Modern Hockey Statistics: An Introduction

Michael Schuckers

schuckers@stlawu.edu

St. Lawrence University

Statistical Sports Consulting, LLC

What is Statistics

Webster's New Collegiate Dictionary

statistic: a single term or datum in a collection

Statistics: a branch of mathematics dealing with the collection, analysis, interpretation and presentation of masses of numerical data

analytics: **new term**

essentially to extract meaning from masses of numerical data

Hockey Statistics

This talk

Some new statistic(s)

Some Statistics/Analytics

Note some insights that ‘Hockey Statistics’ has gleaned (including some of my own work)

Statistics/Analytics in Sports

Where hockey is, analytically/statistically:

Behind Baseball,

Basketball,

Football (both European and American)

2014 Hockey's 'Summer of Analytics'

Hirings in New Jersey, Carolina, Toronto, Washington, Florida, Edmonton

Other active teams: Chicago, Los Angeles, Tampa Bay, San Jose?, Nashville?

http://espn.go.com/espn/feature/story/_/id/12331388/the-great-analytics-rankings

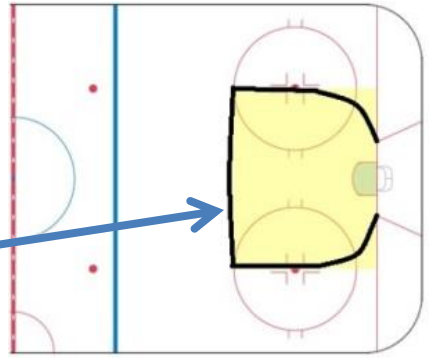
Hockey

- Distinguishing Features
  - Very fluid (think basketball, soccer)
  - Players on ice for short bursts (40 seconds)
  - Penalties result in team being without a players for length of penalty (but can return in some cases)
  - OT (in NHL) played different than rest of the game (4v4)
  - Tie at end of OT in regular season ends in shootout

Some terms

Even Strength(EV) : Both teams have 5 players on the ice

Scoring Chance: Generally



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

IMAGE: http://www.broadstreethockey.com/2011/11/1/2526089/scoring-chance-summary-games-1-11

Some terms

Points:

For Teams, Points = 2 for Win, 1 for OTL, 0 for regulation loss

For Players, Points =1 for Goal or Assist

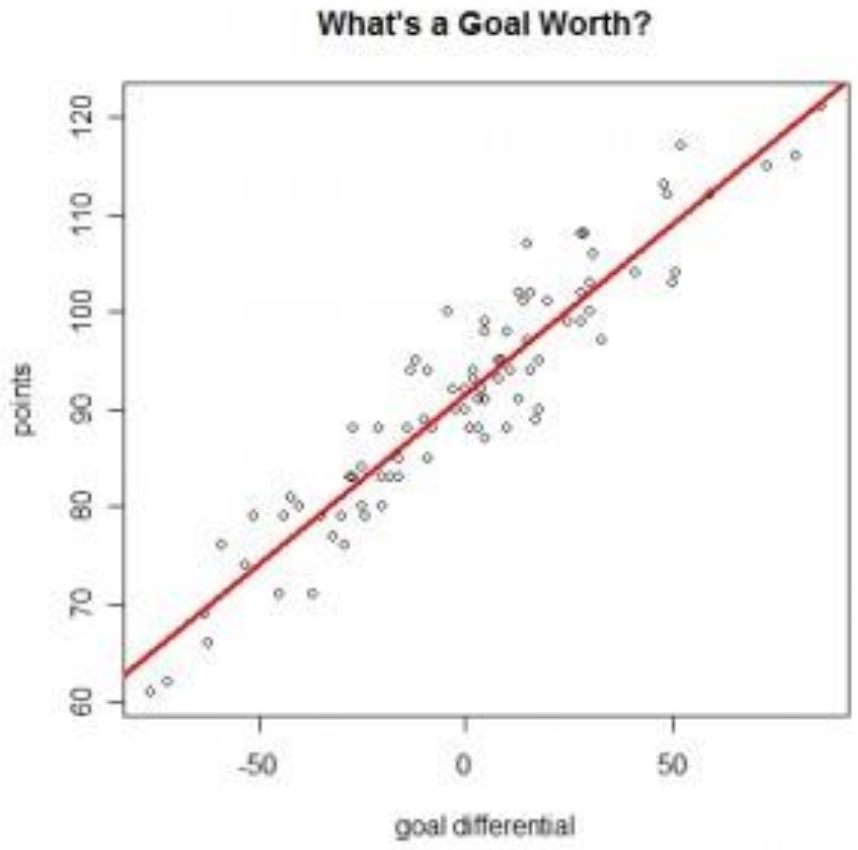
SV% is Save Percentage and is percentage of shots faced that are not goals

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.

