What is run differential good for?

Phillip G. Rogers

July 2021

1 Introduction

It is well known that run differential can be used to predict a team's future success in baseball and other sports alike. This can be particularly informative when it tells a different story than the corresponding win-loss record. For example, a team with a 26-26 record and a -134 run differential might be expected to perform under .500 going forward. The strongly negative run differential suggests that they have been lucky to win as many games as they have lost so far.

As with any statistical predictor, there are exceptions to the rule. Occasionally a team regresses despite a run differential that suggests that they should improve—or vice versa. So how good of a predictor is run differential? How sure can we be that a team's future performance will match the prediction of their current run differential? This is an empirical question, so let's look to the data to answer it!

2 Data

In order to assess run differential as a predictor of future success, we only need basic game statistics such as teams and scores. However, we need this information for every game in a season. This will allow us to track a team's record and run differential game by game. I collected these statistics for every regular season MLB game from 1960 to 2019 from retrosheet.org.

The key to analyzing this data is to create 'snapshots'. A snapshot is a moment in a season for a particular team from which we can calculate past and future performance. Relevant past performance includes record and run differential up to that point in the season. Future performance is constituted by their record from that moment through the end of the season (note: not their overall record, but just from that point on).

We don't have to choose a particular moment in the season for snapshots. Instead, we can just take a snapshot for every team after every game they play. So for just one season of baseball, we have thousands of snapshots representing different teams and different points in the season.

3 From run differential to prediction

How do we make sense of all this data? The first step is to turn run differential into a predicted winning percentage. It turns out that several statisticians have devised accurate formulas for this task, and these formulas are collectively known as Pythagorean expectation. Here is the simple but effective formula that we utilize:

$$\frac{1}{1+(RA/RF)^{1.83}}$$

Now we can turn the current run differential of a team at each snapshot into a prediction for that team's winning percentage going forward.

Before we continue, it is important to recognize that not every situation is ideal for assessing the usefulness of run differential. For example, if a team is .500 and their run differential is 0, then the run differential will predict the exact same record going forward! So we might focus our attention on situations in which current run differential and winning percentage are somewhat at odds with each other.

We quantify this information with a metric that we call 'offset'. Offset is the absolute difference between a team's current winning percentage and the future winning percentage that is predicted by their current run differential. For example, a team could be .500 yet have a poor run differential that predicts a future winning percentage of .425; the offset for this team (at this snapshot) is 0.075, and that represents a pretty significant gap between their record and run differential.

In the following analyses, we subset the data to include only snapshots with an offset of 0.05 or greater. In this way, we eliminate data points for which it would be difficult to tell whether current record or current run differential is a better predictor of subsequent success. In fact, the entire endeavor can be thought of as a test of which of these is a better predictor. We will return to this framework when we analyze the results.

The record versus run differential framework will also help to assess alternative hypotheses that suggest run differential is not as useful as we might hope. For example, in defense of a team whose record outperforms their run differential, one might argue that the team is clutch: they find a way to win close games, whether through timely hitting or excellent relief pitching. As a result, the argument goes, run differential is not informative! If they are clutch, they will continue to be clutch, and there is no reason to believe that they will regress despite their poor run differential. Our analysis will show to what extent this may be true.

4 The analysis

Our analysis centers around future performance: the winning percentage of a team from the moment of the snapshot through the rest of that regular season. We have now described two benchmarks (predictors) by which to measure that future performance: the current record and the predicted record based on run differential. We'll call the latter RDP (run differential prediction).

Across snapshots, records and offsets differ. In order to analyze the predictions of our two benchmarks, we will normalize our data using these predictors as reference points. For each snapshot, future performance is scaled such that its value is 0 if it is identical to the current winning percentage and 1 if it is identical to the RDP. We can then assess whether the data trends toward 0 or 1.

The data, representing 60 years of professional baseball, is plotted in Figure 1. The x-axis represents the number of games played at the time of the snapshot. As such, points on the left side of the chart are snapshots very early in the season, while points on the right side are very late in the season.

