Why We Should Use the Bullpen Differently

A look into how the bullpen can be better used to save runs in Major League Baseball.

Andrew Soncrant

Statistics 157 Final Report

University of California, Berkeley

April, 2016

Abstract

The bullpen is one area of a baseball team where managers have a lot of control. They can choose from 6-10 pitchers based on handedness, velocity, specialties, and experience. They usually make these decisions to select the pitcher most likely to succeed in whatever situation the game presents. This logic fails, however, when it comes to handling save situations. Typically, in save situations, managers go to their closers. Closers are supposed to be the best relief pitchers on their teams, and are entrusted with the last inning of close games when their teams are ahead. This report suggests an alternative method of bullpen use, one that matches relief pitchers to situations befitting their skill using the Leverage Index (Li).

The leverage index attaches a number to game situations, and represents how important a situation is to the outcome of a game. High leverage situations translate to higher pressure situations, where the outcome of the situation is more likely to reflect the outcome of the game. We will first show that a team's performance in high-leverage situations, calculating individual and team leverage-performances (LevP), is an indicator of that team's success throughout the season. We will then use the leverage index to show that for a number of teams, employing a system where all bullpen pitchers, including closers, are used in leverage situations matching their skill level will result in fewer runs allowed. Specifically we will employ a system where average pitchers are allowed to pitch in save situations, and "closers" pitch in earlier situations that have high leverage.

Table of Contents

Introduction

In today's game, when their team is ahead, the better a relief pitcher is, the later in the game they pitch. This strategy came about for a number of reasons, the biggest reason being as a team advances through a game, holding onto their lead becomes more important. By the 9^{th} inning, for instance, if you blow your lead then you lose the game. The idea that some situations have a bigger impact on the outcome of a game is not a new one, and has been quantified by the Leverage Index (Li). Li is a statistic assigned to situations, and players, in order to measure how important a situation is to the result of a game. For pitchers, their associated Li is the average leverage of the game situations in which they appeared in during the season. Not surprisingly, situations that occur later in games tend to have a higher leverage than situations that occur early, and better relief pitchers tend to have a higher Li.

Leverage is one statistic we use in our analysis. Throughout this report we will also refer to saves and save situations. For the purposes of this paper, we will define a save as any time that a team is ahead by one, two, or three runs going into the ninth inning. If the team successfully ends the game in the ninth inning, the pitcher who records the final out is awarded a save. In today's game, saves carry a lot of weight towards things like salary, playing time, and bonuses in player contracts. We will also refer to some general pitching statistics. Runs (R) refers to the number of total runs that a pitcher allows to score while they are pitching. Innings pitched (IP) refers to the total number of innings a pitcher pitches in a season, and R-9 is the number of runs a pitcher allows, divided by their innings pitched, and then multiplied by 9 (also called "runs per nine").

Lastly, a statistic that will use will be leverage-performance (LevP), calculated by taking a pitchers R-9 and dividing by their Li. This acts as a measure of how well pitchers, and teams, perform in situations based on leverage.

Methodology

Most data for this project comes from Baseball-Reference, where I gathered individual and team data from 2015. I originally had a large list of data frames. Each frame corresponds to a team, with each row in the frame being for each pitcher on the team. The data frames were slimmed down to only include relief pitchers with at least 10 IP and to only include relevant statistics (IP, R, R-9, Li).

Part 1

The first part of this project is to show that LevP is a meaningful statistic, that is, that it is correlated to a team's overall performance. Using this data I was able to calculate individual leverage-performance statistics by dividing each pitcher's R-9 by their Li. Then I calculate a team LevP by taking the average of each individual's LevP, weighted by their IP. Using OLS Regression, we regress team wins onto team LevP (Figure 1, 2).

Part two of this project was to show that in a system where pitchers are used in leverage situations matching their abilities, teams can save runs. As mentioned, we will use average pitchers in save situations, and use closers in high leverage situations arising earlier in the game. This, in turn, has its own two parts.

- Calculate the runs lost by having average pitchers pitch in save situations rather than closers

- Calculate the runs added by having closers pitch in high-leverage situations earlier in games

In order to perform both of these parts, we must make some assumptions. First, we assume that LevP is linear. LevP itself is a measure of how many runs a pitcher would allow at leverage situation of one. To calculate the same at a leverage of two, we just multiply that pitcher's LevP by two. Next, we make an educated guess as to how often certain save situations occur. As mentioned, a save situation arises when a team is up by one, two, or three runs going into the ninth inning. This leaves us with three types of save situations: one run, two run, and three run saves. We assume, from analyzing game state frequencies from Retrosheet:

- 1 run: 50% (Li = 2.9)
- 2 run: 30% (Li = 1.6)
- 3 Run: 20% (Li = 0.8)

Part 2A

Notice that the average leverage of a save situation, weighted by these frequencies, comes out to 2.09. Now we calculate the number of runs lost by having ordinary pitchers pitch in save situations, rather than closers. Then, because of our assumptions, calculating the expected runs lost for each team is just algebra:

LevP = team leverage performance

SV = # save opportunities per team

RA = expected runs allowed

RA = (LEVP)*(2.09)*(SV/9)

League-wide, the average expected runs allowed comes out to an average of 55.33. Lastly, from this number we subtract the number of *actual* runs surrendered by closers, which tells us how many runs *worse* the rest of the bullpen would perform than the closer. For each team, we call this number *runs lost (RL).* The league's average RL is 23.5.

