Threat

Micah Blake McCurdy
hockeyviz.com

Vancouver, BC
Vancouver Hockey Analytics Conference
March 3, 2018
Introduction

Tired: Numbers
Wired: Pictures
Introduction

Vancouver with the Sedins at 5v5 generate 45 unblocked shots per hour, 5% more than league average.
Introduction

Tired: Numbers
Wired: Pictures
Aim

Isolate individual skater impact on shots, both for and against.
New Thing

Treat maps as first-class objects, instead of single-numbers like rates or counts.
Isolation

Control for the most important aspects of play which are outside of a player’s control:

- Linemates
- Zone usage
- The score (!)
- Competition faced.
Isolation

Control for the most important aspects of play which are *outside* of a player’s control:

- Linemates
- Zone usage
- The score (!)
- Competition faced. (Not yet, ask me later)
Least-Squares Regression

- $\alpha$ a collection of observations
- $X$ a design matrix
- $\beta$ a collection of (imagined) individual isolated impacts

$$X\beta = \alpha$$
Least-Squares Regression

- $\alpha$ a collection of observations
- $X$ a design matrix
- $\beta$ a collection of (imagined) individual isolated impacts

\[ X\beta = \alpha \]

\[ X^T X\beta = X^T \alpha \]
Least-Squares Regression

- $\alpha$ a collection of observations
- $X$ a design matrix
- $\beta$ a collection of (imagined) individual isolated impacts

\[
X\beta = \alpha
\]

\[
X^TX\beta = X^T\alpha
\]

\[
\beta = (X^TX)^{-1}X^T\alpha
\]
Least-Squares Regression

- $\alpha$ a collection of observations
- $X$ a design matrix
- $\beta$ a collection of (imagined) individual isolated impacts

$$X\beta = \alpha$$

$$X^TX\beta = X^T\alpha$$

$$\beta = (X^TX)^{-1}X^T\alpha$$

Elements of $\alpha$ and $\beta$ can be taken from any inner product space and the usual proof goes through.
Dreamy Wishful Thinking Interlude

What if we had observations \( \alpha \) from \textit{every} possible combination of \( k \) players from our team of \( n \)? What would that get us?

It would make \( X \) very simple:

\[
X_{jp} = \begin{cases} \frac{1}{k} & \text{if } p \text{ is in the } j\text{-th } k\text{-combination of } n \\ 0 & \text{otherwise} \end{cases}
\]
Closed form solutions

Commonly, least squares solutions of $X\beta = \alpha$ are obtained by:

- Minimizing $||X\beta - \alpha||$ with a fancy optimiser.
- Numerically computing $\beta = (X^T X)^{-1}X^T \alpha$ with clever linear algebra.

However, in our case, (because $X$ is highly structured) we can work it out by hand:

$$\beta = (X^T X)^{-1}X^T \alpha$$
Montage with bad music (1/3)

\[
(X^T X)_{pq} = \begin{cases} 
\frac{1}{k^2} \binom{n-1}{k-1} & \text{if } p = q \\
\frac{1}{k^2} \binom{n-2}{k-2} & \text{if } p \neq q 
\end{cases} = \begin{cases} 
\frac{1}{kn} \binom{n}{k} & \text{if } p = q \\
\frac{k-1}{kn(n-1)} \binom{n}{k} & \text{if } p \neq q
\end{cases}
\]

\[
(X^T X)^{-1} = \begin{cases} 
\frac{n}{n-k} \left( \frac{(n-1)k + 1 - k}{n-k} \right) \binom{n}{k}^{-1} & \text{if } p = q \\
\frac{1-k}{n-k} \binom{n}{k}^{-1} & \text{if } p \neq q
\end{cases}
\]
Montage with bad music (2/3)

\[ \beta_p = \left[ (X^T X)^{-1} X^T \alpha \right]_p = \sum_j \left[ (X^T X)^{-1} X^T \right]_{pj} \alpha_j \]

