Momentum and Winning Streaks

Brian Connolly

www.ideajungle.com

BrianConnolly at-sign ideajungle dotcom

I'm interested in whether the "spirit" of a team has any effect on its play. In my first analysis of this, I looked at dramatic games. These are games where a home team comes from behind to win in the bottom of the ninth. It seems clear that if team mood, self confidence and belief have any effect on subsequent play, we'd expect teams to do better immediately after these wins than they do before. That study demonstrated that there isn't any evidence of this in past history. Teams don't seem to do any better after a dramatic win than they do before.

In this analysis, I evaluate a different aspect of this. If mood affects performance, we'd expect teams that begin winning to have a tendency to continue winning. Similarly, teams that start losing and develop a poor attitude will continue to do so. If the sole determinant of a teams ability is its annual winning percentage, P, than the expectation of a win in any particular game would be P. But if "momentum" matters, than if we pick an arbitrary game after a win , we'd expect the probability of a win to be greater than P. Similarly, in games that follow a loss, we'd expect the probability of winning to be less than 1-P.

This leads to the following principle: if "momentum" matters, we'd expect streaks to be longer than they would by chance. A team would, in general, be more "streaky" than a model of the team that was based on purely on a series of random wins and losses with the same overall win percentage.

Approach

Imagine using a coin to model a teams performance in a given year. If the team's overall winning percentage w/(w+l) was .600, we could bias the coin so that it lands heads up 60% of the time. If we throw the dice 30 times, we might come up with a pattern like:

WLWWLWWWWLLWWLWLLWLLWWWWLWLWLW

One of the criteria statisticians use to evaluate random phenomena is a test for "serial independence." As above, they want the probability of each event in a series to be the same regardless of the outcome of the prior event. One test of whether these coin tosses are serially independent is what's known as a "runs test." A runs test can be done when any series of experiments has only two possible outcomes. Let n_1 be the number of outcomes of type one, and let n_2 be the number of outcomes of type two. In our case we consider type one as a win and type two as a loss. A "run" is defined as a series of consecutive wins or losses, allowing for runs of length 1. So in the above example, n_1 is 18, n_2 is 12, and the number of runs is 19. The runs test uses the fact that when that when n_1 and n_2 are both > 10, then the expected number of runs,

$E[\rho] =1 + 2n_1n_2/(n_1 + n_2)$.

is normally distributed, with variance:

$\sigma^2 = 2n_1n_2(2n_1n_2-n_1- n_2)/[(n_1 + n_2)^2 (n_1 + n_2 -1)]$

In our example, the number of runs we expect in our coin toss is ρ=15.4, and it's variance σ^2 is 6.65. The actual value that we observe is 19. Since R is normally distributed, and since 19 is only 1.4 standard deviations from R, there is no reason to reject the hypothesis that the above coin is random.

