

Predicting Run Production and Run Prevention in Baseball: The Impact of Sabermetrics

Philip Beneventano

Ernst & Young
200 Clarendon St.
Boston, MA 02116, USA.

Paul D. Berger

Bentley University
175 Forest St.
Waltham, MA 02452, USA.

Bruce D. Weinberg

Bentley University
175 Forest Street
Waltham, MA 02452, USA.

Introduction

The game of baseball is very popular in many countries, especially the United States, Japan, South Korea, China, and a variety of Central American countries. As a result, the revolution in statistics that began a few decades ago, and continues today, has spearheaded a revolution in the analysis of baseball data and its use in decision making about which players a team should pursue and what value a player has to a baseball team.

The revolution started with Bill James, a man who, in the 1970s, worked as a security guard during the day and had no professional baseball experience. He started composing his extremely popular *Baseball Abstract* books in the late 1970s and coined the word “sabermetrics.” The word *sabermetrics* comes from an acronym of the Society of American Baseball Research and represents an analysis of the game of baseball using detailed performance data, rather than qualitative methods based on such numbers as a player’s height and weight, “look on his face,” and relatively simple statistics such as batting average (BA - number of hits divided by number of at-bats). While more traditional professionals and fans might view popular statistics like batting average and strikeouts as indicators of performance, current “sabermetricians” tend to use different, more detailed statistics, such on-base percentage (OBP - ([hits plus walks plus hit by pitches] divided by [at-bats plus walks plus hit by pitches plus sacrifice flies]) and frequently create their own measures to analyze which players (or teams) are best.

There is no clear line between popular or “conventional” statistics and sabermetric measures/statistics, but as a general rule, sabermetric statistics are not seen in box-scores of games. A box score of a game from April, 2011 is can be seen in Figure 1. Examples of the more common conventional statistics are batting average, runs batted in (RBIs), stolen bases, pitcher wins, and pitcher saves.

Even though box scores do not include many sabermetric statistics, these numbers can be combined using the necessary calculations (and along with taking the ball park of the at-bat into account), resulting in more advanced and complex statistics, to arrive at the new sabermetric number. Essentially, sabermetric statistics use conventional statistics in carefully-chosen combinations, to calculate measures *thought to more accurately gauge a player’s value or relative worth*. Bill James wrote in his 1979 *Baseball Abstract* that “it is startling... how much confusion there is regarding how a hitter or team should be measured” (James, 1979). James especially cited batting average as a poor measure of a team’s performance; he pointed out that many teams, ranked by BA, resulted in ranking teams with higher run totals below teams scoring fewer runs, *even though the purpose of a team’s offense is not to compile a high batting average, but to score more runs!*

Another example comes from the use of runs batted in, or RBI, to evaluate players. While getting a hit, walk, sacrifice, or sacrifice fly to allow a runner to score is very valuable, RBI fails to account for the fact that not all hitters get the same opportunities to hit with runners on base. A leadoff hitter (especially in the National League, where the pitcher almost always hits in the ninth and final slot in the lineup) will likely not get as many chances to hit with runners on base as the third or fourth hitters in the lineup. Joe Carter had 115 RBI in 1990, but his anemic on-base percentage of .290 meant that he performed about the same as a replacement player would have, if given the same opportunities for RBIs. His high RBI total was due merely to the ability of players (e.g., Jack Clark [.441 OBP] and Tony Gwynn [.357 OBP]) hitting right ahead of him to get on base at a high rate.

Saves are also often similarly disregarded in the sabermetric community because some closers have many more opportunities for saves. (Primarily, a save is awarded if a pitcher enters the game with his team ahead by three runs or fewer and pitches the rest of the game without his team losing the lead; this “3 or fewer runs ahead” varies if there are men already on base; a save is also awarded if the pitcher enters when his team is (way) ahead, and stays ahead, and pitches the last 3 or more innings) A pitcher (“closer”) may watch his effective outings go unnoticed (in terms of saves) if his team loses the game very often due to poor offense, or his team is “so strong” that it wins many games by a great margin, making the pitcher ineligible to receive a save.

Perhaps the best quote to sum up sabermetric thinking comes from Baseball Prospectus writer Jonah Keri, who states that “If you know how to think properly about the statistical side of baseball, you can have insights that elude even some professionals” (Click & Keri, 2006). Still, sabermetric analysis is criticized by many people associated with baseball, such as former 10-time All-Star second baseman and Sunday Night Baseball announcer Joe Morgan. Morgan, when asked about the rise of advanced statistics, claimed that he had “a better understanding about why things happen than the computer, because the computer only tells you what you put in it” (Craggs, 2005).

