Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 2, 2024
A Regional Conference of the Society for Industrial and Systems Engineering

NHL Moneyball Goalie Analysis

Ryan Rast and James Enos

Department of Systems Engineering,
United States Military Academy,
West Point, NY 10996

Corresponding author's Email: james.enos@westpoint.edu

Author Note: Cadet Ryan Rast is a member of the Army Sprint Football team as well as a senior student in the Department of Systems Engineering at the United States Military Academy. This paper is written as a component for the requirements to graduate with honors in Systems Engineering.

Abstract: Winning doesn't have to come at a heavy cost. The Oakland Athletics nearly perfected this idea with Moneyball. The A's targeted players whose performance would contribute to wins but were undervalued by the rest of the MLB. In doing so, the A's were able to compete with teams with over double the payroll that they had. The NHL is no different. With an increasing use of data analytics, NHL teams are competing to find the perfect roster while obeying league cap space rules.

Keywords: Corsi, Fenwick, Cap Space, linear regression

1. Introduction

The National Hockey League is one of the largest sports leagues in the United States, with a revenue of almost three billion dollars (Jones, 2023). However, nearly half of the teams in the NHL have negative operating revenues (Badenhausen, & Ozanian, 2020). All teams must endure the rising cost of player salaries. In traditional belief, higher-paid players should deliver better performance. In every sport, athletes go overlooked; identifying players who are currently underpaid but performing at high levels should be the target of all teams to decrease costs. Another factor teams must consider is Cap Space. Cap Space is the amount of money each team is allowed every season and changes yearly based on the overall NHL revenue. Teams must delicately balance these two factors to save money while still creating a championship-caliber team on the ice.

In an ever more competitive league, teams have invested heavily in statistical analysis of players in order to put the best possible players on the ice. Both teams and the NHL have begun to track every aspect of the game down to real-time puck movement on the ice. Having good players is critical to a team's success and finding one is a difficult task required by all general managers. Getting one for cheap is even more difficult. Analyzing NHL goalie and defensemen data from the last five seasons and controlling for team effects created a linear regression predicting win percentage. From the model, key variables were identified and used for a second linear regression model with salary as the response. Using both equations from the models, salary, and win percentage could be predicted based on a goalie and defenseman performance. The best goalies and defensemen in the NHL could be identified using these two equations.

This paper will cover key governing factors within the NHL Salary Cap, leading to a literature review of modern and historical data analysis that teams use to evaluate player talent. The methodology and results will explain how the player analysis was conducted using linear regression models to identify undervalued players. Finally, the conclusion will identify specific players teams should target to capitalize on undervalued but high-performing players.

2. Background

The goal of every hockey general manager is to produce a Stanley Cup-caliber team with the budget and players they have. Hockey has experienced an analytical overhaul within the last decade. Teams now track everything down to puck and player locations for every second of the game. Like any sport, it's impossible to definitvely predict what the future may have, specifically at the player level. This creates a challenge for building teams, as prior performance may not reflect future performance.

2.1 Salary Cap

A salary cap is a way to level the playing field between different franchises. Many professional leagues institute these regulations to prevent teams with larger markets from dominating year to year as they can afford better players while smaller market teams are left taking risks on unproven athletes. The disparity between the top-earning team and the bottom-earning team is drastic. The Arizona Coyotes were last in valuation in 2020 at 285 million dollars, while the New York Rangers were valued at 1.65 billion dollars, a difference of 1.365 billion dollars (Badenhausen, & Ozanian, 2020). The salary cap was established in the NHL shutdown of 2005. The salary cap is variable year to year as it is not a fixed number but rather a percentage of the overall National Hockey League revenue. Since the implementation of the salary cap in 2005, the cap has increased from thirty-nine million dollars to 83.5 million in 2023. The salary cap is a hard cap which means teams don't pay a fine for going over the cap limit but rather must play with a smaller roster so that they make the salary cap (Murphy, 2023). Another rule within the salary cap is no individual player can exceed 20% of the team's salary cap (Vollman, 2016). Although it is not as straightforward as one may assume. It's the cap hit that must remain below that 20% threshold. A cap hit is calculated by the average salary over the length of the contract. A strategy that can be employed is adding years of low-paying season to bring the average down, keeping it below the cap hit. Another rule for NHL player contracts is that no player can be signed for more than seven years, or eight, if staying with the same team. A key strategy teams can use to keep their salary cap down is with ELCs which stand for entry-level-contracts "rookie contracts" (Vollman, 2016). These Entry Level Contracts are maxed at $950,000 and have a signing bonus cap of 10% of the initial salary, making ELCs the least consuming contracts for a team's cap space (News Room, 2023). Finally, a team must pay every player at least the league minimum, which, as of the 22-23 season, is $750,000 a year (Fleming, 2022). Although a team cap space is hard set at the discretion of the NHL, a team can use the strategies just stated, as well as techniques known as cap relief, to free up cap space. Uniquely enough, the salary cap disappears in the postseason, which has caused some controversy after Tampa Bay Lighting placed a player with a cap hit of 9.5 million dollars on the injured reserve, pulled him up for the postseason, and went on to win the Stanley Cup (Murphy, 2023).