Worth of a goal:
Goal Diff vs Points (corr=0.94)



$$\widehat{Pts} = 91.6 + 0.33\ GOAL\ DIFF$$

What's a Goal Worth

In Wins

$$\widehat{Pts} = 91.6 + 0.33\ GOAL\ DIFF(GD)$$

- 0.33 => +1 GD = 1/3 pt, 3 GD's = 1pt
- 6 GD's = 2pts = 1 win

In Dollars

- Slope = (59.4-12.6)/(91-52) = $1.2MM/point
- ~ $0.4MM/Goal
- ~ $2.4MM/Win (Done in 2010/11 likely gone down)

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

Corsi/Fenwick

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)

NHL added these to it's website Feb. 23, 2015.

- Shots Attempted (SAT), i.e. Corsi
- Unblocked Shots Attempted (USAT), i.e. Fenwick
- http://espn.go.com/nhl/story/_/id/12355124/nhl-unveils-new-enhanced-stats-element-official-website

Corsi/Fenwick/SAT/USAT

Even Strength (5v5) Team Corsi % and Fenwick % correlate Close (0,1 goal margin) moderately with

- future win percentage
- future goal scoring rates
- (Good predictors of future performance)

Corsi # and Fenwick # correlate with Scoring Chances #

Also proxies for possession?

2013-14 Regular Season Fenwick % Close (w/in 1 goal)

Rank	Team	Fenwick%	Standings	Finish
1	LA Kings	56.7%	W6	Stanley Cup Champions
2	Chicago	55.2%	W5	West Finals (lost to LA Kings)
3	San Jose	54.6%	W4	Lost 1st round to LA Kings
4	Boston	54.1%	E1	Lost 2nd round to Montreal
5	New Jersey	53.9%	E10	Did not make playoffs
6	New York Rangers	53.5%	E5	Lost in Stanley Cup Finals
…				
12	Ottawa	50.8%	E11	Did not make playoffs
…				
15	Pittsburgh	50.2%	E2	Lost to NY Rangers (2nd round)
….				
30	Buffalo	41.0%	E16	Did not make playoffs

2014-15 Regular Season Fenwick % Close (w/in 1 goal)

Rank	Team	Fenwick%	Finish (Standings 3/18/15)
1	New York Islanders	55.0%	??? (E4)
2	Tampa Bay	54.0%	??? (E3)
3	Winnipeg	53.6%	??? (W7)
4	Pittsburgh	53.3%	??? (E5)
5	Nashville	53.1%	??? (W3)
6	Chicago	52.8%	??? (W4)
7	Los Angeles	52.7%	??? (W8)
8	Boston	52.3%	??? (E8)
…			
21	Ottawa	49.4%	??? (E9)
23	New Jersey	47.6%	??? (E12)
30	Buffalo	37.2%	??? (E16)

Keeping the puck in the offensive zone

May 10, 2013

- http://video.nhl.com/videocenter/console?id=247757&catid=35&startTime=0

http://video.nhl.com/videocenter/console?id=247757

PDO

PDO = Team SV% + Team SH%

SPSv% (Shooting% Plus Save % on NHL.com)

Team SV% tends to regresses to goalie(s) career SV% (and league average SV%)

Team SH% tends to regresses to league mean SH%

PDO is generally considered a measure of luck

Regression to the mean

Idea:

Extreme values tend to be more similar to averages over time

Results = skill + randomness

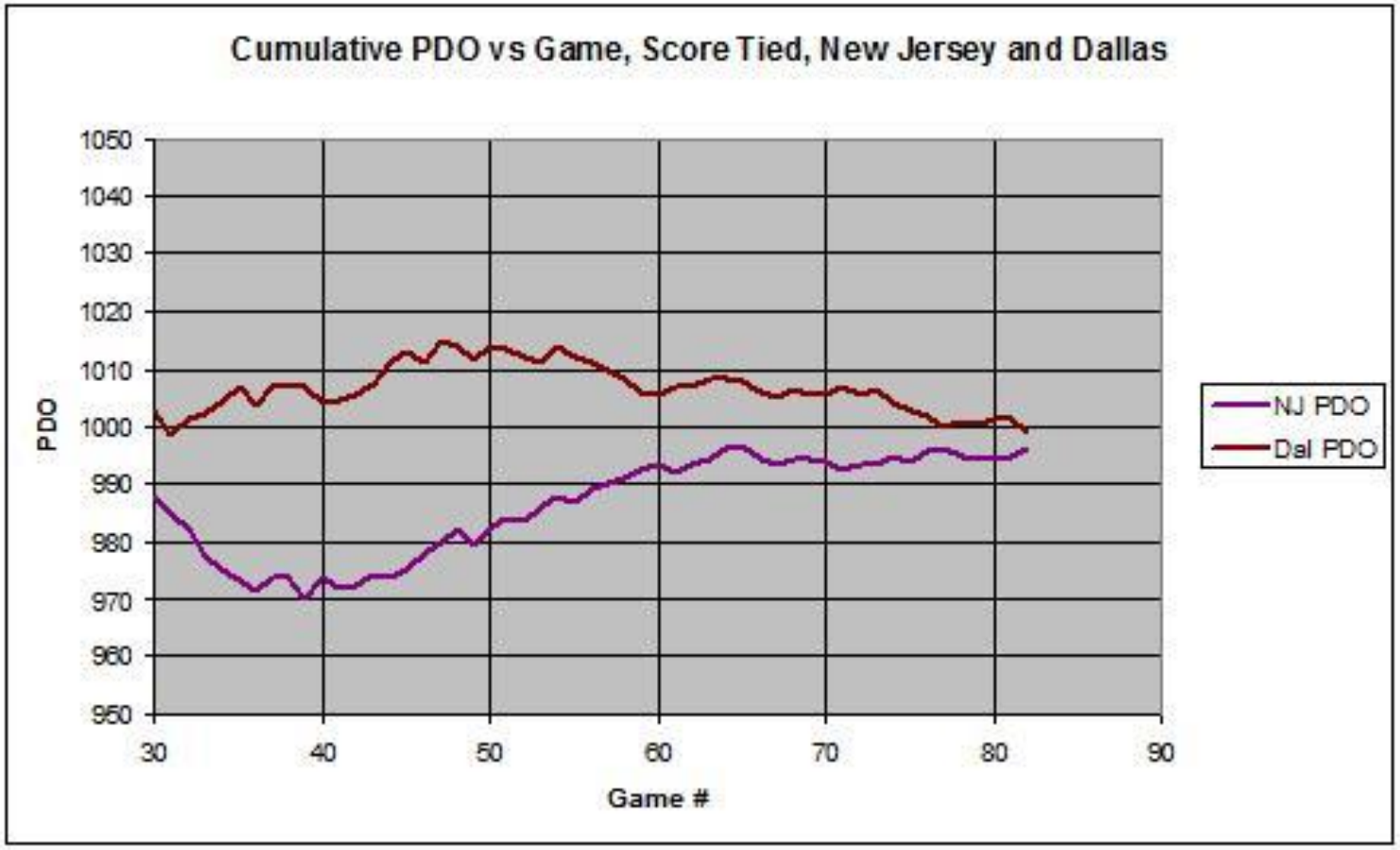
- Very high end (*very low end*) results due to good (*bad*) results in both skill and luck.
- Randomness, i.e. luck/chance evens out over time.

Examples:

Sophomore Jinx

Sports Illustrated Cover Jinx

PDO: Classic Example
2010-11 NJD, DAL



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

2013-14 PDO/SPSv%

Rank	Team	SPSv%	Playoff Seed	Results
1	Anaheim	103.4	W1	Lost 2nd round
2	Boston	102.8	E1	Lost 2nd round
3	Colorado	102.2	W2	Lost 1st round
4	Toronto	101.4	N/A	
5	Montreal	101.4	E4	Lost East. Finals
6	Detroit	101.4	E8	Lost 1st round
7	Tampa Bay	101.0	E3	Lost 1st round
…				
27	New Jersey	98.6	N/A	
28	New York Islanders	97.9	N/A	
29	Buffalo	97.9	N/A	
30	Florida	95.9	N/A	

Source: Stats.hockeyanalysis.com

2014-15 PDO/SPSv%

Rank	Team	SPSv%	Standings (3/18/15)
1	New York Rangers	102.1	E1
2	New Jersey	101.4	E12
3	Pittsburgh	101.1	E5
4	Nashville	101.1	W3
5	St. Louis	101.0	W1
6	Montreal	100.9	E2
…			
10	Buffalo	100.5	E16
…			
16	Winnipeg	100.1	W7
17	Chicago	100.0	W4
18	Los Angeles	99.9	W8
…			
30	Edmonton	97.5	W14

Source: Stats.hockeyanalysis.com

Compared to what?

Shots Attempted (SAT, Corsi)% and Unblocked Shots Attempted (USAT, Fenwick)% when game is close (within one goal)

Better predictors future performance in NHL than

- Goal %
- Shots on Net%
- Past Winning %

These findings are robust and have been shown many times in multiple circumstances and many years.

