Do Clutch Hitters Exist?

David Grabiner

SABRBoston Presents Sabermetrics

May 20, 2006

http://remarque.org/~grabiner/bosclutch.pdf

(Includes some slides skipped in the original presentation)

Two possible definitions

Do some hitters have an ability to hit better in specific situations defined as clutch (e.g., late and close, runners in scoring position)?

Do some hitters have an ability to contribute more to wins than is indicated by their raw statistics?

The win-expectation definition is more complete because it includes all game-clutch and run-clutch situations; if there is an ability, the win-expectation ability should be larger but may be harder to measure.

Three types of answers

There are players with an ability to perform X better in the clutch. (Statistically: X is the standard deviation of ability.)

The players who have performed best in the clutch in the past can be expected to have an ability of X. (Statistically: there is a correlation of Y between past and current clutch performance.)

We can identify players based on factors other than clutch performance who have an ability of X. (Statistically: there is a correlation of Y between skill A and clutch performance.)

Reasons for clutch ability

Ability to perform better under pressure (This is what fans think of as clutch hitting, but we can't measure it independently.)

Adjustment to the game situation (Example: pitchers throw more balls with first base open.)

How the player can be used (Example: good lefty hitters face one-out lefty pitchers with the game on the line.)

How the player was actually used (Example: cleanup hitters bat with more men on base.)

How to measure clutch performance?

With a situational definition, take your favorite statistic, measure it in clutch and non-clutch situations for each player, and the difference is clutch performance. (Average clutch performance will usually be negative because the pitchers in clutch situations are better than average.)

Note that clutch performance is defined as the *difference* between ability in clutch and non-clutch situations. A good hitter whose ability is independent of the game situation will have good clutch statistics but will appear as only average in clutch performance in these studies.

To measure clutch performance with a value-added model, we need to compute expected wins both from the raw statistics and from the situational data.

Given a complete offensive statistic (such as Linear Weights), we can estimate the runs above average added by a hitter, and given the relation between wins and runs, we can estimate the wins above average added by each run. A walk is .33 LWTS runs, and if ten runs is an additional win, it is .033 expected wins.

Given a model for the game, we can compute expected win probability in every situation, and measure the change caused by the hitter. For example, leading off the top of the eighth in a tie game, the win probability is .500. If the batter walks, the win probability becomes .567 [Tango/Lichtman/Dolphin], so the walk was worth .067 wins.

The difference between these two measures is clutch performance; if a player produces 3 wins from LWTS but 4 wins in expected win probability, he produced +1 clutch win. (We can adjust for leverage; a good hitter who bats in high-leverage situations will produce more wins than expected.)

The studies

We will look at many different studies. As the data has improved (thanks to Retrosheet), it has become possible to measure smaller effects, which has allowed several recent studies to conclude that an ability exists. And the most recent study, by Nate Silver in *Baseball Between the Numbers*, shows a strong enough ability to be important, and to suggest other studies looking for the causes.

Cramer, 1977 *Baseball Research Journal*

Mills and Mills, *Player Win Averages* used play-by-play data to calculate expected wins and losses in 1969 (and unpublished 1970), and defined

$$\mathrm{PWA} = \frac{\text{wins added}}{\text{wins added} + \text{losses added}}$$

This automatically adjusts for leverage.

The study computed a linear regression between BWA (essentially OBP$\times$SLG) and PWA; clutch performance is the difference between predicted and actual PWA. There was no correlation between 1969 and 1970 clutch performance.

This study suffered from a lack of data; if the ability is as strong as Silver's study suggests, two seasons would show a correlation of $.05 \pm .10$.

Elias Baseball Analyst, yearly

Elias defined late-inning pressure situations (LIPS) as situations in the seventh inning or later with the score tied, or the batting team down by three runs or less, or by four runs with the bases loaded. Clutch performance was measured by the difference between LIPS and overall batting average.

A study was presented every year, but with the methodology always changing, and never a mention of statistical significance. The best *Analyst*, in 1988, probably had the best study, but the observed splits were not statistically significant; there was a 6% probability that a split that large could occur by chance.

G., 1991, `rec.sport.baseball`

This was the best I could do with the Elias data. Elias listed the players who gained or lost the most BA in LIPS in 1979-1988 or 1985-1989, and I checked the difference in subsequent years through 1991, weighted by AB. 83 player seasons were included.

