Composite Tailored Regression Modeling For Evaluative Ratings in Professional Hockey

John Doe

June 16, 2016

Abstract

Evaluative methods in ice hockey alternative to traditional scouting are relatively new and fairly flawed. Basic statistics such as Plus-Minus and Points offer little insight, while more recently developed metrics and statistical models provide an improved albeit incomplete summary of team or player quality. I posit that each of the processes that yield goals, and thereby influence game outcomes, may be modeled statistically and propose deriving a total rating from the summation of an agent's influence on each of these processes. Non-overlapping partitioning of these processes and tailored selection of model specifications ensure that each aspect is diligently and optimally accounted for. Conversion to a common currency for aggregation is trivial as it is implied in the selection of processes that directly or otherwise impact goals. I test the validity of this model and explore practical applications.

Acknowledgements

I would foremost like to thank Andrew Thomas, Sam Ventura and Alexandra Mandrycky, without the patience and helpfulness of whom I would not have developed the tools necessary for the undertaking of such a project. Additionally, I thank Micah McCurdy, Prasad Patil and Michael Lopez for their guidance along the way.

Introduction

1.1 Purpose

I aim to derive a complete rating for agents[1] participating in hockey games that is both descriptive of total impact and actionable in predicting future outcomes. A success in this venture would involve one or more applications that are preferable to the common alternative. Such applications may include evaluation of talent, predicting game outcomes, gambling or storytelling.

1.2 History

The advent of improved, more complete game records in the NHL spurred the development of new statistics and performance metrics. In the years preceding this era, Plus-Minus (since 1967), Goals and Points were standard tools used for player evaluation. The Plus-Minus statistic is defined by the differential of even-strength or shorthanded goals scored for and against a player's team while the player is on the ice. A major limitation of this measure is the relative infrequency of goals. Advances in the league's event recordings in 2007 led to variants of Plus-Minus making use of shots rather than goals. Corsi refers to all shots attempted, while Fenwick denotes attempted shots that are unblocked. These may be expressed as differentials or percentages, typically in 5v5 situations, and represent a significant improvement over Plus-Minus due to the increased size of event samples. They are nonetheless rudimentary, and fail to account for numerous confounding factors such as quality of teammates or opponents, game states or quality of shots. Attempts to adjust these metrics in response to such factors have been made, notable examples of

[1] "Agent" will henceforth describe any entity participating in a hockey game by affecting its outcome and encompasses teams, players and coaches.

which are dCorsi (Burtch 2014), Corsi Plus-Minus (Sprigings 2015) and Score-Adjusted Corsi/Fenwick (Tulsky 2012, McCurdy 2014). These enhancements have varying validity and often account only for a portion of the contextual factors. More sophisticated statistical models have come into existence in recent years with promising results. In particular, Goals Versus Threshold (GvT) (Awad 2009), Total Hockey Rating (THoR) (Schuckers 2013), Adjusted Plus-Minus (MacDonald 2011) and WAR (Thomas, Ventura 2015).

1.3 Data

1.3.1 Source

The data used were originally compiled and stored in the *Corsica* database. The Play-By-Play is sourced from official NHL game records dating back to the 2007-2008 regular season while event coordinates are sourced from the Sportsnet and ESPN websites. Salary information was provided by WAR On Ice. Other leagues were excluded from the analysis, as the required data do not exist publicly. Additional cleansing and processing of these data was performed after the initial collection process. Namely, I make use of the methods outlined by Krzywicki for regularization of shot coordinates and the calculation of absolute shot angle from the central line normal to the goal line (Krzywicki 2010). Subsequently, shots are categorized according to custom specifications. Rebound shots are defined as those occurring within two seconds of uninterrupted game time of a prior unblocked shot. Rush shots are defined as those occurring within four seconds of uninterrupted game time of any event having occurred in the shooting team's defensive zone or any giveaway or takeaway event.

1.3.2 Expected Goals

I make use of the *Corsica* shot quality measure introduced in the 2016 article "Shot Quality and Expected Goals: Part 1" (Perry 2016). A given shot's Expected Goal (xG) value is equal to its estimated probability of resulting in a goal. This likelihood estimate is produced using a logistic regression model trained on 7 seasons of shot data. Blocked shots are omitted due to coordinate missingness. The total training sample consisted of 542,569 unblocked shots from the 2007-2008 to 2013-2014 seasons. The fitted model was tested against shots from the 2014-2015 season and parameters were selected in order to optimize out-of-sample predictive validity. The final regression contains 6 variables:

$$\gamma \equiv Shot\ type$$

$$\delta \equiv Distance$$

$$\alpha \equiv Angle$$

$$\rho \equiv Rebound\ *$$

$$\sigma \equiv Rush\ *$$

$$\theta \equiv Neutral\ zone\ *$$

$$*\ Boolean$$

And obeys an equation of the form:

[1] $$P_i = \begin{cases} \text{if } \theta = 1, & 0.00648 \\ \text{if } \theta = 0, & \frac{1}{1+e^{-t_{(i)}}} \end{cases}$$

Where t is the function:

[2] $$t_{(i)} = \beta_{\gamma 0} + \beta_{\gamma 1}\delta_i^3 + \beta_{\gamma 2}\delta_i^2 + \beta_{\gamma 3}\delta_i + \beta_{\gamma 4}\alpha_i^3 + \beta_{\gamma 5}\alpha_i^2 + \beta_{\gamma 6}\alpha_i + \beta_{\gamma 7}\rho_i + \beta_{\gamma 8}\sigma_i$$

And P_i gives the probability of a shot i becoming a goal given the independent variables aforementioned. The Booleans take on a value of 1 if the condition is true and 0 otherwise. Thus, shots taken from the neutral zone are assigned a flat xG value equal their average goal probability in the training data subset. The intercept β_0 and coefficients $\beta_1 \cdots \beta_8$ are unique to each shot type γ.

