie save %)] = -0.30 +/- Per Game

Player B Season Plus/Minus Differential = -0.30 x 82 = -24.6

A forward or defenseman has little control over a goalie's save percentage. By playing on a team with a better goalie, players can substantially increase their plus/minus. In this example, Player A had an 73.8 point advantage over Player B simply because he has a more adept goalie. In addition to goalie ability, linemate shooting percentage plays a major role in a player's plus/minus as well. If a player's linemates have abnormally high shooting percentages, that would skew the data as well. Imagine a scenario where player A's linemates shoot 14% while player B's linemates shoot 6%. Similar to the goaltender example, although players have little ability to affect their linemates shooting percentage, they can be greatly rewarded or punished for it (Lipson).

A Better Measure

If plus/minus has its shortcomings, why do National Hockey League managers still use it to make labor market decisions? The answer is, because there has not been a Bill James to revolutionize statistics for the NHL as he did for MLB. There has not been a single figure that has been able to shift the mindset of management like James did for baseball in the early 2000s. There are signs of change though. Some teams have taken to hiring departments tasked with analyzing players based on more advanced measures. In 2014, Ryan Pike wrote of the developing arms race for statistic departments. At that time, ten teams listed in-house stats personnel on their team directory. Now, nearly every team has dedicated a department to statistical analysis (Pike). Even so, plus/minus still remains the primary statistic used today. If this is the case then what statistics are more highly correlated to success?

In 2006, financial analyst Tim Barnes created a statistic called ‘Corsi’ that improved on some of the inherent flaws of plus/minus. The premise is to evaluate players based on an event that was highly correlated to goal scoring but occurred at a more frequent rate – therefore accounting for the high variability. Barnes decided to evaluate players based on shot attempts, as shot attempts were a precursor to goals and occurred frequently. Barnes defined an attempt as anything that was shot towards the net. This included goals, saves, shots off the post, shots that went wide, and shots that were blocked. He then accounted for both attempts for and against while a player was on the ice. Both statistics were aggregated, then used for a number of evaluative measures. For example, the first and most common metric created with this data was

Corsi For Percentage. To calculate player A's Corsi For Percentage, take his team's shot attempts and divide it by the total shot attempts of both teams while he is on the ice. To illustrate, imagine a player A's team has 10 shot attempts while he is on the ice. Additionally, assume there are 5 shot attempts against him while he is on the ice. The player's Corsi For Percentage would be calculated as: 10/(5+10) = 67%. So, while player A was on the ice, his team recorded 67% of the shot attempts (Mckenzie).

Corsi also negates shooting and save percentage biases created by teammates. Corsi measures shot attempts rather than the result of the attempt. Whether the player's goalie makes a save or linemate scores a goal is irrelevant.

However, analysts questioned the composition of Corsi. They did not believe that blocked shot attempts should be included. Many analysts believed that shot blocking is a skill players possess rather than a random event. They believed that over large sample sizes, unblocked shot attempts would be able to predict future success more accurately than all shot attempts (Gretz). To test this hypothesis, analysts created Fenwick. Fenwick is the same measure as Corsi, with the exception that it does not classify blocked shots as a shot attempt.

Most statisticians were still not content with the limitations of Fenwick and Corsi, though. They argued that both statistics fail to truly account for shot quality. "That some shots are better than others is a core tenet of hockey and indeed any such sport (Perry)." A wrist shot from the blue line is much less likely to create a goal than a slap shot from the slot. But, with Corsi and Fenwick, all shot attempts are treated equally.

One measure that accounts for shot quality is a scoring chance. The NHL tracks scoring chances each game, though they do not set parameters as it is a subjective measurement. Generally speaking, if someone tracking a game believes a shot attempt has a high probability of creating a goal, it is considered a scoring chance. It is clear how subjectivity could produce adverse results. Although imprecise, it was a starting point for the statistics community to investigate shot quality.

In one study, Carolina Hurricane's statistician Eric Tulsky sought to investigate the relationship between shot quality and scoring chances. In the study, Fenwick was regressed against scoring chances to discern whether any information relating to the prediction of future success was missing. Tulsky found that Fenwick and scoring chances had a correlation of 0.80. In other words, Fenwick and scoring chances tell much of the same story. But, upon further investigation, Tulsky noticed that "some players do lie a reasonable distance above or below the fit line, getting scoring chances at a rate like a player whose shot differential is a few percent better (Tlusky)." To investigate further, Tulsky divided scoring chances by Fenwick. This illustrates the shot quality advantage a team has while a particular player was on the ice. For each player in the data set, their 2009 contribution to shot quality was regressed against their 2010 contribution. The results showed no predictability from one year to the next (Tlusky).

