By the Numbers

Volume 11, Number 4 The Newsletter of the SABR Statistical Analysis Committee November, 2001

Study

The Reliability of Pitcher Ball-In-Play Average

Willie Runquist

The reliability of a statistic indicates to what extent the value of the statistic is independent of random chance. Here, the author calculates the reliability of hits-per-ball-in-play, providing further evidence of to what extent this statistic is indicative of the skill of the pitcher or defense.

The reliability of a statistic is the proportion of variance (differences) among players which is not due to random error or other factors which balance out between players over a large number of observations. Many baseball statistics are averages over at-bats, plate appearances, or (in the case of pitchers) batters faced, and the outcomes of these instances are all influenced by random error.

In a paper published in BTN (November 1999), the author suggested a means for estimating the reliability of various statistics as R = 1- Var E /Var P, where Var(P) is the variance among the players' values on the statistic and Var(E) is an estimate of the pooled amount of random error in the statistic on each plate appearance, at-bat, or batter faced. Since most statistics represent a combination of various outcomes, the overall reliability of a statistic will be a function of the reliabilities of the individual outcomes.

Of all of the outcomes of a particular batter/pitcher confrontation, strikeouts and walks are the most reliable. Hits (in the form of batting average), on the other hand, is the outcome most subject to random variation. In our previous research, opponents' batting average had very poor reliability, with R ranging from .34 to .52 in various samples of pitchers. On the other hand, the reliability of bases on balls ranged from .50 to .81, and strikeouts were even more reliable, with R being above .80 in all samples.

It should therefore come as no surprise that hits-per-batter-faced should be less reliable when walks, strikeouts, and HBP are removed from the sample of batters faced for each pitcher. The issue, however, is whether all that remains is error variance; i.e. random differences.

With this in mind, we carried out the following analysis on the 50 pitchers who had the most innings pitched in two separate seasons, 1997 and 2000: We computed R for the ratio of hits-to batters-faced, first, with walks, strikeouts, and hit by pitcher included, and, second, with those variables omitted for each group.

The results were almost identical for the two years of data. R with strikeouts, walks, and hit by pitcher included was .61 for 1997 and .59 for 2000. When only batters who put balls in play were considered, however, the reliability was only .12 for 1997 and .11 for 2000. In terms of statistical significance, we computed the common F statistic by dividing Var P by Var E. With all batters faced included, the F ratios were 2.60 and 2.45 for the two seasons, $p < .001$ in both cases. With only balls in play, the ratios were 1.14 and 1.13 , with $p > .25$.

The amount by which R is reduced depends upon the reliability of the components which remain. For example, removing BB reduces R from .62 to .58. Removing SO reduces it to .33. But the combination of removing both BB and SO is devastating. As

In this issue

could be expected, HBP has almost no effect one way or the other.

It would be statistically legitimate, then, to say that in these two cases there is no evidence of a reliable difference between pitchers when only balls in play are considered. To conclude, however, that the differences between pitchers are therefore random is not a legitimate statistical conclusion, since that is tantamount to accepting the null hypothesis.

We still have found that about 10% of the total variance in these samples represents true differences between pitchers. Any reliability greater than 0.00 would be statistically significant with an infinitely large sample. To base a conclusion of randomness on low inter-season correlations suffers from the same interpretive problems. Moreover, such correlations for hits allowed are not impressive even when all batters faced are included.

Note, however, that even if the "true" value of R were to be zero, it does not justify a conclusion that getting batters out is a random process not determined by the pitcher's skill. Rather, the data simply indicates that in the population under study, the pitchers do not differ to any great extent in these skills. We also have to keep in mind that a number of factors are difficult to control. Pitchers may differ in the quality of batters faced and other non-random variables, which are as likely to reduce differences as enhance them.

Willie Runquist, Box 289, Union Bay, BC, Canada, V0R 3B0, willipeg@island.net ♦

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

To join, or for more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. Or write to him at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

E-mail Changes

If you normally receive "By the Numbers" by e-mail, but you found this issue in your physical mailbox instead, it's probably because your e-mail address changed. If you'd like to switch back to an e-mail BTN, please drop me (Phil) a line with your new e-mail address, and I'll switch you back to the electronic version.

Academic Research: Competitive Balance in the Modern Era

Charlie Pavitt

The author summarizes an academic research study on competive balance in the modern era.

