Daniel Allen
12/2/14
MCS 100
Leonid Pekelis

The Single to Right Field: Why Left-handed batters may be undervalued one Mike Trout Unit

Abstract

This paper presents a scrappy but (I'd argue) coherent analysis of how a specific play in baseball is unaccounted for in common measures of player value, namely Wins Above Replacement. This inspection of the single with a runner on first (5% of all plays) may point to a market inefficiency in professional baseball. Given that a batter singles with a runner on first, a left-handed batter advances the runner to third base .09 more times per plate appearance than does a right-handed batter. Unless front offices are already accounting for this disparity, a question not answerable in the scope of this paper, this arguably leads to a misuse of over $50 million per year. Be warned that many values are rounded arbitrarily, but insignificantly (I hope), for reading ease and simplified methods.

Background

Baseball has a rich history of statistical player evaluation. Bill James brought the field of advanced baseball statistics, or "sabermetrics", to an untapped community of mathematically-inclined baseball nerds in the late '70s, and Michael Lewis published *Moneyball*, the bestseller that documented the effectiveness of the Oakland Athletics' data-oriented player evaluation model, in 2003. Since James, and even more so since *Moneyball*, MLB teams, fans, and media have grown increasingly confident in and reliant on sabermetrics as indicators of past and predicted success of a player.

One of the most commonly used metrics is Wins Above Replacement (WAR), a single number that encapsulates the total contribution of a player to his team. Fangraphs, an online baseball statistics resource and blog-community, describes WAR as the "answer to the question, 'If this player got injured and their team had to replace them with a freely available minor leaguer or a AAAA player from their bench, how much value would the team be losing?'" The existence of WAR as a statistic means that a

major league team looking to maximize the number of wins in a season can easily approximate the number of expected wins by calculating the sum of their players' WARs and a number close to 52, the number of wins expected of a replacement-level team.

The calculation of the batting portion of WAR has grown fairly standardized among the major outlets that calculate it, namely Baseball Reference, Fangraphs, and Baseball Prospectus. WAR covers nearly every possible effect a position player can have on the game with metrics for batting, baserunning, and fielding. This paper will explore one aspect of baseball not (completely) covered by WAR: how a runner's final position depends on the batter. In the popular calculations of WAR, two players receive the same credit for the same "event" type. David Ortiz, a pull-hitting lefty, gets credited an identical number of runs for a single as does Chris Carter, a pull-hitting righty. A pull hitter is a batter who hits balls especially often to one side of the field, particularly the left side for righties and the right side for lefties. Most batters hit the ball to the natural "pull" side of the field more often than to the "opposite field".

Consider the difference between typical singles hit by Ortiz and Carter. Ortiz may hit the ball to the right fielder, while Carter may hit the ball to the left fielder. No big difference there for Ortiz or Carter. On a single with these hit locations, they both make it to first base safely, by definition of the condition. Now consider the same situation with a quick runner on first. Carter hits the ball to the left fielder, who picks up the ball 90 feet from the third baseman. The runner stops short at second, seeing he won't make it to third base before the right fielder's throw reaches the third baseman. Ortiz, however, hits the ball to the right fielder, and the speedster's third base coach waves him on to third, knowing the throw likely won't get there in time from across the field, and more likely won't hit its target with precision, given the distance.

Win Probability Added and Wins Above Replacement

Win probability added (WPA) measures the increase in the batting team's chances of winning the current game during a specific plate appearance. On March 30, 2008, Nationals Park opened for the Washington baseball team, replacing their 2005 - 2007 venue, a 1961 concrete cookie-cutter monstrosity

called RFK Stadium, with a packed crowd of 39,389 to see the Nationals take on their division foes, the Atlanta Braves. Ryan Zimmerman batted in the bottom of the 9th inning, score tied 2-2, 2 outs, bases empty, and knocked the ball into the left field bleachers, winning the game and setting the Nationals on a three-year streak of last place finishes in the NL East.

How valuable was this at bat to the Washington Nationals? Well, the answer is actually pretty simple. The Win Expectancy for the Nationals when Zimmerman stepped to the plate was .536 wins, meaning the Nationals had a 53.6% chance of winning the game. After his at bat, the Nationals had a 100% chance of winning, translating to exactly 1 win. So Zimmerman's at bat was worth 1 - .536= .464 wins for the Washington Nationals. This is Win Probability Added. It is a historical measure of exactly how many wins each player earned for their team. WPA can be negative, too. Zimmerman had a total WPA of 0.39 on the game because his first three at bats lost the team a 7.4% chance of winning.

However, WPA is not necessarily the most useful measure for judging a player's performance, because it is so affected by the context of each play. Consider, for example, if instead of 2-2, the Nationals had been down 20-2. A Zimmerman home run would increase the Nationals chances from about 0% to about 0%, valuing his home run at 0 wins (rather than .466) for the same performance. It is generally accepted in the sabermetric community that "clutch" hitting, or performing differently in different situations, has a minimal effect on a batter's performance, so a Zimmerman home run with the crowd on their feet is no more impressive than a Zimmerman home run in a blowout. Since there is so much variance in leverage of batting situations, while the significance lies in the player's actual actions, we need a measure that more accurately predicts WPA than WPA itself.