In this paper, we report and discuss multiple regression analyses, using various traditional offensive-statistics and sabermetric offensive-statistics for teams over many years, to predict/explain the dependent variable of the number of runs a team scores during a season, and then do the same thing for pitching and defensive statistics (i.e., “runs prevented”) as well, using team earned-run-average (ERA - Number of earned-runs yielded per 9 innings) as the dependent variable. Going into the analyses, our beliefs were that sabermetric measures would do a superior job of describing the relationship between team statistics and runs scored and of runs prevented, compared to traditional measures. These beliefs were proven not confirmed, but in both regression analyses might be said to be “partially confirmed.”

Literature Review, *Moneyball*, and Selected Sabermetric Measures

Michael Lewis’ *Moneyball* (Lewis, 2003), which is about the Oakland Athletics organization and its new statistical analysis of the game, is the most famous book on the sabermetric approach to baseball management. It describes the differences between popular statistical thought at the time and how the Athletics, which were run primarily by general manager, Billy Beane, were able to compete with economic behemoths like the New York Yankees, despite having one of the smallest payrolls in baseball. Indeed, during the latter part of 2011, a popular movie of the same name was made from the book.

Beane was a first-round pick of the New York Mets and never actually made the majors. However, he used knowledge he gained during that time that illustrated (in his view) the flaws in traditional measures of baseball-player and team performance. For example, he detailed how some experienced baseball scouts had become enamored with physical features and even “believed they could tell by the structure of a young man’s face... his future in pro ball,” instead of looking at how the player actually performed on the field of play (Lewis, 2003).

As earlier noted, perhaps the most popular traditional measure of hitting efficacy in baseball is batting average (BA). A batting average above .300 in the major leagues will usually rank a player among the league leaders, while a batting average of .200 might get the player sent back to the minor leagues or released (i.e., essentially, *fired*). Beane argued that BA, in his opinion, does not do a good enough job of describing how much a player contributes to a team. BA does not take into account a player’s ability to get on base through a walk, nor does it take into account a player’s ability to get extra-base hits which will lead to more runs than singles.

Beane and his sabermetric-focused front office championed statistics such as on-base percentage (OBP – defined earlier) and slugging percentage (SP - total bases from hits divided by number of at-bats) because they believed that these measures “...represent the two essential components of creating offense - getting on base and hitting for power” (Click and Keri, 2006). Even in the major leagues today, a walk with a runner in scoring position (2nd or 3rd base) is seen often as selfish (as opposed to trying to hit the ball and score the runs). Players from Latin American countries are taught at a young age that an aggressive approach is favorable. Shortstop Miguel Tejada is noted for saying that “you don’t walk off the island,” which, of course, meant that getting hits is a better way to gain the favor of baseball scouts than having “plate discipline” and a high walk-rate.

Early sabermetricians combined batters’ and teams’ on-base percentage and slugging percentage into a statistic called OPS (On base Plus Slugging.) However, even this measure doesn’t fully describe how effective a team is at scoring runs, and has increasingly been replaced by sabermetric measures which *weight* on-base percentage and slugging percentage (wOBA), although not everyone agrees on the weights, except that on-base percentage should get more weight than slugging percentage. This is because most sabermetricians agreed with Lewis’ (2003) argument that, because a baseball team is afforded only three outs in an inning, avoiding making outs (i.e., getting on base!!) is more important than hitting for power, even though, both, of course, are good. In addition, Lewis argued that wOBA takes into account the fact that not all hits or total bases are “created equal,” and that, on average, a double is not as valuable as two singles. The value of wOBA is purposely constructed to be on the same scale as OBP, so any score that would be a great OBP is also a great wOBA (Slowinski, 2010a).

One traditional statistic that is frequently used to value players and teams is stolen bases. However, sabermetricians claim that stolen base numbers distort a player’s value. Beane did not build his team around speed because his statistical analyses (Lewis, 2003) showed that many players would have helped their team more if they had never attempted a steal. Subsequent to writing *Moneyball*, when asked about this further, he cited the example of Juan Pierre, who stole 45 bases in 2004; however, it was in 69 attempts!! *The benefit of gaining an extra base 45 extra times did not outweigh the 24 extra outs made by attempting to steal.* In addition, a large number of steals often hid the fact that the player did not hit for power, or, sometimes, even have a high batting average. It is not clear who said it first, but there is an old adage that, “you can’t steal first base.”

