Chapter 2
Down With The Triple-Crown: Evaluating On-Field Performance

The first step in valuing player performance is figuring out which aspects of performance ought to be rewarded and how to weight them. It might be tempting to borrow from baseball's traditional wisdom to determine which skills that players possess are most important; however, popular notions of what determines success in baseball are not necessarily so. Baseball fans tend to be capable of recalling performance statistics of their favorite players with ease, but there exists widespread innumeracy regarding their interpretation. Despite the available evidence regarding player contributions to winning, most mainstream baseball commentary judges players with antiquated notions of what constitutes good and bad performance.

For example, nearly every time a batter steps to the plate during a televised game, three numbers are posted on the screen below his name: batting average, runs batted in (RBI), and home runs. Like most children who grew up devoting their summers to following the game, I embraced the popular yardsticks for evaluating players without questioning their utility. But the metrics that constitute the "triple-crown" of hitting are not the best measures of batters' abilities to help their teams win, and it does not take much thought to understand why.

The batting average tells us something about one way that a batter can safely reach base: how frequently he gets a hit. This is useful information; however, it can be misleading for two reasons. First, the batting average ignores other ways that a batter can safely reach base without getting a hit. A player who draws many walks or who is hit by many pitches puts a runner on first base for his team and does not make an out, just as a hit does. Reaching base via a hit does

J.C. Bradbury, *Hot Stove Economics*, DOI 10.1007/978-1-4419-6269-0_2,

have some additional benefits over these other methods of reaching base, such as advancing runners; however, batters who walk frequently have more value compared to batters with the same batting average who rarely walk.

Second, the batting average treats all hits equally, even though hits that allow a hitter to advance multiple bases produce more runs than singles. Between two players with identical batting averages, a player with many doubles and home runs is more valuable than a hitter who hits mostly singles. Managers are obviously aware of this as they frequently keep sluggers with low batting averages in the lineup, because they make up for a lack in consistent hitting with power.

The inclusion of RBI and home runs next to batting average may provide some information about the hitting power of a player, because the more hitting power that a hitter has, the more RBI and home runs he ought to have. This logic is correct as players with many RBI and home runs typically do hit with power, but these metrics are not the best sources of information for measuring extra-base power.

RBI is an especially dangerous statistic to rely upon for measuring power, because it is heavily influenced by factors unrelated to hitting ability. A major determinant of RBI is RBI chances: the more often that a batter steps to the plate with runners in scoring position, the more RBI he ought to have. RBI chances are not random across teams or the batting order. A team that has many hitters that reach base will provide many RBI chances to its team's batters that might not be available on a lesser team. Also, a batter who bats in the middle of the lineup typically bats after several players who frequently reach base and will, therefore, have more RBI opportunities than players at the top and bottom of the order. It's not necessarily an ability (such as power) that causes players to rack up RBI; therefore, crediting hitters for RBI rewards or punishes them for factors beyond their control.

Imagine comparing a child born in an upper-class household in the United States to a child born in a refugee camp in Sudan. Just because the American grew up to be a doctor, while the Sudanese became a bus driver does not mean that the American has more natural ability. Clearly, these children's lives were heavily affected by circumstances beyond their control. It might be that the successful American is the more-talented child; just the high-RBI clean-up hitter might more productive than a low-RBI leadoff hitter, but comparing their overall final outcomes is a poor benchmark for measuring their talent. In the real world, it would be difficult to compare the innate talents of these children. We might give them IQ tests or judge their accomplishments relative to their peers in their environment, but in baseball a comparison between player talents isn't difficult at all.

What about "clutch" hitters who perform better than other players in run-producing situations? RBI might capture this skill, and a player who hits better with runners in scoring position would be more valuable than one who chokes. However, as much as fans like to talk about players who rise to the moment, it doesn't seem that hitters have much control over this type of situational hitting (see the Hot Stove Myth at the end of this chapter for evidence). Therefore, it would be wrong to credit players for any successes or failures that they happen to produce in the clutch.

Nearly every event in baseball is recorded, and has been since Henry Chadwick first invented the box score. A clumsy statistic like RBI isn't the only yardstick available for measuring output that a batter generates beyond his batting average. To gauge hitting power, baseball fans often use a modified batting average that is weighted by the number of bases a batter advances when he gets a hit: two bases for doubles, three bases for triples, and four bases for home runs. This way the hitter receives additional credit for power. This metric is known to most baseball fans as the slugging average (SLG). Slugging average is not a perfect measure of hitting power, but it is much more useful than batting average and RBI. A player's slugging average is not affected by a player's teammates nor by his place in the hitting lineup; thus it permits player comparisons across teams and different lineup slots.[5] The slugging average also has the advantage of including the third leg of the triple-crown, home runs.

The slugging average is just one example of a metric that is superior to the triple-crown statistics for judging hitters. Baseball fans have developed a wealth of statistics for measuring player performance. Determining which of these metrics is the best choice for valuing players requires a developing criteria for choosing the right measures.

