Sabermetrics: The Past, the Present, and the Future

Jim Albert

February 12, 2010

Abstract

This article provides an overview of sabermetrics, the science of learning about baseball through objective evidence. Statistics and baseball have always had a strong kinship, as many famous players are known by their famous statistical accomplishments such as Joe Dimaggio's 56-game hitting streak and Ted Williams' .406 batting average in the 1941 baseball season. We give an overview of how one measures performance in batting, pitching, and fielding. In baseball, the traditional measures are batting average, slugging percentage, and on-base percentage, but modern measures such as OPS (on-base percentage plus slugging percentage) are better in predicting the number of runs a team will score in a game. Pitching is a harder aspect of performance to measure, since traditional measures such as winning percentage and earned run average are confounded by the abilities of the pitcher teammates. Modern measures of pitching such as DIPS (defense independent pitching statistics) are helpful in isolating the contributions of a pitcher that do not involve his teammates. It is also challenging to measure the quality of a player's fielding ability, since the standard measure of fielding, the fielding percentage, is not helpful in understanding the range of a player in moving towards a batted ball. New measures of fielding have been developed that are useful in measuring a player's fielding range. Major League Baseball is measuring the game in new ways, and sabermetrics is using this new data to find better measures of player performance. The trajectory and speed for all pitches are currently being measured using the Pitch F/X system, and this article demonstrates how this new data can be used to measure the quality of a pitcher's fastball. New data measuring the location of all batted balls and the fielders will lead to further improvements in measures of player performance which will help teams better understand player values.

1

Introduction

Baseball fans have always had a love for statistics. Even when professional baseball began in 1876, counts of the basic statistics such as hits, doubles, triples, home runs, walks, strikeouts, and runs were recorded. Pitchers and hitters have always been ranked with respect to measures such as the batting average, the number of home runs, and the average runs allowed. Currently, one of the most prestigious achievements for hitting is the Triple Crown, when a player simultaneously has the highest batting average, slugging percentage, and number of home runs. (The last player to obtain the Triple Crown was Carl Yastrzemski in 1967.)

Sabermetrics is the science of learning about baseball through objective evidence. Sabermetrics poses questions such as "How many home runs will Albert Pujols hit next year?", "Is it easier to hit home runs in particular ballparks?", "Are particular players especially good in clutch situations?" and collects and summarizes relevant data to answer them.

One basic problem in sabermetrics is evaluating the performance of batters, pitchers, and fielders. We give an overview of the traditional measures for evaluating players, describe some of the current measures that have been developed, and look forward to new evaluation methods based on new types of data collection.

Measuring Batting

The traditional measure of batting performance is the batting average AVG that is computed by dividing the number of hits H by the number of at-bats AB:

$$AVG = \frac{H}{AB}.$$

The player with the highest batting average is recognized as the batting champion. But this is a flawed measure of batting performance for several reasons. First, the batting average ignores other ways for a player to reach base such as getting a walk or being hit by a pitch. An alternative statistic that measures a player's ability to reach on-base is the on-base percentage (OBP) that divides the total number of on-base events (hits, walks, and hit-by-pitches) by the number of plate appearances:

$$OBP = \frac{H + BB + HBP}{AB + BB + HBP + SF}.$$

Another flaw of the batting average is it gives all hits the same value and doesn't distinguish between singles, doubles, triples, and home runs. An alternative measure that distinguishes the different hit values is the slugging percentage (SLG) that computes the average number of bases reached for each at-bat:

$$SLG = \frac{1B + 2 \times 2B + 3 \times 3B + 4 \times HR}{AB},$$

where $1B$, $2B$, $3B$, and HR denote respectively the count of singles, doubles, triples and home runs.

There are two important skills for a batter. He wishes to get on base and, when there are already runners on base, he wishes to advance them home. The on-base percentage measures the effectiveness of a batter in reaching base and the slugging percentage is useful in measuring the batter's skill in advancing runners. One way of measuring the combined skill of the batter to get on base and advance runners adds the on-base percentage to the slugging percentage, creating the modern statistic OPS:

$$OPS = OBP + SLG.$$

Four ways of measuring hitting have been presented: the traditional batting average AVG, the on-base percentage OBP, the slugging percentage SLG, and the combined statistic OPS. Which is the best measure? To answer this question, we have to realize that the goal of batting is to score runs, and runs are scored by teams and not individuals. So it is necessary to look at team hitting data to evaluate the worth of different measures.

