

Are Expected Goal Models Biased?

John McCool

University of Pittsburgh

Email: johnmccool31@gmail.com

Introduction

In hockey analytics, shot attempt metrics have grown in popularity in recent years due to their ability to project future player and team performance. However, metrics like Corsi (all shot attempts) and Fenwick (all unblocked shot attempts) do not take into account shot quality. One approach to incorporating shot quality into shot attempt metrics is to weight each unblocked shot attempt by its probability of being converted into a goal. In this approach, a goal expectancy is assigned to unblocked shots based factors such as the distance, angle, and whether the shot resulted from a rebound or off the rush. The resulting metric is typically referred to as "Expected Goals" (xG) pioneered by Brian Macdonald, currently the Director of Hockey Analytics for the Florida Panthers, which was in part based on a soccer expected goal version.

Expected goals (xG) are [better predictors](#) of team and player performance than Corsi and +/- because goals are random and scarce events. For reference, the NHL average expected goal on a given shot hovered around 0.068 based on shooting percentage in 2015-2016. Also, it is important to realize that expected goals are also influenced by pass sequences, quick touches, or traffic in front of the net, and not just by player skill alone. Ryan Stimson, a contributor of Hockey Graphs, created The Passing Project, which tracks the relationship between pass types (i.e behind the net or stretch passes) and shooting percentage.

Expected goals differ from shot attempt metrics like Corsi or Fenwick in that the value assigned to a particular shot attempt comes directly from a statistical model, typically logistic regression. To evaluate players or teams in the 2016-17 season, it might seem reasonable to use an xG model trained on previous seasons (e.g. data from 2007-08 to 2015-16). **However, doing so could potentially lead to a substantial amount of bias.** What if goalies have systematically gotten better or worse at stopping particular types of shots (e.g. better rebound control) or shots from particular areas of the ice over time? What if players have gotten systematically better or worse at converting particular types of shots or shots from particular areas of the ice over time? If this is true, it would be inappropriate to use an xG model trained on previous seasons of data to evaluate players or teams in the current season. To date, no one has explored this potential bias.

Methods

Logistic regression and bootstrapping were used to test for bias in the expected goal coefficients. Both statistical techniques allow us to explore the variation between the distance, angle, rebound, and rush coefficients in logistic regression models for expected goals from the 2007-2008 through the 2015-16 seasons. Changes in these xG coefficients may indicate systematic shifts in how goals are scored in the NHL over time, rendering xG models trained on past seasons of data biased when applied to the 2016-17 season

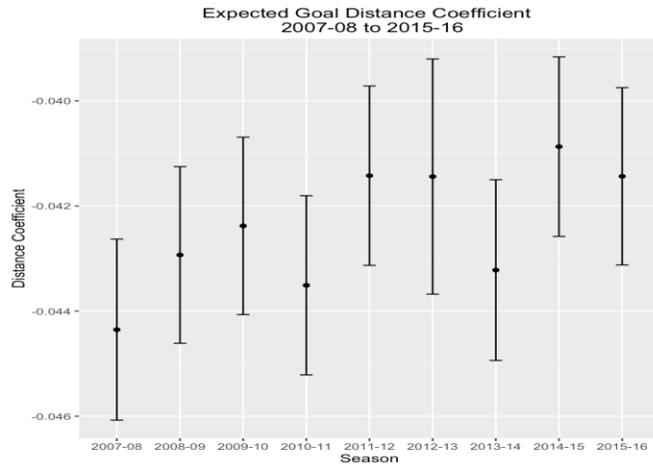
The first part of statistical analysis was carried out using logistic regression. Using this technique, we developed a model that measured the probability or odds of the response (shot outcome) taking on a particular value (1 if goal, 0 if non-goal), which was modeled conditional on the distance of the shot, angle of the shot, whether or not it was a rebound shot, and whether or not it was a shot off the rush. The expected goal coefficient values will be compared to the coefficients obtained in the most recent 2015-2016 season, gathered from 107,872 goal, shot on goal, and missed shot observations. All data comes from [Corsica](#).