This is one in a series of occasional reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at *www.udel.edu/johnc/faculty/pavitt.html*, use it for your research, and let me know what I'm missing.

E. Woodrow Eckard, Free Agency, Competitive Balance, and Diminishing Returns to Pennant Contention, Economic Inquiry, Vol. 39 Number 3, July 2001, pages 430-443

Although the question of competitive balance across teams within each of the major leagues has been studied quite a bit, this piece does make a contribution to our knowledge about it. First, most of the previous work on this issue has looked at competitive balance across the entire twentieth century, with no consistency in findings across studies, whereas this piece concentrates specifically on the past quarter century in order to examine the specific impact of free agency. Second, most previous academic work has based its conclusions on the variation in team W-L Pct. within years (or an analogous measure), with less variation implying that teams are closer together and thus competitive balance is greater. However, the author makes a good argument that the analyst should also consider other indices. One index he uses is the variation in a specific team's W-L Pct. across years, with larger indices meaning that teams vary a lot over years and thus competitive balance is greater. A second additional index is the extent to which league pennant winners differ across years, with "pennant winner" defined as the best league W-L Pct. to allow comparisons before and after divisional play began. Eckard provides a good review of past academic studies but, like most academic baseball researchers, appears unaware of relevant non-academic work, specifically Wood and McCleery (Baseball Analyst 39) and Hadley and Gustafson (By The Numbers Vol. 6 No. 1). This is particularly a problem in the case of Hadley and Gustafson, as they explored the very question Eckard addresses here, finding no change in within-year standard deviations between the 1966-1976 period (the final year of the reserve clause) and the 1982-1992 period.

Splitting periods up a couple of different ways, in order to compensate for the detrimental effects of expansion on competitive balance, Eckard found within-year variation to drop a little for the A.L. and increase a little for the N.L. after free agency began (as Hadley and Gustafson did not distinguish between leagues, the two findings are not necessarily contradictory). However, Eckard found variation in team W-L Pct. across years to rise and concentration of pennant winners to fall after free agency began, both evidence for increased competitive balance.

Now, this thing about "diminishing returns" in the article's title. Eckard proposed that one reason competitive balance still exists in the free-agency era is that fans of generally-successful teams become apathetic to winning, leading to lower attendance, drops in team revenue, and thus less interest in bidding on free agents. In an indirect test of this proposal, Eckard found that teams that remain contenders for several years have indeed suffered attendance decreases over those years. I do however find the proposal rather unconvincing without a demonstration that successful teams do less bidding for free agents over time and, if this is the case, some reason to believe that lower attendance is at its basis rather than, for example, commitment to those players that have contributed to success in the past.

Charlie Pavitt, 812 Carter Road, Rockville, MD, 20852, chazzq@udel.edu ♦

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis - that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any - I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics - confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Jim Box	im.box@duke.edu	Statistics
Keith Carlson	kcarlson2@mindspring.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	HanrahanTJ@navair.navy.mil	Statistics
Keith Karcher	kckarcher@compuserve.com	General
Chris Leach	chrisleach@yahoo.com	General
John Matthew IV	john.matthew@rogers.com	Apostrophes
Duke Rankin	RankinD@montevallo.edu	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

Receive BTN by E-mail

You can help save SABR some money, and me some time, by receiving your copy of *By the Numbers* by e-mail. BTN is sent in Microsoft Word 97 format; if you don’t have Word 97, a free viewer is available at the Microsoft web site (http://support.microsoft.com/support/kb/articles/Q165/9/08.ASP).

To get on the electronic subscription list, send me (Phil Birnbaum) an e-mail at birnbaum@sympatico.ca. If you’re not sure if you can read Word 97 format, just let me know and I’ll send you this issue so you can try

If you don’t have e-mail, don’t worry – you will always be entitled to receive BTN by mail, as usual.

Comments on Methods used in "The New Bill James Historical Baseball Abstract"

Rob Wood

Bill James recently released his new book, "The New Bill James Historical Baseball Abstract." While the book is a significant contribution to the field of baseball analysis, some of James' methods, the author argues, are flawed.

Introduction

The New Bill James Historical Baseball Abstract finally came out in October. It was well worth the wait since James has introduced new methods to evaluate baseball players. And, of course, James is wonderful to read.

His new evaluations are largely based upon a new system he developed called "Win Shares." James' major contribution of this book resides in the research, findings, and insights encompassed in the Win Share formulation.

