

An Expected Goals Model for Evaluating NHL Teams and Players

Brian Macdonald
United States Military Academy
Department of Mathematical Sciences
West Point, NY 10996
Email: bmac@jhu.edu

Abstract

One difficulty with analyzing performance in hockey is the relatively low scoring rates compared to sports like basketball. Fenwick rating (shots plus missed shots) and Corsi rating (shots, missed shots, blocked shots) have been used to analyze players and teams because they have been shown to be better than goals as a predictor of future goals. In this paper, we use variables like faceoffs, hits, and other statistics as predictor variables in addition to goals, shots, missed shots, and blocked shots, to predict goals. Our models outperform previous models with regard to mean squared error of actual goals and predicted goals. The results can be interpreted as expected goals and can be used in adjusted plus-minus models instead of goals. We use ridge regression to estimate a player's contribution to his team's *expected* goals per 60 minutes, independent of his teammates, opponents, and the zone in which his shifts begin. We also give adjusted plus-minus estimates based on goals, shots, Fenwick rating, and Corsi rating and use these results alongside the results for expected goals to provide an additional means by which NHL analysts, decision-makers, and fans can measure how valuable a player is to his team.

1 Introduction

Low scoring rates are a source of difficulty when analyzing team and player performance in hockey, particularly when using less than a season's worth of data. The randomness and scarcity of goals limit one's ability to properly judge current performance and predict future performance of teams and players using goals alone. Shot differential, Fenwick rating (shots and missed shots) differential, and Corsi rating (shots, missed shots¹, and blocked shots) differential have become popular in the hockey analysis community for analyzing the performance of teams and players. One reason hockey analysts use these statistics is that one can obtain better predictions of a team's future performance by using shots, Fenwick rating or Corsi rating instead of goals. Specifically, a team's current shot differential, Fenwick differential, and Corsi differential are each better than a team's current goal differential at predicting goal differential when a half season's worth of data is used [1]. These statistics are less scarce and less random than goals, and they are good indicators of territorial advantage and possession advantage at the team level. One benefit of using these statistics to evaluate a player's performance is that, for the most part, goalies will not have a big impact on a player's defensive ratings. Further details about the benefits of using Fenwick rating and Corsi rating are discussed in [1], [2], [3], [4], and [5].

The first goal of this paper is to answer the following question: Can this predictive performance be improved further if we include additional statistics like hits, faceoffs, etc., as predictor variables, along with some combination of goals, shots, missed shots and blocked shots? We form new models using these statistics and show that they perform better than previous models in terms of mean squared error

¹Note that by "missed shots" we mean an attempted shot that went wide of the net, over the net, or hit the post

of predicted goals and actual goals. We focus strictly on team offense, and we restrict our attention to 5-on-5 situations in which both goalies are on the ice. Unless otherwise specified, all statistics throughout this paper are given as “per 60 minutes” rate statistics during even strength situations. For example, when we say “Goals” or “GF”, we mean “Goals For per 60 minutes of playing time at even strength.”

The results of our models can be interpreted as the “expected goals” that a team will score based on the various statistics mentioned above. There are many uses for an expected goals statistic. For example, expected goals could be used to evaluate team offensive and defensive performance over a half season’s worth of games.

Expected goals can also be used to evaluate players. The second focus of this paper is the use of expected goals in an adjusted plus-minus model to evaluate players. Adjusted plus-minus models were introduced in basketball ([6], [7], [8], [9]) as a way to estimate a player’s contribution to his team, independent of the strength of his teammates and opponents. One of the main downsides of adjusted plus-minus is the large error bounds on the estimates of player performance that are obtained from the model. One reason for the large errors is tendency of some teammates to play together often, which causes collinearity to be present in the data. For example, Daniel Sedin is on the ice with his twin brother Henrik 93% of the time, and their adjusted plus-minus estimates have very large error bounds as a result. Another source of large errors in hockey are the low goal scoring rates.

