By the Numbers

Volume 11, Number 3 The Newsletter of the SABR Statistical Analysis Committee August, 2001

Summary

Academic Research: Free Agent Performance

Charlie Pavitt

The author summarizes academic research on whether signing a long-term contract affects a free agent's performance, and on ranking Japanese ballplayers in light of an update to a classic performance stat.

These are two 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.

In this issue

Ted S. Sturman and Ryan Thibodeau, Performance-Undermining Effects of Baseball Free Agent Contracts, Journal of Sport and Exercise Psychology, Vol. 23, 2001, pages 23-36

Do long-term contracts lead to motivational losses and performance decrements among established players? Sturman and Thibodeau think so. They examined year-to-year performance changes for 33 free agents who signed contracts lasting at least two years that provided at least a 30 percent annual raise (as calculated by average annual salary over the contract length) up to at least one million dollars. Of the 33 players, 21 switched teams and 12 did not. The performance data consisted of batting average, home runs, RBI, and stolen bases for each of the two years before and two years after the contract signing. Basically, what they claimed to find was that performance in all four indices decreased the year after the contract signing, with the percentage of the raise associated with the size of the decrement for BA (they did not examine the other three measures in this regard). All of this provides evidence that player motivation may have been lessened that year by the substantial raise in the new contract. Further, with the exception of SBs, performance rebounded the second year after the contract signing, implying that motivation loss was for one year only.

- As I was reading this, I began to note my objections to the authors' basic claim. To their credit, in further analyses (limited to BA only) they attempted to answer these objections, with different degrees of success for each. Let me go through each in turn.

- In their basic analysis, they did not distinguish between players who did and did not switch teams. Players who switch teams could have one-year decrements due to adjustment rather than motivation decrement. In subsequent analysis, the authors found a BA decline for both sub-samples, although it was greater for those who switched teams.

- Most of the players in the sample were in their 30s, so that a performance decrement across the sample would be expected. However, this does not explain why BA went back up in the second year after the contract signing, and the authors found younger players to have larger declines in BA than older, contrary to this objection.

- Simply said, regression to the mean. If players had a particularly good year before their contract ran out, they would be more likely to get the 30% raise the authors required for inclusion in the study, and we would expect performance to return to normal the year after. The authors' work is susceptible to this problem. They did find BA declines from first pre-contract season to first post-contract season to be greater for players with higher pre-contract

- season BAs, which they admit is evidence for regression to the mean, but then present other analyses of BAs, which they claim to support the motivation change hypothesis instead. Indeed, average BA immediately after the new contract (.268) was substantially lower than that for the other three years in the study (.282, .287, .283). However, average HRs and RBIs were substantially lower two years before the new contract (16.8 and 68.8, respectively) than one year (18.7 and 76.6) before, and those for the year after the new contract (16.6 and 65.9) were comparable to those two years before. For these two performance indices, the authors' results are clearly due to regression to the mean and thus spurious.

Ano Katsinori, Modified Offensive Earned Run Average with Steal Effect for Baseball, Applied Mathematics and Computation, No. 120, 2001, pages 279-288

This one is for specialized interests only. In a classic old piece, Cover and Keilers (Operations Research, Vol. 25, 1977, pages 729-740) came up with a method for estimating the expected number of runs an offensive player would score if he batted in all nine lineup positions.

They did not include stolen bases and caught stealing in their method; the present author adds it, and shows how the change affected the resulting offensive ERA for ten Japanese players during 1996 and for the top ten lifetime. For the trivia fans among you: for lifetime, Sadaharu Oh was number one both with and without the SB adjustment at more than 11 runs a game. As for 1996, Ichiro, then of the Orix Blue Wave, was having a normal season for him (.356 BA, 24 2Bs, 16 HRs, 65 BBs), with the adjustment (35 SB versus 3 CS) raising him from 8.88 to 9.10. Among the players listed, this still put him behind two others, also in the midst of normal seasons; outfielder Hideki Matsui of the Yomiuri Giants at 10.33 (.314 BA, 34 2Bs, 38 HRs, 75 BBs) and third baseman Akira Eto of the Hiroshima Toyo Carp at 9.49 (.314 BA, 19 2Bs, 32 HRs, 80 BBs).

Note

Thanks to the organizational work of Larry McCray and Stephen Lyman, a group of about eight Washington, D.C.-area sabermetricians have begun meeting every two months or so for the purpose of discussing our work. At the first meeting, Stephen described his soon-to-be-published (and then reviewed here) work on elbow and shoulder pain in Little League pitchers, and Clay Davenport talked about his inter-league comparisons of offensive performance. At the second meeting, David Tate presented his studies of ideal lineup placement for given offensive players, and we batted around ideas for an analysis I will be doing of week-to-week consistency in offensive performance. Each meeting was intended to last two hours, but went on for a third. Bill James once wrote that the best thing about the rise in interest in statistical baseball research was the friendships that formed out of it, and I can see that beginning to happen here. I am describing this happy development in the hopes that similar groups spring up in other areas. All it takes is for someone to take the initiative to contact other members of our division in their area (as listed in the Membership Directory or their local chapter's listings) and set up a meeting. A note of caution: we have found it useful to schedule an agenda of research to be discussed to keep the meeting from degenerating into a discussion of trivia (or as Larry puts it, "my player is better than yours").

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

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.

Roberto Alomar and Range Factor

Duke Rankin

Does second baseman Roberto Alomar deserve his stellar defensive reputation? The author examines Alomar's career range factors as evidence of his skills.

Introduction

The conventional wisdom concerning Roberto Alomar is that he is a great second baseman with great range. This has been the conventional wisdom since Alomar's rookie season; Hanke (1989) called Alomar, "a great defensive player with phenomenal range (p. 196)." As the Gold Gloves have accumulated, however, even Alomar's proponents have recognized that his phenomenal range hasn't always translated into phenomenal range factors (hereafter abbreviated RF).[1]

This recently drew the attention of ESPN.com columnist Rob Neyer (2001). To analyze Alomar's range, Neyer compared Alomar's RF with his backups in Toronto and Baltimore/Cleveland. In Toronto, Alomar had a slightly lower RF (4.84) than his backups (4.99).[2] In Baltimore/Cleveland, Alomar had a higher RF (5.12) than his backups (4.60).[3] According to Neyer, the RF difference in Baltimore/Cleveland translates into 80 plays per 154 games, a "huge, titanic" difference. Neyer concludes "Alomar is an outstanding second baseman."

I do not find Neyer's (2000) arguments comparing Alomar to his substitutes convincing. The first -- Toronto had better defenders than Alomar riding its bench -- argues against Alomar's range. Think about it: did Pittsburgh ever have two guys on its bench better at turning the double play than Mazeroski? Of course not. The reverse argument -- Alomar has been better than his backups in Cleveland and Baltimore -- does not argue strongly for Alomar's range: Alomar should be better than his backups, his backups were miserable second basemen, and it's easy to look good when you compare yourself to poor ballplayers. So just how good is Roberto Alomar's range?