Despite these results, statisticians postulated that shot quality was important, but it needed to be measured objectively. To do so, analysts created Expected Goal Models. The premise is to find a way to quantify the quality of a player's shot. Although inputs vary in each model, statistician Emmanuel Perry's results are the most conclusive. Included in his model are variables such as:

- Shot type (Wrist shot, slap shot, deflection, etc.)
- Shot distance (Adjusted distance from net)
- Shot angle (Angle in absolute degrees from the central line normal to the goal line)
- Rebounds (Whether or not the shot was taken from a rebound)

- Rush shots (Whether or not the shot was a rush shot)
- Strength state (Whether or not the shot was taken on the powerplay)

Each shot sub-type (wrist shot, slap shot, deflection) or characteristic is regressed independently of the other variables. "The rationale here is that each shot subset should respond to the variables differently (Perry)." For example, factors such as shot distance or shot angle will affect a wrist shot differently than these variables will affect a slap shot. Both shot distance and angle are assumed to be non-linear. To clarify, it is expected that shot quality does not change at a constant rate as you move further from the net. Once these assumptions are made, a player's shots are evaluated using a logistical regression to arrive at an expected quantity of goals for a given player during a given season. You can then create an expected goal percentage by taking Expected Goals For while a player is on the ice and dividing it by total Expected Goals For while the same player is on the ice – similar to Corsi For Percentage. Results show that this model has a much higher correlation to actual goals scored (0.750) than Fenwick (0.586) and even shots on goal (0.619), while shots on goal has the advantage of only including goals and saves (Perry).

Persisting Limitations

One limitation inherent in all three statistics is the ability to isolate players from their teammates. Hockey is an interconnected game where teammates are constantly using one another to create plays. It becomes difficult to evaluate one player without his teammate bias affecting results. There have been attempts to remedy this situation with Team Relative Corsi and Corsi Plus/Minus differential but neither has produced conclusive results. The scope of this paper will only include Corsi Percentage, Fenwick Percentage and Expected Goal Models.

Chapter 2: Developing Regression Models

First Regression: Predicting Success

Before we can know which statistics are underutilized in the labor market, we must first understand which statistics are the best predictors of success. In hockey, on an individual level, success is measured in points. When a player scores a goal or assists on another player's goal, they receive a point. But, simple points do not always tell the entire story. It is important to normalize for playing time to allow for objective comparisons across all players. To better understand, consider two players:

Player A: 40 points, 50 games, 0.80 points per game

Player B: 50 point, 50 games, 1.00 points per game

Strictly from these numbers, even when we normalize for games, player B is preferable. But, consider a scenario where Player A plays 10 minutes per night while Player B plays 20 minutes per night. Because player B plays twice the minutes as player A, he should be expected to record more points given the same amount of games. If instead, the points are normalized for comparison to the same minute rate, the outcome reveals an accurate depiction of the player's success. Below is an example of this concept:

Player A: 40 points/(50 games X 10 mins) = 0.08 points per minute X 60 = 4.8 points per 60

Player B: 50 points/(50 games X 20 mins) = 0.05 points per minute X 60 = 3.0 points per 60

As you can see, the points were first normalized to a per minute value then were multiplied by 60 to get a per hour rate. Points per 60 minutes is a commonly used statistic in the analytics community. It is used to eliminate playing time differences which allows for more accurate comparisons across players.

Furthermore, I will only be using primary points per 60. There are three ways to earn a point: score the goal, be the player that makes the pass to the goal scorer (referred to as the primary assist), or be the player that makes the pass to the player that makes the pass to the goal scorer (referred to as the secondary assist). Much like shot quality, not all assists are created equal. There is a growing consensus in the hockey analytics community that secondary assists are "noise" rather than accurate predictors of future success. The belief is that if a statistic is indicative of future success, then it should persist from one year to the next. In another study by Tulsky, he attempts to prove this claim. Figure B suggests that goal scoring for a player from year 1 to year 2 has a correlation of 0.22. "A correlation of 0.22 actually means quite a bit of fluctuation (just ask Alex Ovechkin), but I think we all believe that scoring goals is a talent, so this tells us that we should consider a correlation near or above 0.22 to show that something is a result of a player's talent (Tulsky)." So essentially, Tulsky is saying that goals should be a measuring stick when evaluating the predictability of points. When evaluating the persistency of primary assists, as Figure C suggests, the correlation drops slightly to 0.18. This indicates that primary assists are nearly as predictable from year-to- year as goals. Finally, when measuring the same for secondary assists, Figure D shows the correlation drop to 0.05. Note that the entire goal of advanced statistics is to predict future success. Although predicting goals and primary assists is an extremely difficult task, it is markedly easier than secondary assists. From Tulsky's research, there is reason to believe that secondary assists are much less predictable than primary assists and goals. Because there is such a drop in predictability and persistency, I will not be including secondary assists. My measure of success will be primary points per 60.

It is important to also note that I will only be using even strength of play data in my analysis. Players that participate on the power play hold an unfair advantage over those who do not. There is a much higher likelihood that a player will earn a point while participating on the power play than one who is playing even strength. Likewise, it is much less likely that a player participating on the penalty kill will record a point than a player playing at even strength. To eliminate this issue I simply removed all points and minutes not played at even strength. Therefore, I am defining success as even strength primary points per 60.

