Matt Halper
12/10/14
Stats 50

The Batting Pitcher:

A Statistical Analysis based on NL vs. AL Pitchers' Batting Statistics in the World Series and the Implications on their Team's Success in the Series

Matt Halper

In this extended abstract, I will examine the implications of the batting pitcher in the World Series through numerous methods of analysis including correlation, Bill James Estimator, Game Theory, NERV, and more. To introduce the concept, Major League Baseball has two sets of rules for the American League and the National League. In the American League, pitchers do not bat, and instead, a designated hitter (DH) bats for them. The vast majority of games are played between teams in the same league. However, in two situations, this is not the case- the first is scheduled interleague play and the second is the World Series.

To begin this research process, I scrolled through box scores and play by plays of the last 20 years. Oddly enough, pitcher-batting statistics are not accumulated together and I found it necessary to in order to find the statistics I wanted. I went through databases of World Series statistics over the last 20 years and accumulated pitcher's performances at the plate during the World Series. I used Excel to collect, organize, and analyze the data I used.

Let us first look at some of the statistics:

NL AB	AL AB	NL BB	AL BB	NL Hits	AL Hits	NL Total Bases	AL Total Bases	NL RBIs	AL RBIs	NL Runs	AL Runs	NL Strikeouts	AL Strikeouts	Winner NL	Winner AL
6	7	0	0	1	0	1	0	0	0	0	0	2	3	1	0
6	7	0	0	0	0	0	0	0	0	0	0	4	3	0	1
5	3	0	0	1	0	1	0	1	0	0	0	3	1	1	0
7	8	1	0	0	0	1	1	0	0	0	1	3	2	1	0
6	3	0	0	1	1	1	2	0	0	0	0	1	0	1	0
6	4	0	0	2	0	2	0	0	0	0	0	3	3	0	1
8	5	0	0	1	1	2	1	1	0	1	0	3	2	1	0
3	5	0	0	1	1	1	1	0	2	0	0	1	2	0	1
8	4	0	0	0	0	0	0	0	0	0	0	1	2	1	0
3	5	0	0	0	0	0	0	0	0	0	0	1	2	0	1
2	4	0	1	1	0	2	1	0	0	1	0	0	3	0	1
9	5	0	0	1	1	1	1	2	0	0	0	4	3	1	0
3	6	0	0	1	1	1	1	1	0	1	0	1	2	0	1
13	7	0	0	1	1	2	1	1	0	2	0	7	4	1	0
5	8	0	0	1	0	1	0	0	0	0	0	1	4	0	1
2	6	1	0	0	0	1	0	0	0	0	0	2	0	0	1
4	6	0	0	2	1	2	1	0	0	1	0	0	2	0	1
7	10	0	0	0	2	1	3	0	2	1	1	2	5	1	0
4	9	1	0	1	1	2	1	0	0	1	0	1	2	0	1
7	6	1	0	0	0	1	0	0	0	0	0	3	1	1	0
114	118	4	1	15	10	23	14	6	4	8	2	43	46	0.5	0.5

Here I found at-bats, walks, hits, total bases, RBIs, runs, strikeouts, and eventually the winner of the series (marked by a "1" in the last column). I took all of these statistics in order to get a plethora of data so as to be as confident as possible in the correlations. From this graph, I automated excel formulas to calculate Batting Average, Slugging Average, and On Base Percentage below.

	NL Average	AL Average	NL Slugging	AL Slugging	NL OBP	AL OBP
2014	0.166667	0	0.166667	0	0.166667	0
2013	0	0	0	0	0	0
2012	0.2	0	0.2	0	0.2	0
2011	0	0	0.142857	0.125	0.125	0
2010	0.166667	0.333333	0.166667	0.666667	0.166667	0.333333
2009	0.333333	0	0.333333	0	0.333333	0
2008	0.125	0.2	0.25	0.2	0.125	0.2
2007	0.333333	0.2	0.333333	0.2	0.333333	0.2
2006	0	0	0	0	0	0
2005	0	0	0	0	0	0
2004	0.5	0	1	0.25	0.5	0.2
2003	0.111111	0.2	0.111111	0.2	0.111111	0.2
2002	0.333333	0.166667	0.333333	0.166667	0.333333	0.166667
2001	0.076923	0.142857	0.153846	0.142857	0.076923	0.142857
2000	0.2	0	0.2	0	0.2	0
1999	0	0	0.5	0	0.333333	0
1998	0.5	0.166667	0.5	0.166667	0.5	0.166667
1997	0	0.2	0.142857	0.3	0	0.2
1996	0.25	0.111111	0.5	0.111111	0.4	0.111111
1995	0	0	0.142857	0	0.125	0
TOTALS	0.131579	0.084746	0.201754	0.118644	0.161017	0.092437