Part 2B

Now we calculate the runs added by having the closer pitch in mid-game situations with a high leverage. But first a note:

- Analyzing game frequencies is very difficult. Short of raw play-by-play data on each game, no such data is available that makes it feasible to analyze how often certain game states occurred in 2015.

In order to combat this issue, we work backwards. We set a leverage at which each team's closer will only pitch at, or above, during the game. We then determine how many innings a team must use their closer at, or above, this leverage until they have successfully gained back more runs than their RL.

In a similar manner, we calculate the expected runs *added* by each team. Only this time, we do so a little differently. First, we create a vector of integers 1-200, to simulate the number of innings that the closer pitches. For each number in this vector, we calculate as before:

Lev = Chosen leverage

LevP = Closer's LevP

Expected = (LevP)*(Innings/9)*(Lev)

As before, we subtract this result from the number of runs the bullpen *actually* surrendered in such scenarios. Though, as mentioned, lack of available data makes this impossible to do directly. Instead, we perform the same calculation as above, only substitute

LevP = *Team LevP*

The result is a vector of 200 numbers, representing the number of runs added for each team. For each team, we then determine at which index the runs added exceeds the RL. This index is the number of innings at which a team must use their closer at or above our chosen leverage in order to save runs. After examining the leverages of certain situations, and my results at a number of leverages, I settled on a chosen leverage of 2.8.

Results

Part 1

After calculating team LEVP and Wins, we plot them against one another.