2.2 Cap Relief

When it comes to cap space, teams must be as effective as they can with their money. Cap relief allows teams to remove some of their costs, enabling them to sign free agents or give bonuses to players performing well. Cap relief becomes crucial when teams find themselves stuck with a player with a large and long contract, yet the player is not meeting expectations on the ice. There are several ways teams can work around this issue. The first is to send the player down to the minors. Sending down players who are underperforming frees up some cap space and a roster spot to bring in someone new. Since rule changes in 2013, the issue with sending a player to the minors set the cap relief at $375,000 above the league minimum for players on one-way contracts, which is the case for most players (Vollman, 2016). In addition, age becomes a factor as players above the age of 35 who are sent down provide no cap relief at all. Finally, sending a player down can be risky as it allows other teams to purchase their contract for little in return from the team who sent them down (Vollman, 2016). The second-way teams can reduce their cap space is by sending players to the injured reserve. A player who is put on the injured reserve still contributes to the cap space, but their replacement does not (Murphy, 2023). The third way to gain cap space is by trades (Vollman, 2016). Trading away players who are underperforming can be a difficult task. In order to make the offer more appetizing, teams will retain part of a player's cap. Using trades can turn cap space into an asset for teams. Another way teams will free up cap space is by buying out a player's contract. The benefit to teams buying out a player's contract will still require the team to carry two-thirds of the deal and they can spread it out over double the contract length making the annual cap hit almost negligible (Vollman, 2016). The last way a team can relieve their cap space is when a player retires. The one notable exception to this is players over the age of 35, when they signed their contract, will still have the players cap on their sheets (Vollman, 2016).

3. Literature Review

The sport of Hockey has evolved over the years. Similarly, hockey analytics have changed. The first analytical methods were simple but allowed teams to evaluate individuals as part of the whole. Teams quickly identified the issue with their current method of evaluating players, which lacked an appropriate number of observations, which could skew the analysis. As a result, modern-day analytics developed the Corsi and Fenwick analyses, which include shots on goal rather than points scored like its predecessor plus-minus.

3.1 Traditional Hockey Analytics

Hockey analytics can trace their early roots to basic statistical measurements, with plus/minus created in the 1950s by the Montreal Canadians (Vollman, 2016). Plus/minus is a very simple statistic that measures the net goals for and against a player while they are on the ice (Strauss, 2023). An example of how plus/minus works is if player x is on the ice for a game, and while on the ice, his team scores 3 goals and gives up one, his plus/minus for the game will be +2. The problem with plus/minus is that a player can have very little to do with a goal but be credited or debited for just being on the ice. Those who support plus/minus statistics argue that over the course of a season, a player's involvement or noninvolvement in goal will average out (Vollman, 2016).