Criteria for Evaluating Performance Metrics

A performance metric should be judged according to three criteria: (1) how it correlates with winning, (2) the degree to which it separates true ability from random chance, and (3) whether or not the information it conveys regarding performance matches reasonable intuition about what constitutes good performance.[6]

In baseball, teams strive to win, but assigning responsibility for wins to individual players on the team is difficult. For example, a common statistic used for judging pitchers, erroneously labeled "wins" (to avoid the confusion between a team win and a pitcher win I refer to the latter as "W") does a poor job of evaluating pitchers. I frequently hear television analysts comment that W's are the best metric for judging pitchers, because winning games is the goal of every team. As Hall-of-Fame player and ESPN announcer Joe Morgan recently stated, "The name of the game, people always want to forget, for pitchers is wins and losses."[7] This conclusion results from semantic confusion. A starting pitcher is credited with a W if he pitches at least five innings, his team is winning when he leaves the game, and his team holds a lead until the end of the game. This is very different from the sole criterion for a team win: the team scores more runs than the opposing team.

The problem with equating W's with wins is obvious: to earn a W the pitcher needs help from his offensive players, his relief pitchers, and the defenders behind him. Awarding a pitcher full credit for a win because he met the criteria for a W overestimates his contribution. A pitcher who pitches on a team with good hitters will receive more W's than if he was on a team with bad hitters. Conversely, a pitcher on a team with bad hitters will earn fewer W's than he would on a team with good hitters. Properly crediting pitchers for their contributions to winning requires using other measures that better reflect pitchers' abilities to help their teams win.

In baseball, the task of breaking down the game into components of responsibility is relatively easy compared to other team sports, because the teams take turns on offense and defense, and pitchers and batters engage in one-on-one contests. We can value offensive accomplishments for their run production and

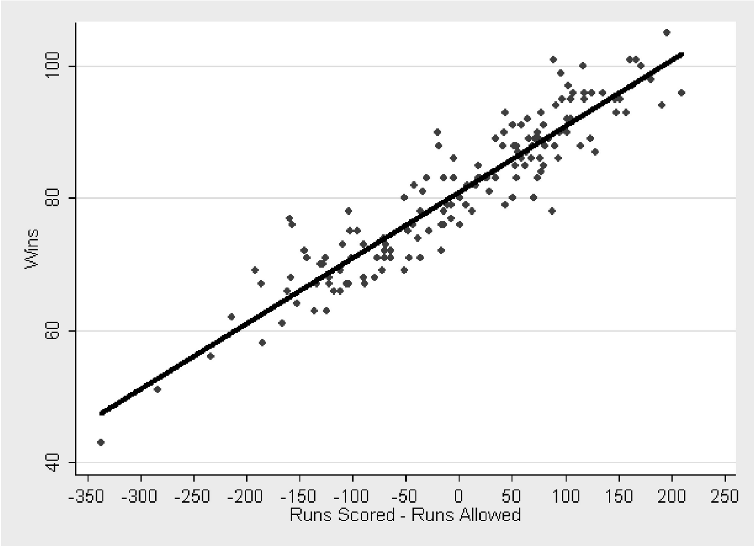

FIGURE 2-1 Relationship Between Wins and Run Differential (2003–2007)

defensive accomplishments for their run prevention. As a hitter produces more runs, or a pitcher prevents more runs, his team's chance of winning increases. Figure 2-1 reveals that there is a tight relationship between team run differential (runs scored – runs allowed) and winning, because the as the run differential rises and falls, so do wins.[8] Evaluating offensive and defensive ability in runs allows us to credit players for the aspects of the game they can control as well as to measure their contributions to winning in a common currency.

Evaluating Hitting

In order to evaluate how well performance metrics meet the first criterion for measuring contributions to winning, let's examine how closely several potential metrics correlate with runs scored on offense and runs prevented on defense, using a sample of team data from 2003 to 2007. The stronger the association between the metric and runs, the better the metric measures player contributions to winning. I chose eight hitting metrics that are sometimes used to judge hitters: batting average (AVG), on-base percentage (OBP), slugging average (SLG), on-base-plus-slugging (OPS), batter's run average (BRA), runs created (RC), regression-estimated runs (LSLR), and linear weights (LWTS).

The first three should be familiar to most baseball fans, and I discussed the batting average and slugging average above. The on-base percentage is the rate at which a player reaches base via a hit, walk, or hit-by-pitch relative to the number of times he steps to the plate. Like the batting average, it does not weight how a player reaches base; but, unlike the batting average, it includes other ways that a hitter can reach base.

The other metrics are commonly used by sabermetricians because they have a stronger correlation with run scoring than the preceding statistics. On-base-plus-slugging, more commonly known by its acronym OPS and popularized by John Thorn and Pete Palmer in *The Hidden Game of Baseball*, is simply the sum of on-base percentage and slugging average. Batter's run average is the product of the two metrics. While adding and multiplying these values together are not intuitive, the combined values correlate strongly with runs scored. Though OPS has its weaknesses, its most-attractive feature is that it is nearly as good an estimator of run production as more complicated metrics while being relatively easy to calculate with information available on the scoreboard.