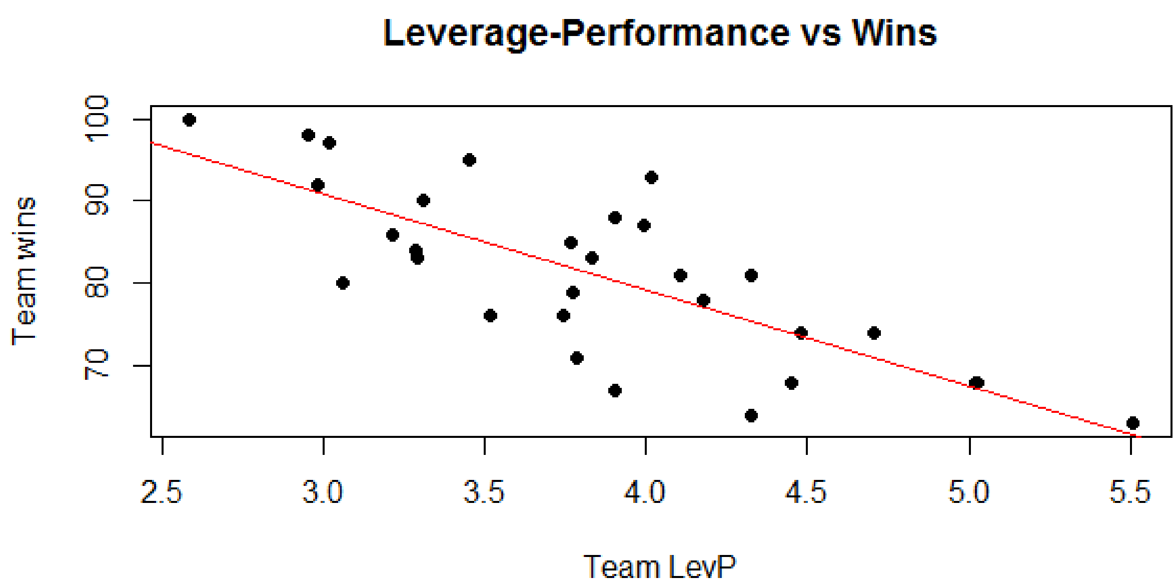

Figure 1

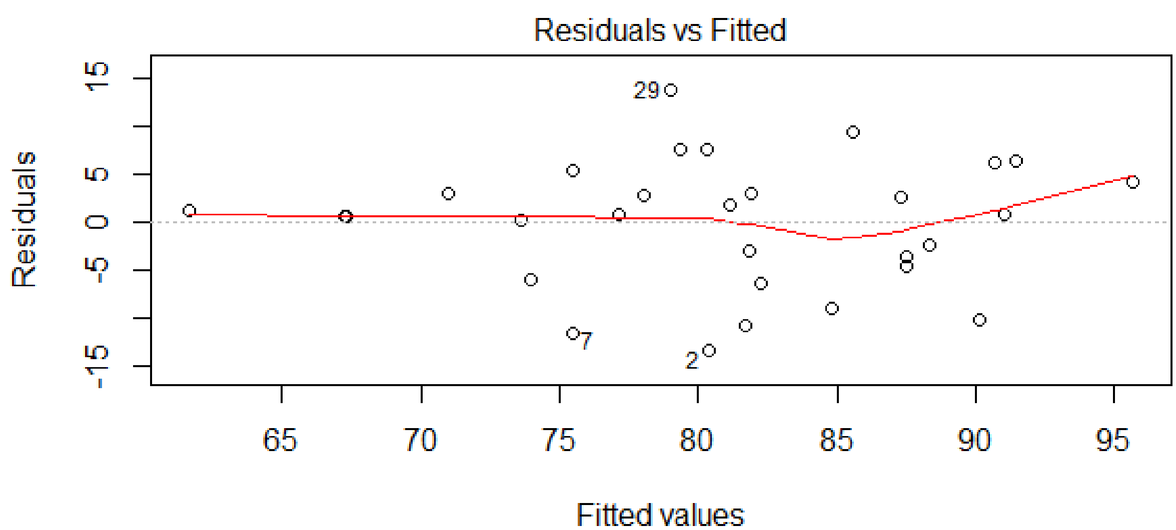

Figure 2

Not surprisingly, we see the general trend that teams with a higher (worse) LevP tend to have fewer wins than teams with a lower LevP. We also examine the residuals (Figure 2). From these graphs we can reasonably assume that LevP is an indicator of regular season success. This is important for our results because it allows us to assume that if we can reduce the number of runs allowed, using LevP, then a team's performance will improve.

Part 2

Calculations from this section result in LevP, RL, and the innings at which a team must use their closer at or above a leverage of 2.8, calculated for each team. These results are shown below, each sorted from best to worst:

Team	Lev Perf
STL	2.584000
PIT	2.950820
LAD	2.983871
CHC	3.016129
TBR	3.061538
HOU	3.216667
SFG	3.285714
WSN	3.289256
NYM	3.307018
KCR	3.452174
SEA	3.519685
CHW	3.741379
LAA	3.767857
ARI	3.775862
MIA	3.787611
MIN	3.833333
ATL	3.902439
TEX	3.905983
NYY	3.990741
TOR	4.019231
BAL	4.103774
BOS	4.178571
CLE	4.322581
CIN	4.324074
OAK	4.450980
SDP	4.480392
DET	4.707547
MIL	5.021739
COL	5.028302
PHI	5.510870

Teams	Runs Lost
SEA	-31
CHC	2
STL	7
BOS	9
WSN	12
HOU	15
TBR	15
CHW	17
TEX	19
TOR	19
CLE	21
NYY	22
PIT	23
BAL	27
LAD	27
MIN	27
ARI	29
ATL	30
KCR	30
MIA	30
LAA	32
PHI	32
SFG	32
COL	33
NYM	36
OAK	37
DET	40
SDP	40
CIN	42
MIL	48

Teams	Innings Needed
SEA	1
CHC	3
SDP	9
BOS	12
STL	12
CHW	15
CLE	20
NYY	21
TEX	21
TOR	21
BAL	24
MIN	25
WSN	25
ATL	27
OAK	27
PHI	27
TBR	27
DET	32
MIA	33
CIN	37
LAA	39
MIL	48
LAD	51
NYM	53
ARI	54
HOU	56
PIT	56
SFG	57
COL	59
KCR	111

Figure 3

The most important of these three figures is the rightmost, which tells us how many innings a team must use their closer in order to begin saving runs. The average number of innings required is 33.43, with a median of 27. The graphic below helps to explain how the numbers in the third figure are calculated. As a sample, we use my favorite team, the Washington Nationals.