$$plus|minus = goals\ for * 1 + goals\ aginst * (-1) \quad (1)$$

The argument against plus/minus is extensive. The first fault with the statistic is that shorthanded goals are included in the calculation (Vollman, 2016). A shorthanded goal or powerplay is when one team has a player in the penalty box due to a foul and is playing with one less athlete on the ice than the other team. Playing a man down gives a massive advantage to the team that is up. The second issue with plus/minus is how much the team factor plays into an individual statistic. Playing for a great or terrible team can heavily sway an individual's plus/minus, which may lead one to think that the individual is either great or terrible when it's his team that is good or bad (Vollman, 2016). The third reason why plus/minus is bad depends on the puck's location in the zone when a player comes on (Vollman, 2016). If a player comes onto the ice primarily when the puck is in their team's zone, they are more likely to have a negative plus/minus than those who come on when the puck is in the opposing team's zone. Goaltenders can also influence a player's plus/minus (Vollman, 2016). Teams with great goaltenders will have better plus/minus. Even if they are a liability on the ice, this is also in reverse. The small sample size is the final reason why plus/minus is not a great analytic. Goals in hockey don't come easy or often lead to very few goals, leading to limited observations that can sway plus/minus drastically (Strauss, 2023). The lack of observations would give way to modern hockey analytics such as Corsi and Fenwick.

3.2 Modern Hockey Analytics

With the faults of plus/minus, new analytic methods were developed. Two of these measurements are Corsi and Fenwick. Corsi came around to solve the main issue of plus/minus, which is the limited number of observations but follows a very similar equation to that of plus/minus.

$$Corsi = shot\ attempts\ for - shot\ attempts\ against \quad (2)$$

A shot attempt is classified as any shot on goal despite the result, whether it be goal, blocked, or missed. Corsi not only solves the issue of limited observations but also removes the variability of the goaltender as all shot attempts are considered. Corsi is excellent for predicting the time of possession (Vollman, 2016). Like in many other sports, the team that controls the ball or puck for most of the game or match usually scores more. A team with a high Corsi score is on the offense and taking more shots on the net than their opponents. Corsi can be displayed in multiple forms, including Corsi per 60 minutes and a percentage, which is shown below.

$$Corsi\ \% = \frac{shot\ attempts\ for}{shot\ attempts\ for + shot\ attempts\ against} \quad (3)$$

Fenwick is similar to Corsi but does not count blocked shots in its calculation. Similar to Corsi, Fenwick is useful for data analysis as it allows for a higher number of observations as goals are a rare occurrence in the sport, and with limited observations, a few goals can have a large impact on the data and may not fully explain how a team or player is performing.

$$Fenwick = (Shots\ on\ goal\ FOR - blocked\ shots) - (Shots\ on\ gaol\ AGAINST - blocked\ shots) \quad (4)$$

3.3 Goaltenders

Goaltenders are so special they deserve their own section. Many of the statistics we use to evaluate the rest of the players on the ice do little to evaluate a goaltender. A quick look: one may assume that save percentage is a good place to evaluate the position, but they would be half correct. One factor many don't consider when evaluating a goaltender is luck

(Vollman, 2018). Goaltender consistency on a year-to-year basis is so random that it is impossible to draw any tendencies or patterns. This is because goals are such a rare phenomenon that a handful of goals can sway a goalie's save percentage enough on a year-to-year basis, making it impossible to predict future performance (Vollman, 2016). Looking at goaltenders' career even-strength save percentages on a 95% confidence interval, almost all goaltenders overlap within the 95% confidence interval, making it statistically impossible to distinguish one from another as to who is the best goaltender without more advanced metrics. Another key factor is the team a goaltender is on. A team that is terrible is more likely to allow higher-quality shots than a good team, which forces teams to take poor-quality shots (Vollman, 2016). One may also assume that if a goaltender is on a bad team, he would see more shots on goal, affecting his save percentage, but there is very little evidence to support this (Vollman, 2016).

4. Methodology and Results

Utilizing data from the NHL and CapFriendly and splicing player performance with their salaries from the 2018-2019 season to the 2022-2023 season, the data was summarized, and using MiniTab, a linear regression was created to identify key variables, with win percentage as the response. The key variables identified were then selected to be utilized in the salary linear regression to calculate a player's true worth. These two linear regressions provided the results of the most overpaid/underpaid players and win percentage differences.