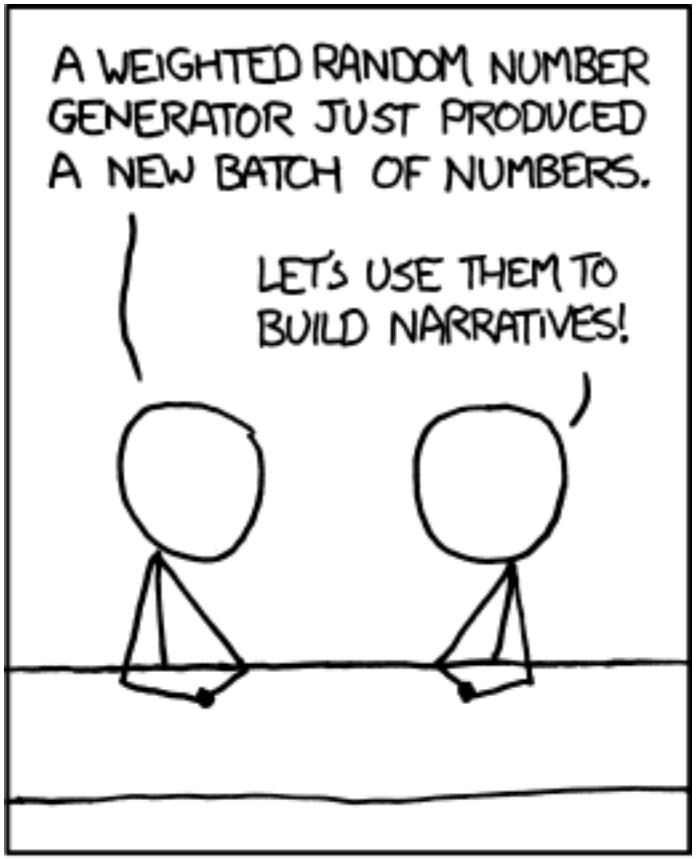
Same with PDO/SPSv

What's missing

Many things:

- Shot Quality
- Penalties
- Quality of PP
- Quality of Penalty Kill
- Quality of Goaltending
- Shooting Pct

Results = Skill + Randomness



Source: http://xkcd.com/904/

It's a binomial (0-1, no goal - goal) game

Hockey is an binomial, np, game

n is number of shots

(m = # of shots for opponent)

p is probability of shot = goal

(r = probability a shot = goal for opponent)

More precisely,

Goal Differential = $n_{EV}p_{EV} + n_{PP}p_{PP} + n_{SH}p_{SH}$
$-m_{EV}r_{EV} - m_{PP}r_{PP} - m_{SH}r_{SH}$

Shot Quality

More precisely,

Goal Differential = $n_{EV}p_{EV} + n_{PP}p_{PP} + n_{SH}p_{SH}$

$$-m_{EV}r_{EV} - m_{PP}r_{PP} - m_{SH}r_{SH}$$

Two parts: n's and p's

Corsi/Fenwick drive n

Can we drive up $p\_$? (and drive down $r\_$?)

Shot Quality

Big Debate:

Define Shot Quality as Average SH%

Shot Quality matters a some, a little or very, very little

Positions Matter: Defensemen (lower SH%), Forwards (higher SH%)

Team SH% regresses to league average

Individual SH% regresses to position average but more slowly

Interjection: Data Quality is an issue

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 Giveways vs Takeaways

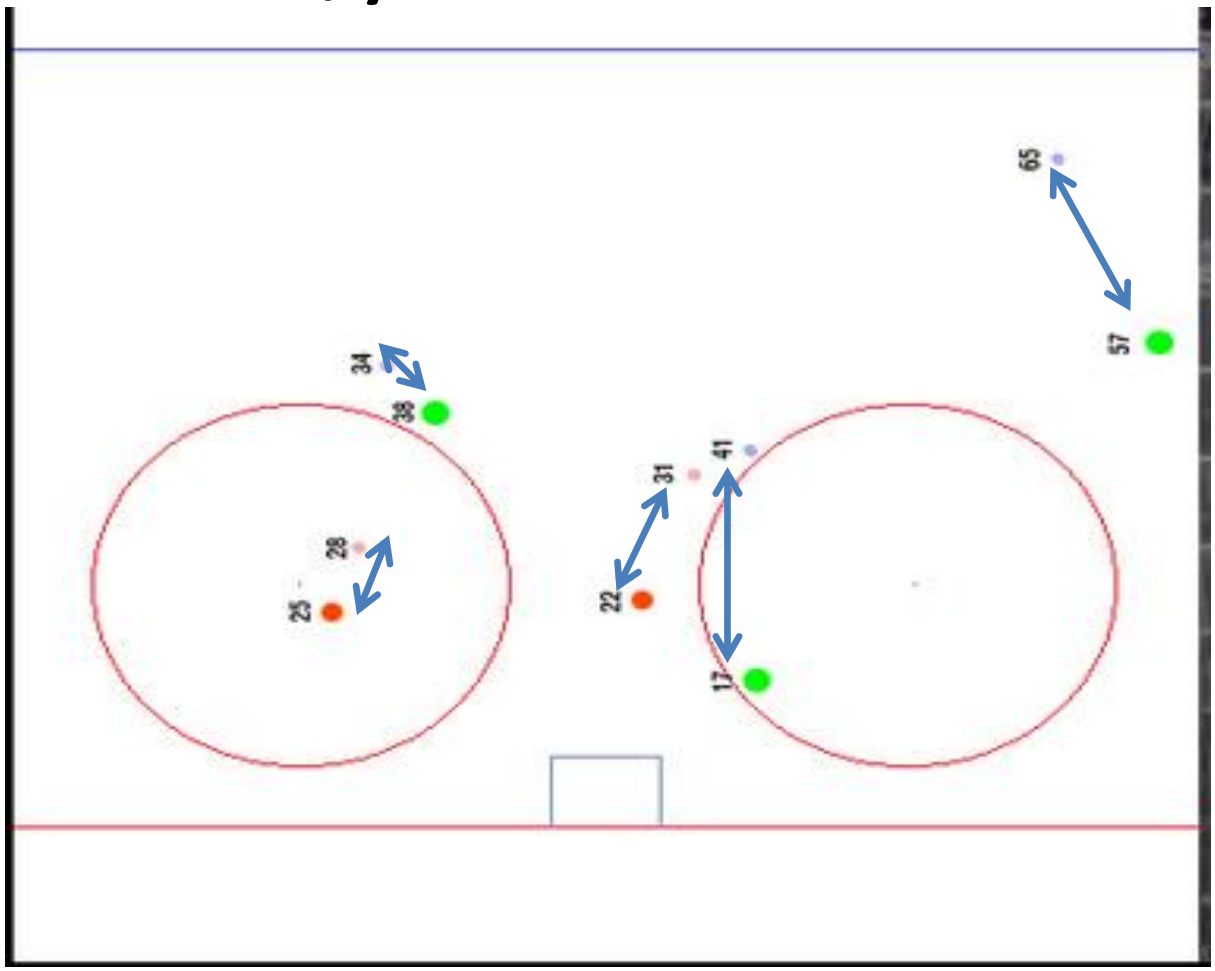
HITS inconsistent Rink to Rink, etc.

