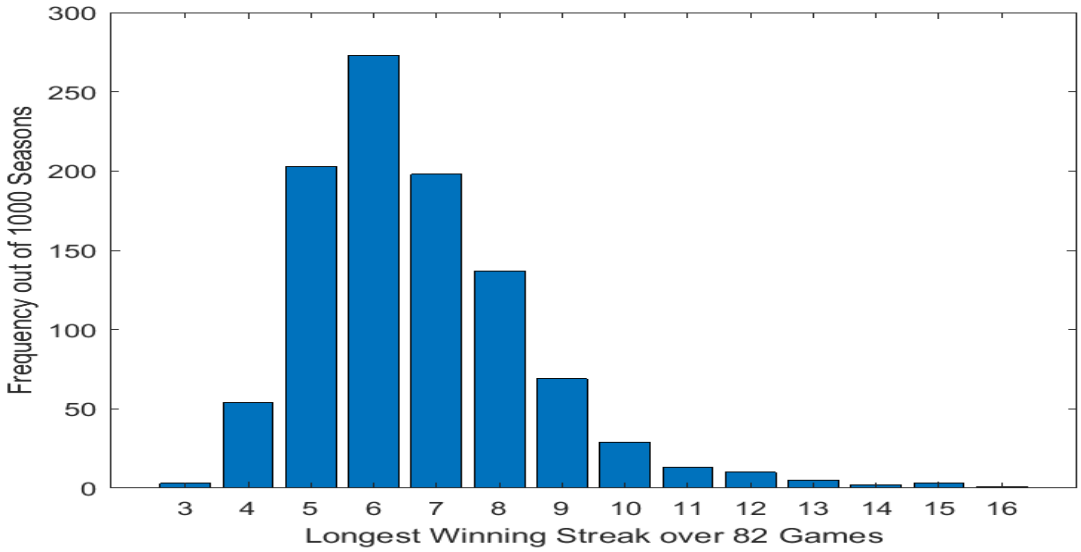
Can Teams Actually be Streaky?

Every sports fan understands that the outcome of a particular game has a degree of randomness to it. If two teams are equally matched and play each other over and over, then there is a pretty good chance that one team will win four, five, or more times in a row just due to random chance. However, if your favorite team is middle-of-the-pack and wins six in a row, you probably say they are on a hot streak!

In fact, suppose my favorite team is .500 and wins half their games and loses half randomly. If I count their longest win streak during a season, the probability they have a streak of at least 5 games is something like **95%.** Maybe even crazier, the probability they win 8 in a row at some point is almost **30%.** I simulated 1000 seasons of 82 games and counted the longest winning or losing streak for a .500 team and plotted the frequency here:



If a .500 team (the Grizzlies?) rattled off 8 straight wins halfway through their season, the media would have a field day, Ja Morant would be called `the unquestioned ROTY' and the talking heads would be talking about the Lakers scary first round matchup against the 8 seed Grizzlies.

There is no doubt that streaks actually exist; teams regularly win 4 or 5 in a row even if they aren't top tier title favorites. We want to determine whether or not streaks are more than

random; we want to know whether teams really get "hot". So we'll consider two alternatives. The first possibility is that streaks are purely random occurrences. The second case is that teams actually go through peaks and valleys in how they perform. Knowing that streaks actually happen, how can we determine which of these two options is the truth?

Let's start with the first case. If teams' streaks are purely attributable to random chance, then knowing a team is on a hot streak doesn't change whether we think they'll win their next game. It's like this: if you flip a fair coin and get 4 straight heads, the next flip isn't more or less likely to be heads or tails. Sometimes 4 straight heads just happens. So, if we use streak information to predict winners we shouldn't be any better at picking the winner than if we didn't know they were on a streak.

What is the alternative? Well, that streaks actually exist. In this second case, if we can identify that a team is on a hot streak, then we probably think they are more likely to win their next game than we otherwise would.

