

- [All 308 blogs](#)



Follow



8+1

1.9M

- [Search](#)

Search



- [NHL](#)
- [Outnumbered](#)

Measuring repeatability in hockey stats

By [Eric T.](#)  [@BSH_EricT](#) on Nov 20 2013, 2:53p [5](#)



Just because a guy shoots 18.9% for a year doesn't mean you should expect him to do it forever. - Vincent Pugliese

It's crucial for many analyses to know how repeatable a performance is. Let's walk through why that's trickier than it sounds.

[Tweet Share on Twitter \(13\)](#) [Share Share on Facebook](#) [Share Share on Google Plus 5 Comments](#) [*Rec Recommend this Post](#)

"You aren't as good as you think when you win, or as bad as you think when you lose."

It's an old sports cliché, and it's simple to understand: We all have good days and bad days, and our performance on any one given day -- whether that's at the the office or on the ice -- isn't necessarily a great indicator of our future performance.

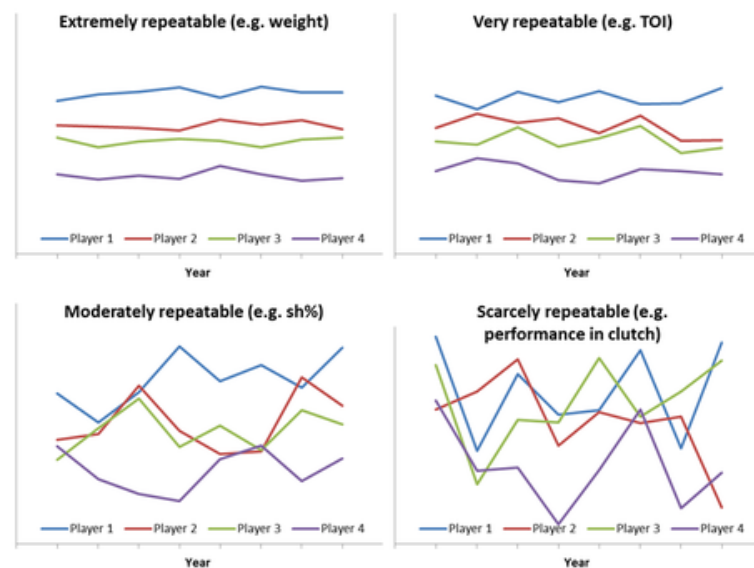
That's where statistical analysis can help. Some performance measures are much more repeatable than others, and measuring exactly how much variability a statistic has can be important for a few reasons.

For example, when we make projections, it's [possible](#) -- and [often critical](#) -- to account for variance by pulling our estimate in toward the mean; the amount we adjust by is determined by this assessment of how repeatable the statistic is. Similar adjustments also show up in our estimates of how performance changes with age, [as we saw in our last piece](#).

So let's examine that repeatability measure more closely.

Repeatability and survivorship

Measuring a stat's repeatability seems pretty straightforward -- all we have to do is look at how highly correlated players' performances are from one year to the next. If the metric is highly repeatable, players will put up the same numbers year after year; if it's heavily subject to variance then we'll see their numbers bounce all over the place.

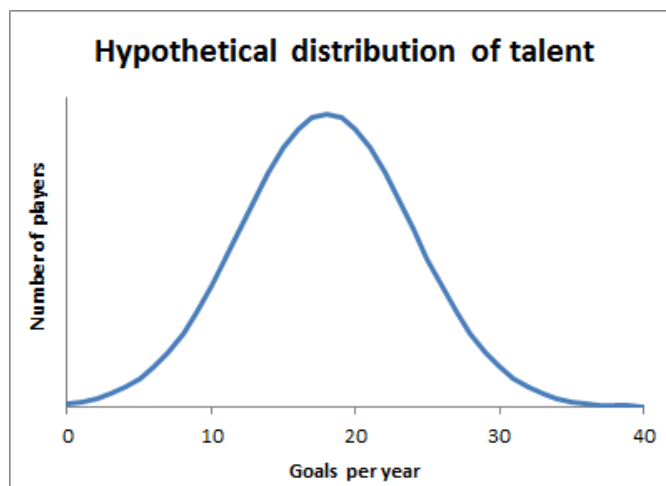


But when we're looking at how correlated performances are in year 1 to performances in year 2, what happens to people who didn't play both years? Obviously they can't factor

into this analysis.