Table 1 – Range factors [RFs] for Roberto Alomar in comparison to league-leading RFs

Year	Team	[n]	Alomar RF	Rank	League Leader	RF
1988	SD	12	5.44	2	Gant, Atl	5.52
1989	SD	12	5.18	2	Oquendo, StL	5.42
1990	SD	12	5.13	1	Alomar, SD	5.13
1991	Tor	12	4.88	5	Sojo, Ana	5.18
1992	Tor	13	4.43	13	Baerga, Cle	5.47
1993	Tor	13	4.63	11	Baerga, Cle	5.38
1994	Tor	11	4.25	8	Reed, Mil	5.51
1995	Tor	12	4.98	4	Alicea, Bos	5.18
1996	Bal	12	5.10	4	Vina, Mil	5.43
1997	Bal	10	4.61	4	Spezio, Oak	4.76
1998	Bal	12	4.85	2	Easley, Det	5.17
1999	Cle	13	4.71	10	Cairo, TB	5.36
2000	Cle	11	5.01	7	Velarde, Oak	5.61

where [n] = the number of starting secondbasemen [≥ 90 games] in the league.

Methods

I analyzed Alomar's range factors in two contexts. First, I compared Alomar's unadjusted RF to his peers by ranking the seasonal RFs of all starting major-league second basemen during Alomar's 13-year major-league career. Second, I ranked Alomar's seasonal RFs after adjusting for two potential effects on range factor: turf and balls-in-play.

To determine the effect of turf on RF, I divided all AL starting second baseman seasons between 1991 and 2000 into two categories: turf parks (Toronto, Minnesota, Seattle, Kansas City) and grass parks (all others; data from STATS Inc, 2000), then compared the following

[1] Range factor (RF) is traditionally defined as chances (assists plus putouts) divided by games played. Neyer (2201) apparently reports RF values based on innings played. I have used the more traditional RF values based on games played throughout this study.

[2] According to Neyer (2001), Alomar's backups in Toronto were primarily Domingo Cedeno and Alfredo Griffin.

[3] Neyer (2001) does not identify Alomar's backups in Baltimore/Cleveland. During Alomar's tenure in Baltimore/Cleveland (1996 - 2000), eight men have played at least 7 games at second base on the same seasonal roster as Roberto Alomar: Billy Ripken, Jeff Huson, Manny Alexander, Jeff Reboulet, Aaron Ledesma, Enrique Wilson, Bill Selby and John McDonald. Their career fielding totals at second base through year 2000: 1528 games, 4.03 RF, .985 FP.

variables using two-tailed t-tests: number of games played per season, fielding average, double plays per game, and range factor. Within range factor, I also examined assists and putouts. Putouts are more likely to be fly balls and therefore less likely to exhibit a turf effect than assists, which are more likely to be ground balls.

To test the effect of balls-in-play (hereafter abbreviated BIPs), I used team pitching statistics to estimate BIPs for all starting second baseman seasons during Alomar's 13-year major league career.[4] Second basemen playing behind pitching staffs that allowed above-average BIPs would be expected to make more than an average number of defensive plays, but these second basemen must share the extra plays with other defenders. Because second basemen make approximately 15% of their team's defensive plays, I multiplied team BIPs by 0.15, then divided this seasonal value by 162 games to estimate BIPs per game.[5]

I analyzed only starting second basemen, defined as those players appearing in at least 90 games (or 55% of the schedule for short seasons) at second base. If a team did not have a player with 90 games at second base, the position was considered vacant and excluded from the study. No team had more than one player with 90 games at second base in any given season. All data were downloaded from *www.baseball-reference.com*.

Table 2 – Range factors (RFs) for Roberto Alomar, adjusted for turf (a unilateral 0.172 RF bonus for turf seasons) and for balls-in-play (BIPs, as determined on a seasonal basis using team pitching statistics)

Year	Team	[n]	Turf-adjusted RF	Rank	BIP-adjusted RF	Rank
1988	SD	12	5.44	2	5.46+	1
1989	SD	12	5.18	2	5.19+	2
1990	SD	12	5.13	2	5.08	2
1991	Tor	12	5.05*	4	5.01+	5
1992	Tor	13	4.60*	13	4.58+	13
1993	Tor	13	4.80*	9	4.75+	6
1994	Tor	11	4.42*	8	4.38+	8
1995	Tor	12	5.15*	2	4.98	4
1996	Bal	12	5.10	4	5.06	4
1997	Bal	10	4.61	6	4.75+	1
1998	Bal	12	4.85	3	4.67	6
1999	Cle	13	4.71	9	4.82+	7
2000	Cle	11	4.70	8	4.90+	3

where [n] = the number of starting secondbasemen in the league,
[*] = RF affected by the turf adjustment, and
[+] = a BIP adjustment that increased Alomar's RF.

Results – Unadjusted RF

Over the course of his 13 years in the majors, Alomar's unadjusted RF in comparison to his peers can be divided into roughly three periods: the San Diego years, the Toronto years, and the Baltimore/Cleveland years (Table 1).

Alomar had excellent range in San Diego. After ranking second in 1988 and 1989, Alomar led the NL in RF in 1990. It was close; Alomar's RF of 5.1314 was one chance better than Robbie Thompson's RF of 5.1268. In addition, Alomar's 1990 RF was the second lowest league-leading total in the 13-year sample, and ranked behind AL second basemen Johnny Ray (RF = 5.36) and Harold Reynolds (5.18).

Alomar's RF in Toronto was clearly below average. In 1992, Alomar ranked last in the AL, trailing the league leader, Carlos Baerga, by 168 plays per 162 game schedule. Alomar would have ranked last in 1993 except for Torey Lovullo and Bill Spiers -- two men who barely played the 90 games necessary to qualify as starters, and neither of whom would be full-time second basemen again. In 1994, Alomar edged Joey Cora and Brent Gates by a combined four chances to rank 8th in the league. Only the retiring Harold Reynolds was clearly worse.

After Toronto, Alomar's RF ranks improve. Alomar had his best AL season in 1998, ranking second in the league. Since moving to Cleveland, Alomar's RF has ranked 10th and 7th in the AL.

Over his 13-year career, Alomar's RF has ranked, on average, 5.54 in a field of 12. Over his 10 years in the AL, Alomar has ranked, on average, 6.70 in a field of 12. In comparison, the highest mean ranks for AL second basemen between 1991 and 2000 (n > 3 seasons as a starter) belong to Jody Reed (mean RF rank = 2.3, n = 3), Carlos Baerga (2.40, n = 5), Luis Alicea (4.0, n = 3) and Harold Reynolds (5.25, n = 4 including the disastrous 1994 season).

