By the Numbers

Volume 16, Number 1 The Newsletter of the SABR Statistical Analysis Committee February, 2006

Review

Two New Detailed Sabermetrics Books

Phil Birnbaum

A review of two new sabermetrics books – "Baseball Between the Numbers," and "The Book."

There haven't been a lot of detailed books on sabermetrics in the last decade or so. But this year, there are at least two – "Baseball Between the Numbers," and "The Book".

"Baseball Between the Numbers" (BBTN) consists of about 30 unrelated chapters, each by a member of the "Baseball Prospectus" gang. The book is subtitled "Why Everything You Know About the Game Is Wrong," but, actually, the book mostly confirms sabermetric wisdom. Chapter 1.1 (the chapters are numbered after innings and outs), titled "What's the Matter With RBI," pretty much runs as you would expect.

"The Book" is a self-published, in-depth study of baseball strategy by Tom Tango ("Tangotiger"), Mitchel Lichtman, and Andrew Dolphin (I'll call them TMA), all of whom have written extensively on sabermetrics, mostly online.

In this issue

If there's a theme running through both books, it's that they both investigate mostly questions for which we already roughly knew the answer – but go into much more detail to quantify the results.

Take the use of relievers, for instance. It's pretty much accepted, among most sabermetricians, that inserting your closer to start the 9th with a three-run lead is counterproductive – better to use him in the 8th with the score tied, where the game is really on the line. But both books take this much further.

In BBTN, Keith Woolner starts by bringing in the concept of "leverage." In the 8th inning of a tie game, a home run changes the chances of winning about two-and-a-half times as much as usual, and so that situation has a leverage of 2.54. Woolner gives us a three-page chart, showing leverage for every base-out situation of every inning, and every score differential.

Up by three runs to lead off the ninth, the leverage is only 0.41 – this situation is less than half as important as average. So, obviously, it's kind of silly to use your closer in this situation. It's much better to save him for that 8th inning situation, with the 2.54 leverage.

This brings up another issue of strategy. Suppose that up by one run in the 7th, your starter loads the bases with nobody out. Is this a good situation for your stopper? The leverage is 1.45, which is reasonably important. But if you bring in your stopper now, if you need him later in the 9th, in an even more important situation, he won't be available. And, also, he may not be available tomorrow if you put him in today. So what should you do?

Analyzing play-by-play data, Woolner found the average maximum leverage per game was 1.66. So unless the leverage exceeds 1.66, you should save your stopper for what will likely be a more important situation tomorrow.

But even if tomorrow is an off-day, you want to wait, since the leverage might increase even in this game. Woolner calculates that, once leverage has exceeded 1.66 before the 9th, there's a 50/50 chance that it will go on to exceed 2.32. So Woolner's recommendation under these simplified assumptions is:

- before the 9th, bring in your closer if the situation exceeds 2.32; or
- in the 9th, bring in your closer if the situation exceeds 1.66.

Woolner goes on to calculate that by using closers in this fashion, teams can improve, on average, by 1.6 wins.

The reliever usage question is also addressed in "The Book." In Chapter 8, TMA calculate that from 1999-2002, the best nine stoppers in the majors were used 19% of the time in situations when the leverage was only 0.7. Since 19% is much more than the number of tune-up appearances for these elite relievers, they must have been used in "real" situations that happened to be low-leverage – three runs up in the 9th, for instance.

TMA figure these stoppers are being used, on average, in situations averaging 1.9. They argue that it would be easy to adjust their appearances to bring their average up to 2.3. This is a 20% gain, so if Troy Percival was worth four extra wins to his team, he'd now be worth five. That's a one win gain.

If these reliever analyses make the books seem similar, well, maybe they are a bit, but not all that much. BBTN covers a broader range of topics, including the effects of steroids and the economics of new stadiums. "The Book", on the other hand, sticks to detailed analyses of in-game strategies (as evidenced by its subtitle, "Playing the Percentages in Baseball").

Baseball Between the Numbers

By The Baseball Prospectus Team of Experts
Edited by Jonah Keri

Basic Books, 454 pages, $24.95 (US)
ISBN 0465005969

The Book – Playing the Percentages in Baseball

By Tom M. Tango, Mitchel G. Lichtman, Andrew E. Dolphin

TMA Press, 382 pages, $18.95 (US)
www.insidethebook.com

Indeed, "The Book" is dense, numeric, and detailed. There are probably close to 100 separate studies in here, each taking only a couple of pages, each exceptionally well-explained, and each worthy of BTN. Take, for instance, the chapter on sacrificing – when is it a good idea and when isn't it? In only 50 pages, TMA analyze the data for runners on first and no outs, and find (among other things) that:

- The sacrifice *attempt* is much more valuable than the successful sacrifice, because of the possibility of the batter reaching base.
- When good hitters bunt early in the game, their results are much better than when lesser hitters bunt. This suggests that the difference is the element of surprise, and therefore the readiness of the defense is a crucial factor in deciding whether to sacrifice.
- Also, sacrifices late in the game are successful at a significantly lower rate than sacrifices early in the game, again suggesting that the readiness of the defense is an important factor.
- A sufficiently poor hitter (as measured by a "wOBA" of less than .300) can bunt even if the defense is expecting it.
- For both good hitters and bad hitters, the run expectation after a sac bunt is about the same as before the sac bunt. This suggests that the defense is properly playing the odds in both cases.
- If two on-deck hitters are equally proficient offensively, but one has a high walk rate and OBP, and the other has a lower walk rate and OBP (but a higher SLG), the sacrifice is more productive if the lower OBP hitter is on deck.
- The bunting skill and foot speed of the batter are sufficiently important that a good bunter can (productively) bunt any time of the game.
- The speed of the runner on first does not make the bunt any more productive, except in the 9th inning of a close game, where a faster runner is better.
- For non-pitchers, the count on the batter affects the results of the bunt approximately the same way it affects the results if the batter swings away, so the count is not particularly relevant in deciding whether or not to bunt.
- Pitchers are signficantly worse bunters than non-pitchers.

- Only poor-hitting pitchers (relative to other pitchers) should bunt regularly. Good-hitting pitchers should bunt only occasionally. The same applies with a runner on second and no outs. With one out instead of zero outs, only the worst hitting pitchers should bunt. With runners on first and second with no outs, all pitchers should bunt.

- With a runner on second and no outs, the successful sacrifice (batter out, runner advancing) is always negative, except in the 9^{th} inning when tied or down by a run.

I didn't have to do much rereading to summarize these findings, because every study in "The Book" recaps its conclusions in a little box – there's one of these every three or four pages, which gives you some idea of how much new information TMA have uncovered.

But one annoying aspect to the book is its insistence on denominating almost every study in a statistic called "wOBA." Its advantage is that it's denominated in roughly the same terms as OBA, so you can intuitively see whether a number is good or bad.[1] The disadvantage is that I'm not all that used to OBA in the first place, and I don't know how to easily convert wOBA to runs. I would rather have seen them use OPS. (But the other book, BBTN, likes to use "eqA," a similar stat scaled like batting average, which is only a bit easier.)

BBTN also has a chapter on the sacrifice; there, James Click calculates that a batter can benefit his team by bunting when his average is below .191 and/or his OBP is below .206. Comparing this to TMA's "good-hitting pitchers should bunt only occasionally," the two books appear to confirm each other.

One case where the two books disagree is on the pitching rotation – four-man, or five? In BBTN, Keith Woolner says that starters actually did better on three days rest than on four days rest, dropping their ERA 8 points. But TMA says the opposite, that pitchers on three days rest see a *rise* in their opponents' wOBAs from .352 to .369. I don't know what 17 points of wOBA means to ERA, but it must be significant – TMA says it doesn't make up for the additional innings the staff aces pitch under the four-man rotation.

Whom to believe? My gut suggests going with the short rotation and giving the extra innings to the better pitchers. But if sabermetricians don't trust Joe Morgan's gut, they certainly shouldn't trust mine.

The measure of books like these, I think, should be what they teach us about baseball. As I wrote earlier, most of their contents flesh out details about what we already knew. But there are still some important new findings in each. I'll list some of them:

- Derek Jeter's defense improved dramatically from 2003 to 2004, and the improvement carried over to 2005. (BBTN)

- Players are significantly less effective as pinch hitters than as starters – their OBP drops from .337 to .313, even after correcting for the pitchers they face. Designated hitters exhibit a similar drop about half as large. (TMA)

- Players appear to perform a bit better in the last year before free agency. (BBTN)

- In the draft, high-school player picks are turning out to be almost even in worth with college picks (high-schoolers used to have significantly less expected value). In the first round, high-school pitchers don't pay off very well, but are almost equal to high-school hitters and college pitchers after that. College hitters remain significantly most valuable.[2] (BBTN)

- Batting the pitcher eighth can add "a couple" of runs per year. (TMA)

- Coors Field doesn't just increase ball travel – it decreases the number of pitches that are swung on and missed. This is statistical evidence confirming the belief that the pitchers' delivery is affected by the thin air. (BBTN)

- A strategy of always pinch hitting for your fourth and fifth starter (and having them relieve each other in alternate games) can create .42 extra runs per game – as much a signing a superstar player. (TMA)

[1] A rough description of wOBA: take the most accurate Linear-Weights-type formula available. Remove outs, but increase every other event by the value of the out. Multiply the result by 1.15, and divide by PA.

[2] This is an update of Bill James' famous 1985 study – but, incredibly, in the entire chapter, BBTN doesn't mention the earlier study even once! This is the most significant example of a frustration I had with both books – many of their studies replicate previous research, but references to prior work are few.

- In this "juiced era" of steroid use, sudden power spikes are most evident among players of moderate power (10-30 home runs). Low and high-power hitters do not show any more spikes than in previous eras. (BBTN)

- Batters who tested positive for steroids (and who presumably stopped using afterwards) subsequently dropped by .010 BA/.014 OBP/.006 SLG. Pitchers who were caught saw their ERAs rise by 0.13. (BBTN)

- There is no evidence that certain batters "own" (have the ability to be more successful against) certain pitchers, except, of course, for the general skills of the hitter and pitcher. (Both books, independently)

- Pitchers are better than average their first time through the order; about average their second time through; and worse than average their third time through. This suggests that either batters "learn" or pitchers get tired. (TMA)

- There is very slight evidence of a very small bit of clutch hitting skill (BBTN); or, there is evidence of clutch hitting skill with a standard deviation of .008 points of OBP, but not enough data to be able to say who the best clutch hitters are. (TMA)[3]

- In terms of revenue, an extra win is worth up to $4.4 million for a team on the cusp of making the playoffs, but only $750,000 for a below-average team where the win won't matter, or a 100-game team where the 101st win won't make a difference. (BBTN)

- Almost all pitchers are worse from the stretch, but to differing extents. The average intrinsic drop is .005 of wOBA, with a standard deviation of .005. There is not enough data to be able to draw conclusions about which pitchers are which. (TMA)

- The evidence suggests that relievers can, with no ill effects, handle pitching more than one inning, and pitching on less rest – up to a total 40% increase in workload. (TMA)

This review is not meant to choose one book over the other; both books are interesting and valuable, and I'd recommend both books wholeheartedly. "Baseball Between the Numbers" would have more appeal to the serious fan with a casual interest in sabermetrics; its subject matter is more varied, and its spends more time talking about the results and what they tell us. It discusses the issues intelligently, and even those sections that aren't strictly about sabermetrics or empirical research (like the chapter on salary caps) are thought-provoking and suggest further study.

On the other hand, "The Book" is for the hardcore sabermetrician. I've never seen so many sabermetric studies crammed into one place. And they're done *right* – that is, when there's a subtle reason why the results might have to be adjusted, TMA anticipate it and deal with it. The book also acknowledges that certain managerial decisions have a tactical side; BBTN doesn't. In fact, BBTN includes a ranking of managers by their strategic decisions, and argues that even the best managers cost their team a game a season by (for instance) attempting a sacrifice even when the percentages are against them. But TMA notes that you have to bunt sometimes, even when it looks like a bad move, to keep the defense honest; otherwise, they'd play the third baseman back, and batters would never be able to take advantage and slap a base hit over his head into left field. "The Book" discusses all this in a chapter on game theory.

Or take their analysis of steals. At one point, they find the breakeven success rate for a certain steal is 69%, and the real-life success rate is 68%. At first glance, it looks like the managers are doing things exactly right, no? But that's not the case. If the breakeven point is 69%, managers should be sending runners when the success rate will be anywhere between 69% and 100%. The *lowest* steal attempt will be at 69%, and so the overall rate should be well over 70%.

