Corsi and Fenwick: Advanced Statistics in the NHL

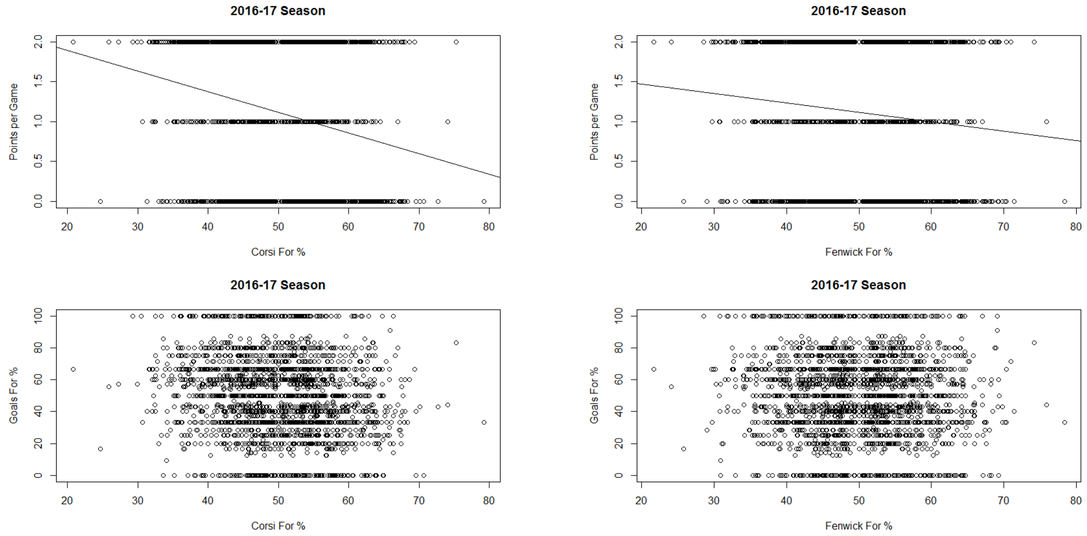
Taylor Paerels, Applied Mathematics/Statistics
Mentor: Dr. Adam King
Kellogg Honors College Capstone Project

Introduction

In the National Hockey League, the newest and most prevalent advanced metrics are **Corsi** and **Fenwick**. Corsi was standardized by Vic Ferrari, but named after former goalie and goaltending coach Jim Corsi. It involves the tracking of shot attempts taken over the course of a game. Shots can either reach the net, get saved by the goaltender, get blocked by opposition players, or miss entirely. Fenwick, a variant of Corsi edited by Matt Fenwick, records unblocked shot attempts, all shots not stopped by an opposing skater. Both metrics can be viewed for or against any particular team.

Corsi = Missed Shots + Blocked Shots + Shots On Goal

Fenwick = Missed Shots + Shots On Goal
= Corsi - Blocked Shots



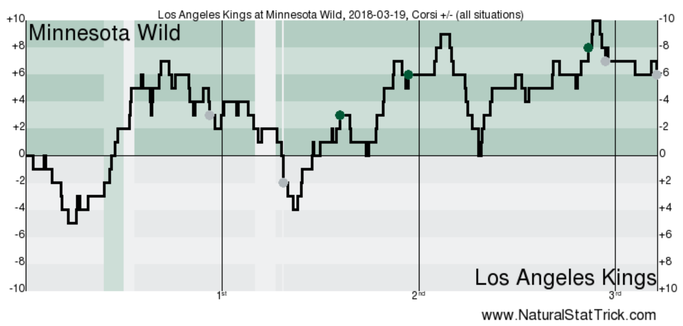
Methods

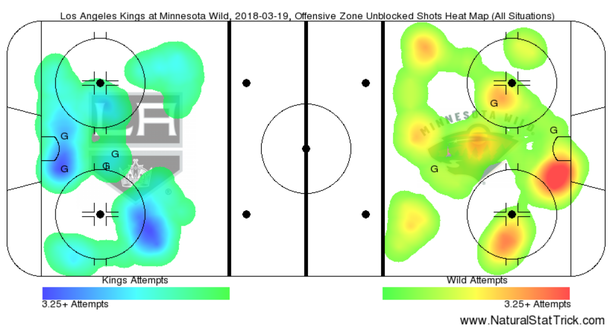
In determining which “fancy stat" was a better predictor of success, several testing methods were used including simple linear regression and multiple linear regression. Rather than begin with a sample size of four years’ worth of 30 teams' 82 games schedules, the 2016-17 season was observed by itself in order to look for trends which may hold up over a longer period of time. In the models, the categories were compared relative to outputs of points earned per game and the percentage of goals scored per game. Due to the NHL’s standings system, points, rather than wins, determine playoff seeding. A win of any kind results in two points earned, while a loss in regulation yields zero points. Meanwhile a single point can be earned through a loss occurring in sudden death overtime or a shootout (where each team sends a skater one-on-one against the opponent's goaltender to see who can score more over 3 rounds, with more skaters competing if the score remains tied and the format becomes sudden death). This predicted data set *did* show whether or not a team won, but because of the small number of discrete response values, problems arose. Initial testing yielded probabilities as high as 30 percent.