Figure 1: Each bin represents a 2.5% interval, where the marker size is proportional to number of observations. Bins totaling fewer than 100 observations were discounted. The data used here are not present in the training subset.

The validity of the model was verified out of sample using a series of tests. One such test involved iterating a process of selecting a random sample of 1,000 shots and comparing the predicted and observed goal percentage. The R^2, P-value and Residual Standard Error were recorded for each of 300 iterations and averaged. The parameters selected for the final regression yielded optimal results in this test. Additionally, the correlation displayed between predicted and observed Fenwick Shooting Percentage (FSh%) of xG bins in figure 1 confirms the model's ability to provide actionable insight. The density of shots from the 2015-2016 regular season belonging to various xG bins obeys an intuitive pattern as demonstrated in figure 2:

Figure 2: Heat map of 2015-2016 shots coloured by xG levels.

1.4 Desired Properties

Before the construction of the model, I take stock of properties an optimal rating system should possess. The intention is to remain mindful of these qualities and attempt to satisfy as many conditions as is possible while prioritizing the total validity of the model.

- **Practicality**

  Vitally, an informative evaluation method must have practical applications. More strictly, it must represent an advantage over common alternatives sizeable enough to justify its relative cost or complexity. As previously discussed, these applications may include, but are certainly not limited to, prediction or player talent evaluation. In different terms, one should be able to apply a successful model to inform fruitful decisions in the context of hockey operations or competition such as gambling.

- **Flexibility**

  The proposed model must possess the property of being applicable to the end of evaluating varieties of agents. Most importantly, an optimal rating system should function in identical fashion and be equivalently interpretable for both teams and players. Further, the quality of being component-based is sought. An example of such is WAR (Thomas, Ventura 2015), in which the overall rating is equal to the sum of multiple components representing either an offensive or defensive contribution. Use in hockey leagues other than the NHL is of secondary concern, as data limitations represent an equally difficult hurdle for all sophisticated statistical models.

- **Scalability**

  A drawback of regression-based models is that they often require large sets of data. Corsi, by comparison, involves nothing more than addition and division of event counts. Thus, it can easily scale between multiple seasons, games or even individual shifts by simple aggregation. The ability of a rating system to represent single-game performance is contingent on this property. Computational requirements may also raise concerns in the case of particularly complex models.

- **Reproducibility**

  Others should be able to reproduce the results obtained regardless of software or computational means. Theoretically, the method should also be applicable to any hockey league provided the necessary data are publicly available.

- **Interpretability**

  Accessibility is dependent on the ease of interpretability of information yielded by the model. Despite the desire for a component-based product, I expect that ratings produced by any successful model should exist as single quantities. The nature of these ratings should tap into one’s ability to intuit quantities in familiar formats. In accordance to my own preferences, ratings below a given baseline should exist as negative quanta, where the preferred baseline is league average.

Model

2.1 Overview

What will henceforth be known as the grand model is an approximation of goals added resulting from the combination of the influences of an agent on each of four unique processes. At minimum, there are four distinctly separate processes in hockey by which goals are directly or otherwise produced. The grand model is built on the assertion that each of these processes can themselves be modeled with agents serving as explanatory variables. An agent can impact the occurrence of goals by influencing:

A. The rate of shots for or against a team

B. The likelihood of a shot becoming a goal

C. The rate of penalties taken for or against a team

D. The game state or context of imminent play

All varieties of agents are not responsible for every process necessarily. For instance, I begin with the assumption that goaltenders do not possess the ability to alter the rate of shots. They do, however, affect the goal probability of shots taken against them. The involvement of non-goalie agents in the course of play may positively or negatively impact the baseline rate at which shots accumulate against teams involved in a game. We may model this process and an agent's influence thereon while controlling for factors that overlap with other processes or do not serve to measure the quality of agents under observation.

Figure 3: Diagram of the four main processes and their products.

This first process is approached as a survival problem[2], using a Cox proportional hazard regression. This method was first employed in 2013 to model the rate of goals and assign team and player ratings (Thomas et al. 2013). Here, an attempted shot serves as a positive observation. Separately, shooting efficiency is modeled using a logistic regression with agents once again acting as explanatory variables. Thus, the quantity and quality of shots are modeled separately with no overlap. This principle of segregation in tandem with the tailored selection of optimal model specifications for each process comprise the main tenets of the grand model.