4.1 Model Development

The NHL tracks multiple metrics, including Age, Games Played, Win %, Shutouts, Goals against average, Save %, Goals allowed per 60 minutes of play, expected goals allowed per 60 minutes of play, and expected goals saved per 60 minutes of play. The model took these statistics over the last five seasons with a minimum game played every season of 25 games. The model then took those statistics by player and made a stat sheet for their last five seasons in the NHL, along with their average salary over those five years. Since hockey is a very dynamic sport and heavily team-based, the model included a team effect metric to remove the variability that a team may have on a goalie's performance. This team metric is the Corsi %, as detailed. earlier. Since goalies can be traded or moved teams, the model calculated season Corsi % from the 2018-2023 seasons and made an average Corsi % for those five years. If a goalie played for multiple teams in the five-season span, the model average the Corsi % to give each goalie their team associated Corsi % in the five-season span. Using MiniTab, the model created a linear regression with win % as the response as the goal for any team is to win games, and winning games leads to playoff berths and potential Stanley Cup Champions. Using the linear regression, the model was able to identify the key variables associated with goalie win % with a linear regression with a 59.13% r2 and test r2 of 67.06% when using 30% of the data as a test. The key variables identified were Age, Games Played, Goals against average, Save %, Goals allowed per 60 minutes of play, expected goals allowed per 60 minutes of play, and expected goals saved per 60 minutes of play. For defensemen, the analyzed variables were Hits, Blocked shots, takeaways, points, and assists. The r2 for this model was 43.9%.

$$\begin{aligned}\text{Average of W\%} = {} & 5.56 - 0.00296 \text{ Average of AGE} + 0.000287 \text{ Sum of GP} - 0.00674 \text{ Sum of GAA} \\ & - 4.84 \text{ Average of Sv\%} + 3.34 \text{ Average of GA60} + 3.66 \text{ Average of GSAx60} \\ & - 3.64 \text{ Average of xGA60} + 0.00719 \text{ Average of Corsi \%}\end{aligned} \tag{5}$$

Using the identified key variables, the model response to Salary. The r^2 for the goalie linear regression was 55.73%, and the test r^2 was 47.5%. While for defensemen was 2.91%. These models are not very good at predicting salary for goalies or defensemen. This means that there are players who are being overpaid and underpaid based on their ability to help their teams win games. Using the linear regression equation, the model then can calculate the true salary a player should make and identify if they are being overpaid or underpaid. Then, subtracting their current salary from their true salary, the model identified overpaid and underpaid players based on their ability to help their teams win. Players who were underpaid would have a positive difference in how much they were underpaid, while players who were overpaid would have a negative amount displaying how much they were overpaid.

$$\begin{aligned}\text{Average of SALARY} = \; & -64495508 + 231406 \text{ Average of AGE} + 37447 \text{ Sum of GP} - 357851 \text{ Sum of GAA} \\ & + 57861780 \text{ Average of Sv\%} - 39318527 \text{ Average of GA60} \\ & - 41870280 \text{ Average of GSAx60} + 40377986 \text{ Average of xGA60} \\ & + 76460 \text{ Average of Corsi \%}\end{aligned} \tag{6}$$

To validate the model the model had to be tested against the four linear regression assumptions: Linear relationship, Independence, Homoscedasticity, and Normality. Testing the independence, a Durbin-Watson test was conducted with a correlation not likely as the value was 1.8, and a value of 2 indicates no autocorrelation. Values between 1.5 and 2.5 are not a concern of correlation. To test for homoscedasticity, a fitted value vs residual plot was created and shows small signs of non-linearity; however, it is more aligned with no problem. The data proved to follow a roughly normal distribution.

4.2 Results

Using equation six above, the model was able to calculate the true salary a player should be paid for their performance on the ice. Taking a player's current salary and subtracting the predicted salary from it gave us the dollar amount a goalie or defenseman is being overpaid or underpaid. Out of the 90 goalies used, 53 were underpaid, 37 were overpaid, and 70 of 150 defensemen were underpaid.

