Code ▾

Examining Repeatability in Different Rush Situations

Jason Baik

6/7/2018

Key Takeaways

(TODO: Revise this part)

Introduction

Have you ever wondered about the numbers behind offensive player advantage (odd man) rushes? Are the number of shots repeatable over an entire season? What about shooting percentages? If they are repeatable, do zones of the primary pass (a pass right before a shot) affect this repeatability? And what about rebounds and rebound shooting percentages (the goals scored off of rebounds)?

To investigate these questions, I will be using the Passing Project Data from the 2015-2016, 2016-2017, and 2017-2018 NHL seasons organized by Ryan Stimson (https://twitter.com/RK_Stimp) and Corey Sznajder (https://twitter.com/ShutdownLine). The data was collected by a group of volunteers who tracked every pass leading to a shot. According to Ryan's MIT Sloan Sports Analytics Conference paper (http://www.sloansportsconference.com/wp-content/uploads/2017/02/1614.pdf), "Each tracker was required to complete 2 - 3 training games for validation by the project leader to ensure consistency in classification of passes and recording of data. Occasional spot-checking of games was also conducted."

I would like to thank my mentor of the Hockey Graphs Mentorship Program, Sam Ventura, for his time and advice.

Data

<div align="right">

Code

</div>

shot_on_goal	goal	odd_man	shot_type	a1_zone	rebound_shot_on_goal
1	0	NA	W	nl	0
0	0	NA	W	of	0
0	0	NA	T	ol	0
1	1	NA	W	dsl	0
0	0	NA	S	or	0
0	0	NA	W	nc	0
0	0	3-2	T	ol	0
0	0	NA	T	ol	0
1	0	NA	W	dsl	0
1	0	NA	W	nl	0

There are 45 variables and over 150,000 observations in this dataset. Some notable variables that I will be using for my analysis include shot_on_goal (1 for yes and 0 for no), goal (1 for yes and 0 for no), odd_man (number of players involved in player advantages during rushes), shooting_percentage

(number of goals / number of shots on goals), shot_type (one-timer, slap shot, wrist/snap shot, backhand, etc), a1_zone (zone that the primary assist originated), and rebound_shot_on_goal (if a rebound occured and resulted in a shot on goal).

There are two analysis based upon this datset; Ryan Stimson's paper (http://www.sloansportsconference.com/wp-content/uploads/2017/02/1614.pdf) discusses repeatability of certain metrics on a team level and player level. A NHL Numbers post (https://www.nhlnumbers.com/2016/08/10/passing-project-dangerous-primary-shot-contributions) covers repeatability of passes that crosses through the royal road (the line that goes directly through the middle of the ice from one net to the other) and provides visualizations.

In my blog post, I will look at *repeatability* of shots taken during rushes with player advantages, shooting percentages, rebound shots, and rebound shooting percentages in 5 on 5 situations.

Visual Identification of Linear Regression Graphs and Repeatability

This article is filled with scatterplots with linear regression lines. I've grouped the data by teams and plotted the first half statistics from the 2015-2016, 2016-2017, and 2017-2018 seasons on the x-axis, the second half statistics from the 2015-2016, 2016-2017, and 2017-2018 season the on the y-axis, and generated a smoothed line with a confidence band. It is a 95% confidence interval on the slope of the regression line.

- If I can draw a horizontal line through the confidence bands, this means there exists a regression line with a slope of 0. This further indicates that there is *no significant relationship between the x-variable and the y-variable.*
- If I can not draw a horizontal line through the confidence bands, this means that a regression line with a slope of 0 falls outside the 95% confidence band. In other words, there is *a significant relationship between the x-variable and the y-variable.*

In this analysis, I designated the midpoint of each of the three seasons as Jan 1st. All the games before that date are labelled as the first half of the season while the games after that date are labelled as the second half. In terms of the linear regression graphs, the x-variable is a statistic from the first half of the season and the y-variable is a statistic from the second half of the season. As a result, we are really examining the relationship between the past and future of a statistic, or the *repeatability* of a statistic.

