Schedule Adjustment for counting stats

Micah Blake McCurdy
@IneffectiveMath

27 November 2014

We all know that some games are easier to play than others, and we all make adjustments in our head and in our arguments that make reference to these ideas. Three points out of a possible six on that Californian road-trip are good, considering how good those teams are; putting up 51% possession numbers against Buffalo or Toronto or Ottawa or Colorado just isn't that impressive considering how those teams normally drive play, or, err, don't.

These conversations only intensify as the playoffs roll around—really, how good are the Penguins, who put up big numbers in the "obviously" weaker East, compared to Chicago, who are routinely near the top of the "much harder" western conference? How can we compare Pacific teams, of which all save Calgary have respectable possession numbers, with atlantic teams, who play lots of games against the two weak Ontario teams and the extremely weak Sabres?

Intuitively, we know that such adjustments are necessary. For instance, if our favourite team repeatedly puts up 50% possession numbers against teams that we somehow KNOW are 60% possession teams, it would be idiotic to say that our team is "A 50% team", they're clearly holding their own against 60% teams and it's much more credible to say that our favourite team is also a 60% team. Various people have developed schemes for adjusting win-loss records, as well as point systems that are explicitly geared at prediction, such as RPI, or various ELO-style schemes, or ZRatings.

I am very fond of predictive statistics, as connoisseurs of my previous articles will doubtless be familiar, but I view such adjustments as primarily descriptive. I want to be able to look at a length of time and say "during this time, my favourite team played *like a 50% team*", for instance. In this way I hope to escape being deceived by strong results over weak opponents or by weak results against strong opponents. (As it happens, I find that schedule-adjusted measures are, as one might expect, more predictive than the corresponding "raw" measures. But that is another matter, for another day.)

In this article, I focus on adjusting possession statistics for strength-of-schedule. For reasons that will be fully explained in future articles, I intensely dislike measuring possession in percentages, and prefer to maintain separate tallies of offence and defence. The method I present here takes raw measurements of possession, expressed as counts (possibly adjusted for other effects, such as score effects, home/road effects, time-of-game effects, or rink effects), and produces schedule-adjusted counts.

First, fix a set of games to consider. It is not necessary that the number of games played by each team be the same.

To illustrate, let's work through a few steps of a gory calculation, using the state of the NHL 2014-2015 after the games of Wednesday November 26, 2014. As I mentioned earlier, any counting stats can be adjusted, but for simplicity let's look at (raw) 5v5 Corsi events. First, we fix an order for the teams; we'll use alphabetical order:

1	Anaheim Ducks
2	Arizona Coyotes
3	Boston Bruins
4	Buffalo Sabres
5	Calgary Flames
6	Carolina Hurricanes
7	Chicago Blackhawks
8	Colorado Avalanche
9	Columbus Blue Jackets
10	Dallas Stars
11	Detroit Red Wings
12	Edmonton Oilers
13	Florida Panthers
14	Los Angeles Kings
15	Minnesota Wild
16	Montreal Canadiens
17	Nashville Predators
18	New Jersey Devils
19	New York Islanders
20	New York Rangers
21	Ottawa Senators
22	Philadelphia Flyers
23	Pittsburgh Penguins
24	San Jose Sharks
25	St. Louis Blues
26	Tampa Bay Lightning
27	Toronto Maple Leafs
28	Vancouver Canucks

29	Washington Capitals
30	Winnipeg Jets

Next, form the matrix of raw events that are of interest. In our case, we'll use corsi events, but the adjustment method works for anything that can be counted. Specifically, form the 30-by-30 matrix whose (i, j) entry is the average number of events obtained by team i against team j in the games under consideration. Here, we have:

$$S = \begin{bmatrix}
0 & 114 & 0 & 105 & 92 & 0 & 36 & 36 & 43 & 28 & 35 & 0 & 67 & 117 & 39 & 0 & 0 & 0 & 56 & 0 & 0 & 34 & 34 & 44 & 79 & 0 & 0 & 88 & 0 & 0 \\
87 & 0 & 0 & 0 & 46 & 35 & 0 & 49 & 0 & 99 & 0 & 83 & 81 & 46 & 47 & 0 & 39 & 0 & 27 & 0 & 0 & 0 & 0 & 35 & 37 & 52 & 41 & 33 & 79 & 43 \\
0 & 0 & 0 & 110 & 0 & 39 & 0 & 44 & 60 & 0 & 64 & 41 & 49 & 0 & 49 & 143 & 0 & 41 & 63 & 0 & 56 & 51 & 43 & 55 & 36 & 0 & 106 & 0 & 42 & 0 \\
73 & 0 & 61 & 0 & 0 & 24 & 25 & 0 & 33 & 0 & 26 & 47 & 32 & 40 & 27 & 32 & 0 & 0 & 0 & 0 & 0 & 0 & 62 & 55 & 37 & 0 & 69 & 0 & 40 & 30 \\
74 & 39 & 0 & 0 & 0 & 94 & 60 & 0 & 37 & 0 & 0 & 34 & 33 & 0 & 0 & 80 & 95 & 40 & 0 & 0 & 43 & 0 & 0 & 0 & 35 & 74 & 0 & 39 & 80 & 39 \\
0 & 45 & 48 & 55 & 96 & 0 & 0 & 56 & 85 & 45 & 0 & 50 & 44 & 91 & 0 & 0 & 0 & 0 & 72 & 29 & 0 & 0 & 0 & 85 & 0 & 0 & 0 & 44 & 48 & 91 \\
55 & 0 & 0 & 62 & 116 & 0 & 0 & 54 & 0 & 104 & 44 & 35 & 0 & 0 & 0 & 52 & 93 & 0 & 0 & 0 & 87 & 52 & 0 & 48 & 43 & 42 & 67 & 42 & 56 & 70 \\
30 & 50 & 37 & 0 & 0 & 47 & 45 & 0 & 0 & 0 & 0 & 0 & 35 & 0 & 60 & 46 & 0 & 32 & 61 & 39 & 35 & 51 & 0 & 36 & 31 & 0 & 97 & 83 & 49 & 33 \\
29 & 0 & 48 & 54 & 38 & 70 & 0 & 0 & 0 & 45 & 47 & 0 & 0 & 39 & 0 & 0 & 0 & 27 & 0 & 40 & 99 & 90 & 0 & 65 & 0 & 49 & 41 & 0 & 50 & 38 \\
37 & 91 & 0 & 0 & 0 & 45 & 89 & 0 & 59 & 0 & 0 & 53 & 0 & 114 & 108 & 0 & 107 & 44 & 49 & 0 & 0 & 42 & 47 & 62 & 43 & 0 & 0 & 32 & 0 & 0 \\
42 & 0 & 68 & 50 & 0 & 0 & 42 & 0 & 42 & 0 & 0 & 0 & 0 & 42 & 0 & 78 & 0 & 30 & 0 & 44 & 88 & 97 & 33 & 0 & 0 & 28 & 128 & 0 & 27 & 36 \\
0 & 86 & 42 & 79 & 68 & 49 & 39 & 0 & 0 & 64 & 0 & 0 & 0 & 45 & 0 & 33 & 84 & 64 & 0 & 48 & 35 & 58 & 0 & 0 & 0 & 40 & 0 & 168 & 30 & 0 \\
49 & 74 & 39 & 50 & 52 & 48 & 0 & 51 & 0 & 0 & 0 & 0 & 0 & 41 & 39 & 0 & 39 & 31 & 43 & 0 & 42 & 71 & 0 & 76 & 0 & 31 & 0 & 0 & 35 & 0 \\
91 & 46 & 0 & 69 & 0 & 97 & 0 & 0 & 50 & 140 & 33 & 38 & 38 & 0 & 58 & 0 & 36 & 0 & 42 & 0 & 0 & 61 & 42 & 56 & 33 & 0 & 0 & 53 & 0 & 47 \\
40 & 50 & 64 & 58 & 0 & 0 & 0 & 96 & 0 & 89 & 0 & 0 & 51 & 95 & 0 & 64 & 0 & 29 & 0 & 33 & 42 & 35 & 44 & 64 & 0 & 83 & 0 & 0 & 0 & 25 \\