In light of these issues, we use an adjusted plus-minus model similar to those described in [10], [11], and [5], using *expected* goals per 60 minutes as the dependent variable instead of goals per 60 minutes. The results give an estimate of a player’s contribution to his team’s expected goals, independent of the strength of his teammates, opponents, and the zone in which his shifts begin. One main benefit of using expected goals is that there is much more data compared to when ones uses just goals, which helps produce lower error bounds.

We also use ridge regression, as first used by Joe Sill in basketball [12], to estimate each player’s adjusted plus-minus. Ridge regression is a statistical technique that is commonly used when collinearity is present in the data. This method helps reduce the error bounds in the estimates, and typically improves the predictive performance in the model. We also use ridge regression to estimate adjusted plus-minus statistics based on goals, shots, Fenwick rating, and Corsi rating. The combination of ridge regression and expected goals, along with the adjusted plus-minus results based on other statistics, provides a useful means with which to analyze the performance of players.

The rest of this paper is organized as follows. In Section 2, we describe our models and compare the performance of the new models to goals, shots, Fenwick rating and Corsi rating. In Section 3, we describe how to use the expected goals obtained from these models to evaluate players using adjusted plus-minus. A discussion of some of the predictor variables, specifically hits and faceoffs, is in Section 4. In Section 5, we describe some current work and ideas for future work.

2 Two new models

We used both ordinary least squares regression and ridge regression with goals, shots, missed shots, and blocked shots, along with several additional statistics from one half of a season to predict goals scoring rate in the other half of a season. The following is a list of all team statistics that we considered:

Goals Goals scored.

Shots Shots on goal.

Missed Shots Attempted shots that missed the net wide or high, or a shot that hit the goalpost

Blocked Shots Attempted shots that were blocked by an opposing team’s forward or defenseman.

Fenwick rating Shots plus missed shots

Corsi rating Shots plus missed shots plus blocked shots.

Zone starts The number of shifts that begin with a faceoff in the offensive, defensive, neutral zone.

Turnovers Giveaways and takeaways.

Faceoffs Faceoff wins, faceoff losses, faceoff winning percentage, net faceoff wins (faceoff wins minus faceoff losses), and total faceoffs

Hits Hits (a player on the team hit a player on the other team), and hits against (a player on the team was hit by a player on the other team)

Shooting percentage Team shooting percentage (goals divided by shots)

We used data from the last four full NHL seasons. For each team, the season was split into two halves. Since midseason trades and injuries can have an impact on a team’s performance, we did not use statistics from the first half of the season to predict goals in the second half. Instead, we split the season into odd and even games, and used statistics from odd games to predict goals in even games. Data from 2007-08, 2008-09, and 2009-10 was used as the training data to estimate the parameters in the model, and data from the entire 2010-11 was set aside for validating the model. The model was also validated using 10-fold cross-validation. Mean squared error (MSE) of actual goals and predicted goals was our choice for measuring the performance of our models.

We chose the subset of the above predictor variables that yielded the best fit, according to adjusted R^2 , in an ordinary least squares (OLS) model. We then removed two more variables which were not statistically significant. The following predictor variables remained: goals, shots, hits, hits against, and faceoffs. We discuss hits and faceoffs in Section 4. These variables were used in an OLS model and a ridge regression model. We also tested Corsi rating with hits (Corsi rating plus Hits Against minus Hits).

Two measures of the performance of our models are depicted in Figures 1 and 2. In order to