Beane would often place good-hitting, bad-fielding infielders in left or right field to give the offense a boost; his belief was that the additional offense outweighed the decreased defensive abilities. In practice, there are examples that went both ways. Despite some negative experiences, Beane and Lewis showed how a small-market team can use new methods of thinking to outperform teams with much higher payrolls.

The book *Baseball Between the Numbers* (Click and Keri, 2006) did an excellent job of highlighting the importance of both defensive and pitching statistics, which are much more difficult to quantify than offensive measures, and there is frequent debate between fans of traditional defense/pitching measures of performance and fans of sabermetricians. While older, more commonly used defensive statistics, such as number of putouts or fielding percentage, are easy to determine and calculate, sabermetric measures focus instead on a frequently unquantified factor in fielding: *range* (how far to the left and right the player can go and still field batted balls). A player with a perfect fielding percentage of 1.000, but poor range, may actually be a worse fielder than a player who makes the occasional error but has superior range in the field and can get to more balls put in play. Click and Keri (2006) sum up the argument by saying, “...in effect... errors are the same as hits as far as the defense is concerned; they are simply balls on which the defense failed to make the play.”

Further complicating the issue is the inequality of fielding chances two very similar fielders may get. A gifted centerfielder who is on a team with a pitching staff throwing “sinkers” (pitches more likely to induce a ground ball than a fly ball) may not get as many putouts as a worse fielder with pitchers who induce fly balls. *Baseball Prospectus* contributor Clay Davenport developed the statistic of Fielding Runs Above Average (FRAA), which adjusts for factors such as the game’s baseball park (all parks differ in size and shape), total balls in play allowed, groundball or fly ball tendencies. A fielder’s number represents how many runs he saved the team or cost the team relative to a calculated “average” fielder for the league. Derek Jeter, despite winning many Gold Glove awards for his defensive skills, consistently ranks poorly in FRAA. Recently, more advanced defensive metrics, such as Ultimate Zone Rating (UZR), have been created to properly “estimate each fielder’s defensive contribution in theoretical runs above or below an average fielder at his position in that player’s league and year” (Lichtman, 2010).

UZR utilizes play-by-play data recorded by *Baseball Info Solutions* over the past few years to give fielding credit to specific players. The UZR “engine” is able to compile the type, location, and speed of each batted ball and also take into account data such as temperature and ballpark played in. Adjusting for these several factors, it then compares those numbers to how often average players turn batted balls into outs; this negates the effect that some fielders have of “ball-hogging” easy potential outs. The four components of UZR are outfield arm runs, double play runs, range runs, and error runs.

The third and final area that sabermetric analysis can (allegedly) help to better analyze is pitching performance. Conventional wisdom says that wins are the most important statistic for starting pitchers, while saves are most important for relief pitchers, especially “closers.” This idea was embodied as a way of thinking when it was written that baseball is “context-specific and is “a team sport where the team goal is victories” (Yang, 2008). Recently, many baseball fans (those in fantasy baseball leagues in particular) have started using statistics such as WHIP (walks plus hits per inning pitched) and K/9 (strikeouts per nine innings) to properly analyze a pitcher’s worth. WHIP and K/9 determine how well each pitcher truly performs when out on the mound and is not context-specific, while a statistic such as wins is deemed too context-specific, because an excellent pitcher should not be underrated due to being on a team with poor offense (i.e., then, the same set of performances will result in fewer wins).

Many in the baseball world were shocked (and some very pleased!) to see Felix Hernandez win the 2010 American League Cy Young Award (given to the best pitcher in each league as voted on by a collection of baseball writers) because his win-loss record that year was a mediocre 13-12 compared to league rival, CC Sabathia’s 21-7 record. Apparently, the writers were able to ignore number of wins and winning percentage, and see that Hernandez had an ERA that was 0.91 runs lower than that of Sabathia, pitched more innings than Sabathia, and struck out more players per inning than Sabathia. The only reason Hernandez did not win more games (and lose fewer games) was that his team, The Seattle Mariners, had a relatively poor offense, indeed, scoring 74 runs fewer than the second-worst offense in all of baseball. Sabermetric websites such as *FanGraphs* and *Baseball Prospectus* have introduced new statistics which (allegedly) show a pitcher’s true value to his team more accurately than the traditional ERA statistics.