2.2 Specifications

2.2.1 Shot Rates

I use a right-censored Cox proportional hazard regression to model the rate of shots by the home team and away team independently. The shot production process is censored by any player substitution, occurring roughly every 11.5 seconds on average. The predictors for both processes are comprised of situational indicators such as the state of the game with respect to the quantity of skaters on the ice for either team as well as all agents of the

---

[2] The name originates from applications in biology and medicine where the binomial response variable takes the value 1 if a death is observed and 0 otherwise.

type under observation. The equation for the rate of shots by the home team may be written as:

$$[3] \qquad \rho_h = e^{\alpha_h + \sum_p X_p^h \Omega_p + X_p^a \delta_p + \sum_s X_s \sigma_s}$$

Where α_h is the baseline home rate in shots per minute, X_p^h and X_p^a are the indicators for each agent on the side of the home and away teams respectively, Ω_p and δ_p are coefficients representing offensive and defensive contributions respectively for each agent *p*, X_s are indicators for each state in S and σ_s is the corresponding coefficient of that state. S is a vector of length 24, 9 elements of which pertain to skater strength, 7 of which pertain to the game score and 8 of which pertain to the recency of the last face-off and in which zone it occurred. We may then write:

$$[4] \qquad K_{\text{SR FOR}}(p) = (e^{\Omega_p^h} - 1)\alpha_h \psi_h \text{TOI}_p^h + (e^{\Omega_p^a} - 1)\alpha_a \psi_a \text{TOI}_p^a$$

$$[5] \qquad K_{\text{SR AGAINST}}(p) = (1 - e^{\delta_p^h})\alpha_a \psi_a \text{TOI}_p^h + (1 - e^{\delta_p^a})\alpha_h \psi_h \text{TOI}_p^a$$

Giving the goals added on offence [4] and defence [5] by a predictor agent *p*, where ψ_h and ψ_a are the average goal values of home and away shots respectively and TOI_p^h and TOI_p^a is the Time On Ice of the agent *p* in home and away games respectively. The net goals added by an agent *p* is then $K_{\text{SR FOR}}(p) + K_{\text{SR AGAINST}}(p)$. A property of equation [3] is that the exponents of the offensive and defensive parameters $e^{\Omega}$and e^{δ} are interpretable as multipliers on the baseline shot rates.

2.2.2 Shot Quality

Goals added by an agent's influence on shot quantity is obtained by employing the average success of shots as a conversion rate. Independently, an agent may influence the quality of shots, or the probability that they will result in a goal. I model this process using a logistic regression allowing for multiple roles by each agent. Namely, in the case of skaters, agents may influence outcomes by acting as the shooter or by being involved in a supporting role – on the ice as a non-shooter.

Figure 4: Matrix of possible roles to be assumed by agents acting on the likelihood of goals.

As with shot quantity, the impacts of various game states are controlled for by their inclusion as regression parameters and the exponents of the coefficients given by the logistic model are interpretable as multipliers, now on the baseline goal odds. This derivation is quite straightforward. Consider the logistic function:

[6] $$P_i = \frac{1}{1 + e^{-t(i)}}$$

Where t is the linear function:

[7] $$t_{(i)} = \beta_0 + \sum_p X_p^i \beta_p$$

P_i gives the goal probability of a shot *i*, thus:

[8] $$\frac{P_i}{1 - P_i} = e^{t_{(i)}} = e^{\beta_0 + \sum_p X_p^i \beta_p}$$

Is equal to the odds of a goal for shot *i.* It follows that e^{β_p} multiplies the base odds e^{β_0}. We obtain goals added with the formulas:

[9] $$K_{\text{SHOOTER}}(p) = \frac{(e^{\beta_p^{sh}} - 1)e^{\beta_0}}{(1 + (e^{\beta_p^{sh}} - 1)e^{\beta_0})} \text{iCF}_p$$

[10] $$K_{\text{SUPPORT FOR}}(p) = \frac{(e^{\beta_p^{sf}} - 1)e^{\beta_0}}{(1 + (e^{\beta_p^{sf}} - 1)e^{\beta_0})} \text{TCF}_p$$

[11] $$K_{\text{SUPPORT AGAINST}}(p) = \frac{(1 - e^{\beta_p^{sa}})e^{\beta_0}}{(1 + (1 - e^{\beta_p^{sa}})e^{\beta_0})} \text{CA}_p$$

[12] $$K_{\text{GOALIE}}(p) = \frac{(1 - e^{\beta_p^{g}})e^{\beta_0}}{(1 + (1 - e^{\beta_p^{g}})e^{\beta_0})} \text{CA}_p$$

The sum of which gives the net goals added by an agent. iCF, TCF and CA are the individual shots taken, shots taken by teammates while on the ice and shots taken against while on the ice. It should be noted that the starting assumption is skaters can act only as shooters or support, while teams act uniquely as shooter or goalie.