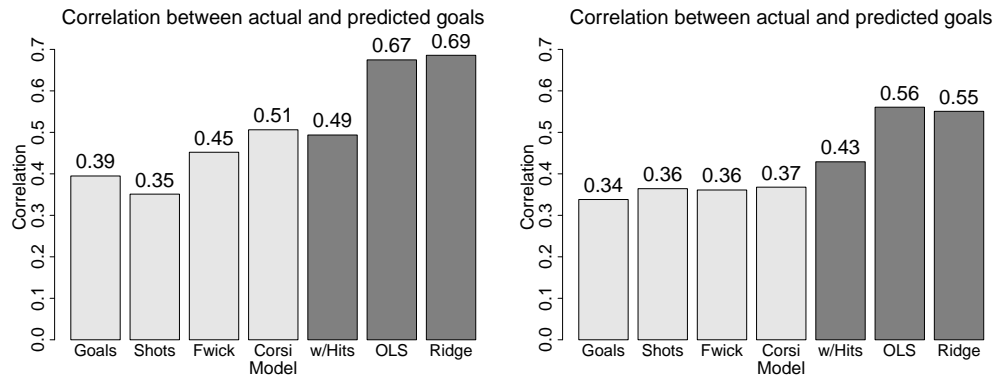


Figure 1: Correlation between actual goals and predicted goals using Goals, Shots, Fenwick rating, Corsi rating, as well as our new models, Corsi with Hits, OLS and Ridge. On the left, correlation the entire 2010-11 season as the validation data is shown. On the right, cross-validated correlation is shown. In both cases, OLS and Ridge have the highest correlation. Here, and throughout the rest of the paper, results for our new models are given in dark grey.

compare our results to those in [1], we show the correlation between the actual goals and the predicted goals using various models in Figure 1. Note the correlations for goals, shots, Fenwick rating and Corsi rating are fairly similar to those given in [1]. The differences are likely because we have restricted ourselves to offense, while in [1] the author considered both offense and defense. Note that shots, Fenwick rating, and Corsi rating tend to have a slightly higher correlation than goals. The correlation between actual goals and predicted goals is highest with our OLS and Ridge models.

The mean squared error (MSE) results are depicted Figure 2. The new models had a lower MSE, indicating the difference between actual goals and predicted goals is lower on average. The ridge estimates perform slightly better than the OLS regression when using the 2010-11 season as the validation data, while OLS performs slightly better when using cross-validation. Since ridge regression typically performs better than OLS regression when the predictor variables are correlated, we will use the ridge regression results for computing the expected goals that we will use in Section 3.

We note that for forming our final model, we removed four outliers, Ottawa and Carolina from 2007-08 and Minnesota and New York Islanders from 2010-11. We also fit the model without removing outliers, and the results for correlation and mean squared error are similar (see Figure 3 the Appendix).

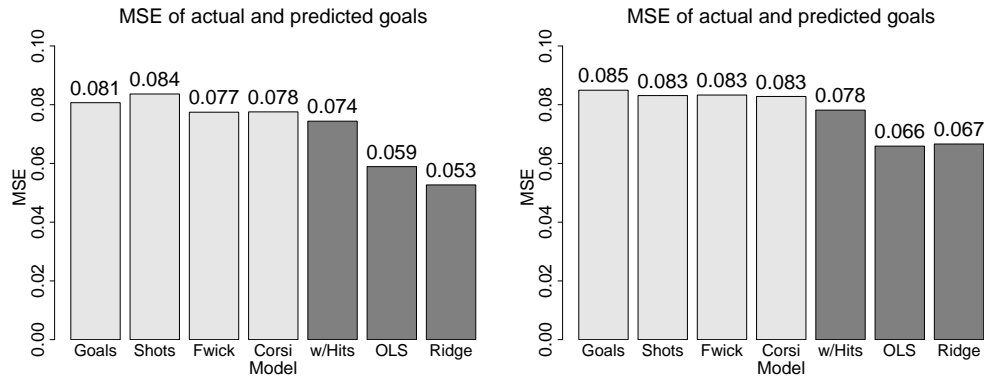


Figure 2: Mean squared error (MSE) of actual goals and predicted goals using Goals, Shots, Fenwick rating, Corsi rating as well as our new models, Corsi with Hits, OLS and Ridge. On the left, MSE using the entire 2010-11 season as the validation data set is shown. On the right, cross-validated MSE is shown. In both cases, OLS and Ridge have the lowest MSE.