[4] I estimated BIPs by multiplying team innings by three and adding hits allowed, then subtracting strikeouts and homeruns allowed.

[5] BIPs per game are equivalent to RF, but the sign must be changed before making the adjustment: BIPs below league average are added to RF values, while BIPs above league average are subtracted from RF values.

Adjusted RF

Unadjusted RF may not be the proper measure of Alomar's true range. Some possible adjustments:

(i) *Turf*. Alomar played on turf in Toronto, and this may have diminished his RF. Of the four variables tested, turf significantly affected only range factor ($p = 0.03$; t-test); games per season ($p = 0.36$), fielding average ($p = 0.40$), and double plays per game ($p = 0.53$) were not significantly different. On average, turf decreased RF by 0.172 chances per game. The effect was more pronounced on assists: turf decreased assists by 0.105 per game, putouts by 0.067 per game.

One way to compensate for turf is to unilaterally add the turf affect (RF = 0.172) to the RF values of turf second basemen, creating a "grass equivalent" RF value. This does increase Alomar's RF in Toronto relative to the league (Table 2). Unfortunately, Alomar has played eight of his 13 seasons on grass, and the turf effect lowers his relative standing during his grass years. Alomar's mean RF rank (5.54) is the same with or without the turf adjustment. The turf adjustment, however, gives Jose Lind the 1990 NL RF title, thereby depriving Alomar of his only league-leading total.

(ii) *Balls in play*. Alomar's pitchers may have allowed fewer balls in play, affording Alomar fewer fielding opportunities. The BIP adjustment attempts to answer the following question: what would Alomar's RF be if he played behind a pitching staff that allowed an average number of balls in play? BIP adjustments can be dramatic. In 1997, for example, Oakland pitchers allowed 279 BIPs above the league average, while Seattle pitchers allowed 198 BIPs below the league average, a difference that could account for an RF difference of 0.44 chances per game for a second baseman.

Alomar has played behind pitching staffs that allowed lower-than-league-average BIPs in nine of his 13 seasons. Because Alomar usually had fewer defensive opportunities than the average second baseman, the BIP adjustment usually improves Alomar's rank RF (Table 2). In 2000, for example, Alomar ranked 7th in unadjusted RF, but 3rd in BIP-adjusted RF. The BIP adjustment also gives Alomar league-leading RFs in 1988 and 1997. Over his 10 year AL career, Alomar's mean RF rank falls from 6.70 to 5.70 in a field of 12, but he remains fifth in mean RF rank among AL second basemen of the period, trailing Reed (mean RF rank = 2.00), Baerga (2.60), Alicea (4.33), and Reynolds (5.25).

Table 3 – Highest range factors [RFs] among starting major league second basemen, 1980 - 2000

Rank	Player	Team	Year	RF
1	Hubbard	Atl	1985	6.27
2	Trillo	Phi	1980	5.90
3	Wills	Tex	1981	5.88
4	Molitor	Mil	1980	5.87
5	Bernazard	CWS	1982	5.81
6	Grich	Ana	1983	5.80
7	J. Cruz	Sea	1981	5.79
7	Grich	Ana	1981	5.79
9	Hubbard	Atl	1980	5.75
10	F. White	KC	1983	5.74
11	D. Garcia	Tor	1980	5.70
11	Sandberg	Cubs	1983	5.70
13	Hubbard	Atl	1982	5.67
13	Herr	StL	1981	5.67
15	Trillo	Phi	1981	5.65
15	Wills	Tex	1980	5.65
15	R. Thompson	SF	1992	5.65
18	Gantner	Mil	1981	5.64
19	F. White	KC	1984	5.61
19	Morrison	Wsox	1980	5.61
19	Gantner	Mil	1983	5.61
19	Velarde	Oak	2000	5.61
23	Garner	Pit	1980	5.58
23	F. White	KC	1985	5.58
25	Oester	Cin	1986	5.57
26	Grich	Ana	1983	5.55
26	Grich	Ana	1984	5.55
28	Sandberg	Cubs	1985	5.54
29	Sandberg	Cubs	1984	5.53
29	E. Young	Cubs	1996	5.53
29	Barrett	Bos	1987	5.53
32	Oester	Cin	1985	5.52
32	Gant	Atl	1988	5.52
32	Randolf	NYY	1980	5.52
32	Vina	Mil	1998	5.52
36	Ray	Pit	1982	5.51
36	F. White	KC	1980	5.51
36	Lind	Pit	1992	5.51
36	Jo. Reed	Mil	1994	5.51
36	White	KC	1981	5.51
36	Garcia	Pit	1994	5.51
42	Reynolds	Sea	1986	5.50
43	Hubbard	Atl	1987	5.48
43	Hubbard	Atl	1984	5.48
45	Baerga	Cle	1992	5.47
46	Sax	LA	1984	5.45
47	R. Alomar	SD	1988	5.44

(iii). *Larger trends*. Because the game has changed, Alomar's RF must be measured against different standards. Clearly, baseball has changed from the early part of the twentieth century, when most of the defensive records for second basemen were established. But how much and how fast? Alomar's RFs are modest even by the standards of the 1980s. Since 1980, Alomar's career-best seasonal RF (5.44 in 1988) ranks 47th, trailing the highest RF of the 1980s -- Glen Hubbard's 1985 RF of 6.27 -- by 134 chances per 162 game schedule (Table 3).

Discussion

As a whole, Alomar's defensive record is excellent: no other AL second baseman has played as well as Alomar over the decade of the 1990s. A few players have exhibited better RFs for a few years -- notably Jody Reed and Carlos Baerga in the early 1990s, and Randy Velarde, Miguel Cairo and Damion Easley in the late 1990s. None of these players, however, have sustained excellence throughout the period. In addition, Alomar's unadjusted RF probably unestimates his range afield. Adjustments such as turf and BIPs generally enhance Alomar's RF.

But that is not the standard against which Alomar has been compared. Alomar's range is supposed to be much better than average. Alomar's RFs during his three years in San Diego were consistently among the best in major league baseball. Over the past 10 years, however, Alomar's RF has been little better than average.

Literature Cited

- Hanke, B., and R. Wood. 1989. The 1989 Baseball Abstract. Mad Aztec Press, Kirkwood, MO.
- James, B. 1982. The Bill James Baseball Abstract 1982. Ballantine Books, New York, NY.
- Neyer, R. 2001. Some arguments from the defense. ESPN.com Baseball, Wednesday, February 14. www.espn.go.com/mlb/columns/neyer_rob/42798.html.
- STATS, Inc. 2000. legends.foxsports.com/demo/ballparks.asp.

Duke Rankin, 136 Indigo Lane, Calera, AL, 35040-4646, rankind@montevallo.edu ♦

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 two 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.

