Joseph D'Onofrio
ECO 247 - Sabermetrics
Research Paper

A Statistical Analysis of the Hit-and-Run

Background/Hypothesis

The hit and run play has been a staple of small ball baseball for as long as the game has been played. The idea of the hit and run is to try to advance the runner one or more bases by having the runner steal on the pitch and having the batter try to hit a ground ball. This can lead to the runner advancing whether it be by a successful steal, a well-placed ground ball, or a hit by the batter.

The advent of the sabermetric era has called many old-school baseball notions into question. The goal of this analysis is to attempt to determine the statistical validity of the hit and run play. As with any analysis, we must first start by presenting a hypothesis. I hypothesize that for the batter, the hit and run play is at least no worse than the standard no-play alternative. For the runner, I hypothesize that the rate of stolen base success will slightly increase when the batter swings and misses on the hit and run attempt.

Data

The data used in this analysis comes from the Retrosheet play-by-play database consisting of every play which occurred in the MLB from 2005-2015. The data was trimmed down to include only hit and run plays, defined as plays where the runner attempted a steal and the batter swung. All data processing was done with a Python script written for this analysis.

Additional data was taken from a Baseball Prospectus article exploring the validity of the hit and run play and formulas/constants for advanced metrics were taken from the FanGraphs website.

Methods

The first thing I wanted to look at was how often a hit and run is actually successful. A success is defined as any time the runner advances one or more bases, regardless of what happens to the batter. This is accomplished by a simple calculation of the number of successes over number of attempts. The success rate for all hit and run plays from 2005-2015 came out to be 0.462, or 46.2%. So, slightly less than half of all hit and run attempts produce the desired result.

The next thing I wanted to know was how the batter was effected by one or more attempted hit and runs in an at bat. Attempting a hit and run could affect a hitter in a number of ways. A stealing runner can put the defense out of position, granting a benefit to the hitter. However, the batter may have to swing at a bad pitch in order to attempt the play, which could result in either a weakly hit ball or a free strike for the pitcher. To investigate this I started by calculating the batting average of at bats which

include at least one hit and run attempt, and I then delved deeper into more advanced metrics to try to further understand the effect on the hitter.

The batting average of hit and run at bats is another simple calculation of hits over at bats. The batting average of hit and run at bats came out to be 0.252, which we can compare to the league average of 0.259. This weakly indicates that the hit and run play has very little effect on the batter. However, since batting average isn't necessarily a great indicator of value, we will turn to more advance metrics, starting with weighted on base average (wOBA):

FANGRAPHS

Weighted On-Base Average (wOBA) is a rate statistic which attempts to credit a hitter for the value of each outcome (single, double, etc) rather than treating all hits or times on base equally. wOBA is on the same scale as On-Base Percentage (OBP) and is a better representation of offensive value than batting average, RBI, or OPS. The weights change slightly with the run environment, but the general formula is:

$$wOBA = \frac{.69\times uBB + .72\times HBP + .89\times 1B + 1.27\times 2B + 1.62\times 3B + 2.10\times HR}{AB + BB - IBB + SF + HBP}$$

This calculation is rather complex, but luckily the data included a column which indicated the amount of wOBA points granted for a given event. From there a simple average of the wOBA points of all hit and run at bats gives us an average wOBA of 0.319 for those at bats. The league average wOBA for 2005-2015 was 0.321, so we now have a slightly stronger indication that a hit and run attempt does not have a large effect on the batter.

Since baseball is all about runs scored, the next thing to investigate is what kind of effect the hit and run play has on runs scored. To do this we will turn to a variant of the Runs Created metric called weighted runs created plus (wRC+):

FANGRAPHS

Weighted Runs Created Plus (wRC+) is a rate statistic which attempts to credit a hitter for the value of each outcome (single, double, etc) rather than treating all hits or times on base equally, while also controlling for park effects and the current run environment. wRC+ is scaled so that league average is 100 each year and every point above or below 100 is equal to one percentage point better or worse than league average. This makes wRC+ a better representation of offensive value than batting average, RBI, OPS, or wOBA.

$$wRC+ = \frac{(wRAA/PA + LgR/PA) + (LgR/PA - (Park\ Factor \times LgR/PA))}{(AL\ or\ NL\ wRC/PA\ excluding\ pitchers)} \times 100$$

While this is generally a catch-all offensive statistic assigned to a batter, we will be calculating is using hit and run at bats to compare the run creation potential of a hit and run to the average at bat. Since the average wRC+ is always 100 we have a nice scale on which to compare to.

wRC+ can be calculated by using the average hit and run wOBA calculated previously and constants/data taken from the FanGraphs website. Constants/data are provided by season and were averaged across the 11-season span. wRAA/PA is calculated as Hit and Run wOBA minus league wOBA divided by the wOBA scale constant and the rest of the data is provided by FanGraphs. Park Factor is assumed to be the average of 1.0 as these at bats span all parks. After calculation, we get a wRC+ of 96.04 for hit and run at bats, just shy of the league average 100 wRC+. By the definition of the wRC+ metric we can now say that at bats which include a hit and run attempt perform 4% worse than the average at-bat.