For my first model, I will be running a simple Ordinary Least Squares regression to determine which statistics are correlated with success. My dependent variable will be even strength primary points per 60, as described above. I will be running three regressions to avoid multicollinearity between variables. Corsi For Percentage, Fenwick For Percentage, and Expected Goal Percentage will each have their own regression. If these statistics have a high correlation with primary points per 60, as I hypothesize, they should also have a high correlation to player salaries, otherwise labor market inefficiencies may exist.

One limitation of this method is the classification of success for defensemen. For forwards, although they are expected to contribute in the defensive zone, their primary role is to create points. For defensemen, this isn’t necessarily the case. Defenseman are expected to limit the other teams scoring opportunities while pushing play towards the offensive zone. Unfortunately, there is no one statistic that is indicative of defensive success. As a result, primary points per 60 will be used as a measure of success for defensemen, as creating points is still an important measure. I will be using a dichotomous variable to account for positional differences. The forward position will be denoted by the number zero, while defensemen will be denoted with a one. The interaction variable will be used measure the effect of the advanced statistics on each position.

The first three regression model will be as follows:

1. Even Strength Primary Points Per 60 = $\beta_o + \beta_1$ * Corsi + β_2 * Position + β_3 * (Corsi * Position)
2. Even Strength Primary Points Per 60 = $\beta_o + \beta_1$ * Fenwick + β_2 * Position + β_3* (Fenwick * Position)
3. Even Strength Primary Points Per 60 = $\beta_o + \beta_1$ * Expected Goals + β_2 * Position + β_3 * (Expected Goals * Position)

I aim to measure how well these advanced statistics correlate to success.

Second Regression: Contract Information

To understand whether or not specific players are being undervalued in the NHL labor market, I compiled data on players who were freely subject to the labor market at some point in their career. In other words, only players without league or team restrictions. In the NHL there are three types of contracts, two of which restrict players from fully participating in the labor market. One example is an entry level contract that players sign when first entering the league. Below is a summary from the National Hockey League Players Association's most recent collective bargaining agreement regarding when a player must sign an entry level contract and for what length. SPC is short for Standard Player Contract.

First SPC Signing Age	Period Covered by First SPC and Years in the Entry Level System and Subject to Compensation Limits
18-21	3 years
22-23	2 years
24	1 year

(Collective Bargaining. pg 23)

If a player is subject to an entry level contract, they can make no more than $925,000 average annual value for the length of the contract (Collective Bargaining. pg 24). This restriction placed on players entering the league creates a price ceiling that makes them ineligible for my study because they are not truly subject to the labor market.

After a player completes their entry level contract, they are either eligible for a restricted or unrestricted free agency contract. According to the same collective bargaining agreement, a couple of requirements must be met in order to be an unrestricted free agent:

i. A player that has played seven seasons in the NHL/AHL or is older than 27 before June 30th

ii. A player that is 25 or older before June 30th and has completed three full NHL season.

(Collective Bargaining. pg 25)

If either of these requirements are met, the player is considered an unrestricted free agent. A player that is deemed a UFA may freely explore the labor market without team or league restrictions. These are the contracts I am interested in because they allow a player to sign with any team for any amount, therefore fully exposing them to the labor market.

The contract data I will be utilizing comes from Matt Cane, an NHL analyst who used a script to scrape data from the now defunct site, WarOnIce. The data set initially included nearly 1,000 unrestricted free agency contracts spanning from 2010 to 2016. This data was trimmed further after eliminating two way contracts. These contracts allow teams to easily maneuver a player between an NHL team and its AHL affiliate. Because players make different salaries depending on which team they are on, I eliminated these contracts. I limited the data one final time by eliminating players that recorded less than 1,000 minutes of ice time in the three years leading up to their contract. This helped eliminate abnormal results from small sample sizes.

The final observations to be used in the regressions are 339 forwards and 208 defensemen contracts.

My data set contains observations from 2010 to 2016. To properly match the evaluation of the player to the resulting contract, I will be using data from the three years preceding their contract year. For example, Jonathan Toews signed a contract in 2011 with an average annual value of $6,300,000. For this observation I aggregated data from the 2009, 2010, and 2011 seasons. Further, Jonathan Toews received another contract in 2016. For this observation, I aggregated data from the 2014, 2015, and 2016 seasons. This method was used for all 547 observations.

Second Regression: Independent Variables

After running my first regression and concluding which statistics have a high correlation with even strength primary points per 60, I will take the same data set and run an OLS regression with each player's average annual salary as the dependent variable. Figure F shows a sample of ten players included in the regression along with their particular data set.