Runs Scored and RBI – Multiple Regression Analysis

Willie Runquist

It is acknowledged that the statistic Runs Produced (defined as half the total of runs plus RBI) is not necessarily an accurate measure of player offensive skill. However, it may still be of interest to find out how RP relates to the components of a batting line. Here, the author uses multiple regression analysis to cast light on this question.

Prolegomena

Run Production (RP) has been defined in *Total Baseball* as a measure of individual run productivity. Usually it is computed as R + RBI - HR. It has, however, been treated with some disdain by baseball analysts. Total Baseball states that RBI (and presumably R as well) is "extremely situation dependent, denying equal access to equal opportunity...". Accordingly, it has been labelled "tainted", "totally flawed", or "discredited". However, such pejorative labels are only valid within the narrow research context of attempts to evaluate a player's merit or ability.

In a broader context, run production is simply a description of one aspect of a player¹s performance, and it is equally valid to determine the relationship of that performance to identifiable variables, whatever they may be.

If we assume that for every run scored by a team, the player who scores the run accounts for one half of that run and the player who drives it in accounts for the other half of the run, then subtracting homeruns makes little sense. Since a player both bats in and scores a run with a homerun, that player should be credited with both halves. Therefore, RP is defined here as (R + RBI) / 2. The purpose of this research is to determine how differences among players in the traditional categories in the batting line: singles, doubles, triples, homeruns, bases on balls, and stolen bases are related to differences in runs and runs batted in.

It is not the purpose of this research to promote RP as a measure of individual excellence.

This study provides a multiple regression analysis of the RP for the years 1901 to 1970. The time span covers various spontaneous changes in the game from the so-called dead ball era, through the roaring twenties and depression, World War II, and early expansion (but not the designated hitter).

The basic procedure was to predict individual runs and runs batted in from the linear combination of six offensive variables. As an empirical procedure, multiple regression analysis provides a direct estimate of the maximum predictability of a linear combination of variables by producing a set of coefficients for each variable if applied to the variables in the analysis will minimize the difference between predicted and obtained values. R2, the squared multiple correlation coefficient, indicates the proportion of variance in the data attributable to this best linear combination. The analysis also provides a measure of the accuracy of the individual predicted values as the standard deviation of the differences between the predicted and obtained values, the root mean square residual (RMS). It indicates the range of approximately two-thirds of the discrepancies.

When evaluating a particular player's performance it is common practice to divide the frequencies by plate appearances, but plate appearances do not produce runs in the same sense that batting variables do. They merely provide the opportunity for those events to occur. For example, if it takes ten home runs to drive in 15 runs, for predictive purposes it does not matter whether those ten homers were hit in 100 or 50 plate appearances the player¹s production will be the same. In evaluating a particular player's performance relative to other players, it may be more equitable to take his plate appearances (or opportunities) into account, but that evaluation is not our concern here.

The number of outs was also not included in the analysis. Outs is a complex variable. Some outs are productive with respect to both runs scored and runs batted in, but conceptually they are neutral in that they usually indicate a non-productive plate appearance. Outs may have a negative impact on team scoring and may be properly considered in a measure of a player's individual performance, but, like plate appearances, ten homeruns should account for the same number of runs whether a player makes 50 outs or 100 outs along the way.

For each of the eight teams in each league the sample of players for these analyses consisted of the nine players for each team that had the most plate appearances in a given season, producing samples of 72 players in each team season. Change of venue for various teams was not considered. Following expansion, the size of the sample would increase thus affecting both R2 and RMS. Therefore, only the ³original² eight teams supplied players for the analysis between 1961 and 1970 keeping sample size at 72. The multiple correlation between the six variables was computed separately for R and RBI for each season as well as for RP and for the American and National leagues. Results for the American and National Leagues, however, did not differ to a great extent. Therefore, the results for the two leagues were pooled. The seasons were then organized into 14 sets of five consecutive seasons. The multiple correlation for each set was computed by pooling the sums of squares for regression and residual for each set in each league.[6]

The values for R2 and RMS are

	01-05	06-10	11-15	16-20	21-25	26-30	31-35
R2	.922	.899	.923	.921	.931	.949	.946
RMS	5.86	6.19	6.58	6.09	6.99	6.70	6.86

	36-40	41-45	46-50	51-55	56-60	61-65	66-70
R2	.939	.934	.937	.931	.932	.939	.951
RMS	7.46	5.96	6.65	6.13	6.06	5.65	5.21

The six variables account generally account for well over 90% of the variance in RP. The values of R2 for individual seasons ranged from 963 to .862 while the range of RMS was from 4.25 to 8.54. Note that a value for R2 of .862 represents a multiple correlation coefficient of .930, while a value of .963 represents a multiple correlation of .981. The individual years may also be divided approximately into six "eras," the dead ball era (1901-1919), the roaring 20s (1920-29), the depression (1930-1940), World War II (1940-1945), post-war pre-expansion (1946-1960) and expansion (1960-70). There was a general increase in R2 through the six eras except for a significant dip during WWII. Changes in RMS were less orderly, but by and large they reflect the correlations. In only one group of seasons did the RMS exceed 7.

Runs Scored and Runs Batted In

The next tables present R2 and RMS values for runs scored (RS) and Runs Batted In (RBI) regressed on the same six variables.

Runs Scored

	01-05	06-10	11-15	16-20	21-25	26-30	31-35
R2	.885	.849	.879	.851	.868	.876	.861
RMS	8.84	9.21	9.13	9.28	10.31	11.03	10.84

	36-40	41-45	46-50	51-55	56-60	61-65	66-70
R2	.870	.854	.876	.881	.877	.885	.882
RMS	10.39	9.85	9.63	9.37	8.90	8.77	8.78

Runs Batted In

	01-05	06-10	11-15	16-20	21-25	26-30	31-35
R2	.759	.724	.766	.756	.778	.858	.830
RMS	10.80	10.69	11.40	10.95	13.29	12.04	12.35

	36-40	41-45	46-50	51-55	56-60	61-65	66-70
R2	.827	.811	.849	.855	.865	.870	.883
RMS	12.39	11.00	11.11	10.65	9.55	10.09	9.22

Both the values of R2 and the RMS of the residuals indicate a much greater accuracy in the prediction of RP than in either of the two components. This is a consequence of RP being the sum of two criterion variables which are poorly correlated and in which the relative weights of the variables differs for the two quantities. The deficiencies of prediction in one criterion for a player will counteract the

[6] Pooling results from different seasons introduces a statistical problem in that some players will contribute more than one set of numbers to the pool. A few comparisons were made between "pure" samples and the total pool. None of those comparisons produced any detectable systematically biases, so the analyses proceeded with the total pool. This procedure also kept the sample size the same from season to season thus assuring some stability in the standard error of the coefficients.

deficiencies in the prediction of the other. The situation is similar to that where team runs are predicted from OPS, where where neither OBA nor SA is highly correlated with the criterion, but the combination is quite successful.