3 Using Expected Goals in Adjusted Plus-Minus

Expected goals can be used to analyze team performance, but here we focus on player performance, and specifically on using adjusted plus-minus with expected goals to evaluate players. Recall that we use an adjusted plus-minus model similar to those described in [10], [11], and [5], using *expected* goals per 60 minutes as the dependent variable instead of goals per 60 minutes. Also, recall that we use ridge regression as in [12] and [5] instead of OLS regression.

In Table 1, we list the top players in E_{EV}^{off} , the **offensive** component of adjusted plus-minus during **even** strength situations based on **expected** goals. In that table we have also listed each player’s G_{EV}^{off} , S_{EV}^{off} , F_{EV}^{off} , and C_{EV}^{off} , which are the **offensive** components of adjusted plus-minus at **even** strength based on **goals**, **shots**, **Fenwick rating**, and **Corsi rating**, respectively. We also give some per 60 minute versions of these statistics in the last 3 columns. For comparison, the top 10 players in G_{EV}^{off} , the

Table 1: The top 10 offensive players in the NHL according to G_{EV}^{off} .

	Player	Pos	Team	E_{EV}^{off}	G_{EV}^{off}	S_{EV}^{off}	F_{EV}^{off}	C_{EV}^{off}	$E_{EV,60}^{off}$	$G_{EV,60}^{off}$	$S_{EV,60}^{off}$
1	Sidney Crosby	C	PIT	8	16	8	9	9	0.43	0.83	0.42
2	Alex Ovechkin	LW	WSH	7	11	11	13	15	0.31	0.46	0.45
3	Daniel Sedin	LW	VAN	7	10	9	9	9	0.31	0.47	0.44
4	Zach Parise	LW	N.J	6	8	9	9	8	0.39	0.49	0.55
5	Pavel Datsyuk	C	DET	6	12	6	6	7	0.27	0.53	0.27
6	Jonathan Toews	C	CHI	6	9	5	4	4	0.28	0.45	0.22
7	Eric Staal	C	CAR	6	6	10	12	13	0.27	0.30	0.48
8	Jeff Carter	C	PHI	6	6	9	10	9	0.27	0.27	0.43
9	Evgeni Malkin	C	PIT	5	8	6	6	6	0.26	0.40	0.31
10	Joe Thornton	C	S.J	5	8	6	6	8	0.20	0.34	0.26

offensive component of adjusted plus-minus based on **goals**, are given in Table 2 in the Appendix.

Not surprisingly, Sidney Crosby is the best offensive player at even strength according to both E_{EV}^{off} and G_{EV}^{off} , despite the fact that he missed significant amounts of time during the last four full seasons. The two lists contain many of the same players. Two players, Eric Staal and Jeff Carter, are in Table 1 but not Table 2 probably because of the number of shots that they generate (recall that shots was one of the variables that we used in our expected goals models).

Nathan Horton may be a surprise as the fifth player in Table 2, as he is typically not regarded as one of the league’s best offensive players. One might prefer to consider a player with significantly better

E_{EV}^{off} or S_{EV}^{off} , like Zach Parise, among the top 5 offensive players at even strength instead of Horton. However, we note the strength of Horton's teammates (2.43 goals per 60 minutes) is fairly low, and when Horton is on the ice, his team scores 3.20 goals per 60 minutes. That is a difference of 0.77 goals per 60 minutes, which is the third best total in the league among players with 600 minutes played². In light of this ranking, Horton's G_{EV}^{off} seems reasonable and he should be considered one of the best offensive players in the league at even strength.

4 Discussion

The results of our model can also be used to analyze the relative value of different box score statistics. For example, in every model that we tried during our model building process, hits were a significant predictor, but not in the way you might think. We now discuss some of the predictor variables that we considered in further detail.

Hits One thing that stood out from the model building process was that hits and hits against were significant statistics. That is not terribly unexpected perhaps, but it is the sign of the estimates for hits and hits against that may be a bit surprising. The coefficients of hits and hits against are negative and positive, respectively. This means that low hits, and high hits against, are good predictors of goals. In other words, the teams more hits against than hits are the teams with higher goals.