Repeatability of odd-man rush shots

First, let's examine if shots taken in rushes with player advantages are repeatable.

Code

Code

Code

Repeatability of Total Shots Facetted by Odd Man Rushes
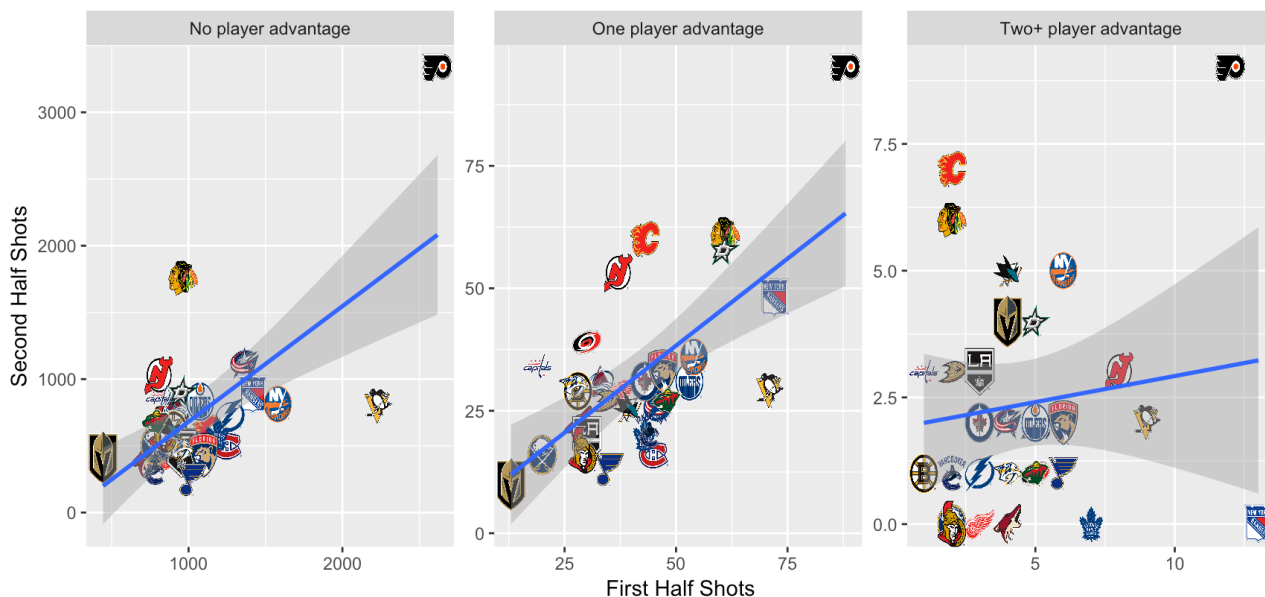

- Shots taken on rushes where the attacking team has more players than the defending team are repeatable over a season when there is no player advantage and one player advantage (Breakaways, 2 on 1 rushes, 3 on 2 rushes, and 4 on 3 rushes).
- This leads me to think that its increadibly hard to predict the number of shots in two-plus player advantages (2 on 0 rushes", 3 on 1 rushes, 4 on 2 rushes, 3 on 0 rushes) over a season.
- Perhaps, teams with an incredible advantage in rushes have too many options at their disposal, and thus these rushes do not always result in shots on net.
- Rather, they could make one extra pass that extinguishes the offensive threat or the extra pass could be intercepted by the defender(s).
- As there is *no repeatability* in two plus player advantage, I will ignore these situations for this analysis.

Repeatability of shooting percentage

Let's examine shooting percentages and see if we can detect the same trend.