Count totals for EVENTS need additional quality control

Video based data is coming (cf. Basketball)

http://video.nhl.com/videocenter/console?id=734244

x,y data issues



http://www.habseyesontheprize.com/2013/2/20/4005122/how-reliable-is-the-nhl-com-shot-tracker

Play By Play File: TOR at BOS

326	3	EV	9:49
10:11	FAC	BOS won Off. Zone - TOR #42 BOZAK vs BOS #46 KREJCI	
327	3	EV	10:00
10:00	HIT	TOR #3 PHANEUF HIT BOS #18 HORTON, Def. Zone	
328	3	EV	10:06
9:54	GIVE	BOS GIVEAWAY - #17 LUCIC, Off. Zone	
329	3	EV	10:41
9:19	HIT	TOR #3 PHANEUF HIT BOS #17 LUCIC, Def. Zone	
330	3	EV	10:55
9:05	SHOT	BOS ONGOAL - #33 CHARA, Wrist, Off. Zone, 56 ft.	
331	3	EV	11:10
8:50	BLOCK	BOS #33 CHARA BLOCKED BY TOR #41 KULEMIN, Deflected, Def. Zone	
332	3	EV	11:12
8:48	GOAL	BOS #33 CHARA(1), Wrist, Off. Zone, 41 ft. Assists: #46 KREJCI(6); #44 SEIDENBERG(1)	

Source: http://www.nhl.com/scores/htmlreports/20122013/PL030145.HTM

Interjection: Data Quality is an issue

Correctable in nearly all cases

- Different rates of events in different rinks
  - Schuckers & Macdonald (2014)
  - Regression on rates of events

- x,y coordinates for shots
  - Differences from rink to rink
  - Some preliminary work in Schuckers & Curro (2013…)

- Video tracking based data is coming (cf. Basketball)

so far …

Value of a win: GD ~ 1/3 pt, Wins cost ~ $1.2M

Corsi & Fenwick predict future goal differential and wins

PDO measures luck

Average Shot Quality might be a small factor

Player Evaluation

- Corsi/Fenwick designed to be team evals
- Used for player evals
  - CorsiRel% = Corsi% (on ice) – Corsi% (off ice)
  - Most typical conventional wisdom
  - E.g. Erik Karlsson, CorsiRel% = 3.6% (2014-15, 5v5 close)

CorsiRel% (2013-14)

Player	Team	CorsiRel%
Backlund, Mikael	Calgary	9.9
Giordano, Mark	Calgary	9.6
Bergeron, Patrice	Boston	8.8
Thornton, Joe	San Jose	8.5
Brodie, T. J.	Calgary	8.0
Crosby, Sidney	Pittsburgh	8.0
Koivu, Mikko	Minnesota	7.5
Kopitar, Anze	Los Angeles	7.5
Stralman, Anton	New York Rangers	7.5
Jagr, Jaromir	New Jersey	7.4

Source: Stats.hockeyanalysis.com, behindthenet.com

CorsiRel% (2014-15 to 3/22/15)

Player	Team	CorsiRel%
Bergeron, Patrice	Boston	9.3
Thornton, Joe	San Jose	8.9
Campbell, Brian	Florida	8.8
Jagr, Jaromir	Florida/NJ	8.3
Datsyuk, Pavel	Detroit	8.0
Koivu, Mikko	Minnesota	7.9
Marchand, Brad	Boston	7.5
Subban, P.K.	Montreal	7.4
Mackinnon, Nathan	Colorado	7.2
Landeskog, Gabriel	Colorado	7.0

Source: Stats.hockeyanalysis.com, behindthenet.com

Context Matters in Player Evaluation

Where shift starts (Zone Starts- ZS)

- Move N-Zone to O-Zone ≈ Replace Avg w/ Toews or Crosby

Who with them on the ice, Quality of Teammates (QoT)