2.2.3 Penalty Rates

The rate of penalties is modeled similarly to the rate of shots. The same survival method is applied here with minor penalties serving as positive observations. Major penalties (resulting in 5 minutes of penalty time) are omitted due to infrequency in order to mitigate complexity, while double-minor penalties were treated as two distinct observations. The coefficients produced by the Cox regression are equivalently interpretable to those in section 2.2.1 with the distinction that, by virtue of multiplying the rate of penalties in place of shots, they are inversely proportional in terms of value added. This fact is addressed in the conversion to goals:

[13] $$K_{\text{P TAKEN}}(p) = (1 - e^{\Omega_p^h})\alpha_h \psi \text{TOI}_p^h + (1 - e^{\Omega_p^a})\alpha_a \psi \text{TOI}_p^a$$

[14] $$K_{\text{P DRAWN}}(p) = (e^{\delta_p^h} - 1)\alpha_a \psi \text{TOI}_p^h + (e^{\delta_p^a} - 1)\alpha_h \psi \text{TOI}_p^a$$

Where each variable is analogous to those in equations [4] and [5], and ψ represents the goal value of a minor penalty, given by the average net goal differential over the duration of powerplay (PP) time resulting from minor penalties. The sum of both values gives the net goals added by an agent.

2.2.4 States

Least intuitively, and partially an artifact of the methodology used heretofore, agents may impact the occurrence of goals in a fourth way. Recall that we've controlled for various game situations in each of the three prior regressions. This precaution serves to adjust for the effect of usage or deployment on skaters – in effect, leveling the playing field. However, it ignores the real possibility that skaters themselves may influence these game states. In actuality, the process described in 2.2.3 is a sub-category of this mechanism. It describes how agents impact the advantage in skaters that may occur over the course of play. A second set of circumstances we've chosen to account for in each process is the score in the game. Agents exert an effect on this state by contributing to goals scored. Thus, the impact is already implicit in the value of goals added – the very product sought in this model. This leaves zone starts. They describe which of the three zones – offensive, defensive or neutral – a face-off to begin a shift occurred in. The value of these zone starts has been stripped from each goals added component yet produced but they may be separately dealt with. I use a multinomial logistic regression to model each possible ensuing zone start succeeding a shift end. Agents act as predictors alongside the various game state elements, including shift starts. Skaters do not earn additional value for creating their own zone starts. Each face-off, with the exception of those following an icing infraction, represents a decision point at which a coach may choose to substitute players. Hence, so-called "earned" zone starts may still indicate preferential treatment. Rather, skaters are attributed the partial goal value of zone starts created for teammates entering play in their place. The multinomial regression yields the following equation:

[15] $$P_{(\mathrm{Y_i=c})} = \frac{e^{\beta_c \cdot X_i}}{\sum_{n=1}^{n} e^{\beta_n \cdot X_i}}$$

Giving the probability of an outcome c, where β_c is a vector of the coefficients associated with the outcome c and X_i is an indicator vector of the variables for an observation i. An agent's goals added follows:

[16] $$\forall c \in n \colon K_p^c = \Delta_p^c \psi_c NS_p$$

Where n is the set all of possible outcomes, Δ_p^c is the difference in probabilities of an outcome c relative to baseline due to the involvement of an agent p, ψ_c is the goal value of outcome c, and NS_p is the total number of shifts played by agent p. n has four elements: three possible zone start outcomes and a shift end which does not result in a face-off.

2.2.5 Cross-Validation and Regularization

The conditions for the regressions described afore are conducive to overfitting. In particular, the large number of predictors calls for a variable selection process. I employ a two-pronged policy of K-fold cross-validation and elastic net regularization to optimize model efficiency on a general out-of-sample data set. For a regression comprised of N observation pairs (x_i, y_i) the elastic net solves (Friedman et al. 2010):

[17] $$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}}\left[\frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \beta_0 - x_i^{T}\beta\right)^2 + \lambda P_\alpha(\beta)\right]$$

Where $P_\alpha(\beta)$ is the elastic net penalty:

[18] $$P_\alpha(\beta) = (1 - \alpha)\frac{1}{2}\|\beta\|_{\ell_2}^2 + \alpha\|\beta\|_{\ell_1}$$

The elastic net mixing parameter α was given the value of 0.25 by simple trial. From here, K-fold cross-validation was performed for a sequence of lambdas of which the optimum was selected.

Figure 5A: The coefficient profile plot of the path for the lambda sequence used for the Cox regression of home team shot rates; Figure 5B: The cross-validation curve for the lambda sequence used for the Cox regression of home team shot rates.

The number of folds and values of lambda to be validated was dependent on the regression and agent type under observation. For teams, 4 folds were used and a sequence of 100 lambdas; for skaters, 3 folds and 60 lambda values.

Results

3.1 Goodness of Fit

The cross-validation process identified the penalty term λ for which the computed mean cross-validated error was minimized for each of the regressions. For the Cox regression models, this error is the partial likelihood deviance. For the logistic models, they are the binomial or multinomial deviance. The absolute mean error was recorded for each cross-validated model, as well as the deviance ratio defined as the fraction of the null deviance explained.

