

A Look at ~~Sports~~ Hockey Analytics

Michael Schuckers

St. Lawrence University
Statistical Sports Consulting, LLC
schuckers@stlawu.edu
@SchuckersM
@EmpiricalSports

November 14, 2014

What the heck is analytics?

- ▶ IBM: discover what is happening, determine why it is happening, predict what is likely to happen and prescribe the best action to take
- ▶ INFORMS: Analytics is defined as the scientific process of transforming data into insight for making better decisions.
- ▶ OXFORD DICTIONARY: The systematic computational analysis of data or statistics
- ▶ SCHUCKERS: less scary word for statistics

Sports Analytics is statistical with emphasize of the importance of:

- ▶ Reliability: Is it repeatable?
- ▶ Validity: Does it correlate with Winning?

What is Hockey?

Distinguishing Features

- ▶ Very fluid (think basketball, soccer)
- ▶ Players on ice for short bursts (50 seconds)
- ▶ Penalties result in team being without a players for length of penalty (but can return)
- ▶ OT played different than rest of the game (4v4)
- ▶ Tie at end of OT in regular season ends in shootout

- ▶ Even Strength(EV) : Both teams 5 players on ice
- ▶ Power Play(PP): When you have more players on the ice than your opponent
- ▶ Shorthanded(SH) : When you have fewer players on the ice than your opponent

- ▶ Offensive Zone (O-zone): Third of the ice closest to opponent goalie
- ▶ Defensive Zone (D-zone): Third of the ice closest to own goalie
- ▶ Neutral Zone (N-zone): Center third of the ice

Points:

Teams, Points = 2 for W, 1 for OT L, 0 for Reg Loss

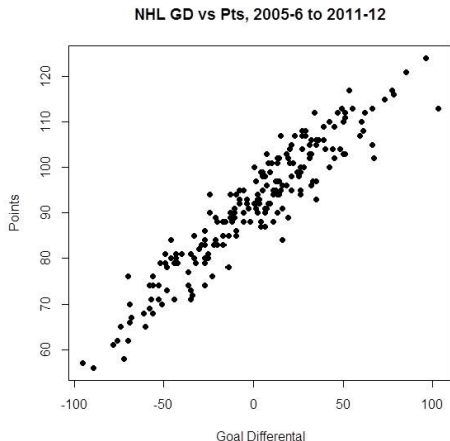
Players, Points = 1 for Goal or Assist

SV% is Save Percentage and is percentage of shots faced that are not goals

SH% is Shooting Percentage and is percentage of shots taken that are goals

RTSS is NHL's Real Time Scoring System which records events, their location, who is on the ice, etc. SHOT, MISS, HIT, BLOCK, FACEOFF, GIVE, TAKE, STOP, etc.

Goal Differential vs Points



$$\hat{Pts} = 91.6 + 0.33(GoalDiff)$$

Correlation = 0.94

What's a Goal Worth

In Wins

$$\hat{Pts} = 91.6 + 0.33GoalDiff(GD)$$

$$0.33 \rightarrow +1 GD = 1/3 pt$$

$$3GD's = 1pt$$

$$6 GD's = 2pts = 1 win$$

In Dollars

$$\text{Slope} = (59.4 - 12.6) / (91 - 52) = \$1.2\text{MM}/\text{point}$$

$$\$0.4\text{MM}/\text{Goal}, \quad \$2.4\text{MM}/\text{Win}$$

(Done in 2010/11, likely more \$\$\$)

Source: <http://www.arcticicehockey.com/2011/10/12/2482642/how-much-do-wins-cost>

Corsi # = Goals + Shots + Missed Shots + Blocked Shots

Corsi % = Corsi For / (Corsi For + Corsi Against)

Fenwick # = Goals + Shots + Missed Shots

Fenwick % = Fenwick For / (Fenwick For + Fenwick Against)

ES Team Corsi % and Fenwick % correlate highly with future win percentage, future goal scoring rates

Corsi # and Fenwick # correlate with Scoring Chances #

Also proxies for possession

Binary Game

Hockey is an binomial, np, game

- ▶ n is number of shots for
- ▶ m is number of shots against
- ▶ p is probability of shot is goal for
- ▶ r is probability of shot is goal against

Goal Differential =

$$n_{EV}p_{EV} + n_{PPP}p_{PPP} + n_{PK}p_{PK} - m_{EV}r_{EV} - m_{PP}r_{PP} - m_{PK}r_{PK}$$

Two parts: n's and p's

- ▶ Corsi & Fenwick drive $n_{_}$
- ▶ Can we drive up $p_{_}$? (and drive down $r_{_}$?)

