

## N = 1

By Russell A. Carleton

Pick a player. What do you know about him? No, I don't mean the lovely human interest stories about the quirky hobbies or his tastes in steak or the story about that one time at band camp. What do you know about what makes him tick as a baseball player? How does he make decisions? Are there patterns that he falls into? Are there weaknesses that can be exploited in his game?

Well?

When you have a hammer, everything looks like a nail, and when you have a big data set, everything looks like a question that can be answered with a large N database query. Sure, there are lots of nails out there that have been pounded down and are still left to pound down, but for those of us who research baseball, I worry that we've grown a little too attached to our hammers. Just about every piece of Sabermetric research contains some variation of the line "I looked for *all* players who..." When you want to tell a story about all players, that's a perfectly reasonable place to start. What's concerning is that they all seem to start there, particularly when it doesn't have to.

Often, I'll be listening to interviews with people who know the game well, and the interviewer will ask a question that starts with "When is the proper time to..." or "How can you tell if a player is..." The answer usually starts with some form of "Well, it depends a lot on the player..." No one thinks this is a strange answer. Yes, sometimes saying "It depends..." is a polite way of saying "I have no idea," but there's a general understanding that players differ from one another. Maybe they differ in ways that are complex and hard to pin down in regressions, but they differ. The folks who research baseball, myself included, will acknowledge this, but then we rarely adjust our research methods to address the issue. Figuring out the best pitch to throw early in the count and the best pitch to throw *to Smith* early in the count are two separate questions, and knowing the answer to the first may not help you very much with the second. Not only that, but there are situations when the answer to the second question would be much more important.

Like when Smith is in the batter's box.

Large N database queries tell you how "baseball players" or some sub-group of them (pitchers, hitters, guys with more than seven letters in their last names) perform. They can reveal larger truths about how human beings—or at least baseball players—behave given a certain set of conditions. Left-handed hitters, as a group, fare worse against left-handed pitching than right-handed pitching. There's a certain geometry to the game, and the angle that a left-handed pitcher throws from is harder to pick up than that of a right-handed pitcher. Looking at all players as a collective tells us something bigger about the game of baseball. Some guys have bigger platoon splits than others, but the platoon effect usually (but, not always!) points the same way. Large N works great in situations like

that.

Let's play with the possibilities a bit. What happens when there is some variable that makes some hitters better, some worse, and for some it just doesn't have an effect. In a large N database model, a standard ordinary least squares (OLS) regression might find that there's no overall effect for that variable, because the good and bad wash each other out. A researcher might come across that regression and think that there's nothing to be gained from looking at the matter further. In doing so, he's missed a variable that moves the needle quite a bit, just not in a consistent manner across all players. Large N models are good at discovering large, aggregate truths in baseball (and the rest of life), and those can be good to know about. But when it comes to setting a strategy, teams do not send an amalgamation of all MLB hitters to the plate. They send a guy with a specific last name stitched onto the back of his jersey. Is he one of the hitters who tends to get better, worse, or stay the same in this situation? Large N won't help us here.

What if, instead of running an analysis with all players, we limited ourselves to just looking at plate appearances featuring only David Ortiz or just John Jay or strictly Nick Swisher. Sure, the data set would be smaller, but suddenly, a lot of those issues about players differing significantly from one another disappear. Big data isn't always the best answer. Sometimes, smaller, more personal data is better. Interestingly enough, this sort of design isn't completely foreign to baseball analysis. Announcers will often quote a player's splits (e.g., how he fared against lefties vs. righties) in describing a player or in discussing whether the manager is either brilliant or silly to bring in a lefty to face him. The problem with a lot of splits is that for a hitter who has 600 plate appearances, the announcer might be quoting his on-base percentage against righties based on a 350 PA sample, and against lefties based on 250 PA. I've done research on how many plate appearances it takes before a statistic becomes a reliable estimate of true talent. OBP, for example, takes about 460 PA before it reaches a point of reliability. In some sense, the announcer is comparing one unreliably small sample to another. But he is comparing one man against himself. While there may be concerns about measure reliability, there's little concern about the two samples being biased in some hidden way based on the people in each group. It's the same guy 600 times over.

But suppose that we could have the best of both worlds: the purity of a sample composed of a single player and enough statistical power that we could confidently draw conclusions. And on top of it, we could actually do something useful with it. A moment ago, I mentioned research that I had done on when various stats become reliable enough to be considered a true reflection of a hitter's talent level. For big-ticket numbers such as OBP or SLG, the answer is that it can take a few hundred plate appearances. The problem with OBP is that while the batter certainly has a say in whether he can draw a walk or hit the ball hard, there are other factors out there that might get in the way. Sometimes a player will hit the ball hard, but right at someone. Sometimes the defender makes a fantastic diving catch leaving the batter with a really loud out. When we come down to things that are much more within a batter's control, such as the decision of whether or not to swing, things are easier to estimate. For example, I've found that an estimate of a

player's swing rate can become reliable within about 50 PA.

But what might affect a hitter's decision of whether or not to swing? Well, maybe he's feeling a little jumpy because the game is on the line. He might be the kind of guy who gets nervous and starts hacking. He might be the kind of guy who gets nervous and gets super cautious because it really matters here. He might be the kind of guy for whom it doesn't matter. As the opposing pitcher (or manager), I might like to know that information. If I know a hitter won't swing at all, I can groove one down the middle for a strike. If I know for sure that he will swing, I can bounce a curve two feet in front of home plate and catch a nice breeze as the batter flails away.