The most obvious feature of these results is that the six variables do a poorer job of predicting RBI than Runs particularly during the "dead ball era." While it may be the case that RBI are in fact less dependent than Runs on the six variables, there may be other reasons for the relatively poor showing. The proportion of unpredictable variance as well as RMS are both susceptible to random variance, i.e. variance in the RBI[1]s not systematically related to the variables in question (as well as random variance in the variables). The percentage of runs that were batted in before 1920 was highly variable from season to season and averaged just over 80%. Modern values consistently exceed 90%. While some of this difference may be accounted for by the generally greater number of unearned runs during the early period, it is also possible that scorers were not diligent in crediting batters with RBI since it was not an official statistic.

Coefficients

If the variability among players differs widely from one variable to another, the magnitude oif a coefficient for one variable cannot be directly compared with another. This was true for these variables. The standard error of the raw coefficients varied from an average of .06 for singles to .63 for homeruns. However, by carrying out the regression analysis on standard scores, the problem is alleviated since the standard deviation is then 1.00 for every variable. The transformation has no effect on the value of R2, and the predicted values are also the same when transformed back to their original units.

Unlike the commonly used "linear weights," however, even the standardized coefficients have no particular theoretical meaning beyond their ability to predict the criterion variables since they reflect not only the unique variance contributed by the particular variable, but also the variance that the variable shares with the other variables in the analysis. Therefore, the actual values are dependent on the variables included in the analysis, particularly if those variables are intercorrelated. The coefficients do reveal changes in the predictive power of the variables.

This analysis was carried out on pooled data from the five-year periods with the combined leagues. Each of the coefficients in the table is therefore based on an analysis of the performance of 720 players.

Runs Scored

	01-05	06-10	11-15	16-20	21-25	26-30	31-35
1B	.466	.401	.381	.340	.364	.126	.382
2B	.102	.135	.140	.286	.185	.282	.171
3B	.155	.122	.090	.112	.062	.149	.137
HR	.124	.044	.106	.145	.183	.265	.252
BB	.237	.329	.306	.309	.177	.153	.253
SB	.146	.166	.191	.094	.144	.164	.119

	36-40	41-45	46-50	51-55	56-60	61-65	66-70
1B	.172	.266	.340	.393	.432	.413	.290
2B	.203	.128	.252	.167	.121	.175	.137
3B	.134	.165	.103	.106	.136	.111	.112
HR	.226	.197	.343	.323	.363	.369	.319
BB	.171	.114	.215	.276	.251	.207	.145
SB	.123	.087	.126	.073	-.056	.195	.129

Runs Batted In

	01-05	06-10	11-15	16-20	21-25	26-30	31-35
1B	.227	.281	.308	.290	.292	.067	.196
2B	.389	.421	.305	.374	.167	.376	.221
3B	.199	.145	.186	.179	.095	.122	.151
HR	.250	.222	.225	.321	.328	.588	.521
BB	-.124	.080	-.093	-.092	-.004	-.013	-.109
SB	.093	.024	-.068	-.007	-.055	.007	-.037

	36-40	41-45	46-50	51-55	56-60	61-65	66-70
1B	.187	.158	.240	.195	.207	.285	.225
2B	.251	.173	.206	.165	.172	.128	.180
3B	.073	.112	.093	.056	.010	.073	.032
HR	.605	.447	.581	.678	.771	.742	.735
BB	.094	-.002	.044	-.007	-.025	.038	.046
SB	-.073	-.074	-.064	-.011	-.010	-.074	-.105

The pattern of coefficients differs considerably for R and RBI. For R singles produced the most consistent high values followed by bases on balls. Generally, then, players not surprisingly score more runs the more often they reach base. While hits appear to be more important than walks, this may because players who walk a lot because they are "pitched around" do not score as often. This is reflected in the fact that coefficients for walks decreases in the later years where homeruns are more important. Homeruns showed the biggest change through the years, with the coefficients beginning to increase in the early 1920's, dipping during World War II then showing a large increase in the post war seasons. Except for occasional blips, stolen bases and triples do not contribute much to a player¹s run scoring. Triples are probably not frequent enough to have much effect, particularly in the modern era, while steals have become a specialty in which only a few players indulge.

Runs Batted In show some rather consistent changes through the years. While singles remain at a fairly consistent level, it is not the primary predictor of driving in runs. At first it was doubles that had the largest coefficients, but the value of doubles in predicting RBI decreased as homeruns became more salient. By the late 1920s, homeruns was a consistent and overwhelmingly better predictor of RBI than any other variable. Bases on balls ranked last from the start and over half of the values were negative. From a causal point of view this rather makes sense since a walk in a potential RBI situation is generally a non-productive outcome, and players who walk a lot may be "top of the order" types who get fewer RBI opportunities or potential run producers who are frequently "passed". Triples contributed somewhat to RBIs in the first two decades, but declined to almost nil as the value of homeruns increased . The coefficients for stolen bases were almost universally negative. Speedy players are seldom put in a position to drive in runs.

Runs Produced is the average of R and RBI. RP can therefore be predicted by averaging the predicted values for R and RBI. It may also subjected to a separate multiple regression analysis, but the raw coefficients in that analysis will be the average of the raw coefficients of the two components.

Coefficients and the Nature of the Residual

The residual variance (1- R2) consists of variation produced by variables not included in the analysis. Some of these variables have a random effect. In any given sample of players or games, the random variable will produce differences among players, but in an infinite number of players playing an infinite number of games, their net effect will be zero. Other variables not included are those which are either correlated with the variables included in the analysis or with the criterion. If they are only correlated with the criterion, they have no effect on the analysis other than to increase R2. Often, however, simply accurately predicting R2 is of minor interest. The most powerful aspect of regression analysis is its ability to isolate the amount of variance attributable to a given variable. It is somewhat paradoxical that this goal is thwarted by the presence of variables not included in the analysis, which can bias an obtained coefficient to the extent that it is correlated with that variable. In the present case, such variables as hit by pitcher, caught stealing and other base running gaffes, safe on fielder's choice, and safe on errors would very likely be systematically related to a player¹s result, and their inclusion would aid prediction to the extent that they are not correlated with variables already included. The most critical missing variables are those which generally would fall under the rubric of "contextual" variables. Since both scoring runs and batting them in requires cooperation from one¹s teammates, (except for home runs), and since position in the batting order affects how much of one¹s teammmates talents may be utilized, these variables have been suggested as important sources of variance in a player¹s totals and hence have been deemed sufficient cause to "discredit" R, RBI or their combination as a measure of a player¹s real contribution.

