Expected Goals: Blending Shot Quantity and Quality To Evaluate Teams and Players

Sam Ventura

Department of Statistics
Carnegie Mellon University
www.war-on-ice.com
@stat_sam

2015 Ottawa Hockey Analytics Conference
February 7, 2015

The State of the Art

Corsi (CF%): Most predictive of future Goals-For% (GF%) for **teams**

Fenwick (FF%): Most predictive of future GF% for **defensemen**

Scoring Chances (SCF%): Most predictive of future GF% for **forwards**

What is Corsi, and why is it useful?

A team's Corsi (CF%) is the **percentage of all shot attempts** (blocked, missed, on goal) that are **directed towards the opposing team's net**

CF% uses **all shot attempts**, meaning more information is **taken into account** than when just looking at goals or shots.

What are scoring chances, and why are they useful?

A scoring chance is a **shot-attempt** that has a **high probability** of resulting in a goal

SCF% **focuses on the most important events** we observe – those with $P(ShotAttempt = Goal)$ above some threshold

Idea: Combine the best of CF% and SCF%

Issue with Corsi: **"All events are created equal"**

Issue with SCF: **"Throw away events that aren't useful"**

Idea: **Weight all events by their usefulness**

This is not a novel idea

Very similar to "weighted shots" (Macdonald et al, 2012)

Other similar approaches: Ryder (2004), Johns (2004), Krzywicki (2005, 2009, 2010), Awad (2009), Schuckers (2011), and probably several others

Designated Math Slide

Let $Y_i \in \{0, 1\}$. Then, the expected value of Y_i is:

$$E(Y_i) = \sum_{y=0}^{1} y * P(Y_i = y) = 0 * P(Y_i = 0) + 1 * P(Y_i = 1) = P(Y_i = 1)$$

Now, let Y_i be a random variable that counts how many goals occur on the i^{th} shot attempt, i.e.:

$$Y_i = 1 \text{ if goal, } Y_i = 0 \text{ if non-goal.}$$

Then $E(Y_i) = P(\text{Shot Attempt } i = \text{Goal})$

Now, aggregate across a group of N shot attempts:

$$E(\sum_{i=1}^{N} Y_i) = \sum_{i=1}^{N} P(ShotAttempt_i = Goal)$$

In other words, a teams' or players' **"expected goals for"** is equal to the **sum of the goal-probabilities** of all their on-ice shot-attempts

Estimating the Probabilities

In order to estimate the probabilities, I use:

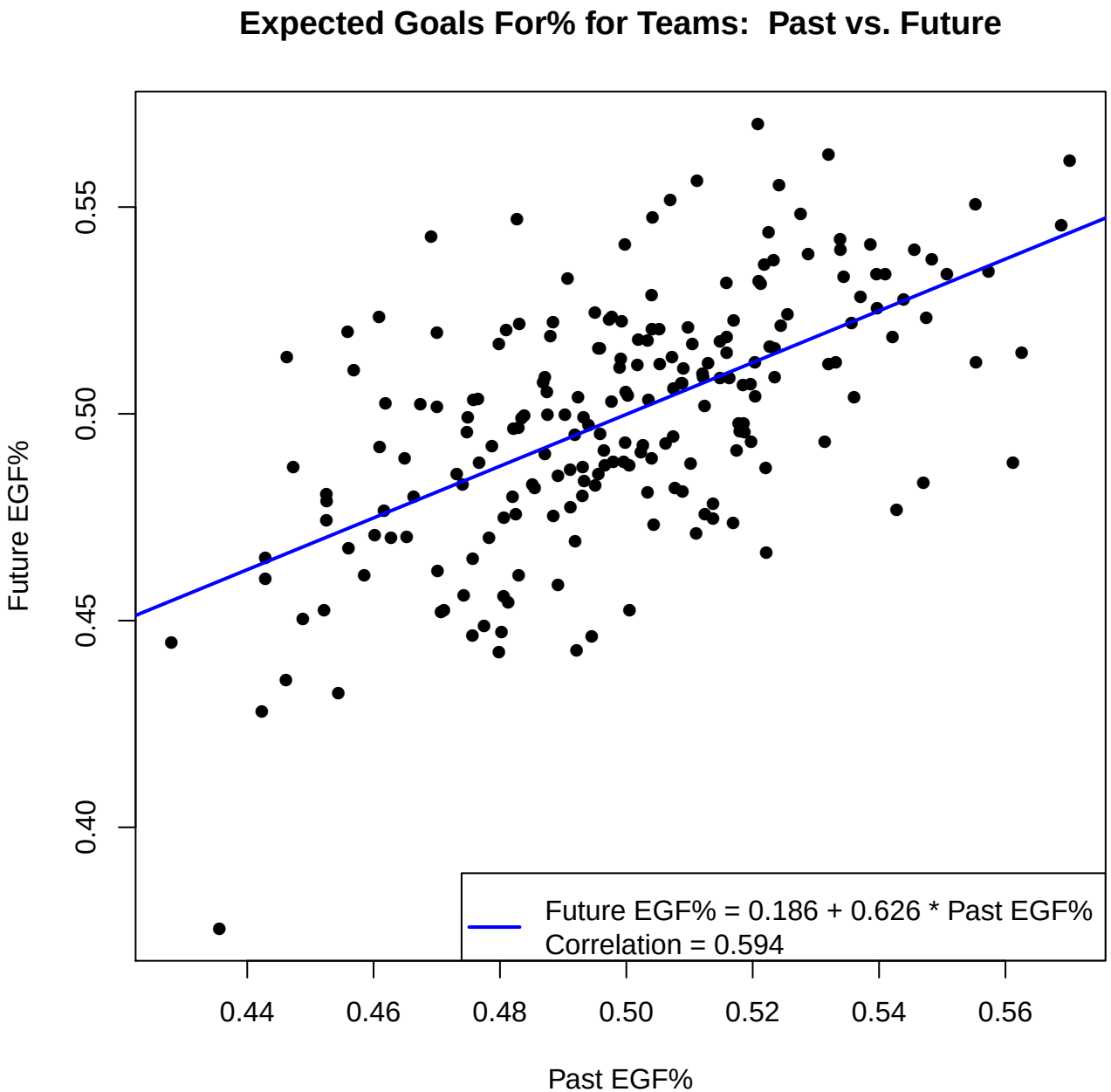
1. Model: Logistic Regression
2. Data: All shot attempts in the NHL for a given season
3. Explanatory Variables:
   - Location (x,y)
   - Distance From Goal
   - Shot Type (wrist shot, backhand, etc)
   - Shot feature (rush, rebound)
   - Some transformations and interactions of the above variables
4. Response Variable: Goal (1 if yes, 0 if no)

