

Hockey's Pythagorean Theorem

Contributed by BringBackZezel
Monday, 20 August 2007
Last Updated Monday, 20 August 2007

It's almost that magical time of year again....when training camps open and everyone comes up with their predictions and line combos.

It's also the time of year when most of the predictions and line combos are based on EA Sports NHL series player ratings and kool-aid fantasies. To those who this describes I'd like to say the following:

Yes, we know that every team is going to win the cup this year, and that career 8 goal per season guy is going to finally break out and bury 40 goals this year. Furthermore, we all understand that if we analyze your premonition, then we are insulting your manhood and as such it's a reasonable next step to question our general intelligence and threaten bodily harm. We get it. You know everything about hockey.

Now that we've got that out of the way, I'd like to take a statistical look the upcoming season. I'd like to take a step back from figuring out who's going to center Paul Kariya to look at how the team will perform as a whole.

As I have discussed before, teams typically don't drastically change from year to year. Teams that allow 300 goals don't cut it back to sub-200 the next year, and teams that score 200 goals bump it up to 300.

With that reality in mind, the goal of this article is to look at the statistics of the NHL and how they are good for predicting future success.

The Pythagorean Theorem

What prompted this article was the "sabrmetric" study of baseball and Bill James' "Pythagorean Theorem". For those of you unaware of the James' theorem, it basically states that there is a direct correlation between runs scored, runs allowed, and winning %. It works remarkably well in baseball, as they play a 162 game season and as anyone who has ever studied statistics will tell you, the larger the sampling the greater the accuracy of the study.

I've always wondered how it would apply to the NHL, so I started crunching numbers...

First, I took only the last two seasons into account as in the time since the lockout, the game has changed (I'll explain this in a bit). James' Theorem states that the expected winning percentage = $RS^2/(RS^2+RA^2)$ (Note: RS = Runs Scored, RA = Runs Allowed). Over the course of a season, this will usually calculate how many games a team will win within 5 or so games. In fact, usually over 2/3rds of MLB teams fall within one standard deviation with this equation.

So how does this apply to the NHL? I took the GF and GA for the last two years in the NHL and used them in James' equation. Because that gives you an estimated winning percentage, I then multiplied that percentage times 82 games to give me an estimated number of wins for the season. Of course in the NHL wins only matter if you're tied in points, so I then multiplied the number of wins by 2 for the number of points per win....and I found that except for the 05-06 Minnesota Wild, ALL teams ended with more points than what was estimated. In most cases teams were 10 or so points behind their real results.

How could this be? Of course my first thought was perhaps the equation just doesn't apply to the NHL, or perhaps the sampling size of only 82 games per season wasn't enough. Then I came to a realization: the numbers are skewed not

because it's faulty logic, but rather because I made the mistake multiplying my estimated winning percentages by the number of games and points per win. In reality, due to overtime losses in the NHL, there are more than 164 points available for every team throughout the season. Technically, there are 3 points per night available, so a single team could actually participate in as many as 246 points being won in a single season (with of course their maximum winning capped at 164).

In reality, the estimated winning percentage should be multiplied by the number of games played (82) and then by the average number of points won per game. To get that number, I took the total number of points won in the NHL last season (2741) and divided by the number of games played (1230). I came up with 2.228 points per game being won (meaning nearly one out of every 4 regular season games went to OT). This is the reason why I only used 2 seasons worth of data in this study. Is that if we go back further when teams could be held to a tie then there were less points won per season and as such it skews the stats away from the current 'shoot-out, winner every night' NHL (But I should add that it also actually makes the equation MORE accurate and not less).

When I use this points per game number I found something very interesting. Not only did the numbers line up more accurately with the actual team point finishes, but they're actually MORE accurate than James' original MLB equation. In looking at the last two complete MLB seasons (2005, 2006), 19 teams were more than 1 standard deviation (σ) away from the average (meaning statistically speaking the result is almost a perfect normal distribution of values). In the last 2 NHL seasons (05-06, 06-07), only 15 teams finished outside one σ ;...and only 2 teams finished outside 2 σ ; in two seasons combined. Further validating this calculation is the fact that the standard deviation in the NHL is less than the wins deviation in MLB. For baseball, 1 σ is 4.3 games. For the NHL, 1 σ is 5.16 points (or 2.31 games using the 2.228 ppg average).