While we have been able to account for over 90% of the variance in RP without recourse to the context variables, it is likely that some of the residual variance is attributable to them. Earlier, we commented on the fact that it is the actions of the player, not his opportunities that produce runs. Plate appearances do not effect this relationship because they are causally independent of these events, but the same cannot be said of the position in which a batter bats or the effect of his teammmate¹s actions. Unfortunately, batting order position is completely confounded with what the player does. His position is largely determined by his past performance which is again correlated with his present performance. Players who have either a perceived or historical ability to get on base despite the absence of power hitting are often placed at the head of the order, while power hitters are placed behind these players in the hope that their power will be more productive. Attempts to correct a players numbers for batting order position are themselves badly biased since they tend to rob the player of the very numbers that put him there in the first place

These corrections are logically akin to a statistical correction procedure called analysis of covariance. It is widely understood that one cannot correct averages for a covariate that is correlated with the independent variable. It is in fact, as some have labelled it, creating a set of data that cannot exist. This not only applies to batting order, but park effects when a team selects players for their perceived ability to hit or pitch in their home park. Correcting for a covariate is only appropriate when the players are "assigned" to batting order position, or to parks randomly, or to other conditions that are deemed correctable randomly. In baseball, such practices never occur.

The result of this confounding is that the coefficient for a particular variable includes not only its unique variance and the variance it shares with other variables in the analysis, but also variance it shares with variables not in the analysis. Therefore, we do not know, for example, whether a large coefficient for doubles indicates that they are causally important in driving in runs, or players who hit doubles were put in situations where their talent to do same would be put to an advantage.

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

Neal Traven Address Change

Neal Traven, co-chair of the Statistical Analysis Committee, has moved as of March 8, 2001. Here is his new address:

Neal Traven
4317 Dayton Ave. N, #201
Seattle, WA
USA
98103-7154

Home phone 206-632-0093
Work phone 206-364-9700 ext. 2021

beisbol@alumni.pitt.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@home.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.

Does Good Hitting Beat Good Pitching?

Tom Hanrahan

Does good hitting beat good pitching? The log5 method, developed by Bill James in the early 80s, was designed to demonstrate how offensive outcomes are affected by the quality of the pitcher and hitter. However, the question lingers, and, so, here, the author looks at empirical data to find out if there is a "good hitting beats good pitching" effect beyond what James predicted in 1983.

A Bit of Background

"Good pitching will beat good hitting:" this adage is much older than I, and so it is with much humility and respect that I attempt to debunk it, while yet affirming it to be true in some senses.

First, when we ask this question, we need to define what we mean. Does the phrase "good pitching will beat good hitting" mean

- when two good teams collide, the team with the better pitching will more often win?
- if Pedro is on the mound and he has his good stuff, it doesn't really matter who is batting?
- the run-generating abilities of Vlad Guerrero or Todd Helton impact the outcome of an at-bat less than the run-preventing abilities of an ace pitcher?
- all of the above?
- something else?

The question could be answered differently, depending on what is meant.

Also, when looking for evidence to answer this question, we need to take care not to confuse causal factors with coincidental ones. For example:

Fact: There are fewer runs scored in your typical post-season game than in the regular season. Is this evidence for the old adage? Or is it proof that fewer runs are scored when you don't need a 5th SP and when it's colder outside?

Fact: Teams which allow 150 fewer runs than the league average will win more games and pennants than teams which score 150 more runs than the league average. Does this prove pitching's superiority? Or does it imply that it's easier to build a winning team in a pitcher's park, or show that by allowing fewer runs your ratio of runs scored to runs allowed is better than by scoring extra runs?

The Question Defined

Here is how I will NOT define the question: *Is a run saved more valuable than a run gained?*

The answer to *this* question I believe is "yes," but it isn't relevant to our topic. Many studies have shown that team's ratio of run scored to runs allowed correlates very well with a team's wins. It is very easy to see that if the league average for a team is 800 runs per year, that lowering your runs allowed by 100 (down to 700) makes your ratio 1.143. In order to achieve a team ratio of 1.143 by scoring additional runs, the team runs scored would have to be 800 * 1.143 = 914. This means 100 runs saved has the same value in this example as 114 runs scored, so these runs saved are 14% more important. This has relevant applications to simulation games, MVP awards, and debates of Roger Clemens' value versus Barry Bonds', but it has no real bearing on what happens when Clemens faces Bonds.

Here is how I *will* define the question: *When superior batters face superior pitchers, are the results different than expectations predicted by using typical mathematical models of their abilities?*

In other words, Mark McGwire hits lots of home runs (when he's healthy). Greg Maddux does not allow many home runs. If McGwire hits 3 times as many dingers as the typical batter, but Maddux only allows one-third as a typical pitcher, what happens when they face each other? The expectation would be that given these inputs, McGwire would hit an average amount of home runs in his at bats against Maddux[1].

[1] Actually, the math is more complicated than this, but using proportions this way is a pretty good estimate. The next section explains the actual calculations involved.

If in reality McGwire hits FEWER home runs in his appearances against Maddux, then we can say that in this instance, good pitching has indeed stopped good hitting.

Specific Methods

I approached the question 2 ways. The first was a study based solely on batting average. I chose this as my first metric because the data were available ("available" meaning Phil Birnbaum obtained a large set of data for me from Retrosheet; thanks, Phil!), because of its common use, and because Dan Levitt had used this method in the Feb 99 BTN ("The Batter/Pitcher Matchup", pgs 18-20), so I would have a point of comparison. The second study was an attempt to look at the whole picture of run scoring and prevention, since it is possible that there could be effects other than batting average. For this second study I found records of individual matchups of batter versus pitcher, and was able to compare expected to actual on base and slugging averages.

In Dan Levitt's paper, he quoted a formula used by Bill James in his 1983 Baseball Abstract which yields the expected batting average, given inputs of the hitter, pitcher, and league averages. Once again this is reproduced here:

Expected batting average =

$$\frac{(\text{BAT AVG})*(\text{PIT AVG})/(\text{LG AVG})}{(\text{BAT AVG})*(\text{PIT AVG})/(\text{LG AVG}) + (1 - \text{BAT AVG})*(1 - \text{PIT AVG})/(1 - \text{LG AVG})}$$

He used a sample of all hitters with at least 446 PAs and all pitchers with at least 100 batters faced in the 1995 season, and found the formula accurately predicted reality.

The database I used here is from the years 1984-1996. There is a total of almost 1.4 million at bats involved when both leagues are combined. The database recorded all batters who hit for a certain average during a specific year, and how they fared when batting against pitchers who allowed a certain batting average in that year. The averages are divided into "bins" every .005 points.

Copying Dan's approach, I divided all hitters and pitchers into 3 groups (good, average, and poor) according to batting average by combining the "bins". This yielded a block of 9 comparisons for each league (good vs. good, good vs. average, etc.). For each block, I computed the batting average expected by the formula, to that actually achieved. The results? The formula predicted the actual totals very, very accurately.