Wins Above Replacement (WAR) is effectively a predictor for WPA. It is a combination of wins added in batting, running, and fielding.

Weighted On-Base Average

The batting portion of WAR, Weighted On-Base Average (wOBA), attempts to measure the average value of all of a batter's plate appearances in expected WPA given an event without context. It is

then scaled to league-level On-Base Percentage (OBP), a more common observational measure with number ranges well understood by the baseball community. For example, the average home run increases the home run-hitting team's chances by 13%, so wOBA assigns each "home run" a weight of .13, which is then scaled among other wOBA's to a mean equal to the mean OBP. wOBA fails to capture the effect of the hit location of a single, which would make sense as if all batter's hit singles to different locations at different rates, but the Methodology section will show that propensity to advance the runner to third base on a single can be regressed on a batter's attributes, namely handedness.

Methodology

I divided my project into four steps:

1. Gather and summarize the data.
2. Calculate the average value added of advancing the runner to third base.
3. Find the players who advanced the runner to third base more often.
4. Calculate the value added of those players.

First, I gathered the data I would need. For Win Expectancy, I used a spreadsheet provided by Fangraphs that has calculated the historical odds of the home team winning for each situation with the following parameters: inning, batting team, score differential, outs, and baserunners. For play-by-play data, I downloaded event files from Retrosheet, from 1920 to 2013.

Next, I gathered some quick statistics to get an intuition for this scenario. I ran a Python script that gathered all instances of singles with a runner on first. This accounted for about 5% of all plays (480,810 in 1920 - 2013 database), Among these plays, here are the resulting base situations.

Frequency of resulting base situations given single with runner on first

Resulting base situation	Frequency
1st and 2nd	59.7%
1st and 3rd	24.5%
1st, 2nd, and 3rd	9.6%
No runners	2.3%
2nd and 3rd	2.1%
2nd	0.63%
1st	0.6%
3rd	0.6%

As we might expect, about 95% of the resulting base situations are either 1st and 2nd, 1st and 3rd, or bases loaded (1st, 2nd, and 3rd). The other 5% can be explained by errors and runners getting thrown out. The runner advanced to second about 69% of the time and to third about 27% of the time.

I then calculated the value added of a single in different relevant situations:

Value of a single

Context	Average WPA
Any single	4.2%
Runner on first	5.9%
No runner on first	3.4%
Runner on first, reaches second	5.1%
Runner on first, reaches third	7.6%

As we see above, the difference in WPA between a single where the runner reaches second and a single where the runner reaches third is about 2.5%, or .025 (1/40) wins.

Now that I had the true value added of the play, I wanted to calculate each player's probability of achieving this advance for each plate appearance. First I calculated the historical averages, that is, each player's number of singles with a runner on first where the runner advanced to third base divided by the

total number of singles where the runner advanced to either second or third base. Note that this is a simplification; it does not including situations where the runner was thrown out or scored.

Most likely batters to advance the runner to third base (1990 - 2013), minimum sample = 200

Batter	Bats	1B makes it to 3B	Batter	Bats	1B makes it to 3B
Jim Edmonds	L	0.481727575	Carlos Beltran	L	0.418269231
Dave Magadan	L	0.477876106	Jeromy Burnitz	L	0.416666667
Bobby Higginson	L	0.459854015	Barry Bonds	L	0.416666667
Mo Vaughn	L	0.453038674	Ken Griffey, Jr.	L	0.413202934
Mark Grace	L	0.444444444	Jim Thome	L	0.411392405
Rafael Palmeiro	L	0.442831216	Todd Walker	L	0.410788382
Todd Helton	L	0.435215947	Fred McGriff	L	0.408730159
Tony Gwynn	L	0.433255269	Wally Joyner	L	0.405797101
Will Clark	L	0.43220339	Larry Walker	L	0.403100775
Carlos Delgado	L	0.431818182	Jason Giambi	L	0.401869159
Rusty Greer	L	0.431095406	Kirby Puckett	L	0.400809717
Harold Baines	L	0.426035503	Cliff Floyd	L	0.399141631
Eddie Murray	L	0.423791822	Garret Anderson	L	0.398268398
Cecil Fielder	L	0.423236515	J.T. Snow	L	0.395348837
Darrin Fletcher	L	0.421940928	Chipper Jones	L	0.394444444

Least likely batters to advance the runner to third base (1990 - 2013), minimum sample = 200

Batter	Bats	1B makes it to 3B	Batter	Bats	1B makes it to 3B
Benito Santiago		0.157738095	Kenny Lofton		0.221288515
Otis Nixon		0.157894737	Gary Gaetti		0.223076923
Reggie Sanders		0.159851301	Rich Aurilia		0.227891156
Luis Castillo		0.172248804	Brian Jordan		0.229607251
Scott Brosius		0.184466019	Nomar Garciaparra		0.231060606
Wil Cordero		0.1886121	Juan Encarnacion		0.237864078
Ed Sprague		0.193069307	Andruw Jones		0.237918216
Bernard Gilkey		0.199095023	Ron Gant		0.238493724
Raul Mondesi		0.202531646	Deivi Cruz		0.240143369
Tony Womack		0.205479452	Magglio Ordonez		0.24137931
Miguel Tejada		0.207746479	Marquis Grissom		0.243735763
Doug Glanville		0.212871287	Javy Lopez		0.244147157
Craig Biggio		0.213483146	Jose Guillen		0.245192308
Greg Vaughn		0.217857143	Glenallen Hill		0.245901639
Jay Buhner		0.219409283	Jeff Cirillo		0.246376812