Big Debate:

- ▶ Define Shot Quality as Mean SH%
- ▶ Shot Quality matters some or very, very little
- ▶ Positions Matter: Defensemen, Forwards
- ▶ Team SH% tends to regress to league average
- ▶ Individual SH% regress to position average but more slowly (and in some cases Stamkos, Crosby not at all)

Schuckers' Defense Independent Goalie Rating (DIGR) SV% is function of two things: Goalie Ability Shots Distribution Faced

$$SV\% = \int P_g(x, y, t, s \dots) S_g(x, y, t, s \dots) dx, dy, dt, ds \dots$$

where P_g is the performance of the goalie at s, x, y, t, \dots

S_g is the Shot distribution (pdf) at s, x, y, t, \dots

Adjusted Save Pct

$$\text{Aver.GoalieSV\%} = \int \bar{P}(x, y, t, s \dots) S_g(x, y, t, s, \dots) dx, dy, dt, ds \dots$$

$$\text{ShotNeutralSV\%} = \int P_g(x, y, t, s \dots) \bar{S}(x, y, t, s \dots) dx, dy, dt, ds \dots$$

$$\text{DIGR} = \int \hat{P}_g(x, y, t, s) \bar{S}(x, y, t, s \dots) dx, dy, dt, ds$$

where DIGR is Defense Independent Goalie Rating
(Schuckers, 2011) There are issues with NHL Data (x,y)'s



4



Play By Play

Thursday, November 13, 2014
Attendance 16,839 at Rexall Place
Start 7:43 MST; End 10:18 MST
Game 0240
Final



3



OTTAWA SENATORS
Game 16 Away Game 8

EDMONTON OILERS
Game 17 Home Game 9

#	Per Str	Time: Elapsed Game	Event	Description	OTT On Ice	EDM On Ice
1	1	0:00 20:00	PSTR	Period Start- Local time: 7:43 MST	27 88 61 4 65 41 C C R D D G	23 27 8 19 22 30 C C L D D G
2	1	0:00 20:00	FAC	OTT won Neu. Zone - OTT #27 LAZAR vs EDM #27 GORDON	27 88 61 4 65 41 C C R D D G	23 27 8 19 22 30 C C L D D G
3	1	0:08 19:54	STOP	OFFSIDE	27 88 61 4 65 41 C C R D D G	23 27 8 19 22 30 C C L D D G
4	1	0:08 19:54	FAC	OTT won Neu. Zone - OTT #27 LAZAR vs EDM #27 GORDON	27 88 61 4 65 41 C C R D D G	23 27 8 19 22 30 C C L D D G
5	1	0:34 19:28	HIT	OTT #88 HOFFMAN HIT EDM #22 AULIE, Off. Zone	27 88 61 4 65 41 C C R D D G	23 27 8 19 22 30 C C L D D G
6	1	0:46 19:14	SHOT	EDM ONGOAL - #14 EBERLE, Snap, Off. Zone, 44 ft.	93 8 9 2 5 41 C R L D D G	14 93 87 2 21 30 C C L D D G
7	1	1:09 18:51	HIT	EDM #21 FERENEC HIT OTT #93 ZIBANEJAD, Def. Zone	93 8 9 2 5 41 C R L D D G	14 93 87 2 21 30 C C L D D G
8	1	1:11 18:49	SHOT	OTT ONGOAL - #9 MICHALEK, Backhand, Off. Zone, 20 ft.	93 8 9 2 5 41 C R L D D G	14 93 87 2 21 30 C C L D D G
9	1	1:12 18:48	STOP	PUCK FROZEN	93 8 9 2 5 41 C R L D D G	14 93 87 2 21 30 C C L D D G
10	1	1:12 18:48	FAC	EDM won Def. Zone - OTT #7 TURRIS vs EDM #27 GORDON	7 90 16 4 65 41 C R L D D G	23 27 8 5 88 30 C C L D D G
11	1	1:43 18:17	HIT	EDM #23 HENDRICKS HIT OTT #7 TURRIS, Neu. Zone	7 90 16 4 65 41 C R L D D G	23 27 8 5 88 30 C C L D D G

source:

<http://www.nhl.com/scores/htmlreports/20142015/PL020240.HTM>

Hockey Analytics (©2014 Michael Schuckers)

There are issues with NHL Data (x,y)'s

- ▶ Data from NHL's RTSS feed has significant issues x,y coordinates often far from ground truth (video analysis) especially in certain rinks (MSG)
- ▶ Home bias in Giveaways vs Takeaways
- ▶ HITS inconsistent Rink to Rink, etc.
- ▶ Count totals for EVENTS need additional quality control
- ▶ Spatial tracking data is coming (cf. Basketball) I hope

More on this later ...