It's a subtle point, but "The Book" is on it; they explain it casually in one sentence, and get on with their analysis. It's not a big deal, but it's the reason I like the book so much – you get the impression that these guys take their work seriously enough to get it right. It feels like they've been refining this stuff for years, they've thought everything through, and that's why they're able to explain it so well. Even when it gets a little too detailed for my taste, it's still a pleasure to read.

Phil Birnbaum, birnbaum@sympatico.ca ♦

[3] TMA do not explain in full how they figured this out, but they give a sketch.

A Comparison of Catcher Evaluation Statistics

Keith Carlson

A recent book evaluating catchers prompts the author to review the results of several catcher ratings published over the past few years.

McFarland has recently published a book by William McNeil about catching and catchers; it is entitled *Backstop* and subtitled "A History of the Catcher and a Sabermetric Ranking of 50 All-Time Greats." This is an impressive book and I recommend it. Personally, I like lists—especially when the underlying methodology is explained clearly. McNeil does a good job of explaining, but this note is really not a review. There is much to quibble with in the book, especially for the sabermetrically-inclined, but rather than get involved in nitpicking, I thought his listings would take on greater meaning if compared with those of other analysts, especially those that are quantitative in nature. McNeil discusses Pete Palmer's system as well as Bill James', but there are other ranking systems that barely get mentioned, if at all. The table on the next few pages summarizes McNeil's rankings along with those of seven analysts including Palmer and James, all of whom consider both offense and defense. Some of the systems are run-based (James, Palmer, Total Baseball 8, and Davenport) while the others use some kind of a point system, which vary in degree of sophistication. For details on the methodology underlying all of these rankings, the reader is referred to the original sources. There are many more lists in other publications and on the internet, but there aren't many that are based on an evaluation of the *overall* performance of the player, i.e., including both offense and defense. You can argue indefinitely about which of these systems are sabermetrically solid, but I will leave that subject for another day.

Table Columns

In the table, "G" is the number of games *caught* by the player during the period 1920-2004 (from the Lahman database). The order of players listed is first the McNeil ranking of his chosen 67 followed by an alphabetical listing of the others. An asterisk indicates a player in the Hall of Fame. Plus signs are players who were still active in 2004. A double asterisk in a ranking column means that the player did not make that analyst's list probably because he was active at the time the list was made and/or he didn't meet the eligibility requirement.

McNeil -- William F. McNeil, *Backstop* (McFarland & Company, Inc., 2006). McNeil's focus is on the skills of the catchers rather than value to the team. As a result, he deemphasizes longevity in his evaluations. He selects 67 catchers for analysis but focuses on the top 50. He excludes those who played most or all of their careers in the 19th century and the Deadball Era (before 1920). He ranks his selected players with a 2-to-1 weighting ratio for offense-to-defense. The ranking that I show for him in the table, however, is the 1-to-1 weighting that he gives in an appendix. I think this makes his list more comparable with the others. His lists are based on player data through 2003.

Rosciam -- Chuck Rosciam, *Encyclopedia of Catchers* at http://www.baseballcatchers.com. His ranking is based on data through 2002 and is limited to catchers with at least 800 games caught. I would describe it as a sophisticated point system. His list has been renumbered to exclude those who played the bulk of their careers before 1920.

James -- Bill James, *The New Bill James Historical Baseball Abstract* (Free Press, 2003), pp. 370-432. His list is based on the Win Shares system but includes a subjective element as well as consideration of peak versus career value. It's not clear if he has a cutoff on number of games caught. His results are based on data through 2000.

Palmer -- Pete Palmer and Gary Gillette, editors, *The 2005 ESPN Baseball Encyclopedia* (Sterling Publishing Co., Inc., 2005). This ranking is based on BFW (batter/fielder wins) for all catchers on the previous three lists (McNeil, Rosciam and James) plus all others with more than 750 games caught. The authors use data through 2004 in their calculations.

TB8 -- John Thorn, Phil Birnbaum and Bill Deane, editors, *Total Baseball*, eighth edition (Sport Media Publishing, Inc., 2004). This ranking is based on TPR (Total Player Rating), a measure first developed by Palmer and presented in the previous seven editions. The methodology in the eighth edition follows that developed by Palmer but the method of measuring defense appears to have been modified for catchers. Although TPR and BFW for catchers are highly correlated (.92), there are substantial differences for some players. The differences reflect Palmer's newly developed estimates of stolen base-caught stealing data and also the fact that he includes measures of pitcher performance, which TB8 does not do. The catchers included are the same as those on the Palmer list, but the rankings are based on data through 2003.

(continued after table)

	G	McNeil	Rosciam	James	Palmer	T-B 8	Dvprt	Faber 1	Faber 2	Peters
Hartnett, Gabby *	1793	1	2	9	1	2	6	5	4	3
Campanella, Roy *	1183	2	5	3	12	16	21	7	13	5
Dickey, Bill *	1708	3	1	7	7	7	7	2	7	6
Berra, Yogi *	1699	4	4	1	6	4	5	1	1	1
Cochrane, Mickey *	1451	5	3	4	9	8	12	4	8	4
Bench, Johnny *	1742	6	6	2	2	3	1	3	2	2
Carter, Gary *	2056	7	8	8	4	6	3	6	5	7
Piazza, Mike +	1429	8	7	5	3	1	10	**	**	**
Freehan, Bill	1581	9	15	12	23	17	16	12	12	18
Hoiles, Chris	819	9	19	**	34	31	60	**	**	**
Parrish, Lance	1818	11	31	19	15	36	11	38	31	20
Battey, Earl	1087	11	41	50	27	20	57	11	9	10
Howard, Elston	1138	13	30	15	40	42	33	15	16	14
Torre, Joe	903	14	13	11	14	9	8	10	3	9
Fisk, Carlton *	2226	15	14	6	5	5	2	8	10	8
Tenace, Gene	892	15	33	23	10	11	19	77	70	21
Rodriguez, Ivan +	1688	17	9	13	8	13	4	**	**	**
Sundberg, Jim	1927	18	43	32	29	27	14	58	46	22
Ferrell, Rick *	1806	18	17	29	32	35	25	36	32	45
Scioscia, Mike	1395	18	44	36	22	46	35	22	18	27
Munson, Thurman	1278	21	29	14	11	15	17	14	14	11
Roseboro, Johnny	1476	22	50	27	52	69	36	25	29	65
Crandall, Del	1479	23	42	30	45	48	26	20	26	12
Lombardi, Ernie *	1544	24	20	22	18	12	22	13	11	15
Ruel, Muddy	1317	25	32	51	54	50	47	32	35	46
Johnson, Charles +	1141	25	18	**	53	73	52	**	**	**
Pena, Tony	1950	27	57	34	26	100	18	19	15	24
Porter, Darrell	1506	27	23	18	17	18	13	34	36	26
Simmons, Ted	1771	29	21	10	16	47	9	54	49	**
Lopez, Javy +	1238	29	28	**	13	61	28	**	**	**
Danning, Harry	801	29	36	55	49	10	78	9	6	13
Mueller, Ray	917	32	77	**	68	43	84	59	59	**
Hargrave, Bubbles	720	32	**	85	46	60	114	83	91	**
Lollar, Sherm	1571	34	16	31	28	33	27	78	75	39
Tettleton, Mickey	872	34	35	37	41	26	37	17	19	17
Ferguson, Joe	766	34	**	79	31	19	69	66	50	19
Rosar, Buddy	934	37	61	**	81	84	88	80	68	66
Cooper, Walker	1223	38	49	33	33	32	40	21	24	16
Daulton, Darren	965	39	26	25	30	25	48	75	67	34
O'Dea, Ken	627	40	**	**	65	80	122	**	**	**
Kendall, Jason +	1205	41	10	**	21	22	31	**	**	**
Haller, Tom	1199	42	24	26	25	28	42	33	37	38
O'Farrell, Bob	1249	42	34	46	39	41	54	60	65	29
Tebbetts, Birdie	1108	42	76	64	108	130	91	67	83	48
Phelps, Babe	592	45	**	98	50	38	109	**	**	**
Westrum, Wes	902	45	51	104	43	64	98	105	113	**
Bailey, Ed	1064	47	38	39	51	39	50	53	55	30
Triandos, Gus	992	48	80	56	64	53	72	55	47	36
Burgess, Smoky	1139	49	25	28	37	34	45	45	44	23
Sanguillen, Manny	1114	50	55	42	61	71	46	18	17	57
Lopez, Al	1918	51	52	41	69	30	23	30	22	49
Yeager, Steve	1230	52	103	78	75	96	75	63	85	**
Boone, Bob	2225	53	69	21	105	132	15	16	21	44
Hogan, Shanty	908	54	27	94	67	67	101	50	57	40

	G	McNeil	Rosciam	James	Palmer	T-B 8	Dvprt	Faber 1	Faber 2	Peters
Edwards, Johnny	1392	55	70	53	66	79	41	35	28	43
Hegan, Jim	1629	56	83	44	100	146	53	26	62	37
Davis, Spud	1282	57	12	71	55	23	56	27	20	31
Mancuso, Gus	1360	58	82	74	77	101	58	42	66	50
Santiago, Benito +	1911	59	56	**	86	105	20	69	61	54
Wilson, Dan +	1270	60	68	**	101	116	64	**	**	**
McCarver, Tim	1387	61	59	24	42	56	38	28	23	42
Wilson, Jimmie	1351	62	73	63	125	112	94	41	45	47
Cerone, Rick	1279	63	119	101	136	143	95	48	48	**
Hundley, Randy	1026	64	112	109	126	141	117	**	**	**
Seminick, Andy	1213	65	58	57	38	29	43	84	77	28
Lieberthal, Mike +	962	66	**	**	56	66	59	**	**	**
Taubensee, Eddie	871	67	65	**	98	85	128	**	**	**
Alomar, Sandy +	1237		84	68	138	118	70			
Ashby, Alan	1299		90		107	90	76	56	56	
Ausmus, Brad +	1424		54		85	93	39			
Azcue, Joe	868			117	82	91	118	49	34	
Bassler, Johnny	730			47	57	37	107	92	87	
Bateman, John	953				140	133	136	110	106	
Benedict, Bruce	971				133	122	124	82	76	
Berry, Charlie	657									
Berryhill, Damon	590							101	107	
Borders, Pat +	976				109	121	110	91	100	
Brenly, Bob	705			102				93	80	
Brown, Dick	614									
Cannizzaro, Chris	714									
Casanova, Paul	811				148	147	148	111	111	
Courtney, Clint	802		75	123	123	81	139	98	81	
Dalrymple, Clay	1003		92	107	90	78	83	90	92	
Davis, Jody	1039		71	90	91	82	82	62	53	55
Dempsey, Rick	1633		86	43	59	74	34	31	54	67
Desautels, Gene	699									
Diaz, Bo	965		99	97	96	108	102	61	60	70
Diaz, Einar +	598									
Downing, Brian	675									
Duncan, Dave	885				141	134	138	97	105	
Dyer, Duffy	634									
Early, Jake	694									
Essian, Jim	642									
Etchebarren, Andy	931				94	110	121			
Evans, Al	647									
Fabregas, Jorge	595									
Fitz Gerald, Ed	651									
Fitzgerald, Mike	748									
Flaherty, John +	987				146	137	142			
Fletcher, Darrin	1143		60		120	106	99			
Foote, Barry	637									
Fordyce, Brook +	591									
Fosse, Ray	889			106	89	95	119	64	58	
Garagiola, Joe	614									
Gedman, Rich	979			105	92	99	100	88	94	56
Girardi, Joe	1247		89		142	144	86			
Gooch, Johnny	758				102	117	144	113	115	
Grote, Jerry	1348		93	66	78	109	66	47	39	80
Harper, Brian	688			99	93	87	106	76	72	60

