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 $P(\text{ShotAttempt} = \text{Goal})$ 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
Let $Y_i \in \{0, 1\}$. Then, the expected value of $Y_i$ is:

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

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:

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

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
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?

Expected Goals For% for Teams: Past vs. Future

Future EGF% = 0.186 + 0.626 * Past EGF%

Correlation = 0.594
Are Expected Goals Predictive of Future Goals For?

Past EGF% vs. Future GF%

Future EGF\% = 0.164 + 0.672 \times \text{Past EGF}\% 

Correlation = 0.411
### Best and Worst EGF% Teams, 2014-15 Season

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

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, correlation 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!