Code

Repeatability of Shooting % Facetted by Odd Man Rushes
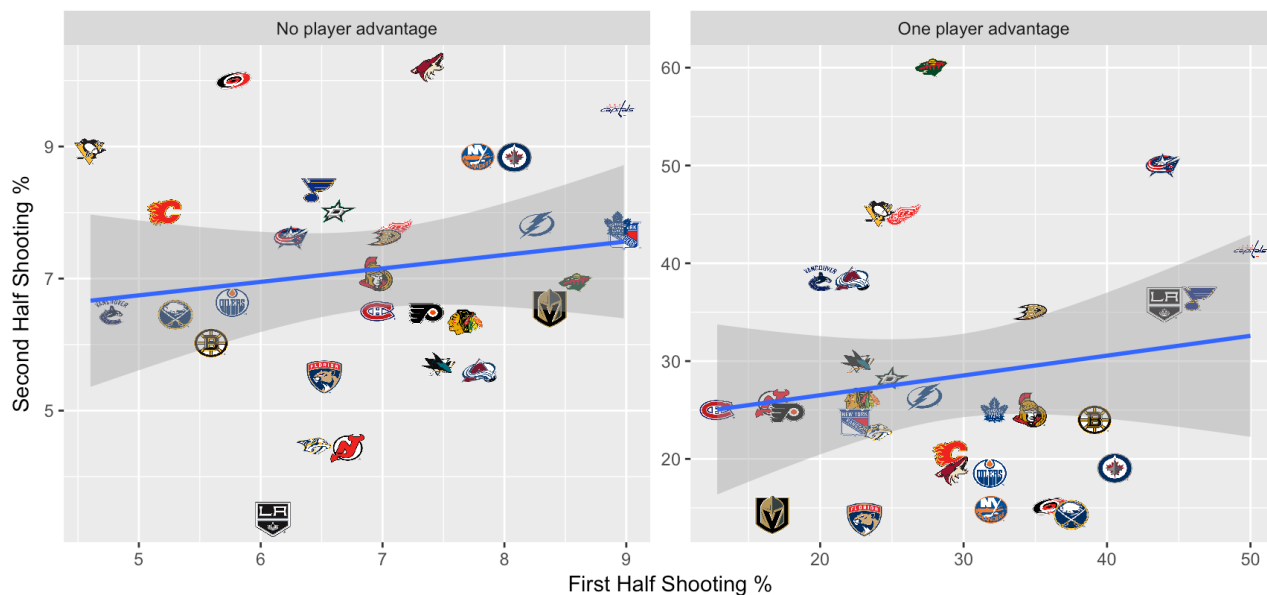
- The scales for the two facetted graphs are vastly different: shooting percentages across the whole season are *significantly higher* in one player advantage rushes, which is reasonable given the offensive freedom players have.
- Interestingly, shooting percentages are indeed not repeatable over a season during no player advantage rushes since I can draw a horizontal graph through the confidence band.
- Shooting percentages are not repeatable during one player advantage rushes as I can draw a horizontal graph through the confidence band.

Shooting percentages by Zone of Primary Pass

Now, I observe the repeatability of shooting percentages facetted by the zone the primary pass (the pass right before the shot) originated from.