Runs created, regression-estimated runs, and linear weights are estimators that convert many specific things that players do into expected runs scored, but they differ in their methods for estimating the impact of baseball events. Runs created was developed by Bill James; though, it has many variations I report its simplest formula: the sum of hits and walks, times total bases, divided by the sum of at-bats and walks. Regression-estimated runs uses historical team data to estimate the impacts of singles, doubles, triples, home runs, walks, and hit-by-pitches, stolen bases, and caught stealing on run scoring. The method uses *multiple regression analysis* to weight individual factors according to how much they impact runs. Multiple regression analysis uses changes in many variables across many observations to generate weights to account for the impacts of each factor (see Appendix A for further explanation). For example, the technique estimates that a singe is worth 0.62 runs and a double is worth 0.76 runs.[9]

Linear weights is similar to regression-estimated runs in that it assigns weights to the things that individual players do to produce runs; however, instead of estimating weights of baseball events from team outcomes using regression analysis, it uses play-by-play data to estimate expected runs that typically result from baseball events. This method was developed by operations research analyst George Lindsey. Thorn and Palmer expanded on Lindsey's work to update expected run-value weights from a more recent and larger sample of games. I use the "batting runs" linear weights formula to value hitters in later chapters, because it generates expected run values for nearly all the things that hitters do, including stealing bases.[10]

Figure 2-2 shows the graphs of eight different metrics and their correlations with run scoring. In each graph, the dark trend line maps the direction of the relationship; and, for all the metrics, better performance is associated with more run scoring. The line represents the linear "best fit" of the relationship between performance metric and runs scored calculated by minimizing the prediction error based of metric.[11] In most cases, the actual runs scored and the performance metric for teams do not fall on the line but are close to it. The further the dots are from the line, the weaker the relationship is between the metric and runs scored; dots clustered closely around the line indicate a stronger relationship. The graphs reveal that batting average is the metric least associated with scoring runs. On-base percentage and slugging average have a stronger association with run scoring than batting average, but are less correlated with run scoring than the more-advanced metrics.

The second criterion for choosing a performance metric is how well it reflects skill rather than luck. Though a player may have been involved in events that directly helped or hurt his team's run scoring, his performance was not necessarily the result of an ability, which the market ought to reward.

Imagine that you want to identify the best investment strategist in your neighborhood to handle your retirement. You decide to find your richest neighbor and ask his advice; after all, it's reasonable to assume that a good investor ought to be wealthy if he is good at managing money. However, upon finding this person you learn that he amassed his fortune by hitting the Powerball Jackpot. This doesn't reveal that playing the lottery is a smart business strategy: he got lucky. This example reveals why relying on metrics that are heavily influenced by luck to measure skill can be misleading. It is important not to reward or punish players for outcomes beyond their control when valuing players.

One method for gauging how well a metric captures talent versus luck is to observe how it fluctuates from one year to the next. Real skills should persist over time, while luck ought to fade away.[12] Table 2-1 reports correlation coefficients from season to season for hitters with more than 400 plate appearances in