The y-axis represents our normalized future performance values. As a point of reference, dotted white lines are shown at 0 (current record) and 1 (RDP). Points near 0 represent teams that continued to play at about the same winning percentage after the snapshot, despite an RDP that predicted their record would change (for better or worse). Points near 1 represent teams that played close to their RDP after the snapshot. (Note that this RDP could have been better or worse then their record at the snapshot, as long as it differed from their record by at least 0.05.) Points below 0 represent teams that played better even though their RDP predicted that they would play worse, and vice versa. Points above 1 represent teams that outperformed their RDP relative to their current winning percentage (for better or worse).

A red regression line is included to show the overall trend of the data. For the most part, this line stays between the benchmarks of 0 and 1, but it is significantly closer to 1 than 0. This means that current run differential is a better predictor of a team's future success than their current record!

We can also observe how reliable RDP at certain points in a season. One might assume that, early in a season, there aren't enough games to reliably use something like RDP to predict future success. Yet our regression line actually rises toward and even past 1 as it approaches the left side of the plot. This means that even early in a season, RDP is an excellent predictor of future success.

We can make sense of this finding by framing the analysis as a competition between two predictors: current winning percentage and current run differential. You will notice that early-season data points are distributed more widely on the y-axis. This is a representation of the variation and uncertainty that comes from small data sets (snapshots based on very few games). Yet this uncertainty does not just apply to run