As pictured above, the NL has dominated the AL in every category. That is not unexpected at all since National League pitchers tend to have much more experience and practice at the plate simply because of the rules of each league. National League Slugging and On Base Percentage are both nearly double that of the American League. The differences are striking, and actually close to consistent with the regular season statistics. As you can see, each league has won ten World Series titles in the last 20 years. I went on and calculated the correlation of pitchers batting performance with the results of the series. The results were peculiar:

CORRELATION WITH WINS						
	NL Average	AL Average	NL Slugging	AL Slugging	NL OBP	AL OBP
	-0.496324	-0.209976	-0.487776	-0.233778	-0.584883	-0.111763

In all the cases, the correlation was actually negative. In other words, the better the pitchers hit, the less often their team won the series. There are numerous reasons for why this could be happen. What I take away from these correlations is that so far it does not seem statistically likely that NL pitchers superior batting is making the difference in series that they win. The negative correlation is most likely attributed to a limited sample size rather than some reason why batting success would lead to team failure.

This might seem surprising at first when we use methods of analysis like the Bill James Estimator:

$$p \sim \frac{PS^{\alpha}}{PS^{\alpha} + PA^{\alpha}}$$

For baseball, the constant alpha is two. Using the estimator to pin National League and American League pitchers in the World Series over time against each other, we get

seemingly severe results. Over the last 20 years, National League pitchers have scored eight runs, while American League pitchers have scored two.

NL:

$$.941 = \frac{8^2}{(8^2 + 2^2)}$$

AL:

$$.059 = \frac{2^2}{(2^2 + 8^2)}$$

Looking at just the numbers, we can take that the National League has over a 94% chance of winning to the American League's 6%. Purely from this estimator, it looks like this particular run differential would cause some sort of movement in the World Series. However, when spread over 20 years, the estimator holds much less significance because it is diluted with thousands more at-bats from other players at superior batting positions. With the Bill James Estimator, we run into the issue of separating statistics where they do not necessarily separate and are more accurately viewed in the context of the specific situations over time. Inherently in this issue exists a similar story- there is so little pitcher production in a World Series that even after 20, the numbers are surprisingly low.

I want to attempt to bring the subject into the realm of game theory to explain why this batting superiority exists in general before diving in and uncovering the implications. It is difficult to assign exact values of gained or lost utility for this situation, but there is an inherent game theory involved. The decision for whether to make pitchers participate in batting practice is a decision general managers and coaches make all the time. Since pitchers are distributed to National League and American League teams without their batting merits considered, we can assume that a league full of pitchers on each side would field very close batting statistics and averages if they practiced the same amount. Below is a chart displaying the game:

	AL batting practice	AL no batting practice
NL Batting Practice	Batting: AL = NL Pitching: AL-, NL-	Batting: NL+, AL- Pitching: AL+, NL-
NL No Batting practice	Batting: AL+, NL- Pitching: AL-, NL+	Batting: NL = AL Pitching: AL+, NL+

Essentially, if National League pitchers participate in batting practice and American League ones do not, National League pitchers will gain positive utility at the plate compared to American League pitchers. However, National League pitchers would also seemingly lose a little utility in pitching performance in regards to American League

