A sparse regression approach for evaluating and predicting NHL results

Brendan Ames

Department of Mathematics
The University of Alabama

Alabama Program in Sports Communication
Fifth Annual Sports Symposium
January 27, 2017

Introduction

Propose new statistical models for predicting and analyzing NHL results.

Approach is based on linear and Poisson regression with sparsity inducing penalty terms.

Penalization allows greater interpretation of results, revealing most discriminative statistics.

Joint with Nicholas Laffey (undergraduate researcher).

Challenges

Analyzing and predicting hockey results presents several challenges:

- Low scoring rates cause significant randomness and variance in results.

- Continuous nature of play and frequent changes of possession place cause difficulties in establishing dominance of play.

- Significance of summary statistics vary due to context, causing counterintuitive results.
  - **e.g.** teams with lead tend to be outshot, teams with high hit counts do poorly.

Current State of the Art

Use shot attempt counts as surrogates for possession.

- Fenwick score: shots on goal plus missed shots on goal.
- Corsi score: shots, missed shots, and blocked shots.
- Limited evidence that high Fenwick/Corsi differential correlates with wins.

MacDonald 2012: Expected Goals model based on linear regression.

- Fits linear model to goals scored using predictor variables: shots, missed shots, blocked shots, Fenwick rating, Corsi rating, zone starts, turnover rates, faceoff rates, hits, shooting percentage.

Our Approach

Extend method of MacDonald 2012 to much wider variety of statistics.

Fit wins to 1011 predictor variables given by exhaustive list of statistics.

Predictor variables spread across contexts

- regular strength, power play, trailing, leading, tied, etc.

Penalized Regression Models

Data is severely undersampled:

- Have significantly fewer teams/season to train model with than statistics.
- Traditional linear regression leads to overfitting.

Use **elastic net regression** to eliminate overfitting:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_1 + \gamma \|\boldsymbol{x}\|_2^2.$$

- $f(\boldsymbol{x})$ is ordinary least squares or Poisson error term.
- $\|\boldsymbol{x}\|_1$ term induces **sparse** solutions, yielding greater interpretability of results (more later).

Regular Season Models

We fit Poisson elastic net models to predict number of regular season wins.

Used data from 2014-2015, 2013-2014, 2011-2012, 2010-2011, and 2009-2010 seasons to train the models.

Test models by predicting results for 2008-2009 season.

Regular Season Results

	West Poisson		Actual		East Poisson		Actual	
1.	CHI	44.26	SJ	53	BOS	48.63	BOS	53
2.	DET	43.71	DET	51	NJ	44.07	WSH	50
3.	SJ	42.14	VAN	45	WSH	43.26	NJ	51
4.	VAN	40.64	CHI	46	FLA	39.33	PIT	45
5.	CGY	37.45	CGY	46	PIT	39.17	PHI	44
6.	STL	35.66	STL	41	CAR	37.83	CAR	45
7.	CBJ	34.93	CBJ	41	PHI	37.77	NYR	43
8.	MIN	34.88	ANA	42	BUF	37.14	MTL	41
9.	ANA	34.03	MIN	40	MTL	33.83	FLA	41
10.	EDM	31.59	NSH	40	NYR	31.18	BUF	41
11.	DAL	31.25	EDM	38	OTT	30.36	OTT	36
12.	NSH	30.46	DAL	36	ATL	29.29	TOR	34
13.	LA	30.12	ARI	36	TOR	26.30	ATL	35
14.	ARI	28.20	LA	34	TB	25.81	TB	24
15.	COL	25.99	COL	32	NYI	24.06	NYI	26

Playoff Results

Train models using to predict playoff proficiency based on regular season results.

Use team statistics from the years 2009-10, 2010-11, 2011-12, and 2013-14 as training data.

Test models using 2007-08, 2008-09, 2012-13, and 2015-16 seasons.

Method	**Success Rate**	**Method**	**Success Rate**
Poisson Ridge	**0.6833**	**Poisson Elastic Net**	**0.6000**
OLS Ridge	**0.6833**	Wins	0.6000
Goal Differential	0.6833	**OLS Elastic Net**	**0.5778**
Fenwick	0.6833	PDO	0.4833
Expected Goals	0.6667	Team Salary	0.4667
Corsi	0.6167		

Interpreting the models

We retrained Poisson models with added restriction that model contains 8-15 coefficients.

Metric	Regular	Playoff	Type
Salary Cap Space		-0.018540109	Salary
% Salary Cap Spent on Goalies	0.1459301		Salary
Goals For	0.3704185	-0.014109219	Offensive
Goals For %	3.2139014		Offensive
Shooting %	0.7025324		Offensive
High-Danger Scoring %	0.1172388		Offensive
Hits Taken		0.027732577	Offensive
Penalty Killing %		0.060864055	Defensive
Save %		0.041979715	Defensive
Scoring Chances Against		-0.008634176	Defensive
Goals Against		-0.113854819	Defensive
Goal Differential	2.0927952		Combined
PDO	0.4012918		Combined
High-Danger Scoring Chances Differential	0.2354926		Combined
Scoring Chances Total		-0.00464826	Combined

Future Research

Predictions obtained by fitting to goals/shots instead of wins.

Repeat analysis with other sports.

Predict player career performance from amateur statistics.

- Identify traits of successes/busts?