One unfortunate aspect of the book is that he is waiting until his next book (due out in Spring, 2002) to fully describe the win share system. Nevertheless, I herein would like to comment on several aspects of James' evaluation methods that warrant scrutiny, independent of the details of win shares.

Bill James is one of the pioneers of sabermetrics and I look forward to all of his writings and contributions. As in many disciplines, progress is often made by discussing others' methods, and ultimately leading to greater understanding and refinements. So please consider my critiques in that spirit, recognizing full well that it is always easier to critique than to create.

Active Players

One of the factors James uses to evaluate players is their average number of win shares per season. While I will have more to say about this below, there is an obvious bias this factor introduces. Generally speaking the career path of a player over his career is an upside down U; a player's performance rises in the early part of his career, reaches a peak around ages 27-29, and then declines until he retires. James is one of the analysts who showed how important and prevalent this aging pattern is among baseball players.

Since players' average performance typically declines over the final portion of their careers, evaluations of active players currently "past their peak" will be biased upward. If we re-did our evaluations after they retired, their average win shares per season would be lower.

But, you say, after they are retired their career value aspects will be higher. Perhaps James struck the right balance in evaluating active players? Two comments. First, James states that he tries to be very conservative in evaluating active players. Second, I do not think he found the right balance, as I turn to next.

Balancing Career Value and Peak Value

As mentioned, James attempted to balance career value aspects with peak value aspects. In fact, his system leans more heavily on peak value than on career value, as we will see.

Although the formulas are not yet available, it is clear that win shares are zero-based numbers in one sense. Every player who plays a major league game receives positive win shares. Inside the formulas there is an attempt to measure the player's contributions relative to the league average, I think, but in the end players' win shares always go up over time. Thus, even if a player has a poor season and is in reality hurting the team, he will be credited with a positive number of win shares.

So you can see that win shares may be considered "biased" toward career value at the expense of peak value. James recognizes this and therefore considers several evaluation factors more reflective of the player's peak value and gives them a great deal of weight. I think this is appropriate.

What I think warrants scrutiny is the way James attempts to balance these factors. To give the punch-line at the outset, I think James' evaluation system may be flawed. I am not saying that the balance between career and peak value that James achieves is different than the balance that I personally desire. What I am saying is that James' balance seems to be internally inconsistent.

To set the background: a player who achieves 50 win shares in a season just had one of the greatest seasons of all time (Babe Ruth in 1920 or 1921). A 40 win share season is a superstar season (Willie Mays in 1962 or Willie McCovey in 1969). A 30-win-share season is a very good year (Ernie Banks in 1958 or Billy Williams in 1970) and is likely to receive serious consideration for the league MVP. A 20-win-share season is a good year (Roberto Clemente in 1962 or Tony Perez in 1974). A 10-win-share season is nothing special.

In order to highlight the career vs. peak value issue, let me create two different players and see who James' evaluation system would say is the greater overall player. The table below presents their seasonal win share totals over their respective careers.

For simplicity, assume that each player plays 162 games each season, and that the two players have exactly the same "peak value". In terms of James' evaluation factors, this means that they have the same average win shares in their top 3 seasons and the same average win shares in their best five consecutive seasons.

	Player A	Player B
Year 1	30	30
Year 2	40	40
Year 3	40	40
Year 4	40	40
Year 5	40	40
Year 6	40	40
Year 7	30	30
Year 8		32
Year 9		31
Year 10		15
Career Total	260	338

Player A falls off a cliff after his peak and is out of baseball one year after his last great season; Player B has a more gradual decline, putting in two additional MVP-type seasons, before retiring. As you can tell, I stacked the deck in favor of Player B.

According to the evaluation scheme James uses to summarize a player's "value" (based upon his seasonal win shares), Player A ranks higher than Player B. How can this be? Player B has the exact same career as Player A, except he played 3 more seasons, two of which he contended for the league MVP, say, and the last of which he was, say, a league average player.

Well, the math underlying the evaluation scheme is straightforward. Player B's final sub-par season is more than enough to pull down his average win share per season more than his career win share total goes up! How can that be – that seems impossible? Well, James does not use the player's raw career win share total. He converts the career win share totals into a figure that could be combined with seasonal average win share figures "without dwarfing them." A player's career win share total is harmonized first. James takes the player's career win share total, divides by 10, and then finds the harmonic mean of that number and 25.