Distance Coefficient

Shot distance is a good predictor of goal probability. Unsurprisingly, shot attempts closer to the net have a higher probability of getting past the goaltender. As an example, the Montreal Canadiens had a -0.47

correlation between expected goals and the distance of their shots last season, which includes the statistics of their recently signed players during the offseason.

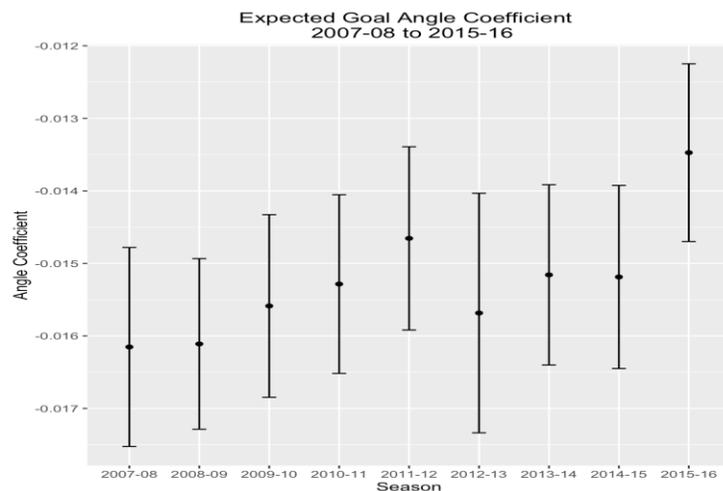
The log-odds of a goal for the distance coefficient was -0.041 in 2015-16. In other words, for every one-foot decrease in the distance coefficient, the log-odds of a goal (relative to a non-goal / save) fell by 0.041. This coefficient had the highest absolute Z-value (-48.1) suggesting that it is strongest predictor of expected goals in the model (p -value < .001).



We can see that while there are no significant changes in the distance coefficient across the past nine seasons, there appears to be a slight upward trend in its value between 2007-08 and 2015-16 moving from -0.044 to -0.041. Based on this increase, we might interpolate that there is now a higher expected goal probability on shots from further distances. Interestingly, for consecutive pairs of seasons such as 2010-11 and 2011-12 or 2013-14 and 2014-15, the distance coefficients make substantial (though not significant) jumps across these seasons.

Angle Coefficient

We also found that increasing the angle of a shot slightly decreases the log-odds of a goal by -0.013 units (p -value < .001). There was also a -0.20 correlation between angle of the shot and expected goals last season.

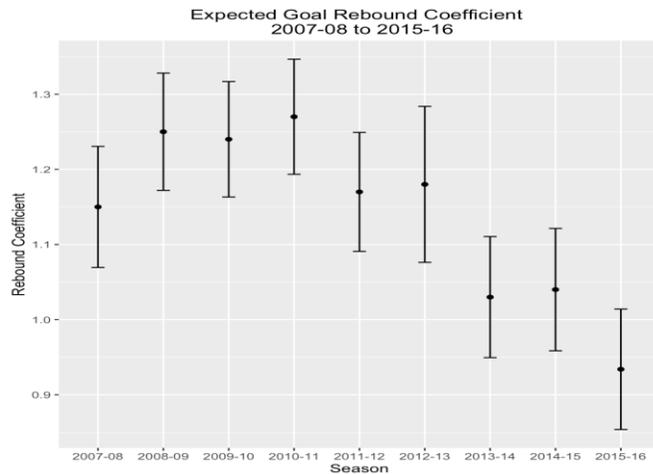


The time series shows an increasing trend in the distance coefficient since 2007-08. The coefficient moved from -0.016 to -0.015 before jumping to -0.013 in 2015-16. The large increase for the 2015-16

season is concerning, since the angle coefficient in 2015-16 falls outside of the 95% confidence interval from all previous seasons except the lockout-shortened 2011-12 season. This graph provides some potential evidence of a systematic shift in what affects the likelihood of an unblocked shot attempt resulting in a goal in recent seasons. At the very least, some caution should be used when using xG models from previous seasons to obtain xG statistics for the current season.