	07-08	08-09	09-10	10-11	11-12	12-13*	13-14	14-15	15-16	DF	Avg.
A_T	0.01503	0.01586	0.01425	0.01293	0.01217	0.01248	0.01287	0.01251	0.01220	~70	0.01337
B_T	0.00847	0.00881	0.00562	0.00680	0.00720	0.01027	0.00757	0.00650	0.00777	~33	0.00767
C_T	0.00756	0.00941	0.00970	0.00812	0.00859	0.00794	0.00732	0.00798	0.00698	~52	0.00818
A_S	-	-	-	-	-	-	0.21755	0.21800	0.21880	1111	0.21812
B_S	-	-	-	-	-	-	0.03262	0.02565	0.03582	535	0.03136
C_S	-	-	-	-	-	-	0.50700	0.51265	0.51205	153	0.51057
D_S	-	-	-	-	-	-	0.12440	0.11830	0.11940	1505	0.12070

* Lockout year (48-game regular season)

Table 1: Average maximum cross-validated deviance ratio of the regressions used in the grand model. A-D as labeled in section 2.1, T for Teams as agents, S for Skaters as agents. DF = Degrees of Freedom.

3.2 Ratings

An agent's rating *K* is equal to the sum of each *K*-component for which their agent class is responsible.

[19] $$K_p = \sum_n K_n^p$$

Where $K_1, K_2, \ldots, K_n$ are the class-dependent *K*-components. *K* is interpretable as the approximate net goals added by an agent relative to a neutral agent – one that has

Figure 6: The class dependency of *K*-components for the three major agent types.

neither a positive nor negative influence on any process contained in the grand model. Using the specifications described in section 2.2.2, only 16.4% of goaltenders evaluated were assigned non-zero coefficients. While the hypothesis that goaltenders may impact this process is supported, they do not act as foci in the particular model used for teams and skaters. A separate specialized regression was instead used to

	A	B	C	D
Teams	100.00%	57.92%	97.08%	-
Skaters	99.06%	54.26%	27.96%	100.00%
Goaltenders	-	16.41%	-	-

Table 2: The percentage of agents by class with non-zero *K*-components in each of the processes A-D as labeled in section 2.1.

generate goalie ratings. In this new logistic model, the observations are limited to shots on goal. This dismisses the assumption that goaltenders can cause shots to be blocked by skaters or miss the net. Agents acting as goalie or shooter serve as predictors, along with the set of game states and the xG value of shots. Lastly, the coefficients are regularized using the ridge method, and the model is cross-validated using 4 folds.

3.2.1 Team Ratings

Team *K* ratings are pro-rated by calculating goals added per 82 games in order to account for lockout or otherwise shortened seasons. The Chicago Blackhawks of 2012-13 are far and away the best team observed in this analysis. In 6 of 8 seasons where a Stanley Cup was awarded, the winner

2010-11	*K*	2011-12	*K*	2012-13*	*K*	2013-14	*K*	2014-15	*K*	2015-16	*K*
S.J	51.2	STL	63.0	CHI[S,P]	103.8	BOS[P]	67.1	T.B	45.6	L.A	48.7
VAN[P]	43.9	PIT	59.3	PIT	62.6	CHI	49.7	L.A	31.7	DAL	28.9
BOS[S]	34.5	BOS	50.6	L.A	60.1	S.J	46.0	CHI[S]	26.9	PIT	27.0
CHI	28.8	DET	41.0	BOS	58.4	ANA	41.9	NYI	24.8	ANA	25.2
PIT	24.9	VAN[P]	39.5	NYR	52.0	L.A[S]	38.5	NYR[P]	21.9	WSH[P]	21.0
…		…		…		…		…		…	
ANA	-31.5	TOR	-24.3	EDM	-38.9	NYI	-31.6	TOR	-39.3	OTT	-20.4
NYI	-33.3	NYI	-26.1	BUF	-38.9	TOR	-55.0	COL	-43.0	ARI	-21.3
MIN	-34.7	CBJ	-42.8	CAR	-39.3	EDM	-61.2	ARI	-63.2	N.J	-26.3
EDM	-43.6	T.B	-43.3	CGY	-59.8	FLA	-68.5	EDM	-70.5	TOR	-30.8
COL	-62.8	MIN	-63.0	FLA	-94.8	BUF	-93.0	BUF	-108.7	COL	-39.5

* Lockout year (48-game regular season).
[P] Won President's Trophy (most regular season points).
[S] Won Stanley Cup.

Table 3: The top and bottom ranked teams for each of the last 6 seasons.

was among the top-5 teams by this metric. 5 times, they were top-3 teams, and thrice the best overall. In each of 9 seasons tested, the President's Trophy winner was among the top-5 teams as rated by *K*.

3.2.2 Player Ratings

The distribution of player ratings does not vary considerably between forwards and defencemen, as shown in figure 7. The standard deviation of *K* is 5.53 for forwards and 5.56 for defenders, while it is 9.30 for goaltenders.

Figure 7: The distribution of offensive and defensive *K*-components for forwards and defencemen and the distribution of total *K* for goaltenders. Mean and standard deviation shown in red.