I have done lots of arbitrary stuff like that in my own research. However, before I put a lot of credence in formulas that had funky stuff like that going on, I tried to make sure that they did not exhibit any aberrant behavior. I think James' "harmonizing" scheme probably does a good job in most cases. However, it seems that he did not think through all the ramifications.

Don't jump to the conclusion that the anomaly in my example is really a reflection that a 15 win share season is implicitly considered a poor season (and therefore should rightly pull down the evaluation of the player). In fact, James' evaluation of Player A would actually go down when he plays at an MVP-type level in his final season. So even a 30 win share season can pull down a player's rating!

I know this is merely a cooked up example, but I think my point is generally valid. James has used a conversion scheme that combines absolutes (career totals) and relatives (averages) in a way that is very difficult to disentangle. In such a situation, we should not be surprised that he did not do a perfect job, since there may in fact be no perfect solution.

James has publicly stated that he did not want to maintain separate rankings by career value and peak value. So he was forced to come up with a method to combine the two aspects that would guide him as he presented the player comments in terms of the rankings of the all-time best players at each position. All I am saying is that the method he chose is not the one I would have chosen.

In private communication James has told me that he does not want to get drawn into the type of argument I present above. To James, the players should be the focus, not the specific formulas used in the ratings. Admittedly, in this section and throughout the article, I am focusing on the formulas.

Timeline Adjustment

James believes that the quality in major league baseball has improved steadily over time. I concur. As a reflection of the improved quality, James believes that earlier players were able to dominate their leagues more than modern players. Again, I concur.

Let me tread carefully here. There are basically two approaches to proceed given these beliefs. One way is to work within the standard sabermetric framework and derive an "internal" adjustment mechanism that attempts to account for the improved quality of play. The other way is to step outside the standard sabermetric framework and make an "external" adjustment. I prefer the first approach; James chooses the second approach.

The first approach attempts to stay within the boundaries of our standard sabermetric relativistic measures. Here is the thinking behind this approach. A player's job is to help his team win games, pennants, world series, etc., so we evaluate players based upon how well they helped their teams win games, etc. Value, then, is necessarily a contemporaneous concept. The belief that players today are better than players were 125 years ago is completely irrelevant to a discussion of the impact a player 125 years ago had on his team winning games 125 years ago. Right?

Yes, but it is relevant to think about any biases in our evaluation methods. The implicit assumption behind our relativistic methods is that a player who is 20% better than his league average 125 years ago would perform 20% better than league average today (and vice versa). I share the belief that this assumption is probably flawed due to the reasons James describes related to improved quality. Of course, many of our methods of evaluation use a player's performance relative to the league average as a reflection of his contributions to his team winning games. So I do think it is appropriate to make an adjustment for this "sparseness" effect (125 years ago the MLB population was very "sparse" whereas today it is very "tight").

Remember that a strict interpretation of this approach maintains that the *only* relevant effect of improved quality of play is the change in the ability of players yesteryear to more easily stand out from their colleagues (league average). I will come back to this below.

I researched this sparseness issue earlier this year. I calculated the standard deviation of all regulars' OPS in each league from 1871-2000. The sparseness conjecture would suggest that this standard deviation has steadily declined over time. Indeed it has. However, the decline was steep in the early days, and then has tapered off since 1900 or so. To get a sense of the decline, I estimated a curve to fit the data.

The best fit curve is a parabola that declines sharply in the early part of the curve, and then declines only gradually over the rest of the curve. The parabola has equation:

`Lg Std Dev = (0.234536) * ((Year-1870)^-0.11316)` where ^ is the power operator.

The data and details are available upon request. Suffice it to say here that we can use this curve to monitor the sparseness effect over time. For example, comparing 2000 to 1871 we see there is a significant effect. Players in 1871 were much more able to stand apart from their league average than were players in 2000. The data indicates that 1871 players got an artificial boost of 73% due to the large degree of sparseness back then. Not to take this into account would lead us to evaluate 1871 stars as far greater than 2000 stars, due entirely to the differing degrees of sparseness exhibited in the respective professional baseball populations.

But sparseness is seen to shrink quickly. By 1900 sparseness has shrunk by 47% relative to 1871. Stated another way, comparing 2000 players to 1900 players, the data indicates that 1900 players received an artificial boost of only 18% due to the relative sparseness. Comparing 2000 to 1950 shows that 1950 players received an artificial boost of about 5 percent. By modern times, sparseness is barely changing from year to year; comparing 2000 to 1990 shows that 1990 players received an artificial boost of less than 1 percent.