We can measure the ability of a team to score by R/G, the average number of runs scored per game. For each team in the 2008 season, we collect R/G, the team batting average, the team on-base percentage, the team slugging percentage, and the team OPS. Figure 1 displays a scatterplot of batting average against R/G. As expected, there is a positive trend in this graph – teams with higher batting averages tend to score more runs. But the points are widely scattered which indicates that AVG is not a good predictor of runs scored. Only 46 percent of the total variation in runs scored can be explained by batting average, so 54 percent of the variability in runs scored is due to other differences between the thirty teams.

Figure 2 shows a scatterplot of a good hitting measure OPS and R/G. Here we see a strong positive trend in the graph indicating that this measure that combines the on-base percentage and the slugging percentage is a good predictor of runs scored. In fact, 89 percent of the total variation in runs scored can be explained by the differences in OPS.

Major League Baseball and the baseball media can be slow to adopt new measures of performance. But OPS is one measure that has become popular and is useful in comparing players. For example, Ichiro Suzuki hit for a batting average of .351 in the 2007 season. It appears that he was a much better hitter than his teammate Raul Ibanez, who only had a .291 batting average. But if we compare their OPS (Table 1), we see that although Suzuki was more successful in getting on base, Ibanez was more successful in advancing runners. The two players had similar values of OPS in the 2007 season, which means they created the same number of runs for their teams. An OPS of 1.000 is a standard of great performance. Only nine players in baseball history have had a career OPS exceeding 1.000. Babe Ruth is arguably the greatest hitter in baseball history with the highest career OPS value of 1.164.
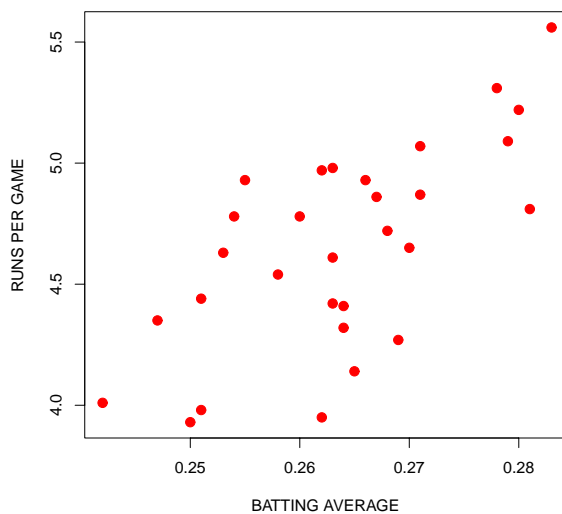
Figure 1: Scatterplot of batting average and runs scored per game for all teams in the 2008 baseball season. There is a relatively weak positive association pattern in the scatterplot, indicating that batting average is a weak predictor of runs scored.

Name	AVG	OBP	SLG	$OPS = AVG + SLG$
Ichiro Suzuki	0.351	0.396	0.431	0.827
Raul Ibanez	0.291	0.351	0.480	0.831

Measuring Pitching

How does Major League Baseball currently evaluate pitchers? A traditional measure of pitching is the pitcher's win/loss percentage

$$WIN\% = \frac{W}{W + L},$$

where W and L are respectively the number of team wins and team losses credited to the pitcher. Another traditional measure is the earned run average (ERA), which is the average number of earned runs allowed by the pitcher in 9 innings:

$$ERA = 9 \times \frac{Earned\,Runs}{Innings\,Pitched}.$$

Both measures have some problems. A pitcher can have a high winning percentage not because he is a great player, but because his team tends to score many
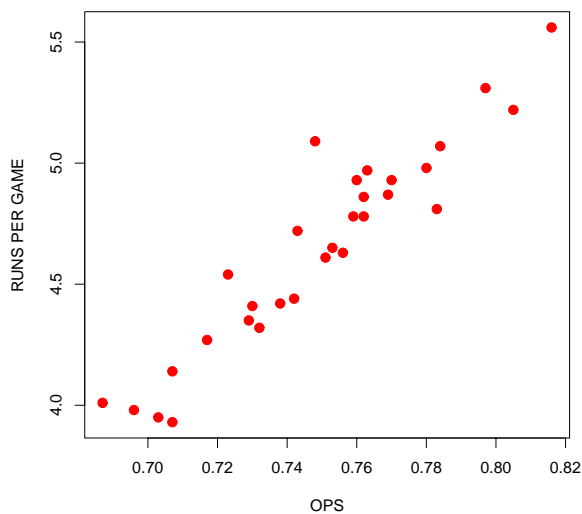
Figure 2: Scatterplot of the OPS measure and runs scored per game for all teams in the 2008 baseball season. There is a strong positive association pattern in the graph, indicating that OPS is a good predictor of runs scored.

runs when he is pitching. The earned run average seems like a good measure – after all, a pitcher's objective is prevent runs scored and the ERA gives the number of runs allowed per game. But there are two problems with an ERA. Teams prevent runs by good pitching and good fielding and the ERA reflects the combined effort of the pitcher and his teammates. The ERA does not separate the ability of the pitcher from the ability of the fielders. Another problem with the ERA is there is a chance element in preventing runs. A pitcher cannot control the path of a ball that is placed in-play, a ball hit in the fair territory of the baseball field. So a pitcher may have a high ERA not because he is a poor pitcher, but because many balls happened by chance to fall as base hits.

One way of demonstrating that a particular hitting or pitching statistic is a good measure of ability is to compare its value in one season with its value of the statistic in the following season. If there is a strong relationship between the two statistics, this particular measure is a good measure of the player's ability and we can it to make a good prediction of its value for the following season.

Is the earned run average a good measure of a pitcher's ability? For all the pitchers who started at least 25 games in both the 2007 and 2008 baseball seasons, Figure 3 shows a scatterplot of the 2007 ERA and the 2008 ERA. Although there is a positive trend in the scatterplot, only 9 percent in the variability in the 2008 ERA values can be explained by the variability in the 2007 ERA values. This is a remarkable finding. It indicates that a pitcher's

5

ERA is controlled largely by factors such as the team's fielding and lucky balls that fall for base hits that are outside the control of the pitcher.
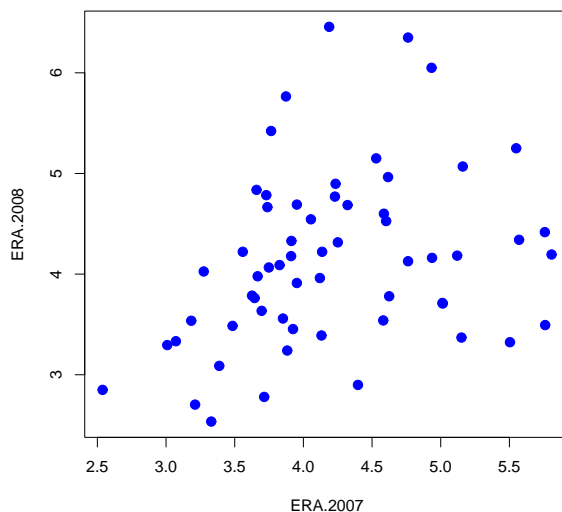

Figure 3: Scatterplot of the 2007 ERA and the 2008 ERA for all pitchers who started at least 25 games in both seasons. There is a weak association pattern in the graph, indicating that the ERA is a weak measure of a pitcher's ability.

Sabermetricians have searched for alternative measures of pitching performance that are independent of the contributions made by fielders. The pitcher's ability to strike out batters does not depend on his teammates, and so a pitcher's strikeout rate:

$$SO/9 = 9 \times \frac{SO}{IP},$$

where SO and IP are respectively the counts of strikeouts and innings pitched may be a better measure of a pitcher's ability. Figure 4 displays a scatterplot of the 2007 strikeout rate and the 2008 strikeout rate for our collection of starting pitchers. We see a strong relationship in the graph that indicates that strikeout rate is a better good measure of a pitcher's ability than ERA; 69% of the variability of the 2008 strikeout rates can be explained by the 2007 strikeout rates.