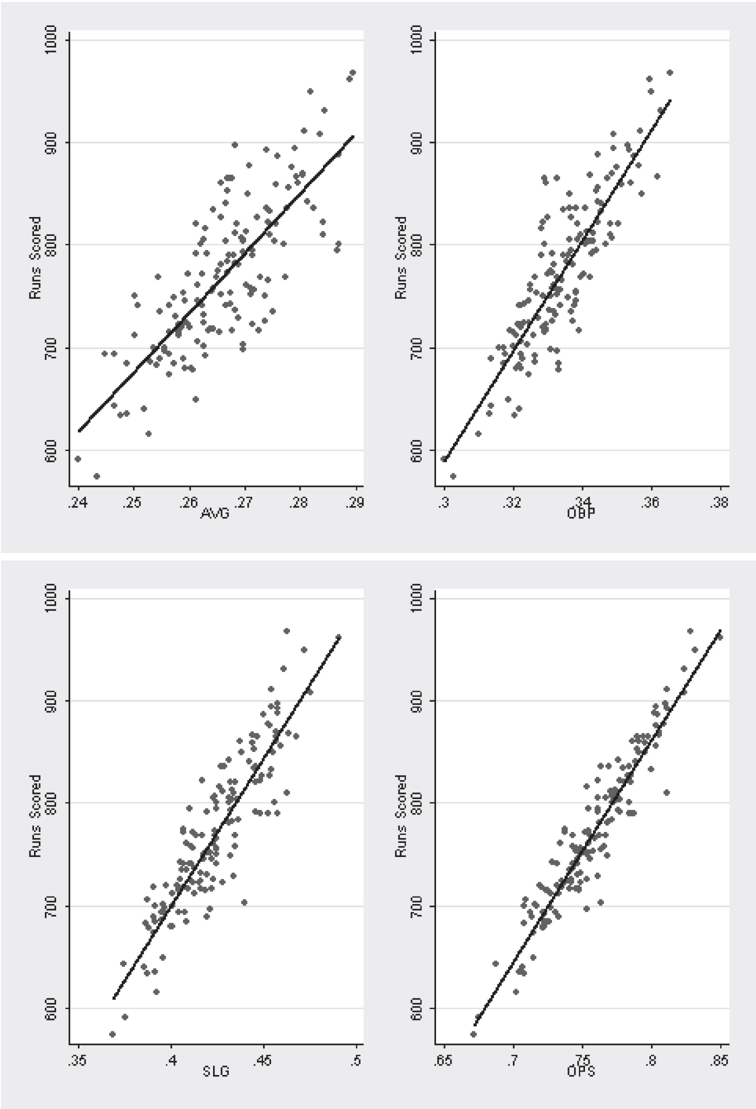

FIGURE 2-2 Correlation Between Metrics and Runs Scored (2003–2007)

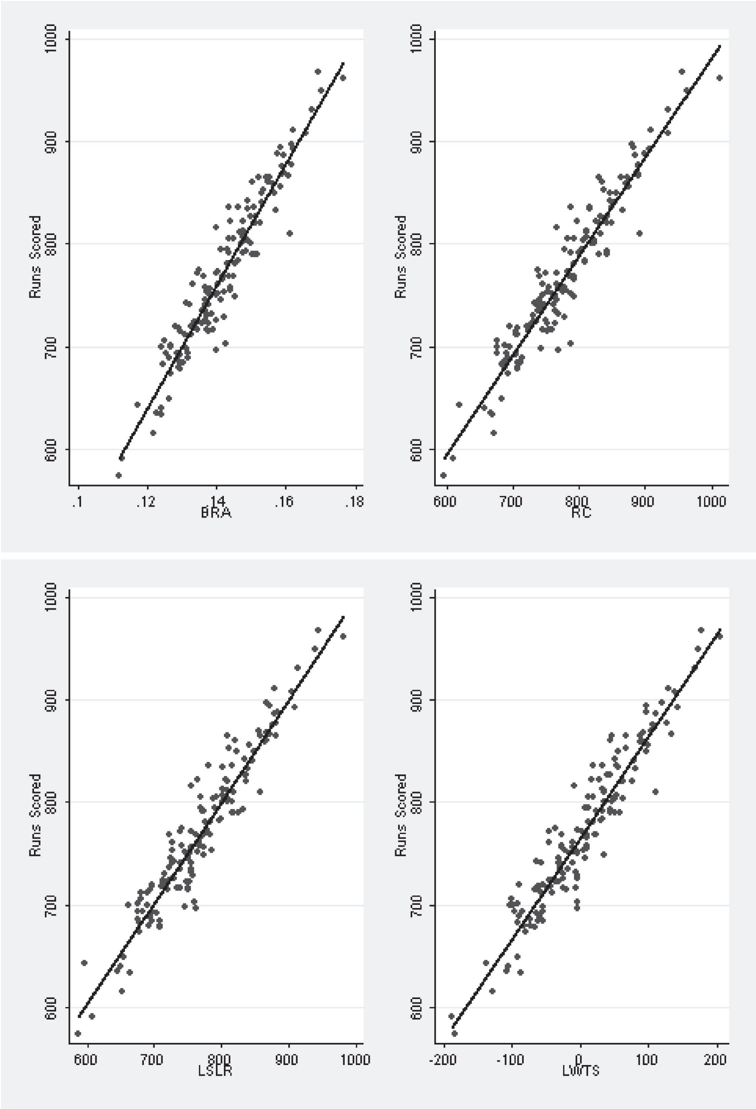

FIGURE 2-2 (continued)

back-to-back seasons. The absolute value of the correlation coefficients range from zero to one as player performances range from less to more similar across seasons.[13] The higher the correlation, the more stable player performance is in this area and the more likely it reflects ability than luck.

Metric	Correlation
AVG	0.4139
OBP	0.6542
SLG	0.6333
OPS	0.6388
BRA	0.6635
RC	0.5782
LSLR	0.5347
LWTS	0.6311

TABLE 2-1 Performance Correlation from Season to Season for Hitters (2003–2007)
>400 PAs for consecutive seasons

All the metrics vary similarly from season to season; however, batting average is the least stable of the bunch. This is not surprising, because hits are heavily influenced by random bounces on the field. Slow bleeders often dribble between fielders, and line drives may be hit directly at defenders, but in most cases bleeders result from bad hitting and "at-'em balls" reflect good hitting. Over a period of time, these occurrences normally even out, but occasionally luck can accrue in one direction. Avoiding volatile performance areas limits potential bias from luck. The high variance of batting average is one of the reasons that I do not like to use it for evaluating players.