The next step in this analysis is to determine if there is a benefit to the steal attempt of the runner of a hit and run steal over a no-swing straight steal. For this we will turn to a pair of tables included in a Baseball Prospectus article which uses another variant of the Retrosheet data to analyze hit and runs from 2003-2011.

Outs	Stolen Base Success Rate		
	Batter Takes	Batter Swings	Difference
0	79.0%	60.4%	-18.6%
1	77.3%	64.3%	-13.0%
2	79.0%	81.1%	+2.2%

Count	Stolen Base Success Rate		
	Batter Takes	Batter Swings	Difference
0-0	80.5%	58.3%	-22.2%
1-0	76.8%	60.7%	-16.1%
2-0	84.3%	72.4%	-11.9%
3-0	90.2%	n/a	n/a
0-1	77.7%	72.8%	-4.8%
1-1	75.6%	59.3%	-16.2%
2-1	72.6%	62.3%	-10.3%
3-1	68.9%	63.9%	-5.0%
0-2	82.5%	84.1%	+1.6%
1-2	79.5%	84.7%	+5.2%
2-2	78.4%	72.3%	-6.1%
3-2	57.0%	54.0%	-3.1%

A hit and run appears to have a significant negative impact on the runner's steal attempt in a majority of situations with the exception of when there are two outs or if it is a 0-2 or 1-2 count. The increase for the two strike counts could be a result of the higher volume of off-speed pitches thrown in these counts, as a curveball in the dirt is an easier pitch to steal on. On these counts a hit and run may be a favorable

play, as the batter is already in a bad position and the runner will have a higher chance of a successful steal. On most other counts however, a hit and run seems to have a large negative impact on steal success.

Summary & Conclusion

The picture to the right is the output of the Python script which summarizes the above analysis. These stats, plus the stolen base tables above seem to indicate that for most situations, the hit and run play has a small negative effect when compared to the average no-play at bat.

```
Hit and Run success rate: 0.4624601850447444

BA for AB with HnR attempts: 0.25184170255169225
League Average BA 2005-2015: .259

wOBA for AB with HnR attempts: 0.3187907064230499
League Average wOBA 2005-2015: .321

wRC+ for AB with HnR attempts: 96.0427907543348
League Average wRC+ is always 100
```

Hypothesis Clause 1: *For the batter, the hit and run play is at least no worse than the standard no-play alternative.*

The wOBA and wRC+ numbers would seem to indicate that the hit and run is just slightly worse than the no-play alternative. We will reject this clause of the hypothesis with the caveat that a hit and run play is very close to the average at bat, but just slightly worse.

Hypothesis Clause 2: *The rate of stolen base success will slightly increase when the batter swings and misses on the hit and run attempt.*

In the average case, the rate of stolen base success decreases significantly when the batter swings. We will again reject this clause of the hypothesis but make an important note that there are a few situations where the swing does increase the success rate.

Overall, I think it is fair to say that in the average case it is probably better to not use the hit and run. However, it is far from a bad play. It should not be the go-to for the average possible hit and run situation, but it can be useful in niche situations. One example of this could be if a struggling hitter is in a 1-2 count with a fast runner on base. There is a fairly small loss in opportunity cost when a poor hitter swings in this count, and the runner picks up about 5% success probability on the steal. If the runner advances in this manner, the play can be considered a success. Further analysis is required to determine in exactly what specific situations a hit and run is beneficial and what situations it is not. However, we that the hit and run has been used in the past to a slight net negative effect when compared to the no-play alternative.

Sources

Retrosheet Info: http://www.retrosheet.org/datause.txt
http://www.retrosheet.org/eventfile.htm

Baseball Prospectus Article: http://www.baseballprospectus.com/article.php?articleid=15713

FanGraphs metric formulas: http://www.fangraphs.com/library/offense/wrc/
http://www.fangraphs.com/library/offense/woba/