0 & 0 & 115 & 50 & 77 & 0 & 39 & 57 & 0 & 0 & 78 & 48 & 0 & 0 & 44 & 0 & 0 & 0 & 0 & 86 & 0 & 118 & 52 & 0 & 54 & 41 & 53 & 38 & 27 & 34 \\
0 & 36 & 0 & 0 & 120 & 0 & 87 & 0 & 0 & 110 & 0 & 100 & 57 & 36 & 0 & 0 & 0 & 0 & 0 & 0 & 104 & 0 & 42 & 0 & 93 & 0 & 48 & 48 & 0 & 133 \\
0 & 0 & 50 & 0 & 46 & 0 & 0 & 36 & 35 & 39 & 33 & 32 & 22 & 0 & 24 & 0 & 0 & 0 & 0 & 34 & 44 & 45 & 33 & 41 & 85 & 32 & 0 & 44 & 82 & 82 \\
56 & 51 & 52 & 0 & 0 & 90 & 0 & 106 & 0 & 42 & 0 & 0 & 57 & 38 & 0 & 0 & 0 & 0 & 0 & 41 & 0 & 65 & 162 & 121 & 0 & 74 & 41 & 0 & 57 & 37 \\
0 & 0 & 0 & 0 & 0 & 45 & 0 & 47 & 36 & 0 & 35 & 42 & 0 & 0 & 27 & 114 & 0 & 31 & 60 & 0 & 0 & 36 & 72 & 48 & 88 & 76 & 108 & 0 & 0 & 45 \\
0 & 0 & 42 & 0 & 36 & 0 & 83 & 37 & 89 & 0 & 75 & 32 & 29 & 0 & 23 & 0 & 72 & 45 & 0 & 0 & 0 & 0 & 0 & 0 & 95 & 38 & 58 & 39 & 0 & 57 \\
57 & 0 & 34 & 0 & 0 & 0 & 46 & 34 & 89 & 66 & 66 & 53 & 70 & 58 & 48 & 90 & 0 & 54 & 29 & 38 & 0 & 0 & 40 & 0 & 0 & 40 & 0 & 0 & 0 & 0 \\
47 & 0 & 48 & 91 & 0 & 0 & 0 & 0 & 0 & 48 & 31 & 0 & 0 & 25 & 38 & 44 & 31 & 30 & 131 & 111 & 0 & 41 & 0 & 0 & 0 & 0 & 163 & 0 & 0 & 42 \\
43 & 67 & 44 & 111 & 0 & 40 & 45 & 63 & 103 & 59 & 0 & 0 & 101 & 45 & 41 & 0 & 0 & 33 & 105 & 42 & 0 & 0 & 0 & 0 & 0 & 53 & 0 & 70 & 37 & 36 \\
94 & 56 & 50 & 39 & 47 & 0 & 40 & 39 & 0 & 34 & 0 & 0 & 0 & 53 & 0 & 54 & 86 & 76 & 0 & 87 & 65 & 0 & 0 & 0 & 0 & 0 & 0 & 51 & 49 & 37 \\
0 & 55 & 0 & 0 & 98 & 0 & 38 & 0 & 55 & 0 & 36 & 41 & 41 & 0 & 92 & 44 & 0 & 37 & 82 & 92 & 47 & 35 & 0 & 42 & 0 & 0 & 56 & 32 & 39 & 43 \\
0 & 36 & 87 & 105 & 0 & 0 & 37 & 112 & 39 & 0 & 117 & 0 & 0 & 0 & 0 & 45 & 48 & 0 & 33 & 95 & 34 & 0 & 104 & 0 & 0 & 41 & 0 & 0 & 0 & 0 \\
106 & 52 & 0 & 0 & 42 & 39 & 52 & 92 & 0 & 52 & 0 & 186 & 0 & 27 & 0 & 51 & 44 & 52 & 0 & 0 & 46 & 0 & 0 & 20 & 33 & 33 & 0 & 0 & 44 & 0 \\
0 & 79 & 25 & 71 & 97 & 54 & 27 & 46 & 36 & 0 & 38 & 60 & 49 & 0 & 0 & 41 & 0 & 74 & 38 & 0 & 0 & 0 & 0 & 55 & 45 & 50 & 0 & 42 & 0 & 0 \\
0 & 22 & 0 & 35 & 58 & 97 & 38 & 61 & 34 & 0 & 52 & 0 & 0 & 37 & 33 & 53 & 123 & 80 & 36 & 48 & 45 & 0 & 39 & 40 & 49 & 43 & 0 & 0 & 0 & 0
\end{bmatrix}$$

For example, consider Ottawa (number 21) and Winnipeg (number 30). If we look in the 21st row, in the 30th column, we see 57, that is, the number of Corsi events Ottawa generated against Winnipeg in the only game they have played this year. If we look in the 30th row, in the 21st column, we see 45, which is the number of Corsi events Winnipeg generated against Ottawa. It was a fun game. Notice that many entries are zero, since many pairs of teams have not played one another this year. Of course the diagonal entries are all zero, since no team has played itself. Note also that some entries are quite high, since some teams have played one another more than once.

We will need to know the average number of Corsi events generated by one team in a game, in this sample it is 43.7. We call this number m (for "mean"). We want to reward teams that do "better than expected [against a given team]" and punish teams that do "worse than expected [against a given team]". In the set of games being considered, teams allowing fewer than m events are 'strong' defensively; teams generating more than m events are 'strong' offensively, and so on.

We want to use this matrix S, which will not change, to estimate the relative offensive and defensive ability of all the teams. We begin by implicitly assuming that they are all equally good at offense and defense, we will discover the truth in time.

To make our first adjustment, we first add up all of the events generated by a given team; that is, add up all of the entries in a given row. This produces a vector whose entries are the total number of events for each team over the sample at hand, then we divide the entries in this vector by the number of games played against each team; and finally we divide the vector by m, the average number of events generated by a team in the sample.

In symbols, we form F_1 defined by:

$$(F_1)_i = \frac{1}{m} \cdot \frac{\sum_{j=1}^{30} S_{ij}}{\sum_{j=1}^{30} N_{ij}}$$

In our example, this vector is:

$$\begin{bmatrix} 1.04 & 0.96 & 1.09 & 0.74 & 0.89 & 1.07 & 1.17 & 0.89 & 0.95 & 1.06 & 0.91 & 1.07 & 0.93 & 1.03 & 1.05 & 1.01 & 1.11 & 0.87 & 1.13 & 0.99 & 0.93 & 0.99 & 1.00 & 1.13 & 1.00 & 1.00 & 0.97 & 1.01 & 1.01 & 0.98 \end{bmatrix}$$

Where I've rounded to only two decimal places for ease of explanation.

This gives a vector whose entries are all the relative offensive strength of each team, compared to average. The highest entry is 1.17, in the seventh spot; this is Chicago. The Blackhawks are generating, on average, 17% more than average events. The lowest entry is 0.74, the fourth entry; this is Buffalo. They are generating 26% fewer corsi events than average. This vector F_1 is our first estimate of offensive strengths F_1.

Similarly, we make a calculation of defensive strength by summing the columns of S, that is, the shots allowed by a given team. Dividing the entries of this vector by the number of games played between the relevant teams, and then dividing by the average number of events m gives a vector whose entries are a measure of each team's defensive ability.

In symbols, we form A_1 defined by:

$$(A_1)_j = \frac{1}{m} \cdot \frac{\sum_{i=1}^{30} S_{ij}}{\sum_{i=1}^{30} N_{ij}}$$

In our example, A_1 is:

$$\begin{bmatrix} 1.01 & 1.04 & 0.95 & 1.24 & 1.12 & 1.00 & 0.90 & 1.11 & 1.01 & 1.11 & 0.84 & 1.02 & 0.98 & 0.99 & 0.87 & 1.06 & 0.98 & 0.89 & 0.97 & 0.99 & 0.99 & 1.07 & 0.93 & 1.04 & 0.95 & 0.92 & 1.12 & 0.99 & 0.95 & 0.95 \end{bmatrix}$$

The highest entry is 1.24, the fourth entry; this is Buffalo again. The sabres allow 24% more corsi events than average. The lowest entry is 0.84, the eleventh entry; this is Detroit. The Red Wings allow 16% fewer events against than average. Let us call this vector defensive strengths A_1.

We can use these vectors (F_1, A_1) to further refine our estimates of strength. Now that we have an idea of the strengths of the teams, we can re-interpret their event totals in that light. Define F_2 using a slightly modified formula:

$$(F_2)_i = \frac{1}{m} \cdot \frac{\sum_{j=1}^{30} S_{ij}/(A_1)_j}{\sum_{j=1}^{30} N_{ij}}$$

The key difference is the weighting of the events against a given team by that team's defensive ability. For instance, let us say we are going to calculate $(F_2)_{30}$, that is, Winnipeg's offensive ability. The first non-zero term in the sum in the numerator is $S_{30,2}/(A_1)_2 = 22/1.04 = 21.15$. Team 2 is Arizona, and they are slightly softer than the league average, allowing 4% more events than average. Therefore, Winnipeg's 22 events against them are counted as 21.15 events, since events against Arizona are slightly easier to come by than we would expect had Winnipeg played an "average" team. Slightly later in the computation, another term in the sum in the numerator is $S_{30,7}/(A_1)_7 = 38/0.90 = 42.22$. Team 7 is Chicago, who are quite stingy, allowing 10% fewer events than average, so Winnipeg's 38 events in their two games against the Blackhawks are counted as 42.22 events. Carrying on in this way, we compute that F_2 is:

$$\begin{bmatrix} 1.03 & 0.95 & 1.08 & 0.75 & 0.90 & 1.06 & 1.15 & 0.90 & 0.94 & 1.08 & 0.90 & 1.07 & 0.92 & 1.02 & 1.04 & 1.01 & 1.10 & 0.88 & 1.13 & 1.00 & 0.94 & 1.01 & 0.99 & 1.13 & 0.99 & 1.01 & 0.97 & 1.01 & 1.00 & 0.98 \end{bmatrix}$$

Comparing this vector to F_1 reveals that very little has changed (especially when rounding to two decimal places), but a few teams have moved. For example, Chicago's entry of 1.17 has been revised to only 1.15, since Chicago's opponents have been, taken together, not especially strong defensively and thus Chicago's high totals are not quite as impressive as they appeared at first.

Similarly, to adjust our measure of defensive ability, form the weighted sum of all the entries in a column of S, where the weights are the inverses of the vector of strengths, F_1, computed previously, divided by the number of games against each opponent, divided by the average statistic m. Call this vector A_2, our refined vector of coefficients of defensive strength. In symbols, this is:

$$(A_2)_j = \frac{1}{m} \cdot \frac{\sum_{i=1}^{30} S_{ij}/(F_1)_i}{\sum_{i=1}^{30} N_{ij}}$$

For example, let's examine Anaheim's defensive strength, that is, at computing $(A_2)_1$. One of the terms that appears in the sum in the numerator is $S_{21}/(F_1)_2 = 87/0.96 = 90.63$. Team number 2 is Arizona; their offense is below average ($0.96 < 1$), and so the 87 events that Anaheim allowed in their two games against Arizona are counted as 90.63 events instead. Later in the calculation, the 55 events Anaheim allowed against Chicago (offensive ability $(F_1)_7 = 1.17$) are counted as only $55/1.17 = 47.01$ events. Continuing in this way, we compute A_2 is:

$$\begin{bmatrix} 1.01 & 1.03 & 0.96 & 1.22 & 1.11 & 1.00 & 0.91 & 1.10 & 1.01 & 1.08 & 0.85 & 1.01 & 0.97 & 0.98 & 0.87 & 1.06 & 0.98 & 0.89 & 0.96 & 0.99 & 0.99 & 1.07 & 0.92 & 1.04 & 0.96 & 0.92 & 1.13 & 0.98 & 0.96 & 0.95 \end{bmatrix}$$

As in the offense case, we see only small changes. Buffalo's coefficient has decreased from 1.24 to 1.22, they have faced strong offensive opposition, and so their defensive woes are not *quite* as woeful as they first appeared. It is crucial to point out that Buffalo have faced stronger opposition *than the league average*, not merely strong compared to Buffalo. It is quite possible for the worst team in a given ranking to have an "easy" schedule and find that strength-of-schedule adjustments make them look still worse.

So, we have a way of taking estimates (F_i, A_i) of offensive and defensive strength, and turning them into better measures (F_{i+1}, A_{i+1}). We repeat this process many, many times (a hundred times is usually enough, computers help here) until the output vectors are imperceptibly different from the input vectors, and we consider these vectors (F, A) to be the true schedule-adjusted offensive and defensive skill of the teams as shown over the games in question.

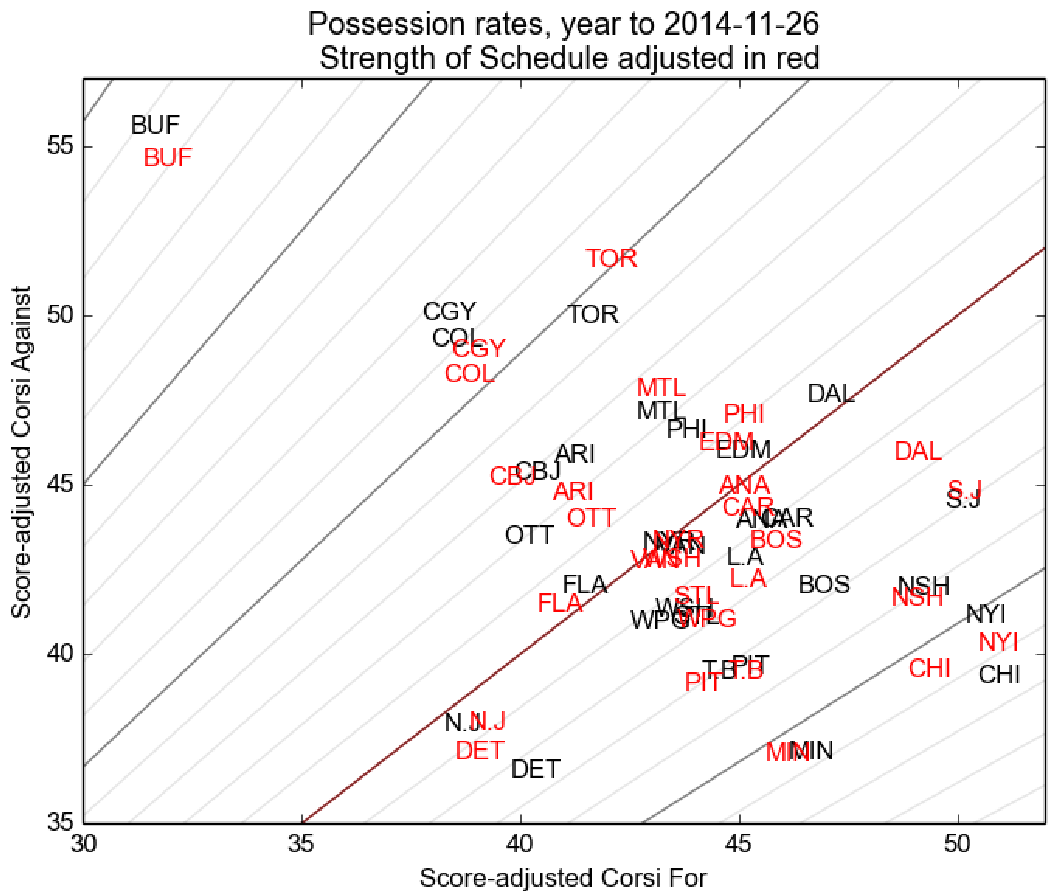
Once we have the vector F, describing offensive ability (relative to the sample at hand), and the vector A, describing defensive ability, it is simple to compute the schedule-adjusted statistic–simply multiply each element of F by the average statistic m to obtain the offensive schedule-adjusted statistic, and multiply each element of A by m to obtain the defensive schedule-adjusted statistic.

In our example, this produces vectors:

$$F = \begin{bmatrix} 45.2 & 41.6 & 46.6 & 32.7 & 39.8 & 46.1 & 49.5 & 39.2 & 40.8 & 48.1 & 38.8 & 46.6 & 40.0 & 44.5 & 45.1 & 44.1 & 48.2 & 38.6 & 49.4 & 43.5 & 41.9 & 44.7 & 42.8 & 49.6 & 43.5 & 44.2 & 42.9 & 43.8 & 44.0 & 43.5 \end{bmatrix}$$

$$A = \begin{bmatrix} 45.0 & 44.5 & 42.8 & 53.1 & 48.1 & 43.8 & 39.5 & 47.5 & 43.9 & 46.7 & 37.2 & 44.4 & 42.2 & 42.5 & 38.1 & 46.8 & 42.4 & 38.7 & 41.5 & 43.3 & 43.8 & 47.1 & 40.3 & 45.7 & 42.2 & 40.0 & 50.5 & 42.2 & 42.7 & 41.8 \end{bmatrix}$$

Those cursed souls who follow me on twitter will already have seen the results of these schedule-adjustments, since I tweet plots like the below every day. This example shows 5v5 score-and-venue adjusted corsi, in black, and schedule-adjusted score-and-venue adjusted corsi, in red. Buffalo is so dire (by both measures) that I have omitted them from the chart to permit the schedule-adjustments to be seen.



Most teams are more or less unchanged by the adjustment, which makes sense — most teams have played a representative set of opponents and put up ordinary numbers. However, several notable things jump out all the same. Dallas, for instance, has a much better offense than appears from their raw counts, since they've been putting up decent numbers against, on average, very stingy teams. Their defensive numbers also improve somewhat. Toronto has been putting up offense as expected but concede more events than they should considering their weak schedule so far. Boston is substantially inflated by weak opposition, Winnipeg the opposite.

Of course, early in the season, the matrix S will have a lot of zeros in it, since many teams will not have played one another. Even now, around a quarter of the way through the season, many of the entries are zero. When "too many" of the entries of S are zero (where the technical sense of "too many" is beyond the scope of this article), the process I describe above will not settle gently on one value, but instead behave chaotically. In this case, there is no schedule-adjustment. To make adjustments, one has to guage the strength of one's opponents, and to do so requires using *their* opponents, which are evaluated using their opponents, and so on, and if there are insufficiently many games played, then there is no information with which to perform these evaluations.

Another infuriating consequence of the lockout, in case we were somehow short of such, is that the shortened schedule contained no regular-season inter-conference play. This means that it is only possible to perform schedule-adjustments within conference. How vexing.

As a mathematical addendum: The process of forming (F_{n+1}, A_{n+1}) from (F_n, A_n) is an endomorphism of $R^{30} \times R^{30}$, which I call S^*. If we adopt the convention that F_0 and A_0 are vectors whose every entry is one, we can write:

$$\begin{aligned} (F_{n+1})_i &= \frac{1}{m} \cdot \frac{\sum_{j=1}^{30} S_{ij}/(A_n)_j}{\sum_{j=1}^{30} N_{ij}} \\ (A_{n+1})_j &= \frac{1}{m} \cdot \frac{\sum_{i=1}^{30} S_{ij}/(F_n)_i}{\sum_{i=1}^{30} N_{ij}} \end{aligned}$$

Performing the adjustment amounts to computing a fixed point of S^*. As mentioned above, when S is sparse, S^* may not have a fixed point in the component of $R^{30} \times R^{30}$ containing 1×1; ideally we should be able to characterize the fixed points of S^* using the properties of S. The map S^* strongly resembles a Markov process, for which fixed points are well-understood, but I cannot seem to rewrite it as such. The main obstacle is the inversion of the entries of A_n and F_n at every step, which does not play nicely with matrix multiplication. Brouwer's fixed point theorem does not appear to apply. Alas.