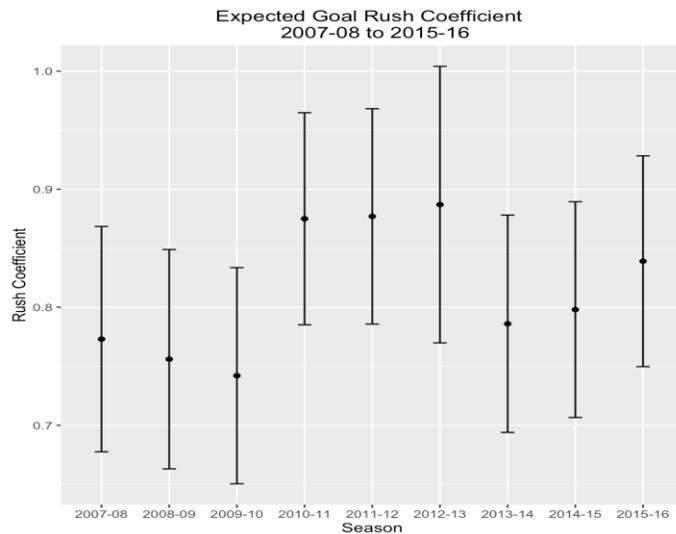
Rebound Coefficient

The log-odds of a goal from rebound shots was 0.934 last season (p -value < .001). As a binary/indicator variable, the rebound coefficient has a slightly different interpretation than the distance and angle coefficients. For instance, we would say that the log-odds of a goal (vs. a non-goal / save) increased by 0.934 compared to a non-rebound shot situation in 2015-16.



This graph is particularly concerning. The time series plot shows that there has been a significant drop in the rebound coefficient, especially in recent seasons. This coefficient reached a peak in 2010-11 (1.27) before steadily falling to 0.934 in 2015-16. The decreasing trend might suggest that rebound chances are having less of an influence on expected goal probability. It is also possible that goalies are stopping more rebound shots during recent seasons. The coefficients for several seasons fall outside of the 95% confidence interval around the coefficients of the previous season (e.g. 2011-12, 2013-14, 2015-16). These significant differences in the rebound coefficient from season-to-season are very concerning, and provide some evidence that it may be inappropriate to use xG models trained on previous seasons to obtain xG statistics for subsequent seasons.

Rush Coefficient



[Rush shots](#) increased the log-odds of an expected goal by 0.839 compared to “normal” shot attempts in 2015-16 (p -value<0.001). The coefficient had a high amount of variation (likely due to the lower sample size for rush shots), similar to that of the rebound coefficient. The expected goal log-odds for the rush coefficient increased by 0.04 units last season relative to 2014-15, which was only a moderate change compared to a 0.10 drop between the 2012-13 and 2013-14 seasons or the large increase from the 2009-10 to 2010-11 seasons. The unpredictability of this coefficient could possibly be explained through goalie, defensive, or offensive performance in rush situations.

Overall, we expected to find minimal variation between the expected goal coefficients. On one hand, there was relatively low deviation in the distance and angle coefficients. On the other, the rebound and rush coefficients had an unusually high amount of variation. Some of this variation could be related to the changing pace and style of play of the game or simply random sequencing.

Finally, note that we used a parsimonious xG model here (with no interaction terms). More complicated xG models (for example, those that include interaction terms involving rebound and/or rush shots) may suffer coefficient instability across seasons even more significant than what we discovered above. We encourage other analysts to perform similar tests of their own xG models and, perhaps more importantly, to be clear about which seasons their models were trained when publishing xG results.

Bootstrapping

The second half of the analysis explores the coefficient variability using a bootstrap analysis.

