Future Expectations for Over-Performing Teams

1. Premise/Background

In late September, 2007, an issue arose on the SABR Statistical Analysis bulletin board premised on the substantial over-performance of the 2007 Arizona Diamondbacks in wins vs. Pythagorean expected wins. Somebody (on or off the list, I'm not sure which) suggested that this over-performance was a fluke, and Mike Emeigh argued that the Diamondbacks over-performance might *not* be a fluke, but that "IMO, the Diamondbacks' actual performance this year is more reflective of their team quality going forward than their Pythagorean performance."

This led to an active discussion which, over the course of the next two weeks, splintered into a hundred different issues as such discussions are inclined to do. On October 13 I posted the following, as a part of one of those splinter discussions:

> Well, the reason you can say that pythagorean under-performance is a fluke is that it has a persistence of near zero. Teams that over-perform in one season have basically no tendency to over-perform again the next.

To which Mike Emeigh responded:

> That's true—but that's been studied to death. The question I posed is different:
>
> Will the Diamondbacks ACTUAL record NEXT year be closer to their actual record this year, or their Pythagorean record this year? IOW, are they more likely to win 90 in 2008, or 79?

To which I then responded:

> I understand that you think there's an important difference here. I don't believe there is.
>
> Take two teams, one of which scores 750 runs and allowed 800 runs and wins 76 games, as we would expect, and one of which scores 750 runs and allows 800 but wins 76 games (*my typo. I meant to say 85 games, rather than the second 76.)* Numerous studies have shown that there is NO difference in there (sic. ..my typo again) expected performance the next season. .at least this is my understanding; I haven't studied the issue myself since the 1980s, and may be behind the curve. . ..somebody may have found that there is some small difference in expected wins the next year. . . .it doesn't matter whether you are saying that they will not relapse next year because
>
> a) they will continue to outperform pythagorean expectations, or
> b) they will improve their runs scored/runs allowed ratio.
>
> because in EITHER case, you are still asserting that they have a higher expected wins next year than another team with the same runs/opposition runs ratio. Which, as I say, you could well be right, but I'm skeptical.

I then decided that if I was going to speak publicly about this issue I should know a little more what I was talking about, so I decided to spend a few hours studying the issue.

What I am doing in this study is NOT intended to investigate Mike Emeigh's thesis that some over-performing teams are actually much better than their runs/opposition runs ratio would suggest. To investigate that thesis would be a complicated task, and I suspect it would be impossible to refute the hypothesis given the limitations of real-world data. I was not checking out *his* statement; rather, I was double-checking my *own* statement that "Numerous studies have shown that there is NO difference in their expected performance the next season." I was wondering, "Is that really true, that there is no difference in their expected performance next season? Or, in saying that, was I relying on my hazy recollections of studies that I did a long time ago which might not meet the standards of modern sabermetrics anyway? Do I really *know* that there is no difference in their expected performance the next season?

And, to give away the main conclusion of the study. . .I was wrong; there IS a difference. But let me proceed in an orderly fashion.

II. General Method

I started with a spreadsheet containing the runs scored, runs allowed, wins, losses and games played of all teams in major league history through 2007 I will also make this spreadsheet, "Pythagoras through 2007" available to any researchers by posting it along with this report.

Worksheet 1 of that spreadsheet is a simple list of the teams and data. It also includes Innings pitched and I3 (thirds of an inning), which I saved just on the off chance that it might become useful, although it never did.

I then figured the expected winning percentage for each team, using the variation of Pythagorean wins with the exponent of

{(Runs + Opposition Runs) / Games } ^ 0.286

Which creates a higher exponent for teams which score and allow more runs per game, and has been shown to be more accurate that the original Pythagorean formula.

I then identified the 100 most over-achieving and 100 most under-achieving teams of all time in wins vs. expected wins. Similar things were also done by other people posting to the discussion group, and their list may be as good as mine, but my list of the top over-achieving teams was:

Team and Year	Expected Wins	Actual W	Overperformance
1. Detroit Tigers, 1905	64.87	79	+14.13
2. Arizona Diamondbacks, 2007	65.21	77	+11.79
3. New York Janquis, 2004	89.27	101	+11.73
4. New York Mets, 1984	78.33	90	+11.67
5. Kansas City A's, 1955	51.51	63	+11.49
6. Brooklyn Dodgers, 1954	80.71	92	+11.29

7. Cincinnati Reds, 1970	90.77	102	+11.23
8. New York Mets, 1972	71.83	83	+11.17
9. Arizona Diamondbacks, 2007	**78.90**	**90**	**+11.10**
10. Brooklyn Dodgers, 1924	81.25	92	+10.75

The list, given in the second worksheet of the spreadsheet, ranks all teams in history 1 through 2,516, with the bottom teams being:

2512. Pittsburgh, 1911
2513. Baltimore, 1967
2514. Pittsburgh, 1984
2515. Pittsburgh, 1986
2516. New York Mets, 1993 -14.36 games

My sympathies to any 96-year-old Pirate fans in the audience.