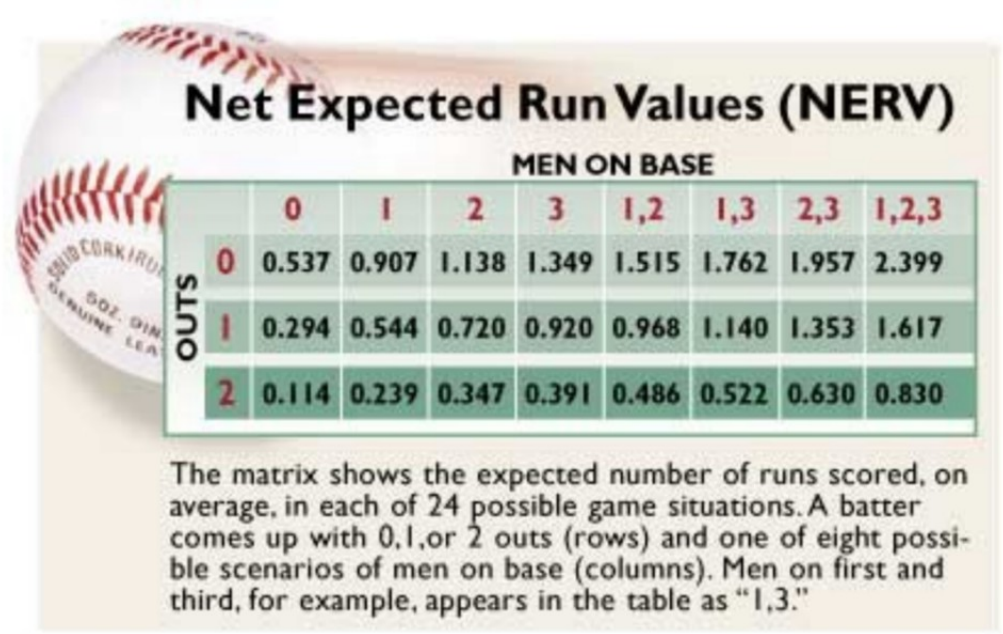
pitchers. The same goes the other way around. If both leagues do or do not participate in batting practice, they will together see an equal increase or decrease in batting "performance" or increase/decrease in pitching performance, depending on the situation. The *difference* is in how much or little utility is gained or lost through this performance. National League starters average 62 plate appearances during the year. American League Starters average seven. From this information, we can gather that the increase in utility for National League pitchers participating in regular batting practice is significantly higher than that of American League pitchers. A certain amount of batting practice is worth the utility lost from less pitching practice for the National League. It is less the case for the American League. This situation dictates where pitchers end up at on the game theory chart. I estimate it is somewhere around where the "X" lies:

	AL batting practice	AL no batting practice
NL Batting Practice	Batting: AL = NL Pitching: AL-, NL-	Batting: NL+, AL- Pitching: AL+, NL- X
NL No Batting practice	Batting: AL+, NL- Pitching: AL-, NL+	Batting: NL = AL Pitching: AL+, NL+

In terms of competitive advantage, this arrangement makes sense. Does it make sense for AL pitchers to practice? Probably not. National League? Probably.

From here, I want to transition to some deeper statistics that help paint the story more clearly and answer some of our questions. Using both NERV and wWPA in a specific situation, we can see where some of the implications or lack of implications may lie. Below is a graph of values for NERV for reference:

NERV:



Net Expected Run Values (NERV)

OUTS	MEN ON BASE: 0	1	2	3	1,2	1,3	2,3	1,2,3
0	0.537	0.907	1.138	1.349	1.515	1.762	1.957	2.399
1	0.294	0.544	0.720	0.920	0.968	1.140	1.353	1.617
2	0.114	0.239	0.347	0.391	0.486	0.522	0.630	0.830

The matrix shows the expected number of runs scored, on average, in each of 24 possible game situations. A batter comes up with 0,1,or 2 outs (rows) and one of eight possible scenarios of men on base (columns). Men on first and third, for example, appears in the table as "1,3."

Winning Team Win Probability Addition (wWPA) is a percentage that I found very helpful in my analysis. For each plate appearance, the result is calculated through a series of formulas to find the percent increase or decrease a plate appearance had on the probability at that moment of the winning team winning. This is similar to the "darkest before dawn" concept we talked about earlier in class.

Let's take game four of the 1998 World Series. The American League Yankees are up 3-0 on the National League Padres. The matchup is between both team's aces, Kevin Brown for the Padres against Andy Pettitte of the Yankees. From the ending box score, Kevin Brown goes a solid 1 for 2 with a single and no strikeouts while Andy petite ends up going 0-2 with two strikeouts and a sacrifice bunt.