	G	McNeil	Rosciam	James	Palmer	T-B 8	Dvprt	Faber 1	Faber 2	Peters
Hassey, Ron	946		67	89	88	59	79	71	69	52
Hayes, Frankie	1311		47	75	114	45	68	96	78	61
Hayworth, Ray	677									
Heath, Mike	1083			108	117	102	80	95	88	
Hemsley, Rollie	1482		87	69	124	126	61	87	95	41
Hendricks, Ellie	602									
Henline, Butch	608									
Hernandez, Ramon +	699									
Herrmann, Ed	817				99	77	133	86	73	
Hill, Marc	687									
Hundley, Todd	1096		72	119	84	55	90			
Karkovice, Ron	918				76	94	92			
Kendall, Fred	795				147	140	147	112	108	
Kennedy, Terry	1378		63	52	127	86	49	44	38	
Kreuter, Chad	892				95	88	112			
Landrith, Hobie	677									
Laudner, Tim	657									
LaValliere, Mike	850		74	91	60	65	81	73	90	
Lopata, Stan	695			92	58	44	89	108	103	
Macfarlane, Mike	1058		64	84	74	76	65			
Manwaring, Kirt	993				144	136	132			
Martin, J.C.	692									
Martinez, Buck	1008				132	139	130	100	109	
Masi, Phil	1101		62	93	87	70	73	74	74	59
Matheny, Mike +	1107				137	142	97			
May, Milt	1034		81		80	72	77	52	42	74
Mayne, Brent +	1143		85		134	128	108			
McCullough, Clyde	989		95	114	110	98	115	102	93	
Melvin, Bob	627									
Miller, Damian +	693									
Mitterwald, George	796				97	111	127	114	112	
Moore, Charlie	894			96	103	119	96	106	104	
Moss, Les	720									
Myatt, Glenn	734									
Myers, Greg +	890				122	107	123			
Nixon, Russ	722									
Nokes, Matt	689			124						
O'Brien, Charlie	782				73	97	113			
Oliver, Joe	1033		96		111	125	105			
O'Neil, Mickey	643									
O'Neill, Steve	687		39	54	48	63	62	37	51	69
Ortiz, Junior	702									
Owen, Mickey	1175		91	88	128	123	104	65	89	
Pagliaroni, Jim	767				70	57	126	103	99	
Pagnozzi, Tom	827			120	130	114	116			
Perkins, Cy	952		78	113	113	103	135	107	101	62
Picinich, Val	781		98		104	92	129	116	116	
Posada, Jorge +	955				20	21	32			
Pytlak, Frankie	699			112						
Rader, Dave	771				131	115	143	109	110	
Reed, Jeff	1071		94		129	124	120			
Rice, Del	1249		100	87	112	127	85	57	71	77
Rodgers, Buck	895				135	135	137	104	97	
Rodriguez, Ellie	737									

	G	McNeil	Rosciam	James	Palmer	T-B 8	Dvprt	Faber 1	Faber 2	Peters
Romano, Johnny	810		40	73	36	24	74	70	64	
Roof, Phil	835				118	113	145			
Ryan, Mike	632									
Schalk, Ray *	822		22	35	24	68	30	24	27	58
Schang, Wally	961		11	20	19	14	24	29	33	32
Servais, Scott	792				116	104	140			
Severeid, Hank	744		37	70	106	83	87	51	43	76
Sewell, Luke	1562		79	59	139	148	93	72	98	63
Sims, Duke	646									
Slaught, Don	1237		66	67	72	51	63	46	40	75
Smith, Earl	706			100	47	52	125	89	102	
Smith, Hal	648									
Snyder, Frank	717		46	61	62	58	67	23	25	73
Spohrer, Al	731									
Stanley, Mike	751				79	40	55			
Stearns, John	699							99	86	
Steinbach, Terry	1381		45	38	63	49	29	39	52	68
Surhoff, B.J.	704									64
Swift, Bob	980				145	145	146	115	114	
Taylor, Zack	856				143	131	141	81	84	
Tillman, Bob	725									
Todd, Al	752				119	120	134	85	82	
Tresh, Mike	1019			125	115	129	111	79	96	
Trevino, Alex	742									
Valle, Dave	902		97		83	89	103	94	79	
Varitek, Jason +	788				44	62	71			
Virgil, Ozzie	677									
Wagner, Hal	626									
Walbeck, Matt	651									
White, Sammy	1027		88	111	121	138	131	68	63	71
Whitt, Ernie	1246		53	72	35	54	51	40	30	33
Wilkins, Rick	650									
Wynegar, Butch	1247		48	65	71	75	44	43	41	77

Davenport -- Clay Davenport, *Davenport Translations* at www.baseballprospectus.com. For his ranking I used WARP3 (wins above replacement with various adjustments). Davenport's methodology is not all that clear but his results are readily available and have been for quite some time. The catchers are the same as those on the Palmer list, except that the data are through 2005.

Faber 1, Faber 2 -- Charles F. Faber, *Baseball Ratings*, second edition (McFarland & Company, Inc. 1995). Two ratings are shown for Faber and both are based on a point system using both batting and fielding data: Faber 1 includes a "handling pitcher" bonus which is derived from the team's win percentage and Faber 2 excludes the bonus. This study is somewhat dated but still of interest, in my opinion. Faber's criteria for eligibility are quite involved, but note he shows very few players with less than 750 games. His results are based on data through 1994.

Peters -- Frank P. Peters, *REAL Major League Baseball* (Self-published, 1996). This listing is also based on a point system, but more straightforward than Faber's. Batting points are constructed using conventional data but for fielding points he simply uses Palmer's fielding runs in *Total Baseball* (fourth edition). His cutoff point is 3400 plate appearances using data through 1995.

Comments on the table

What are we to make of all these rankings? Are McNeil's latest rankings out of line when compared with those of other analysts? Answering such questions is difficult because of different methodologies and, to some extent, different objectives, namely ability versus value. Nonetheless, some patterns emerge and some general conclusions can be drawn.

The “top seven” on McNeil’s list are all in the Hall of Fame, and there is almost universal agreement among the other analysts. Possible exceptions are Palmer, TB8, and Davenport in the ranking of Campanella; their systems reward or penalize longevity (good hitters are rewarded but poor hitters can be penalized). To a lesser extent Cochrane is also penalized. If a catcher’s defensive skills decline or stay unchanged over the years, continued success (or mediocrity) at the plate can raise or lower the player in these three rankings. Davenport’s methodology, in particular, seems to be most extreme in rewarding or penalizing longevity because it reflects runs above replacement rather than above average. Fisk appears relatively low in McNeil’s rankings, and this is because of his de-emphasis on longevity. Fisk had many years of solid, but not spectacular performance.

After you get beyond the top ten or so, the rankings go every which way. Hoiles appears to be ranked a little high and maybe Roseboro also. On the other hand, Munson and Simmons seem to be ranked a little low. Catchers like Torre, Tenace, Ferguson and Tettleton are special cases and difficult to pinpoint because they spent so much of their careers at other positions; all of the systems appear to use career batting performance in their evaluation, regardless of position. The largest differences between McNeil and the others are in the ranking of O’Dea, Phelps, Mueller and Hargrave, all of whom appear obviously overrated, and maybe Rosar and Tebbetts as well. .

Are there any viable candidates overlooked by McNeil? I think the answer is no. Looking at the other rankings and noting whether any of those unranked by McNeil are in the top 50 for any of the others, we see that those mentioned most often are Steinbach, Whitt and Wynegar. Omitting these players is not an egregious error, but they could well be candidates to replace the questionable ones on McNeil’s list.

Turning now to the active players. At this time Piazza and Rodriguez appear to be locks for the Hall of Fame, even if they might be in the declining phase of their careers. Aside from these two stars, among those actives on McNeil’s list Javy Lopez and Jason Kendall could move up, although Lopez is 35 years old. Posada is not on his list of the top 50, but McNeil recognizes him as a candidate who could move on to the top 50 list. He also receives support from other analysts.

Concluding comments

With only a few exceptions, the McNeil ranking appears sound. Given that there are different criteria for evaluating catchers, and also different methodologies, it is difficult to take issue with his list as a whole. One would think that evaluating offense would be consistent from one system to another since so much work has been done with these metrics. It is not obvious that this is the case, however. On the defensive side, there are obviously vastly different methodologies, since this is such a murky area for analysis. Most of the systems provide the breakdown between offense and defense, so it is possible to do a similar comparison of defensive rankings. A factor in McNeil’s favor is that he explicitly accounts for caught stealing in his system, as does Palmer (in fact, it appears that he uses Palmer’s estimates), but as everyone knows there is a lot more to catching than throwing out basestealers. How you measure these attributes is another matter. Does McNeil’s choice of Hartnett as number one defensively stand up when compared with others? Such a comparison waits for another day.

Keith Carlson, kcsqrd@charter.net ♦

Are Runs Scored and Runs Allowed Independent?

Ray Ciccolella

A recent derviation of the Pythagorean Theorem makes the assumption that runs scored and runs allowed in a game are independent of each other. The assumption makes for convenient calculations, but how true is it? This study investigates.

In an article elsewhere in this issue of BTN, Steven J. Miller of Brown University demonstrates how to derive the Pythagorean Formula using a Weibull distribution. While I cannot do his paper full justice, the most interesting portion of the paper to me was the author's conclusion that runs scored and allowed in a game are statistically independent once you adjust for the fact that they cannot be equal. This conclusion is counterintuitive to me, as I think there are at least three "environmental" factors that should drive runs scored and allowed to be correlated. In each game the ballpark, the weather conditions, and the home plate umpire are the same for each team. In addition the net impact of tactical in-game decisions, such as utilization of the bullpen, subs, sacrifices, and stolen base attempts might tend to increase the correlation of runs scored and allowed.

For this analysis I made the assumption that runs scored and runs allowed are independent than analyzed several relationships that should follow if that assumption were true. What I concluded is that runs scored and allowed are not independent but the degree of correlation is far less than I expected and can be probably be explained by the rules of the game concerning when the home team bats or not. Below I describe my methods and data in more detail.

Method 1 – Margin of Victory and One-Run Games

My first approach was to find the actual margin of victory and compare it to a randomly generated average margin of victory assuming runs scored and allowed are independent. I also checked the percentage of games that ended with a 1 run margin of victory for both the actual results and my randomly generated results.

I used the 1999 season for my data set since I had already the game by game scores from each team from some previous analysis. The runs scored per game for this season are in-line with figures for entire period of 1996 though 2002, though towards the higher end of the range. Also the distribution of games by margin of victory is also similar to the distributions from 1996 through 2002, as is the percentage of games decided by 1 run. In short, there is nothing in the data to indicate that 1999 is anything but a normal year for my purposes.

For each team I randomly selected a runs scored number and runs allowed number based on that team's actual distribution of runs scored and allowed. For each pair, I calculated the difference between runs scored and allowed. I replicated 5 seasons worth of results for each team.

If the randomly selected pair turned out to be equal I categorized it as an extra inning game. From a sample of about 250 extra inning games over 3 seasons (1999 through 2001) I determined that the average (mean) margin of victory in extra inning games was about 1.5 runs and that 73% of extra inning games resulted in a one-run margin of victory.

I'll use the Cardinals as an example. In one of the five replicated seasons they had 146 games in which runs scored and runs allowed were not equal. For these games the total absolute difference between runs scored and allowed was 533 runs and 39 of these games had a one-run margin of victory. They had 15 games (out of 161) in which the randomly selected runs scored and runs allowed were equal. For the 15 "extra inning" games I added 23 runs to the absolute difference (15 multiplied by 1.5 rounded up to 23) and 11 games with a one-run margin of victory (73% multiplied by 15).

The adjusted totals for the Cardinals then become an average (mean) absolute difference of 3.45 runs per game (533 plus 23 divided by 161) with 31% of their games having a margin of victory of one run (39 plus 11 divided by 161). I followed this same procedure for each team. Table 1 shows the tabulated results.

The actual average margin of victory is less than what I observed in the randomly generated results for 26 of the 30 teams and across both leagues. Also there were fewer games decided by one run in the randomly generated results than actually occurred in 1999. Interestingly, the Rockies, with the largest park factor (by far), had the largest discrepancy between actual and randomly generated average margin of victory.

If runs scored and runs allowed were independent I would not have expected these results. I was, however, surprised at how close the totals were and how close the actual and expected results were for many teams.