Baseball Between the Numbers (Click and Keri, 2006) contains an extensive chapter about the problems with ERA, which “...[does] lessen the problem of teammate reliance, but does not eliminate it,” due to the still-existing effects of defense, quality of the bullpen, and even luck, can have on a pitcher’s ERA. A group of writers, led by Tom Tango, helped create *FIP*, which stands for Fielding-Independent Pitching. FIP is a formula created using only a pitcher’s strikeouts, walks given up, home runs allowed, and innings pitched, plus a constant, to give a value on the same scale as the familiar ERA value which fans are more familiar with. This statistic takes into account that a pitcher has very little control of batted balls [put] into play (BABIP), so it accounts only for the balls a pitcher has not allowed in play; note that a home run is *not* a ball put into play. The key is that statistics like FIP, which are part of a group of statistics called defense-independent pitching statistics, recognize the extremely small influence pitchers have on BABIP. We believed (going into this study) that the sabermetric statistics such as FIP and WHIP will be prominent in predicting ERA, because they get to the root of how runs are given up, unlike statistics such as wins or saves which can strongly distort how well the pitcher performed and how effective he was at preventing runs from being scored.

We have not come across multiple regression models using a combination of more common, conventional statistics and newer, complicated (saber)metrics for either offense or defense and pitching. Basic analysis of the relative importance of on-base percentage and slugging percentage has been done, such as the work by V. Wang in the *Society for American Baseball Research Journal* as reported in the New York Times (Schwartz, 2007), and some more complex measures were considered by by Amour and Levitt (2003). Wang found “that weighting on-base percentage 80 percent more than slugging percentage worked best” when attempting to find the closest correlation between OBP and SLG weights and runs scored. This study corroborates earlier work that suggested the importance of on-base percentage, which, in turn, demonstrates the importance of not making an out, relative to hitting for power.

It should be noted that many individual teams or baseball statisticians have their own proprietary regressions or models known only to them to value and predict the performance of players or teams as a whole.

Moneyball was an extremely rare look into what tools front office and general managers use to adjust their roster, as most teams are unwilling to share exactly how they operate behind the scenes. However, a large percentage of the work in the sabermetric field is open to the public and can be seen and discussed on websites such as *Baseball Prospectus* and *FanGraphs* for free or for a very limited monthly fee.

The Internet has allowed for sabermetric insights to be found by unlikely sources; the year-to-year instability of batting average on balls in play was originally discovered not by an employee of a team or by a statistician, but by Robert McCracken, who was working as a paralegal, and who frequented the *rec.sport.baseball* forum between 1999 and 2001 (Passan, 2011). McCracken was able to create documents and spreadsheets with decades of data and share it with others on the forum to help promote and refine his new findings. Other sports have their own equivalents to sabermetrics and advanced statistical analysis, although their team aspects makes the task far more difficult, and most would agree that the statistical analyses for these other sports are less-well developed than baseball analyses. Basketball has APBRmetrics (named after the *Association for Professional Basketball Research*), and football has the *Advanced NFL Statistics* and *Football Outsiders* websites that develop statistics such as DVOA, or defense-adjusted value over average, to evaluate the success of plays while accounting for the score, field position, and many other factors.

Methodology

We perform two different stepwise-multiple-regression models using SPSS to discover the conventional and advanced sabermetric statistics that have strong relationships to 1) run production and 2) run prevention. Each data point represents one team for one year. The first regression model has a dependent variable of runs scored and independent variables of both conventional and sabermetric offensive statistics; the primary goal is to see which statistics have what relationship to scoring runs. The second regression has a dependent variable of ERA, and the independent variables are both conventional and sabermetric pitching and defensive numbers. The primary goal is to see which statistics have what relationship to allowing (earned) runs. To obtain the hitting, pitching, and defensive data, we used helpful sites such as *espn.com* and *fangraphs.com*, which provide both conventional statistics and advanced (saber)metrics for all players and teams for several years, free of charge.

The offensive analysis used data from all teams for a 10-year period (2001-2010). Given the turnover of player personnel year-to-year, and the fact that there were no rule changes over the 10-year period, the data were treated as 30x10=300 cross-sectional data values. Since UZR (recall: Ultimate Zone Rating for measuring performance of a fielder) was developed in 2002, the values for defensive and pitching statistics are comprised of team data for each year for the 9-year period between 2002 and 2010, resulting in 270 data values.

