By the Numbers

Volume 12, Number 1 — The Newsletter of the SABR Statistical Analysis Committee — February, 2002

Review

"Curve Ball:" How to Think Like a Statistician

Charlie Pavitt

While "Curve Ball" is more statistical than sabermetrical, it is nonetheless sabermetrics' "best book-length effort to date," the author writes in this review.

For more than fifteen years now, if anyone were to ask me where to go for a good introduction to or review of sabermetric research, I had only one answer: Thorn and Palmer's *The Hidden Game of Baseball*. There were no alternatives. That is, until now. Albert and Bennett's *Curve Ball* will stand aside *The Hidden Game* as a must-read for anyone interested in statistical baseball research.

Let me begin by discussing *Curve Ball*'s intent. As the authors admit on page 343, it "is not...a complete guide to sabermetrics." This is certainly the case; in fact, the authors have only a few things to say about pitching and nothing at all about fielding. Yet, when then they claim it to be merely "a loosely connected collection of quantitative essays on baseball statistics," Albert and Bennett have shortchanged their effort. In my eyes, what Albert and Bennett have attempted and, with only a couple of exceptions, succeeded in doing is showing the reader how to think like a statistician when analyzing batting data. The reader learning about methods for evaluating offensive performance simultaneously learns about means and standard deviations, probability distributions and confidence intervals, regression equations and standard errors, and the logic of statistical inference and hypothesis testing. Another of my favorite sabermetric books, Willie Runquist's *Baseball By The Numbers* (not to be confused with this newsletter), has a similar intent but does not perform the task quite as systematically. In fact, one aspect of *Curve Ball* that I prefer to *The Hidden Game* is that the former is a general argument for a particular way of thinking about baseball data whereas the latter is a specific argument for a particular method (Linear Weights). From a pedagogical standpoint, I do believe the authors erred in not introducing the correlation coefficient when describing the relationship among indices, although they use it on page 108. Otherwise, I could literally see their book being used as a statistics textbook for physical education majors and other students who find the baseball content interesting and understandable. One further service this book provides is lucid descriptions of research previously published in statistics journals in forms incomprehensible even to a reader with standard social science statistics training (if I am indeed representative of that group).

In this issue

The book begins with an analysis of the statistical presumptions underlying well-known table-top baseball games which will fascinate anyone familiar with those games (the two with which I grew up, All-Star Baseball and APBA, are well represented). In Chapters 2 and 3, Albert and Bennett use simple offensive (BA, SA, OBA) and pitching (K, BB) data to introduce basic statistical concepts. Chapter 4 includes a nice discussion of Albert's previously published work on situational effects while teaching the reader about the logic of model testing. Chapters 6 and 7 read at first as mere reworkings of the Thorn and Palmer comparisons among indices for measuring offense (Runs Created, Total Average, Linear Weights, etc.), but function as lead-ins to some very thoughtful comparisons among these indices, in Chapter 8 going well past *The Hidden Game*'s presentation. Chapter 9 does a nice job of considering clutch play in the context of more general ideas about measuring performance. Chapter 10 considers methods for predicting the odds of attaining batting milestones, both single year and career; in the latter case, it would have been nice to see them evaluate the James/STATS methods discussed in each year's *Major League Handbook*. Chapter 11 looks at the odds of winning the World Series for teams of differing skill levels. In an afterword, the authors emphasize the role of chance in determining player and team performance outcomes.

There is, unfortunately, one big clunker here; the authors' treatment of batting streaks in Chapter 5. Rather than presenting a general analysis of batting data, they examine the pre-All Star-break 1999 batting record of Todd Zeile, which was unusually inconsistent over time during those months. This can lead the reader to conclude that there is good evidence for streakiness as a general characteristic, when in fact existing research suggests that only select players, if any, exhibit non-random streakiness. The problem, which has cropped up in earlier work (their response to Albright's work in Volume 88 of the Journal of the American Statistical Association), is that Albert and Bennett are biased in favor of the existence of streaks and slumps beyond random variation for a non-statistical and, I must add, very bad reason. As basketball players in their youth, they perceived themselves as "having the hot hand" on days that they made their shots and "out of rhythm" on days they missed, and claim that as evidence that "a hot hand refers to a feeling -- it's an intrinsic characteristic of our shooting ability" (page 144). Social psychological theories such as cognitive dissonance and self-perception explain quite well our tendency to label ourselves consistently with our perceptions of our behaviors ("I hit a few shots early, so I must be hot today"), and how such perceptions may persist even when subsequent behavior should disconfirm them. In other words, Albert and Bennett's "feelings" are easily-explained psychological biases which likely have little basis in reality.

I can only hope that readers of this book do not trust the authors' comments in this regard. I also hope that every reader of this newsletter has the opportunity to read this book. It is probably our little discipline's best book-length effort to date.

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.

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.

Ichiro and the MVP

Duke Rankin

Ichiro Suzuki's selection as American League MVP was one of the most contoversial in recent memory – in on-base plus slugging, one of the premier measures of offensive performance, he ranked a large 299 points behind league leader Jason Giambi. In this article, the author puts the 299-point discrepancy in historical context, and then analyzes whether other aspects of Ichiro's performance help his case.

I don't know what happened to your blood pressure when Ichiro Suzuki won the 2001 AL MVP, but I, for one, was so tired of the Itchy for Superman mantra that I had to be restrained. Sure, Suzuki led the league in BA and steals, and in addition played gold-glove defense. But he also hit for little power, seldom walked, and played a corner position. Suzuki's OPS total was 299 points behind Giambi, the league leader. Given a choice between Giambi and Suzuki, who deserved the award?

To structure this analysis, I asked the following questions: First, how often do players 299 OPS points behind the league leader win the MVP? And second, is it possible that other aspects of Ichiro's game could counterbalance this disparity?

The historical context

For each major league season beginning in 1931, I compared the MVP to the league leader in OPS. Although MVP votes predate the study, the 1931 season was both the first year of the present BBWAA voting system, and a reasonable approximation of the lively ball era. When a pitcher won the MVP vote, I substituted the highest vote getter among position players.

Only 12 BBWAA votes (n = 142; 71 in each league) exhibit OPS differentials of 200 points or greater (Table 1[1]). The 299 point differential between Suzuki and Giambi is the fourth highest in the history of the BBWAA vote.

Table 1: Largest OPS disparities between the the MVP and the league leader

Year	MVP	pos	OPS	Leader	pos	OPS	difference
1934	Cochrane	c	.840	Gehrig	1b	1.172	332
1962	Wills	ss	.720	F. Robinson	rf	1.045	325
1944	Marion	ss	.686	Musial	rf	.990	304
2001	Suzuki	rf	.838	Giambi	1b	1.137	299
1954	Berra	c	.855	T. Williams	lf	1.113	258
1942	Gordon	2b	.900	T. Williams	lf	1.147	247
1960	Groat	ss	.765	F. Robinson	1b	1.002	237
1931	Frisch	2b	.764	Hornsby	2b	.996	232
1955	Berra	c	.819	Mantle	cf	1.042	223
1947	DiMaggio	cf	.913	T. Williams	lf	1.133	220
1941	DiMaggio	cf	1.083	T. Williams	lf	1.286	203
1973	Rose	lf	.838	Stargell	lf	1.038	200

In general, catchers and middle infielders were the beneficiaries of the voting. Only two corner outfielders received the MVP when trailing the league leader by 200 or more OPS points: Pete Rose in 1973 and Ichiro Suzuki in 2001.

The 1934 vote sets the pattern: a light-hitting (but high BA) middle defender on a championship team winning the award over a heavy-hitting corner defender on a non-champion. Led by Greenberg, Gehringer, Bridges and Rowe, the Tigers won the pennant by 7 games. The pitchers both won over 20, but Lefty Gomez of the Yankees was better, leading the league in wins, winning percentage, innings and ERA. Hank Greenberg was a great player but Gehrig posted better numbers, winning the triple crown. Charlie Gehringer was brilliant – second in the league in BA, fifth in RBI, excellent in the field – but lacked the gaudy power numbers. That apparently left Cochrane, an established star who hit .320, as the MVP.

[1] Data in this and other tables are from www.baseball-reference.com unless otherwise noted.

The 1944 MVP illustrates a second recurrent theme in the data: the absence of a clear-cut MVP candidate. The Cardinals won three pennants between 1942 and 1944. Their two best players – Mort Cooper and Stan Musial – had already been honored, winning MVPs in 1942 and 1943. In 1944, Musial's power numbers were down – 12 dingers and 94 RBI. Cooper pitched very well – 22 wins and the third best ERA in the league – but had posted better numbers in the past. Marion did not distinguish himself statistically -- he hit for a reasonable average but no power, rarely walked and stole very few bases, fielded adequately but not as well as the best shortstops in the league. The wartime National League of 1944, however, contained very few MVP candidates. Dixie Walker led the NL in hitting, but he played for the seventh-place Dodgers and posted numbers almost identical to Musial. Bill Nicholson led the NL in dingers, runs and RBI for the fourth-place Cubs, but he failed to hit outside the war years. Bucky Walters pitched well for the third-place Reds, but posted numbers very similar to Mort Cooper. A gritty shortstop on the championship team was probably a good choice.

Yogi Berra received two MVPs despite differentials of over 200 points, but Berra's value should not be underestimated. During the seven years between 1949 and 1955, Berra was arguably the best defensive catcher in the AL, leading the league in assists three times and double plays five times. In addition, the Yankees led the league in ERA three times. Offensively, no one was close. In 1955, Berra hit .272 with 27 dingers and 108 RBI. The next best catcher, Sherm Lollar, hit .261 with 16 dingers and 61 RBI. In 1954, it was Berra with a .307/22/125 performance, Sammy White only .282/14/75. In 1951, Berra was .294/27/88, and Jim Hegan .238/6/43. Berra simply dominated AL catchers.

Ted Williams appears on the 'screwed' portion of Table 1 four times, and it's difficult to escape the conclusion that this was, to some degree, a reflection of the personal animosity between Williams and the BBWAA. Williams led the AL in OPS ten times but won only two MVP awards; Williams had seven seasons where he led the MVP winner by over 100 OPS points. The 1947 vote was the famous one in which Williams won the triple crown, but lost the MVP by one point because two writers from the Boston area left Williams completely off their ballots.

The most suspect MVP vote, however, occurred in 1962, when Maury Wills mesmerized baseball by stealing 108 bases, breaking Ty Cobb's 1915 record of 96. In reality, Wills wasn't very good. Although he hit .299, Wills had only 29 extra base hits and 51 walks in 695 AB. Defensively, he was terrible: below average in range factor, next to last in errors, and tied for last in fielding percentage (though he inexplicably received the Gold Glove). Unlike the Marion MVP in 1944, the NL had several excellent candidates: Willie Mays played on a champion, led the league in dingers, placed second in RBI, and was probably the best defensive outfielder in the league. Tommy Davis led the league in hitting and RBI; Frank Robinson led in OPS; and Don Drysdale won 25 and led the league in innings and strikeouts. Any of those guys would have done just fine.

Since 1931, eleven right fielders have won the MVP without leading their league in OPS (Table 2). Three other right fielders were the top vote getters among position players. Six MVP right fielders finished at least 100 OPS points behind the league leader:

Table 2: OPS for MVP right fielders vs. OPS for league leader

Year	MPV	OPS Leader	Difference
2001	Suzuki	Giambi	299
1998	Sosa	McGwire	198
1996	J. Gonzalez	McGwire	187
1961	Maris	Cash	156
1966	Clemente	Allen	131
1958	Jensen	T. Williams	111
1968	Rose*	McCovey	62
1998	J. Gonzalez	Belle	59
1957	Aaron	Musial	56
1934	P. Waner*	Collins	40
1974	Burroughs	Allen	37
1942	Slaughter*	Ott	6
1988	Canseco	Boggs	5
1960	Maris	Mantle	5

Asterisk indicates position player with highest MVP vote (if pitcher was MVP). Years listed are those for which the OPS leader and MVP are different.

1. In 1958, Jackie Jensen finished 111 points behind OPS leader Ted Williams. Jensen led the AL in RBI, but the MVP should have gone to Mickey Mantle, who played on a champion and led the league in dingers, bases, walks and runs.

2. In 1966, Clemente finished 131 points behind OPS leader Dick Allen. Allen led the league in home runs and RBI, but played only 137 games in the field. Clemente put up pretty good numbers himself – 29 dingers and 119 RBI. In addition, Clemente was as famous for his defensive skills as Allen was for his emotional baggage.

3. In 1961, Roger Maris finished 156 points behind OPS leader Norm Cash. Maris hit for the asterisk as the Yankees crushed the American League.

4. In 1996, Juan Gonzalez finished 187 points behind OPS leader Mark McGwire. Gonzo put up pretty good numbers for the division champion Rangers – second in slugging, second in RBI, fifth in dingers – while McGwire finished a mysterious seventh in the voting.

5. In 1998, Sammy Sosa finished 198 points behind Mark McGwire. McGwire hit 70 dingers, but Sosa hit 66 dingers of his own and led the league with 158 RBI and 134 runs.

6. Finally, in 2001, Suzuki finished 299 points behind Giambi.

Giambi and Suzuki: a direct comparison

The historical context suggests the 299 OPS disparity between Suzuki and Giambi is unusual but not necessarily unprecedented. Is it possible that other aspects of Suzuki's game can compensate for the OPS disparity? First, however, we should place the 299 point disparity into a context. A simple comparison:

	OBP	SLG	OPS
Giambi	.477	.660	1.137
Suzuki	.381	.457	0.838
Somebody else	.245	.262	0.507

Some of you may recognize the "somebody else" line as the career totals for Mario Mendoza, the legendary shortstop who created the Mendoza line, the point at which a player hits so poorly that no level of secondary statistics or locker room presence can justify a starting job. Mendoza finished his career 331 OPS points behind Suzuki's 2001 season. Suzuki hit almost as poorly compared to Giambi as he hit successfully compared to Mario Mendoza.

Bill James (1984 *Abstract*, pp. 189-191) has written a cautionary tale about creating a single statistic to evaluate ballplayers; it should be required reading for fledgling sabermetricians. But what if, for the sake of argument, we adjust OPS by adding estimates of performance from other areas in which Suzuki is clearly superior to Giambi? Here are the adjustments:

Defense

Suzuki's defensive statistics were arguably the best among AL right fielders (Table 3). Chris Richard actually led the league in range factor and fielding percentage, but played only 69 games in right field. Among right fielders with at least 90 games, Suzuki led the league in range factor and fielding percentage.

Suzuki's assists, double plays and zone rank were close to league average.

The mean range factor for right fielders was 2.09 chances per game. Suzuki's range factor was 0.26 chances per game higher than the mean, suggesting that, over the course of the season, Suzuki caught 40 balls more the average right fielder. Suzuki's 0.26 range differential is also quite similar to the career marks for right fielders such as Roberto Clemente (0.23) and Hank Aaron (0.20). Of course, Aaron's career mark includes 105 outfield games at age 39 -- at Suzuki's age of 27, Aaron was 0.67 chances

Table 3: Defensive statistics for 2001 AL right fielders

Player	Team	G	PO	A	E	DP	FP	RF	ZR
Salmon	ANA	125	253	13	3	5	.989	2.20	.872
Richard	BAL	69	155	5	0	1	1.000	2.68	.824
Nixon	BOS	83	141	3	5	3	.966	1.97	.919
Ordonez	CHW	155	285	11	5	0	.983	2.01	.863
Gonzalez	CLE	119	214	10	3	3	.983	2.06	.906
Encarnacion	DET	63	121	3	2	0	.984	2.11	.929
Dye	KAN	92	175	6	3	0	.984	2.05	
Lawton	MIN	67	129	4	2	1	.985	1.99	.887
O'Neill	NYY	130	210	1	4	0	.981	1.74	.856
Dye	OAK	61	96	7	3	1	.972	1.72	.911
Suzuki	SEA	152	335	8	1	2	.997	2.35	.877
Grieve	TAM	64	117	0	3	0	.975	1.95	.885
Ledee	TEX	60	110	1	3	0	.974	2.23	.906
Mondesi	TOR	149	263	19	8	2	.972	1.92	.828

ZR scores from http://sports.espn.go.com per game above the league average.

Suzuki also throws well. To estimate the number of outs created with his arm, I added his assists to his double plays and divided the total by games played. Suzuki's 'arm index' was 0.0658 plays/games, good for 9th in the AL (Tim Salmon led the league with 0.144 plays/games), and slightly below the AL mean of 0.080 plays/game. Over the course of the season, Suzuki threw out three fewer base runners than the average right fielder.

Finally, Suzuki committed only one error. Over the course of the season, Suzuki caught 5 balls that other outfielders would have misplayed into an error. Combining range, arm and fielding percentage, Suzuki's defense contributed approximately 42 more outs than the average right fielder.

Giambi's defensive statistics are not as good as Suzuki's (Table 4). His range factor compares favorably with the AL mean of 9.39 chances/game. Most plays at first, of course, involve catching throws from infielders. Bill James (1983 *Abstract*) uses assists per game as a measure of range at first base. Giambi's assist/game ratio of 0.55 is 0.09 assists below the league mean of 0.64. Over the course of the season, Giambi made 14 fewer assists than the average first baseman. Giambi's fielding percentage of .992 is almost exactly the league mean of .993 -- a difference of one error over the course of the season.

By implication, Giambi is also unable to field balls that an average first baseman would turn into unassisted putouts. For the sake of argument, I'm guessing the magnitude is about the same as the assist total: 14 balls per season, or a total of 28 plays per season that the average AL first baseman would have made. In all, the defensive difference between Suzuki and Giambi is approximately 70 catches per season compared to average players at the same positions.

How to add this to OPS? I will attempt to convert Suzuki's defensive contributions into offensive contributions. First, I will assume preventing a single is the functional equivalent of hitting a single: for each single Suzuki saves defensively, I will award him an extra single offensively. Some of the balls batted past first basemen and right fielders, however, would go for extra bases. Without any data on the subject, I'm guessing the number of singles would equal the number of doubles: for 70 defensive plays, 35 singles and 35 doubles.

Table 4: Defensive statistics for 2001 AL first basemen

Player	Team	G	PO	A	E	DP	FP	RF	ZR
Spiezio	ANA	105	819	74	1	64	.999	10.15	.895
Conine	BAL	80	646	45	4	61	.994	9.27	.831
Daubach	BOS	106	839	75	11	71	.988	9.09	.846
Konerko	CHW	144	1277	90	8	120	.994	9.76	.853
Thome	CLE	148	1176	78	10	105	.992	9.36	.805
Clark	DET	78	647	48	3	69	.996	10.23	.850
Sweeney	KAN	108	946	88	12	124	.989	9.99	.842
Mientkiewicz	MIN	148	1263	69	4	95	.997	9.44	.868
Martinez	NYY	149	1143	99	5	105	.996	8.64	.881
Giambi	OAK	136	1224	75	11	107	.992	9.94	.875
Olerud	SEA	158	1210	121	9	116	.993	8.89	.848
Cox	TAM	78	569	48	1	64	.998	8.66	.839
Palmeiro	TEX	113	905	83	8	112	.992	8.93	.858
Delgado	TOR	161	1520	103	9	167	.994	10.15	.844

ZR scores from http://sports.espn.go.com.

To adjust Suzuki's OPS, I will add 70 base hits and 105 total bases to Suzuki's batting line:

	AB	H	TB	BB	BA	OPB	SLG	OPS
Before	692	242	316	30	.350	.381	.457	0.838
Adjusted	692	312	421	30	.450	.477	.608	1.085

Stolen bases

Next, I will add Suzuki's stolen bases. Subtracting Suzuki's 14 CS from his 56 SB yields a net total of 42 bases, or 40 more bases than Giambi (2 SB, 0 CS). The value of a CS, in terms of lost productivity, however, is not equivalent to a SB. The conversion factor is CS = SB*1.75, for a net increase of 31 bases, or 29 more than Giambi. I will simply add these bases to Suzuki's total.

	AB	H	TB	BB	BA	OBP	SLG	OPS
Before	692	312	421	30	.450	.477	.608	1.085
Adjusted	692	312	450	30	.450	.477	.650	1.127

Clutch hitting

Finally, I will adjust for hitting with runners in scoring position. Suzuki led the league in this category, hitting .449. However, Giambi hit well with runners in scoring position, too:

	AB	H	TB	BB	BA	OBP	SLG	RBI	OPS
Suzuki	136	61	74	17	.449	.509	.544	55	1.053
Giambi	113	40	73	51	.354	.531	.646	72	1.177

With runners in scoring position, Giambi had 19 extra base hits (12/ 0/7) compared to seven for Suzuki (6/2/1). Giambi drove in 0.43 runs per plate appearance with runners in scoring position, Suzuki 0.36 runs. Pitchers were three times more likely to take the stick out of Giambi's hands compared to Suzuki.

On the one hand, both Giambi and Suzuki have already had these totals incorporated into their OPS statistics. On the other hand, there is a quantitative difference between the two players: to reach Giambi's mark of 1.177 OPS, Suzuki needed 12 additional singles in his 136 AB (BA = .536, SLG = .632, OPS = 1.168). I will penalize Suzuki the 12 singles from his batting record. I have no data to support this adjustment; hopefully, it reflects the bonus value of Giambi's performance in the context of the game.

	AB	H	TB	BB	BA	OBP	SLG	OPS
Before	692	312	450	30	.450	.477	.650	1.127
Adjusted	692	300	438	30	.433	.460	.633	1.093

The summary so far: adjusting for defense, base stealing and clutch hitting gives Suzuki an adjusted OPS of 1.093, compared to Giambi's OPS of 1.137. Suzuki had a much better season than I thought, and made roughly similar contributions to winning as Giambi. But the difference still represents over 20 points in batting average -- the difference between a .350 hitter and a .370 hitter.

Of course, the award is for Most Valuable Player, not for Best Offensive Performance. Value is a nebulous term that can manifest itself in many ways. For the purposes of this discussion, I will ask two questions: which player had the greater influence on the pennant race? And second, which of the two ballplayers would be easier to replace?

Influence on the pennant race

In 2001, Seattle won its division by 14 games over Oakland, while Oakland won the wild card by 17 games over Minnesota:

	W	L	PCT	GB
Seattle	116	46	.716	--
Oakland	102	60	.630	14
Anaheim	75	87	.463	41
Minnesota	85	77	.525	(17)

Suzuki's 128 RC is the equivalent of approximately 13 offensive wins; if Seattle had played without a right fielder, it still would have won the AL West. Giambi's 173 RC is the equivalent of 17 offensive wins; if Oakland had played without a first baseman, Oakland would have tied Minnesota for the wild card. Removing Suzuki and Giambi from the pennant race does not unambiguously change the pennant race or the teams in the playoffs.

Finding a replacement

Teams do not play without a player at each position. Between Giambi and Suzuki, Suzuki may have been easier to replace. At least one other AL right fielder hit as well as Suzuki: Magglio Ordonez created 122 runs, only one half of an offensive win behind Suzuki's 128 RC (Table 5). Trot Nixon, Juan Gonzalez and Jermaine Dye were all within two offensive wins.

Suzuki also consumed an enormous number of outs: 467 outs in 738 PA, 8th highest in the AL. Although Suzuki's 128 RC ranked 9th in the AL, the 467 outs produce an RC/out ratio of 0.274, third-best among AL right fielders. On a ratio basis, both Ordonez and Gonzalez were more productive than Suzuki.

Ordonez, Gonzalez and Dye all had average defensive numbers. Suzuki was easily the best defensive right fielder in the league, and the difference was substantial: Suzuki's defensive won/loss percentage of approximately .670 converts to one or two defensive wins above average. Offensively, Suzuki could be replaced. Defensively, Suzuki was the best in the league. Largely because of his glove, Suzuki was probably 2 wins better than the best available replacement right fielder.

Giambi clearly created more offensive wins than any other AL first baseman (Table 6). The next best first baseman, Jim Thome, was still three offensive wins below Giambi's total. Rafael Palmeiro and Carlos Delgado were each four offensive wins below Giambi's total, and John Olerud 6.5 wins. Giambi was also very efficient -- his RC/out ratio of 0.482 was easily the best ratio among AL first basemen. Only one player in the league hit as well as Giambi, and he was a shortstop.

Defensively, Giambi was average at best, but none of the outstanding hitters at first base were much better; it appears unlikely that any of Thome, Palmeiro or Delgado would have substantially improved on Giambi's defense. Giambi was at least three wins better than the best available replacement first baseman – very close to the two wins estimated for Suzuki.

Table 5: Hitting statistics for AL right fielders

		OPS	RC	RC/out
Salmon	ANA	.748	73	.192
Richard	BAL	.770	67	.177
Nixon	BOS	.881	107	.270
Ordonez	CHW	.914	122	.282
Gonzalez	CLE	.960	114	.302
Encarnacion	DET	.700	49	.148
Lawton	MIN	.835	62	.219
O'Neill	NYY	.789	75	.189
Dye	OAK	.813	102	.232
Suzuki	SEA	.838	128	.274
Grieve	TAM	.760	84	.203
Ledee	TEX	.654	27	.141
Mondesi	TOR	.794	90	.199

Table 6: Hitting statistics for AL first basemen

		OPS	RC	RC/out
Spiezio	ANA	.764	67	.196
Conine	BAL	.829	92	.241
Daubach	BOS	.859	74	.239
Konerko	CHW	.856	101	.232
Thome	CLE	1.040	143	.373
Clark	DET	.856	77	.241
Sweeney	KC	.916	113	.279
Mientkiewicz	MIN	.851	100	.254
Martinez	NYY	.830	95	.217
Giambi	OAK	1.137	173	.482
Olerud	SEA	.873	109	.259
Cox	TAM	.750	46	.172
Palmeiro	TEX	.944	136	.306
Delgado	TOR	.948	135	.319

[data from sports.espn.go.com].

Conclusion

Did Giambi really deserve the MVP? Giambi may have played better than Suzuki, but the difference was slight at best, and the questions of replacement and influence seem ambiguous. Suzuki, on the other hand, was clearly a leader on the best team in the league, and he played on a division winner. Both Giambi and Suzuki were deserving of the MVP award, but to the victors go the spoils: Suzuki was probably a better choice than Giambi. Excuse me while I go check my blood pressure.

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

Strength of Opposition a Starting Pitcher Faces: Part I – Introduction of Method

Rob Wood

What is the best way to calculate the average strength of a group of starting pitchers pitching one game each? Averaging their winning percentages fails to take into account that some winning percentages are based on small sample sizes, and therefore unreliable. Weighting the average by number of starts is also misleading because each pitcher should have equal weight regardless of how often he pitched in games outside the sample. Here, the author provides an alternative statistic, based on a bayesian approach, that estimates a pitcher's intrinsic winning percentage even in the face of a very small number of starts.

Sabermetricians have used a pitcher's ERA as the best measure of how well he pitched in any season. Of course, ERAs need to be considered in their context. Sabermetrics provides methods to take into account the era and the ballparks in which the pitcher's ERA was achieved. After all, a 2.90 ERA means different things in 1930 vs. 1968, or in Dodger Stadium vs. Coors Field.

In Dick Thompson's fabulous review of Wes Ferrell's career, he uncovered evidence that Ferrell's ERA might have been high for contextual reasons other than era and parks. Specifically, in reviewing Ferrell's career, it seemed that Ferrell was often pitted against the toughest opponents (both teams and opposition starting pitchers), whereas it seemed that Lefty Grove was often pitted against easier opponents. Dick raises an interesting point. A pitcher's ERA depends upon who he faced over the course of the season, and if there are systematic factors causing one pitcher to face tougher opposition than another, these factors should be taken into account when comparing their ERAs.

By using game-by-game Retrosheet data, the esteemed Tom Ruane looked into the specific case of Ferrell vs. Grove. It turns out that there is little evidence that, *over the course of their entire careers*, Ferrell systematically faced tougher opposition than did Grove. While with Cleveland, Ferrell did indeed face tough opposition; however, Ferrell's opposition with the Red Sox was easier than Grove's.

Leaving aside the specific Ferrell-Grove comparison, Dick raises a challenging methodological issue that needs to be addressed. How do we measure the "strength" of the opposition that a pitcher faces in a season? After all, we would like to know if he faced more difficult or easier opposition than, say, his pitching teammates did. Remember that the fact that he does not face his own team (e.g., Herb Pennock did not have to face Babe & Lou and the vaunted 1927 Yankee offense) is already captured in the "park factor".[1]

Tom Ruane looked at several statistics -- first, the average win percentage of the teams that the pitcher faced in the season (OPct); and second, the average win percentage of the opposition starting pitchers that the pitcher faced in the season (PPct)[2]. Then, these two measures then compared to the average win percentage of all the teams that the pitcher's club faced in the season (TPct). A pitcher can be said to have faced tougher opposition than expected if OPct or PPct are significantly greater than TPct.

While this seems to be a reasonable approach, problems arise when a pitcher faces another pitcher who only starts a few games in the season. Let's see why. Suppose an opposition starter goes 0-1 for the season. Maybe he was an emergency starter, maybe he got injured after that game, or maybe he was a September call-up. This throws a wrinkle into the calculation of PPct above, since his win percentage of .000 goes into the average along with all the other pitchers' win percentages.

For simplicity, suppose a pitcher faced the following four pitchers: 20-10, 20-10, 20-10, and 0-1. The average win percentage of these pitchers is .500 (the average of .667, .667, .667, and .000). Everyone would agree that in this case .500 is not a good reflection of the strength of the collection of pitchers our hero faced. The average method ignores how "reliable" the win percentages are likely to be (sample size issues).

Several people recommended that instead of the average of the win percentages, it would be better to take the win percentage of the combined records of the pitchers faced. In the example above, the opposition starters combined for a 60-31 record, which is a .659 win percentage. This is a much better reflection of the strength of the opposition, though it may appear to be too high.

[1] This is why park factors are calculated for a team's pitchers and hitters separately.
[2] In the database that Tom used, each game's losing pitcher was not available, so Tom used the win pct of the team in games that the opposition starter started. This does not affect what follows.

Unfortunately, examples can be readily found for which the combined win percentage method gives an answer we are not happy with. Suppose the pitcher faced the following pitchers: 30-0, 0-10, 0-10, and 0-10. The combined record method gives .500 (the win percentage of 30-30). This is surely too high, since our hero actually faced one unbeatable foe and three bums. The combined method ignores the fact that each game has a separate winner and loser.

In this article, I will introduce a method that is the best of both worlds. The method will automatically take into account sample size issues as well as the fact that each game has a winner and loser. In addition, the method will combine the information contained in OPct (the strength of the opposition teams) and PPct (the strength of the opposition starters) in one single measure.

The method will be based upon statistical inference methods sometimes called "Bayesian updating" methods. Bayesian inference starts with a prior belief, obtains new sample observations, and then constructs a posterior distribution that combines (i.e., updates) the prior beliefs with the new information contained in the sample. One important aspect of this updating process is a decision on how much weight we should put on our prior beliefs and how much weight we should put on the new observations. I will return to this below.

As a silly example of Bayesian updating, suppose that you believe that 90% of all lawyers carry briefcases. On the way to your office, you see 10 lawyers and only 5 of them have briefcases. What do you believe now? Bayesian inference provides the means by which you should properly update your prior beliefs (90%) with the new information (50%) to give you your new beliefs (??%). Clearly, the stronger your prior beliefs, for example the more evidence it is based upon, the less you change your beliefs. Suppose you have been counting lawyers and briefcases everyday for 10 years. Then your prior is very solid and a new sample of 5 of 10 briefcases is unlikely to change your belief much. On the other hand, suppose the 90% belief was only based upon yesterday's count of 9 of 10. Then you would be likely to update your beliefs so that you now believe that 70% of lawyers carry briefcases (14 of 20).

To use this methodology in our case, let's call the opposition starting pitcher's W-L record the sample (the new data) and his team's W-L record as the prior[3]. The posterior then is a reflection of how strong we think the opposition team really is when the starter in question starts against us. For example, suppose you are about to face the 1927 Yankees and they are starting a pitcher who is 0-2 for the season. What is a good measure of the strength of your opposition on that day? Surely we want to take into account the fact that the 1927 Yankees are a great team (as reflected in their great win percentage) as well as the fact that the bum who is pitching has not demonstrated an ability to win in the big leagues (though the evidence is only two starts). Bayesian inference provides the formulas to come up with the best measure.

Remember we need a way to reflect our prior beliefs. This needs to be a distribution that has a mean reflecting our "current belief" and a variance reflecting how strongly we believe our current belief. Variance reflects how much uncertainty there is around the mean. The smaller is the variance, the less uncertainty, so the greater is our confidence in the prior, and the less we update it based upon new data.

The standard prior distribution used in Bayesian inference of a binomial probability[4] is a beta distribution. A beta distribution takes values between 0 and 1. With a beta prior, the posterior distribution of a binomial probability also is a beta distribution. We will also make use of the additional fact that under very general conditions, the *mean* of the posterior distribution minimizes Bayes risk when the loss function is quadratic. The mean of the posterior distribution is

```
Best Pred = E[p|x,n] = (x+a)/(a+b+n)
```

where the new sample data is x "successes" out of n trials, and a and b are the parameters of the prior beta distribution. In our case, x is W and n is W+L, meaning that the new information (the opposition starter's record) is W wins and L losses.

We can solve for a and b using the formulas for the mean and variance of a beta distribution.

```
E[y] = a/(a+b);  Var[y] = (a*b)/[((a+b)^2)*(a+b+1)]
```

A reasonable assumption is that the mean of the prior distribution should be set equal to the observed win percentage of the opposition team. Call this Q. For simplicity, let's use the variance of the proportion of successes of a binomial process for the variance of the prior distribution.

Then E[y] = Q and Var[y] = (Q*(1-Q))/T where T is the number of "prior observations". That is, if T is very large, our uncertainty around our prior belief on Q is very small, and the less we would change our estimate with new data. On the other hand, if T is very small, our uncertainty around our prior belief on Q is large, and the more we would change our estimate with new data.

---

3 For this discussion I don't think it matters if we remove the pitcher's own decisions from the team record.

4 A team's wins in a season is often treated as a binomial process with the probability of winning each game being the binomial probability p.

Equating the means and variances and solving for a and b in terms of Q and T yields the following:

```
a = Q*(T-1)   and b = (T-1)*(1-Q).
```

Plugging these values into our formula of the mean of the posterior distribution yields:

```
Best Pred = [W+(Q*(T-1))]/(W+L+T-1)
```

The formula above is the best updated prediction of a team's winning percentage when a specific starter starts. We need the following information. Q is the win percentage of the opposition team in all of its games. W is the specific starter's wins in the season, and L is the specific starter's losses in the season. The only remaining variable we need is T.

Let's see how the formula behaves with T. Suppose T is one. Then you can see that our best prediction is the pitcher's own win percentage (W/(W+L)). As T gets larger, the prediction moves from the pitcher's own win percentage to the team's win percentage. This makes intuitive sense since T was a reflection of the confidence we had in the prior.

I have played around with the formula at some length. For seasonal win percentages, I have found that a reasonable value for T is around 15. Roughly speaking, this implies when a pitcher reaches 15 decisions, our best estimate of his "true" win percentage (i.e., the "true" win percentage of his team when he starts) is the average of his own win percentage and the win percentage of his team in all of its games. As the number of his decisions increases, the more faith we put in his own win percentage; and below 15 the more faith we put in the team win percentage.

I have coded up an Excel spreadsheet that has the relevant formulas in it that allows me to play around with various cases and see the result using different values for T. I am comfortable that a value of 15 is sensible. I would be happy to send the spreadsheet to anyone interested.

Where does all this leave us? We now have a formula that estimates the "strength" of each opposition starter that a pitcher faces over the course of a season. The formula automatically takes into account the number of games associated with the opposition pitcher's win percentage. The next step is to take the average of these strength estimates over all the pitchers our hero faced in the season.[5]

Let's look at how the Bayesian method measures the opposition's strength in the two examples listed above. If our hero faced the following pitchers 20-10, 20-10, 20-10, and 0-1, the Bayesian method says that the strengths of these four pitchers are .614, .614, .614, and .467, respectively. The average opposition is therefore .577. Recall that we thought that the average .500 was too low but the combined method's score of .659 was too high.

The other example I mentioned was that our hero faced 30-0, 0-10, 0-10, 0-10. The average of .250 was too low, while the combined method's score of .500 was too high. The Bayes method gives four estimates of .841, .292, .292, .292, for an average of .429. In both examples, I consider the Bayes estimate to be very reasonable. The Bayes examples above used T of 15 and a Q of .500. In reality, I recommend using the opposition teams' actual win percentages for Q.

To summarize, I am proposing a method of estimating the strength of any pitcher's opposition during a season. The method automatically takes into account the strength of the opposition teams as well as the strength of the specific opposition starting pitchers. The method was specifically developed to properly address small sample issues that can bedevil other "ad hoc" formulas since many pitchers have gone 0-1, 0-2, etc. The method is based upon Bayesian statistical inference and yields a formula that is very simple to program and to understand.[6]

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

[5] If our hero faced the same opposition pitcher more than once, the best prediction of the opposing starter's win pct should enter the average more than once.
[6] I would be happy to work with anyone who has access to historical game-by-game records. We could look at the strength of the opposition faced by various pitchers such as Christy Mathewson, Pete Alexander, Walter Johnson, Lefty Grove, Bob Feller, Warren Spahn, Sandy Koufax, Bob Gibson, Tom Seaver, Roger Clemens, Greg Maddux, et al.

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.

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