Table 1 – American League

Player Category	Poor Batters (avg < .253)	Average Batters .252<avg<.283	Good Batters (avg>.282)	All Batters
Good Pitchers (opp avg < .253)				
At Bats	93137	106936	81393	281466
Hits	19215	25320	22272	66807
Batting Avg	.2063	.2368	.2736	.2374
Average Pitchers (.252 < avg < .283)				
At Bats	93321	111247	86180	290748
Hits	21995	29988	26020	78003
Batting Avg	.2357	.2696	.3019	.2683
Poor Pitchers (opp avg > .282)				
At Bats	63793	77928	61820	203541
Hits	16822	23437	20664	60923
Batting Avg	.2637	.3080	.3343	.2993
All Pitchers				
At Bats	250251	296111	229393	775755
Hits	58032	78745	68956	205733
Batting Avg	.2319	.2659	.3006	.2652

Table 1 gives the data for the American League, Table 2 the National. Here is how to calculate the formula-predicted batting average: find the opponents' average for good pitchers in the rightmost column of all batters in Table 1 (.2374). Then find the average for good hitters in the bottom row of all pitchers (.3006). An extra decimal place is used to reduce rounding errors. The overall league average (bottom row, right) is .2652. Using the formula, the predicted batting average resulting from these circumstances is (.2374*.3006/.2652) / [(.2374*.3006/.2652) + (.7626*.6994/.7348)] = .2704. The actual average achieved was .2736. The good batters actually hit 3 points better than expected against the good pitchers, over the course of over 81000 at-bats.

Table 3 gives a summary of how each batter versus pitcher combination fared as compared to the formula prediction, when the data from both leagues is combined. Every category was within 5 points of batting average of the predictions. For the sample sizes used, up to 3 or 4 points difference could be explained by random chance. The only difference that is statistically significant is for the case of good batters facing poor pitchers. The standard deviation of the batting average, as calculated using the binomial method, for both leagues combined would be .0013 ([.3294 * (1-.3294) / 129223] ^ 0.5). The hitters batted 4 points lower, which is 3 standard deviations lower. However, even though this is statistically significant, it is not practically significant -- a MLB manager may have a difficult time finding a way to take advantage of the knowledge that good hitters pound lousy pitchers by a whole 4 points less than expected.

Table 2 – National League

Player Category	Poor Batters (avg < .253)	Average Batters .252<avg<.283	Good Batters (avg>.282)	All Batters
Good Pitchers (opp avg < .253)				
At Bats	63914	80218	55085	199217
Hits	13011	19084	14810	46905
Batting Avg	.2036	.2379	.2689	.2354
Average Pitchers (.252 < avg < .283)				
At Bats	57020	74962	55227	187209
Hits	13121	19744	16503	49368
Batting Avg	.2301	.2634	.2988	.2637
Poor Pitchers (opp avg > .282)				
At Bats	64197	89963	67403	221563
Hits	16878	26386	21905	65169
Batting Avg	.2629	.2933	.3250	.2941
All Pitchers				
At Bats	185131	245143	177715	607989
Hits	43010	62514	53218	161442
Batting Avg	.2323	.2660	.2995	.2655

Table 3 – Actual Results Compared to Expected Batting Average Differences

Player Category	Lg	Poor Batters	Average Batters	Good Batters
Good Pitchers	AL	0	-1	3
Average Pitchers	AL	1	1	-2
Poor Pitchers	AL	0	1	-3
Good Pitchers	NL	-1	2	2
Average Pitchers	NL	-1	-1	1
Poor Pitchers	NL	4	-1	-5
Good Pitchers	MLB	-1	0	3
Average Pitchers	MLB	0	0	0
Poor Pitchers	MLB	2	0	-4

So what does this mean? I conclude that in terms of batting average, good pitching does *not* beat good hitting. Batters hit pretty much what was expected, given the quality of pitching they faced. If anything, there is a hint of evidence that the good hitters may fare a bit better than expected against good pitchers.

A Different Approach

I mentioned previously a second method of study, based on on-base and slugging averages. This was undertaken since it is certainly possible that the ability to produce or prevent runs would show up in the batter/pitcher matchup in ways that are not represented by batting average alone. I actually started along this track first, before I had access to the Retrosheet data used above. Stats, Inc. produces an annual book entitled *Match-Ups!* The 1997 version gives individual batter vs. pitcher data for all players active at the end of the 1996 season.

Following the same approach as in the other study, I divided both hitters and pitchers into groups labeled good, poor, and average. I found all hitters active as of September 1996 who had more than 3000 career at-bats, and pitchers with greater than 800 innings pitched. The hitters were grouped by Runs Created per Game (the formula used in the Bill James Historical Abstract, first edition), and the pitchers by ERA. There were 110 hitters and 85 pitchers in the study, of which 36 hitters and 27 pitchers were each classified as "good[2]". Their totals, both composite and for pitchers divided by group, are given in Table 4 (the other pitchers' totals are only shown for comparison). They are not used in the study, as I only was interested in the matchup of good hitters against good pitchers; besides, it would have been 9 times as much work to record and calculate all of the other combinations. (The main focus only requires we know the stats of the good pitchers and hitters and the league averages.) It should be noted that the ERA of the pitchers in the study is better than the overall league average. Obviously, pitchers and hitters who had a large amount of playing time are typically better than "average", so the labels given to the groups might be misnomers in a sense. This study may actually be using some of the top quarter or fifth of all pitchers and hitters in this period, but that does not affect the methods used or the conclusions.

Table 4 – Batter vs. Pitcher Matchup Data Career Records, 1984-1996

Player Category	Innings Pitched	Weighted ERA
Good Pitchers (ERA < 3.53)	43116	3.28
Average Pitchers (3.53 < ERA <3.90)	41105	3.74
Poor Pitchers (ERA > 3.90)	42105	4.16
All Pitchers With > 800 IP	126326	3.72

Player Category	AB	R/G	OPS
Good Hitters (R/G > 5.9)	178654	6.28	.860

The main difference between this method and the first is that the groupings are based on career statistics, not individual seasons. Reasonable arguments can be made about the advantages of either method.

Data for pitchers' SLG allowed was not available. However, this can be derived from the opponents' OBA and ERA. The league averages for the period 1984-1996, weighted by player-years active in study, were OBA = .3274, SLG = .3988, and ERA = 4.074.[3] From this we can relate ERA to the offensive components by the formula ERA = OBA * SLG * 31.20 for this period, using the accepted structure for Runs Created. Since this formula works for leagues (the RMS error in computing ERA for individual seasons using the formula was 0.08, or 2% of the ERA), I have assumed it works for individual pitchers.

