A Redefinition of Hockey Goalie Save Percentages Using Weights Generated by Spatial and Temporal Functional Tensor Analysis*

Barinder Thind† and Jiguo Cao‡
Statistics & Actuarial Science
Simon Fraser University.
(USRA Research Completion)

Matthew Reyers§
Statistics & Actuarial Science
Simon Fraser University.

Brad Smallwood¶
Financial Analyst
VanCity Mutual Funds.
(Dated: February 11, 2019)

In recent years, the usage of analytics in hockey has seen a dramatic increase. The advent of such tools has completely revolutionized the way in which we evaluate players. In particular, the goaltender position has been viewed through the lens of various statistical methods. In this paper, we introduce a weighted save percentage metric that incorporates results generated by a functional boosted decision tree model; the technique considers spatial and temporal information to optimize a set of features such that they best optimize the emphasis placed on every individual shot taken throughout a game. This method is then specified across games for specific goalies to demonstrate merit; that is, goalies are shown to follow a particular trajectory as evident by their realizations season to season.

KEYWORDS: Markov Chains, Functional Data Analysis, Boosted Decision Tree, Gradient Descent

I. INTRODUCTION

In this paper, we look to analyze a large and detailed NHL data set with the ultimate goal of finding a new measure to evaluate goalies with. The data set included temporal and spatial information - i.e. there was information available on where a shot was taken from and at what time in the game the shot was taken. Both of these factors, along with others, were considered to create a new shot quality metric.The new shot quality metric was then used as a weight on saves a goalie makes throughout the game. This allowed us to take into account the probability of a shot becoming a goal when evaluating a goalie's save percentage.

In particular, the shot quality metric is developed using a boosted decision tree model that showed robustness and high accuracy under various cross-validation techniques. A plethora of factors were taken into account to demonstrate this fact. The temporal aspect is explored individually where a decision is made on its impact. In the Methodology section, a description of the data is given, some initial analysis and statistics are presented, the shot quality metric is detailed, and lastly, the distribution outlining the save percentage is shown. In the Results section, the methods presented are applied to two goalies, retroactively. This allowed us to show that validity of the model as the results are shown to match the actual trend of the goalies. Finally, the conclusions and some future considerations are presented.

* Author Emails
† bthind@sfu.ca
‡ jiguo.cao@sfu.ca
§ mreyers@sfu.ca
¶ bsmallwo@sfu.ca

II. METHODOLOGY

A. Data Description

The data set is vast - it includes information for over 5000 games. Each of these games contains specifics on which players are on the ice, when they change, who takes a shot or a penalty, and whether the shot got to the net or not. The shots are then further sub-categorized into "type-of-shot"; this is broken into wrist shots, slap shots, tip-ins, deflections, wrap-arounds, and snap shots. Furthermore, we did some feature engineering to add an angle variable. This variable uses the given Cartesian co-ordinates to help us transform the data into its polar counterpart - the main advantage being that the analysis is made a bit simpler as we do not have to deal with negative values. Figure 1 highlights the angle calculation.

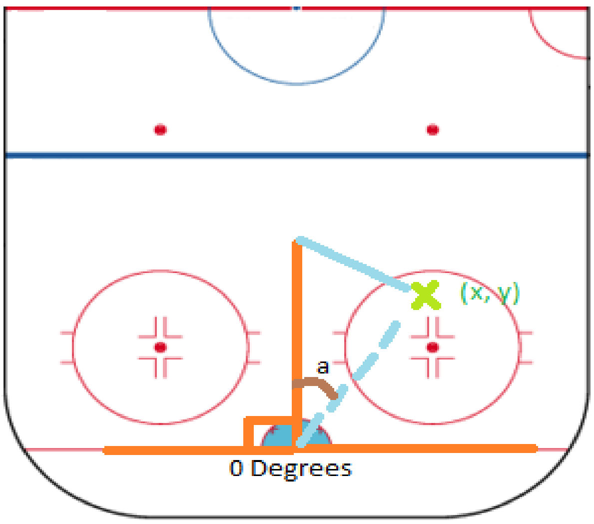

FIG. 1. Angle Calculation: Shots that come directly from in front of the net are "0 degrees" and any deviations to the left or right move up absolutely from zero. In this example, x is where the player would be shooting from and *a* would be the associated angle.[a]

[a] Angle $= arctan(\frac{y}{x})$

Another variable we considered was the number of consecutive shots in a single zone entry. For example, if a team has possession of the puck for an extended period of time in the attacking zone, then the likelihood of a goal occurring increases with time. The rationality behind this is that teams get tired and are prone to make more mistakes. The current players on the ice are unable to switch out and the goalies constant movement results in fatigue. Lastly, we looked at the temporal information - was their a higher probability of a goal occurring at a particular point in the game? And if so, why?