Here we will analyze each plate appearance by the pitchers. In the first at-bat, Kevin Brown faces Andy Pettitte in the bottom of the second. The situation is bases are loaded with two outs. Kevin Brown tries to bunt and gets thrown out to end the inning. This blown opportunity in the stat book or the traditional stats we saw, but definitely not positive for the Padres nevertheless. In this one bases loaded situation the Padres lose 8% probability for victory and -.83 expected runs.

wWPA = 8%
NERV= -.830 runs

Pettitte's first at-bat comes in the top of the third. There are no men on and two outs. He strikes out. He only loses 1% of victory for the Yankees and about one-tenth of a run.

wWPA = -1%
NERV = -0.114

The second at-bat for Brown comes in the 5th inning once again facing Pettitte. He singles in centerfield with no one on. Quilvio Veras grounds out to the pitcher the next at-bat and the inning is over. During that at-bat, he provided 2% more probability for the Padres victory and provided .125 of expected runs.

wWPA = -2%
NERV = 0.239 – 0.114 = +0.125

Pettitte's second at-bat comes in the 5th against Brown. Runners are on 2nd and 3rd with two outs in the inning. Pettitte strikes out swinging. In this at-bat, Pettitte provides the Padres 8% higher chance of victory and loses .63 expected runs for the Yankees.

wWPA = -8%
NERV = -0.630

While Kevin Brown is replaced, Pettitte stays in and has a third plate appearance. There is a man on 1st with one out. Pettitte sacrifice bunts to move Ricky Ledee to 2nd in scoring position. He gives the Padres a 2% larger probability of victory.

wWPA = -2%
NERV = 0.347 – 0.544 = -0.197

The resulting totals are accumulated for the pitchers below:

Kevin Brown total:
wWPA = 6% during his 2 at-bats
NERV = -0.705

Andy Pettitte total:
wWPA = -11% during his 3 plate appearances
NERV = -0.941

The Yankees won this game to win the series 4-0. For this matchup, the truth is neither pitcher batted well or contributed positive utility from the plate. And like usual, the National League pitcher was better at batting than the American League one. Essentially Kevin Brown was less of a negative utility than Pettitte when both at the plate. But in reality while Kevin Brown's stats at the plate, 1 for 2 with a single and no strikeouts look pretty good on paper, they are a lot closer to Pettitte's contributions than what we might first expect. This is not necessarily exclusive to this particular situation at all. If we multiply Kevin Brown's contributions by 1.5 to match his plate appearances with Pettitte, we find that his wWPA = 9% and his NERV = -1.058, which is much closer to Pettitte's contributions, especially in the context of the last 20 years, where the American League has only *one* more plate appearance than the National League in the

World Series. Essentially, the intricacies of the situation seem to often balance out the few and far between pitcher at-bats in the World Series.

To conclude, while National League pitchers are significantly better performing hitters historically in the World Series, there is little data to suggest that this difference in performance between the two leagues' pitchers has influenced the outcomes of the series as a whole.

There are numerous likely explanations for this. In the end, there are not enough at-bats to show significant differences. While the stats for the most part speak for themselves, during the course of the World Series the advantage seems to become lost in the numbers. Over the last 50 years, World Series' have gone into exactly 6.0 games on average. From that, there are only 3 games at NL parks. On average, there are about two at-bats for pitchers in those games. Six at-bats in a series that usually ranges from 130 to 250 at-bats per team has not shown its significance enough to conclude that it influences the series as a whole.

Another explanation for this is the fact that American League teams usually have a big-name contracted slugger as a designated hitter, while National League teams are forced to improvise when they play interleague or in the World Series, and require a DH. Traditionally and logically, these American League DHs perform better than National League ones. This is no doubt a reason why these statistics even out in some areas.

Lastly, while there have been many World Series, the sample size simply is still not very large. Pitchers only bat a few times per World Series and there is only one World Series per year. Also, this has only been the case since 1973 when the DH was implemented- that limits the data even further. Baseball is a sport about the numbers, but it requires vast amounts for a confident and conclusive result.