Batting average is a major component of on-base percentage and slugging average; therefore, whenever I see a player's numbers in those areas rising above or falling below expectation, I immediately look to the batting average to see if it foretells a coming reversion. If a player's isolated on-base percentage (on-base percentage – batting average) and isolated power (slugging average – batting average) significantly deviate from past performance, then I normally expect the player to return to career form. For example, in 2004, Chipper Jones of the Atlanta Braves batted a measly .248, which was quite a departure from his .309 career batting average. At the time, many commentators thought Jones's career was nearing its end, and that his reduced production was a product of age. Instead, over the next four seasons Jones would bat .332 and win the batting title in 2008. Looking closer at Jones's numbers in 2004, it should have been clear that his down year was an anomaly. He was walking and hitting for power at his career rates; the problem was that his batting average was approximately 60 points below his career norm. Chipper Jones was unlucky on his hits in 2004; and, when all his other numbers remained stable, a rebound should have been expected.

It is important to acknowledge that other metrics being more stable than batting average does not mean that they are immune from luck. Bad and good luck are more likely to be prevalent in the batting average than the other metrics. Other metrics are also subject to random fluctuations; therefore, care must be taken when inferring skills from any performance metric.

At this stage it appears that several offensive metrics are highly correlated with run production and are similarly stable. Any of these metrics would do a fine job at estimating player value, but because I have to choose one metric for valuing hitters, I use the one that makes the most intuitive sense, which the third criterion requires.

OPS, battter's run average, and runs created measure batting skill, but do not include stolen bases. While, stealing bases is not as useful as it is often portrayed in the media, it is a valuable part of many players' games. For example, Carlos Beltran of the New York Mets has attempted to steal over 300 bases in his career, while being caught just twelve percent of the time. Few players can steal at such a high rate, but those who do offer quite a bit of value to their team. Regression-estimated runs and linear weights include stolen bases, but because linear weights uses average outcomes from game states it is better for evaluating individual players.

Regression-estimated runs suffers from a problem known as *omitted variable bias*, which occurs when factors omitted from the analysis are accidentally weighted by factors included in the analysis. Economist Ted Turocy noticed that when stolen bases, caught stealing, and triples are included in the regression model—variables correlated with player speed, and speed is not controlled for explicitly in the model—that the regression estimates assign incorrect weights to the included the factors. Therefore, regression-estimated runs are likely to generate biased weights of player contributions to run production. Linear weights don't suffer from this bias, because it credits the expected value from each event determined from play-by-play outcomes rather than estimating weights from the sum of team performance.

Evaluating Pitching

Measuring pitcher contributions to winning requires a slightly different approach. Unlike hitters, baseball fans typically judge pitchers according to how well they prevent runs using the earned run average, more commonly known as ERA. Denominating performance in runs is an advantage, but ERA is inferior to hitting metrics in other areas. While ERA suffers from some issues in attributing runs to different pitchers, its main problem is that it is heavily polluted by factors beyond the pitcher's control.

First, let's look at the impact of different pitcher performance metrics on run prevention at the team level. Figure 2-3 includes several graphs that map the relationship between runs allowed and several pitching performance metrics: ERA, strikeout rate, walk rate, and home run rate. ERA is far and away the best measure of run prevention, but this is expected. The way runs are credited to teams, only unearned runs—runs that were produced because of errors by the defense—are not included. Earned runs allowed and runs allowed are virtually the same thing at the team level. This is why the second criterion for evaluating performance metrics is so important.

For things that pitchers do to prevent runs without fielders—dish out strikeouts, issue walks, and give up home runs—the relationship with runs allowed is not particularly tight. This occurs because more than 70 percent of plate appearances result in a ball hit into the field of play, which requires the help of fielders. That this is a major component of a pitcher's ERA is unfortunate because pitchers do not appear to have much ability to affect this part of their game. Outcomes from balls in play are heavily random, which makes ERA unstable.

Table 2-2 lists the correlations from season to season for several pitching statistics. In particular, I focus on the main components of ERA: strikeouts (K9), walks (BB9), home runs (HR9), and batting average on balls in play (BABIP). Strikeout and walk rates are much more stable over time than ERA, while the home run rate stability is similar to that of ERA. BABIP measures the percentage of balls handled by fielders that become hits, and it is much less stable than ERA's components.[14]