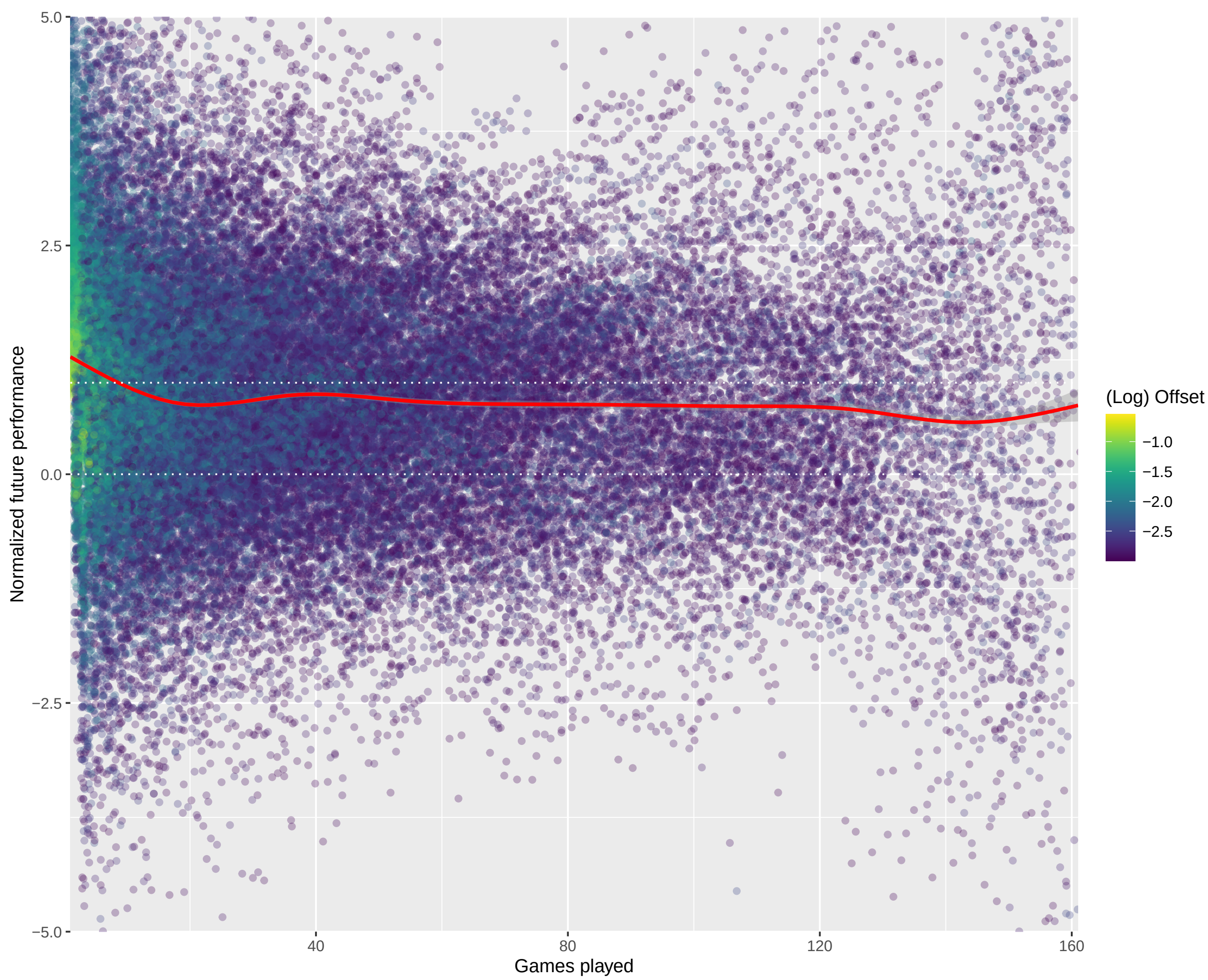

Figure 1: Normalized future performance of baseball teams as a function of games played from 1960 to 2019. A normalized score of 0 means the team's winning percentage after the snapshot matched it's current winning percentage at the time of the snapshot. A score of 1 means the team's winning percentage after the snapshot matched the prediction based on their run differential at the snapshot.