That wouldn't be a problem if they were roughly the same as everyone else, but they aren't -- good players are much more likely to play again next year than bad ones are. The result can be a significant impact on our assessment of repeatability. The simplest way to explain is with a thought experiment.

Let's suppose the measure of talent we're interested in is goal-scoring, and that simple variance might lead any given player's goal total to fluctuate by a couple of goals from year to year. If we're looking at everyone in the NHL, the talent distribution might look like this:



In our thought experiment, people bounce up and down by a few goals per year, but that's small compared to the range of talents we're looking at, so the data looks pretty repeatable -- a good player will outperform an average player year after year :



But what if we're not looking at everyone in the league? Let's imagine we focus in on just the top half of the league, for whatever reason.

Random chance still makes people bounce up and down by a few goals per year, but now a few goals is the difference between good and average. So in this smaller data set, an average player can randomly bounce from the top to the bottom from one year to the next:



The result is that goal scoring looks less repeatable. We're no longer quite as certain where to rank a player with just one year's data.

That's how survivorship issues can affect our measure of repeatability. Cutting half the league in this example is obviously extreme, but it's definitely true that a lot of bottom-end performers lose their jobs. That restricts our sample for year-over-year repeatability to a narrower range of talent, which makes it look like variance plays a larger role than it really does.

My goal here is to work out how significant that effect could be.

Setting up a simulation

The challenge with inquiries like this is that we have incomplete knowledge. If a player does poorly one year and loses his job, we'll never know what he would have done the next year. It's impossible to know whether he was a bad player or a decent player who just had a bad year.

For this kind of situation, simulations can be a powerful tool, since we can know things about our simplayers that we would never know about real ones. So I set up a simulation to study this survivorship issue; here's how it worked:

- 1) Each of my 100,000 simplayers has a randomly assigned score for his talent.** Talent follows a normal distribution with a symmetric bell curve (like the one we used for goal scoring above), and doesn't change from one year to the next.
- 2) Each player also has a randomly assigned score for his luck in simyear 1.** Luck is also normally distributed in this simulation. Moreover, we'll set the bell curve for luck to be half as wide as the one for talent. That means that we're simulating a measure that is driven more by talent than luck -- it's almost impossible for a player with high-end talent to get so unlucky that he performs like a bad player for a year.
- 3) Each player has a randomly assigned score for his luck in simyear 2.** This value is assigned exactly the same way as for his luck in year 1, but it's chosen completely independently -- running hot in year 1 doesn't make a player any more or less likely to be lucky in year 2.
- 4) A player's observed performance in any given year is the sum of his luck score for that year and his talent score.** With real players, we only know this observed performance level, but in our simulation we know exactly how much skill went into that performance.

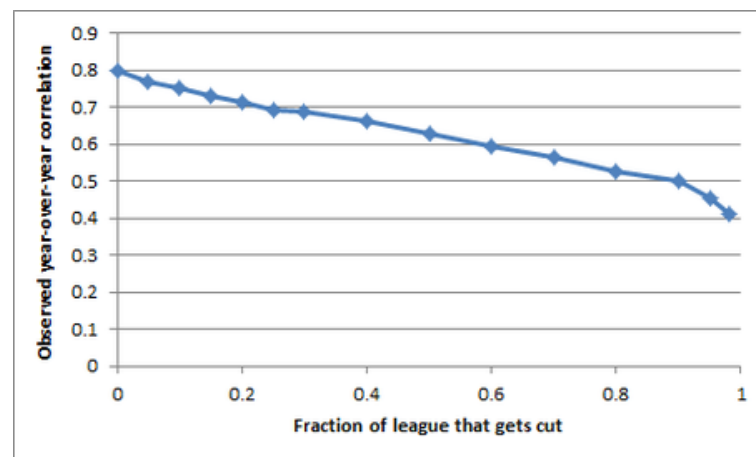
The results

Over my 100,000 simulated players, the correlation between observed performance in year 1 and observed performance in year 2 was 0.8006. Arithmetically, we would expect 0.8, so it seems like the simulation works the way it should.