Table 1: Top Overpaid and Underpaid Goalies

Rank	Most Under Paid	Amount Under Paid	Most Over Paid	Amount Over Paid
1	Jacob Markström	$ 2,578,020.59	Carey Price	$ 8,059,308.35
2	Juuse Saros	$ 2,236,361.86	Sergei Bobrovsky	$ 4,318,602.80
3	Mike Smith	$ 2,093,614.10	Henrik Lundqvist	$ 3,179,753.79
4	Darcy Kuemper	$ 2,032,299.21	Cory Schneider	$ 3,165,489.80
5	Curtis McElhinney	$ 2,017,177.70	Tuukka Rask	$ 2,394,554.88

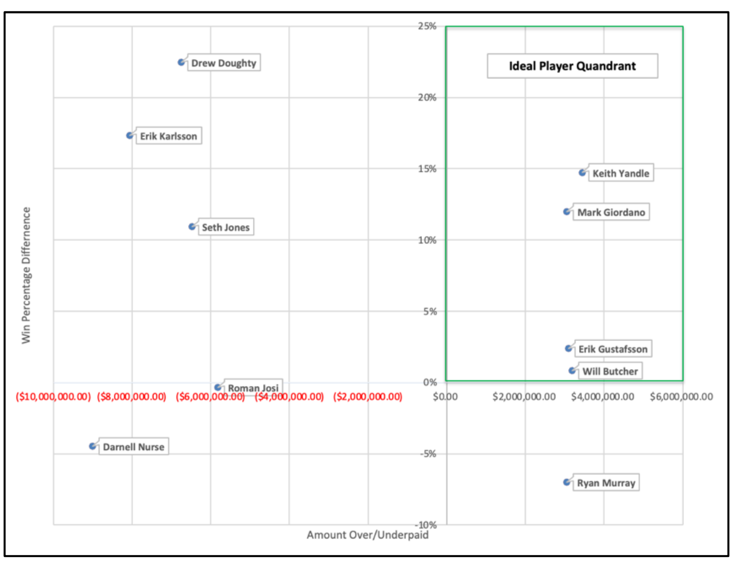

Figure 1. Top Overpaid/Underpaid vs Win Percentage Differences - Defensive Players

ISBN: 97819384962-4-0

In addition to predicting salary, we also predicted win percentage using equation five above. Using the predicted win percentage and subtracting their current win percentage, we found a goalie's win percentage difference. 40 out of the 90 goalies had a positive win percentage, while 50 had a negative win percentage. 78 of the 150 defenders had a positive win percentage, nine had no difference, and 62 had a lower predicted win percentage.

Table 2: Highest Predicted Win Percentage (Goalies)

Rank	Player	Current win percentage	Predicted win percentage	Win Percentage Difference
1	Jeremy Swayman	71%	73%	2%
2	Daniel Vladar	70%	72%	2%
3	Filip Gustavsson	71%	71%	0%
4	Tuukka Rask	73%	69%	-4%
5	Pavel Francouz	75%	68%	-7%

Table 3: Highest Predicted Win Percentage (Defensemen)

Rank	Most Under Paid	Current Win%	Predicted Win%	Win% Difference
1	Rasmus Andersson	58%	78%	19%
2	Brent Burns	63%	74%	11%
3	Noah Hanifin	70%	73%	2%
4	Jeff Petry	50%	71%	21%
5	Brett Pesce	61%	70%	9%

However, targeting purely underpaid players would not be the best strategy possible as when calculating predicted win percentage, some underpaid players returned a negative win percentage difference between their current win percentage and predicted win percentage. Instead, the players you would want to target would be players who are underpaid and return a high win percentage difference. Tables 4 and 5 identify the best five players who are underpaid and return the highest win percentage. Analyzing team win percentage over those five years, the highest win percentage was 65%, and 18 of the 32 teams had a win percentage at or above 50%. Adding players with a predicted win percentage greater than the team should lead to more wins for a team in turn increasing the chances of a championship.