Average Shot Probability (ASP)

Even Strength (5v5) ASP, shots faced by away team

(loess smoother for each shot type)

Correlation

0910 vs 1011 : $\hat{\rho} = 0.6092$

1011 vs 1112 : $\hat{\rho} = 0.6928$

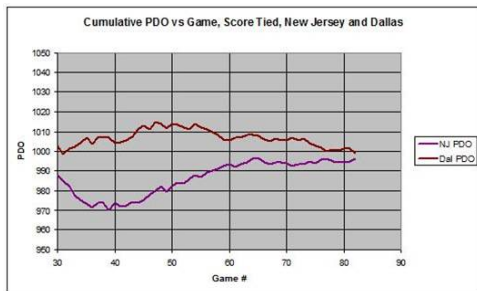
1112 vs 1213 : $\hat{\rho} = 0.4637$

Does it matter?

Typical *Max - Min* these years 23.5 Goals \approx 4 wins.

$PDO = (Team SV\% + Team SH\%)/100$
Team SV% regresses to goalie(s) career SV%
Team SH% regresses to league mean SH%
PDO is generally considered a measure of luck

PDO Example



source:<http://www.arcticicehockey.com/2011/4/11/2103499/an-abject-lesson-for-dallas-in-the-law-of-averages>

2014-15 5v5 PDO

Team	PDO
PIT	103.5
NSH	103.2
LAK	102.4
CGY	102.3
STL	101.6
⋮	⋮
CAR	97.7
WPG	97.5
EDM	97.4
BUF	97.1
CBJ	96.6

source: stats.hockeyanalysis.com

Context of Events

Player Evaluation:

- ▶ **Where shift starts (Zone Starts- ZS)** Move N-Zone to O-Zone \approx Replace Avg w/ Toews or Crosby
- ▶ **Quality of Teammates** Who with them on the ice (QoT)
- ▶ **Quality of Competition** Who against them on the ice (QoC)
- ▶ **Score Effects** What is current score (Score Effects): Teams that are ahead get better outcomes
- ▶ **Home Ice** Better events at home a la *Scorecasting*

Does QoC matter? Yes (events, shifts) and No (seasons)

Scale of these events matter!

Joint Work with Brian Macdonald (West Point)

NHL's RTSS system has Rink Effects (REs)

- ▶ Recording of SHOTs (along with x,y coord's), HITS, GIVE, TAKE, BLOCK, MISS inconsistent from rink to rink
- ▶ Noted by many analysts including: Boyle, Zona, Desjardins, McCurdy, Fischer, and Awad.

Few Analytical Solutions proposed

- ▶ GLOBALLY: Schuckers and Curro (2013) THoR does marginal adjustment across events
- ▶ SHOTs (x,y or distance): mean adjustment (SEVERAL), Krzywicki (2010), CDF adjustment (Schuckers & Curro, 2013), PDF adjustment (Thomas & student, ???)
- ▶ SHOTS (count): Pfeiffer (2014) averaging of SHOT/GOAL ratio
- ▶ EVENTs not SHOTS: *crickets*

General Approach: Schuckers and Macdonald

Log-linear model of 5v5 EV Non-Empty Net events per 3600s, Y
Different fit for each year (6 years)

Fit using 10-fold cross-validation lasso via `glmnet` in R

$\ln(Y+\epsilon) = \text{mean effect} + \text{rink effect} + \text{home ice} + \text{team effect FOR}$
 $+ \text{team effect VS} + \text{avg score diff} + \text{home*rink ('homer effect')}$

- ▶ Ran model for SHOTs, HITs, MISSs, BLOCKs, TAKEs, GIVEs
- ▶ Ran also for CORSI, FENWICK, TURNs
- ▶ Observational unit is team-game (2 obs per game)