A pitcher's strikeout rate is an example of a *defense independent pitching statistic* or $DIPS$, a measure of pitching performance that does not involve the fielders. Other examples of $DIPS$ are home runs allowed (HR), hit batters (HBP), and walks (BB). One pitching measure based only on these statistics
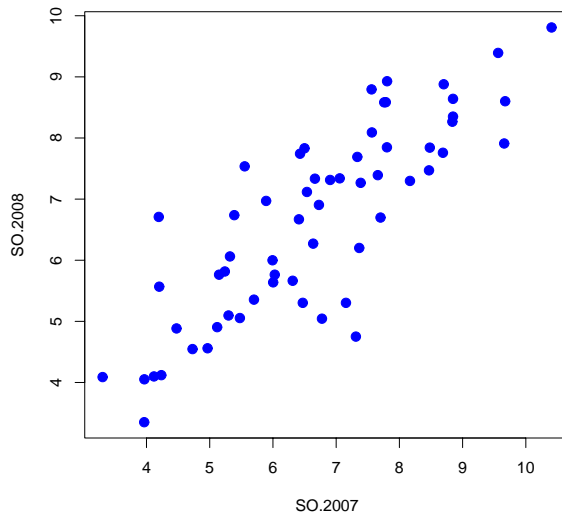
Figure 4: Scatterplot of the 2007 strikeout rate and the 2008 strikeout rate for all pitchers who started at least 25 games in both seasons. There is relatively strong association pattern in the graph, indicating that the strikeout rate is a good measure of a pitcher's ability.

is called the defense-independent component ERA or $DICE$:

$$DICE = 3.00 + \frac{13 \times HR + 3(BB + HBP) - 2 \times SO}{IP}.$$

Figure 5 shows a scatterplot of a pitcher's 2007 $DICE$ statistic with his 2008 ERA. There is a stronger relationship between $DICE$ and the 2008 ERA than the 2007 and 2008 ERAs (Figure 3). So, the defense-independent component measure $DICE$ is a better predictor of the following year's ERA than the present year's ERA. It can be shown that 15 percent of the variability in the 2008 ERA can be explained by the 2007 $DICE$ measure.

Measuring Fielding

Sabermetricians are also interested in finding good measures of fielding performance. The basic fielding measures are the counts of put-outs (PO), assists (A), and errors (E). The traditional measure of fielding ability is the fielding percentage $FLD\%$
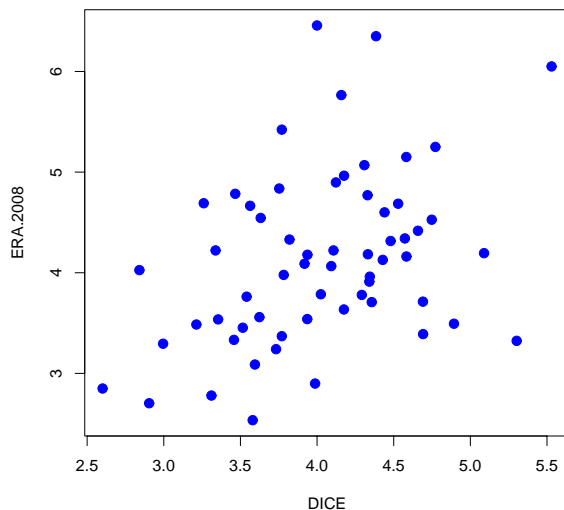
$$FLD\% = \frac{PO + A}{PO + A + E}.$$

Figure 5: Scatterplot of the 2007 dependent independent pitching measure $DICE$ and the 2008 ERA for all pitchers who started at least 25 games in both seasons. The 2007 $DICE$ measure is slightly better than the 2007 ERA in predicting the following season's ERA for a pitcher.

This statistic measures the proportion of fielding plays that were made success-fully. However, this measure ignores balls outside of the reach of the fielder where a play was not made. It is important for a fielder to have good *range*, that is, the ability to move a long distance for the ball that is placed in-play, and the fielding percentage says nothing about a fielder's range.