Now let's return to Bill James' time line adjustment. As I said above, James uses an "external" adjustment. Players receive an amount added to their win share for each season in the amount of their year of birth minus 1800, divided by 10. This approach implicitly assumes a regular effect throughout baseball history. Let's see the impact on a typical (before the adjustment) 10-point season throughout history. Comparing to a 2000 season, an 1871 season would be deemed artificially 80% better, a 1900 season would be 59% better, a 1950 season would be 23% better, and a 1990 season would be 4% better.

Remember that my sparseness analysis found these percents to be 73%, 18%, 5%, and 1%, respectively. So the external approach James uses leads to roughly the same overall magnitude of the effect of improved quality as compared to the internal approach I advocate. However, compared to the sparseness effect I found, by using his regular decline measure, James' approach "penalizes" players from 1890-1925 by about 30%, players from 1926-1950 by about 20%, players from 1951-1975 by about 10%, and players from 1976-1990 by about 5%.

The implications of James' time line adjustment appear to be significant. Just to mention a few of the numerous cases, James ranks Craig Biggio ahead of Nap Lajoie, Sal Bando ahead of Jimmy Collins, Barry Larkin ahead of George Davis, Jim Fregosi ahead of Bill Dahlen, and Tim Raines ahead of Ed Delahanty.

To recap, James saw a need to make an adjustment for the improved quality of play throughout baseball history. He chose to make an "ad hoc" adjustment external to his evaluation framework. I cannot quibble with that approach, since by its nature it is a reflection of a personal belief.

On the other hand, I try to pay strict adherence to our standard sabermetric relativistic methods. Internal to this framework, I have found that improved quality of play over time has led to biases in our standard estimates of a player's value due to the ease by which yesteryear's stars could stand out from their colleagues. By performing an analysis of this sparseness issue, I developed a time line adjustment internally consistent with our standard sabermetric formulas.

When the time line adjustments from the two different approaches are compared, the overall magnitudes are similar, but the amount of the adjustments are significantly different since 1900. The consequence of James' approach appears to me to be an undervaluing of olden stars to the benefit of modern stars.

19th Century Innings Pitched

When James first developed his win share system, he found that all the top seasons were by pitchers from the 1876-1892 era. Guys like Old Hoss Radbourne with 679 innings and 60 wins in 1884. Pitchers in those days did pitch a ton of innings, no question. James was faced with a quandary. What can he do to "fix this?"

As we all know, the load teams have asked their starting pitchers to shoulder has steadily declined as baseball conditions and strategies have evolved. The introduction of overhand pitching, the curveball and other stressful pitches, the 60-foot pitching distance, the lively ball, 5-man rotations, specialized relief pitchers, pitch counts, etc., have made teams choose to pitch their top starting pitchers fewer innings in virtually every succeeding generation throughout baseball history.

James' solution was to discount all pitchers' win shares in seasons before 1893 by 50%. I am confident that this is equivalent to dividing all these pitchers' innings by 2 in those seasons. Now, I am all for adjusting old-time pitchers' innings. However, what I would have preferred is a mini-study to determine a reasonable adjustment mechanism. The ideal mini-study would include a "model" of what is going on, an analysis to estimate the parameters of the model, a conversion of the parameters into adjustment factors, and then a review of the results of the adjustments to see if they are reasonable.

Several years ago I looked into this issue and thought that a reasonable thing to do would be to adjust a pitcher's innings based upon the average of the top three innings pitched totals in the league divided by the historical average of the top three innings pitched totals. I am not saying that this is necessarily a great method (though I think it may be at least a very good one). The advantage is that there is something behind my adjustments, and that the adjustments are fluid and automatically react to changing conditions. James' method is presented as "I have divided pre-1893 innings by two since then I like the results better."

As above, James basically skipped the intermediate steps of the ideal approach and jumped directly to an adjustment factor. The advantage of the approach I am advocating is that linking adjustments to a model allows the audience (and analyst himself) to question the assumptions, the data, and the analytical methods, besides simply questioning the results. The method of jumping directly to the answer works if the analyst has phenomenal intuition, but nobody's intuition is so good that they can always come up with the right answer.