Who against them on the ice, Quality of Competition (QoC)

What is current score (Score Effects)

- Teams that are ahead get better outcomes
- Teams play different with a lead (less risky??)
- Score matters more in 3rd (interaction)

Home or Away

- Better events at home a la *Scorecasting*

Adjusting for Context

Broader Hockey Analytics Community

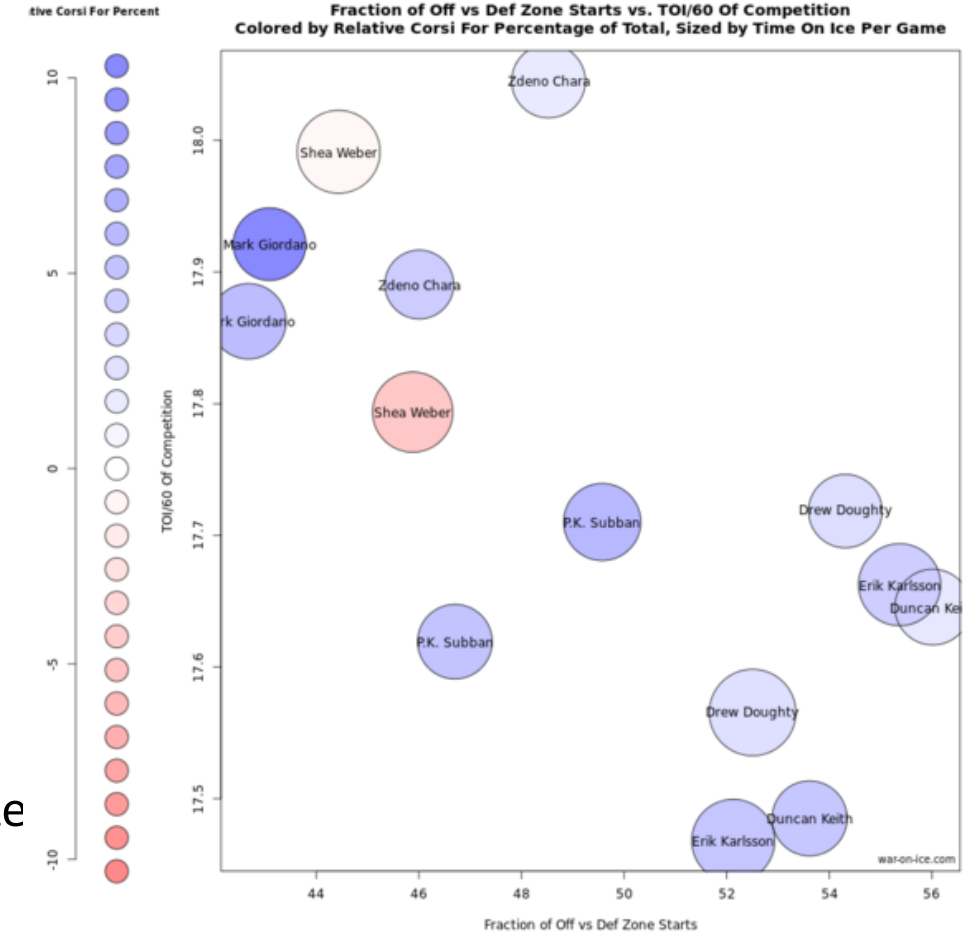
X-axis=% of Off Zone Starts

Y-axis = Quality of Competition

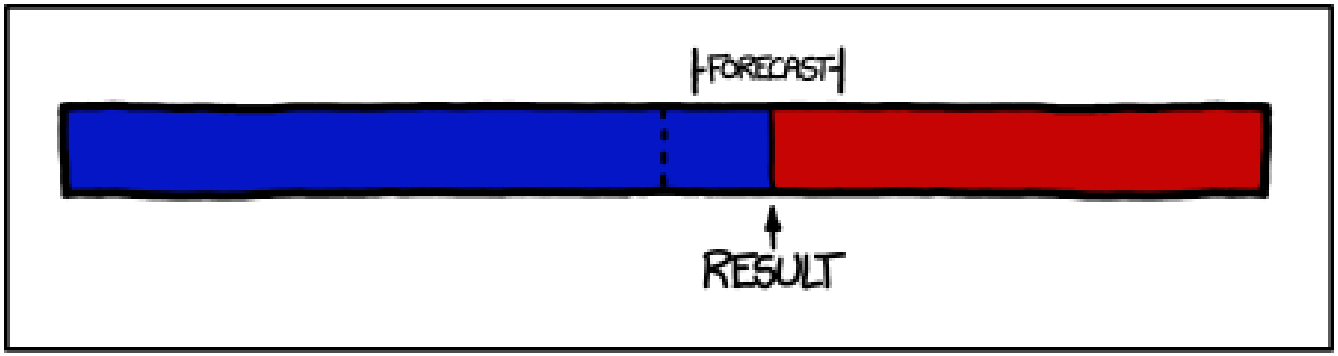
Size of Bubble = Time on Ice

Color of Bubble = CorsiRel%

CorsiRel% preferred on sample size



http://flamesnation.ca/2015/3/16/why-mark-giordano-should-still-be-in-the-norris-conversation



BREAKING: TO SURPRISE OF PUNDITS, NUMBERS CONTINUE TO BE BEST SYSTEM FOR DETERMINING WHICH OF TWO THINGS IS LARGER.