Code

Repeatability of Shooting % During Odd Man Rushes Facetted By Lanes of passes
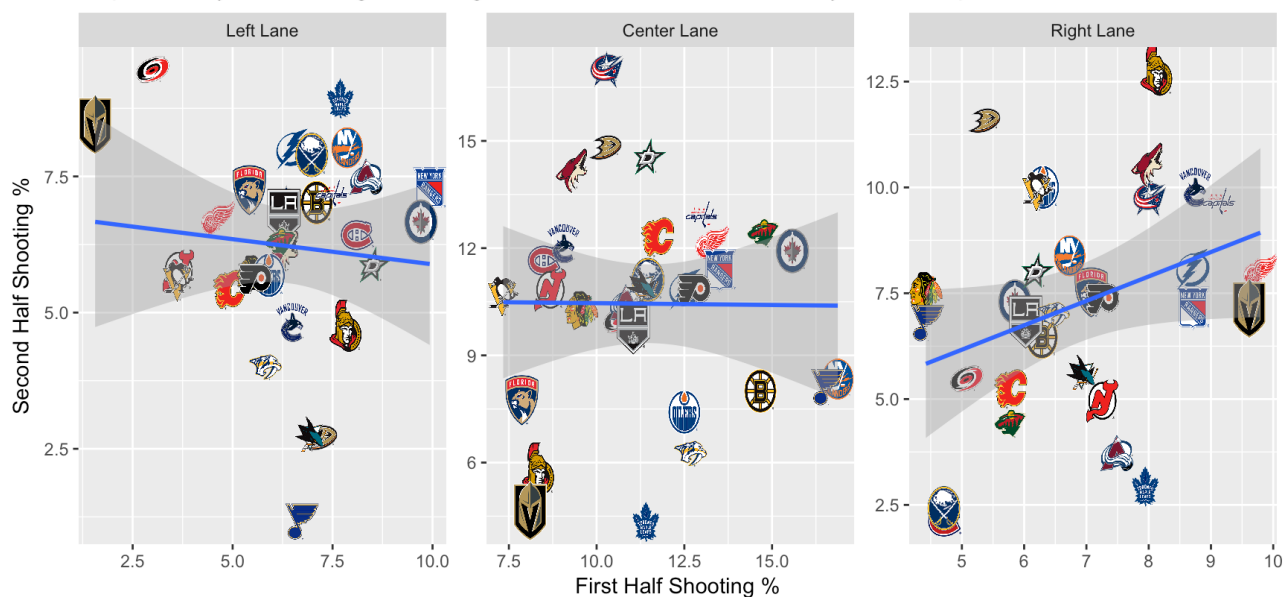
- The center lane facet shows no repeatability between first half shooting percentage and second half shooting percentage, but is the closest out of all three graphs to being repeatable.

- None of these graphs show any signs of repeatability since I can draw a horizontal line through all these graphs.

- As a result, *the lane of a primary pass does not help explain the discrepancy in repeatability of shooting percentages in odd man rushes.*

Repeatability of rebound shots

During the course of a game, there is bound to be a rebound. Most "trash" goals scored in the "dirty" areas are scored off of a rebound. Let's examine the repeatability of rebound shots over a season.