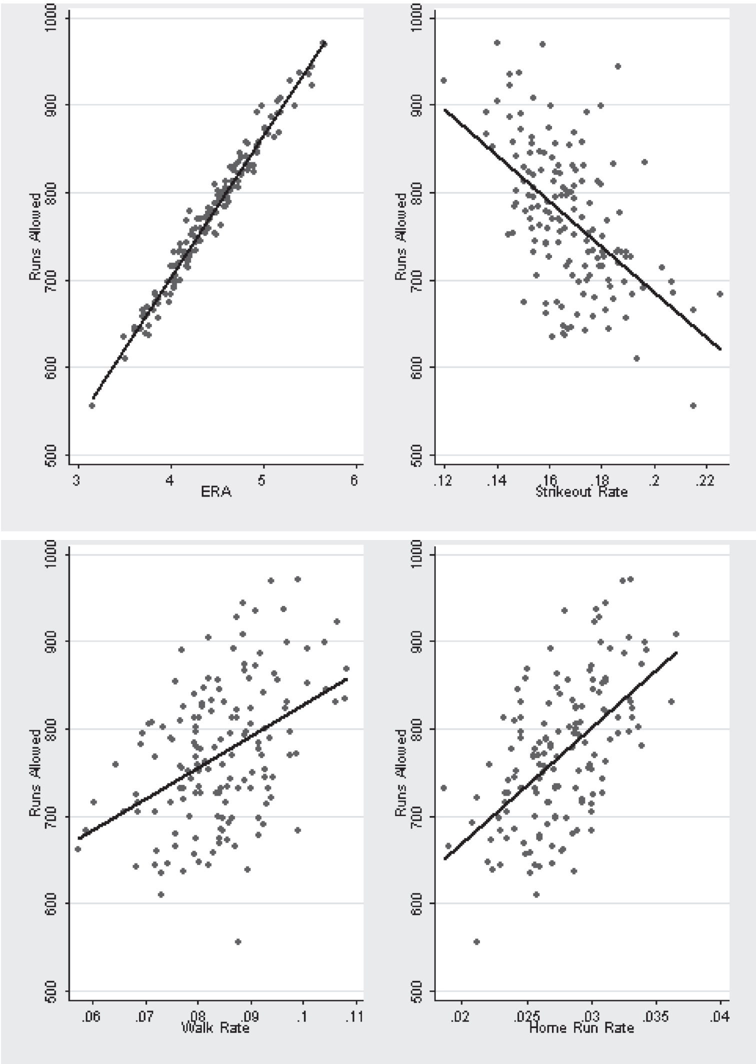

FIGURE 2-3 Correlation Between Metrics and Runs Allowed (2003–2007)

The heavy influence of the unstable BABIP on ERA caused sabermetrician Voros McCracken to develop a new metric for evaluating pitchers without looking at the hits they allow on balls in play: he called it the defense-independent

Metric	*Correlation*
ERA	0.30
K9	0.77
BB9	0.69
HR9	0.32
BABIP	0.18
DIPS ERA	0.54

TABLE 2-2 Performance Correlation from Season to Season for Pitchers (2003–2007)
>400 BFP for consecutive seasons

pitching statistics (DIPS) ERA. McCracken contended that pitchers had little ability to impact whether or not a ball put in play would become a hit. Because hits allowed on balls in play are a major determinant of ERA, the statistic is misleading. His DIPS ERA uses plays during which fielders do not participate in defense to predict performance of pitchers from season to season. It turns out that a DIPS ERA actually does a better job of projecting a pitcher's future ERA than his past ERA. By removing the noise generated on balls in play, we can better gauge pitcher quality and reward pitchers for ability rather than luck.

The last row of Table 2-2 lists the season-to-season correlation for the DIPS ERA that I use to value pitchers in Chapter 4, and it proves to be more stable than raw ERA. Though the DIPS components appear to be only moderate predictors of run prevention on their own, together they do a decent job, and they convey more information about pitcher quality than ERA. Figure 2-4 maps the predicted runs allowed estimated from the DIPS components relative to actual runs allowed. While the correlation with run prevention isn't as strong as raw ERA, this is expected because ERA is merely reporting what did happen on the field, which was heavily influenced by luck. The information provided by the defense-independent performance provides a sufficient prediction and dampens the impact of luck.[15]

The intuition criterion for pitching metrics is easily met by DIPS ERA. That pitchers who strike out many batters and give up few walks and home runs prevent more runs than pitchers who perform poorly in these areas is consistent with expectations regarding run prevention.