Source: https://xkcd.com/1131/

Statistical Method

Regression based methods

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Y is outcome (response, dependent variable)

X_i 's are predictors (explanatory variables, independent variables)

Adjusting for Context

Statistical Model (Even Strength):

Outcome ~ mean effect + home effect

+ all the players on the ice +

+ zone start + score effect +

Outcome is Corsi or Fenwick (+1,-1)

'Big Data' (97K Fenwick events per season, 130K Corsi, 1200 players)

Based upon Schuckers and Curro (2013), Hurlbut (201?)

Ridge Regression (account for Multicollinearity)

Adjusted Fenwick (Even Strength, 2011-2014)

Name	Team	Position	Counts	Adjusted Fenwick
Patrice Bergeron	Boston Bruins	C	3486	3.40%
Justin Williams	Los Angeles Kings	R	3537	3.24%
Daniel Sedin	Vancouver Canucks	L	3307	3.01%
Pavel Datsyuk	Detroit Red Wings	C	2623	2.75%
Alexander Steen	St. Louis Blues	C	2549	2.75%
Joe Thornton	San Jose Sharks	C	3955	2.66%
Anze Kopitar	Los Angeles Kings	C	3900	2.65%
Jonathan Toews	Chicago Blackhawks	C	3195	2.46%
Marc-Edouard Vlasic	San Jose Sharks	D	4579	2.45%
Patric Hornqvist	Nashville Predators	R	2938	2.45%

Adjusted Fenwick (All Events, 2011-2014)

Name	Team	Position	Counts	Adjusted Fenwick
Justin Williams	Los Angeles Kings	R	4237	3.67%
Daniel Sedin	Vancouver Canucks	L	4291	3.33%
Patrice Bergeron	Boston Bruins	C	4512	3.21%
Pavel Datsyuk	Detroit Red Wings	C	3466	3.16%
Joe Thornton	San Jose Sharks	C	5192	2.74%
Alexander Steen	St. Louis Blues	C	3361	2.70%
Patric Hornqvist	Nashville Predators	R	3482	2.70%
Anze Kopitar	Los Angeles Kings	C	5296	2.51%
Chris Kunitz	Pittsburgh Penguins	L	4863	2.48%
Logan Couture	San Jose Sharks	C	4395	2.41%

Adjusted Fenwick

Things we can learn about Fenwick from our model:

Change where shift starts: Neu Zone to Off Zone +4%

Home team gets about +2%

Team moves lead from k goal to k+1 goals (k>1) -1%

(Similar results for all adjusted models we consider here.)

-

Adjusted Fenwick over Adjusted Corsi

Validity

Fenwick slightly better

Reliability

Corsi larger sample size

Our analysis, out of sample year to year correlation

	Adj.Fenwick	Adj.Corsi
Even	0.47	0.28
All	0.47	0.29

Adjusting for Context

Statistical Model (Even Strength):

Outcome ~ mean effect + home effect

+ all the players on the ice +

+ zone start + score effect +

Outcome is Net Expected Goals every event

'Big Data' (250K events per season, 1200 players)

Schuckers and Curro (2013), Hurlbut (201?)

Total Hockey Rating (THoR)

THoR:

Net Expected Goals (NP20) =P(Home Goal in 20 sec) – P(Away Goal in 20 sec)

EV model, by event, ~250000 events per season, 1200 players/season

Ridge Regression, PENL, TURN, x,y gridded adjustment CDF

EV/PP/SH model to come….

NP20 (EV)

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

THoR Top Players, Even, 2011-14

Name	Team	Position	# Events	THoR Wins Over Average
Anze Kopitar	Los Angeles Kings	C	13042	13.29
Patrice Bergeron	Boston Bruins	C	11918	8.64
Gabriel Landeskog	Colorado Avalanche	L	9623	8.58
Joe Pavelski	San Jose Sharks	C	11781	8.27
Chris Kunitz	Pittsburgh Penguins	L	9541	8.21
Dan Hamhuis	Vancouver Canucks	D	11164	7.68
Logan Couture	San Jose Sharks	C	10551	7.56
Marc-Edouard Vlasic	San Jose Sharks	D	12201	7.48
Henrik Sedin	Vancouver Canucks	C	11485	7.40
Erik Karlsson	Ottawa Senators	D	12770	7.33

Even Strength, Wins Over Average (Exp Goal Diff per Event * # of Events adjusted by pos)
Ridge Regression deal with multicollinearity (linemates)