A simple way of measuring a fielder's range is to consider the count of plays that are made. Fielders with more range tend to be involved in more plays. One range measure is the range factor per innings $RF/9$, which is the number of plays (assists and putouts) made for each nine innings:

$$RF/9 = 9 \times \frac{PO + A}{IP}.$$

Let us use these two fielding measures to compare two modern great short-stops Omar Visquel and Derek Jeter. Both have been awarded the Gold Glove award for fielding excellence as a shortstop; Visquel won the award in the nine seasons 1993 to 2001 and Jeter won the award for the three seasons 2004 through 2006. Figure 6 displays the fielding percentages of both players plotted as a function of their age. Smoothing curves help us see the general patterns in the graph. The fielding percentages increase from left to right, which means that both fielders are more successful in making plays as they mature. Visquel's

8

fielding percentages are substantially higher than Jeter, indicating that Visquel is more successful (by about one percent) in making plays. Figure 7 shows the range factors for both players plotted against age. A player's range tends to decrease with age, so older players are less able to reach balls that are not hit directly to them. Also note that Visquel displays a much greater range than Jeter. Generally, Visquel is able to reach about a half additional play than Jeter for each nine innings. This is a significant difference. Visquel is able to prevent more runs scored for his team than Jeter. In fact, Omar Visquel is generally considered to be one of the best fielding shortstops in baseball and Derek Jeter is considered (at least by sabermetricians) to be one of the worst fielding shortstops despite the fact that he won three Gold Glove awards.
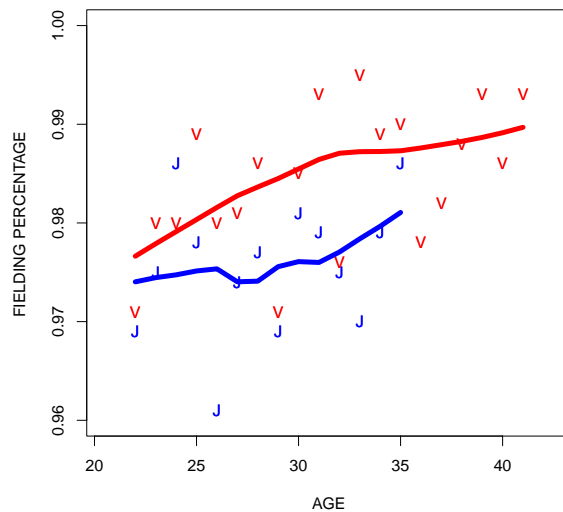

Figure 6: Fielding percentages of Omar Visquel and Derek Jeter plotted as a function of age where the plotting symbol is the first letter of the player's last name. Lowess smoothing curves are used to show the basic patterns in the percentages. It is clear that Visquel was more successful in making fielding plays.

New Measurements, New Data and Measures of Performance

Baseball is continuing changing in its use of statistics. Major League Baseball (MLB) is measuring the game in new ways and the new data generated can help improve our measurements of player performance. During the 2007 baseball
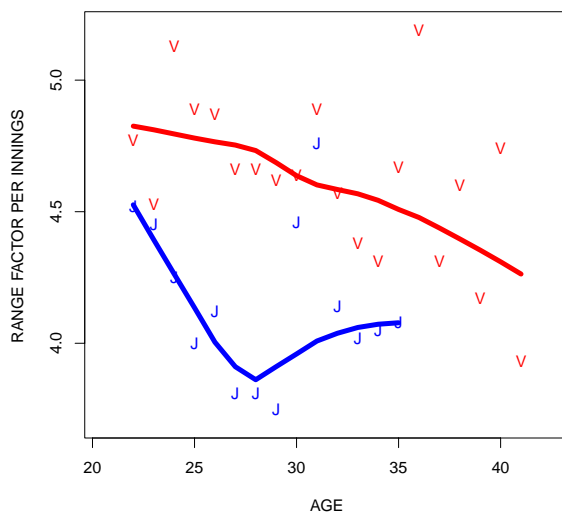
Figure 7: Range factors of Omar Visquel and Derek Jeter plotted as a function of age where the plotting symbol is the first letter of the player's last name. Lowess smoothing curves are used to show the basic patterns in the range factors. This graph shows that Visquel had a superior fielding range in his career.

season, MLB began a systematic effort to record detailed information about the pitches that are thrown. All baseball stadiums were equipped with video cameras that would track each pitched ball and determine precise its trajectory. From the measurements made from the cameras, one is able to learn about the speed of each pitch at its release point and at the point where it reaches home plate. Also one can measure the amount and angle of the "break" of the pitch. This technology is known as the PITCH F/X system.