Independent Variables

Table 1 lists the independent variables the stepwise regression with dependent variable Runs Scored was able to select from, indicating also whether the variable is considered “conventional/traditional” or “sabermetric,” as judged by sabermetricians and the vast majority of baseball analysts; there are 6 variables of each Type:

“CONVENTIONAL” MEASURES	“SABERMETRIC” MEASURES
<ul style="list-style-type: none"> • Number of Home Runs • On-base (OB) % • Slugging % • Number of Stolen Bases • Number of RBI’s • Batting Average 	<ul style="list-style-type: none"> • Isolated Power (slugging % minus BA) • % Plate Appearances ending in a Walk • % Plate Appearances ending in a Strikeout (“Strikeout %”) • wOBA (Weighted on-base %) • Batter’s BA of Balls Put into Play • OPS+ (OB% + Slugging % relative to League Average, adjusted for Ballpark) (Baseball Statistics Glossary, 2011)

Table 1: Independent variables for regression with dependent variable: RUNS SCORED

Table 2 lists the independent variables the stepwise regression with dependent variable Earned Run Average was able to select from; again, there are 6 variables of each type:

“CONVENTIONAL” MEASURES	“SABERMETRIC” MEASURES
<ul style="list-style-type: none"> • BA Against • Number of Double-Plays Turned • Fielding % • Strikeouts per 9 Innings • Walks per 9 Innings • Home Runs Allowed per 9 Innings (HR/9) 	<ul style="list-style-type: none"> • BABIP (% of Pitcher’s Batted Balls that are Hits) • FIP (Fielding Independent Pitching) • xFIP (Same as FIP, except that the pitcher’s Home-run Rate is Replaced by the League Average – considered by some to be superior due to extremely high variability from year-to-year for Individual Pitchers) • WHIP (Walks + Hits Given up per 9 Innings) • Left-on-base (LOB)% (Percent of Runners that Get on Base that are Left on Base) • UZR (Ultimate Zone [fielder] Rating)

Table 2: Independent variables for regression with dependent variable: EARNED RUN AVERAGE

It was anticipated that these mixes of conventional independent variables and sabermetric independent variables would be broad enough to find models to correlate highly with run production and run prevention,

Results and Discussion

Runs-Scored (Offensive)Model

According to the offensive model, the best predictor of runs scored for a team (Y) is weighted on-base average (wOBA), a primary statistic used by sabermetricians. On its own, wOBA had an $R^2 = .896$. This is a relatively high value and means that wOBA, on its own, is estimated to explain about 90% of the variability in number of runs a team scores. Weighted on-base average was followed in the stepwise regression (in order) by strikeout percentage, slugging percentage, and on-base percentage. The final model is (coefficient rounded to nearest integer):

$$Y_{\text{predicted}} = \text{Number of Runs Scored} = -903 + 2226 X_1 - 184 X_2 + 1116 X_3 + 1501 X_4$$

where:

- X_1 = weighted on-base average
- X_2 = % of plate appearances ending in a strikeout
- X_3 = slugging percentage
- X_4 = on-base percentage

The value of R^2 for the above “final model” = 95.3%. The only variable with a negative coefficient is strikeout percentage; the negative sign, of course, makes intuitive sense, since teams that strikeout fewer times per plate appearance, put more balls in play, and thus, have more opportunities for hits (and, consequently, runs scored!). The positive coefficients for the other three variables also make intuitive sense. It might also be noted that, while all of the X’s are on a (0 – 100) scale (percents), the coefficient of X_2 is substantially lower than those of the other three X’s. That coefficient was larger in magnitude when it entered the equation, but got smaller when X_3 entered, and yet smaller when X_4 entered (while still retaining a p-value < .05, and hence, not deleted from the model.) This is not too surprising, since, *ceteris paribus*, the more a player strikes out, the less frequently he puts the ball into play, and, correspondingly, the lower his slugging percentage and on-base percentage. The main original contention of this paper – that the sabermetric variables would dominate the final model’s independent variables – is not confirmed, or, perhaps, may be said to be “partially confirmed.” The final model contains two sabermetric variables and two conventional variables; however, the *first two variables selected* by the stepwise procedure were sabermetric variables, and together, they provided a value of $R^2 = 93.0\%$.