B. Exploratory Data Analysis

We first consider time. We needed a reasonable way to measure the impact of time on the likelihood of a goal. Clearly, the raw *shot quality* was not impacted by the time, but perhaps shots in the second period were more likely to go in? The approach taken was to categorize time into one minute bins. This results in 60 categories - for each of which we can evaluate the amount of goals scored across many games. To demonstrate the result, we take a sample of a 1000 games and the aftermath is presented in Figure 2.

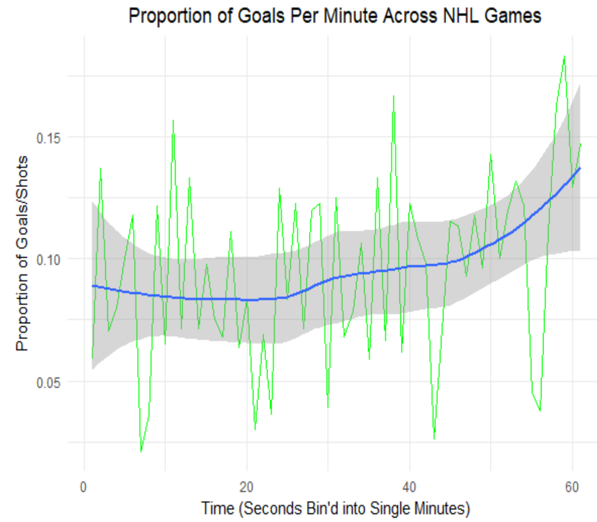

FIG. 2. Proportion of Goals Scored Per Minute of a Game

As is evident, the amount of goals being scored at any point in the game seems to be relatively consistent. The local regression curves standard errors only deviate substantially near the end of the game. This could be due to empty net goals - it is much easier to score on a shot if their no goalie in the net! A *Monte Carlo* bootstrap can be used to flush out some of the noise in Figure 2.

Next, we can look to see how different types of shots have different likelihoods of going in! A common question might be: Is a wrist shot more likely to result in a goal than a slapshot? The data was sampled from and, for 50 games, the following results were seen:

Probability of a Goal

Backhand	Deflected
0.12598425	0.38709677
Slap	Snap
0.04990403	0.12050740
Tip-In	Wrist
0.21705426	0.09533074

FIG. 3. Probability of a Goal given a Particular Type of Shot

In the given table, the proportions are associated with the chance that when that type of shot is taken or occurs, how likely has it been seen to go in. In other words, this is not the proportion of goals separated across the different types of shots. It might be the case that there are only 38 deflection-type goals in a 100 shots, but there are 450 slapshot goals, just over 10,000 shots. In essence, this is the conditional probability of a goal given a type of shot: $\mathbf{P(Goal \mid ShotType}_i\mathbf{)}$ for $i = 1, 2, ..., 6$.

Next, we explore the distance variable which we will use for our analysis. A simple regression will show that distance is highly correlated with the likelihood of a goal. Visually, we can see in Figure 4 that, as distance increases, goals become much less likely.

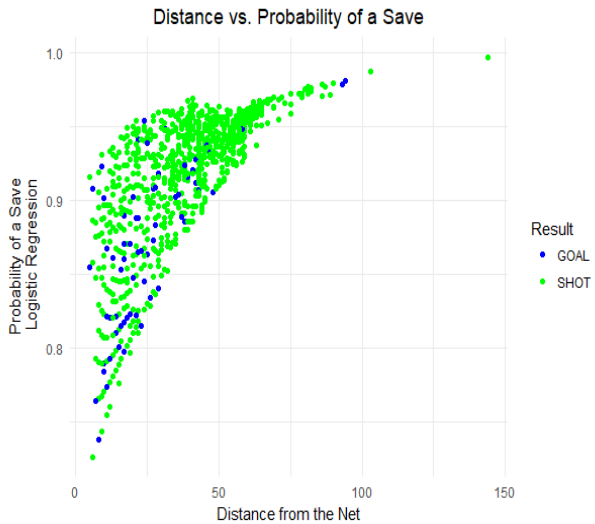

FIG. 4. Distance against Probability of a Shot being a Goal

Here, we can see that, as distance increases, the logistic regression predictions on whether a shot is saved (or not), goes up steadily. The actual results can also be seen, with a large concentration of the goals coming in short distances from the net. Similarly, we can plot the angle against the same probabilities to see that there is an optimal ban of angles to shoot from which result in the most goals:

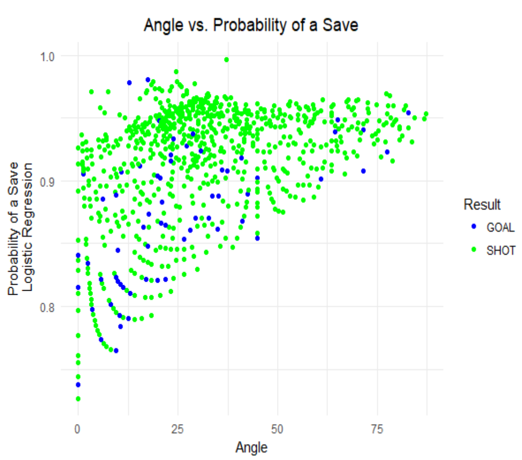

FIG. 5. Angle against Probability of a Shot being a Goal

More specifically, we see that a large majority of the goals occur between 0^o and 40^o deviations from having a direct view at the net. This information can then be combined to view how the distribution of proportions looks as displayed on a make-shift hockey rink:

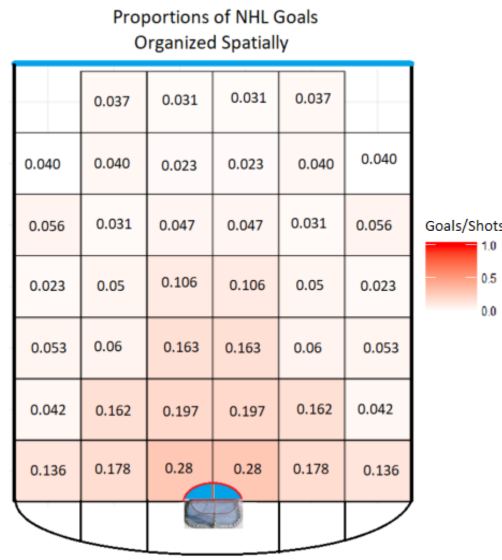

FIG. 6. A Top-Down View of a Hockey Rink Displaying Goal Proportions

The majority of goals seem to come from the directly in front of the net and from short distances. The conditional probability of a goal given that the shot was a deflection is 4 times that of a wrist shot and 5 times that of a slap shot. Most players aren't shooting from the sharp corners unless they are looking for deflections; this means that when shots are taken, they are very likely to go in due to deflections. This likely results in the higher shooting percentages from those areas. Also note that the mirror image is a result of the lack of separation between the direction of the angle. As mentioned earlier, there isn't a distinction made between whether they angle is garnered from the left or right of the center, so shots from each side are confounded. This restricts inference in some domains however, our analysis remained unaffected. [1]

C. Stochastic Boosted Decision Trees

In this section

D. Adjusted Save Percentages

In this subsection, the new save percentages are defined for which goalies will be re-evaluated. As mentioned earlier, at this point, we will have a shot quality metric that takes in various parameters (distance, angle, type of shot, etc.) and outputs a new value for that occurrence of the shot. These new values will be used as a part of the Adjusted Save Percentage Metric (ASPM) defined as follows:

$$\text{ASPM} = \frac{\sum_{i=1}^{k} \sum_{i=1}^{n} \frac{\alpha_{ki} I(s_k i \neq goal)}{n}}{k}$$

Where $I(s_k i \neq goal)$ is defined to be 0 when the s_i^{th} shot is a goal, and 1 otherwise. Note here that the index i refers to the i^{th} shot of the k^{th} game. There is also an additional condition on the expectation of α:

$$\mathbf{E}(f_\alpha(\alpha)) = \frac{\sum_{i=1}^{n} shot_i}{n}$$

Where $shot_i$ is i^{th} shot evaluated through the weighting metric proposed in the previous section. The purpose of this condition is to make it so that the difficulty of the average shot is weighted as such in this adjusted save percentage metric.

III. RESULTS

A. Analysis for a Goalie Towards an Upward Trend

in this

B. Analysis for a Goalie Towards a Downward Trend

in this

IV. CONCLUSION

test

A. References

[1] Badi Baltagi. *Econometric analysis of panel data.* John Wiley & Sons, 2008.