Table 1 – 1999 Average (mean) Margin of Victory and Percent of 1 Run Games

	Mean Margin of Victory		% of 1 Run Games	
Team	Actual	Random	Actual	Random
Diamondbacks	3.37	3.79	30%	25%
Braves	3.77	3.71	31%	26%
Cubs	3.51	3.88	31%	24%
Reds	3.35	3.70	28%	25%
Rockies	3.65	4.47	29%	22%
Marlins	3.57	3.69	27%	22%
Astros	3.57	3.52	25%	27%
Dodgers	3.47	3.67	30%	28%
Brewers	3.48	4.08	30%	21%
Expos	3.55	3.81	27%	26%
Mets	3.58	3.70	28%	25%
Phillies	3.98	3.92	27%	25%
Pirates	3.53	3.61	26%	25%
Cardinals	3.04	3.59	33%	25%
Padres	3.41	3.55	33%	27%
Giants	3.60	3.69	31%	28%
NL Total	3.53	3.77	29%	25%
Angels	3.72	3.81	30%	24%
Orioles	3.53	3.84	26%	25%
Red Sox	3.77	3.72	25%	25%
White Sox	3.94	4.02	24%	23%
Indians	3.72	4.16	28%	20%
Tigers	3.97	3.98	25%	23%
Royals	3.66	3.83	27%	25%
Twins	3.48	3.81	28%	23%
Yankees	3.82	3.92	21%	21%
A's	3.81	3.90	25%	26%
Mariners	4.09	4.26	27%	24%
Devil Rays	3.54	3.91	29%	23%
Rangers	4.06	4.09	25%	23%
Blue Jays	3.77	3.96	27%	21%
AL Total	3.78	3.94	26%	23%
TOTAL MLB	3.64	3.85	28%	24%

Method 2 – Winning Percentage versus Runs Scored in a Game

For my second approach I compared actual and expected wins at each level of runs scored assuming the independence of runs scored and allowed. I'll use the Cardinals as example with the key information displayed in Table 2 below. Brackets in the last column indicated actual wins were less than expected wins.

In 1999 the Cardinals scored zero runs in a game 4 times, one run 16 times, two runs in 19 games, and so on. They allowed zero runs 3 times, one run 10 times, etc. When the Cardinals scored zero runs they won zero games, when they scored one run they won 2 games, when they scored two runs they won 3 games, etc.

If runs scored and allowed were independent, how many games would we expect them to win for at each level of runs scored? For example, how many games could we expect them to win when they scored 4 runs? The Cardinals allowed fewer than 4 runs in 48 games (3 plus 10 plus 17 plus 18). They also allowed exactly 4 runs in 28 games. Counting those games as half wins makes an additional 14 games, totalling 62 games (48 plus 14) out of 161, or 38.51%.

The Cardinals scored 4 runs in a game 22 times. Assuming they win 38.51% of those games that translates to 8.47 projected wins. I followed this logic across all levels of runs scored (which is why there are expected wins at zero runs scored) to create the table above. I then performed the same calculation for all teams with the results summarized in Table 3 below. Brackets in the last column again indicate actual wins were less than expected wins.

At the total level, expected wins are almost exactly equal to actual wins but there are patterns in the differences based on runs scored. In general at low and high levels of runs scored actual wins are less than expected wins but that is offset by the results at 2, 3, 4, and 5 runs scored. Table 4 below shows the delta between actual minus expected wins by team but with the runs scored in a game grouped.

Table 2 – 1999 Cardinals Results by Runs Scored and Allowed

Runs	Count RS	Count RA	Exp Wins	Actual Wins	Delta
0	4	3	0.04	0	(0.04)
1	16	10	0.80	2	1.20
2	19	17	2.54	3	0.46
3	16	18	3.88	3	(0.88)
4	22	28	8.47	8	(0.47)
5	18	28	10.06	10	(0.06)
6	17	13	11.67	14	2.33
7	16	11	12.17	9	(3.17)
8	15	10	12.39	11	(1.39)
9	6	8	5.29	5	(0.29)
10	4	6	3.70	3	(0.70)
11+	8	9	7.78	7	(0.78)
Total	161	161	78.78	75	(3.78)

Table 3 – 1999 Actual and Predicted Wins by Runs Scored, All Teams Combined

Runs	Count RS	Count RA	Exp Wins	Actual Wins	Delta	Error
0	193	193	3.92	0	(3.92)	0%
1	430	430	35.67	27	(8.67)	-32%
2	546	546	100.05	105	4.95	5%
3	612	612	185.52	191	5.48	3%
4	626	626	268.24	283	14.76	5%
5	555	555	307.37	323	15.63	5%
6	452	452	298.62	289	(9.62)	-3%
7	393	393	293.40	277	(16.40)	-6%
8	320	320	261.63	269	7.37	3%
9	222	222	194.35	199	4.65	2%
10	158	158	144.45	134	(10.45)	-8%
11+	347	347	334.27	330	(4.27)	-1%
Total	4854	4854	2,427.49	2427	(0.49)	0%

Overall teams won fewer games than expected if runs scored and allowed were independent when they scored 2 runs or less and when they scored 6 runs or more. They won more games than expected when they scored 3 to 5 runs in a game. The results are not universal as there are teams that buck this trend and there are differences between the leagues but the overall pattern seems clear.

These results again indicate to me that runs scored and runs allowed are not completely independent. If teams win less often than we would otherwise expect when they score a high number of runs that is means they must allow more runs than normal in those games.

As a check I performed the same types of calculations but used the runs scored and runs allowed distribution for all teams combined (instead of team by team) from the 1990, 1996, and 2005 seasons. These results are shown in Table 5 below and they match the overall pattern seen in Tables 3 and 4.

Method 3 – Comparison to Previous Work

In the February 1999 edition of "By the Numbers" summary data on the relationship between runs scored and allowed was presented (article by Clifford Blau, data from Tom Ruane). Paraphrasing from that article, the most common number of runs to score when losing is one less than the number allowed.

I performed the same analysis, on a much smaller scale and obtained essentially the same results. My sample consisted of 12 teams randomly selected from the 1999 through 2003 seasons. These teams had an actual winning percentage of .495 versus a Pythagorean predicted percentage of .501, and 26.4% of their games were decided by one run. Their runs scored and allowed averages (mean, median, and mode) were in-line with league averages. I believe these sample teams fairly represents a typical group of teams from this time period.

I reviewed the pattern of runs scored and allowed for these teams and found multiple instances that were inconsistent with runs scored and runs allowed being independent. A few examples are shown below.

- Except for shutout losses, the most common number of runs to score when losing was one less than the number allowed for all levels of runs scored.

- When winning, the most common number of runs to allow was one less than the number scored for all levels of scoring up to 6 runs.

- Teams that scored seven runs allowed 6 or 8 runs more often they allowed 5 runs even though 5 runs allowed was more common overall.

- Teams that scored 9 runs allowed 8 runs more often than they allowed 4, 5, 6, or 7 runs even though 8 runs allowed occurred less frequently overall.

- Teams that scored 1 run allowed 2 runs more frequently than they allowed 3 and 4 runs even though allowing 3 and 4 runs occurred more frequently overall.

- Teams scoring 6 runs allowed 5 more often than they allowed 3 and 4 runs.

- Teams that scored zero runs allowed 1 run more often they allowed 2 and runs even though 2 and 3 runs allowed were both more common than 1 run allowed.

If runs scored and runs allowed were completely independent, we would not expect to see results like these.

Table 4 – 1999 Actual Less Predicted Wins by Runs Scored Groups

	0 to 2	3 to 5	6 +	Total
Diamondbacks	(0.06)	6.54	(5.38)	1.11
Braves	(0.54)	5.24	2.38	7.08
Cubs	0.91	(3.03)	2.78	0.65
Reds	1.49	4.34	(3.71)	2.13
Rockies	(1.18)	5.63	(5.07)	(0.62)
Marlins	0.46	(0.43)	(2.52)	(2.48)
Astros	(0.55)	0.96	(0.37)	0.04
Dodgers	0.17	(3.52)	(1.62)	(4.97)
Brewers	1.55	2.78	(4.02)	0.31
Expos	(1.07)	0.67	(0.69)	(1.08)
Mets	(1.21)	2.47	0.52	1.79
Phillies	2.28	(0.97)	(3.71)	(2.41)
Pirates	(2.51)	(1.66)	4.71	0.54
Cardinals	1.63	(1.41)	(4.00)	(3.78)
Padres	1.41	(2.93)	1.13	(0.38)
Giants	(1.65)	3.20	(1.04)	0.52
NL Total	1.14	17.89	(20.61)	(1.57)
Angels	1.73	(0.06)	(0.01)	1.66
Orioles	1.43	1.48	(8.85)	(5.94)
Red Sox	(4.37)	5.02	0.40	1.06
White Sox	(3.03)	0.63	1.55	(0.84)
Indians	0.52	1.17	2.85	4.54
Tigers	(1.94)	1.02	(0.37)	(1.28)
Royals	(1.60)	(7.24)	0.99	(7.85)
Twins	1.14	(5.34)	2.09	(2.12)
Yankees	0.83	3.82	0.62	5.27
A's	0.70	1.73	0.18	2.61
Mariners	(1.54)	4.52	(4.12)	(1.15)
Devil Rays	(0.61)	5.15	(4.71)	(0.17)
Rangers	(3.17)	6.87	1.02	4.73
Blue Jays	1.11	(0.81)	0.25	0.55
AL Total	(8.79)	17.97	(8.11)	1.08
MLB Total	(7.65)	35.87	(28.71)	(0.49)
Total Games	1169	1793	1892	4854
Error	(0.7%)	2.0%	(1.5%)	0.0%

Summary and Discussion

In this article I examined whether runs scored and allowed are independent. I utilized multiple methods and found that my results were inconsistent with the assumption that runs scored and runs allowed are independent. I did have to make several assumptions in my analysis, especially how to handle "tie" games. While I think these assumptions represent reasonable choices and tradeoffs there are probably other reasonable alternatives that, if I had chosen those, might have yielded different results.

I was, however, surprised at my results to some extent. I actually expected that the lack of independence would appear stronger. My results and overall trends were not observed in all teams. For the results shown in Tables 3, 4, and 5 the sign of the error is the same for both low scoring and high scoring games. I would have expected the signs to be opposite. Lastly, the percentage errors shown in Table 3 and the average deltas shown in Table 5 are really fairly small.

So what is different from what I found and what Mr. Miller concluded? First, we approached the question differently. It could be, for his purposes, the degree of independence is enough for his derivation. Second, the strength of the correlations I found is probably weak, and certainly weaker than I expected.

Table 5 – Actual versus Expected Winning Percentage by Runs Scored

Runs Scored	1990 Actual	1990 Exp	1996 Actual	1996 Exp	2005 Actual	2005 Exp	Avg Delta
0	-	0.031	-	0.022	-	0.027	(0.027)
1	0.090	0.118	0.070	0.084	0.082	0.104	(0.021)
2	0.239	0.245	0.204	0.184	0.254	0.218	0.017
3	0.389	0.390	0.324	0.310	0.347	0.350	0.003
4	0.573	0.536	0.466	0.441	0.487	0.483	0.022
5	0.681	0.661	0.551	0.561	0.615	0.606	0.006
6	0.758	0.761	0.690	0.668	0.697	0.714	0.004
7	0.847	0.834	0.766	0.756	0.824	0.799	0.016
8	0.878	0.885	0.773	0.822	0.832	0.861	(0.028)
9	0.898	0.922	0.825	0.874	0.899	0.908	(0.027)
10	0.944	0.945	0.890	0.914	0.900	0.939	(0.021)

I suspect that most of what I found in the correlation of runs scored and allowed is caused the enforced correlation of runs scored created by the home team batting or not batting in the ninth inning. Also if the home team bats in the bottom of the ninth, or in extra innings, they stop batting as soon as they take the lead.

In addition I think park effects influence what I found to some degree. I believe if I had broken out the runs scored and allowed distributions by home and road games some of the correlation I did find would be reduced further.

To crudely examine this point I did a quick test on one team with an extreme home park effect, the 2001 Colorado Rockies. I replicated my Method 1 from above but separated home and road games. What I found was that the randomly generated mean absolute difference between runs scored and allowed was almost exactly the same but slightly less than the actual results. The Rockies' actual average absolute difference between runs scored and allowed in road games was 3.46 while my randomly generated results were 3.42. For home games the actual figure was 4.43 while the random results were 4.35. When I did not separate home and road game runs scored and allowed the actual results for the Rockies was an average gap of 3.94 runs but the gap for the randomly generated results was 4.08. While results from one team certainly are not proof, this quick test showed that at least part of the correlation I found was probably created by park effects.

Lastly, there is clearly a lot of noise in this data. Teams use, and face, pitchers of different quality each game, the quality and style of play of opponent changes frequently, and there is the normal game to game variation that is part of each player's and team's performance. All these factors, and probably others, might impact the independence of runs scored and allowed.