Code
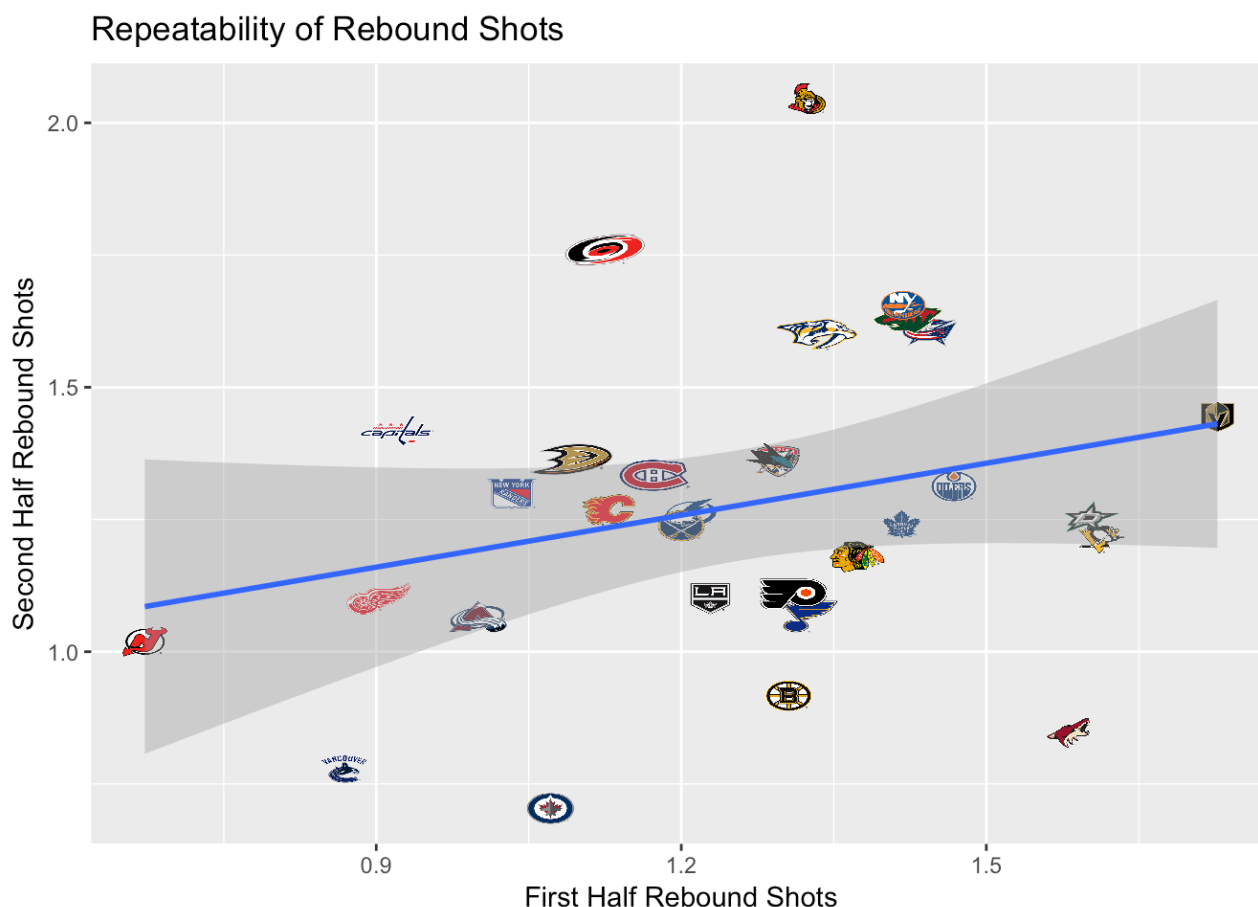
- This graph clearly shows that rebound shots are not repeatable over a given season.
- This suggests that rebounds are intrinsically hard to predict over an entire season. Teams face different styles of goaltenders and even different goaltenders for each team if the team plays its backup or AHL goaltender. Perhaps, goaltenders have varying abilities to control rebounds.

Repeatability of rebound shooting percentage

To dig deeper into the effectiveness of rebounds in offensive situations, I've defined *rebound shooting percentage as the percentage of rebound goals scored.* Let's now examine its repeatability.

Code

Repeatability of Rebound Shooting %
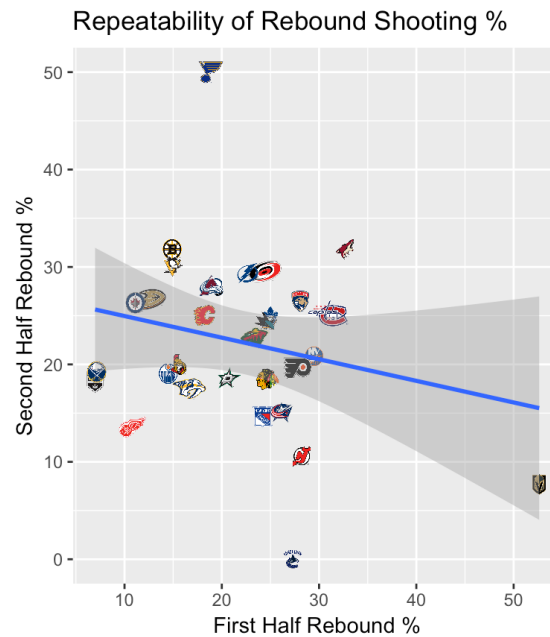
- Overall, rebound shooting percentages are not repeatable since I can draw a horizontal line through the confidence band.
- This again suggests that rebounds are a random occurence during a hockey game.
- Furthermore, over an entire season, there exist unpredictable offensive situations that lead to rebounds and thus, make it impossible to predict.

Conclusion

TODO