THoR Top Players, Even, 2011-14

Name	Team	Position	# Events	THoR Wins Over Average
Anze Kopitar	Los Angeles Kings	C	16809	19.85
Joe Thornton	San Jose Sharks	C	15006	17.25
Chris Kunitz	Pittsburgh Penguins	L	11782	16.92
Jordan Eberle	Edmonton Oilers	C	11176	16.23
Logan Couture	San Jose Sharks	C	12904	15.75
Joe Pavelski	San Jose Sharks	C	15267	15.12
Jakub Voracek	Philadelphia Flyers	R	10553	14.61
Patrice Bergeron	Boston Bruins	C	14979	14.17
Corey Perry	Anaheim Ducks	R	11999	14.14
Tyler Seguin	Dallas Stars	C	11663	14.04

Even Strength, Wins Over Average (Exp Goal Diff per Event * # of Events adjusted by pos)
Ridge Regression deal with multicollinearity (linemates)

THoR versus Adjusted Fenwick

Validity

Fenwick slightly better

Reliability

THoR better especially with all events

Our analysis, out of sample year to year correlation

	Adj.Fenwick	THoR
Even	0.47	0.49
All	0.47	0.79

Rating Goalies

Shot Types	Hard	Easy	SV%
Goalie A	88% (n=500)	92%(1500)	0.910
Goalie B	90%(1750)	93%(250)	0.904

Who's the better goalie??

How to fix this?

Rating Goalies

Shot Types	Hard	Easy	Adjusted SV%
Goalie A	88% (**1125**)	92%(**875**)	0.8975
Goalie B	90%(**1125**)	93%(**875**)	0.9131

Adjusted SV% to account for difficulty of shots (in the league)

Most analysts believe that difference in average Quality of shots faced by teams does not differ Much.

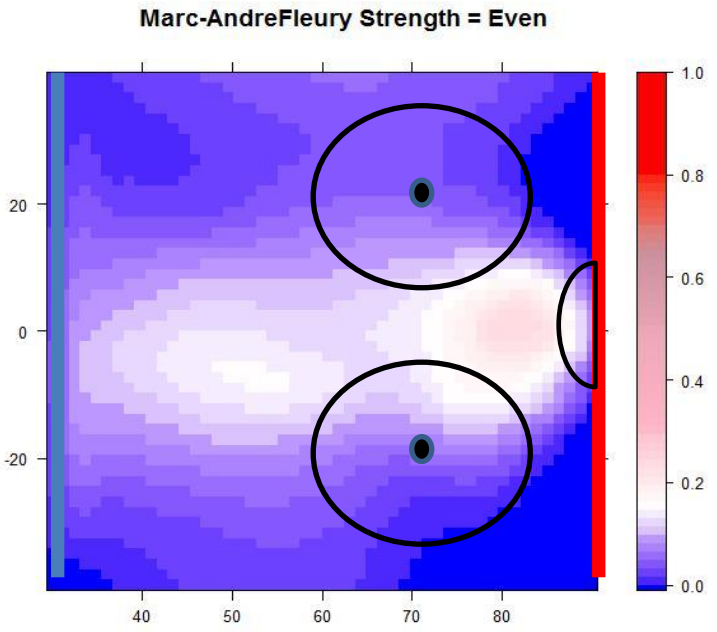
Shot Quality

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

- Goalie Ability
- Shots Distribution Faced

DIGR maps goal probabilities spatially (x,y) all shot types (slap, snap, etc) for all goalies then finds expected SV% for all shots taken for all shots in a given year

There are issues with NHL Data (x,y)'s
Time to revisit….



Recap

GD ~ 1/3 pt, Wins cost ~ $1M

Corsi & Fenwick predict Goal Diff and wins

PDO = (SV% + SH%)/100 and is proxy for luck

Context Matters (ZS, QoT, QoC, Home, Rink, Score)

More sophisticated methods exist (THoR, etc.)

Shot Quality a small factor, Hockey more n than p

Additional Topics

- Pull the goalie more….Beaudoin and Swartz
- Drawing PENL's seems to be repeatable
- Average – Replacement = 0.5 wins
- Referees are robots (Schuckers 2012)
- Draft Value Pick Charts exist (Schuckers 2011, Schuckers and Argeris 2015)
- Central Scouting does good ($3 to 8 M/yr) job
- ~60 Faceoff wins = 1 goal
- Shootouts are crapshoots
- League Translations (AHL to NHL, SM-Liiga to NHL)

Great news on campus

SLU has a new Statistics major!

Approved by Faculty 2/2015

Endorse by Board of Trustees 2/2015

Pending approval by New York State

Some Links

Index of Hockey Papers

www.statsportsconsulting.com/hockey-index/

Hockey Postings

www.statsportsconsulting.com/category/hockey

Thank you!

schuckers@stlawu.edu

@SchuckersM

@EmpiricalSports

Special Thanks to Chris Wells, Collen Knickerbocker, Dennis Lock, Ed Harcourt, Matt Generous, Jim Curro, Tom Pasquali, Zach Nelson, Jake Hurlbut, Danny Driscoll, Brian Macdonald