Acknowledgments

I would like to thank Clifford Blau for his comments, particularly concerning the impact of home team batting in the ninth inning. I would especially like to thank Steven J. Miller for his extensive correspondence on this topic, his feedback, questions, and suggestions. I would like to thank Sal Baxumusa for sending me Dr. Miller's paper originally.

Ray Ciccolella, rciccolella@austin.rr.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 are all welcome.

Articles should be submitted in electronic form, either by e-mail or on CD. 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. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

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

Send submissions to:
Phil Birnbaum
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca

Announcement: Baseball, Statistics, and the Role of Chance in the Game

The Ohio Section of the Mathematical Association Association is sponsoring a short course on "Baseball, Statistics, and the Role of Chance in the Game" at Mt. Union College, Alliance, Ohio on July 7-9, 2006.

Jim Albert, Professor of Statistics at Bowling Green State University, will introduce several explorations of baseball data including the search for the ultimate batting statistic, looking for true streakiness and clutch ability, modeling run production by a Markov Chain model, and comparing great players such as Babe Ruth and Barry Bonds who played in different eras. This workshop will be directed both to instructors who wish to infuse their teaching of probability and statistics with applications from baseball and to baseball fans interested in learning about sabermetrics.

For more information, visit the meeting website at www.muc.edu/~zwilliml/shortcourse .

A Derivation of James' Pythagorean Projection

Steven J. Miller

Under certain statistical assumptions about the distribution of a team's runs scored and runs allowed, Bill James' Pythagorean Projection can be shown to follow mathematically. Here, the author explains.

1. Introduction

The goal of this paper is to show how Bill James' Pythagorean Win-Loss Formula follows from reasonable assumptions about how teams score and allow runs in a game. He observed that if RS_{obs} and RA_{obs} are the observed average number of runs a team scores and allows per game in a baseball season, then RS_{obs}^{γ} / $(RS_{obs}^{\gamma} + RA_{obs}^{\gamma})$ is an excellent predictor of a team's won-loss percentage, usually accurate to about four games a year. At first γ was taken to be 2 (which led to the name "Pythagorean," after the right-triangle formula), but fitting γ to the results of many seasons led to the best γ being about 1.82.

By modeling the runs scored and allowed per game for a team as continuous independent random variables drawn from Weibull distributions, we can provide a theoretical basis for this formula and value of γ; hopefully our model will be of use for additional investigations as well.

Our assumptions mean that we model a baseball game as follows. We first choose one number (the runs scored) from one continuous probability distribution, and then we choose another number (the runs allowed) independently from another continuous probability distribution. All we then need to do is calculate the odds that the first number is bigger than the second; this of course then leads to predicting the won-loss percentage for the team.

There are many probability distributions. While distributions such as the exponential or normal (i.e., bell curve) are common and well known, they only have one or two shape parameters; the Weibull has three. This makes it much easier to fit the observed baseball data with a Weibull distribution than with some of the better known distributions[1]. Further, the exponential decays too slowly to be realistic for baseball; it leads to too many games with large scores. By choosing our parameters appropriately, a Weibull has a much more realistic decay, which leads to very few games with a team scoring or allowing more than 20 runs. In the appendix, we illustrate how flexible the Weibull is by showing how appropriate choices of the parameters lead to terrific fits with the observed run.

The Weibull distribution is defined as follows: if $f(x;\alpha, \beta,\gamma)$ is the probability density of a Weibull with parameters (α,β,γ), then

$$f(x;\alpha,\beta,\gamma) = \left(\frac{\gamma}{\alpha}\right)\left(\frac{x-\beta}{\alpha}\right)^{\gamma-1} e^{-\left(\frac{x-\beta}{\alpha}\right)^{\gamma}} \quad \text{if } x \geq \beta, \text{ and 0 otherwise.}$$

For example, let's assume that for our favorite team the runs scored per game is given by a Weibull with parameters (5, 0, 2). Then the probability that a team scores between 1 and 3 runs in a game is just

$$\int_1^3 \left(\frac{2}{5}\right)\left(\frac{x}{5}\right) e^{\left(\frac{-x^2}{5}\right)} dx \approx 27\% .$$

The main consequence of our model is the following:

[1] There is also a technical reason for choosing a Weibull distribution: Weibull distributions lead to integrations that we can easily solve!

For a given team, suppose that the runs scored are drawn from a Weibull with parameters (α_{RS}, β, γ) and that the runs allowed are independently drawn from another Weibull with parameters (α_{RA}, β, γ). Choose these values so that the mean of the runs scored Weibull is RS_{obs} and the mean of the runs allowed Weibull is RA_{obs}. Then

$$\textbf{\textit{Won-Loss Percentage}} \quad = \quad \frac{RS_{obs}^{\gamma}}{RS_{obs}^{\gamma} + RA_{obs}^{\gamma}} \;.$$

The data for the American League teams of 2004 supports our assumptions on how runs are scored and allowed per game. Thus the assumptions of our model are met, and the Pythagorean Formula holds for some exponent γ. The actual value of the exponent γ is determined by the Weibull parameters. This means we find the exponent not by analyzing the outcomes of a team's games, but rather by analyzing the distribution of their runs (scored and allowed) per game. Using the 2004 data (obtained from *Baseball Almanac* [1]) and averaging over the 14 teams, we find using the method of maximum likelihood, the mean of γ is 1.74 with a standard deviation of .06, in terrific agreement with the numerical observation that $\gamma = 1.82$ is the best exponent. Using the method of least squares, our estimate is similar -- mean 1.79 with standard deviation .09.

Details of the calculations (as well as the programs used to read in the teams' data and perform the analysis) are available from the author, or see [6] for an expanded version of this paper with all the proofs and calculations.

2. Numerical Results: American League 2004

We analyzed the 14 teams[2] of the American League from the 2004 season. For each team we used two separate methods, the method of maximum likelihood and the method of least squares, to simultaneously find the best fit Weibulls of the form (α_{RS},-.5,γ) and (α_{RA},-.5,γ) (taking $\beta = -.5$ is a technical point, due to the fact that in a baseball game scores must be integers). What this means is we varied the parameters to find the Weibulls that are closest to the observed distributions of runs scored and allowed.

We then compared the predicted number of wins, losses, and won-loss percentage with the actual data. For brevity, we will show only the results from the method of maximum likelihood:

Team	Obs. Wins	Pred. Wins	Obs. W/L Percentage	Pred. W/L Percentage	Games Diff	γ
Boston Red Sox	98	93.0	.605	.574	5.03	1.82
New York Yankees	101	87.5	.623	.540	13.49	1.78
Baltimore Orioles	78	83.1	.481	.513	-5.08	1.66
Tampa Bay Devil Rays	70	69.6	.435	.432	0.38	1.83
Toronto Blue Jays	67	74.6	.416	.464	-7.65	1.97
Minnesota Twins	92	84.7	.568	.523	7.31	1.79
Chicago White Sox	83	85.3	.512	.527	-2.33	1.73
Cleveland Indians	80	80.0	.494	.494	0.00	1.79
Detroit Tigers	72	80.0	.444	.494	-8.02	1.78
Kansas City Royals	58	68.7	.358	.424	-10.65	1.76
Los Angeles Angels	92	87.5	.568	.540	4.53	1.71
Oakland Athletics	91	84.0	.562	.519	6.99	1.76
Texas Rangers	89	87.2	.549	.539	1.71	1.90
Seattle Mariners	63	70.7	.389	.436	-7.66	1.78

For the exponent γ, the mean from the 14 teams is 1.79 and the standard deviation is .089. Note how close this is to the numerically observed best exponent of 1.82. The mean absolute difference between observed and predicted wins here was 5.77; for the least squares method (not shown), the corresponding figure was 4.19. These results are consistent with the observation that the Pythagorean Formula is usually accurate to about four games in season.

We performed χ^2 tests to determine the goodness of the fit from the best fit Weibulls from the method of maximum likelihood. This test measures how accurately a Weibull with given parameters models a team's run production. There are 20 degrees of freedom for our tests, and the critical thresholds are 31.41 (at the 95% level) and 37.57 (at the 99% level). If the Weibull with these parameters is a good model for the team's performance, then 95% of the time we should obtain a value of 31.41 or less. Thus small values of the χ^2 statistic indicate a

[2] The teams are ordered by division (AL East, AL Central, AL West) and then by number of regular season wins, with the exception of the Boston Red Sox who, as the 2004 World Series champions, are listed first.

good fit. Conversely, if we were to obtain a large value of the χ^2 statistic, that would indicate a poor fit, which would mean our assumption that the run production is given by a Weibull is false.

We also did a χ^2 test to examine the independence of the runs scored and runs allowed per game. This is a crucial input for our model; unfortunately, the real world is a bit more complicated than our idealized model. In our model we consider runs scored and allowed as independent events drawn from continuous distribution. A nice consequence of this is that there is zero probability that the two numbers will be equal; however, in the real world, baseball scores must be integers. Further, runs scored and allowed per game can *never* be equal. This leads to slightly more complicated language.

What this means is that we have to use a more advanced test than a standard χ^2 test if we want to study whether or not runs scored and allowed per game are independent. Of course, in a game these two number *cannot* strictly be independent because a baseball game cannot end in a tie: if we know our team scores 5 runs in a game, we may not know how many runs they allow but we do know they did not allow exactly 5 runs! We shall discuss this issue further in the conclusion.

Statistical theory cannot allow us to conclude that runs scored and allowed per game are independent, because clearly they are not (as remarked above, if we know we score 5 runs then we know we do not allow 5 runs). What we *can* conclude is that, except for this forced condition, the runs scored and allowed per game behave as if they are independent. We are therefore investigating the following question: Given that runs scored and allowed cannot be equal, are the runs scored and allowed per game statistically independent events?

Instead of a standard χ^2 test, we actually have an incomplete two-dimensional contingency table; see [3] for the theory of such an analysis, and [4, 6] for more on the independence of runs scored and allowed in baseball. There are 109 degrees of freedom for our tests, and the corresponding critical thresholds are 134.4 (at the 95% level) and 146.3 (at the 99% level). Again, this means that our assumption about the independence of runs scored and allowed is validated any time we observe a value of 134.4 or less; if we observe a value significantly greater than 134.4, then the runs scored and allowed are probably not independent.

We summarize our results below; the first column of numbers is the χ^2 test for the goodness of fit from the best fit Weibulls, and the rightmost column is the χ^2 test for the independence of runs scored and runs allowed.

Team	RS + RA X^2: 20 d.f.	Independence X^2: 109 d.f.
Boston Red Sox	15.63	83.19
New York Yankees	12.60	129.13
Baltimore Orioles	29.11	116.88
Tampa Bay Devil Rays	13.67	111.08
Toronto Blue Jays	41.18	100.11
Minnesota Twins	17.46	97.93
Chicago White Sox	22.51	153.07
Cleveland Indians	17.88	107.14
Detroit Tigers	12.50	131.27
Kansas City Royals	28.18	111.45
Los Angeles Angels	23.19	125.13
Oakland Athletics	30.22	133.72
Texas Rangers	16.57	111.96
Seattle Mariners	21.57	141.00
95% confidence	31.41	134.4
99% confidence	37.57	146.3
95% Bonferroni	41.14	152.9
99% Bonferroni	46.38	162.2

Except for the Blue Jays, Mariners, and the White Sox, all test statistics are well below the 95% critical threshold. As we are performing multiple comparisons, chance fluctuations should make some differences appear significant when they are not[3]. For example, if we flip a fair coin 10 times then the probability that all 10 tosses are heads is quite small, namely $1 / 2^{10} = 1 / 1024$, or about .1%; however, if we consider 1024 sets of 10 tosses of a far coin, then we *do* expect to observe one set of 10 tosses as all heads (and if we consider 1,000,000 sets of 10 tosses, we would be astonished if there was no set of 10 heads).

[3] If the null hypothesis is true and 10 independent tests are performed, there is a 40% chance of observing at least one statistically significant difference at the 95% confidence level. More simply: *given enough events, unlikely things should happen!*

What this means is that the values of the χ^2 statistics are too small – the critical thresholds need to be increased. We must therefore adjust the confidence levels. We use the common, albeit conservative, Bonferroni adjustment method for multiple comparisons[4]. For example, for 20 degrees of freedom the critical threshold is now 41.14 and not 31.41, and for 109 degrees of freedom it is 142.9 instead of 134.4. The Blue Jays and White Sox just miss being significant at the 95% level.