Bootstrapping allows us to get an approximate distribution of the distance, angle, rebound, and rush coefficients for each of the 2007-08 through 2015-16 seasons.

As an overview, bootstrapping is a statistical method that draws repeated samples from a population of interest with n observations (in our case, all unblocked shot attempts in a season). Bootstrapping gives us the sampling distributions of each of the coefficients in the model. In other words, it helps to answer the question, “If we replayed the 2015-16 season, how different would our expected goal model be?”

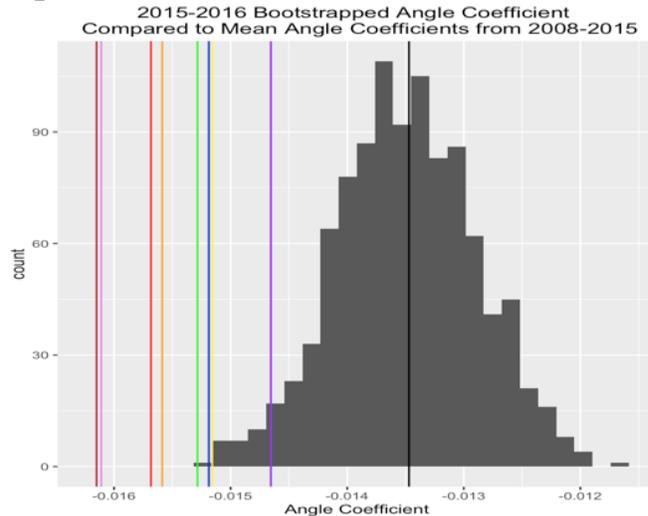
Bootstrapping also allows us to estimate the confidence interval for the unknown parameter estimate (θ) and the probability that θ lies within the bounds of the interval, without any of the parametric assumptions from logistic regression. In other words, it helps to answer the question, “How much variability would there be in the rush shot coefficient if we replayed the 2015-16 season many times?” The following procedure was used for the bootstrap analysis:

1. Resample dataset (usually more than 100,000 unblocked shot attempts per season, marked as goal or non-goal) with replacement 1000 times. Each of the 1000 (re-) samples of the data is called a “phantom” sample (or bootstrapped sample).
2. Estimate the expected goals model with logistic regression on each sample, recording the coefficients for each iteration.

3. Create histograms of each coefficient. We'll call these the bootstrapped distribution of the coefficients.
4. Repeat this for all seasons for which we have data on unblocked shot attempt locations (2007-2016).

We performed bootstrap analysis on each xG coefficient with 1,000 re-samples over the past nine seasons. For reference, we compared the mean coefficient from the 2007-2015 seasons to the 2015-16 bootstrapped distributions as a measure of coefficient variability across NHL seasons. [1]

Angle Coefficient



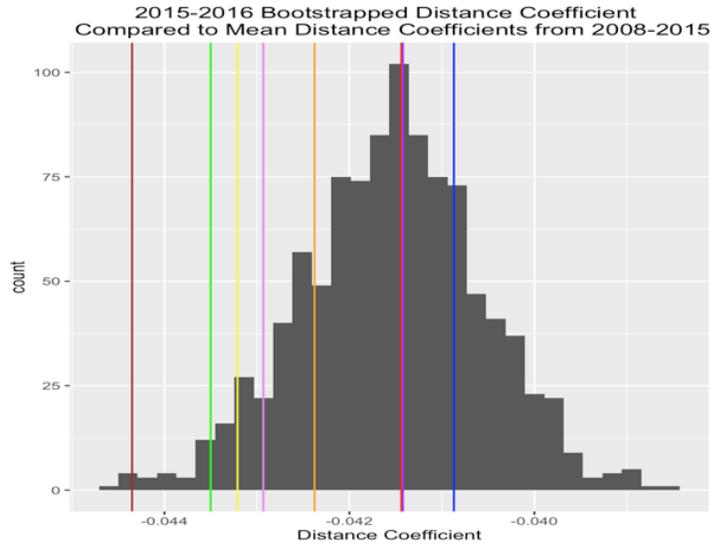
Legend:

- 2015-16 Black
- 2014-15 Blue
- 2013-14 Yellow
- 2012-13 Red
- 2011-12 Purple
- 2010-11 Green
- 2008-09 Orange
- 2009-10 Violet
- 2007-08 Brown

From the previous article, we found that increasing the angle of a shot slightly decreases the log odds of a goal. The coefficient, however, marginally increased to -0.0134 last season, which is significantly higher than the previous eight seasons. The bootstrapped coefficient distribution (2015-16) is approximately normally distributed, with 95% of the coefficients between (roughly) -0.015 and -0.012.

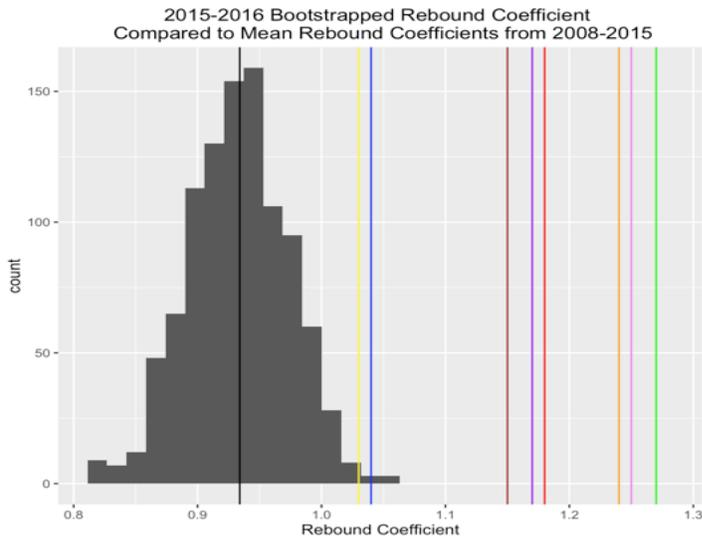
Aside from 2011-12, none of the angle coefficients from previous seasons fall inside last season's 95% bootstrap confidence interval for the 2015-16 season. **This means that if we used an xG model trained on the 2014-15 season to obtain player and team evaluation metrics for the 2015-16 season, the model would be biased in how it assigns goal expectancies for the 2015-16 season.** Finally, four angle coefficients lie completely outside of the bootstrapped distribution, the furthest of which are the -0.0162 and -0.0161 values from 2007-08 and 2008-09 respectively.

Distance Coefficient



The bootstrapped distance coefficients exhibit less variability than that of the angle coefficients. Last season's distance coefficient compares similarly to 2014-15, 2011-12, and 2013-14. The distribution also shows no evidence of an increasing or decreasing trend in the coefficient. Furthest from the center of 2015-16 distance distribution are the 2007-08, 2010-11, and 2013-14 mean coefficients. That being said, the distance coefficient range is just -0.01 indicating that it is a relatively stable xG predictor.