I saw a lot of lefties at the top of list and righties at the bottom of the list, so that was a good sign. The other thing I noticed was that almost all of the top hitters were power hitters. I realized, then, that many of these batters may have been strategically placed immediately after quick runners in the lineup. So, to account for the help those runners may be giving the hitters, I created a new statistic which measures the difference between the hitter's historical propensity to advance the runner to third and the average historical propensity of his runners to advance to third, regardless of the batter:

Paboveexpected = Phistorical / Pbatterrunner

Here are some players ranked highly and lowly by Paboveexpected:

Batters most likely to perform better than their runners

ID	Difference	Bats
nichb101	1.696157944	L
hende103	1.673987907	L
bouce101	1.658054615	L
maybj101	1.620460525	L
eastl101	1.608784236	L
dalrc101	1.585507802	L
herre102	1.579061713	L
gentj101	1.573663694	L
mincd101	1.546110897	L
sorrp001	1.530867415	L
spenj101	1.524876727	L
may-m001	1.520962885	L
carbb101	1.514755304	L
davia002	1.503651742	L
baile101	1.502511316	L
coviw101	1.499614443	L
heywj001	1.491952523	L
kingj101	1.490503623	L
mathe101	1.480473079	L
epstm101	1.471123665	L

Batters least likely to perform better than their runners

ID	Difference	Bats
tavef101	0.453239899	R
abbok002	0.45687191	R
cotth001	0.507087536	R
sandr002	0.518136783	R
jackd002	0.519160369	R
zimmd101	0.519938503	R
benam001	0.52826527	L
nixoo001	0.531156686	B
sists101	0.538267691	R
wigga001	0.53922461	B
rizzp101	0.560237516	R
byrdm001	0.576355988	R
ceder001	0.576751958	B
paquc001	0.581274123	R
deerr001	0.585084431	R
pedrd001	0.589960541	R
snydc001	0.593332557	R
sodee101	0.596413079	R
duncm001	0.60321151	R
verbe101	0.603495977	R

Summarizing these results, we get that the average left-handed hitter advances the runner to third about .06 more times per plate appearance than an average hitter, and a right-handed hitter advances the runner to third about .03 less per plate appearance than an average hitter. Given that the value of advancing the runner to third is worth .025 wins more than advancing the runner to second, and that a single with a runner on first occurs in 5% of MLB plays, lefties are worth an average of .06*.025*.05 = .000075 wins per plate appearance more than WAR would predict, and righties are worth an average of .

03*.025*.05 = .0000375 wins per plate appearance less than WAR would predict. For a player given 600 plate appearances, typical for a starting position player, .000075 wins per plate appearance amounts to 600*.000075 = .045 extra wins per season, not a whole lot, and .0000375 wins per plate appearance amounts to 600*.0000375 = .0225 fewer wins per season.

Conclusion

The resulting numbers, .045 extra wins for lefties and .0225 fewer wins for righties, do not seem like much, but a small increase in talent goes a long way in a game with as little variance and as much money as baseball. Using the $7 million market value determined by Lewie Pollis of the saber-blog Beyond the Box Score, .045 wins are worth a little over $300,000, and .0225 wins are worth over $150,000. This means a left-handed starting player with a career length of 15 years is underpaid $4.5 million while a right-handed starter with the same career length is overpaid $2.25 million. An approximation for the total amount of money misspent in the MLB due to this discrepancy, with about 180000 plate appearances in 2013, about 33% of whose batters were lefties and 54% were righties, would be (.33*180000*.000075 + .54*180000*.0000375 wins)*$7 million = 8.1*$7million = $56 million, about the value AL MVP and top WAR earner Mike Trout contributed to the Los Angeles Angels of Anaheim in 2014.

Numbers in a more digestible format

Average # wins added by a lefty over the average player, per plate appearance: .000075

Average # wins added by a righty over the average player, per plate appearance: -.0000375

Average # wins added by a switch hitter over the average player, per plate appearance: 0

Value of 1 WAR (per Fangraphs): $7 million

Total absolute Win Probability Added not accounted for by WAR per season: 8.1

1 Mike Trout Unit (MTU): $56 million

Total value not accounted for by WAR per season: $56 million

Sources

Paper:

http://www.beyondtheboxscore.com/2013/10/15/4818740/how-much-does-a-win-really-cost

http://www.fangraphs.com/boxscore.aspx?date=2008-03-30&team=Nationals&dh=0&season=2008

http://www.fangraphs.com/library/misc/war/

Data:

retrosheet.org

ftp://ftp.baseballgraphs.com/wpa