Anyway (moving on to the third spreadsheet in the worksheet) I then filled in the "next season performance" for all of these teams which had a "next season" for us to work with. (2,409 of the 2,516 teams had a next season, the exceptions being the 2007 teams, the last-season teams from the Federal League, the Player's League, and the other defunct leagues, and your occasional odd pre-1903 team that went out of business.) This enabled me to look at the average next-season performance of historically over-achieving and historically under-achieving teams.

The first relevant data is shown on lines 2520 and 2521 of the Worksheet 3 of this file. The "over-achieving" teams in the study—100 of them—

1) Scored an average of 659 runs
2) Allowed an average of 684 runs
3) Won an average of 81.78 games
4) Lost an average of 70.35 games,
5) Had an expected winning percentage of .482
6) Had an actual average winning percentage of .537 (aggregate wining percentage of .538)
7) Overachieved by an average of 8.300 games.

We have follow-up season data for 96 of these 100 teams. In the following seasons, these 100 teams again exceeded their Pythagorean expectation, but by an average of only 0.474 wins.

The "under-achieving" teams in the study

1) Scored an average of 695 runs
2) Allowed an average of 683 runs
3) Won an average of 69.16 games
4) Lost an average of 84.06 games,

5) Had an expected winning percentage of .511

6) Had an actual average winning percentage of .454 (aggregate wining percentage of .451)

7) Under-achieved by an average of 8.68 games.

We have follow-up season data for 98 of these 100 teams. In the following seasons, these 100 teams again fell short of their Pythagorean expectation, but by an average of only 0.244 wins.

This gives us the first answer to one of the questions I wanted to study here, which was "Is the persistence of Pythagorean over achievement zero, or near zero?" It appears more likely, based on this study that it is NEAR zero, not actually zero. It appears that there is *some* persistence to the tendency to over- or under-achieve vs. Pythagorean expectations, thus, that such performance is not entirely a fluke.

III. Targeted Method

But I'm not supposed to be reporting results yet; I'm supposed to be explaining what I did.

The real question I was trying to investigate here was "Do over-performing teams have *any* tendency to be better, in following seasons, than teams of the same runs/runs allowed ratio who are not over-performers or who are under-performers?" This was the tedious part of the study. . the rest of this stuff just took me a half-hour or something. The hard part, which consumed a few hours, was to "match" over-performing teams with very similar teams which were not over-performing.

My method was

1) To eliminate from the study any teams which had no follow-up season to be studied.

2) To take the 100 most over-achieving teams of all time (remaining in the data).

3) To try to identify, for each one of those teams, a team with a nearly identical combination of games, runs scored and runs allowed, but with a very different won-lost record.

In other words, the most over-achieving team of all time was the 1905 Detroit Tigers, who were outscored by 97 runs (511-608) but finished 79-74, five games over .500. They were matched with a team they actually competed with, the 1905 St. Louis Browns, who scored 508 runs (three less than the Tigers), allowed 608 runs (the same as the Tigers) but finished 54-99. (For 1905 different sources have different numbers of runs scored and allowed by these teams, although all of the data is similar to this.)

The second most-overachieving team, the 2005 Diamondbacks, scored 696 runs, allowed 856 runs, but finished 77-85. They were matched with the 2006 Tampa Bay Devil Rays, who scored 689 runs, also allowed 856, but finished 61-101.

These are near-ideal matches. Many teams do have near-ideal matches, and some teams do not. . more on that in a moment. The formula that was used to identify "matching" teams is shown in columns AL through AQ of Worksheet 4 of the file "Pythagoras through 2007". The resulting matches are shown in what would be Worksheet 5 of the file, which I have re-titled "Results".

As I said, many teams have near-ideal matches. For example, the 2005 World Champion Chicago White Sox scored 741 runs, allowed 645 runs, giving them 91.3 expected wins. They actually won 99, over-achieving by 7.7 wins.