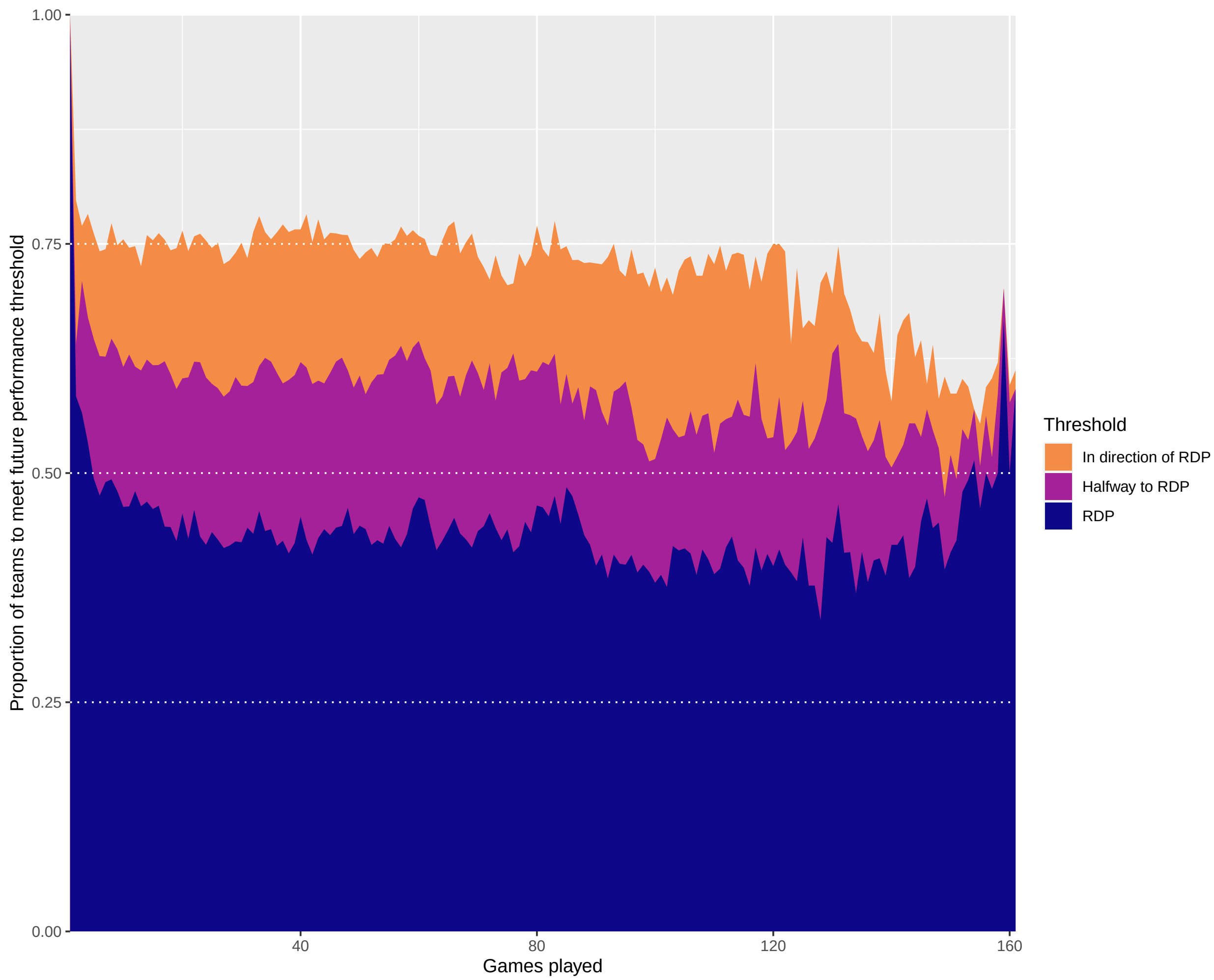

Figure 2: Proportion of teams from 1960 to 2019 that meet thresholds based on run differential prediction (RDP) as a function of games played.

differential, but also to records! In fact, it is wins and losses that become less reliable more quickly when dealing with smaller data sets. In other words, run differential can be fluky, but records are even more fluky.

5 How well (often) does it work?

So far, we have demonstrated that run differential is a better predictor of future success–regardless of sample size–than a team's current record. But how often can we rely on that prediction? To answer this question, we can calculate the percentage of data points that meet certain thresholds. This is illustrated in Figure 2.

We can start with a focus on the first threshold, shown in orange. This represents the proportion of teams whose winning percentage changes in the direction of RDP. Across much of the plot, this proportion hovers around 75 percent. This means that if RDP suggests a team will get better (by at least 0.05 winning percentage), there is a 75 percent chance that the team will show some improvement going forward (and vice versa). Keep in mind that if RDP was completely unrelated to future success, this proportion would be around 50 percent.

The second (middle) threshold represents the halfway point between current winning percentage and RDP. For at least the first half of the season, this proportion is around 60 percent. This means that well over half of teams perform as close or closer to their RDP going forward than to their current winning percentage. Put another way, the RDP is a better predictor of future success than current winning percentage.

The third threshold is the RDP. The proportion of data points that meet this threshold is above 40 percent through the first half of the season. This means that 40 percent of teams play to or exceed the expectations of RDP (for better or worse). This is pretty remarkable.

Finally, the astute reader will notice that these proportions begin to decline after midseason and generally converge near 50 percent at the far right hand side of the plot. This is another artifact of the uncertainty (flukiness) of wins and losses! Just as snapshots early in the season suffer from small "pre" sample size, snapshots late in the season suffer from small "post" sample sizes. For example, a snapshot taken after 160 games is then used to predict the final 2 games, and a team might win those 2 games even if they were pretty bad all season. So the accuracy of predictions will vary widely when they are tested on such small data sets.

6 Conclusion

Using decades of MLB data, we've shown that run differential is a better predictor of a baseball team's future success than its current record. This trend is even stronger when you apply it very early in the season. We've also shown what this means in practice: 75 percent of teams perform in the direction predicted by RDP, and 40 percent reach or exceed that prediction. So if you want to know how your team will play going forward, remember that the game is all about scoring runs and preventing the other team from doing the same.