\[ = \sum_j \sum_q \left( (X^T X)_{pq}^{-1} X_{qj} \right) \alpha_j = \sum_j \sum_q \left( (X^T X)_{pq}^{-1} X_{jq} \right) \alpha_j \]

\[ = \sum_j \sum_{q \text{ in } j} \left( (X^T X)_{pq}^{-1} \frac{1}{k} \alpha_j \right) \]

\[ = \sum_{j \text{ with } pq \text{ in } j} \left( (X^T X)_{pq}^{-1} \frac{1}{k} \alpha_j \right) + \sum_{j \text{ without } pq \text{ in } j} \left( (X^T X)_{pq}^{-1} \frac{1}{k} \alpha_j \right) \]

\[ = \frac{(k - 1)e + d}{k} \sum_{j \text{ with } p} \alpha_j + e \sum_{j \text{ without } p} \alpha_j \]

\[ = \frac{(k - 1)e + d}{k} \left( \frac{n - 1}{k - 1} \right) \overline{\alpha}_j \text{ with } p + e \left( \frac{n - 1}{k} \right) \overline{\alpha}_j \text{ without } p \]
Montage with bad music (3/3)

\[
\frac{(k - 1)e + d}{k} \binom{n - 1}{k - 1} = \frac{1}{k} \frac{k}{n} \binom{n}{k} [(k - 1)e + d]
\]

\[
= \frac{1}{n} \binom{n}{k} \left[ (k - 1)n \frac{1 - k}{n - k} \binom{n}{k}^{-1}
\right.

\[
+ n \frac{(n - 1)k + 1 - k}{n - k} \binom{n}{k}^{-1}
\]

\[
= (k - 1) \frac{1 - k}{n - k} + \frac{(n - 1)k + 1 - k}{n - k}
\]

\[
= \frac{(k - 1)(1 - k) + (n - 1)k + 1 - k}{n - k}
\]

\[
= \frac{k(1 - k) + (n - 1)k}{n - k} = \frac{k(n - k)}{n - k}
\]

\[
= k
\]
\[
e\left(\begin{array}{c}
n - 1 \\ k
\end{array}\right) = \binom{n-1}{k} n \frac{1-k}{n-k} \binom{n}{k}^{-1}
\]
\[
= \frac{(n-1)! nk! (n-k)!}{k! (n+k-1)! n! (n-k)} (1-k)
\]
\[
= 1 - k
\]

just as desired.
Closed form solutions

\[ \beta_p = k \cdot \left( \text{Average of Entries in } \alpha \right) - (k - 1) \cdot \left( \text{Average of Entries in } \alpha \text{ Without } p \right) \]
Closed form solutions

\[ \beta_p = 5 \left( \text{Average of Entries in } \alpha \text{ With } p \right) - 4 \left( \text{Average of Entries in } \alpha \text{ Without } p \right) \]
Closed form solutions

\[ \beta_p = 3.1 \left( \text{Average of Entries in } \alpha \text{ With } p \right) - 2.2 \left( \text{Average of Entries in } \alpha \text{ Without } p \right) \]

Ridge regression with \( \lambda = 0.5 \). This requires standardization.
Observations

We need observations for every combination of 5 players.
Observations

We need observations for every combination of 5 players.

When observations aren’t available, impute them:
▶ by amalgamating all possible subsets of four players,
▶ by amalgamating all possible subsets of three,
▶ &c.

By hook or by crook, manufacture an $\alpha$ for every combination.
Adjustments

The regression adjusts for teammates, but not for score or zone usage. To account for them, we adjust each $\alpha$ for these things individually.
Score Adjustment

Home Team Losing ———— Tied ———— Home Team Winning

45.1 shots per hour
45.2 shots per hour
43.9 shots per hour
41.9 shots per hour
39.1 shots per hour
37.8 shots per hour
35.3 shots per hour
32.1 shots per hour
35.0 shots per hour
36.7 shots per hour
39.5 shots per hour
41.8 shots per hour
43.2 shots per hour
42.9 shots per hour
Zone Adjustment

Defensive Zone—Neutral Zone——On The Fly——Offensive Zone
Treat shot densities as first-class objects.