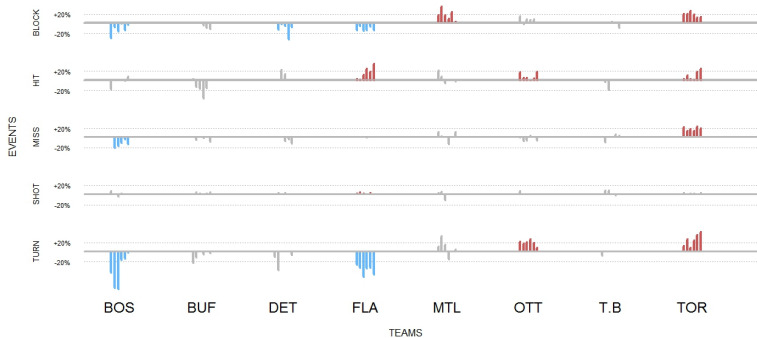
Raw data from `nhlscrapr` for 2007-8 to 2012-13

Full paper (soon) to JQAS and arXiv

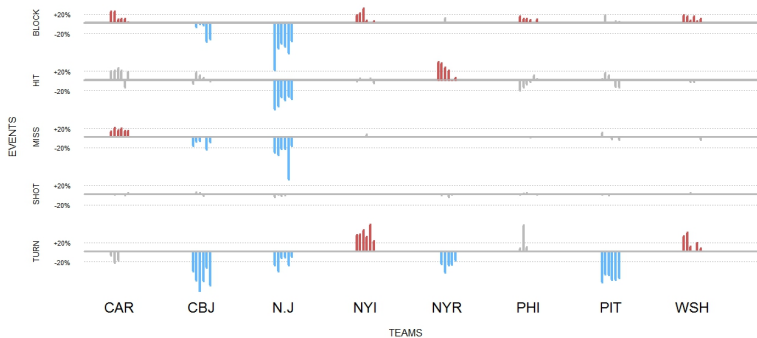
Rink's with significant, persistent problems across ALL of
BLOCK, HIT, MISS: EDM, LAK, NJD, TOR
OTT significant and persistent in HITs, GIVEs, TAKEs,

Significant & persistent means sum of sign of RE's was ≥ 5
or ≤ -5 across six years.

Atlantic Division



Metropolitan Division



Rink Effects Adjustments

For every event, weight that event by $1/RE$ for that rink.

RE for HITS in OTT is 1.091 so each HIT is weighted $1/1.091 = 0.917$.

Table 1 : Significant Rink Effects for HITS.

Rink	Effect
L.A	1.298
NYR	1.274
PHX	1.163
DAL	1.197
FLA	1.175
TOR	1.132
OTT	1.091
COL	0.848
EDM	0.805
MIN	0.783
CGY	0.639
N.J	0.592
N.J*HOME	1.196
DAL*HOME	1.073
TOR*HOME	0.906

Table 2 : Adjusted HIT Counts for 2012-13 NHL Regular Season

Name	Team	Adjusted Hits	Raw Hits	Differential
M Martin	NYI	228.9	234	-5.1
C Neil	OTT	193.4	206	-12.6
S Ott	BUF	182.6	187	-4.4
L Schenn	PHI	181.8	187	-5.2
C Clutterbuck	MIN	181.3	155	26.3
L Smid	EDM	173.0	151	22.0
L Komarov	TOR	169.7	176	-6.3
D Backes	STL	159.9	158	1.9
R Clune	NSH	159.0	159	0.0
M Fraser	TOR	147.7	153	-5.3
E Kane	WPG	143.0	147	-4.0
Z Rinaldo	PHI	140.4	143	-2.6
M Lucic	BOS	137.4	139	-1.6
R Reaves	STL	137.0	135	2.0
D Brown	LAK	136.4	156	-19.6
K Clifford	LAK	135.8	155	-19.2
R Callahan	NYR	132.0	154	-22.0
J Petry	EDM	128.2	112	16.2
B Boyle	NYR	126.9	145	-18.1
D MacKenzie	CBJ	124.2	122	2.2

Player Evaluation

Joint Work with:

Dennis Lock (2008),
Matt Generous (2009),
James Curro (2012),
Jake Hurlbut (2015)

Thanks to C.J. Knickerbocker, Ed Harcourt (St. Lawrence U)

Context of Events (Again)

Player Evaluation:

- ▶ **Where shift starts (Zone Starts- ZS)** Move N-Zone to O-Zone \approx Replace Avg w/ Toews or Crosby
- ▶ **Quality of Teammates** Who with them on the ice (QoT)
- ▶ **Quality of Competition** Who against them on the ice (QoC)
- ▶ **Score Effects** What is current score (Score Effects): Teams that are ahead get better outcomes
- ▶ **Home Ice** Better events at home a la *Scorecasting*

Does QoC matter? Yes (events, shifts) and No (seasons)

Scale of these events matter!