There are two possible explanations. First, typically the team doing the hitting does not have the puck. So hits and hits against contain information about possession. Hits against indicate possession of the puck, and hits indicate that the opposition had the possession of the puck. Good puck possession teams have more hits against than hits simply because they have the puck more often than their opponents.

Another potential reason is that a player who applies a hit often takes himself out of the play temporarily, along with the player he hit. A hit typically means the other team has the puck, and after a hit, the play can momentarily be 4-on-4 instead of 5-on-5. This is even more true on the power play. If a penalty killer hits someone on the other team, the play is temporarily 4-on-3 instead of 5-on-4, even if it is only for a half of a second. Even fractions of a second can matter, especially at the NHL level, and especially in the defensive zone.

We remark that these results do not necessarily indicate that hits are bad, or that players should stop finishing their checks. But it does provide some evidence that hits, hits against, and puck possession are related, and that poorly timed hits can impact goal scoring.

Total faceoffs Interestingly, total faceoffs were a significant variable. Intuitively, offensive zone faceoffs should be a significant predictor of goals, but the importance of total faceoffs is not as obvious. One reason is that total faceoffs contain some information about goals, since every goal results in a neutral zone faceoff. But even when we excluded the neutral zone faceoffs that followed a goal, total faceoffs were still significant.

Another possible reason total faceoffs are significant is that the flow of a game is more structured after a faceoff, since a team that wins a faceoff can run an organized play. A third possible reason is that typically after a faceoff there is not a lot of dead time in the play when both teams are changing players "on the fly". One team may hold the puck behind their net for 5-10 seconds while the substitutions occur, and this break in activity reduces the goals scoring rate (for both teams). On the other hand, after a faceoff, there is typically not a break in activity like this as often.

We tried the model without total faceoffs. The performance of both the OLS model and the ridge model was reduced without the total faceoffs, however both models still performed better than shots, Fenwick rating, or Corsi rating alone. See Figure 6 in the Appendix. In the end we included faceoffs because performance increased with this variable, and while they may be not an obvious explanation for why they should matter, it appears that they do indicate an increase in goal scoring rate.

²Incidentally, Crosby is first in this measure also, with 1.03 goals per 60 minutes.

Faceoffs Wins We note that if a predictor variable, such as faceoff wins or net faceoff wins, for example, was removed from our model, it does not necessarily mean that the statistic is not important. It may just mean that the statistic does not provide any information that is not already provided by other predictor variables. In the case of faceoff wins, one would think that they are important because possession is important. But several other statistics, namely shots, missed shots, and hits against, are all indicators of possession. Faceoffs wins, net faceoff wins, or faceoff percentage may not be adding much additional information about possession.

5 Future work and Conclusions

A natural idea for future work would be to develop similar models for goals against or net goals. One added complication that goalies are involved, but one could take steps to account for the strength of a team's goalies. We could also do similar studies for special teams situations. The significance of some statistics could be different for special teams situations as opposed to even strength situations. One problem with studying special teams would be the lack of data compared to even strength.

Another approach we could take is to split the season in half using a random sample of 41 games, instead of using odd and even games, fit the model, and repeat this process many times. This approach may yield more robust results. A statistic like total faceoffs may have come up as significant in our model by chance, but may not be significant if the data were split randomly and the models were fit several times. Drawing any firm conclusions about a statistic like faceoffs could be deferred until after this approach is taken.

One could also partition each season into smaller chunks of games. One may be interested in finding the variables that are the best predictors of performance in 10-game chunks, for example. The significant variables would likely change when using smaller chunks of data. For example, shots would likely become a stronger indicator than goals using a smaller chunks of games. Different sized partitions are studied in [13].