Result: EGF and EGA

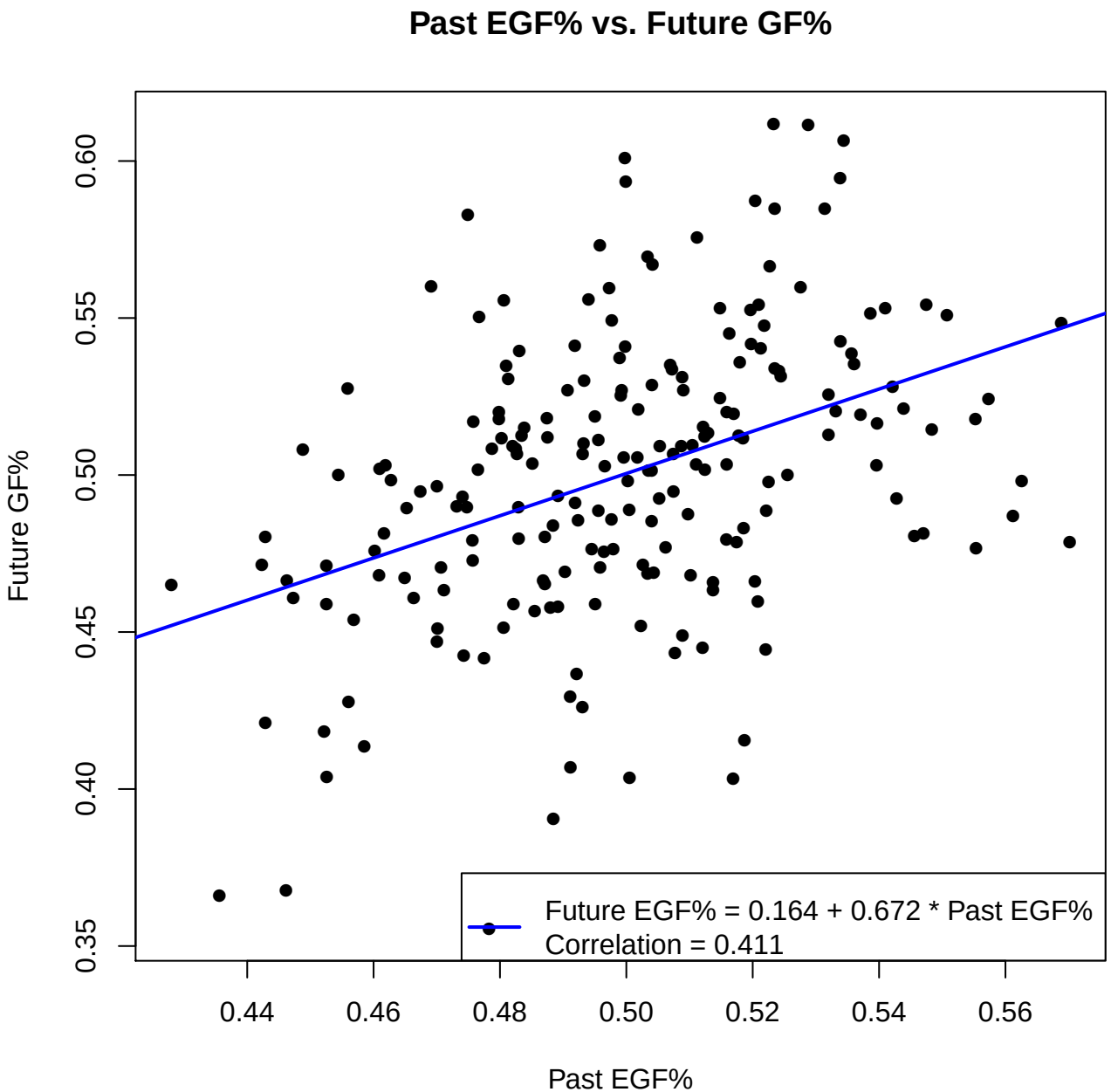
We can calculate the following for any team or player:

- EGF = Expected Goals For
- EGA = Expected Goals Against
- Natural extensions: EGF/60, EGA/60, EG+/-, EGF%, EGF% Rel

Are Expected Goals Predictive of Future Expected Goals?



Are Expected Goals Predictive of Future Goals For?



Best and Worst EGF% Teams, 2014-15 Season

	EGF	EGA	EG+/-	EGF%
BUF	68.77	114.50	-45.73	0.38
COL	81.24	109.37	-28.13	0.43
TOR	88.74	109.57	-20.84	0.45
CGY	83.21	99.77	-16.56	0.45
CBJ	69.79	79.17	-9.39	0.47
EDM	82.65	93.38	-10.74	0.47
...	...	...	...	...
CHI	99.95	87.10	12.86	0.53
NYI	102.18	88.22	13.96	0.54
NSH	90.26	77.20	13.06	0.54
MIN	94.18	80.13	14.06	0.54
DET	87.05	71.04	16.01	0.55
T.B	100.95	81.14	19.81	0.55

Passes the three laws of modern hockey statistics!

Best EGF+/- Players

(see online for player tables/charts)

Conclusions

Expected goals combine advantages of Corsi and Scoring Chances

For teams, orrelation with future GF% = 0.411

For teams, correlation of Corsi% with future GF% = 0.408

Future Work

Calculate EGF, EGA, EGF%, EG+/-, etc for players

Test predictivity of EGF% for players

???????

Get hired

Questions?

Contact me!

Email: sventura@stat.cmu.edu
Twitter: @stat_sam

Thanks!