Here is one illustration of the use of this new data to develop a new measure of pitching performance. One basic pitch in baseball is the fastball that is thrown at a high speed. A successful pitcher is able to throw the fastball at precise locations around the strike zone. A batter often will swing at a fastball. He will typically swing and miss at a "good" fastball thrown high in the strike zone, but be successful in connecting with a fastball that is thrown in the middle of the strike zone. By using the PITCH F/X data, we can measure the "sweet spot" where the batter has a high probability of connecting with the fastball. Figure 8 shows the location of the sweet spot in the strike zone for four pitchers, Paul Maholm, Cliff Lee, Johan Santana, and Edinson Volquez – these graphs are based on pitching data from the 2008 season. One sees that batters are generally successful in making contact with a fastball thrown low in the strike zone. The size of the sweet spot is an indication of the quality of the pitcher's

fastball. Cliff Lee and Edinson Volquez had effective fastball since the batter's sweet spot is small. In contrast, Paul Maholm had a less effective fastball and batters were able to make contact with balls pitched in a larger area.
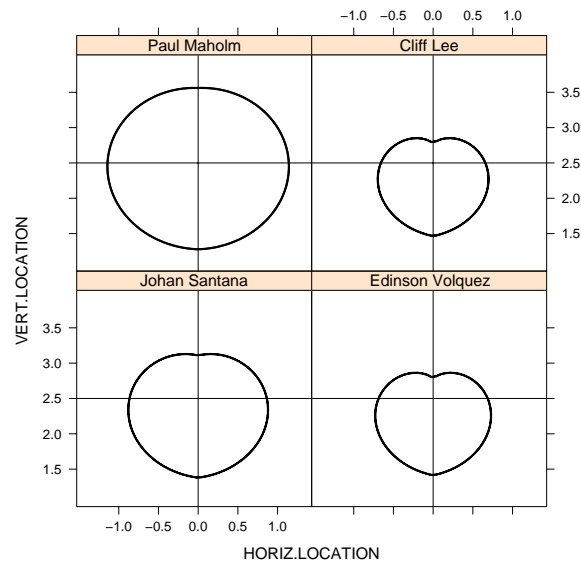

Figure 8: Contour plot of probability of making contact of a fastball for four pitchers. The line corresponds to the probability of 0.78; the region inside the contour line is the "sweet spot" for the batter where the probability of contact exceeds 0.78. The pitcher Edinson Volquez with an excellent fastball has a small sweet spot.

MLB is currently testing a new measurement system that will record the exact speed and location of the ball and every player on the field during a game. This system will have many applications, but it will make it much easier to measure fielding ability. A fielder is responsible for covering a particular section of the playing field and this new system can record the success and failure of each play by the fielder in his zone. Allan Swartz in a *New York Times* article, says that this new system will allow "the most digitized of sports to be overrun anew by hundreds of innovative statistics that will rate players more accurately, almost certainly affect their compensation and perhaps alter how the game itself is played."

Further Reading

The book *Curve Ball* by Albert and Bennett provides a good introduction to sabermetrics and describes statistical thinking in the context of baseball.

Baseball Between the Numbers: Why Everything You Know About the Game Is Wrong by Baseball Prospectus and *The Book: Playing the Percentages in Baseball* by Tango, Lichtman, and Dolphin describe a variety of sabermetrics topics such as clutch ability, baseball strategy, how much should players be paid, the impact of ballparks on statistics, and what statistics tell us about steroid use. The book *The Numbers Game: Baseball's Lifelong Fascination with Statistics* by Swartz gives a history of the use of statistics in baseball including some of the important people such as Henry Chadwick and Bill James.

Biography

Jim Albert is a Professor of Mathematics and Statistics at Bowling Green State University. He is a Fellow of the American Statistical Association and past editor of *The American Statistician*. His research interests include Bayesian inference for categorical data, the teaching of statistics, and the application of statistical thinking in sports. He has written books on statistical thinking in baseball, Bayesian inference, and on the use of the R statistical system to perform Bayesian computations. Away from academia, Jim is an active tennis player and enjoys music and following the Philadelphia Phillies.

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403 `albert@bgnet.bgsu.edu`