REGRESSION

Advanced Models for Skater (Forwards & Defencemen) Evaluation:

- ▶ Gramacy, Taddy, Jensen (2013)
- ▶ Schuckers and Curro (2013)
- ▶ Macdonald (2012)
- ▶ Thomas, Ventura, Jensen, Ma (2013)
- ▶ Burtch (2014)

Questions:

- ▶ Response
- ▶ Covariates/Predictors
- ▶ Model form
- ▶ Link function
- ▶ Estimation (OLS, Lasso, Ridge)

Total Hockey Ratings

- ▶ Response = NP20
- ▶ Covariates/Predictors= Yes
- ▶ Model form = Linear
- ▶ Link function = Identity
- ▶ Estimation (OLS, Lasso, Ridge)=Ridge

Total Hockey Ratings(THoR)

$NP_k = P(\text{Goal for Home team in next } k \text{ secs after event}) - P(\text{Goal for Away team in next } k \text{ secs after even})$

Exceptions: SHOT and PENL

Been using $k=20$

NP_{20} for THoR

Event	Shot Type (if relevant)	Location	NP20
SHOT	Backhand	Off	0.1348
SHOT	Wrist	Off	0.1096
SHOT	Slap	Off	0.0697
TURN (to Home Team)		Off	0.0362
FAC		Off	0.0167
MISS	Wrist	Off	0.0159
HIT (by Home)		Off	0.0039
FAC		Neu	0.0026
HIT (by Home)		Neu	-0.0008
TURN (to Home Team)		Neu	0.0264
FAC		Def	0.0005
HIT (by Home)		Def	-0.0060

Terms in model:

- ▶ Home Ice
- ▶ Rink Effect
- ▶ Zone Start
- ▶ Score Effect
- ▶ Home Players
- ▶ Away Players
- ▶ PP, PK & interactions

Probability of Winning Given Out ___ Your Opponent Data from 2009 to 2013

Conditions	THoR	Corsi	Fenwick	Shots
5v5	0.519	0.530	0.461	0.406
5v5 within 2	0.537	0.568	0.520	0.452
5v5 within 1	0.573	0.591	0.580	0.493
5v5 tied	0.589	0.607	0.620	0.538

Year to Year Corr Player Ratings from 2009 to 2014 Seasons

Model	1 Yr	2 Yr	3 Yr
Adj Fenwick	0.47	0.41	0.41
Adj Corsi	0.33	0.32	0.28
THoR	0.80	0.76	0.77

Using THoR Model with Corsi and Fenwick as 1,-1 response

Top 10 Total Value 2011-14

Adjusted Corsi:

PBergeron, CKunitz, JWilliams, JBoychuk, MGiordano,
PDatsyuk, PHornqvist, SWeber, DSedin, JGorges

Adjusted Fenwick:

JWilliams, DSedin, PDatsyuk, PBergeron, ASteen,
PHornqvist, JThornton, LCouture, EKarlsson, GLandeskog

THoR:

AKopitar, BRichards, EKarlsson, AOvechkin, JThornton,
MDuchene, PBergeron, EStaal, DSedin, JEberle

None of these include SH%

Recent

- ▶ Microstat Tracking
- ▶ Regression
- ▶ Adjust for more factors

Future

- ▶ New Data (SportVU), likely proprietary
- ▶ Will Be Smooth & **Multivariate**
- ▶ More Regression
- ▶ Adjust for more factors
- ▶ Is Here

Thank You

`schuckers@stlawu.edu`

Links

- ▶ <http://www.habseyesontheprize.com/2013/2/20/4005122/how-reliable-is-the-nhl-com-shot-tracker>
- ▶ <http://www.coppernblue.com/2011/10/17/2495447/dear-oilers-find-new-official-scorers>
- ▶ <http://www.arcticicehockey.com/2010/10/18/1756880/clean-up-your-act-madison-square-garden>
- ▶ <http://www.coppernblue.com/2010/3/30/1396334/rtss-stats-giveaways-takeaways-and>
- ▶ <http://www.inlouwetrust.com/2010/7/8/1559914/blocked-shots-the-new-jersey>
- ▶ <http://www.puckprospectus.com/article.php?articleid=351>