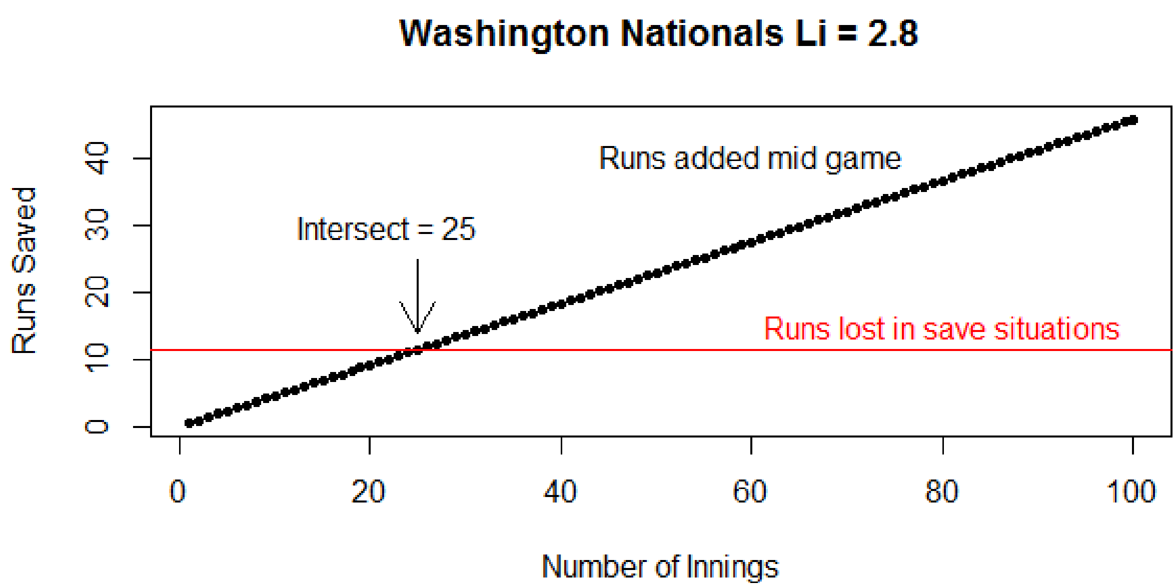

Figure 4

The RL line is flat in this graph because in 2015, the number of save situations for each team was a set number, and so the expected number of runs lost do not vary with innings pitched.

Analysis

These preliminary calculations, though by no means perfect, seem to indicate that for a number of teams, using the suggested bullpen method would save runs over the course of a season. There are a number of in-game situations with a leverage of over 2.8[1]. Just to name a few:

- Any bases loaded situation after the 5th inning
- 7th inning, one out, one run lead, runners on second and third
- 8th inning, no out, one run lead, runner on third

While not common on their own, collectively situations at or above a leverage of 2.8 occur with some regularity. In my opinion, assuming that 30+ innings will occur at or above this leverage every season is reasonable. It follows that at least half of the league stands to benefit from using the bullpen in this manner.

One question that arises from these results is, "what makes this proposed system effective for some teams, but not for others?" One way of answering this is to look at the best and worst teams in terms of the number of innings they need to use the new system: The Kansas City Royals and the Seattle Mariners.

> The Royals had arguably the most effective bullpen in baseball in 2015. As a result, the difference between the LevP of their entire bullpen, and their closer, was very small at just .87. Thinking in terms of Figure 4, the slope of their Runs Added line is very flat. This translates to a large number of innings needed before they begin to save runs, because the difference between their closer and the rest of their bullpen isn't very pronounced.

The Seattle Mariners, on the other hand, had one of the worst bullpens in baseball, including their closers. With a RL less than zero, we see that their regular bullpen would have performed *better* than their closer(s). As such, any amount of time using this new system would have saved them runs.

The rest of the teams in the league, however, fall somewhere in the middle, and with no obvious way to calculate why. The formula for calculating results is very straight forward, but a few things make the results of any individual team difficult to interpret:

- The linear assumption of LevP may not be absolutely correct.
- Each team had a different number of save situations, varying the number of innings under which they would need to use this new system.
- The health of a team's relief pitchers influence playing time, and it's possible some pitchers were not pitching when they should have been, even by today's standard.
- Random chance is always a factor. Runs are not always evenly distributed by players or at certain leverages. Some teams may have over or under-performed.
- Different teams face a different level of competition. Strength of schedule doesn't account for streaking teams (good or bad) which can affect runs allowed
- This is not a method that exactly matches players with situations befitting their skill. It is a rough approximation, which only substitutes average pitchers into save situations. Some teams would vary more, or less, from such a general strategy.

Conclusion

These results are not perfect. A lot of assumptions had to be made along the way: the definition of a save, the linearity of LevP, the frequencies of save situations, and that calculating expected runs allowed the way we did is acceptable.

Concluding that leverage-performance is important, however, is much more concrete, and is in and of itself a reason to consider switching to the system I am proposing. Save situations tend to occur at a lower leverage than the level at which I am proposing we use closers, and so continuing to use closers in these situations is giving up runs that could be saved elsewhere in the game. As a result, when considering this to be a rough model for what would happen if teams switched to a system that assigns pitchers to situations befitting their skill level, we come to the conclusion that at least half of the league stands to benefit from such a switch.

Sources

- Retrosheet.org
- Baseball-Reference.com
- http://www.insidethebook.com/li.shtml (1)