The data validates our assumption that, given that runs scored and allowed cannot be equal, the runs scored and allowed per game are statistically independent events, and that the parameters from the method of maximum likelihood give good fits to the observed distribution of scores. In the appendix, we provide plots for two AL East teams.

Using our best fit parameters, we can estimate the mean number of runs scored and allowed per game. This provides an additional test to see how well our theory agrees with the data. Since there are so many games in a season, we may use a z-test to compare the observed versus predicted means. The critical z-values are 1.96 (at the 95% confidence level) and 2.575 (at the 99% confidence level); any number smaller than 1.96 is powerful support for the Pythagorean Formula.

Team	Obs RS	Pred RS	z-stat	Obs RA	Pred RA	z-stat
Boston Red Sox	5.86	5.80	0.24	4.74	4.83	-0.35
New York Yankees	5.54	5.47	0.24	4.99	4.95	0.12
Baltimore Orioles	5.20	5.26	-0.22	5.12	5.08	0.16
Tampa Bay Devil Rays	4.43	4.41	0.12	5.23	5.21	0.09
Toronto Blue Jays	4.47	4.51	-0.18	5.11	4.95	0.59
Minnesota Twins	4.61	4.74	0.32	4.41	4.48	-0.28
Chicago White Sox	5.34	5.40	-0.22	5.13	5.05	0.34
Cleveland Indians	5.30	5.18	0.40	5.29	5.25	0.09
Detroit Tigers	5.10	5.06	0.18	5.21	5.13	0.27
Kansas City Royals	4.44	4.48	-0.13	5.59	5.46	0.48
Los Angeles Angels	5.16	5.10	0.22	4.53	4.59	-0.22
Oakland Athletics	4.90	4.85	0.18	4.58	4.63	-0.19
Texas Rangers	5.31	5.29	0.05	4.84	4.82	0.08
Seattle Mariners	4.31	4.29	0.10	5.08	5.03	0.18
95% confidence			± 1.96			± 1.96
99% confidence			± 2.58			± 2.58

We note excellent agreement between all the predicted average runs scored per game and the observed average runs scored per game, as well as between all the predicted average runs allowed per game and the observed average runs allowed per game.

3. Conclusions and Future Work

Bill James' Pythagorean Won-Loss Formula may be derived from very simple and reasonable assumptions, and the parameters can easily be obtained by fitting to seasonal data. For the 2004 AL teams the fits were (basically) always significant, our assumptions were validated, and the best fit exponent γ was about 1.79 (or 1.74, depending on method), in excellent agreement with the observed value of 1.82. Importantly we find the exponent γ not by fitting the Pythagorean Formula to the observed won-loss percentages of teams, but rather from an analysis of the scores from individual games. Thus we now have a theoretical justification for using the Pythagorean Formula to predict team performances!

While our simple model is quite effective, it would be interesting to do a more micro analysis and incorporate additional effects, especially an inning-by-inning analysis. Ballpark effects can be added (see for example [8, 9]), as well as the effects of interleague play, or the effects of jumping out to a big lead (or deficit). The data seems to support the conclusion that a lot of these effects average out. For example, if a team has a big lead it might pull some of its stars to give them some rest; this will should decrease their run production for the rest of the game. Conversely, a team that is trailing by a lot late in the game is unlikely to use their ace reliever, instead viewing this as an opportunity to test some of their prospects. Thus there should be some relations between runs scored and allowed at various times in the game, and this would be a fascinating future project (one which the author is willing to pursue with an interested collaborator).

[4] The Bonferroni adjustment for multiple comparisons divides the significance level by the number of comparisons. For 20 degrees of freedom the adjusted critical thresholds are 41.14 (at the 95% level) and 46.38 (at the 99% level); for 109 degrees of freedom the adjusted critical thresholds are 152.9 (at the 95% level) and 162.2 (at the 99% level).

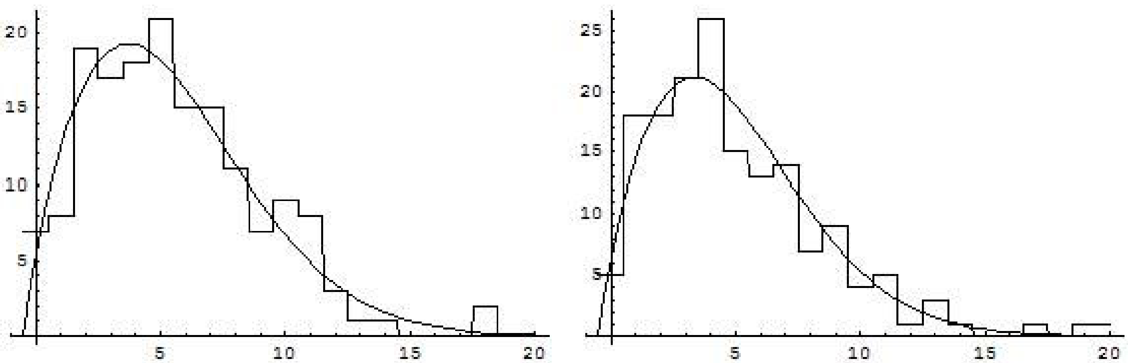
As baseball games cannot end in a tie, runs scored and allowed are never equal; however, they can be equal after 9 innings. One avenue for future research is to classify extra-inning games as ties (and record which team won), and adjust for games when the home team doesn't bat in the ninth.

Acknowledgements

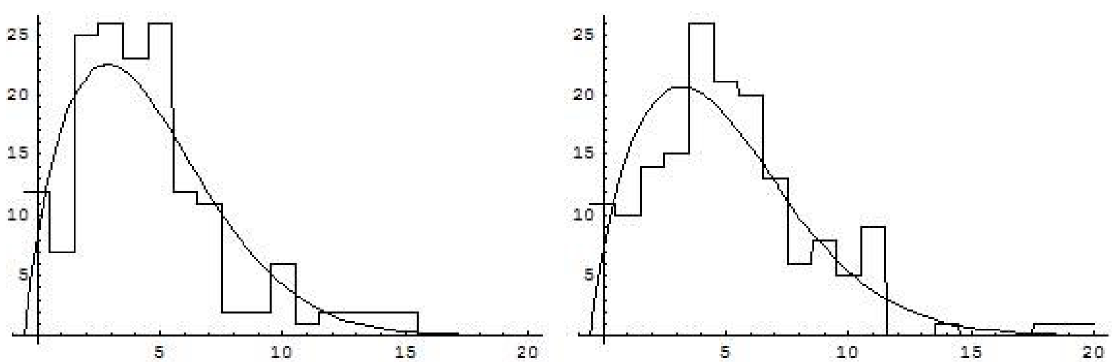
I would like to thank Russell Mann and Steven Johnson for introducing me to the Pythagorean Formula in baseball, Kevin Dayaratna for inputting much of the baseball data, Jeff Miller for writing a script to read in baseball data from the web to the analysis programs, Eric T. Bradlow for helpful comments on the expanded version [6], Phil Birnbaum for comments on an earlier draft, and Ray Ciccolella for discussions on the independence of runs scored and allowed.[5]

Appendix – Plots of Best Fit Weibulls

Below we plot the best fit Weibulls against the observed histograms of runs scored and allowed for two of the teams of the AL East in 2004; the other plots are similar and are available at [6] or from the author. We chose these teams as they had the combined worst and best fits, and are thus representative of the quality of how well the Weibulls fit the data.



Plots of Runs Scored (pred vs obs) & Runs Allowed (pred vs obs) for New York Yankees (Best fit of AL East)



Plots of Runs Scored (pred vs obs) & Runs Allowed (pred vs obs) for Toronto Blue Jays (Worst fit of AL East)

[5] See Ciccolella's article in this issue. –Ed.

Even a quick, visual inspection shows how good of a fit we have. The Weibulls capture the key features of runs scored and allowed per game: they start with low probability for games where the runs are near zero, increase to a maximum and then quickly decrease to zero probability for games with a large number of runs.

Bibliography

[1] Baseball Almanac, http://baseball-almanac.com.
[2] J. Angus, http://forums.mlbcenter.com/showthread.php?t=14121.
[3] Y. M. M. Bishop and S. E. Fienberg, *Incomplete Two-Dimensional Contingency Tables*, Biometrics 25 (1969), no. 1, 119-128.
[4] Ray Ciccolella, *Are Runs Scored and Runs Allowed Independent?*, preprint.
[5] Bill James, *The Bill James Baseball Abstract 1983*, Ballantine.
[6] S. J. Miller, *A Derivation of the Pythagorean Won-Loss Formula in Baseball*, preprint, http://arxiv.org/abs/math.ST/0509698.
[7] L. Dean Oliver, http://www.rawbw.com/~deano/helpscrn/pyth.html.
[8] M. J. Schell, *Baseball's All-Time Best Hitters*, Princeton University Press, Princeton, NJ, 1999.
[9] M. J. Schell, *Baseball's All-Time Best Sluggers*, Princeton University Press, Princeton, NJ, 2005.

Steven J. Miller, Department of Mathematics, Brown University, 151 Thayer Street, Providence, RI, 02912, sjmiller@math.brown.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.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Receive BTN by Internet Subscription

You can help save SABR some money, and me some time, by downloading your copy of *By the Numbers* from the web. BTN is posted to http://www.philbirnbaum.com in .PDF format, which will print to look exactly like the hard copy issue.

To read the .PDF document, you will need a copy of Adobe Acrobat Reader, which can be downloaded from www.adobe.com.

To get on the electronic subscription list, visit http://members.sabr.org, go to "My SABR," and join the Statistical Analysis Committee. You will then be notified via e-mail when the new issue is available for download.

If you don't have internet access, don't worry – you will always be entitled to receive BTN by mail, as usual.

A Breakdown of a Batter's Plate Appearance – Four Hitting Rates

Jim Albert

A player's batting line results from of his skill at the plate, combined with a hefty helping of good or bad luck. In this study, the author starts by assuming that batting talent is actually a linear combination of four different basic skills. Then, based on those skills, he derives a statistical method to separate the player's performance into a skill component and a luck component.

1. Introduction

There has much effort in the sabermetrics literature in measuring the performance of a batter. The goal of any hitter is to help create runs for his team and so any good batting measure should be highly correlated with the number of runs scored. It has been shown that a batting average has a relatively weak relationship with runs scored, and that there exist much better measures such as runs created and OPS that are highly correlated with runs.

Albert (2005) recently criticized the batting average from a different perspective. Any player's hitting statistic such as AVG = H/AB, a strikeout rate SORATE = SO/AB, or a walk rate BBRATE = BB/PA is an estimate at the player's probability of that particular play. For example, BBRATE is an estimate at the player's chance of getting a walk in a plate appearance. If one observes hitting rates for many players, one will see much variation between the rates. There are two explanations for this variation. First, players have different probabilities or talents to perform the batting event; for example, players have different abilities to draw a walk. Second, part of the variation in the hitting rates is due to inherent chance or luck variation. Albert (2005) showed that hitting statistics differed with respect to the amount of luck relative to the amount of talent. The strikeout rate is an example of a talent measure. Most of the variation in strikeout rates for players in a single season is due to differences in the players' strikeout probabilities. In contrast, a batting average is an example of a "lucky" statistic. Much of the variation we see in players' batting averages for a season is to due to chance and actually players have similar probabilities of getting a hit.

In this paper we present a way of subdividing the plate appearances of a batter by removing in turn walks, strikeouts, home runs, and "in park" hits. This subdivision leads to the definition of four batting rates. Each batting rate measures a particular skill of a hitter such as the skill in getting a walk, the skill in not striking out, the skill of hitting a home run, and the skill in getting ball in-play to fall in for a hit. These four rates can be combined linearly to obtain a simple measure of the batter's performance. We show in Section 2 this combined measure compares favorably with other good measures such as runs created and OPS in predicting runs scored. The talents of a player to draw a walk, to not strikeout, to hit a home run, and to get a batted ball to land for a hit can be measured by the respective probabilities of these events. By using data from the recent 2005 season, we estimate in Section 3 the collection of probabilities for all nonpitchers for each batting skill. This study is helpful for understanding that some batting rates are reflective of the talents of the hitters and other rates are due more to chance variation. We compute the estimated probabilities for all regular players for the 2005 season and give tables showing the best and worst hitters with respect to each batting skill.