Table 4: Best Goalies Available

Rank	Player	Underpaid	Win Percentage
1	Jeremy Swayman	$ 221,817.66	73%
2	Filip Gustavsson	$ 42,221.46	71%
3	Pavel Francouz	$ 1,023,760.61	68%
4	Jake Oettinger	$ 1,387,966.57	67%
5	Jaroslav Halák	$ 944,332.47	66%

ISBN: 97819384962-4-0

Table 5: Best Defensemen Available

Rank	Player	Underpaid	Win Percentage
1	Radko Gudas	$559,286.83	69%
2	Jon Merrill	$2,581,861.76	68%
3	Matt Roy	$395,950.65	67%
4	Colin Miller	$1,779,890.36	67%
5	Mike Reilly	$1,022,117.19	67%

5. Conclusion

The two goals of every NHL team are to win and turn a profit. At the end of the day, the NHL is a business. Using the two models developed, a general manager would be able to identify players who are undervalued and have a predicted positive win percentage impact on the ice. Teams could also use this model when negotiating contracts with players who are calculated as overpaid to restructure their contracts, which would, in turn, increase available cap space for other players or decrease overall cost. Out of the 90 goalies analyzed over the last five seasons, 19 were undervalued and had a positive win percentage impact. Out of the 150 defensemen analyzed 34 were undervalued and had a positive win percentage. Teams who are attempting a playoff run may not care about money and purely about wins and want the goalies who will significantly increase their win percentage. There are three goalies who have a win percentage increase of over 10% and 24 defensemen with a win percentage greater than 10%. This would drastically increase a team's chances of winning any given game. However, in the long run, it is unsustainable to do this, as two of those three goalies were calculated as overpaid, and 12 of the 24 defensemen were overpaid. When NHL teams are building their teams in the offseason, targeting underpaid players will increase the available cap space a team can work with, allowing them versatility for going after other position players. The best goalies currently available after analyzing the last five NHL seasons are Jeremy Swayman, Filip Gustavsson, Pavel Francouz, Jake Oettinger, and Jarislav Halák. The best defensemen available are Gustav Forsling, Greg Pateryn, Derrick Pouliot, Radim Simek, and Keith Yandle.

This project proposes a new way to calculate player salaries based on their ability to help their team win. By doing so, NHL teams can identify players who are underpaid but will have a positive impact on their team’s performance while also saving the team money. Further research is necessary to identify forwards who are undervalued and could contribute significantly to a team's performance, in turn increasing the team’s win percentage.

6. References

Fleming, Connor. "How Do The Minimum MLS And NWSL Salaries Compare To Other U.S. Sports?" *The 18*, 28 Feb. 2022, the18.com/soccer-entertainment/minimum-salary-in-pro-sports. Accessed 23 Oct. 2023.

Herman, Jeff. "Better Know a Statistic: Corsi & Fenwick." *The Hockey Writers*, 4 Nov. 2014, thehockeywriters.com/corsi-fenwick-stats-what-are-they/. Accessed 23 Oct. 2023.

Jones, Wayne. "What Is the Yearly Revenue of the NHL? (plus, Where It Comes from)." *Hockey Answered*, 1 Jan. 2023, hockeyanswered.com/what-is-the-yearly-revenue-of-the-nhl/. Accessed 23 Oct. 2023.

Ozanian, Mike, and Kurt Badenhausen. "NHL Team Values 2020: Hockey’S First Decline In Two Decades." *Forbes*, 9 Dec. 2020, www.forbes.com/sites/mikeozanian/2020/12/09/nhl-team-values-2020-hockeys-first-decline-in-two-decades/?sh=410a083970dd. Accessed 23 Oct. 2023.

Room, News. "NHL Entry-level Contract, Explained: How Much Can Rookies Make on Their First NHL Deal?" *Grand Athlete*, 28 Jun. 2023, grandathlete.com/ice-hockey/nhl-entry-level-contract-explained-how-much-can-rookies-make-on-their-first-nhl-deal/. Accessed 23 Oct. 2023.

Strauss, Travis. "The Ultimate Guide to Understanding Plus-Minus in Hockey." *Ice Hockey Central*, 19 Feb. 2023, icehockeycentral.com/the-ultimate-guide-to-understanding-plus-minus-in-hockey/. Accessed 23 Oct. 2023.

Vollman, Rob . *Stat Shot a Fan's Guide to Hockey Analytics*. *ECW Press*, 2018.

Vollman, Rob . *Stat Shot a Fan's Guide to Hockey Analytics*. *ECW Press*, 2016.