For the good pitchers, their OBA allowed = .3009. SLG is defined as = ERA / OBA / 31.2 = 3.283 / .3009 / 31.2 = .3497. We infer this as the slugging percentage allowed by the good pitchers.

```
Good pitchers   OBA = .3009    SLG = .3497
Good hitters    OBA = .3790    SLG = .4807
League avg      OBA = .3274    SLG = .3988
```

Formula predicted results of good pitchers vs. good hitters:

```
OBA  = (.3009*.3790/.3274) / (.3009*.3788/.3274 + .6991*.6210/.6726) = .3505

SLG  = (.3497*.4807/.3988) / (.3497*.4807/.3988 + .6503*.5193/.6012) = .4287
```

[2] Hitters were R. Alomar, Biggio, Bagwell, Baines, Belle, Barry Bonds, Bonilla, Boggs, Buhner, Canseco, W. Clark, E. Davis, Dykstra, Grace, Greenwell, Griffey, Gywnn, R. Henderson, Knoblauch, Larkin, E. Martinez, McGwire, McGriff, K. Mitchell, E. Murray, Molitor, Olerud, O'Neill, Palmeiro, Raines, Sandberg, Sheffield, Strawberry, Tartabull, Tettleton, Thomas; Pitchers were Aguilera, Appier, K. Brown, Clemens, Cone, Drabek, Eckersley, Eichhorn, Fassero, Glavine, Gooden, Hershiser, Honeycutt, M. Jackson, Key, D. Martinez, G. Maddux, McDowell, Orosco, Pena, Schilling, L. Smith, Smoltz, Viola.

[3] "League averages weighted by player-years" means this: Of all of the good hitters and pitchers in the study, 27 of them were active in 1984, 33 in 1985, and all 63 of them by 1991. So, the league averages for the time span in the study are estimated by [27 * (1984 avg) + ….+ 63 * (1996 avg)] / (27 +……+ 63). This won't give as exact an answer as using actual innings pitched, but it should be very close. As an example, a simple average of ERA for the 13 years is 4.05; this compares closely to the weighted average of 4.07, so I strongly suspect further refinements would make very little difference.

What were the actual results? Drumroll, please.

```
In 16564 at bats of good hitters versus good pitchers: OBA = .3501,  SLG = .4545
```

Table 5 shows that good hitters had an OBA when facing good pitchers almost exactly as expected by the formula prediction. The SLG, however, was 26 points higher! Is this a significant difference statistically? Calculating a standard deviation of SLG is not as simple as batting average, since there are multiple event possibilities, but by Monte Carlo simulation I came up with a sigma(SLG) for 16564 at bats = .0058. This means that a difference of 26 points is definitely statistically significant (over 4 standard deviations!). The larger chance for error here is not mathematical, but rather is in my assumption of how to infer slugging average allowed for pitchers. However, it seems unlikely to me that the formula would be off by as much as the 6% difference between expected SLG and actual SLG, since the error for predicting league ERAs for each year was only 2%.

Table 5 – Good hitters records against good pitchers defined by career data, 1984-1996

	Actual	Predicted	Difference
On Base Average	.3501	.3505	-.0004
Slugging Average	.4545	.4287	+.0258

Conclusions

Does good pitching stop good hitting? No more so than anyone should expect. A pitcher who allows few hits and walks will surrender proportionately fewer of these regardless of the class of hitter at the plate. Manny Ramirez is likely to garner more hits against poor pitchers, but he will still hit his "share" against good ones.

There is some evidence to suggest that it is possible, although I'm not sure I should say "likely", that in terms of power, good hitting might actually beat good pitching. I recommend a study that focuses on home runs, using a larger sample than my second study here, to see if this is true.

Tom Hanrahan, 21700 Galatea St., Lexington Park, MD, 20653, HanrahanTJ@navair.navy.mil. ♦

A Promising New Run Estimator – Base Runs

Brandon Heipp

Runs Created and Linear Weights are two examples of run estimators, statistics that attempt to derive runs scored from the components of a batting line. Here, the author summarizes a new statistic developed by David Smyth, which is potentially more accurate than others.

Run estimators are a topic that is extremely common in sabermetrics, much to the annoyance of some who would like to get past the issue and move on to something else. Personally, I am fascinated by them. Anyway, most new run estimators proposed are simply linear weights formulas with slightly different coefficients than previous method. However, some recent work by David Smyth appears to have promise and be an alternative to preexisting measures.

Smyth wrote an article on his Base Runs method which is available at *http://www.baseballstuff.com/fraser/articles/basenew.html* about a year ago. He has recently updated his work on the Strategy and Sabermetrics forum at FanHome.com *(www.fanhome.com)*.

Base Runs is a team context formula, like Runs Created. It tries to measure the interactivity between offensive events, not just the simple linear values. The basis for BsR is that runs scored equals home runs plus the number of baserunners times the percentage of baserunners who score. As an equation, this is:

$$\frac{A \times B}{B + C} + D$$

A represents baserunners, B is advancement, C is outs, and D is Home Runs.

B/(B+C) is the estimated percentage of baserunners who score, an estimator that Smyth found gives good empirical results.

Smyth gives these equations for the factors:

```
A =  H + W + HB – CS -HR
B = (2.5*TB – H – 5*HR + 2*SB + .05*(W+HB))*X
C =  AB – H + SF
```

Where X is set so that Base Runs equals Runs for whatever unit is being tested(team, league, etc.), and is historically around .535. A version that I have rigged up is useful in that it looks at just the basic stats(AB, H, TB, W):

```
A = H+W-HR
B = (2*TB – H – 4.5*HR + .05*W)*X
C = AB-H

X is historically around .81.
```

The most intriguing part of Base Runs is that Smyth claims that it is more accurate at the extremes than other run creation methods. Here are estimates for a few extreme teams using the basic versions of Base Runs, Extrapolated Runs, and Runs Created. Team A makes 500 outs, no baserunners. Team B hits 500 HR, no outs. Team C draws 500 walks, no outs:

Team	BsR	XR	RC
A	0	-48	0
B	500	720	2000
C	500	170	0

In each case, BsR gives a more reasonable estimate than either XR or RC. But to be a useful formula, it must predict runs with similar accuracy in normal conditions. Here are the Root Mean Square Errors for each of the formula, predicting team runs in the period 1980-2000:

Method	RMSE
XR	23.7
BsR	23.9
RC	25.8

The accuracy of Base Runs is comparable to that of Extrapolated Runs, and superior to that of Runs Created, at least in this sample.

One final note about Base Runs is that it is a team method, like Runs Created, and can cause distortion when applied to players. To properly evaluate players by Base Runs, a theoretical team mechanism like the one Bill James uses in his New RC is necessary.

If you are interested, please check out Smyth's work at the aforementioned sites. Thanks to him for his research and his willingness to share it with all.

Brandon Heipp, 17381 Creighton Drive, Chagrin Falls, OH, 44023, patriot@csuvikings.com ♦

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