### **Warning! Gory Mathematical Details Ahead!**

Let's put together some data sets. Plural. I selected all players (see, I just did it...) who had more than 500 plate appearances (excluding intentional walks) in 2013. There were 139 such players. For each player, I created a data set of what he did in each plate appearance on the first pitch. Using just pitches on an 0-0 count controls for any count-related tendencies to swing (e.g., taking all the way on 3-0, needing to expand the strike zone with 2 strikes, or the fact that some guys have certain pitches that they like to swing on). I coded the pitch as either a swing (foul ball, ball in play, or swinging strike) or a take (ball, called strike). Since we know we have at least 500 observations per player, that should be plenty to power a regression.

(Side note: For the super-initiated, you are probably already mouthing abbreviations like GLM or HLM or MLM. Yes, we could go there, but not today.)

I used game leverage as a proxy for the importance of the situation. Leverage is a concept developed by researcher Tom Tango. It assigns a scalar value (1 is average, 2 is twice as important as the average situation, 0.5 is half as important) to each situation in a game, based on the inning, score, number of outs, and placement of any runners. It measures how much the probabilities of each team winning might shift as a result of what happens in this next at bat. Not surprisingly, situations that take place in the late innings of close games, particularly with runners on, have higher leverage values. Situations in 15-3 games have low leverage values because we already know who's going to win the game.

For each player, I ran a binary logistic regression (swing vs. no swing) as predicted by the leverage of the plate appearance. Each player gets his own regression, so it's the same overall tendency to swing on all the data points. The regression just tells us which way to tilt it for leverage. The results had a few surprises in them. Most players became more patient as the game becomes more tense, but not everyone. In fact, of the 139 players, 37 showed a significant tendency toward swinging more often in higher leverage situations, relative to how often they normally swing. No one had a statistically significant effect in the other direction (although a couple of hitters came close).

Below, we see the more extreme cases from 2013, again with a minimum of 500 PA, and in parentheses, their expected swing rate at a leverage index of 1 (a situation of average importance) and a leverage index of 2 (a situation twice as important as average).

<b>Top 5 Hitters Who Become More Likely to Swing in High Leverage</b>	<b>Top 5 Hitters Who Become Less Likely to Swing in High Leverage</b>	<b>Top 5 Hitters Who Were Most Unaffected by Leverage</b>
Justin Upton (32%, 48%)	Nate Schierholtz (28%, 23%)	Mitch Moreland (14%, 14%)
Kyle Seager (23%, 38%)	Nolan Arenado (30%, 26%)	Adam Laroche (24%, 24%)
Matt Dominguez (23%, 37%)	Dan Uggla (32%, 28%)	Ben Zobrist (26%, 26%)
Adam Jones (42%, 55%)	Carlos Gomez (52%, 48%)	Jonathan Lucroy (14%, 14%)
Will Venable (29%, 42%)	Prince Fielder (29%, 25%)	Josh Hamilton (41%, 41%)

Justin Upton gets jumpy when faced with the first pitch of a big at bat. Nate Schierholtz gets more passive. Mitch Moreland apparently doesn't notice. But now, when setting a game plan against Justin Upton, a team has a little bit of extra information about how he changes with the game.

To make sure that these weren't just random fluctuations, I found all players who had at least 500 PA in 2012 and ran the same analyses. There were 87 players in both the 2012 and 2013 data sets. I calculated the predicted swing rates for both seasons in all 87 cases at a leverage value of 0.5, 1.0, and 2.0. Across those 87 players, the rates from 2012 and 2013 correlated at .89 for a leverage of 0.5, .88 for a leverage of 1.0, and .71 for a leverage of 2.0. These effects are rather consistent from year to year.

### **Little Data, Big Possibilities**

We've just discovered some interesting things about some players. The kicker is that we only looked at one influencing variable (leverage) and its effect on one decision (whether or not to swing on a 0-0 pitch). Imagine if we looked at other variables and outcomes using this same  $N = 1$  paradigm. Do certain hitters change their approach when facing certain types of pitchers? Do these same basic patterns hold for other counts? Is contact rate affected? Are there certain players who seem to show the same patterns across several of these types of analyses? Now, we can start to build a profile of a player, not just based on his splits, but on how the situation around him affects him. Our analyses will only be applicable to a single hitter (or pitcher... or manager) but they hold the potential to reveal a greater truth about him. Call it Saber-scouting, if you like.

I prefer to think of it in a different way. Maybe it's because of the emphasis on large  $N$  research that's been so prevalent across the discipline of Sabermetrics for so long, but the

field has marginalized these types of individual differences. Maybe it's the fact that a lot of what Sabermetrics has set out to combat has been narrative-driven pseudo-research that relied on "just trust me" or sample sizes that are laughably small to make a point about a player. But we've come to ignore, or worse, dismiss the thought that players might react to situations in different ways. The point of Sabermetrics shouldn't be to destroy anything which smells of narrative, but to promote good, solid research methods in the study of baseball. There are perfectly good, methodologically sound ways to look at individual players and to find interesting things about them. This sort of research is something that Sabermetrics should already naturally be doing.

I'm often asked what I think "the next big thing" in Sabermetrics will be. It's an awful question, because there are about ten "next big things" and if I pick one, it makes me look as though I'm denying the power of the other nine. I think people who ask this are expecting some sort of major strategic response ("Look for guys with high OBPs, because they'll be underpriced!") I often answer that the next big thing is understanding how each player works, and instead of strategies that can be applied across entire organizations or in every situation, understanding the nuances of each player on a case-by-case basis. The applications from the team's point of view are obvious, and for those who wish to write publicly, it makes for some great long-form profile work analyzing a single player. So, if you're looking for a way to make your mark, maybe the best thing to do would be to brush up on some good ideographic research methodology, and shorten your data set to the smallest sample of all.  $N = 1$ .