These "post-win share" adjustments are the not main focus of James' evaluation system, Win Shares are. So it may be that I am focusing too much attention on the "post" evaluation scheme.

Impact of Designated Hitter

This isn't really a critique of James' evaluation system, per se, though I thought I would include one more issue here. James ranks Craig Biggio as the 35th greatest player ever. Everyone is entitled to his own opinion; I just want to make sure that the evidence on which the opinion rests is valid. In making the case for Biggio as "the best player in major league baseball today", James compares him to Ken Griffey Jr.

In particular he compares their 1998 seasons. It turns out that Biggio created slightly more runs than did Junior that year. James does not let it rest there, though. He says that this is actually strong evidence in Biggio's favor since he did this in a league (NL) in which the average team scored 4.60 runs per game, whereas Junior played in a league (AL) in which the average team scored 5.01 runs per game, almost 10% more.

But the AL used the DH that year while the NL forced weak-hitting pitchers to bat. It is easy to show that the DH factor can account for the 10% difference in team runs. The question then is whether it is appropriate to compare the players' runs created to their respective league average runs scored when one player played in the NL and the other played in the AL-DH. It turns out that it is not appropriate to do so. I led a thread on SABR-L in late October through early November that convincingly demonstrated that not taking into account the DH would unduly penalize AL hitters by about 10%. (See the SABR-L archives for the details.)

Either Bill James has a blind spot when it comes to evaluating AL hitters in the presence of the DH or he wrote the Biggio-Junior comment without thinking clearly about the issue. Either way, I don't want the comment to go unquestioned. Perhaps more importantly, I am not sure if James' win share methods properly take into account the impact of the DH. I suppose they do, but we won't be sure until we see the formulas.

Conclusions

The New Bill James Historical Abstract is a fantastic book and another significant contribution from one of the giants in sabermetrics. In this article I tried to raise some concerns about James' methods. Where I could I also tried to indicate what improvements I would recommend that James (and others) include in subsequent research. I hope my comments are taken in the spirit in which they are intended. As in most disciplines, commenting upon the methods of others' research is an important part of advancing sabermetrics.

Rob Wood, 2101 California St. #224, Mountain View, CA, 94040-1686, robert_wood@standardandpoors.com ♦

Book Reviews Wanted

Every year, a number of books and magazines are published with a Sabermetric slant. Many of our members have never heard of them. Our committee members would like very much to hear when this kind of stuff comes out.

If you own a copy of any baseball book of interest, we'd welcome a summary or a full-length review. The only restriction, please: the book should have, or claim to have, some Sabermetric content.

For a sample of what we're looking for, check out David Shiner's review three issues ago, or Gabe Costa's review in the issue before that.

Send reviews to the usual place (see "Submissions" elsewhere in this issue). Drop me a line if you want to make sure no other member is reviewing the same publication, although multiple reviews of the same book are welcome, particularly for major works. Let me know which book you're doing, so I don't assign the same book twice.

And if you're an author, and you'd like to offer a review copy, let me know – I'll find you a willing reviewer.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work (but no death threats, please) are all welcome.

Articles should be submitted in electronic form, either by e-mail or on PC-readable floppy disk. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

I usually edit for spelling and grammar. (But if you want to make my life a bit easier: please, use two spaces after the period in a sentence. Everything else is pretty easy to fix.)

If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does, and please include your byline at the end with your address (see the end of any article this issue).

Deadlines: January 24, April 24, July 24, and October 24, for issues of February, May, August, and November, respectively.

I will acknowledge all articles within three days of receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to:
Phil Birnbaum
18 Deerfield Dr. #608, Nepean, Ontario, Canada, K2G 4L1
birnbaum@sympatico.ca

More On the Relief Pitcher's ERA Advantage

Phil Birnbaum

For various reasons, a reliever should have a lower ERA than the equivalent starting pitcher. How much is that ERA advantage? Here, the author updates a previous study that attempted to answer this question.

In the February, 1988 *Baseball Analyst*, I presented a study estimating the relief pitcher's inherent ERA advantage over a starter. Since then, I've discovered some omissions in my study, which I re-present here.

Before I start, let me make explicit what I'm trying to measure, which is this: the amount by which a relief pitcher's ERA would tend to be lower than that of a starter of exactly the same talent. Or, expressed another way, if Joe Average starter were to be moved to the bullpen, by how much would we expect his ERA to go down assuming that his talent level does not change?