Runs-Prevented (ERA) Model

According to the ERA model, the best predictor of runs prevented for a team (Y) is WHIP, a primary statistic used by sabermetricians. On its own, WHIP had an $R^2 = .940$. WHIP was followed in the stepwise regression (in order) by LOB%, HR/9, Fielding%, and Number-of-double-plays. The final model is:

$$Y_{\text{predicted}} = \text{ERA} = -12.483 + 2.889X_1 - 9.564X_2 + 1.006X_3 - 19.022X_4 - .001X_5$$

where:

$X_1 = \text{WHIP}$

$X_2 = \text{LOB\%}$

$X_3 = \text{HR/9}$

$X_4 = \text{fielding \%}$

$X_5 = \text{Number of double plays}$

The value of R^2 for the above “final model” = 98.8%, which, of course, is a very high value. Two of the coefficients are positive, those of WHIP and HR/9, and each result is intuitive. A team that gives up more hits and/or walks per inning, *ceteris paribus*, would be expected to have a higher value for ERA; the same can be said for HR/9. Three of the coefficients are negative, LOB%, Fielding %, and Number-of-double-plays. All of the negative coefficients would seem to make intuitive sense. A higher percent of base runners left on base means that fewer of them score. As was mentioned earlier, there is some controversy about how meaningful the statistic, fielding%, is, it is certainly not surprising that, *ceteris paribus*, a higher fielding% leads to a lower ERA; indeed, in a simple regression with ERA and fielding%, the coefficient was also positive. The result for Number-of-double-plays might appear to be intuitive, but a few words are in order, here. When somebody is on base, “inducing” a double play is a *good* thing!! However, in a way, there is a “catch;” a double play can occur only when at least one player is on base, so that a team that converts more double plays also has more base runners – a *bad* thing! The key is the “*ceteris paribus*” part of the explanation. Once we are holding WHIP constant, then, indeed, converting more double plays reduces ERA. Interestingly, and different from fielding%, a simple regression of ERA and Number-of-double-plays is *negative*, indeed reflecting the increased number of base runners (and correspondingly, more runs scored). It might be noted that Number-of-double-plays was the last variable to enter the stepwise regression, and, while statistically significant ($p < .05$), added only about .001 to the final R^2 .

As an aside, all the coefficients for the *runs-prevented* (defensive and pitching) model are much smaller than the coefficients for the *runs-scored* (offensive) model. This is simply due to scale. The run-prevented model is predicting a team’s earned run average (virtually always between 3 and 6), whereas the runs-scored model is NOT per 9 innings, but a total for the season (virtually always between 5000 and 9,500). It is simply a matter of baseball tradition that offenses in baseball are rarely ranked by the number of runs scored *per game*, but rather by the total number of runs scored for the season. Conversely, you will rarely see a reference to a team’s total earned runs allowed when discussing its pitching prowess. Of course, one can convert the two dependent variables to the approximately the same scale by multiplying ERA times 162, the number of games in a season – it is approximate since some games go more than 9 innings, and some seasons for some teams are comprised of fewer than 162 games (by no more than a couple of games), due to postponements late in the season (due to weather) that are not made up since the games have zero affect on the division winners or wild-card winner. The main original contention of this paper – that the sabermetric variables would dominate the final model’s independent variables – is, again, not confirmed. The final model contains two sabermetric variables and *three* conventional variables, although, as in the previous model, the *first two variables selected* by the stepwise procedure were sabermetric variables (WHIP and LOB%), and together, they provided a value of $R^2 = 96.1\%$.

Further Discussion

It was surprising to see a traditional defensive statistic, fielding%, make the final stepwise model while UZR was not included. After all, UZR has been highly heralded as the “poster child” for how sabermetric measures are so superior to traditional measures. However, one possible reason for this may be its relative newness; while scorers (usually, baseball writers, hired to do so) have been classifying plays as hits or errors for decades, UZR is constantly being updated, so data from the first several years in the study may not have been *exactly* the same measure, and, hence, not have the same predictive power, as UZR data from, say, the 2010 season. In addition, the fact that it is park-adjusted, unlike most of the other statistics in the model, may also have affected its exclusion from the final model.

While this fact would likely not affect the UZR of infielders, it can drastically alter the UZR's given to outfielders. As the *FanGraphs* description of UZR points out, some "left fielders play primarily in Fenway Park," with a distance to the left field wall of 310 feet, "while some play in PETCO Field," which is the most run-suppressing ballpark in Major League Baseball (Slowinski, 2010b). While UZR theoretically accounts for this, it is not clear whether it is a "very precise" accounting.