The mean is approximately 1 goal above zero and 1 below zero for forwards and defencemen respectively. This disadvantage can be attributed in large part to the inability of defenders to add goals by acting as shooters relative to a neutral agent. The model is agnostic to positions within an agent class, making defencemen poorer shooters in comparison to the total mean ability of the group. Only Nick Holden in 2013-14 managed a positive shooter *K*. Table 4 contains the best and worst 5 players at each position over each of the last three years according to total *K*. The best single season for a skater

	2013-14	*K*	2014-15	*K*	2015-16*	*K*
Forward	J. Toews	28.2	N. Kucherov	24.2	P. Kane	25.1
	A. Kopitar	27.2	J. Toews	22.8	L. Eriksson	23.5
	C. Perry	24.9	V. Tarasenko	19.6	E. Kuznetsov	22.5
	J. Jagr	22.1	N. Foligno	18.8	P. Bergeron	21.0
	J. Williams	19.6	J. Tavares	18.6	T. Toffoli	21.0
	…		…		…	
	E. Nystrom	-13.5	N. Thompson	-11.0	R. Reaves	-10.3
	A. Chiasson	-16.0	D. MacKenzie	-11.1	J. Stoll	-11.3
	S. Ott	-17.6	T. Vanek	-11.5	P. Gaustad	-11.5
	N. Yakupov	-19.0	C. Stewart	-11.9	M. Boedker	-13.8
	J. Spezza	-20.3	C. Paquette	-11.9	D. Stafford	-17.3
Defence	M. Niskanen	24.5	K. Shattenkirk	25.1	C. Parayko	30.4
	M. Giordano	22.5	A. Stralman	23.1	J. Klingberg	24.2
	B. Seabrook	17.0	T. Barrie	21.2	R. McDonagh	19.8
	A. Pietrangelo	14.0	K. Letang	15.9	P.K. Subban	19.3
	R. Ellis	13.8	D. Keith	15.5	M. Ekholm	17.8
	…		…		…	
	D. Girardi	-13.1	N. Kronwall	-11.5	B. Hutton	-11.0
	A. Edler	-14.2	D. Engelland	-11.5	J. Merrill	-11.1
	D. Boyle	-14.6	N. Guenin	-12.8	J. Bouwmeester	-11.9

	A. MacDonald	-19.2	D. Seidenberg	-13.1	B. Seabrook	-12.7
	M. Weber	-22.7	T. Daley	-16.0	D. Girardi	-14.5
Goalie	H. Lundqvist	27.2	C. Price	32.9	B. Elliott	30.2
	C. Price	24.0	B. Holtby	22.4	C. Crawford	17.7
	S. Varlamov	22.6	C. Schneider	18.2	C. Schneider	15.6
	T. Rask	20.6	M.A. Fleury	16.7	T. Greiss	15.5
	K. Lehtonen	18.7	H. Lundqvist	16.4	H. Lundqvist	13.4
	…		…		…	
	D. Ellis	-15.4	N. Backstrom	-15.1	J. Bernier	-13.5
	A. Lindback	-16.3	C. Johnson	-15.9	M. Condon	-14.2
	R. Berra	-16.6	R. Emery	-17.5	G. Sparks	-14.2
	K. Poulin	-19.0	M. Smith	-19.4	J. Hiller	-21.2
	D. Dubnyk	-20.1	B. Scrivens	-38.7	P. Rinne	-26.5

* Playoffs excluded

Table 4: The top and bottom ranked players by position for each of the last three seasons.

belongs to Colton Parayko in 2015-16. Carey Price holds the best rating among all players for his 2014-15 season performance. Ben Scrivens in the same year is a distant worst, at 38.7 goals below baseline.

3.3.3 Multi-Year Ratings

The regressions were limited to single seasons because memory requirements for multi-year computations exceeded the limits of the system used. Aggregate ratings for agents spanning multiple seasons are reported as rates of total goals added, by game played or time on ice. Here, games played is preferred as it accounts for the value added by playing a greater quantity of minutes. The greater relative importance of defencemen due to this fact becomes implicit. Let $K/82$ define the approximate goals added by an agent per 82 games played relative to a neutral agent. We find that Pavel Datsyuk

Forward	$K/82$	Defence	$K/82$	Goalie	$K/82$	Team	$K/82$
P. Datsyuk	19.1	K. Letang	13.5	C. Price	32.0	L.A	39.3
J. Toews	18.6	J. Klingberg	12.8	B. Elliott	22.3	CHI	28.2
J. Jagr	17.8	R. McDonagh	11.3	C. Schneider	21.4	BOS	27.3
P. Bergeron	17.7	P.K. Subban	10.6	H. Lundqvist	21.3	ANA	24.9
J. Thornton	16.4	M-E. Vlasic	10.0	S. Varlamov	13.6	STL	21.9
L. Eriksson	15.9	M. Giordano	9.7	J. Halak	13.2	T.B	21.6
T. Toffoli	14.2	A. Stralman	9.3	S. Mason	13.1	S.J	20.3
J. Pavelski	14.2	K. Shattenkirk	9.3	C. Crawford	13.0	PIT	20.2
C. Perry	14.0	E. Karlsson	8.3	C. Talbot	12.3	DAL	15.6
P. Kane	13.9	D. Keith	7.9	B. Holtby	10.2	NYR	12.8
…		…		…		…	
E. Nystrom	-7.0	J. Bouwmeester	-8.8	M. Smith	-5.4	CBJ	-11.3
L. Korpikoski	-7.2	M. Irwin	-8.8	J. Howard	-7.7	OTT	-12.4
D. Stafford	-7.2	N. Guenin	-8.9	J. Bernier	-10.2	N.J	-16.1
T. Glass	-7.4	B. Strait	-9.0	K. Ramo	-10.3	FLA	-20.0
G. Campbell	-7.6	D. Girardi	-9.4	A. Niemi	-10.8	COL	-22.0
P. Gaustad	-7.7	D. Seidenberg	-10.4	J. Hiller	-14.0	ARI	-26.3
S. Ott	-8.2	L. Smid	-10.8	J. Reimer	-14.3	CGY	-27.2
J. Lupul	-8.8	J. Merrill	-11.0	C. Ward	-15.0	TOR	-41.6
C. Paquette	-10.3	A. Edler	-11.2	P. Rinne	-15.4	EDM	-47.2
N. Yakupov	-11.7	S. Gonchar	-11.3	B. Scrivens	-24.8	BUF	-70.3