There are three factors that contribute to such an ERA difference: (a) the tendency of relievers to enter a game with fewer than three outs remaining in the inning; (b) games ending in the middle of innings; and (c) differences in the frequency of intentional walks between starters and relievers.

We take each of the three factors in order.

Entering the game with men already out

I discussed this in depth in my previous article, but briefly, a relief pitcher has an advantage over a starter when he enters a game with men already out. Consider two pitchers; one comes into three games with two out in the 7th and pitches the third out, while another comes in once at the beginning of the inning and pitches the entire 7th. Now, suppose both pitchers give up two singles before recording each out. In each of his three appearances, the first pitcher gives up two singles, then gets the next batter out, gets out of the inning, and has no runs charged to him. But the second pitcher has given up six singles before the third out, and has probably given up a few runs.

Each pitcher has given up six singles in 1.0 innings of work. But the first pitcher has an ERA of zero, while the second has an ERA in the double digits.

Pitchers who often come in part way through an inning will have an ERA advantage over those who don't, because they get the bases cleared more frequently. And who comes in most in the middle of an inning? Relievers, obviously.

How much is the reliever's advantage in this situation? In the AL, about 14 points (that is, 0.14 runs per nine innings). For the reasoning and calculations behind this number, see my other article.

Games ending with runners on base in the middle of an inning

In the bottom of the 9th, with the score tied, the home team puts runners on first and second with nobody out. The next batter hits a double, scoring a runner, and leaving runners on second and third. But, as soon as that run scores, the game ends, and those runners on second and third never get a chance to score, thus saving the pitcher a couple of earned runs that probably would have scored had the inning continued. Who saves those runs? Relievers, overwhelmingly; very rarely is the starter allowed to stay in the game after putting the winning run on in the 9th or later.

So how much is the advantage? Well, let's start with a particular case, the one in the previous paragraph. The game ended with no outs and runners on 2nd and 3rd. From the famous Run Potential Table (I'm using Sabermetric Review, July '87, page 10), second and third with no outs is worth 1.972 runs; if nobody was on, .518 runs would be expected to score with no outs. So, the pitcher saved himself (1.972 - .518) = 1.454 runs. That's a considerable number considering 1.454 runs makes a difference of about .14 in ERA for a stopper who pitches 100 innings a season.

To get a figure for the league-wide reliever advantage, we've got to add up every time this happens in a season and how much each one is worth. Using Project Scoresheet data for the 1987 American League, I got the following figures:

```
49.3 "fewer than 3 outs when winning run scored" runs were left on base.
```

Of those,

```
45.1 runs were the responsibility of a reliever;
 2.7 runs were the responsibility of the starter; and
 1.5 runs belonged to both.
```

To simplify matters, I proportioned the "both" figure (which arose when some of the runners left on base had been put on by the starter, and others by the reliever) so as to get final numbers of

```
46 runs saved by the reliever, and
 3 runs saved by the starter.
```

Assuming starters pitch 13697 innings per year, and relievers 6487 (these were the actual 1985 AL figures), this gives us "savings" figures of

```
.064 runs / 9 innings for relievers, and
.001 runs / 9 innings for starters,
```

or an advantage to the relievers of about 6 points.

Intentional Walks

Although relievers have an ERA advantage over starters for the reasons discussed in the above two sections, they also have a *disadvantage* because they issue many more intentional walks, which increase run potential and actual runs. For example, walking a batter with a runner on third and two outs to get to a weaker hitter will probably give the defensive team a better chance to get out of the inning without runs scoring. But, if the batter does manage to hit a double or a home run, you've got an extra run scoring that wouldn't have otherwise. The question is, how many extra runs is that intentional walk worth?

I'm not going to give a completely accurate answer to that, because I'm going to simplify the question a bit, by making these assumptions:

1. Every hitter on a team hits equally well;
2. Every inning is played out to completion.

Both these assumptions, are, of course, not correct, and both would lead to too-high a calculated value for the intentional walk, because (1) stronger batters are often walked to get to weaker batters, producing fewer runs than would result if the batters had the same ability, and (2) lots of intentional walks occur with the winning run already on base in the bottom of a late inning, which means that the intentional walks can't score at all unless a home run is hit.

For objection (1), I offer no excuse except that the calculation would be impossible otherwise, and for (2), this is the proper assumption *for this study* because we have already corrected for runs which never score in the previous section.