As we had expected, traditional statistics such as batting average, home runs, and stolen bases were not included in the stepwise regression model for offensive statistics. Proponents of batting average argue that a hit is more valuable than a walk because many hits go for extra bases and advance runners more than one base. However, wOBA (*weighted* on-base percentage), which was in the final model, indeed entering first) takes this into account. As previously mentioned, stolen bases can be accompanied with a high number of caught-stealing attempts, and this will hurt, not help, the team score more runs. Finally, while the value of home runs is captured in slugging percentage, it is captured as well as wOBA. Sabermetricians would likely be surprised that FIP and/or xFIP were not in the final model, as these are also heralded as primary sabermetric measures. One could speculate whether any of the reason for this is that both of these measures are highly variable from year-to-year. Out of the five variables in the final runs-prevention model, three were pitching statistics and two were fielding statistics. So, these were split somewhat evenly between the two types of measures. This might be seen to belie the thoughts of Billy Beane as related in *Moneyball* – that defense is not that important in the grand scheme of things (Lewis, 2003).

Conclusions and Directions for Further Research

We believe that the multiple regression models are potentially useful in determining exactly how run production on offense and run prevention on pitching and defense have relationships with both older conventional measures and newer sabermetric measures. These models can help teams properly determine exactly how valuable players are to the teams and potentially find undervalued players to sign. Unfortunately, we were unable to fruitfully use data from the potentially useful Field FX system from Sportvision, which became available only at the beginning of the 2011 season; the system is considered to be interesting new development for sabermetric fielding statistics, and may revolutionize the capturing of defensive statistics. Essentially, "Sportvision's FieldFX camera system records the action while object-recognition software identifies each fielder and runner, as well as the ball" then provides data on every movement involved in the play (Carey, 2010). Sportvision is the company responsible for the ubiquitous yellow first-down line in every football telecast and created the "glowing hockey puck" for NHL telecasts. FieldFX will have the capability to make the somewhat crude and confusing park aspect of UZR a thing of the past, and use several other factors (including humidity, precipitation, and length of the grass) that are not included at all in UZR (Lichtman, 2010).

Future regressions may also include other types of sabermetric defensive statistics that are gaining popularity. For example, John Dewan's Defensive Runs Saved (often abbreviated as DRS or +/-) uses film and computer analysis to determine "how many more or fewer successful plays a defensive player will make than league average" (Slowinski, 2010c). Although statistics have changed significantly from the first baseball box-scores over a hundred years ago, it is important to remember that someone is always out there finding a way to properly value and analyze exactly how well a player or team is performing by developing new methods. From the days of Bill James to today, there is no question that sabermetrics are changing the way baseball players and teams are properly valued.

References

- Amour, Mark, and Daniel Levitt (2003), *Paths To Glory: How Great Baseball Teams Got That Way*. Washington, D.C.: Potomac Books, Inc.
- Carey, Bjorn (2010), "New Camera System Takes the Guesswork out of Baseball Stats," *Popular Science*, February 12.
- Click, James, and Jonah Keri (2006), *Baseball Between the Numbers: Why Everything You Know about the Game Is Wrong*. New York: Basic Books.
- Craggs, Tommy (2005), "Say It Ain't So, Joe." *San Francisco News, Events, Restaurants, Music*. July 6.
- James, Bill (1979), *The Bill James Baseball Abstract, 1979*, New York: Ballantine Books.
- Lewis, Michae (2003), *Moneyball: the Art of Winning an Unfair Game*. New York: W.W. Norton.
- Lichtman, Mitchel (2010), "The FanGraphs UZR Primer." *FanGraphs*. May 19.
- Passan, Jeff (2011), "Sabermetrician In Exile." *The Post Game*. January 25.

Schwarz, Alan (2007), "New Baseball Statistic, With a Nod to an Old Standard." *New York Times*. February
 Slowinski, Steve (2010a), "WOBA." *FanGraphs Baseball*. February 15.
 Slowinski, Steve (2010b), "UZR." *FanGraphs Baseball*. February 15.
 Slowinski, Steve (2010c), "DRS." *FanGraphs Baseball*. February 15.
 Yang, Alan (2008), "Wins Are For Losers, Part Eleven Million." *Fire Joe Morgan*. August 27.