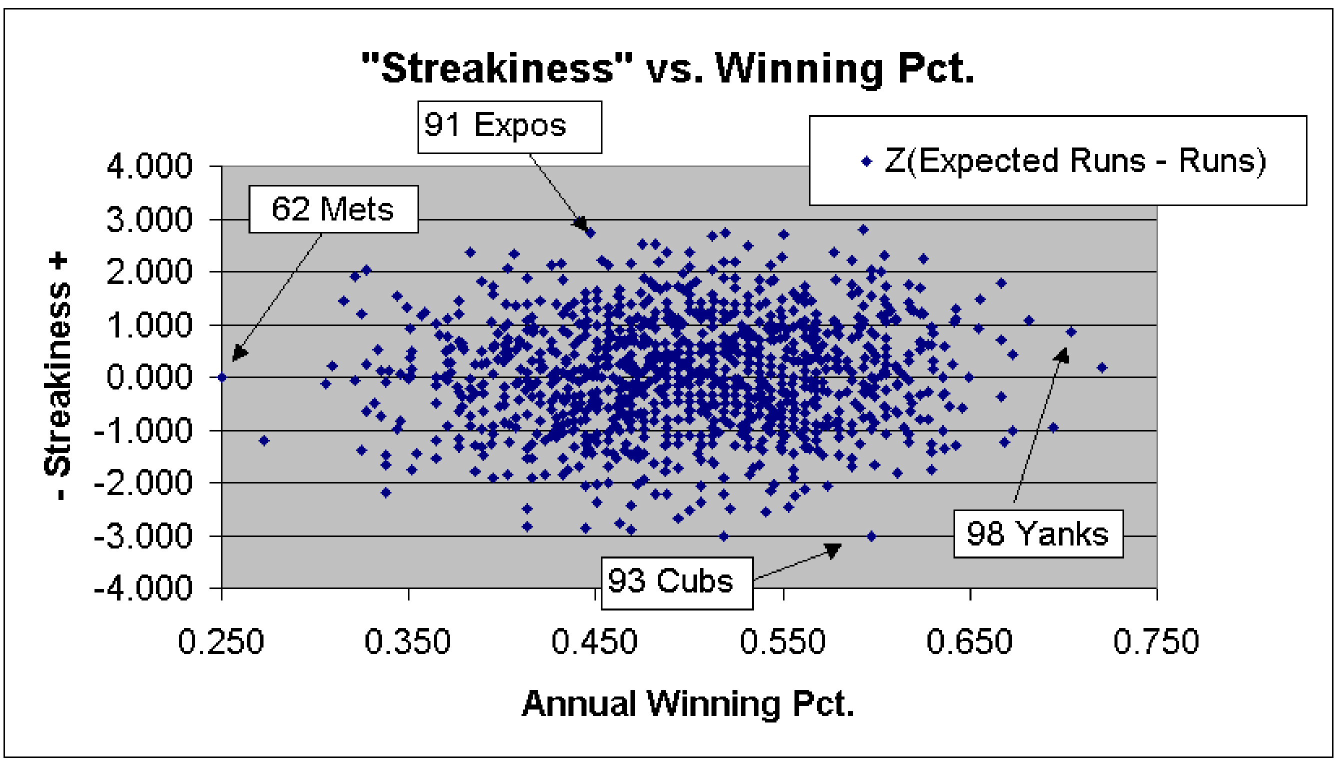
Methodology

The study was based on Retrosheet game logs from 1950-2000. The number of wins, losses, and won-loss runs was computed for each team. The concept was to treat each season as a "coin toss" experiment, by comparing the actual number of runs with the theoretical expected values from the formulas above. If team are indeed streaky, if prior results bias future results we would expect to see some significant deviation from the theoretical ranges.

There is one inaccuracy worth noting. There have been 262 tie games in the fifty years. Since a change in result is considered a run in my method, a loss-tie-loss or a win-tie-win will always count as two runs. This means that the total number of runs across all seasons and all teams might be overstated by at most 524 (one extra run per year for each team participating in the tie). This overstates the 91238 runs across all teams by at most 0.68%

Results

The following graph is a scatter plot of all 1173 seasons. The annual winning percentage of each team is plotted against Z, where Z is the number of standard deviations difference between the expected number of runs and the actual number.



The teams in the upper half have positive Z, meaning they have fewer runs, and are thus more "streaky" than expected. The teams in the lower half have more runs than expected, and are less streaky. In other words, these less streaky teams have more, and hence shorter runs of consecutive wins and losses.

Conclusions

Based on this evidence, there is no reason to believe that teams have hot and cold streaks *as a function of mood.* The plus and minus 3 lines are seasons where a team had actual runs more than 3 standard deviations different than the theoretical expected value. Only a few teams approached these values. Since the expected runs for each team in a year is is normally distributed, we'd expect to get an actual value within 3 standard deviations more than 99% of the time. It's not surprising that there were a handful of values near these limits out of 1173 samples.

It is important to qualify the meaning of this apparently negative result. It's clear that good teams will string together more wins, and bad teams will string together more losses. The theoretical formula accounts for this. A team with a .250 winning percentage is expected to have 61 runs, many of which will be long series of losses. Similarly, a team with a .750 winning percentage will also have 61 runs, but theirs will have a similar bias towards winning streaks. What this result means is this: There is no reason to believe that teams string together more wins or more losses than the theoretical, random model would predict.

The negative result is surprising. To return to the earlier discussion. To say that the probability of a win on any particular day in a season is the same as the season's winning percentage is suspect, just for scheduling reasons. On any given day, a team is more likely to play yesterday's team than it is to play the other teams in the league. So a win today should be more likely when it follows after a win yesterday. The runs test is just one of a number of tests of randomness. Perhaps other tests will show that results are not serially independent.

-- May, 2002