2. A decomposition of a plate appearance

A player comes to bat for a plate appearance. Either the player walks or doesn't walk – his chance of walking is estimated by the walk rate (BB+HBP)/PA. (Note that we combine walks and hit by pitches in the formula since each event has the same result of getting the batter to first base without creating an at-bat.) Removing walks from the plate appearances, we next record if the batter strikes out or not. We define the strikeout rate as the fraction of strikeouts to the number of at-bats or SO/AB. With walks and strikeouts removed, we next record if the batter hits a home run or not. The home run rate is defined to be the fraction of home runs for all plate appearances where contact is made by the bat. That is, HR rate = HR/(AB – SO). With the walks, strikeouts, and homeruns removed, we only have plate appearances where the ball is hit in the park. Of these balls put in-play, we record the fraction that fall in for hits – we call this the in-play hit rate or "hit rate" (H-HR)/(AB-SO-HR). This hit rate has been described as the *Ball-In-Play Average* by Woolner (2001) and the *Batting Average On Balls In Play* (BABIP) by Silver (2004). Much of the interest in this statistic is due to McCracken (2001)'s study that indicated that pitchers have little control over the outcomes of balls put into play that are not home runs.

Since these rates are defined by sequentially removing walks, strikeouts, and home runs from the plate appearances, they measure distinct qualities of a hitter. Specifically, these rates measure (1) the talent to draw a walk, (2) the talent to avoid a strikeout, (3) the talent to hit a ball out of the park (a home run), and (4) the talent to hit a ball "where they ain't". In contrast, traditional hitting statistics confound some of these talents. A batting average confounds three batter talents: the talent not to strikeout, the talent to hit a home run, and the talent to hit an in-play ball for a hit. An on-base percentage confounds the batter's talent to draw a walk with his talent to get an in-play hit and his talent to hit a home run. Since a batter's ability encompasses all of these talents, these four rates may provide useful detailed information about the hitting ability of a player.

3. Team data – predicting runs scored

Since a team, not an individual, produces runs, one can see the usefulness of these rate statistics in predicting runs by looking at team data. We collect hitting statistics for all 30 teams in the 2005 season and compute the four rate statistics for each team. We run a stepwise regression where we use these four rates to predict runs scored per game. We first add the home run rate variable with a $R^2 = .45$ – this variable explains 45% of the variation in the runs scored per game. Then we add the hit rate with a total R^2 of .54, the strikeout rate with a total R^2 of .73, and finally the walk rate with a total R^2 of .83. It is interesting that each rate explains a significant portion of the variation in the runs data, even when the other variables are included in the model. The final model is

```
Runs per game =      -  3.2
                     + 13.2 (walk rate)
                     - 12.3 (strikeout rate)
                     + 40.9 (HR rate)
                     + 24.5 (hit rate)
```

How does this model compare with other "bad" and "good" predictors of runs scored? If one uses batting average to predict runs scored, then the R^2 value is .50, and if we use SLG, the R^2 value is .63. These values are not surprising, since AVG and SLG are relatively poor predictors of runs scored. But if we use the better statistics OPS = OBP + SLG, runs created, and LSLR to predict runs scored, the R^2 values are respectively .77, .82, and .86. (LSLR, or least-squares linear regression, is the result of computing the best linear combination of singles, doubles, triples, home runs, and walks in predicting runs.) So our new measure has a similar correlation with runs scored as the good measures OPS, runs created, and LSLR.

Table 1 – Average values of R^2 for using different statistics to predict team runs scored per game for seasons 1950 through 2005

	Runs Created	OPS	LSLR	4 Rates	5 Rates
R^2	87.7	87.9	90.5	89.6	90.9

To see if this is generally true across seasons, we looked at team data for the seasons from 1950 through 2005 and performed this regression study for each season. Table 1 shows the average (mean) R^2 value for runs created, OPS, and LSLR for these 56 seasons. Also we tried the "five rates", consisting of walk rate, strikeout rate, home run rate, singles rate, and doubles + triples rate to predict runs per game. Note that generally our four rates are superior to both runs created and OPS in predicting runs scored. LSLR generally does better than four rates, but the average improvement in R^2 is only about 1%. It is surprising that "five rates" that distinguishes singles from doubles/triples does only marginally better than "four rates" – the average improvement in R^2 is only 1.3%. This indicates that there is little added value in distinguishing singles from extra-base hits once walks, strikeouts, home runs, and in-play hits are recorded.

4. Estimating probabilities of each type of batting skills for 2005 players

Each player has four rate statistics. There statistics are estimates of the following probabilities:

p_{walk} = probability of a walk in a plate appearance
p_{so} = probability of a strikeout in an at-bat (walks excluded)
p_{hr} = probability that a batted ball is hit for a home run
p_{hit} = probability that ball put in-play falls in for a hit

Each player has four batting probabilities. For each type of skill (walking, not striking out, hitting a home run, and having a batted ball fall in for a hit), we estimate the batting probabilities for all nonpitchers using data from the 2005 season.

We use the random effects model used earlier in Albert (2005) and Albert (2006). For a particular skill, say drawing a walk, let $p_1, \ldots, p_N$ denote the probabilities of this skill for the *N* players. We let the probabilities { p_i } come from a "talent distribution" of the functional form

$$g(p) \propto p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1.$$

Given the talent p_i for the *i*th hitter, the observed number of walks y_i in n_i opportunities is assumed to follow a binomial (coin-tossing) distribution with probability of success p_i. Using the data for all players' data for the 2005 season, we fit this model by finding the values of the probabilities $\{p_i\}$ and the talent numbers *a* and *b* that make the observed data most likely. There are two aspects of this model fitting that are relevant here. First, the estimated values of the numbers *a* and *b* are informative – they tell us something about the location and spread of the talents of the players. Second, we are able to use the estimated values of *a* and *b* to estimate the talents { p_i } for all players.

Table 2 gives the estimated values of the talent distribution for each of the four skills. The talent dimension of the particular statistic can be measured by the sum of the two estimated parameters *a* +*b*. If the estimated *a* + *b* is small, then it indicates that the observed rates for that skill are reflective of the different abilities of the players. On the other hand, if the estimated *a* + *b* is large, the particular "rate" is heavily determined by chance. In this case, much of the variation in the rate statistics for players is determined by luck rather than differences in the talents of the players. Since the estimated *a* + *b* for walking, striking out, and hitting a home run are small, these rates are driven more by the talents of the players. On the talent/chance scale, the strikeout rates are most reflective of the abilities of the players. At the other extreme, the in-play hit rates have a large estimated value of *a* + *b* – these rates are driven more by chance variation.

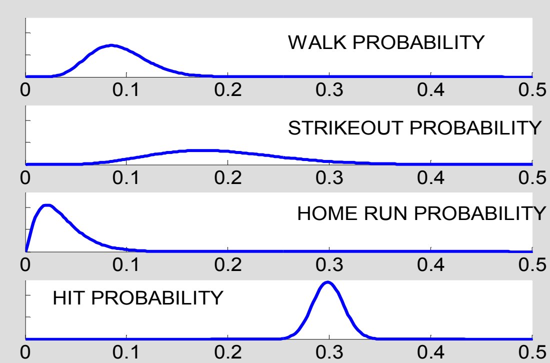
Table 2 – Estimated values of the talent distribution and the average probability for the four rate statistics

Skill	a	b	Mean Talent
Walking	9.5	92.9	.093
Striking out	7.4	31.0	.193
Hitting a home run	2.4	64.5	.036
Getting an in-play hit	268.1	628.4	.299

The above conclusions are displayed graphically in Figure 1. This figure shows the estimated talent distribution for each of the four rates. Note that the strikeout probability distribution is very wide, indicating that players have variable talents to strikeout. The in-play hitting probabilities have a very narrow talent distribution – these probabilities fall in a narrow interval about the average value of .3.

Is this a good model for predicting the observed 2005 rates? We answer this question by seeing if predicted data simulated from the model resemble the actual 2005 data. We illustrate this method for the walk rates. We first simulate a set of walk probabilities from the fitted beta talent distribution with *a* = 9.5 and *b* = 92.9. Then we simulate walk data for the players using these simulated probabilities. Figure 2 displays histograms of the simulated walk rates for "regular" players with at least 300 plate appearances for eight of these simulations. The actual 2005 walk rates for these regular players are graphed as a histogram with solid bars. Note that the simulated rates from the fitted model generally resemble the actual rates in general shape, average, and spread. Displays such as this one were used to confirm that these random effects models were suitable fits for all of the rate data.

Figure 1 – Estimated distribution of talents of all 2005 non-pitchers corresponding to the four hitting rates



The above estimated talent distributions give us some understanding about the variation in talent for a particular batting skill. We can also use the estimated values of *a* and *b* to obtain estimates for the batting probabilities for any player. We illustrate the computation of these estimates for Andruw Jones.

Suppose we are interested estimating the home run probability for Jones in the 2005 season. His home run rate for this season was 51/(586 – 112) = .108. But there is a general phenomenon called the regression effect that says that this rate, since it is extreme at the high end, overestimates Jones' probability of hitting a home run. It is desirable to adjust his home run rates downwards towards the average home run rate for all players. For our random effects model, it can be shown that an estimate at a player's home run probability is given by

$$\frac{HR+\hat{a}}{AB-SO+\hat{a}+\hat{b}}$$

where $\hat{a}$ and $\hat{b}$ are the estimated values of a and b from the fitted talent distribution. In this case, since $\hat{a}$ = 2.4 and $\hat{b}$ =64.5, the estimate at Jones' home run probability is

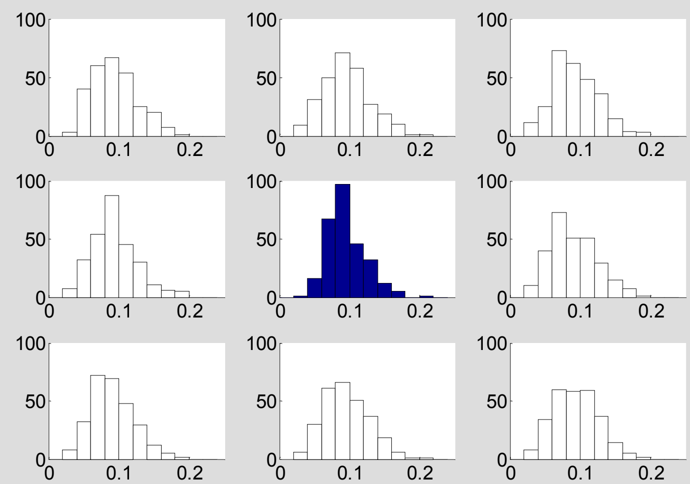
$$\frac{51+2.4}{586-112+2.4+64.5}=.099\,.$$

Table 2 – Correlation coefficients for estimated probabilities for regular players in 2005 season

	Walk Rate	SO Rate	HR Rate	Hit Rate
Walk Rate		0.37	0.47	0.16
SO Rate	0.37		0.55	0.13
HR Rate	0.47	0.55		-0.03
Hit Rate	0.16	0.13	-0.03	

We adjust Jones' actual home run rate of .108 slightly downward to .099. We only make a slight adjustment since we know that home run rate is a talent measure and not heavily affected by chance variation.

Figure 2 – Simulated walk rates from the random effects model for players with at least 300 plate appearances



Histograms with clear bars display rates simulated from the fitted model; histogram with solid bars displays the actual 2005 walk rates

As a second example, let's estimate Jones' in-play hit probability. His observed hit rate in 2005 was (154 – 51)/(586 – 112 – 51) = .243. But we know that hit rates are driven by chance variation and we wish to make a larger adjustment to estimate his probability. For the in-play hit talent, the estimated values of the talent distribution are given by $\hat{a}$ = 268.1 and $\hat{b}$ = 628.4, and so the estimate of Jones' probability is

$$\frac{154-51+268.1}{586-112-51+268.1+628.4}=.281\,.$$

We see that we are making a large adjustment to Jones' hit rate in this case. This says that we believe that Jones' hit probability is more likely to be much larger than his observed rate of .243.