So back to the original question: how much is an IBB worth?

The answer can again be found from Run Potential tables. For example: with two outs and a runner on third, the run potential is .389. After the IBB, we have two outs with first and third occupied, for a run potential of .532. The walk was thus worth .143 runs. If we take every intentional walk, and average the difference in run potential resulting from each (I again used Project Scoresheet 1987 AL data), we get the result that *an intentional walk is worth 0.162 runs.*

Now we need to know how many more IBB's relievers issue than starters. The figures for 1987 are as follows (except innings pitched, which are 1984):

```
Starters:     160 IBB in 13697 IP        .105 IBB / 9 IP
Relievers:    320 IBB in  6487 IP        .444 IBB / 9 IP
------------------------------------------------------------------------
Difference:                              .339 IBB / 9 IP
```

And since an IBB is worth 0.162 runs, the relievers ERA *disadvantage* is 0.339 * 0.162 = 0.55 runs / 9 IP, or about *5.5 points.*

Subtracting this disadvantage from the advantages calculated in parts one and two, gives *a total relief pitcher advantage of 15 points for the American League*. Dividing by 1.1 to get earned runs gives an advantage of *about .135 in ERA.*

`In the AL, the relief pitcher has an ERA advantage of about .135.`

Comparing the Two Leagues

So, in the AL, the relief pitcher's ERA advantage is about 13 points. How does the NL advantage compare to this?

1. The reliever's advantage from entering the games in mid-inning is only 8 points in the NL. That's because relievers frequently come in at the beginning of an inning, mostly because pitchers are often removed between innings for pinch-hitters, rather than exclusively in the middle of innings when they get into trouble.

2. If we assume that the pattern of winning-runs-scoring-with-less-than-three-outs is the same in the National League, we can calculate (using NL run-potentials, but the same AL play-by-play data and innings-pitched totals) that the relief pitcher's advantage is about 5.6 points.

3. In 1985, NL pitchers issued 60% more intentional walks than AL pitchers (.4 per game NL, .25 per game AL). But, using NL run potentials, a National League intentional walk is worth only .128 runs, or 79% as much as in the AL. So, the relief pitcher disadvantage is 1.6 * .79, or 1.26, times as much as in the AL. That works out to .07.

So, the total NL relief pitchers advantage is .066 runs per 9 innings, or, dividing by 1.1 to adjust for unearned runs, .06, or 6 points, or less than half the AL figure.

`In the NL, the relief pitcher has an ERA advantage of about .06.`

Comparison to the 1977 Bill James study

In the 1977 *Baseball Research Journal*, a study by Bill James concluded that the relief pitcher's ERA advantage under National League conditions (No DH) is about 20 points. This study, on the other hand, gives a result of 6 points. Why the difference?

Well, Bill's study and mine measure different things. Mine, as I mentioned in the introduction, measures the advantage of a reliever against a starter of equal *talent*. Bill's, on the other hand, measured the reliever's advantage versus a starter of equivalent *statistics*.

What Bill did was find groups of pitchers with similar rates of hits per game and walks per game, then divide each group into starters and relievers and compare their ERA's. But when you compare a starter and reliever with an equal number of walks per game, the reliever's walks will include many more intentional passes than the starter's. Since an IBB is worth much less, in terms of runs, than an unintentional walk, this gives an artificial advantage to the reliever in the comparison. Also, since Bill did not adjust for relievers' inherent tendency to give up more intentional walks than starters, intentional walks wind up *adding* to the advantage figure rather than reducing it.

If we were to use Bill's method of comparison for this study, what would we get as the advantage of intentional walks? If an approximately average starter and reliever give up an equal number of walks per game, the reliever will have the "advantage" of .542 more of his walks/9 innings being intentional (NL conditions). As calculated earlier in this article, an intentional walk is worth .128 runs, while an unintentional walk (calculated the same way) is worth .297. Thus, the reliever "saves" .542 * (297 - .128) = .092 runs per 9 IP. Adding to this the .08 and the .056 caused by entering in mid-inning and leaving men on at end-of-game respectively, gives us an advantage of .228. Dividing by 1.1 for earned runs gives .207, which is comparable to Bill's .20.

Phil Birnbaum, 18 Deerfield Dr. #608, Nepean, Ontario, Canada, K2G 4L1, birnbaum@sympatico.ca ♦