Table 5: Best and worst aggregate $K/82$ for each agent class and position from 2013-2016.

supplies the most positive impact per game of action among all regular skaters (having played 2,000 or more minutes) over the past 3 seasons. Carey Price is the best-ranked starting goaltender (having played 100 or more games) in that span.

3.3.4 Validity

Various tests were performed in order to verify that, and to what extent, the ratings produced by the model were representative of the quality of agents under observation. Verifying that K is descriptive of quality is only superficial in evaluating its validity, as it is a purely retrospective venture. As stated in section 1.1, the property of informing predictions for future outcomes is paramount. To begin, the correlation between teams' point

Figure 8: A comparison of the relationship between various metrics and team points. Z-Scores are used to standardize the measures.

totals and their *K* ratings was examined in comparison with two alternate evaluative measures. *K* was far more descriptive of success in the regular season than 5v5 Corsi-For percentage (CF%) and expected Goals-For percentage (xGF%) with a reported R^2 value of 0.6828. For perspective, teams' Goals-For percentage (GF%) in all situations produced an R^2 value of 0.8807 in the same test. Next, box and whisker plots were produced to visualize the distribution of players' *K* ratings categorized by position and according to Time On Ice percentage (TOI%) and Salary bins. With the

Figure 9: Box and whisker plots for players' *K* by season from 2013-2016, coloured by position and grouped by TOI% (left) and Salary (right) bins.

apparent exception of top-paid defencemen, players appear to be generally rewarded with more ice time and more lucrative contracts as *K* increases. This trend appeals to the notion NHL coaches and General Managers are astute talent evaluators, though we expect some exploitable inefficiency on their part. Because player ability is assumed relatively constant, we expect a statistically significant correlation between *K* in consecutive seasons. This idea is satisfied by comparing the linear relationship between year-over-year player-*K* ratings against other evaluative metrics. While the reported

Metric	Residual Standard Error	R^2	P-Value
K	5.79 on 994 DF	0.1799	$< 2.2e^{-16}$
+/-	10.20 on 1008 DF	0.1015	$< 2.2e^{-16}$
5v5 CF%	3.24 on 1008 DF	0.3653	$< 2.2e^{-16}$

Table 6: Summary of the linear fit between regular players' (having played 41 or more games) ratings according to various metrics in consecutive seasons.

correlation coefficient is stronger for 5v5 CF%, it is worth noting that this measure does not adequately account for the influence of teammates. Lastly, a simple forward-looking experiment was conducted to ensure that the strength of teams' rosters as approximated by total player-*K* was successful in predicting regular season performance to a statistically significant degree. A crude model was constructed using each team's opening-night roster, where total player-*K* is the sum of each player's *K* in the previous (Y-1) season. Where this value was absent, a flat rating of zero was assigned. This roster quality measure served as a predictor in a linear regression where regular season points is the dependent variable. The reported P-Value in this test was 0.000077, and the Residual Standard Error of 12.63 was slightly smaller than that of Y-1 team 5v5 CF% at 12.74. It is reasonable to expect that this predictive method can be greatly improved upon by using multi-year *K* and a multivariate model.

Discussion

4.1 Applications

The primary applications of *K* ratings are prediction and team or player evaluation. Simple predictive models were designed in order to verify that the use of information yielded by this experiment is advantageous to common alternatives. Using data from the past 8 seasons, each of regular season points per 82 (P/82), 5v5 CF%, 5v5 xGF% and *K*/82 in the previous season were used as predictors for regular season team points. Each metric was tested on 8 folds. In each, one season was kept as a testing subset and the remaining 7 to train the linear regression model. The residual standard error produced in the testing data was averaged over all 8 folds, giving the Mean Residual Standard Error (MRSE) summarized in table 7. The best results were obtained using Y-1 *K*/82 indicating potential predictive validity at the

Y-1 Measure	P/82	5v5 CF%	5v5 xGF%	K/82
MRSE	9.66	9.84	9.79	9.30

Table 7: 8-fold Mean Residual Standard Error of various linear models. Each Y-1 measure was used as a lone predictor for regular season team points.

team level. Supporting evidence is provided by a second test in which the outcomes of playoff series were predicted using basic logistic regression models. The same metrics were tested, this time both teams' regular season ratings serving as covariants. An iterative random sampling process was used, whereby 20% of the total data were kept as a testing set and the remaining 80% served as training data. In each of 10,000 iterations the logarithmic loss was calculated for fitted probabilities in the random testing set and these values were subsequently averaged. At 0.6488, $K/82$ once again performed the best of the 4 metrics tested. For comparison, 5v5 xGF% was second best at 0.6848 and 5v5 CF% was the worst at 0.6966.