We noted that many of the predictor variables were correlated, which is one reason we chose the ridge regression model as our final model. Other methods, such as principal component regression or partial least squares regression, are commonly used when the predictors are correlated, and those models might improve performance. These methods are used in a forthcoming article [14]. In both of those methods, new uncorrelated predictor variables that are combinations of the original predictor variables are formed. It is typically harder to interpret the results for specific predictor variables (like we did for hits) with these models, which is one reason we chose to use OLS and ridge regression first. But if the main goal is to predict goal scoring, these models, and others like them, could prove useful. Bayesian techniques, non-parametric techniques, or time-series techniques, for example, could give better predictions.

The results of any of these models could be interpreted as expected goals. We believe that the use of expected goals in a ridge regression to estimate adjusted plus-minus, coupled with the results based on goals, shots, Fenwick rating and Corsi rating, can be useful to NHL teams, analysts, and fans as they evaluate the performance of teams and players.

References

- [1] JLikens, "Shots, Fenwick and Corsi," February 2011. <http://objectivenhl.blogspot.com/2011/02/shots-fenwick-and-corsi.html>, Accessed 09-03-2011.
- [2] V. Ferrari, "Zone Time," August 2008. <http://vhockey.blogspot.com/2008/08/zone-time.html>, Accessed 09-16-2011.
- [3] V. Ferrari, "Possession is Everything," May 2009. <http://vhockey.blogspot.com/2009/05/possession-is-everything.html>, Accessed 09-16-2011.

- [4] G. Desjardins, “Frequently Asked Questions #3: What is a Corsi Number,” October 2009. <http://www.arcticicehockey.com/2009/10/8/1076788/frequently-asked-questions-3-what>, Accessed 09-03-2011.
- [5] B. Macdonald, “Adjusted Plus-Minus for NHL Players using Ridge Regression,” *Submitted*, December 2011.
- [6] D. Rosenbaum, “Measuring How NBA Players Help Their Teams Win,” April 2004. <http://www.82games.com/comm30.htm>.
- [7] D. Lewin, “2004-2005 Adjusted Plus-Minus Ratings.,” 2007. <http://www.82games.com/lewin3.htm>.
- [8] E. Witus, “Count the Basket,” 2008. <http://www.countthebasket.com/blog/>.
- [9] S. Ilardi and A. Barzilai, “Adjusted Plus-Minus Ratings: New and Improved for 2007-2008,” 2008. <http://www.82games.com/ilardi2.htm>.
- [10] B. Macdonald, “A Regression-Based Adjusted Plus-Minus Statistic for NHL Players,” *Journal of Quantitative Analysis in Sports*, vol. 7, no. 3, p. 29, 2011.
- [11] B. Macdonald, “An Improved Adjusted Plus-Minus Statistic for NHL Players,” *Proceedings of the MIT Sloan Sports Analytics Conference*, March 2011.
- [12] J. Sill, “Improved NBA Adjusted +/- Using Regularization and Out-of-Sample Testing,” March 2010. <http://www.sloansportsconference.com/research-papers/2010-2/past-years/improved-nba-adjusted-using-regularization-and-out-of-sample-testing/>. Accessed 3-1-2011.
- [13] B. Macdonald, “Predicting Goals Scoring Rates in the NHL,” *In progress*, 2012.
- [14] B. Macdonald, “Predicting Goals Scoring Rates in the NHL using Principal Component Regression and Partial Least Squares,” *In progress*, 2012.