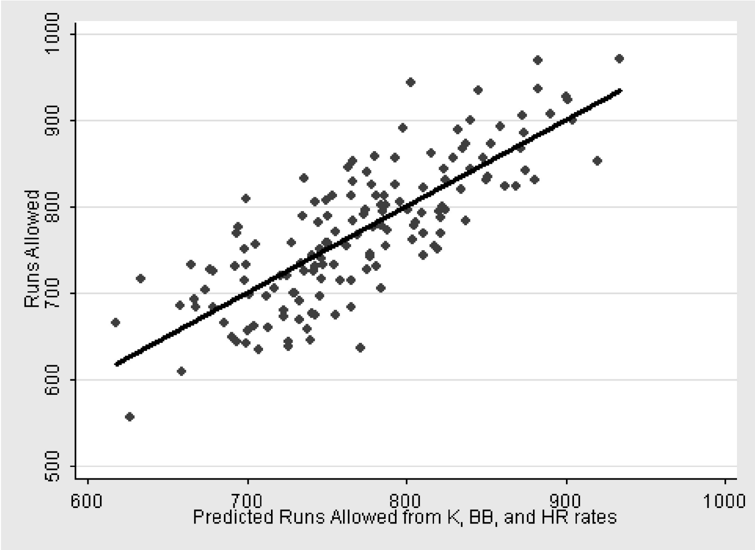

FIGURE 2-4 Predicted Runs Allowed from DIPS ERA

Evaluating Fielding

Defense behind pitchers also plays a role in preventing runs, and many players are prized for their defensive skill at turning likely hits into outs. Unlike hitting and pitching, few metrics exist for quantifying fielder contributions. Errors, the mainstream benchmark for measuring defensive ability, are deeply flawed by the fact that errors are based on the subjective judgment of scorers. Skilled scorers, trained to evaluate defense objectively, might be able to judge the fieldability of all balls hit into play consistently, but that is not how scorers assign errors. The basic criterion for recording an error on a play is that if a fielder looks like he should have made the play and did not record an out, he is credited with an error. This means that there are two ways to avoid errors: make plays or don't make it look like a play could be made. A fielder with stone feet won't come within fielding range of a ball that most average fielders would easily turn into an out; yet, an error won't be recorded because it didn't look like he could have gotten to the ball. On the other side, excellent fielders who flub plays that normal fielders wouldn't come close to making are credited with errors.

There have been a few attempts to generate alternative fielding metrics, but most suffer from inherent biases in their construction or calculation.[16] In 2004, Baseball Info. Solutions began recording a new fielding metric known as plus/minus. To remove some subjectivity from defensive analysis, plus/minus is calculated by mapping batted balls on a television screen using a grid. Players are then held responsible for fielding balls relative to their peers and are graded in terms of plays made above (plus) or below (minus) average. John Dewan, the founder of STATS, Inc. and Baseball Info Solutions, explains the objectivity of the recording:

> Every play is entered into the computer where we record the exact direction, distance, speed, and type of every batted ball. Direction and distance is done on a computer screen by simply clicking the exact location of the ball on a replica of the field shown on the screen . . ..
> The computer totals all softly hit groundballs on Vector 206, for example, and determines that these types of batted balls are converted into outs by the shortstop only 26% of the time. Therefore, if, on this occasion, the shortstop converts a slowly hit ball on Vector 206 into an out, that's a heck of a play, and it scores at +0.74. The credit for the play made, 1.00, minus the expectation that it should be made, which is 0.26. If the play isn't made—by anybody—it's −0.26 for the shortstop . . .
> Add up all the credits the player gets and loses based on each and every play when he's on the field and you get his plus/minus number (rounded to the nearest integer).[17]

According to *The Fielding Bible*, each play that a player makes above/below average costs the team between 0.56 and 0.76 runs, depending on the position played. Thus, the plus/minus numbers can be used to estimate how many runs a player adds or subtracts with his defense. While Dewan continues to tweak his system, and certainly there will be improvements, the basic core of what is being done is correct, and I feel safe using these measures to quantify the fielding prowess of position players. Because of its newness, I cannot evaluate its variance as I have done with the hitting and pitching metrics. However, having reviewed the system and viewing some of the raw data, I am satisfied that plus/minus provides a sufficient proxy for fielding skill, and, therefore, I will use it to value the defensive ability of fielders.

Summing Up

Performance metrics should be evaluated according to three criteria: (1) how well they correlate with winning, (2) the extent to which they measure ability instead of luck, and (3) if they make intuitive sense. Hitters who get on base and move around the bases with power and speed produce runs, and linear weights does a good job of measuring the many things that hitters do to produce runs. Pitchers who keep the other teams' hitters off the bases by striking out batters, preventing walks, and keeping the ball in the ballpark prevent runs. Though recording outs on fielding balls is a major determinant of runs allowed, pitchers appear to have little influence over preventing hits on balls in play. Therefore, a DIPS ERA should be used to estimate pitcher's run-prevention capabilities. Fielding is difficult to evaluate, but the newly developed plus/minus metric provides an objective measure of defense that avoids many of the problems that plague older measures of fielding ability.

Properly quantifying players' on-field contributions to winning is only the first step in valuing players, and there is still a bit more left to be done. Player performance isn't constant over time, it changes in a predictable rise-and-decline pattern due to aging. Valuing players requires knowing not only how good players are now, but how good they will be in the future. The next chapter looks to the past to see how performance tends to change with age. With this information, it's possible to generate long-term projections so that players can be valued over time.

Hot Stove Myth: Some Players are Clutch

Baseball announcers love to praise players for their ability to come through when the game is on the line. The problem is that our identification of players as "clutch" or "chokers" is largely based on inferences drawn from selective sampling of performance. It's true that players have hero and goat moments, but that doesn't mean that players who have them possess some sort of clutch skill that we can count on them to draw upon at the appropriate moment.