So this is exactly how we'll try to answer the posed question. If we try to identify streaks, are we more accurate in picking the winner of any given game? The short answer is probably shocking: no. Streaks as we typically understand them in sports are a myth.

The long answer? I'm going to give the same answer your 10^{th} grade algebra or freshman physics teacher would give you: it's complicated. It depends on what you consider an improvement in predictive ability.

The Method

We use our Ensemble Ratings (hyperlink) in two different forms to predict winners. Remember that Ensemble Ratings use game data to assign each team a rating so that the difference between two team's ratings predicts the margin of victory. The ratings are determined by finding the numbers that best explain the margins of victory we have already seen (the games that have already been played!).

We can similarly define what we call Ensemble Recent Ratings that are allowed to place more emphasis on games that happened recently. For example, we could choose to weight each of the most recent 100 NBA games and give them a weight of 2. What this means is that if Toronto

beat Miami by 10 points and it was one of the 100 most recent NBA games, then our ratings pretend that Toronto and Miami played twice with Toronto winning both by 10.

We will look at the prediction accuracy of Ensemble Recent Ratings versus the unweighted Ensemble Ratings and use this to answer our question.

Why Would this Work?

Suppose we weight the 100 most recent games by a factor of 7. Since there have been roughly 800 games played so far this season, this gives equal weight to the first 700 results and the most recent 100 results. The most recent 100 games corresponds to about 6-8 games per team. If a team is on a hot streak (winning 6 in a row, winning 7 of the last 8, etc.) then their Ensemble recent rating will essentially have them as 'undefeated' for half the season. That is, their Ensemble Recent Rating should be much higher than their Ensemble Rating.

A bit more concretely, suppose my favorite team is 36-28 after having won 8 straight. Then, my Ensemble Rating will reflect their true .562 winning percentage. This team could be expected to be a fair competitor with, say, the 2020 Philadelphia 76ers, for instance. However, if the last 8 games are weighted by a factor of 3, then my Ensemble Recent Rating will reflect a 52-28 record instead of 36-28. This team's rating would be much closer to that of the Utah Jazz than Philadelphia.

By weighting recent games more heavily, we can distinguish which teams are on streaks and predict them to have a higher probability of winning.

What do the Numbers Say?

If you look back to our post introducing Ensemble Ratings (hyperlink), you'll notice that weighting the previous 100 games by a factor of 5 actually gave us better predictive accuracy than regular Ensemble Ratings. In fact, the following table shows the accuracy in picking the winner using different recency weighting schemes (remember as of writing this there have been 803 NBA games in 2020 and baseline Ensemble Ratings are 65.4% accurate this year).

	Most Recent 50	Most Recent 100	Most Recent 200
Weight by 2	66.3%	66.3%	65.8%
Weight by 5	66.3%	66.3%	65.2%
Weight by 10	66.4%	65.9%	64.9%

But wait a minute. You said that using streaks didn't actually increase predictive ability?

That's true and I argue that the above table supports what I said.

Really, I said it's complicated. If we look at the differences between adjusted and unadjusted ratings we see that New Orleans has the largest increase while Milwaukee has the largest decrease. In the New Orleans case, the extra weight being added to the recent games likely picked up on the team's improved performance as a result of the return of Zion Williamson. This is not attributable to a streak of any kind.

Milwaukee, on the other hand, saw a massive decrease in Ensemble Recent Ratings as a result of losing 4 of 5 before the quarantine stoppage. However, after digging into the data, none of Milwaukee's win/loss projections have changed due to this extra weighting. That is, Milwaukee was already so dominant that they were still predicted to win each of their last 5 games even when taking their losing streak into account.

In short, I actually attribute the improved prediction accuracy of recency adjusted rankings to picking up on team changes quicker. What does this mean for our original question? Betting on the hot team or cold team to continue their streak gives you less than 1% improved accuracy in choosing the winner. And, after taking into account other `recency' effects such as players getting injured/coming back from injury, there is likely virtually no benefit in using streaks to help in picking your winners/losers.