Figure 1

J. Reyes ss	4	0	0	0	0	0	1	0	0	.000
W. Harris lf	4	1	2	0	0	0	0	0	0	.500
D. Wright 3b	4	0	0	0	0	0	2	0	1	.000
C. Beltran rf	4	1	1	1	0	0	0	0	0	.250
A. Pagan cf	4	0	1	0	0	0	0	0	0	.250
I. Davis 1b	3	0	0	1	0	1	1	0	0	.000
B. Emsw 2b	2	0	0	0	0	1	0	0	1	.000
J. Thole c	2	0	0	0	0	0	1	0	0	.000
S. Hairston ph	1	0	0	0	0	0	1	0	2	.000
M. Nickeas c	0	0	0	0	0	0	0	0	0	.000
M. Pelfrey p	1	0	0	0	0	0	0	0	0	.000
B. Boyer p	0	0	0	0	0	0	0	0	0	.000
L. Duda ph	1	0	0	0	0	0	0	0	0	.000
P. Beato p	0	0	0	0	0	0	0	0	0	.000
C. Hu ph	1	0	0	0	0	0	0	0	0	.000
T. Buchholz p	0	0	0	0	0	0	0	0	0	.000
Totals	31	2	4	2	0	2	6	0	4	

Batting		Fielding	
2B - W Harris (1, J Johnson); C Beltran (1, J Johnson).		DP - 1 (B Emsw-J Reyes-I Davis).	
RBI - C Beltran (1), I Davis (1).			
Runners left in scoring position, 2 out - B Emsw 1, S Hairston 1.			
Team LOB - 4.			

Florida										
	AB	R	H	RBI	HR	BB	K	SB	LOB	Season Avg
C. Coghlan cf	4	0	0	0	0	1	0	0	2	.000
O. Infante 2b	3	0	1	0	0	1	0	1	1	.333
H. Ramirez ss	4	1	1	0	0	0	0	1	2	.250
M. Stanton rf	2	1	1	0	0	1	0	0	2	.500
S. Cousins rf	1	0	1	0	0	0	0	1	0	1.000
G. Sanchez 1b	4	1	2	1	0	0	0	0	3	.500
L. Morrison lf	3	2	1	1	1	1	0	0	2	.333
J. Buck c	4	1	2	4	1	0	0	0	0	.500
D. Murphy 3b	4	0	1	0	0	0	0	0	0	.250
J. Johnson p	3	0	0	0	0	0	2	0	1	.000
M. Dunn p	0	0	0	0	0	0	0	0	0	.000
C. Hensley p	0	0	0	0	0	0	0	0	0	.000
G. Dobbs ph	1	0	0	0	0	0	0	0	1	.000
L. Nunez p	0	0	0	0	0	0	0	0	0	.000
Totals	33	6	10	6	2	4	2	3	14	

Batting		Base Running	
2B - M Stanton (1, M Pelfrey); G Sanchez (1, M Pelfrey); J Buck (1, P Beato).		SB - O Infante (1, 2nd base off M Pelfrey/J Thole), H Ramirez (1, 2nd base off B Boyer/J Thole), S Cousins (1, 2nd base off P Beato/M Nickeas).	
HR - L Morrison (1, 8th Inning off T Buchholz 0 on, 0 Out), J Buck (1, 4th Inning off M Pelfrey 3 on, 1 Out).			
RBI - G Sanchez (1), L Morrison (1), J Buck 4 (4).			
2-out RBI - G Sanchez.			
Runners left in scoring position, 2 out - C Coghlan 1, G Sanchez 3, L Morrison 1.			
GIDP - H Ramirez.			
Team LOB - 7.			

NY Mets										
	IP	H	R	ER	BB	K	HR	WHIP	Season ERA	
M. Pelfrey (L, 0-1)	4.1	4	5	5	4	2	1	1.85	10.38	
B. Boyer	0.2	1	0	0	0	0	0	1.50	0.00	
P. Beato	2.0	3	0	0	0	0	0	1.50	0.00	
T. Buchholz	1.0	2	1	1	0	0	1	2.00	9.00	

Florida										
	IP	H	R	ER	BB	K	HR	WHIP	Season ERA	
J. Johnson (W, 1-0)	6.2	3	2	2	2	3	0	0.75	2.70	
M. Dunn (H, 1)	0.1	0	0	0	0	1	0	0.00	0.00	
C. Hensley (H, 1)	1.0	1	0	0	0	1	0	1.00	0.00	
L. Nunez	1.0	0	0	0	0	1	0	0.00	0.00	