Results

SLR Models:
Corsi For vs Total Points: p-value = 1.039×10^{-5}, adjusted R^2 = 0.1453
Fenwick For vs Total Points: p-value = 1.131×10^{-6}, adjusted R^2 = 0.1757
Shots For vs Total Points: p-value = 3.022×10^{-8}, adjusted R^2 = 0.2233
Thus, all of these results are statistically significant. However, their adjusted R^2 values show that less than a quarter of the data is explained by the predictors.

MLR Models:
Divided Corsi For vs Total Points: p-values = 0.687 (MS), 0.505 (BS), 1.99×10^{-6} (SOG), 6.464×10^{-7} (overall)
adjusted R^2 = 0.2176
Divided Fenwick For vs Total Points: p-values = 0.406 (MS), 1.01×10^{-6} (SOG), 1.634×10^{-7} (overall)
adjusted R^2 = 0.2213
Only the p-values (in both models) for shots on goal and overall, are statistically significant.
Although none of these R^2 values are not inherently bad, they are difficult to use in trying to make predictions about success.



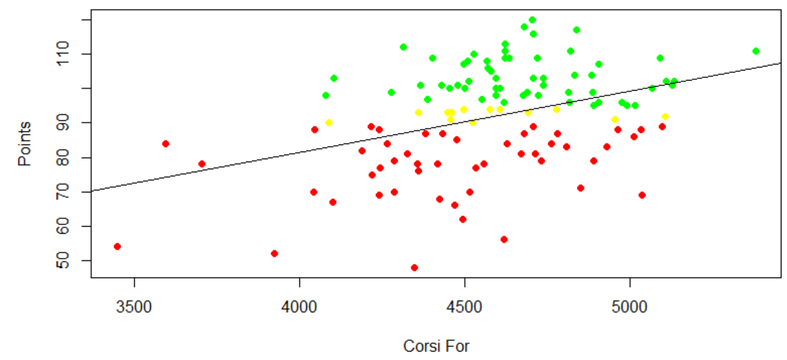


The website Natural Stat Trick, a play on the phrase “natural hat trick” – three consecutive goals scored by the same player, tracks Corsi graphically in two ways. On the left, a game flow map which shows the change in Corsi over the course of a game. In this example, upward movements are shot attempts by the Minnesota Wild, while downward movements are attempts by the Los Angeles Kings. Vertical bars represent power plays/penalty kills. The graph on the right is a heat map showing where shots were taken relative to the goal.

Discussion

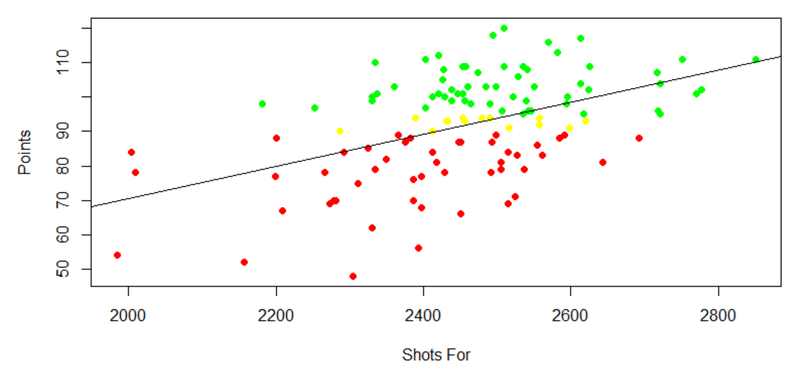
One potential change which could be made to try linking offensive and defensive numbers would be to view the typical Fenwick For values in addition to the number of blocked shots “against" from Corsi Against. Essentially, this would mean looking at the number of times a certain team takes a shot, which is either on goal or missed, plus the number of times that the same team blocks shots taken by opposing players. While this would not work as well to explain the effectiveness of a team offensively, it would be interesting to note whether a team is subsequently more successful with a more well-rounded game plan. Furthermore, testing could involve the use of the strength variants. It is possible that one team's success correlates with their time spent on the power play and so Corsi and Fenwick values from that data set may be more convenient. Conversely, a team may be quite successful while playing at even strength, but may struggles on the power play, leading them to experience having a stronger correlation between their Corsi and Fenwick during normal 5 skaters versus 5 skaters segments. Finally, the ultimate measure of success is whether or not a team wins a championship. The playoffs in the NHL involve the top eight teams in each conference competing in best of seven series. Theoretically, the team with the best statistics entering the postseason should advance the farthest, but this is not always the case.

Special thanks to my family, especially my sister, for their help and support on this project.



Total Points = 10.452785 + 0.017745(Corsi For)

Total Points = -2.995729 + 0.027749(Fenwick For)



Total Points = -22.833515 + 0.046627(Shots For)

All statistics were obtained from naturalstattrick.com