Basically rather than guessing at how many points a team will have next season, if you analyze how many goals they'll score and how many they'll allow you can roughly calculate how many points they'll have within about 5 points.

As an aside, The NHL had the exact same number of OTL in 05-06 as 06-07, meaning the total number of points and the average points per game were identical (2741 and 2.28 respectively). This will likely not happen again this upcoming season, but it does mean that the 2.28 estimate of points per game will likely be VERY close to the actual number for the end of the season.

I've also found that by lowering the exponent on the original equation to somewhere in the 1.855 range, it yields a slightly more accurate estimate (lowering the σ to around 5.02), but given the relatively small sampling size of only two seasons and the fact that it didn't change the number of teams that fell outside the σ , I felt that it simply wasn't enough to alter the equation at this time. Perhaps in another couple seasons we'll come back to this.

GOALS FOR 2005-2006 VS. 2006-2007

In looking at the 05-06 season in comparison to the 06-07 season, I found that overall goals were down 342, or an average of 11.4 per team. Edmonton had the greatest decline in goal scoring with 61 less than the season before. Calgary had the greatest increase with 40 more goals than 05-06. There were 18 teams that scored less, and only 12 that scored more. Keeping in mind that overall goal scoring was down more than 11 goals per team on average and adjusting for that difference, we find that in comparison to the league average Calgary was up 51 goals and Edmonton was down 50.

GOALS FOR 2005-2006 VS. 2006-2007

Of course because goal scoring was down 342 goals, that means that there were that many less goals scored. Pittsburgh led the league in GA decline by dropping the GA a whopping 70 goals (more on this later). Meanwhile, unsurprisingly, Philadelphia allowed 44 MORE goals in a season when goal scoring declined. Adjusted for goal deflation, those numbers come out to the Penguins at +59 and the cross-state Flyers at -55.

GF-GA DIFFERENTIAL FROM 2005-2006 TO 2006-2007

Another interesting stat to look at (especially considering the relation of GF and GA to ultimate standings performance) is the differential of GF to GA. In 05-06, Ottawa was a +103 in that department, while the lowly Blues were a -95. In 06-07, Ottawa and Buffalo tied at +66, while Philadelphia pulled up the rear at -89. Since the goal of this is to use trending data to predict future performance, the more important numbers are how the goal differentials in 06-07 compare to 05-06.

It will be no surprise to anyone paying attention that the team that showed the most improvement last season versus the season before is Pittsburgh, but the level of their improvement is historic. They added 33 goals to their season total (in a season when goal scoring was down over 11 per team) and cut their goals allowed to 70 goals less than the previous season. That gives the Pens a total differential improvement of 103 goals. The second place team was the Islanders at +56 with a more traditional improvement of 18 more goals scored, 38 less goals allowed. The Blues were 3rd at +55 using a similar formula. From there, we see a drop off to San Jose in 4th with +35 and a number of teams in the mid-20s. The standard deviation for this differential was 39, meaning compared to the statistical norms, the Penguins had about a 0.5-2% chance of doing exactly what they did. In fact, when you compare one season to the next, only 4 teams have had a positive differential of more than 66 goals...the 05/06 Rangers (86), 05/06 Hurricanes (71) and 02/03 Stars (74). Pittsburgh blew them all away at +103...and did so without a major rule change.

WHAT DOES THIS MEAN?

Well, it means a couple of things...

- I have too much time to analyze numbers and a distaste for random guesses at performance.
- It's likely that the best case for the Blues is another improvement of around +55, but based on averages, more likely they'll be around +25 to +35 with the amounts being equally split.

When we plug those numbers into the equation this is what we get:

+55

242 GF

227 GA

97 Points

+30

229GF

239GA

87 Points

Here's what it comes down to: The Blues need to improve by 55 goals (a combination of more GF and less GA) to make the playoffs next season. If you don't agree with me, then you're an idiot and I'll beat you up.