Now we can explore the impact of survivorship bias. Suppose the NHL has 20% turnover, with the worst performers getting replaced in any given year. In the NHL, we can only

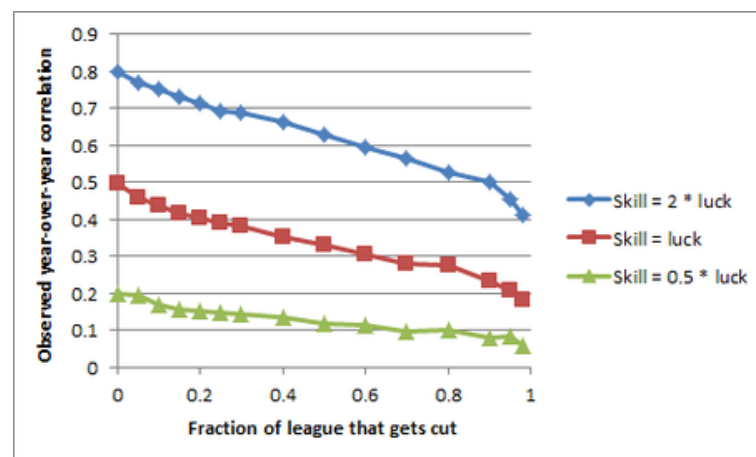
calculate repeatability for the 80% who kept their jobs, but in our simulation we know what the other 20% would have done if they'd played another year. That lets us see how survivorship issues are affecting our assessment of repeatability.

Here's how the repeatability correlation we observe in practice changes depending on how much turnover there is:



The simulation shows us that in a league where the bottom 20% get cut, we'd observe a repeatability of 0.71 instead of 0.80. Survivorship bias plays a modest but significant impact on our projections -- by compressing the talent pool to just the top 80% of the players, we make luck appear to be a larger fraction of observed performance than it is for the league as a whole.

So what can we do about that? Well, we can look at how much turnover there is in the NHL and use a curve like this to estimate the true repeatability score. We can make charts like the one above for a variety of different repeatability scenarios:



So if we observe that the league has 20% turnover and a year-over-year correlation of 0.40 for some stat, then that puts us on the red curve and we could see that the repeatability if there were no cuts would be 0.50.

Unfortunately, there's another complication to consider.

These curves assume that the cuts were made exactly along performance lines. But that's not generally going to be true; there are team depth charts to consider, and more importantly it's going to depend on what we're measuring. The players who get cut are reasonably likely to be near the bottom of the list in a metric like ice time or points, but won't necessarily be near the bottom in, say, hits or shorthanded assists.

This affects how much we're compressing the league talent pool and impacting the observed repeatability. If players who have great numbers in whatever metric we're studying are just as likely to get cut as players who have terrible numbers, then the cuts won't impact our measured repeatability at all.

I've put this into my simulation, but it's complex enough that I can't easily publish a table for everyone to use. Fortunately, there's an alternative solution.

Skip the sim

Using a simulation to work out exactly how much luck there is might be analytically best in some cases, but it's not the only answer. If we know that a player is going to get a second year, we can use that information in our estimate of what his talent is likely to be.

Instead of pulling his numbers towards the overall league average, we can pull them towards the average of players who stick around for a second year. Then we don't have to worry about what the players who got cut would have done.

Let's work through an example. I once looked at [aging curves for players' shot rates](#) without directly assessing repeatability; let's look at how I'd calculate the repeatability as part of updating that study.

The simple correlation for year-over-year shot rate on players used in that study is 0.74, but we now know that's not quite right. There's definitely some survivorship bias in play -- the players who dropped out of the study were appreciably lower than the ones who stuck around (7.17 SOG/60 vs 7.73). So here's what we can do:

1. Use a simulator. Input how likely a player is to survive to year 2 and let the simulator run a variety of true underlying correlations until it finds one where our observed result would be 0.74. The answer we get is 0.79, so we regress our talent estimates 21% of the way towards the average for all players.
2. Skip the simulator: Use the observed 0.74 repeatability and regress our talent estimates 26% of the way towards the average second-year performance of the players who played two years in a row.

In effect, we're choosing whether to alter the mean we pull guys towards or to how far we pull them. But in either case, the effective result is a slightly higher estimate of talent for the guys who play both years than we would have gotten if we'd ignored survivorship.

Summary

Survivorship bias does have an effect on our measured repeatability, and it can be significant if we're looking at a talent that's thought to be important, that's heavily tied to whether a player keeps his job.

Our simulation gives us a way to correct for this, or we can just remember that survivorship has changed our talent pool a little and so we aren't pulling people towards the overall NHL mean, but towards the mean for players who don't lose their jobs.

[Tweet Share on Twitter \(13\)](#) [Share Share on Facebook](#) [Share Share on Google Plus](#) [5 Comments](#) [*Rec Recommend this Post](#)

Latest News