The statistic aAAv is the players average annual salary over the length of their contract. I will once again use a dichotomous variable interaction to account for position. The set of regressions will be as such:

1. Player aAAV = $\beta_o + \beta_1$ * Corsi + β_2 * Position + β_3 * (Corsi * Position)
2. Player aAAV = $\beta_o + \beta_1$ * Fenwick + β_2 * Position + β_3* (Fenwick * Position)
3. Player aAAV = $\beta_o + \beta_1$ * Expected Goals + β_2 * Position + β_3 * (Expected Goals * Position)

Once both regressions are complete, I can compare significance values and correlation coefficients to conclude whether inefficiencies exist.

Chapter 3: Results and Analysis

Model 1 Results:

Dependent Variable: Primary Points Per 60	(1)	(2)	(3)
Corsi	0.047***		
Fenwick		0.046***	
Goals For			0.044***
Position	0.717**	0.744*	0.871**
Corsi:Position	-0.031***		
Fenwick:Position		-0.031***	
Goals For:Position			-0.034***
Constant	-1.121***	-1.059***	-0.996***
Observations	547	547	547
R^2	0.729	0.718	0.819
F Statistic (df = 3; 543)	486.272***	461.278***	819.281***
Note:	*p**p***p<0.01		

The results of the first three models were definitive. All three advanced statistics were statistically significant with a P value of 0.01. Stated differently, there is approximately a 1.0%

chance that the result of this sample is incorrect for the population. Each regression's correlation was strong as well. The most surprising result was the drop in R^2 from the Corsi model to the Fenwick model. Literature suggested that Fenwick was a better predictor of success. My results indicate the opposite as Corsi has a slightly stronger correlation. The equations, including the interaction were as follows:

1. PP/60 = -1.121 + 0.047*CF% + 0.717*POS – 0.031*CF%*POS
2. PP/60 = -1.059 + 0.046*CF% + 0.744*POS – 0.031*CF%*POS
3. PP/60 = -0.996 + 0.044*GF% + 0.871*POS – 0.034*GF%*POS

A few important conclusions to note. First, the Corsi model rewards incremental increases in advanced statistics more than the other two. The Corsi For % coefficient is 0.047. This means that for every 1% increase in Corsi, a forward is expected to have 0.047 more primary points per 60. Additionally, it appears that positional difference is a bigger determinent for the Expected Goals model than it is for the other two. The coeffieicnt of the position variable is significantly higher in the GF% model than it is Corsi or Fenwick. Even more interesting though is that the coeffient of position for all three equations is positive. Because the variable is dichotomous, the coefficient represents the increase in PP/60 of being a defensmen (1) rather than a forward (0). In reality, given a specific Corsi For %, forwards will almost always have a higher PP/60 than defensmen. The equation accounts for this through the interaction variable. Although the interaction variables' coeffiecient is small, it will have a significant effect in practice. Because I am using whole number percentages in my formula (I.E. 50% rather than 0.50), the interation term will have a huge effect on the PP/60 valuation. The negative coefficient will only harm defensmen as the interaction will disappear for forwards (Pos=0). One final observation is that the second term (GF, CF, FF) will have a huge impact on expected PP/60 for forwards. The third and fourth terms include position, which will disappear for any forward. To test the model, I chose a forward and a defensmen with identical PP/60 and input their values into the Expected Goal % model.

Forward: David Legwand – Expected PP/60 = -0.996 + 0.044*(46.19) + 0.871*(0) – 0.034*(46.19)*(0) = 1.036

Defenseman: Erik Karlsson – Expected PP/60 = -0.996 + 0.044*(50.57) + 0.871*(1) – 0.034*(50.57)*(1) = 0.3807

Over the three year span, both players actually had 0.92 PP/60. It is clear the model favors forwards over defensemen. Given a forward and defensemen with exactly identical GF%, a forward is expected to have 0.837 (0.871-0.034) more PP/60 than a defensemen. This result reflects the notion that the two positions have different expectations for creating offense.

Overall, the models make one thing clear, all three statistics are extremely reliable when predicting player success.

Model 2 Results:

Dependent Variable			
Annual Salary			
	(1)	(2)	(3)
Corsi	232,726.300***		
Fenwick		224,674.700***	
Goals For			187,185.700***
Position	3,371,345.000	3,547,840.000	534,633.200
Corsi:Position	-66,494.130		
Fenwick:Position		-70,386.780	
Goals For:Position			-9,952.099
Constant	8,273,169.000***	7,868,500.000***	5,988,942.000***
Observations	547	547	547
R^2	0.144	0.122	0.204
F Statistic (df = 3; 543)	30.345***	25.146***	46.422***
Note:	*p**p***p<0.01		

When analyzing the second model, there are some key results to note. Unsurprisingly, there is a positive relationship between these advanced statistics and player salaries. Additionally, each of the three variables is statistically significant at a P value of 0.01. Once again, strong conclusions can be drawn from the results. The R^2 shows that there is minimal-to-no correlation between these statistics and a player's salary. Of the three, the Expected Goals Model has the highest correlation. The equations for each regression is as follows:

1. Player AAV = -8,273,169.000 + 232,726.300*CF% + 3,371,345.000*POS - 66,494.130*CF%*POS
2. Player AAV = -7,868,500.000 + 224,674.700*FF + 3,547,840.000*POS - 70,386.780*FF%*POS
3. Player AAV = -5,988,942.000 + 187,185.700*GF% + 534,633.200*POS - 5,988,942.000*GF%*POS

One interesting result is the Expected Goals model values positional difference more than the other two. Specifically, the interaction term coefficient for the Expected Goals model is nearly $6,000,000 while the same coefficient doesn't eclipse $1,000,000 in the other two models. This model also punishes defensemen substantially. The reason could be that points are less important when determining a defenseman's contract than a forward. Forwards are expected to score more while defensemen are more concerned with preventing goals. Given the same Expected Goals % for both positions, a forward is expected to earn $5,454,308.8 (5,988,942-534,633.2) more than a defenseman.

Chapter 4: Concluding Remarks

To definitively state that there are inefficiencies in the National Hockey League labor market is impossible. The premise of this paper was to merely investigate the possibility that inefficiencies may exist – similar to baseball in the early 2000s. Given the statistical significance and correlation these advanced statistics have with primary points, it is clear they are important measures for predicting future player success. Their subsequent lack of correlation to player salaries over the last 10 years is a clear indication that inefficiencies may exist. Much like the Athletics, if hockey general managers can exploit these inefficiencies, there could be substantial returns for minimal investments.