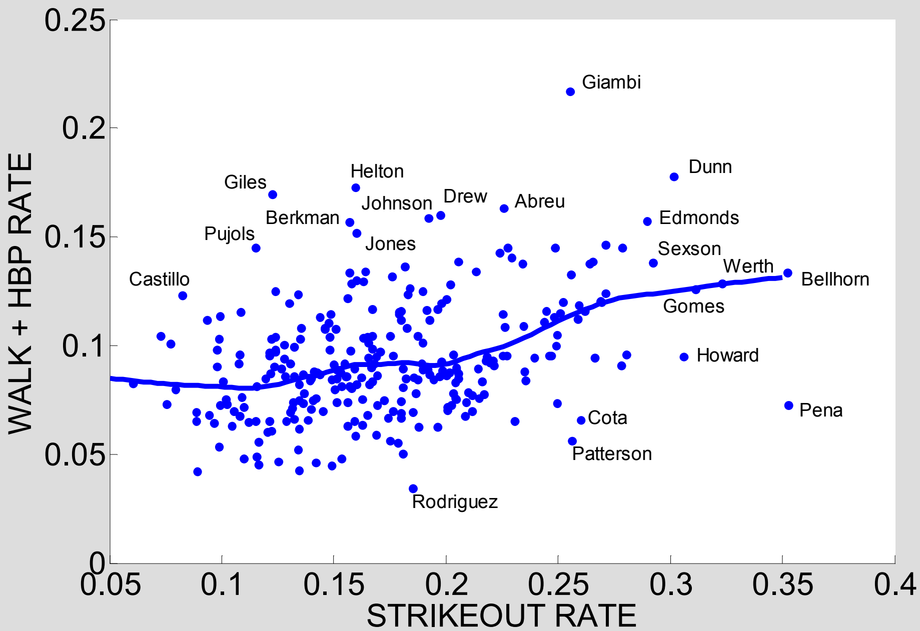
Using these formulas, we estimate the walk, strikeout, home run, and hit probabilities from the hitting data for all non-pitchers in the 2005 season. To help understand the relationships between these rates, Table 3 presents correlation coefficients between all pairs of rates for the players with at 300 at-bats (the regulars). We see that walk rate, strikeout rate, and home run rate are all positively correlated – this means, for example, that players that tend to walk frequently also tend to hit home runs at a high rates. In contrast, hit rate is at most weakly correlated with the other three rates. This means that the talent to place a batted ball for a hit is weakly related with the talent to walk, to avoid a strike out, or to hit a home run.

To take a closer look, Figures 3, 4, and 5 display scatterplots of the estimated talent probabilities for the regular players. Figure 3 plots the estimated strikeout probabilities against the estimated walk probabilities and Figures 4 and 5 plot the estimated strikeout probabilities against the estimated home run and hit probabilities, respectively. A large number of extreme points are labeled with the players' names. For each graph, a smoothing curve is placed on top that shows the general relationship between the corresponding estimated probabilities.

In Figure 3, we see a slight positive relationship between a player's estimated strikeout probability and his estimated walk probability. Jason Giambi, escaping the shadow of Barry Bonds, stood out in 2005 for his large walk rate. There are a number of other players such as Brian Giles, Todd Helton, J. D. Drew, Bobby Abreu and Adam Dunn that were pretty good in 2005 in drawing walks. At the other extreme, Ivan Rodriguez had an unusually small walk rate in 2005.

Figure 4 shows a strong relationship between a player's strikeout tendency and his home run rate. A number of hitters who had large estimated home run probabilities are labeled. There are several players that deviated strongly from the high strikeout rate/high home run pattern. Mark Bellhorn and Jayson Werth had very high strikeout rates but few home runs to show for it. Albert Pujols, Vladimir Guerrero and Aramis Ramirez (not Manny) displayed high home run rates with small strikeout rates.

Figure 3 – Scatterplot of estimated strikeout probabilities and estimated walk probabilities for all 2005 players with at least 300 plate appearances



I think the extreme players labeled in Figures 3 and 4 will not surprise most readers. Strikeout rates, home run rates, and hit rates are all talent statistics and so players that have high values of these rates for one year will tend to have high values of the rates for another year. In contrast, Figure 5 displays a scatterplot of the estimated strikeout probabilities against the estimated hit probabilities. Since hit rates are more influenced by luck, the estimated hit probabilities are in a small range from about .28 to .32. There are some interesting extreme probability estimates that are labeled in the figure. Kenny Lofton, Johnny Damon, and Derek Jeter have large values and Mike Lowell, Andruw Jones, and Steve Finley have small values. Do these players have unusual skill (or lack of skill) in getting balls to fall in for hits? Actually, no since we know that hit-rates are heavily influenced by luck or chance variation.

Figure 4 – Scatterplot of estimated strikeout probabilities and estimated home run probabilities for all 2005 players with at least 300 plate appearances

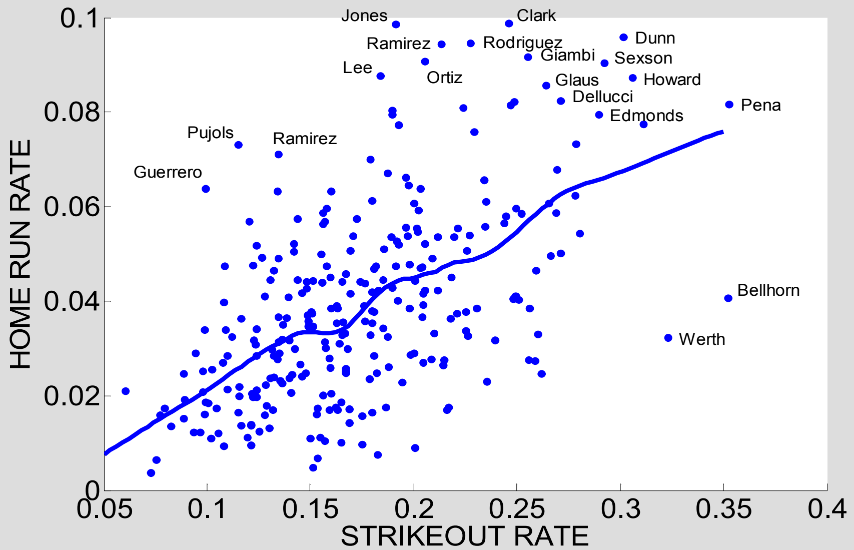
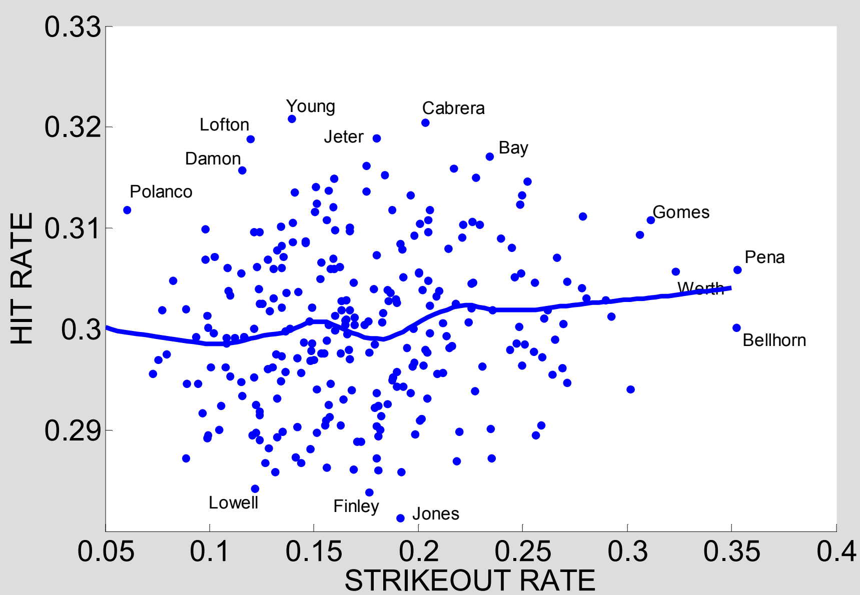


Figure 5 – Scatterplot of estimated strikeout probabilities and estimated hit probabilities for all 2005 players with at least 300 plate appearances



We conclude by presenting in Table 4a to 4d the best and worst 2005 hitters with respect to the probability of each of the four talents. These tables reinforce some the comments made earlier. The strikeout, walking, and home run rankings are meaningful since these are talent measures. In contrast, the rankings for the hit rate are less meaningful due to the large role of chance. Due to the small differences in the estimated hit rates, we suspect that the rankings for hit rate would have a dramatic change for the 2006 season.

Table 4a: Highest and Lowest Probability of Striking Out

.353	Wily Mo Pena	.060	Placido Polanco
.352	Mark Bellhorn	.073	Jason Kendall
.323	Jayson Werth	.076	Juan Pierre
.311	Jonny Gomes	.077	David Eckstein
.306	Ryan Howard	.080	Paul Lo Duca
.301	Adam Dunn	.083	Luis Castillo
.292	Richie Sexson	.088	Freddy Sanchez
.289	Jim Edmonds	.088	Yadier Molina
.280	Victor Diaz	.089	Neifi Perez
.279	Pat Burrell	.094	Mark Loretta

Table 4b: Highest and Lowest Probability of Walking

.217	Jason Giambi	.034	Ivan Rodriguez
.178	Adam Dunn	.042	Neifi Perez
.173	Todd Helton	.043	Robinson Cano
.170	Brian Giles	.045	Garret Anderson
.163	Bobby Abreu	.045	Jose Reyes
.160	J.D. Drew	.046	Jorge Cantu
.159	Nick Johnson	.047	Aaron Miles
.157	Jim Edmonds	.048	Alex Cintron
.157	Lance Berkman	.048	Tony Womack
.152	Chipper Jones	.049	B.J. Surhoff

Table 4c: Highest and Lowest Probability of Hitting a Home Run

.099	Andruw Jones	.004	Jason Kendall
.099	Tony Clark	.005	Scott Podsednik
.096	Adam Dunn	.007	Juan Pierre
.095	Alex Rodriguez	.007	Tony Womack
.094	Manny Ramirez	.008	Jamey Carroll
.092	Jason Giambi	.009	Royce Clayton
.091	David Ortiz	.009	Omar Vizquel
.090	Richie Sexson	.010	Cesar Izturis
.088	Derrek Lee	.010	Willy Taveras
.087	Ryan Howard	.010	Nook Logan

Table 4d: Highest and Lowest Probability of Getting a Hit

.321	Michael Young	.281	Andruw Jones
.321	Miguel Cabrera	.284	Steve Finley
.319	Derek Jeter	.284	Mike Lowell
.319	Kenny Lofton	.286	Jason Phillips
.317	Jason Bay	.286	Justin Morneau
.316	Willy Taveras	.286	Omar Infante
.316	Cory Sullivan	.286	Cristian Guzman
.316	Johnny Damon	.286	Joe Crede
.315	Derrek Lee	.287	Carlos Lee
.315	Alex Rodriguez	.287	Kevin Mench

5. Concluding comments

When one evaluates a batter, one typically thinks of two qualities – the ability of a batter to get on-base and the ability to advance runners towards home. But these two abilities are confounded – for example, a hit or a walk will accomplish both purposes – and so it may be desirable to think of a different set of talents of a hitter that measure distinct entities. Here we have suggested a different breakdown by removing in turn walks, strikeouts, and home runs from the plate appearances. This breakdown motivates the consideration of four probabilities that seem to represent distinct qualities of a batter. All of these qualities are important in predicting runs scored for a team. But some rates are more driven by the talent of the players and other rates are more a byproduct of chance variation. When evaluating a hitter, a scout should focus on ability-measures such as a player's strikeout rate, his walk rate, and his home run rate, and place little confidence in a player's hit rate. One surprising finding from this work is that doubles and triples don't appear to explain much of the variation in runs scored after walks, strikeouts, home runs, and hits are given.

References

- Albert, J. (2005), "Does a Baseball Hitter's Betting Average Measure Ability or Luck?" *Stats*, 44.
- Albert, J. (2006), "Pitching Statistics, Talent and Luck, and the Best Strikeout Seasons of All-Time", Journal of Quantitative Analysis of Sports, Volume 2, http://www.bepress.com/jqas/vol2/iss1/2/
- Albert, J. and Bennett, J. (2003), *Curve Ball: Baseball, Statistics and the Role of Chance in the Game*, Springer-Verlag, revised edition. ISBN 0-387-00193X

- McCracken, V. (2001), "Pitching and defense: how much control do hurlers have?", *Baseball Prospectus*, http://www.baseballprospectus.com.
- Silver, N. (2004), "Lies, Damned Lies: The Unique Ichiro", *Baseball Prospectus*, http://www.baseballprospectus.com.
- Woolner, K. (2001), "Counterpoint: Pitching and Defense", *Baseball Prospectus*, http://www.baseballprospectus.com.

Jim Albert, Department of Mathematics and Statistics, Bowling Green State University, albert@bgnet.bgsu.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
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
Mark E. Johnson	maejohns@yahoo.com	General
David Kaplan	dkaplan@UDel.Edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics