By the Numbers

Volume 13, Number 4 The Newsletter of the SABR Statistical Analysis Committee November, 2003

Summary

Academic Research: Competitive Balance and Coors Air

Charlie Pavitt

The author reviews two recent academic papers – one on the reasons for baseball's recent steady increase in competitive balance, and another on the causes of offense increases in Coors Field.

This is one of a series of reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at its new location www.udel.edu/communication/pavitt/biblioexplan.htm . Use it for your research, and let me know what is missing.

Martin B. Schmidt and David J. Berri, On the Evolution of Competitive Balance: The Impact of an Increasing Global Search, Economic Inquiry, October 2003, Volume 41 Number 4, pp. 692-704

This is another in a long line of studies on competitive balance in professional sports. Research has shown unequivocally that competitive balance in major league baseball improved markedly over the course of the twentieth century. The issue is why. Schmidt and Berri discuss it in terms of the Coase theorem, a proposal to the effect that purposive manipulations to a market have no impact, because the market will work identically either way. Several economists have applied the idea to baseball, with the implication that changes in the manner by which players are distributed among teams that might impact on competitive balance, such as free agency and the amateur draft, have no long-term impact on the team on which a player ends up. This is because under any circumstance a player will end up with the team that values his services the most, and which therefore is the most willing to pay the player to sign with them, or compensate the team owning that player in a trade.

In this issue

One implication of the Coase theorem when directed to baseball is that competitive balance should not change over time as a consequence of changes in player distribution rules. Yet we know that competitive balance has improved. So finding that competitive balance is due to a factor other than these changes would provide indirect support for the Coase theorem. For that, Schmidt and Berri turn to the hypothesis first suggested by Stephen Jay Gould that the decreased variation in performance among major league players across the twentieth century (e.g., the disappearance of the .400 and sub-.200 hitter) is due to the improvement in the skills of the average player relative to the best. This is due in turn to the increase in the population from which major league players were drawn, starting with whites from the Northeast U.S. and then including in turn whites from across the U.S., blacks from across the U.S., Hispanics from more and more countries over time, Asians, and anywhere else on the globe. As a consequence, a certain skill level that was adequate in the 1920s would no longer be sufficient in the 1940s, adequate skill levels in the 1940s would be obsolete in the 1960s, and so on.

Although the original hypothesis was geared in terms of the increase in the skill level of the average player, due to displacement of the relatively poor, this also applies to the average team, because weaker teams have the weakest players as a whole and so would gain the most advantage from an increase in average skill level. Schmidt and Berri's analysis of MLB data show that the 20th century increase in competitive balance indeed did mirror increases in the proportion of black and Hispanic players, with this effect drowning out the impact of changes in the rules for player distribution.

So the Coase theorem appears to be supported. But we know that the Coase theorem cannot account for free agent distribution in baseball. As I wrote above, the Coase theorem implies that players will end up with the team that is willing to pay the most for them, either directly or indirectly through purchasing their contract from another team. Yet free agents often sign with teams other than the one that offers them the most money, because they want to live in a given location or they think they will get more playing time or they think the home ballpark suits them or any number of other reasons. There are arguments for and against changes in the current rules for player distribution, but the Coase theorem is not one of them.

Frederick Chambers, Brian Page, and Clyde Zaidins, Atmosphere, Weather, and Baseball: How Much Farther Do Baseballs Really Fly at Denver's Coors Field?, Professional Geographer, 2003, Volume 55 Number 4, pp. 491-504

The impact of Coors Field on run production is well known. Robert Adair (The Physics of Baseball) and others have assigned at least part of the credit for this impact on Denver's high altitude, estimating that its weaker gravity and thinner air allows fly balls to travel about ten percent farther. However, the authors, all faculty at the University of Colorado's Denver campus, found that mean fly ball distance from 1995 through 1998 (as reported by STATS, Inc.) was only six percent farther at Coors Field (302.8 feet) than the average for other National League parks (284.5 feet). Why the discrepancy?

During the 1997 season, the authors set up two meteorological stations inside Coors Field, with the Rockies' knowledge and permission, and made measurements every 15 minutes during games. The authors found fly ball distance to have no relationship with either humidity or temperature, but a substantial relationship with wind direction; a correlation of -.45 with east winds (which basically come in from right field toward home plate) and +.49 with west winds (which generally blow out toward right field from home plate). Further, Denver has a tendency for east winds to dominate during the afternoon and evening when games are in progress. As a consequence, the wind usually depresses fly ball distances during the game. Of course, when the wind does blow out, the stage is set for some spectacular offense.

Continuing the analysis, the authors computed the average fly ball distance as a percentage of average outfield dimensions for all NL parks, and found that, given its relatively big dimensions, the functional advantage for Coors Field during those four seasons was only three percent overall. This was no greater than for Philadelphia, Los Angeles, and Atlanta, and was less than for St. Louis. Thus Mark McGwire's home run glory years were achieved while "enjoying the advantages of a ballpark that is every bit as conducive to home-run production as Coors Field in terms of how far the average fly ball carries relative to the average position of the outfield fence" (page 503). But the impact of Coors Field on home runs remains valid; .044 home runs per at bat between 1995 and 2002 versus an average of .029 for other NL parks. The authors attribute this impact to other factors, most notably the impact both of thin air on pitch movement (breaking pitches don't break and all pitches are harder to control) and of low humidity on the ball itself (making the ball not only lighter, but drier and slicker, making the ball harder for pitchers to grip).

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

Leadoff Men in "The New Bill James Historical Baseball Abstract"

Herm Krabbenhoft

Bill James' recent Historical Baseball Abstract contained a discussion of the best leadoff men of all time. But, as the author points out, many leadoff men were omitted from Bill James' rankings – and many of the players ranked were not leadoff men at all.

In "The New Bill James Historical Baseball Abstract," James has two separate discussions on the topic of leadoff men.[1,2] My commentary is being made because of serious historical errors in those discussions.

In the first discussion, James focuses on shortstops as leadoff men:[1]

> *Reese was…how do I say this…the best career leadoff man among the shortstops. Of the top 100 shortstops, almost exactly one-fourth were essentially leadoff men.*
>
> *Of the leadoff men, there were three who were probably better leadoff men than Reese, at least in theory. The most effective leadoff man in the group, actually, was Solly Hemus. Hemus, however, was not really a shortstop, and thus was always fighting to stay in the lineup, even after he led the National League in runs scored in 1952.*
>
> *Johnny Pesky was a highly effective leadoff man, more effective than Reese, but lost his best years to World War II, and had a short career. Lyn Lary was a terrific leadoff man, led the American League in stolen bases in 1936 and was a high percentage base stealer, also drew 117 walks in 1936, but he was also in and out of the lineup due to injuries and marginal defense. Ray Chapman was a quality leadoff man, but…*
>
> *Among the shortstops who were leadoff men and who had long careers – Bartell, Crosetti, Bancroft, Rizzuto, Donie Bush, Maury Wills, Campaneris, Aparicio, Patek – Reese was the most effective leadoff man.*

I don't know how or where Mr. James obtained his information to write the above statements. His statements are not consistent with the actual baseball record. Here are the facts.

Of the 100 shortstops that the author lists (ranked by Win Shares), 10 played exclusively or predominantly in the 19th century. Of the 90 20th century shortstops, only 15 were principal leadoff batters for five or more seasons, where a "principal leadoff batter" is defined as the player who leads off the most games for a team in a given season.[2]

That is, only 17% of these shortstops can realistically be considered "essentially leadoff men."

With regard to the statements about Reese, Hemus, Pesky, Lary, and Chapman, here are the facts about their leadoff batter activity:

Reese was a principal leadoff batter for 5 seasons (1940-1942, 1949-1950). In 1940 he was a leadoff batter in 61 of the 84 games he played (73%). In 1941, he was a leadoff batter in 96 of the 152 games he played (64%). In 1942, he was a leadoff batter in 76 out of the 151 games he played (50%). In 1949, he was a leadoff batter in 147 out the 155 games he played (95%). And, in 1950, he was a leadoff batter in 85 out of the 141 games he played (60%). Therefore, for his 5 principal leadoff batter seasons, he was a leadoff batter in 465 games out of the 683 games he played (68%). For his entire career (1940-1942, 1946-1958), he batted leadoff in 538 games.

Hemus was a principal leadoff batter for just three seasons (1951-1953). In 1951, he was a leadoff batter in 69 of the 120 games he played (58%). In 1952, he was a leadoff batter in 132 of the 151 games he played (87%). And, in 1953 he was a leadoff batter in 111 of the 154 games he played (72%). Thus, for these three seasons he was a leadoff batter in 312 of the 425 games he played (73%). For his entire career (1949-1959), he was utilized as a leadoff batter in 372 games (i.e., just 60 games beyond the 312 during the 1951-1953 period).

---

[1] Bill James, "*The New Bill James Historical Baseball Abstract*," pp.598, 649-651 (Simon & Schuster, 2003).
[2] The leadoff statistics presented in this essay are from the author's work, "*The Encyclopedia of 20th Century Leadoff Batters*" (2004).

Pesky was *never* a principal leadoff batter. For his entire career (1942, 1946-1954), he was utilized as a leadoff batter in just 25 games – 10 in 1942, three in 1947, one in 1951, nine in 1952 (four with the Red Sox and five with the Tigers), and five in 1953. It is incomprehensible how Pesky could be classified as "a highly effective leadoff man" when he was hardly ever a leadoff man.

Lary was a principal leadoff batter for four seasons (1935-1938). In 1935, he was a leadoff batter in 93 of the 142 games he played (65%). In 1936 and 1937, he was a leadoff batter in 100% of the games he played (155 and 156 games, respectively). And in 1938, he was a leadoff batter in 138 of the 141 games he played (98%). Thus, during his four principal leadoff batter campaigns he was a leadoff batter for a total of 542 games (91%). For his whole career (1929-1940), he was a leadoff batter in 574 games (i.e., only 32 more than he accumulated during the 1935-1938 period).

Chapman was *never* a principal leadoff batter. For his entire career (1912-1920), he was utilized as a leadoff batter in only 44 games – 23 in 1915, three in 1916, and 18 in 1918. It doesn't make sense to me how Chapman can be called "a quality leadoff man" when he was hardly ever used as a leadoff batter.

Next, considering "the shortstops who were leadoff men and who had long careers," the following chart summarizes the principal leadoff batter careers of these players (listed in alphabetical order).

	Principal Leadoff Batter Seasons				PLOB Composite Results			PLOB Games		LOG
Shortstop	first	last	total	≥ 75%	OBA	OPS	TA	total	%	total
Luis Aparicio	1957	1968	11	5	.302	.641	.579	1188	73	1276
Dave Bancroft	1918	1924	4	3	.373	.748	.709	406	87	614
Dick Bartell	1929	1943	5	3	.360	.752	.715	468	71	615
Donie Bush	1912	1919	7	5	.359	.657	.672	840	83	1000
Bert Campaneris	1965	1973	9	9	.306	.659	.626	1196	91	1440
Frankie Crosetti	1934	1943	8	8	.343	.696	.658	895	84	915
Freddie Patek	1971	1975	5	2	.307	.623	.586	529	75	622
Pee Wee Reese	1940	1950	5	1	.358	.715	.721	465	68	538
Phil Rizzuto	1950	1952	2	0	.379	.771	.746	153	50	495
Maury Wills	1961	1971	10	10	.333	.666	.610	1342	90	1502

First, let me explain the column entries.

The first section (Principal Leadoff Batter Seasons) indicates the "first" and "last" years that the player was a principal leadoff batter. The "total" column gives the total number of principal leadoff batter seasons the player had in his major league career. The "≥ 75%" column gives the number of PLOB seasons in which the player was the leadoff man in at least 75 % of all the games he played (excluding pinch hitting assignments) in those seasons.

The next section (PLOB Composite Results) presents three metrics for evaluating batting performance – any batter's performance, including leadoff batters. The three metrics are On Base Average (OBA), On Base Plus Slugging Percentage (OPS), and Total Average (TA). There are, of course, other metrics that can be used to evaluate batting performance – e.g., batting average, slugging percentage, runs created, win shares, etc. I have chosen OBA, OPS, and TA because they include each of the elements generally deemed crucial for leadoff batter performance – i.e., getting on base and advancing on the bases so as be in (good) position to be driven home by the hitters in the heart of the batting order. OBA deals exclusively with getting on base. OPS deals with getting on base and advancing on the bases via extra base hits. And TA deals with getting on base and advancing on the bases via both extra base hits and stolen bases.

The entries in the OBA, OPS, and TA columns are the composite values for the players during their principal leadoff batter seasons *exclusively*. For example, Luis Aparicio played in the majors for 18 years, from 1956 through 1973. He was a principal leadoff batter in 11 of those seasons (1957-1963 and 1965-1968). Thus, his composite OBA, OPS, and TA entries are those calculated by considering just the 1957-1963 and 1965-1968 seasons. However, the player's pertinent full-season statistics (rather than pure leadoff batter statistics) were used to calculate the composite values.

In the third section (PLOB Games), the player's total number of leadoff games during his principal leadoff batter seasons are given along with the percentage of his principal leadoff batter games compared to all the games he played during his principal leadoff batter seasons. For example, Dave Bancroft was a principal leadoff batter for 4 seasons (1918 and 1922-1924). In those 4 seasons he was a leadoff batter for 406 games, which is 87% of the 467 total games he played in 1918 and 1922-1924.

In the last column (LOG), the player's total number of leadoff games during his entire career is given. For example, Dick Bartell (who was a principal leadoff batter for a total of 468 games during the 1929, 1934, 1937, 1940, and 1943 seasons) was employed as a leadoff batter in a total of 615 games in his career (1927-1946).

Upon inspection of the composite OBA, OPS, and TA information presented in this chart, it does not seem clear-cut to me that Reese would be rated as "the most effective leadoff man" among these players. Among these 10 players, Reese ranked fifth in composite OBA, fourth in OPS, and second in TA.

Also perplexing is the fact that James did not include Eddie Joost, Don Kessinger, and Rabbit Maranville in his list. As shown in the next chart, these players were all "shortstops who were leadoff men with long careers." Furthermore, it seems dubious to me that Reese would be called "the best career leadoff man among the shortstops" when Joost's numbers dwarf Reese's.

	Principal Leadoff Batter Seasons				PLOB Composite Results			PLOB Games		LOG
Shortstop	first	last	total	≥ 75%	OBA	OPS	TA	total	%	total
Eddie Joost	1942	1952	6	4	.384	.789	.813	745	89	954
Don Kessinger	1967	1975	8	6	.318	.640	.541	1059	86	1136
Rabbit Maranville	1913	1932	7	3	.321	.653	.595	753	73	897
Pee Wee Reese	1940	1950	5	1	.358	.715	.721	465	68	538

Moving on, now, to the other section in his book dealing with leadoff batters, James writes the following:[3]

> *How do you rate the greatest leadoff men of all time? You can do it however you want, but here's one way. First, you can estimate how many runs the player should score by what I call the leadoff man formula, which I have printed many times [although no references are given] ...take the number of times the player has been on first base, multiply by .35, his times on second by.55, his times on third by .8, and his home runs by 1. Many players, and most modern leadoff men, will actually score about the number of runs that the formula says they should score.*
>
> *One can turn that into a rating of the greatest leadoff men by*
>
> 1. *Converting the Expected Runs Scored into Expected Runs Scored per 27 outs.*
> 2. *Contrasting that figure with the league average for runs scored per out during the player's career.*
>
> *Obviously imperfect, for many reasons, but still…sometimes it is helpful to take a fresh look at these kind of issues with new methods even if the new methods are imperfect.*
>
> *All of the greatest leadoff men ever, by this method, would be guys who aren't leadoff men, starting with Ted Williams. (Williams, Ruth, Mantle, Barry Bonds, Ty Cobb, Musial, Joe Jackson, Hornsby, Frank Robinson, and Willie Mays.) Why Robinson ranks ahead of Mays I don't know and don't care, but anyway, this is logical on its own terms: if you had two Ted Williamses, and could afford to use one of them as a leadoff man, he would be the greatest leadoff man who ever lived.*
>
> *What we want, of course, are the greatest leadoff men who were actually leadoff men. That list is:*
>
> | | | |
> |---|---|---|
> | *1.* | *Rickey Henderson* | *1.67* |
> | *2.* | *Tim Raines* | *1.64* |
> | *3.* | *Topsy Hartsel* | *1.61* |
> | *4.* | *Lenny Dykstra* | *1.59* |
> | *5.* | *Wade Boggs* | *1.57* |

James then proceeds to mention those players who placed 6th through 20th and then those who ranked 21st through 65th.

There should be – and is – no quibble with the "leadoff man formula" approach to ranking "the greatest leadoff men who were actually leadoff men" (although James applies it to the player's entire career rather than to just those seasons in which the player was actually employed as a leadoff batter). There is, however, no justifiable reason for including players who *aren't* leadoff men and omitting players who are bona fide leadoff batters.

[3] Bill James, "*The New Bill James Historical Baseball Abstract*," pp. 684-685 (Simon & Schuster, 2003).

The following chart lists all of the players that James included as being among "the greatest leadoff men who were actually leadoff men." Also provided is the information (from the 1901-2000 period) relative to their participation as leadoff batters.[3] While no leadoff batter performance information is given, the players are listed in the rank-order that James listed them. Players listed in boldface type are those that had less than 5 principal leadoff batter seasons and/or less than 3 "≥ 75%" principal leadoff batter seasons.

rank	Player	Principal Leadoff Batter Seasons				PLOB Games		LOG
		First	Last	Total	≥ 75%	Total	%	Total
1	Rickey Henderson	1979	2000	22	22	2727	93	2727
2	Tim Raines	1981	1994	12	6	1205	74	1397
3	Topsy Hartsel	1901	1910	10	10	1172	93	1175
4	Lenny Dykstra	1986	1994	9	8	951	87	1095
5	Wade Boggs	1984	1996	7	3	672	64	864
6	Bobby Bonds	1969	1975	7	3	741	69	913
7	**Augie Galan**	1935	1937	2	2	266	88	384
8	Craig Biggio	1992	1999	5	5	784	98	926
9	Eddie Stanky	1945	1951	6	6	856	98	967
10	Pete Rose	1963	1981	15	13	2049	89	2298
11	Don Buford	1966	1972	6	5	628	76	743
12	Roy Thomas	1901	1909	9	8	1076	95	1092
13	**Rod Carew**	1981	1983	2	2	168	76	357
14	Stan Hack	1936	1947	11	10	1228	84	1360
15	**Elbie Fletcher**	-----	-----	ZERO	ZERO	ZERO	ZERO	7
16	Miller Huggins	1904	1915	11	8	1280	85	1303
17	**Lonnie Smith**	1982	1990	4	0	312	55	540
18	Bob Bescher	1909	1916	7	5	825	83	885
19	Billy Hamilton		1900					
20	John McGraw		1901					
21	Eddie Yost	1947	1960	14	13	1690	86	1729
22	Richie Ashburn	1948	1962	11	9	1286	82	1410
23	Lou Brock	1962	1977	14	11	1761	83	1893
24	Davey Lopes	1973	1981	9	9	1129	94	1195
25	**Gary Redus**	1983	1992	5	2	332	61	591
26	Burt Shotton	1911	1918	8	7	1060	92	1148
27	Ron Hunt	1966	1874	6	4	571	72	620
28	Tommy Harper	1965	1974	9	8	1120	88	1144
29	Dom DiMaggio	1940	1952	8	8	1020	91	1039
30	**Johnny Pesky**	-----	-----	ZERO	ZERO	ZERO	ZERO	25
31	George Burns	1913	1923	10	7	1348	91	1389
32	Paul Molitor	1978	1991	13	13	1525	91	1570
33	Max Bishop	1924	1933	10	10	1107	94	1201
34	**Max Carey**	1914	1928	8	2	569	53	696
35	Brett Butler	1981	1997	14	13	1706	88	1845
36	**Ray Chapman**	-----	-----	ZERO	ZERO	ZERO	ZERO	44
37	Earle Combs	1925	1933	8	7	896	79	1054
38	**Pee Wee Reese**	1940	1950	5	1	465	68	538
39	Billy North	1974	1980	5	4	587	82	753
40	Brady Anderson	1989	2000	9	6	1084	87	1210
41	Lu Blue	1922	1931	6	6	711	86	913
42	Ron LeFlore	1975	1982	8	8	980	94	1039
43	Jim Gilliam	1953	1960	8	8	1011	85	1025
44	**Matty Alou**	1966	1970	4	3	479	80	570
45	Donie Bush	1912	1919	7	5	840	83	1000
46	Johnny Temple	1955	1962	8	8	985	94	1050
47	Vince Coleman	1985	1995	11	11	1259	95	1274
48	Mookie Wilson	1981	1990	5	3	462	72	739
49	**Bill Werber**	1939	1942	4	3	412	83	501
50	Mickey Rivers	1974	1980	6	5	697	86	798

rank	Player	Principal Leadoff Batter Seasons				PLOB Games		LOG
		First	Last	Total	≥ 75%	Total	%	Total
51	**Dave Collins**	1976	1981	4	3	381	83	665
52	Willie Wilson	1979	1989	11	8	1271	84	1366
53	Billy Bruton	1953	1964	6	4	687	83	820
54	Maury Wills	1961	1971	10	10	1342	90	1502
55	**Woodie English**	1928	1933	2	1	144	65	310
56	Lloyd Waner	1927	1939	9	7	962	85	1203
57	George Case	1938	1946	9	8	1043	89	1062
58	Red Schoendienst	1945	1958	6	3	607	75	713
59	**Phil Rizzuto**	1950	1952	2	0	153	50	495
60	Bert Campaneris	1965	1973	9	9	1196	91	1440
61	Luis Aparicio	1957	1968	11	5	1188	73	1276
62	Don Blasingame	1956	1965	9	5	916	77	973
63	**Gary Pettis**	1984	1990	6	1	519	65	586
64	Otis Nixon	1991	1998	8	8	900	90	1104
65	Dummy Hoy		1902					

Examination of the player composition of the chart reveals that three of the players – Billy Hamilton (#19), John McGraw (#20), and Dummy Hoy (#65) – would be classified as 19th century players; as indicated, they concluded their principal leadoff batter careers in 1900, 1901, and 1902, respectively.

There are 15 players with their names listed in boldface type, indicating that they were not principal leadoff batters for 5 or more seasons and/or they were not "≥ 75%" principal leadoff batters for 3 or more seasons. That means that 24% of the 62 twentieth century players in James' list should not be classified as "essentially leadoff men." Included among these 15 players are three players who were *never* a principal leadoff batter during their major league careers – Elbie Fletcher, Johnny Pesky, and Ray Chapman.

Fletcher was *never* a principal leadoff batter during his entire major league career (1934-1949). In fact, he was a leadoff batter in only seven games during his whole career. Yet, James ranked him as the 15th greatest leadoff man of all time!

Pesky, as noted above, was *never* a principal leadoff batter during his entire major league career (1942, 1946-1954). He was employed as a leadoff batter in a meager 25 games throughout his career. Unbelievably, James ranked him as the 30th greatest leadoff batter of all time!

Chapman, as pointed out above, was *never* a principal leadoff batter during his entire career (1912-1920). He was utilized as a leadoff batter in just 44 games in the duration of his career. Incredibly, James ranked him as the 36th greatest leadoff batter ever!

It is extraordinary that Fletcher, Pesky, and Chapman, along with the other 12 players who were not principal leadoff batters for at least 5 seasons or who were not 75% principal leadoff batters for 3 or more seasons, would be ranked by James as being among the 65 players comprising the "greatest leadoff men" in ML history.

Another incredible aspect of the James list of "the greatest leadoff men who were actually leadoff men" is that several bona fide leadoff men were omitted. For example – Felipe Alou, Al Bumbry, Max Flack, Harry Hooper, Charlie Jamieson, Eddie Joost, Joe Judge, Chuck Knoblauch, Harvey Kuenn, Nemo Leibold, Kenny Lofton, Tony Phillips, Willie Randolph, Bip Roberts, Jimmy Slagle, Lou Whitaker, and Eric Young. Each of these players was a principal leadoff batter for at least five seasons and a 75% principal leadoff batter for at least four seasons.

In summary, in my opinion, James has done a poor job in discussing leadoff batters – he has (arbitrarily) included players who were *not* "essentially leadoff men" and excluded players who truly were "essentially leadoff men." Without a proper (i.e., complete and accurate) domain of "essentially leadoff men," one should not attempt to rate the performances of them.

Considering all of the above, I conclude the leadoff batter discussions that James gives in his book are of questionable value at best.

Herm Krabbenhoft, BQR9343@aol.com ♦

Correction

Due to an editing error in the August, 2003 issue of BTN, the formula in the footnote on page 6 was not correct. The correct formula is as follows:

$$1-\frac{6\sum D^2}{n(n^2-1)}$$

The editor apologizes for the error.

Informal Peer Review

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

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

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

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

Member	E-mail	Expertise
Jim Box	im.box@duke.edu	Statistics
Keith Carlson	kcarlson2@mindspring.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	HanrahanTJ@navair.navy.mil	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
Duke Rankin	RankinD@montevallo.edu	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

Using Calculus to Relate Runs to Wins: Part I

Ralph Caola

How many extra runs does it take to create an additional win? There are two existing answers to that question: first, the rule of thumb that ten runs equals one win; and, second, Pete Palmer's formula based on the teams' total runs per game. But in this study, the authorgives a more robust answer – one that applies to teams of differing abilities, and one that is derived mathematically, rather than statistically, from the Pythagorean Projection formula on which it is based.

Before Sir Isaac Newton went to college at Cambridge, he knew very little formal mathematics, or so the story goes. To begin learning, he bought a copy of Euclid's *Elements of Geometry*. In the time it takes me to program my VCR, he had invented Calculus to help formulate his monumental Theory of Gravity. Little did Newton know, 300 years later, his Calculus would be applied to a subject as frivolous as baseball.

Ten Runs Per Win?

Bill James' Pythagorean Theorem predicts winning percentage (W%) as a function of runs scored (Rs) and runs allowed (Ra). The formula is:

$$W\% = \frac{Rs^2}{Rs^2 + Ra^2}$$

or, in terms of wins (W),

$$W = G\frac{Rs^2}{Rs^2 + Ra^2}$$

(where G is the total number of games played.)

Since the formula expresses wins in terms of runs, I wanted to find out how many runs per win it would predict and compare the result to Pete Palmer's "10 runs per win" rule of thumb. (See Chapter 4 of *The Hidden Game of Baseball,* in the section titled "Runs and Wins".)

This is where calculus comes in. To derive an expression for runs per win, I held Ra constant and took the partial derivative of W with respect to Rs. This gives the number of incremental wins per run scored. Then, I inverted the result, to get the number of incremental runs scored per win (Rsi/W). Therefore, *the result is the number of runs needed to get one more win*. The result of the differentiation is:

$$\frac{dRs}{dW} = \frac{(Rs^2 + Ra^2)^2}{2GRsRa^2} \qquad \text{(equation 1)}$$

If Rs and Ra are in runs per game, the factor of G is not needed and

$$\frac{dRs}{dW} = \frac{(Rs^2 + Ra^2)^2}{2RsRa^2} \qquad \text{(equation 1a)}$$

Notice what happens when Rs = Ra = 810 runs. 810 runs is 5 runs per game over 162 games.

```
dRs/dW = (4*810^4) / (2*162*810^3)
       = (4*810)/(2*162)
       = 10 runs per win!
```

So James' formula predicts exactly 10 runs per win when Rs = Ra = 810 runs.

Notice what happens when Rs = Ra = R, in general. Substituting into $\frac{dRs}{dW}$, we get:

$$\frac{dRs}{dW} = \frac{4R^4}{2GR^3} = \frac{2R}{G}.$$

This also supports the 10 runs per win rule. Again, when teams score 810 runs,

```
dRs/dW = 2*810/162 = 10 runs per win.
```

More generally, it also tells us that, for a .500 team (Rs=Ra), the number of incremental runs scored per win *is twice the number of runs scored per team per game***.** Table 1 illustrates this at selected values of Rs.

So teams need to score 10 extra runs per season to get one extra win. But, this result is valid only for a .500 team that scores 810 runs (5 runs per game).

Table 1 – Runs Per Win for a Team with a .500 Winning Percentage (Rs = Ra)

Rs and Ra (Runs)	Runs Per Team per Game (Rs/G)	Incremental Runs per Win (Rsi/W)
162	1	2
324	2	4
486	3	6
648	4	8
810	5	10
972	6	12

Varying Runs Scored and Allowed

The previous results were derived for a team with the same number of runs scored and allowed – a team with a .500 record. How many incremental runs scored are needed per win for other winning percentages?

Though Ra was treated as a constant when calculating the derivative, it remains a variable, along with Rs, in the resulting equation. Table 2 shows runs per win for various values of Rs and Ra. In all tables in which it was appropriate, I shaded all values within 10% of 10. This shows that 10 runs per win is a good approximation at many common values of Rs and Ra.

For seasons from 1901 through 2003, more than 97% of all teams scored and allowed between 3 and 6 runs per game and the average was 4.4 runs per game.

In Table 2, notice the cases in which runs scored and allowed are within 0.5 run of each other – for example, Ra = 4.5 and Rs = 5.0, or Ra = 4.0 and Rs = 3.5. For those combinations, incremental runs scored per win are within a run of the sum of runs scored and runs allowed (Rs + Ra). Only when runs scored and allowed differ significantly, is incremental runs scored per win very different from Rs + Ra.

There is a counterintuitive asymmetry in Table 2. For a team that scores 4 runs per game and allows 3 runs, it takes 9 runs to produce a win (8.68), while for a team that scores 3 runs per game and allows 4 runs, only 7 runs (6.51) are required for an extra win.

Table 2 – Runs Per Win given Runs Scored and Runs Allowed

Ra (runs allowed/game)	Rs (runs scored/game)						
	3.0	3.5	4.0	4.5	5.0	5.5	6.0
3.0	6	7	9	11	13	16	19
3.5	7	7	8	10	11	13	16
4.0	7	7	8	9	11	12	14
4.5	7	8	8	9	10	11	13
5.0	8	8	8	9	10	11	12
5.5	9	9	9	10	10	11	12
6.0	9	9	9	10	10	11	12

Why aren't the (4,3) and (3,4) values from Table 2 equal? To find out, consider an extreme example: Team 1 scores 8 runs per game and allows only 2 and Team 2 scores 2 runs per game and allows 8. Equation 1a says the incremental runs scored per win for Team 1 is 72 and the incremental runs scored per win for Team 2 is 18. (The ratio of the two values of the incremental runs scored per win is always the same as the ratio of runs scored to runs allowed (72/18 = 8/2 = 4)).

Here's what I think is going on:

The Pythagorean Theorem predicts a 0.941 winning percentage for Team 1 and 0.059 for Team 2. So, Team 1 will win about 152 of 162 games, and Team 2 will win only 10 games.

Team 2, being so bad, is losing a lot of blowouts. That makes its incremental runs per win high compared to the benchmark value of 10 runs per win. It needs an above average number of extra runs to turn its losses into wins, because its losses are by an unusually large margin.

This is also why, as a team gets worse – as the difference between its runs allowed and runs scored increases – its incremental runs scored per win also increases. You can see this effect by proceeding down column 1 of Table 2.

But, if being bad makes incremental runs per win high, why is Team 1's four times higher than Team 2's? Because Team 1 has so few losses (10) to turn into wins. For example, if Team 1 scores 10 more runs, 1306 instead of 1296, the runs will probably occur in games they would have won anyway. So, Team A has to score a lot more runs before any of them happen to come in its rare losses. Even then, it has to score enough to turn one of those losses into a win.

This is also why, as a team gets better - as the difference between its runs scored and runs allowed increases – its incremental runs scored per win also increases. You can see this effect by proceeding across row 1 of Table 2.

Winning Percentages

We can also determine runs per win using winning percentages instead of runs per game. Starting with the Pythagorean Theorem and solving for Ra, we get:

$$Ra = Rs\sqrt{\frac{1-W\%}{W\%}}$$

Substituting this for Ra in equation 1 gives:

$$\frac{dRs}{dW} = \left(\frac{1}{2}\right)\left(\frac{Rs}{G}\right)\left(\frac{1}{W\%(1-W\%)}\right) \qquad \text{(Equation 2)}$$

Equation 2 is now a way to express Runs Per Win by Runs Scored and Winning Percentage. Table 3 shows the results.

Notice, more runs are needed as winning percentage deviates from .500. For example, 10 more runs scored are needed at Rs = 5 runs per game and .500 winning percentage, whereas 12 more are needed at .300 or .700. Notice also, the number of extra runs needed is symmetric about a .500 winning percentage.

Table 3 – Incremental Runs Scored Per Win for Various Values of Runs Scored per Game and Winning Percentage

Rs (runs/game)	Winning Percentage								
	.300	.350	.400	.450	.500	.550	.600	.650	.700
3.0	7	7	6	6	6	6	6	7	7
3.5	8	8	7	7	7	7	7	8	8
4.0	10	9	8	8	8	8	8	9	10
4.5	11	10	9	9	9	9	9	10	11
5.0	12	11	10	10	10	10	10	11	12
5.5	13	12	11	11	11	11	11	12	13
6.0	14	13	13	12	12	12	13	13	14

For seasons from 1901 through 2003, more than 98% of all teams had winning percentages between .300 and .700.

In Table 3, for winning percentages near 0.500, incremental runs scored per win is twice runs scored (2Rs).

General Exponents

Analysts have tried using exponents other than 2 to make the Pythagorean Theorem more accurate. (Again, see Chapter 4 of *The Hidden Game of Baseball*, in the section titled "Runs and Wins".) The Theorem expressed with a general exponent is:

$$W\% = \frac{Rs^x}{Rs^x + Ra^x}$$

The corresponding expression for incremental runs scored per win is:

$$\frac{dRs}{dW} = \frac{(Rs^x + Ra^x)^2}{(xRs^{x-1}Ra^x)} \qquad \text{(equation 3)}$$

where Rs and Ra are runs per game.

For a .500 team, Rs=Ra=R, and the expression reduces to

$$\frac{dRs}{dW} = R\left(\frac{4}{x}\right)$$

where R is runs per game.

Notice, when we use the typical value for the exponent, x=2, the expression reverts to dRs/dW = 2R.

It has been found that using an exponent of 1.83 makes the Pythagorean Theorem a bit more accurate than an exponent of 2.

With x = 1.83, dRs/dW = 2.19*R.

Tables 4, 5, and 6 repeat tables 1, 2, and 3, but for exponent 1.83 instead of 2.

Table 4: Incremental Runs Scored Per Win For a Team With a.500 Winning Percentage (Rs = Ra) and Pythagorean Exponent 1.83)

Rs, Ra (Runs)	Runs per Team per Game (Rs/G)	Incremental Runs Per Win (Rsi/G)
162	1	2.2
324	2	4.4
486	3	6.6
648	4	8.7
810	5	10.9
972	6	13.1

Table 5: Incremental Runs Scored Per Win for Various Values of Runs Scored and Runs Allowed per Game (Exponent 1.83)

Ra (runs/game)	Rs (runs/game)						
	3.0	3.5	4.0	4.5	5.0	5.5	6.0
3.0	7	8	9	11	14	16	19
3.5	7	8	9	10	12	14	17
4.0	7	8	9	10	11	13	15
4.5	8	8	9	10	11	12	14
5.0	8	9	9	10	11	12	13
5.5	9	9	10	10	11	12	13
6.0	10	10	10	11	11	12	13

Incremental Runs Allowed Per Win

Equation 1 was based on runs scored – but we could have used runs allowed instead. Following the same procedure as before, incremental runs allowed per win is

$$\frac{dRa}{dW} = -\frac{(Ra^2 + Rs^2)^2}{2GRaRs^2}$$

The minus sign means a team must decrease its runs allowed to get an incremental win. Notice that this expression is the same as the one for incremental runs scored per win (dRs/dW), except that Rs and Ra have exchanged places (and the minus sign). This means the table for dRa/dW would be the same as Table 2, except the rows and columns would be interchanged. So, if you wanted to find the incremental runs allowed per win for a team that scores 4 and allows 3 (Rs = 4, Ra = 3), you could look in Table 2 in the Rs = 3, Ra = 4 position and take the negative of that entry.

Differential and Total Runs

In the first part of this article, I analyzed incremental runs scored and allowed per win starting with the Pythagorean Theorem expressed in terms of runs scored and allowed. I did this because it was the most straightforward approach. However, most formulas for runs per win, like the "10 run rule," are not expressed in terms of runs scored and allowed, but as the difference between the two. The ten run rule, expressed mathematically, is

```
W = 81 + Rd/10
```

where 81 is the number of wins needed for a .500 record in a 162 game season and where differential runs, Rd = Rs – Ra.

Table 6: Incremental Runs Scored Per Win for Various Values of Runs Scored per Game and Winning Percentage (Exponent 1.83)

Rs (runs/game)	Winning Percentage								
	.300	.350	.400	.450	.500	.550	.600	.650	.700
3.0	8	7	7	7	7	7	7	7	8
3.5	9	8	8	8	8	8	8	8	9
4.0	10	9	9	9	9	9	9	9	10
4.5	11	11	10	10	10	10	10	11	11
5.0	13	12	11	11	11	11	11	12	12
5.5	14	13	12	12	12	12	12	13	14
6.0	15	14	14	13	13	13	14	14	15

For example, when Rs =750 and Ra = 700, Rd = 750 – 700 = 50 and the 10 run rule predicts

```
W = 81+ 50/10 = 81 + 5 = 86 wins.
```

To calculate the derivative in terms of Rd, I first had to express the Pythagorean Theorem in terms of Rd and another variable instead of in terms of Rs and Ra. To do this, I transformed the Pythagorean Theorem from runs scored and allowed to differential runs (Rd) and total runs (Rt) using

```
Rd = Rs - Ra  and
Rt = Rs + Ra.
```

Solving for Rs and Ra

```
Rs = (Rt + Rd)/2 and
Ra = (Rt - Rd)/2.
```

Table 7: Winning Percentages for Various Combinations of Total Runs and Differential Runs Per Game as Calculated by the Pythagorean Theorem

Rd (runs/game)	Rt (runs/game)						
	6	7	8	9	10	11	12
-2	.200	.236	.265	.288	.308	.324	.338
-1	.338	.360	.377	.390	.401	.410	.417
0	.500	.500	.500	.500	.500	.500	.500
1	.662	.640	.623	.610	.599	.590	.583
2	.800	.764	.735	.712	.692	.676	.662

Substituting these expressions into the Pythagorean equation yields

$$W\% = \frac{(Rt+Rd)^2}{(Rt+Rd)^2 + (Rt-Rd)^2} \quad \text{(equation 4)}$$

Rt and Rd can be expressed in runs or runs per game.

Table 7 shows winning percentages calculated using equation 4 at various values of total and differential runs. For seasons from 1901 through 2003, about 80% of all teams had between –1 and +1 differential runs per game. More than 99% of all teams had between 6 and 12 total runs per game and the average was 8.8 runs per game.

In Table 7, notice that winning percentage is symmetric about .500. For example, at Rt = 11 and Rd = -1 the winning percentage is .410, which is .500 - .090. At Rt = 11 and Rd = +1, the winning percentage is .590, which is .500 + .090.

To calculate incremental differential runs per win, I took the partial derivative of W% with respect to Rd, using equation 4, and inverted. The result is

$$\frac{dRd}{dW} = \frac{\left(Rt^2 + Rd^2\right)^2}{Rt\left(Rt^2 - Rd^2\right)} \qquad \text{(equation 5)}$$

Here, Rt and Rd are in runs per game.

Notice, when Rd = 0, the equation simplifies to

$$\begin{aligned} dRd/dW &= Rt^4/Rt^3 \\ &= Rt \\ &= Rs + Ra \end{aligned}$$

For a team with a .500 record, Rs = Ra = R, Rt = 2R and dRd/dW = 2R.

When both teams score 5 runs per game, dRd/dW = 2R = 10, as previously calculated.

Table 8: Incremental Differential Runs Per Win for Various Values of Differential Runs and Total Runs per Game

Rd (runs/game)	Rt (runs/game) 6	7	8	9	10	11	12
0.0	6	7	8	9	10	11	12
+/- 0.5	6	7	8	9	10	11	12
+/- 1.0	7	7	8	9	10	11	12
+/- 1.5	7	8	9	10	11	12	13
+/- 2.0	8	9	10	10	11	12	13

So, the incremental differential runs per win is equal to the total runs scored by both teams per game (Rt). This is the same result we got in the previous analysis when Rs = Ra (Rd = 0). Rd = 0 implies a team with a winning percentage of .500. So, a team with a winning percentage of .500, in order to get one extra win, needs to increase its difference between runs scored and allowed by the total of it's runs scored and allowed per game.

Another thing we can see from equation 5 is that we get the same result whether Rd is positive or negative. That's because both times Rd appears, it is squared, and the square of any real number is positive. So, it takes as many runs for a .400 team to get an extra win as it does for a .600 team. Table 8 shows incremental differential runs per win for various values of differential runs and total runs per game.

Note that the incremental runs per win are equal to total runs per game (Rt) for Rd's between –1 and +1 run per game. This happens because, for most practical values of Rt and Rd, Rt^2 is much greater than Rd^2.

For example, when Rt=8 and Rd=1,

$$\begin{aligned} dRd/dW &= (Rt^2 + Rd^2)^2 / [Rt*(Rt^2 - Rd^2)] \\ &= (64 + 1)^2 / [8*(64 - 1)] \\ &= 8.4 \end{aligned}$$

If we ignore Rd, we get

$$\begin{aligned} dRd/dW &= (8^2)^2 / 8*8^2 \\ &= 8^4/8^3 \\ &= 8 \end{aligned}$$

The difference between 8 and 8.4 is only 5%. For values of Rt greater than 8 or values of Rd less than 1, the difference is even less. The way I've calculated runs per win is different from the way it's been done in the past. Once again, the ten run rule is:

W = 81 + Rd/10

Another way of stating it is

```
Rd/(W – 81) = 10.
```

This ratio is the *average* runs required per win above 81. By taking the derivative, I've calculated, not the *average* runs required per win, but the number of runs required to get the *next* (marginal) win. The resulting equation indicates the number of incremental runs per win is not constant, but is a function of the total runs per game (Rt).

In *The Hidden Game of Baseball* (again Chapter 4, "Runs and Wins" section), Palmer recognized that the number of runs required per win was not constant and that teams involved in high scoring games needed more runs to produce a win. He arrived at the following equation for runs per win:

$$RunsPerWin = 10\sqrt{\frac{Rt}{9}}$$

So, when Rt = 9, Runs per win = 10.

As shown in Table 9, this formula produces results different from the ones presented here.

Table 9: Comparison of Runs Per Win Predicted by Various Methods as a Function of Total Runs Per Game

Total Runs Per Game (Rt)	dRd/dW (Rd=0)	Ten Run Rule	Rd=10* Sqr(Rt/9)
6	6	10	8.1
7	7	10	8.8
8	8	10	9.4
9	9	10	10.0
10	10	10	10.5
11	11	10	11.1
12	12	10	11.5

Summary

Based on James' Pythagorean Theorem, I derived equations for incremental runs per win. Rather than average runs per win, incremental runs per win is the number of differential runs a team needs to produce its next win.

The equation for incremental runs scored per win is

$$\frac{dRs}{dW} = \frac{(Rs^2 + Ra^2)^2}{2GRsRa^2}$$

the equation for incremental runs allowed per win is

$$\frac{dRa}{dW} = -\frac{(Ra^2 + Rs^2)^2}{2GRaRs^2}$$

and the equation for incremental differential runs per win is

$$\frac{dRd}{dW} = \frac{\left(Rt^2 + Rd^2\right)^2}{\left[Rt\left(Rt^2 - Rd^2\right)\right]}$$

As the latter equation indicates, incremental differential runs per win is a function of total runs (Rt) and differential runs (Rd). Because, in most cases, Rt^2 is significantly greater than Rd^2, we can assume, with a good degree of accuracy

$$\frac{dRd}{dW} = Rt$$

Although the equations are theoretically different from Palmer's "ten runs per win" rule, both support Palmer's rule at many common values of runs scored and allowed and differential and total runs.

Ralph Caola, RJCSB25@aol.com ♦

Submissions

Phil Birnbaum, Editor

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

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

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

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

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

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

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

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

Week-to-Week Consistency Revisited

Charlie Pavitt

Are hitters "streaky" in the sense that a good week of hitting is likely to be followed by another good week? Are they the opposite, where a good week is more likely to be followed by a bad week? Or is it all just random? Here, the author revisits his 2002 study in search of better evidence on this question.

How consistent is offensive performance? I asked this question in an earlier essay published in the May 2002 issue of *By the Numbers.* Guessing that most of my current readers have that issue stored in their baseball attic, I will only summarize here what I there described in detail. Evidence from several studies searching for consistency across consecutive at bats (S. Christian Albright, *Journal of the American Statistical Association,* 1993, volume 88, pp. 1175-1183), games (Steven Copley, described by Bill James in the 1986 *Baseball Abstract*), and five game stretches (1987 *Elias Baseball Analyst*) has been consistent with the claim that streaks and slumps in offensive performance are random fluctuations. As Jim Albert stressed in a response to Albright's work (pp. 1184-1188), the data is also consistent with the claim that streaks and slumps in offensive performance are real phenomena that coincidentally happen to occur at rates that parallel what would occur with random fluctuations. Although present-day methods cannot definitively distinguish between these two claims, as long as data appear random, the former is the simpler and thus more likely explanation.

In order to search for non-randomness in offensive performance, my previous study assembled a data set consisting of 11 years (1991 to 2001) of week-by-week batting and slugging averages, consisting of players who had at least 10 Sunday-through-Saturday weeks in which they registered 10 or more official at bats over at least 4 seasons. I also included a few Kurt Manwaring/Jeff Reed type catchers who had several seasons with consistent 7 to 13 at bat weeks. I tended to cut out stretches in which players only played a couple of weeks in a season before, after, or around injuries.