Appendix

Figure A: Relationship Between Plus/Minus and Minutes

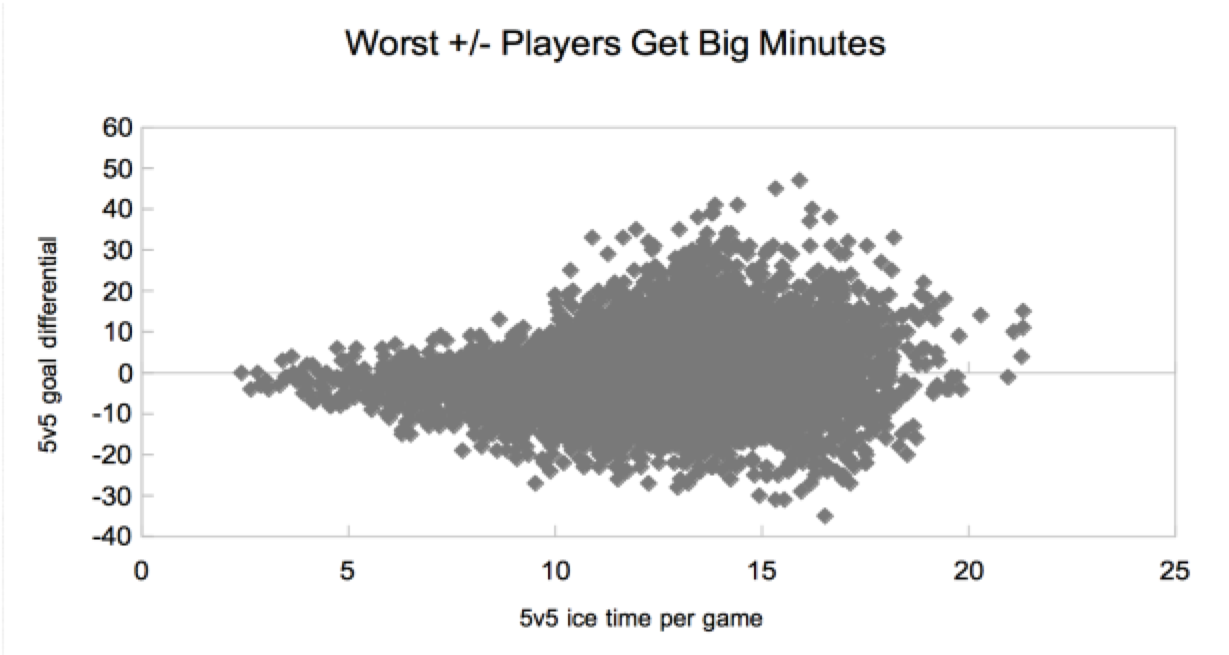

Figure B: Persistency of Year to Year Goal Scoring

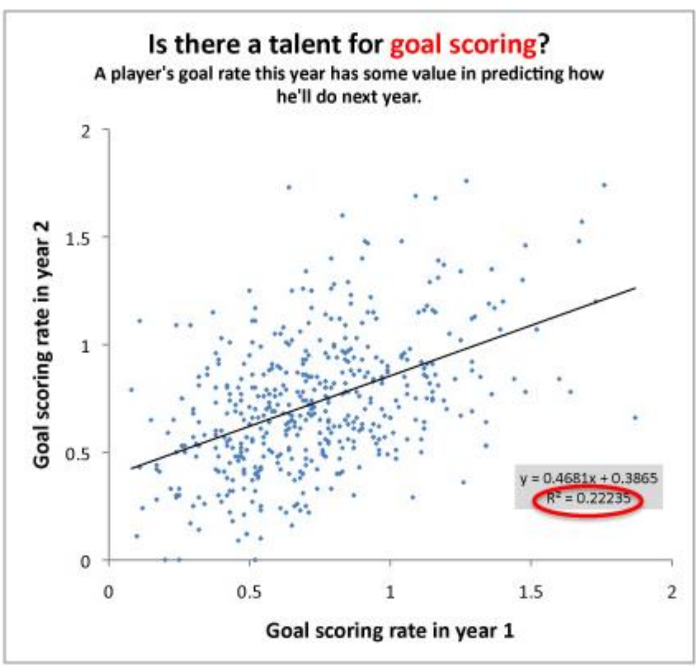

Figure C: Persistency of Year to Year Primary Assists

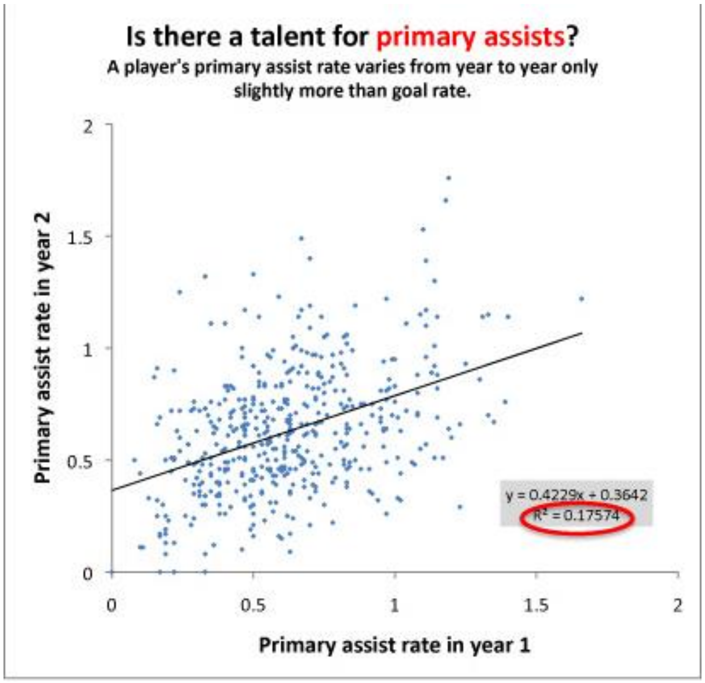

Figure D: Persistency of Year to Year Secondary Assists

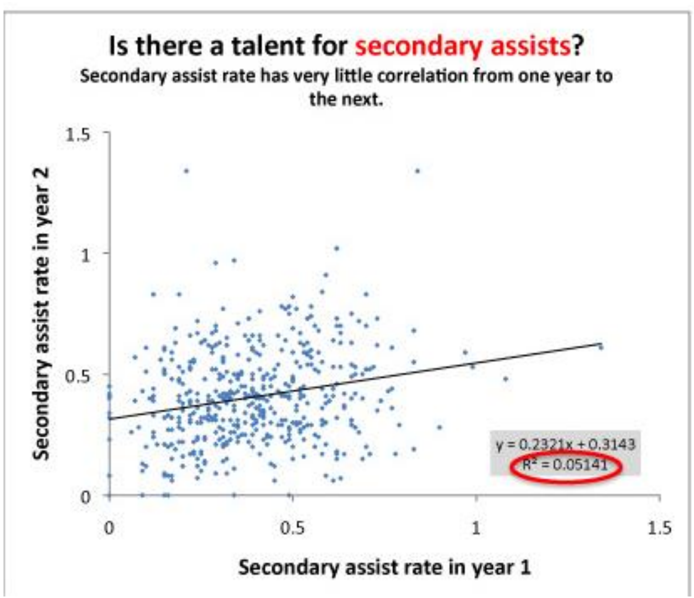

Figure E: Sampling of Statistics

1	Name	Year	Age	Contract.Length	Calc.AAV	FAStatus	GP	TOI/GP	G/60	A/60	A1/60	A2/60	CF%	FF%	xGF%	PDO
2	Jonathan Toews	2016	27	8	10500000	UFA	237	14.03181	0.82	1.21	0.78	0.43	55.3	54.34	53.71	101.14
3	Patrick Kane	2016	26	8	10500000	UFA	212	15.18283	1.14	1.27	0.9	0.37	53.28	52.58	50.66	101.54
4	Evgeni Malkin	2015	28	8	9500000	UFA	160	14.12338	0.94	1.39	0.84	0.55	54.55	53.74	53.58	101.37
5	Sidney Crosby	2014	26	12	8700000	UFA	138	15.07804	1.05	2.24	1.53	0.71	54.67	55.17	58.59	103.39
6	Corey Perry	2014	28	8	8625000	UFA	205	14.63888	1.16	1.02	0.74	0.28	52.72	52.82	54.46	100.89
7	Claude Giroux	2015	26	8	8275000	UFA	211	14.51033	0.64	1.18	0.75	0.43	52.42	51.87	51.15	100.15
8	Eric Staal	2010	24	7	8250000	UFA	233	14.17292	0.95	1.13	0.72	0.41	52.74	52.83	51.98	100.37
9	Ryan Getzlaf	2014	28	8	8250000	UFA	203	14.89369	0.63	1.54	0.99	0.55	51.62	51.81	53.74	101.49
10	Phil Kessel	2015	26	8	8000000	UFA	212	15.0641	1.02	1.11	0.7	0.41	45.26	44.41	44.05	101

Figure F: Correlation of Corsi For Percentage and Primary Points

```
===============================================
                        Dependent variable:    
                    ---------------------------
                               PP.60           
-----------------------------------------------
CF.                          0.047***          
                              (0.004)          
                                               
pos                           0.717**          
                              (0.361)          
                                               
CF.:pos                      -0.031***         
                              (0.007)          
                                               
Constant                     -1.121***         
                              (0.192)          
                                               
-----------------------------------------------
Observations                    547            
R2                             0.729           
Adjusted R2                    0.727           
Residual Std. Error      0.260 (df = 543)      
F Statistic          486.272*** (df = 3; 543)  
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Figure G: Correlation of Fenwick For Percentage and Primary Points

	Dependent variable: PP.60
FF.	0.046***
	(0.004)
pos	0.744*
	(0.381)
FF.:pos	-0.031***
	(0.008)
Constant	-1.059***
	(0.204)
Observations	547
R2	0.718
Adjusted R2	0.717
Residual Std. Error	0.265 (df = 543)
F Statistic	461.278*** (df = 3; 543)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure H: Correlation of Expected Goals For Percentage and Primary Points

	Dependent variable: PP.60
GF.	0.044***
	(0.002)
pos	0.871**
	(0.413)
GF.:pos	-0.034***
	(0.008)
Constant	-0.996***
	(0.100)
Observations	547
R2	0.819
Adjusted R2	0.818
Residual Std. Error	0.212 (df = 543)
F Statistic	819.281*** (df = 3; 543)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure I: Correlation of Corsi For Percentage and Salary

	Dependent variable: Calc.AAV
CF.	232,726.300***
	(26,708.850)
pos	3,371,345.000
	(2,527,537.000)
CF.:pos	-66,494.130
	(50,420.750)
Constant	-8,273,169.000***
	(1,344,356.000)
Observations	547
R2	0.144
Adjusted R2	0.139
Residual Std. Error	1,815,572.000 (df = 543)
F Statistic	30.345*** (df = 3; 543)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure J: Correlation of Fenwick For Percentage and Salary

	Dependent variable: Calc.AAV
FF.	224,674.700***
	(28,206.520)
pos	3,547,840.000
	(2,647,023.000)
FF.:pos	-70,386.780
	(52,734.580)
Constant	-7,868,500.000***
	(1,419,370.000)
Observations	547
R2	0.122
Adjusted R2	0.117
Residual Std. Error	1,838,323.000 (df = 543)
F Statistic	25.146*** (df = 3; 543)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure K:Correlation of Expected Goals For Percentage and Salary

	Dependent variable:
	Calc.AAV
GF.	187,185.700***
	(16,287.430)
pos	534,633.200
	(3,407,710.000)
GF.:pos	-9,952.099
	(68,096.630)
Constant	-5,988,942.000***
	(823,255.800)
Observations	547
R2	0.204
Adjusted R2	0.200
Residual Std. Error	1,750,220.000 (df = 543)
F Statistic	46.422*** (df = 3; 543)
Note:	*p<0.1; **p<0.05; ***p<0.01

Bibliography

"Bill James." *BR Bullpen*. N.p., n.d. Web. 22 Dec. 2013. <http://www.baseball-reference.com/bullpen/Bill_James>.

"Collective Bargaining Agreement." *NHL Players Association.* 12. Jan 2013. Web. 12. Jan. 2017. http://www.nhlpa.com/inside-nhlpa/collective-bargaining-agreement

"Corsi Plus-Minus: individual Player Value Accounting for Teammates." *Donttellmeaboutheart.blogspot*. n.d.Web.27.Jan.2017.< http://donttellmeaboutheart. blogspot.com/2015/07/corsi-plus-minus-individual- player.html>

"Go Figure" *NHL.com.* N.p. n.d. *Web. 21. Nov. 2016.* < http://www.nhl.com/ice/page.htm?id= 26374>

Hawerchuk. "Advanced Metrics: PDO and regression to the mean." *SB Nation.* 11.Dec.2011.Web.29.Dec.2016.< http://www.arcticicehockey.com /2011/12/20/2648333/pdo-regression-to-the-mean-or-why-you-should-ignore-shooting>

Hawerchuck. "PDO: If You Were Going to Understand Just One NHL Statistic." *SB Nation.* 28. Oct. 2011. Web.14. Jan.2017.< http://www.arcticicehockey.com/2011/10/28/2520115/ pdo-if-you-were-going-to-understand-just-one-nhl-statistic>