Adjust observations of a given set of players for score and zone.

Impute observations for combinations who didn’t play together.

Use ridge regression to isolate individual performances.
Verification

- Consolidate impact into convenient units.
- Measure correlation from season to season. (for serious)
- Look at some interesting examples. (for insight)
- Look at some tails of the distribution this year. (for laughs)
To compare different players we weight their isolated shot contributions according to league average shooting percentages from given locations to obtain threat.
Threat

To compare different players we weight their isolated shot contributions according to league average shooting percentages from given locations to obtain *threat*.

Carefully avoiding shooting talent and goaltender talent.
Threat

To compare different players we weight their isolated shot contributions according to league average shooting percentages from given locations to obtain *threat*.

Carefully avoiding shooting talent and goaltender talent.

- Units of threat are goals per hour, purely from individual impact on shot locations.
2017-2018 Daniel Sedin, Observed vs Isolated, Offence

On-ice

44 shots per hour
2.5 threat

Isolated

47 shots per hour
2.5 threat
2017-2018 Henrik Sedin, Observed vs Isolated, Offence

On-ice

43 shots per hour
2.4 threat

Isolated

42 shots per hour
2.1 threat
2017-2018 Daniel Sedin, Observed vs Isolated, Defence

On-ice
39 shots per hour
2.2 threat

Isolated
38 shots per hour
2.0 threat
2017-2018 Henrik Sedin, Observed vs Isolated, Defence

On-ice
39 shots per hour
2.3 threat

Isolated
36 shots per hour
2.1 threat
2017-2018 Mathieu Perreault, Observed vs Isolated, Offence

On-ice

50 shots per hour
2.7 threat

Isolated

58 shots per hour
2.9 threat
2017-2018 Mathieu Perreault, Observed vs Isolated, Defence

On-ice
38 shots per hour
1.8 threat

Isolated
30 shots per hour
1.5 threat
Correlations - Offensive Threat Created

![Scatter plot showing 5v5 offensive measures with correlation coefficient r = 0.50.](image)
Correlations - Defensive Threat Allowed

5v5 Defensive Measures

\[ r = 0.48 \]

Following Year Isolated Threat (Goals per hour)

Previous Year Isolated Threat (Goals per hour)
Correlations - Net Threat

5v5 Net Measures (Offence - Defence)

$r = 0.47$
### Season-to-Season Auto-Correlations

For players who do not change teams, team-seasons for 2015-2016, 2016-2017, and 2017-2018, per hour of icetime:

<table>
<thead>
<tr>
<th></th>
<th>Offence</th>
<th>Defence</th>
<th>Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-ice Goals</td>
<td>0.36</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td>On-ice Unblocked Shots</td>
<td>0.56</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>On-ice Shots</td>
<td>0.62</td>
<td>0.51</td>
<td>0.59</td>
</tr>
<tr>
<td>Isolated Threat</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
</tr>
</tbody>
</table>
## Season-to-Season Auto-Correlations

For players who do not change teams, team-seasons for 2015-2016, 2016-2017, and 2017-2018, per hour of icetime:

<table>
<thead>
<tr>
<th></th>
<th>Offence</th>
<th>Defence</th>
<th>Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-ice Goals</td>
<td>0.36</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td>On-ice Unblocked Shots</td>
<td>0.56</td>
<td>0.52</td>
<td>0.56</td>
</tr>
<tr>
<td>On-ice Shots</td>
<td>0.62</td>
<td>0.51</td>
<td>0.59</td>
</tr>
<tr>
<td>Isolated Threat</td>
<td>0.51</td>
<td>0.50</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Correlation between isolated offensive threat and isolated defensive threat per hour is 0.01. The two performances are independent.
<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>Isolated Threat Against</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matt Stajan</td>
<td>CGY</td>
<td>1.32</td>
</tr>
<tr>
<td>Charles Hudon</td>
<td>MTL</td>
<td>1.38</td>
</tr>
<tr>
<td>Scott Laughton</td>
<td>PHI</td>
<td>1.45</td>
</tr>
<tr>
<td>Greg Pateryn</td>
<td>DAL</td>
<td>1.45</td>
</tr>
<tr>
<td>Mathieu Perreault</td>
<td>WPG</td>
<td>1.47</td>
</tr>
<tr>
<td>Colton Parayko</td>
<td>STL</td>
<td>1.48</td>
</tr>
<tr>
<td>Dmitrij Jaskin</td>
<td>STL</td>
<td>1.48</td>
</tr>
<tr>
<td>Jordan Nolan</td>
<td>BUF</td>
<td>1.52</td>
</tr>
<tr>
<td>Alexander Iafallo</td>
<td>LA</td>
<td>1.52</td>
</tr>
<tr>
<td>Dan Hamhuis</td>
<td>DAL</td>
<td>1.53</td>
</tr>
</tbody>
</table>
## Best Offensive Threat Performances, 2017-2018

<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>Isolated Threat For</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sidney Crosby</td>
<td>PIT</td>
<td>3.56</td>
</tr>
<tr>
<td>Timo Meier</td>
<td>SJ</td>
<td>3.44</td>
</tr>
<tr>
<td>Kris Letang</td>
<td>PIT</td>
<td>3.30</td>
</tr>
<tr>
<td>Conor McDavid</td>
<td>EDM</td>
<td>3.26</td>
</tr>
<tr>
<td>Michael Frolik</td>
<td>CGY</td>
<td>3.24</td>
</tr>
<tr>
<td>Joonas Donskoi</td>
<td>SJ</td>
<td>3.20</td>
</tr>
<tr>
<td>Pierre-Luc Dubois</td>
<td>CBJ</td>
<td>3.20</td>
</tr>
<tr>
<td>Patric Hörnqvist</td>
<td>PIT</td>
<td>3.08</td>
</tr>
<tr>
<td>Tyler Toffoli</td>
<td>LA</td>
<td>3.07</td>
</tr>
<tr>
<td>Markus Nutivaara</td>
<td>CBJ</td>
<td>3.06</td>
</tr>
</tbody>
</table>
# Best Overall Threat Performances, 2017-2018

<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>Isolated Threat Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathieu Perreault</td>
<td>WPG</td>
<td>1.45</td>
</tr>
<tr>
<td>Hampus Lindholm</td>
<td>ANA</td>
<td>1.34</td>
</tr>
<tr>
<td>Kris Letang</td>
<td>PIT</td>
<td>1.33</td>
</tr>
<tr>
<td>Colton Parayko</td>
<td>STL</td>
<td>1.31</td>
</tr>
<tr>
<td>Mark Giordano</td>
<td>CGY</td>
<td>1.19</td>
</tr>
<tr>
<td>Joonas Donskoi</td>
<td>SJ</td>
<td>1.18</td>
</tr>
<tr>
<td>Pierre-Luc Dubois</td>
<td>CBJ</td>
<td>1.18</td>
</tr>
<tr>
<td>Dougie Hamilton</td>
<td>CGY</td>
<td>1.15</td>
</tr>
<tr>
<td>Timo Meier</td>
<td>SJ</td>
<td>1.14</td>
</tr>
<tr>
<td>Adam Lowry</td>
<td>WPG</td>
<td>1.12</td>
</tr>
</tbody>
</table>
Conclusions

- Isolated threat seems to describe repeatable aspects of
  - offensive,
  - defensive, and
  - all-around skater performance

for players who do not change teams.
Future Work

For shot map isolation itself:

- Quality-of-competition.
- Non-linear effects. (Chemistry!)

For a broader evaluation scheme:

- Goalies and shooting talent.
- Special Teams.
Thanks!
Quality of Competition

Shot Rates facing various competitions
Quality of Competition

Shot Rates facing various competitions