Clutch terrors lost .005 in LIPS.
Average hitters lost .008 in LIPS.
Chokers lost .024 in LIPS.

The difference, $.019 \pm .013$, is not statistically significant, but if it is accurate, a top clutch hitter in the past will get one extra clutch hit per year.

G., 1993, `rec.sport.baseball`

The *Great American Baseball Stat Book* defined late and close as the seventh inning or later, with the batting team up by one, the score tied, or the tying run on base, at bat, or on deck. Clutch performance was measured by the difference between OPS in late-and-close and other situations. I correlated clutch performance over a player's previous career (since 1984) with his performance in the current year, taking a correlation weighted by AB over all regulars with at least 250 late-and-close AB. 245 player seasons were included.

Correlation: $.01 \pm .07$.

It would be impossible to identify any clutch hitters by past performance if the true correlation were that low; an ability with standard deviation of 20 OPS points (one clutch hit a year) would account for a correlation of .08, and even that ability would only allow us to identify clutch hitters with a standard deviation of 8 OPS points from their past clutch data.

Looking for other causes

This study also led me to look for other causes of consistent clutch performance. Three of the worst chokers were Boggs, McGriff, and Strawberry, three great left-handed hitters with large platoon splits. If such hitters had the platoon edge 10% less than normal in late-and-close situations (because lefty relievers were brought in to face them), they would lose 15 OPS points to that alone. I didn't have enough data to draw a statistically significant conclusion.

Dolphin, 2003, *Baseball Primer*, now at
`www.dolphinsim.com/ratings/notes/clutch.html`

Dolphin defined a clutch situation as the sixth inning or later, with the score tied, or the tying run on base, at bat, or on deck, and measured clutch performance by OBP. Rather than computing a correlation, he computed the observed standard deviation of clutch performance, and compared it to what would be expected if clutch hitting were completely random. The sample was all players with 1000 clutch PA in 1972-1992 (and some earlier Retrosheet data), a total of 612 players.

Probability of observed bias by chance: .009. Best estimate of clutch ability: .007 in OBP.

The difference of .007 is one extra time on base a year. In a full career (5000 PA), the linear regression between career performance and ability would be .20; that is, a player whose performance was .030 better than average (gaining .021 rather than losing .009) could be expected to have an ability of .006. A difference of .030 is two standard deviations.

Dolphin also noted that half the variance in clutch ability was explained by a negative correlation with SLG. (And it wasn't just clutch OBP; similar studies of other statistics showed the same bias.) A player with SLG .050 better than the league average could be expected to perform .005 worse in clutch OBP than the average.

This is the opposite of what would be expected if sluggers drew semi-intentional walks, but it happened, and that difference is statistically significant and deserves further investigation.

Tango/Lichtman/Dolphin, 2005, *The Book: Playing the Percentages in Baseball*

This study defined clutch as the eighth inning or later, with the batting team down by at most three runs. It used only plate appearances against right-handed pitchers, to eliminate platoon effects. Clutch performance was measured by OBP, and by wOBA, which is essentially 1.15×LWTS/PA. The details of the methodology are not given. The sample was all players with 100 clutch and 400 non-clutch PA in 1960-1992, a total of 848 players.

Standard deviation of OBP ability: $.008 \pm .004$.
Standard deviation of wOBA ability: $.006 \pm .004$.

This is not statistically significant, and if this is the magnitude of the ability, the correlation between career clutch performance and ability is .02.

Silver, 2006, *Baseball Between the Numbers*

This is a value-added study. Win expectations were computed using Keith Woolner's model from *Baseball Prospectus 2005*. Run expectations were computed from Marginal Lineup Value (team Runs Created, minus team Runs Created with an average player replacing the player). Runs were converted to wins by the Pythagorean formula for the team, then multiplied by the average leverage faced by the batter (defined as the relative importance of one run towards winning). Clutch performance was measured by the difference between expected and actual wins added.

The test was the correlation between odd-year and even-year clutch performance; this gives a larger sample size than single-year data. The sample was all players with 2500 odd-year and 2500 even-year PA since 1972, a total of 292 players.

Correlation: $.33 \pm .04$.

Wow!

In order to account for a correlation of .33, the standard deviation of ability must be .58 times the standard deviation of performance in a half-career. In a full career, it would be .71.