Performances in pivotal moments leave lasting memories that we sometimes use to make generalizations about player abilities that are wrong. A walk-off homer may cause us to forget the dozens of other times when player grounded out to end the game in an expected loss. And the pitcher who gave it up maybe never forgiven by fans, even though he's among the league leaders in strikeouts. As a Braves fan, I'll never forget Francisco Cabrera's miraculous single that scored Sid Bream from second base to win the 1992 National League Championship, but I never hope to see the light-hitting Cabrera at the plate or the slow-footed Bream on the base paths when the game is on the line ever again.

Whether some players rise to the moment or shrivel in the spotlight is an empirical question that has been studied by many researchers, and the general conclusion is that players don't seem to have any special clutch ability. For example, statisticians Jim Albert and Jay Bennett find that if any clutch ability exists its impact is small, and it is difficult to identify which players might have clutch skill.[18] Most studies of clutch ability take a set of aggregate performance in clutch and non-clutch situations and compare the outcomes using statistical tools. There is nothing wrong with this method, but it's possible that some of this clutch ability is getting lost in the noise of aggregate data. For example, when we compare hitters in clutch and non-clutch situations, it's difficult to account for the fact that the best pitchers tend to come in the game at that time, and the best players vary from team to team—it's not just pressure that distinguishes the situations. I think it would be better to look at performance at a more granular level to see how players performed in the clutch.

Using a sample of play-by-play data, I estimated the outcome of individual plate appearances while controlling for several potentially influential factors. I looked at three types of outcomes for hitters and pitchers: whether or not the batter gets a hit, whether or not batter gets on-base, and the number total bases the batter advances.[19] As a proxy for any clutch ability that a player might have, I used his performance with runners in scoring position (RISP) during the previous three seasons as an explanatory variable. If clutch performance is a skill, then past performance should be associated with RISP performance in the present. I also controlled for the general ability of the player by including his seasonal overall performance, the quality of the pitcher on the mound, and I identified whether or not the platoon advantage (batter and pitcher have opposite dominant hands) was in effect. The overall performance for hitters and pitchers is measured by batting average for hits, on-base percentage for reaching base, and slugging average for total bases. After accounting for all of these factors, if pitchers or hitters have clutch ability, then past RISP performance should predict present RISP performance.

The results presented in the table below strongly support the hypotheses that neither hitters nor pitchers have clutch ability. The table below reports the estimated impact of each factor on the likelihood of the outcome occurring, where a one-unit change in the predicting variable is associated with an X-unit change in the outcome variable at the average. For example, every one-point (0.001) increase in a batter's batting average is expected to increase a batter's likelihood of getting a hit by 0.00104.

Predicting the Outcome of Plate Appearances

	Hit	*On Base*	*Total Bases*
Hitters			
Past RISP	–0.06162	**0.00018**	0.00012
Batter Performance	**1.04**	**0.98**	**0.93**
Pitcher Performance	**1.152**	**1.031**	**0.983**
Platoon Advantage	**0.014**	**0.040**	**0.039**
Pitchers			
Past RISP	–0.02390	–0.00220	–0.11920
Pitcher Performance	**1.1801**	**0.8815**	**0.9702**
Batter Performance	**1.0148**	**1.0737**	**0.9816**
Platoon Advantage	**0.017**	**0.033**	**0.043**

Bold font indicates that the estimated relationship is statistically significant, meaning that the estimated effect is likely not zero. For the most part, the variables are statistically significant and fit with the general intuition about how they ought to predict the outcomes (e.g., better performance in the past is associated with positive outcomes). But among the clutch variables, in only one case does a player's past clutch performance appear to predict future clutch performance: getting on base for hitter. However, before we declare clutch "on-basing" to be a real skill, we need to look at more than statistical significance.

The estimate shows that every one-unit increase in RISP on-base percentage is associated with a 0.00018 increase in the likelihood of getting on base; thus, a player increasing his RISP on-base percentage by 0.010 (10 "points") increases his on-base probability by 0.0000018. For practical purposes, there is no effect here; especially when compared to the other factors in the model. The performance variables show a nearly one-for-one relationship with outcomes. The platoon advantage predictably increases the likelihood of getting a hit by 1.4 to 1.7 percent and reaching base by 3.3 to 4 percent—that is, 14 to 17 points in batting average and 33 to 40 points of on-base percentage. The expected number of total bases increases between 0.39 and 0.043 bases or 39 to 43 points of slugging average.

Those who wish to cling to the idea that clutch ability exists may identify imperfections in the analysis to justify their continued faith in clutch players. I admit, this study is imperfect, and so are many others that have been done by other researchers who have not found evidence of clutch skill. But, if clutch hitting is something that is so easy for baseball pundits to identify, then why isn't it showing up under a figurative microscope? The sheer number of observations makes statistical significance simple to achieve, yet, past clutch performance does not seem to predict present clutch performance with any reasonable certainty. If clutch ability exists, it is not readily identifiable among players and is, therefore, useless for evaluating players.