6 Appendix

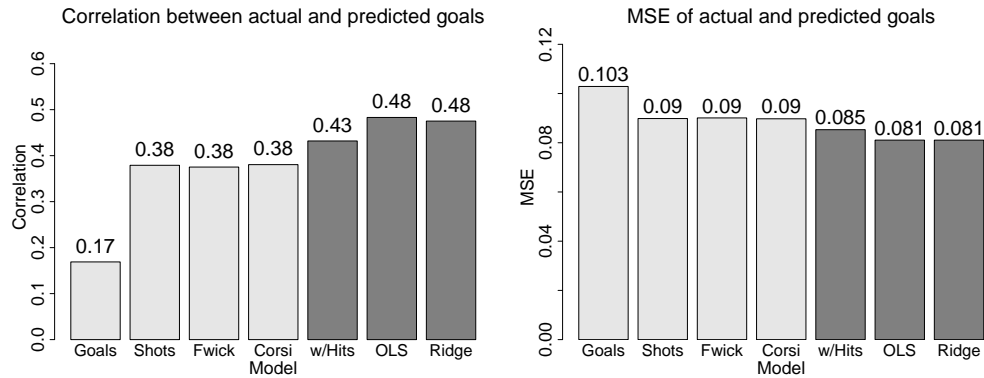


Figure 3: Correlation between (left) and MSE of (right) actual goals and predicted goals using Goals, Shots, Fenwick rating, Corsi rating, as well as our two new models, OLS and Ridge. This figure is similar to the cross-validation results in the right half of Figures 1 and 2, except that outliers were not removed. Note that our new models still outperform the other models when outliers are not removed.

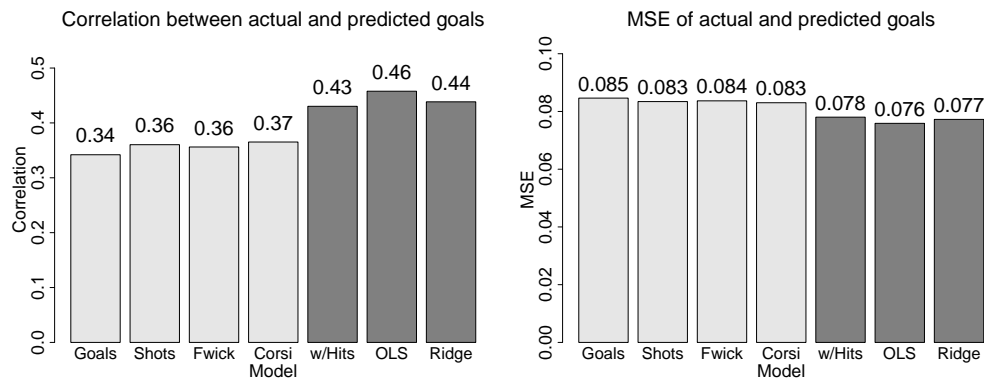


Figure 4: Correlation between (left) and MSE of (right) actual goals and predicted goals using Goals, Shots, Fenwick rating, Corsi rating as well as our two new models (OLS and Ridge). This figure is similar to the cross-validation results in the right half of Figures 1 and 2, except that these are the results when faceoffs are removed. Note that our new models still outperform the other models when faceoffs are not used.

Table 2: The top 10 offensive players in the NHL according to G_{EV}^{off} .

	Player	Pos	Team	E_{EV}^{off}	G_{EV}^{off}	S_{EV}^{off}	F_{EV}^{off}	C_{EV}^{off}	$E_{EV,60}^{off}$	$G_{EV,60}^{off}$	$S_{EV,60}^{off}$
1	Sidney Crosby	C	PIT	8	16	8	9	9	0.43	0.83	0.42
2	Pavel Datsyuk	C	DET	6	12	6	6	7	0.27	0.53	0.27
3	Alex Ovechkin	LW	WSH	7	11	11	13	15	0.31	0.46	0.45
4	Daniel Sedin	LW	VAN	7	10	9	9	9	0.31	0.47	0.44
5	Nathan Horton	RW	BOS	5	10	4	5	6	0.25	0.51	0.22
6	Jonathan Toews	C	CHI	6	9	5	4	4	0.28	0.45	0.22
7	Bobby Ryan	RW	ANA	5	9	4	5	5	0.28	0.58	0.25
8	Evgeni Malkin	C	PIT	5	8	6	6	6	0.26	0.40	0.31
9	Joe Thornton	C	S.J	5	8	6	6	8	0.20	0.34	0.26
10	Zach Parise	LW	N.J	6	8	9	9	8	0.39	0.49	0.55