The standard deviation of career clutch performance is half a win per year, and thus the standard deviation of clutch ability is .35 wins per year. The regression coefficient between career clutch performance and ability is 0.5!

In other words, the three top active clutch hitters, Matt Lawton (+1.42 wins per year), Jason Kendall (+1.41), and Jeromy Burnitz (+1.17), can be expected to be worth more than half a win per year in clutch ability, based solely on past performance.

Where does that ability come from?

This is the only study with a significant sample which deals with run-clutch effects, and thus suggests that run-clutch effects are more significant than game-clutch effects.

The .35 wins is also probably a sum of several smaller effects; the game-clutch effects that other studies have estimated are just one part.

Silver conjectures that much of the ability is due to players adjusting to the game situation. There is a .29 correlation between $(2 \times \mathrm{BB} - \mathrm{K})/\mathrm{PA}$ and clutch performance; strikeouts and walks certainly indicate the ability to adjust to different pitches. This correlation explains 1/6 of the variance in the ability.

Another hypothesis: the lineup

Many players are used consistently in the same lineup spots. The run value of events depends on the base-out situation, and the base-out situations are not evenly distributed over the lineup. Leadoff hitters bat with very few runners on, third-place hitters bat more often with two out, and cleanup hitters bat with a lot of runners on.

Lineup details

[Tango/Lichtman/Dolphin] Here is an abbreviated table from *The Book*, giving the run expectation of various events by batting-order position. This is based on AL data; the NL figures, with the pitcher batting, might show a more significant effect because the leadoff hitter is even more likely to bat following an out. (The tables show spots 1–5 because the bottom of the lineup shows little variation.)

Position	1B	HR	BB	Out
1	.468	1.291	.350	−.298
2	.479	1.349	.340	−.301
3	.469	1.384	.319	−.300
4	.504	1.436	.337	−.320
5	.513	1.438	.348	−.323
All	.487	1.388	.339	−.309

Normalized to −.309 runs per out, to adjust for leverage:

Position	1B	HR	BB
1	.486	1.339	.362
2	.492	1.385	.349
3	.483	1.426	.329
4	.487	1.387	.325
5	.490	1.375	.332
All	.487	1.388	.339

The value of a single doesn't change much, but homers and walks do.

Leverage and clutch performance

Hitters in the 1–3 spots have 3% less run-clutch leverage than average, and in the 4–5 spots have 3% more; note the costs of outs. (Good hitters can still bat at the top of the lineup; a leadoff hitter gets 10% more plate appearances than average, so even if he loses 3% of the value of each plate appearance, it's still valuable to put a good hitter there.)

Silver's study adjusted for game-clutch leverage, but not for run-clutch leverage. Players often bat in the same lineup spot for most of their careers, so the run-clutch leverage will be consistent. A great hitter who produces 30 LWTS runs a year will lose one run a year to leverage if he bats leadoff, and gain one run a year if he bats cleanup.

Situational lineup effects

After leverage adjustments, a home run for the leadoff hitter costs .049 runs compared to its average value, and a walk gains .023 runs. A leadoff hitter who hits 10 fewer homers and draws 40 more walks than the average hitter gains a run and a half. Similarly, a third-place hitter who hits 20 more homers and draws 20 fewer walks than average gains more than a run.

Both effects will be consistent if the player stays in the same lineup spot for his whole career. Richie Ashburn was probably a good clutch hitter; in his non-DH league, he may have gained two runs a year just by being the leadoff batter.

These effects of 1–2 runs aren't much, but if the standard deviation of clutch ability is .35 wins, that's 3.5 runs, so the lineup effects may be significant. These are the type of effects we can look for in follow-up studies.

What now?

Since we have some idea of what to look for, we could now study various effects on run-clutch hitting. We can look at the expected runs, rather than wins, for all the situations in which a player bats, and see which hitters add more expected runs than they would if their performance in all situations were random.

How much of the clutch effect is due to player usage patterns, such as platooning and pinch-hitting? In which situations do which regular players face more or fewer pitchers with the platoon edge?

Are there clutch pitchers, particularly run-clutch? Pitchers make larger adjustments than hitters to the game situation, so we would expect an ability. In addition, there are physical differences; Nolan Ryan was known for pitching worse from the stretch than from the wind-up.