Rebound Coefficient

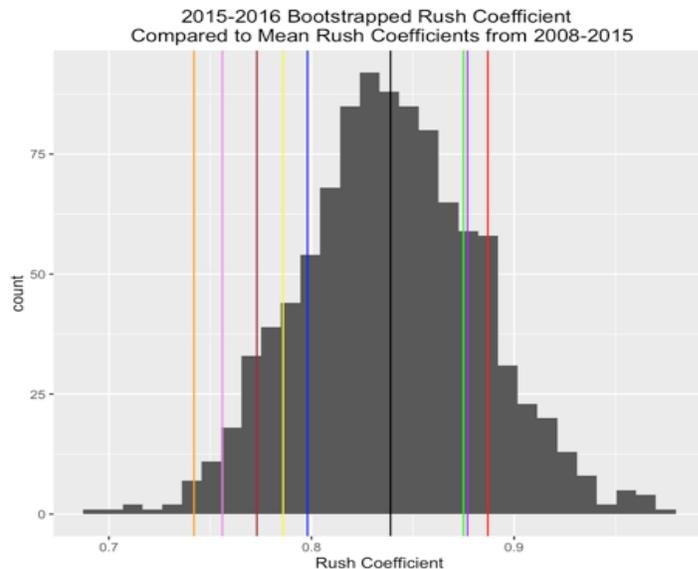


The rebound coefficient (0.934) in 2015-16 fell to lowest value over the past nine seasons. This bootstrapped distribution is approximately normally distributed with 95% of the coefficients ranging between about 0.82 to 1.05. On the right tail of the histogram lie the 2013-2015 rebound coefficients, which are outside of the 95% confidence interval for 2015-16. **Again, this means that an xG model trained on the 2014-15 season would be biased if applied to the 2015-16 season.**

The rebound coefficients follow a decreasing trend in the rebound coefficient since the start of the 2013-14 campaign, as we saw in Part I of this series. For example, the 2012-13 coefficient was 1.18 compared to just 1.03 and 1.04 in the following seasons. All of this might suggest that rebound shots are currently having less influence on expected goal models. Although this is not rule out the possibility xG rebound coefficient could jump all the way to 1.10 or even 1.20 this season, shattering this assumption.

On the more extreme end are the 2008- 2011 rebound coefficients with a combined 1.25 average. In other words, rebounds had more weight in expected goal models during this period. For reference, the 2010-2011 confidence interval (1.15, 1.37) for the bootstrapped rebound coefficient lies completely outside of the last season's interval!

Rush Coefficient



All together, the bootstrapped rush coefficients from 2015-16 are similar to those from previous seasons. The increased variation in the rush coefficient is likely due to the reduced sample size of rush shots in each season. From 2008-2010, the rush coefficient averaged 0.743 before increasing to 0.879 over the next three seasons. Since then, the coefficient has fallen between 0.786 to 0.839.

Conclusions

In summary, by using the above bootstrap analyses, we provide a more rigorous statistical analysis that suggests that bias likely exists in some xG models, especially those trained on prior seasons and applied to subsequent seasons. Significant deviations in the angle and rebound coefficients could potentially alter the accuracy of xG projections that fail to adjust for these changes. We again encourage analysts to be detailed and forthright in describing how and on which seasons their xG models are trained, so that any potential bias in the resulting xG statistics can be accounted for.

References:

- <http://www.corsica.hockey/blog/2016/03/03/shot-quality-and-expected-goals-part-i/>
- <http://www.corsica.hockey/blog/2016/08/13/shot-quality-and-expected-goals-part-1-5/>
- http://statsportsconsulting.com/main/wp-content/uploads/VENTURA_OTTANALYTICS.pdf
- http://statsportsconsulting.com/main/wp-content/uploads/CANE_OTTANALYTICS.pdf
- <http://donttellmeaboutheart.blogspot.com/2015/05/updated-nhl-expected-goals-model.html>
- <http://www.sloansportsconference.com/wp-content/uploads/2012/02/NHL-Expected-Goals-Brian-Macdonald.pdf>
- <https://hockey-graphs.com/2015/10/01/expected-goals-are-a-better-predictor-of-future-scoring-than-corsi-goals/>
- <http://hockeyanalysis.com/2014/07/09/introducing-rush-shots/>
- <https://hockey-graphs.com/2016/09/06/weighted-shot-rates-based-on-the-passing-project/>
- Ryder (2004), Johns (2004), Krzywicki (2005, 2009, 2010), Awad (2009), Schuckers (2011).
- <http://stat.rutgers.edu/home/mxie/RCPapers/bootstrap.pdf>
- <http://www.corsica.hockey/blog/2016/03/03/shot-quality-and-expected-goals-part-i/>
- <http://www.corsica.hockey/blog/2016/08/13/shot-quality-and-expected-goals-part-1-5/>