They are paired with the 1997 Los Angeles Dodgers, who

1) Played the same number of games,
2) Scored one more run (742),
3) Allowed the same number of runs (645), but
4) Finished 88-74.

But some teams do not have real good matches, and wind up being paired with teams which are similar in regard to runs scored/runs allowed per game and expected winning percentage, but pretty different in some other respect. For example, the 1894 New York Giants, who played 137 games, outscored their opponents 940-789 and finished 88-44 (+10.20 games) wind up paired with the 1930 New York Yankees, who played 154 games, outscored their opponents 1062-886 and finished 86-68 (-5.2 games). The two teams are similar in terms of runs scored per game, runs allowed per game and expected winning percentage, but the fact that the one team played 137 games including five ties and the other played 154 games makes it a less-than-ideal match. The 1936 Cardinals and the 2004 Detroit Tigers is a less than ideal match. The 1997 San Francisco Giants and the 1924 St. Louis Cardinals is a less than ideal match. The 2004 Cincinnati Reds and the 1912 Boston Braves is a less than ideal match. The 1907 Chicago Cubs and the 1972 Baltimore Orioles is probably the worst match in the study. I'll let you study the data and reach whatever conclusion you want to about the matches.

IV. Conclusions

In short, my statement that there is no difference in the next-season performance of over-achieving and similar teams is not correct. There is a difference. As I said, I wasn't studying the Emeigh thesis in regard to the Diamondbacks, and I can't comment on whether his hypothesis is true or not. However, based on my study, his hypothesis—which I think I can state generically as "there is a sub-set of over-achieving teams which over-achieve because their runs scored/runs allowed record does not reflect the true quality of the team"—appears to be plausible.

This is the specific data:

1) The 100 most over-achieving teams in history (not counting those which have no next season to be studied) scored an average of 658 runs, allowed an average of 683, yet finished with an average record of 82-71 (82.02 – 70.54) consistent with previously stated data.

2) The 100 teams selected to be nearly identical to those 100 teams scored an average of 658 runs, allowed an average of 689 runs (quality leakage of six runs), but finished with an average won-lost record of 68 – 85 (68.44 – 84.72). The over-achieving teams overachieved by an average of 8.23 wins. Their "matched set" counterparts underachieved by an average of 5.09 wins, so that there was, on average, a separation of more than 13 wins between the two groups.

3) **In the following seasons, the over-achieving teams outperformed their underachieving counterparts by a substantial margin.** In the following seasons the over-achieving teams scored an average of 688 runs, allowed an average of 697 runs, and finished with an aggregate total of 7,636 wins, 7,761 losses. In the following seasons the counterpart teams scored an average of 662 runs, allowed an average of 693 runs, and finished with an aggregate total of 7,237 wins, 8,059 losses. The overperforming teams, in the following seasons, outperformed their counterparts by total of 399 wins, 298 losses—348.5 "games".

It seems to me unlikely that this is a random outcome, but I'll leave that to those of you who know those methods better than I do. I think we're in the neighborhood of four standard deviations deep.

This data and more detail about it can be found in lines 2413 and 2414 of Worksheet Four of the file "Pythagoras through 2007".

V. Observation

A problem with this study is that it essentially exhausts the data, making it difficult to replicate the study. You can't study another group of 100 similarly over-performing teams, because there simply aren't any such teams. This leaves the following options to confirm the conclusion of this study:

1) Study minor league teams (which seems to be entirely useless, because minor league teams have such high turnover that their next-season performance is not a stable or reliable measure.)

2) Look at over-performing teams in minor leagues such as the Pacific Coast League before those leagues became captive slaves of the major leagues (but it would be hard to get ENOUGH data to do that, and good luck finding the data for that.)

3) Study the same in Japanese ball (that should work if you had the data. In a few years researchers will certainly have that.)

4) Keep the 100 over-performing teams in this study, but throw out the 100 matched teams and identify 100 OTHER teams that match them fairly well (but that's would be a weak confirmation if you found the same effect or a weak rebuttal if you didn't.)

5) Replace certain elements of the study with theoretical substitutes. .i.e. expected next-year performance of these teams, rather than actual next-year performance of comparable teams. (That would have been a lot easier, but I didn't do it that way because I thought it whatever I found would be less convincing.)

6) Find "best comps" for the 100 most UNDER-achieving teams, in the same way I have done for the most over-achieving teams.

Appreciate your patience, and look forward to your responses.

Bill James