4.2 Curiosities

This analysis produced several noteworthy findings, many of which have been detailed in previous sections. Others are briefly reported here with the intent of encouraging further exploration. In addition to the superior results yielded by the series prediction model constructed using team $K/82$, it was found that teams with the advantage in this rating won 64.12% of playoff series since the 2008 postseason. This crude algorithm performed within 6 percentage points of the theoretical upper limit for NHL playoff series predictions (Weissbock 2013). For comparison, teams with the home ice advantage won roughly 55% of series in the same sample. A brief investigation was conducted on how the relationship between player K and both TOI% and Salary differs according to player position. Quantile-Quantile plots were produced to compare the distribution of residuals between positions. The greatest inefficiency in evaluation appears to exist at the goaltender position, with defence

Figure 10: Q-Q plots displaying the actual and theoretical residual quantiles for the linear models $K/82$ ~ TOI% (left) and $K/82$ ~ Salary (right).

coming second. We do not infer from this that the inefficiency lies within the allocation of salary or ice time.

4.3 Limitations

Though practical applications of K have been uncovered, the model is not without limitations. Some may be addressed, while others are by-products of complex regression modeling that are inescapable. For one, the process is slow and clumsy. The sheer number of both variables and observations at the skater level, particularly among the Cox regressions, coupled with the cross-validation and regularization policies employed significantly lengthen the execution time. Variables were restricted to players having appeared in 10 or more games in order to mitigate this effect. Using parallel computing on two cores, the grand model for one season of skaters requires roughly 24 system hours to reach completion. Scalability is a second concern. Data requirements and nontrivial run times mean computation of K does not generalize well to arbitrary subsets of games or date ranges. This is especially true of skaters, when it is impractical in most cases to compile ratings at regular intervals. Team and goaltender ratings may be computed at semi-regular intervals such that a predictive model may anchor itself to the most recent iterate or employ a Bayesian approach. Additionally, some K-components are prone to significant variance across seasons. In particular, the goals added by skaters acting as support on the goal probability of shots for or against their team ($K_{\text{SUPPORT FOR}}$, $K_{\text{SUPPORT AGAINST}}$) does not often persist in consecutive seasons. Predictive validity may be improved by altering the regularization parameters or by removing this item entirely.

References

Burtch, Stephen. 2014. "dCorsi – Introductions" *NHL Numbers*, July 19. http://nhlnumbers.com/2014/7/19/dcorsi-introductions.

Sprigings, Dawson. 2015. "Corsi Plus-Minus: Individual Player Value Accounting for Teammates" *Don't Tell Me About Heart*, July 23. http://donttellmeaboutheart.blogspot.ca/2015/07/corsi-plus-minus individualplayer.html.

Tulsky, Eric. 2012. "Adjusting for Score Effects to Improve Our Predictions" *Broad Street Hockey*, Jan 23. http://www.broadstreethockey.com/2012/1/23/2722089/score-adjusted-fenwick.

McCurdy, Micah. 2014. "A Better Way to Compute Score-Adjusted Fenwick" *HockeyViz*. http://hockeyviz.com/txt/shiftsArticle/senstats.html.

Awad, Tom. 2009. "Understanding GVT, Part 1" *Hockey Prospectus*, July 30. http://www.hockeyprospectus.com/puck/article.php?articleid=233.

Schuckers, Michael, and James Curro. 2013. "Total Hockey Rating (THoR): A Comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events" *MIT Sloan Sports Analytics Conference*, March 1. http://statsportsconsulting.com/main/wp-content/uploads/Schuckers_Curro_MIT_Sloan_THoR.pdf.

MacDonald, Brian. 2011. "A Regression-Based Adjusted Plus-Minus Statistic for

NHL Players" *Journal of Quantitative Analysis in Sports*. DOI: 10.2202/1559-0410.1284.

Thomas, Andrew C., and Samuel Ventura. 2015. "The Road to WAR Series" *WAR On Ice*. http://blog.war-on-ice.com/the-road-to-war-series-index/.

Krzywicki, Ken. 2010. "NHL Shot Quality 2009-10" *Hockey Analytics*. http://hockeyanalytics.com/Research_files/SQ-RS0910-Krzywicki.pdf.

Perry, Emmanuel. 2016. "Shot Quality and Expected Goals: Part I" *Corsica*, March 3. http://www.corsica.hockey/blog/2016/03/03/shot-quality-and-expected-goals-part-i/.

Thomas, Andrew C., Samuel Ventura, Shane Jensen, and Stephen Ma. 2013. "Competing Process Hazard Function Models for Player Ratings in Ice Hockey" *arXiv*, March 1. https://arxiv.org/pdf/1208.0799.pdf.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent" *Journal of Statistical Software*, January. https://www.jstatsoft.org/article/view/v033i01.

Weissbock, Josh. 2013. "On Luck, Talent, Parity and Predicting the Upper Bounds With Machine Learning" *NHL Numbers*, October 1. http://nhlnumbers.com/2013/10/1/on-luck-talent-parity-and-predicting-the-upper-bounds-with-machine-learning.