Hohl, Garret. "Behind the Numbers: Why Plus/Minus is the Worst Statistic in Hockey and Should be Abolished." *Hockey-Graphs*. 1. Nov. 2016. Web 27. Nov. 2016. < https://hockey-graphs.com/2016/11/01/behind-the-numbers-why-plusminus-is-the-worst-statistic-in-hockey-and-should-be-abolished/>

Langager, Chad. "Average Goals Scored Per Game in NHL." *Sporting Charts.* N.p. 15. Mar. 2014. Web 22. Dec. 2016. <https://www.sportingcharts.com/articles/nhl/average-goals-scored-per-game-in-the-nhl.aspx>

Lipson, Daniel. “Why Plus/Minus is the Worst Statistic in Hockey.” *SB Nation.* 5. June. 2014. Web. 17. Oct. 2016. <http://www.arcticicehockey.com/2014/6/5/5602668/why-plus-minus-is-the-worst-statistic-in-hockey>

Mckenzie, Bob. “Hockey Confidential.” *Collins; Reprint edition.* 14. Oct. 2014. Web. 2. Aug. 2016. < http://www.tsn.ca/mckenzie-the-real-story-of-how-corsi-got-its-name-1.100011>

“NHL Player Stats/Ratings.” *Stats.HockeyAnalysis*. N.p. n.d. Web.5.Jan.2017. < http://stats.hockeyanalysis.com/>

“NHL Salary Rankings.” *Spotrac.* USA TODAY Sports Media Group. n.d. Web.1.April.2017. < http://www.spotrac.com/nhl/rankings/>

Perry, Emmanuel. “Shot Quality and Expected Goals: Part 1” *Corsica.* 3. Mar.2016. Web. 9.Oct. 2016. < http://www.corsica.hockey/blog/2016/03/03/shot-quality-and-expected-goals-part-i/>

Pike, Ryan. “The NHL’s Advanced Stats Arms Race Heats Up.” *The Hockey Writers.* USA TODAY Sports Digital Properties.6.August.2014.Web.1.April.2017. < http://thehockeywriters.com/advanced-stats-arms-race/>

“Skater Stats.” *Corsica.* N.p. n.d. 4. Jan. 2017. < http://www.corsica.hockey/skaters/>

The Neutral. “Non-Traditional Metrics Glossary: PDO.” *SB Nation*. 2 Aug.2012.Web.1.Feb.2017< http://www.fearthefin.com/2012/8/2/3215351-non-traditional-metrics-glossary-pdo>

“Tlusky, Eric”. “Simplify Scoring: Drop the Pointless Secondary Assist.” *SB Nation.* 15. Mar. 2011. Web. 1. Nov. 2016. <http://www.broadstreethockey.com/2011/3/15/2046512/simplify-scoring-drop-secondary-assists>

"Tlusky, Eric." "Shot Quality Revisited: A Look at the Correlation Between Scoring Chances and Shot Quality." *NHL Numbers.* The Nation Network. n.d. Web.1.April.2017 <http://nhlnumbers.com/2012/6/26/shot-quality-revisited-a-look-at-the-correlation-between-scoring-chances-and-shot-totals>

"2000 Baseball Payrolls: Baseball Chronology." *Baseball Chronology*. N.p., n.d. Web 14. Jan. 2017. http://www.baseballchronology.com/Baseball/Years/2000/Payroll.asp

ACADEMIC VITA

Adam Rastatter
Adamrastatter@gmail.com

EDUCATION

Pennsylvania State University – Schreyer Honors College *2017*

- Bachelors of Science in Economics
- Bachelors of Science in Accounting

EXPERIENCE

International Timber and Veneer – Accountant *May 2016 – August 2016*

- Analyzed and developed an effective strategy to manage inventory based on trend mapping within the timber industry to initiate savings of $50,000
- Oversaw Accounts Payable and Receivables for the firm, enabling me to learn how a company operates through cash inflows and outflows
- Developed a time study analysis while also generating weekly payroll reports allowing me to evaluate contribution margins of each product line
- Aggregated financial data and provided summaries of pertinent information to the CEO and CFO, expediting their decision making processes

PHILANTHROPIC/LEADERSHIP ACTIVITIES

Penn State Dance Marathon *October 2013 – February 2016*

Chief Financial Officer

- Led and inspired a committee of 30 members while organizing fundraising events
- Responsible for ensuring all committee members executed tasks delegated to them for the 46 hour dance marathon
- Assisted in managing funds for the largest student run philanthropy in the world

Merchandise Committee Member

- Maintained merchandise stores with committee members, raising funds for a $10,000,000 philanthropy
- Facilitated surveys of the THON community, providing executives with optimal merchandising tactics

Finance Committee Member

- Effectively worked with a team of 30 students to oversee the funds of the philanthropy
- Used computer software to input money drop-offs allowing the philanthropy to aggregate donations