I will explain the statistical analysis I performed in detail, as it is the basis of the methods used in the present study. It was conducted two ways; through Wald-Wolfowitz runs tests and regression-based time series analysis. A runs test (my sources were W. J. Conover's *Practical Nonparametric Statistics,* pages 349 to 356, and the second edition of William Hay's *Statistics for the Social Sciences*, pp. 775- 777) provides a z-score that represents the extent to which a sequentially-ordered set of data points includes stretches ("runs") that are consistently above or below the median for that data. A positive z-score, or more runs than would occur by chance, would indicate a circumstance in which good and bad weeks alternated non-randomly (analogous to a pattern of coin flips such as HTHTHTHTHT). A negative z-score, or fewer runs than would occur by chance, would indicate a circumstance in which long streaks and slumps alternated (analogous to HHHHHTTTTT). The regression-based time series (my source was Charles W. Ostrom's *Time Series Analysis*) provides a regression equation revealing whether performance improved or worsened across weeks and, more importantly, the Durbin-Watson test, which provides an index representing consistency of performance above or below the regression line. The index ranges from 0 to 4, such that 2 implies completely random variation around the regression line, a significantly large index implies more runs than chance allows, and a significant small index fewer runs than by chance. The time series analysis is more sensitive than the runs test, but technically should be limited to data sets with no missing values. This data has quite a few missing weeks due to injury and the like. The runs test does not have this limitation. The runs test also implies that a player's average performance remains at the median across a season, whereas the Durbin-Watson test can determine whether there is non-random stretches even within the context of a performance that generally improves or worsens from the beginning to the end of a season. For this reason, the two often lead to different conclusions for the same season.

I examined the data at the annual level, including only seasons in which a player had 15 weeks for which I had data. In order to control sample size, I also limited the analysis to players whom I believed were retired. These moves resulted in a total of 549 seasons from 93 players.

The results in a nutshell: The runs tests showed absolutely no tendencies for either consistency or inconsistency. The Durbin-Watson tests revealed a few more large indices than small, indicating a bit of inconsistency across weeks. One possible explanation for this finding is the alternation between home stands and road trips, which may approximate one week in length often enough to produce these findings. Another possibility, of course, is a slight tendency for real inconsistency in performance.

One reason that I limited this study to seasons rather than entire player careers was that the table I had for examining the statistical significance of Durbin-Watson indices (from Jam Kmenta's *Elements of Econometrics*, page 625) was limited to a sample size of 100. In my earlier essay, I asked the readers for information on how to work with larger sample sizes, and in that cooperative sabermetric spirit, I heard from both John Goldsmith and Rob Wood (a SABR Salute to both). After sifting through their suggestions, I settled on an approximation based on the standard normal described in A. C. Harvey's *The Econometric Analysis of Time Series* (page 201 of the second edition). Thus armed, I performed the analysis the way I originally wanted; using all the weeks of player's careers for which I had data. This analysis had an added complexity compared to the seasonal research, because, over the course of years, a batter's performance usually varies in an inverted-U pattern; in other words, improvement for a few seasons to a peak and then decline for a few more until retirement. As mentioned earlier, the use of runs tests presume constancy in a player's average performance, and, to the extent to which a career approximates an inverted-U, will be biased toward consistency. The analysis will find the player to be performing below median and thus unusually poorly for the first and last seasons of his career and above median and so unusually well around the peak. Durbin-Watson indices presume a linear progression in a player's average performance, and so could also misrepresent a player's natural performance trajectory. Also keep in mind that I was not using complete career data for players active before 1991 and after 2001.

I decided to use data for all (retired or active) players who had, in the 1991-2001 interim, amassed at least 4 seasons totaling 75 10 at-bat weeks, with at least 10 weeks in each season. I also dropped any season with fewer than 10 weeks from a player's data, with long-career platoon catchers again an exception. I ended up with a data set of 297 players (for the sake of trivia, Rafael Palmeiro had the most 10 at-bat weeks during that stretch, with 275). The statistical methods were identical to the previous study, with one addition; I added a quadratic term to the regressions in order to look for the expected inverted-U career progressions.

Table 1 – Runs Tests

Significance Level	Batting Average .05	Batting Average .10	Slugging Average .05	Slugging Average .10
Number of z's	297	297	297	297
Significant z's in data indicating inconsistency	4	8	6	11
Significant z's in data indicating consistency	8	15	11	22
Significant z's for each by expected by chance	7.5	15	7.5	15

Data for the runs tests are shown in Table 1. As in the earlier study, the runs tests for batting average showed no overall systematic tendencies for either inconsistency or consistency. There were no more significant z's for consistency, and fewer significant z's for inconsistency, than expected by chance. The distribution of batting average z's was close to symmetrical, with a mean z of -.02 and a median z of 0. The runs tests for slugging average were a bit more complicated. There were as many significant z's for inconsistency as expected by chance. There were a few more significant z's for consistency than chance would allow, but a second-order test of proportions found in Hubert M. Blalock Jr.'s *Social Statistics* showed that the preponderance was no greater than chance for either .05 or .10 significance levels. Nonetheless, the distribution of slugging average z's tended noticeably toward consistency, with a mean z of -.16 and a median z of -.18. The mean z differed significantly from 0 in a one-sample t test, with a probability of occurring by chance of only .006. Turning to the Durbin-Watson's, data can be found in Table 2. Indices for batting average again showed no tendencies for either inconsistency or consistency. There were no more significant indices for consistency, and fewer at .05 for inconsistency, than expected by chance. Once again, the results for slugging average showed more significant indices for consistency than expected by chance, and that preponderance was significant for both the .05 ($z = 2.13$) and .10 ($z = 2.71$) levels.

Table 2 – Durbin-Watson Indices

Significance Level	Batting Average .05	Batting Average .10	Slugging Average .05	Slugging Average .10
Number of indices	297	297	297	297
Significant indices indicating inconsistency	9	32	10	28
Significant indices indicating consistency	15	32	23	44
Significant indices for each by expected by chance	15	30	15	30

I must conclude that both the runs and Durbin-Watson tests show evidence for more consistency in performance than expected by chance for slugging average, but not for batting average. This is a strange finding, given that batting average makes up a good chunk of slugging average, and also given that the runs test indices for the two were correlated at +.367, indicating not surprisingly that they went up and down in tandem. Actually, it is surprising that this correlation is not higher, and I would speculate that whatever factor it is that leads to only a moderate association between ups and downs in hitting for average and hitting for power might also be implicated in the finding that the latter is more consistent than the former.

Data for the regressions themselves are displayed in Table 3. Beginning with the linear component, there were fewer significant regressions than expected by chance indicating either improvement or decline across seasons for batting averages and indicating decline across seasons for slugging average. In contrast, there were quite a few more significant linear regression coefficients indicating improvement in slugging average than chance would allow. In other words, more than a random number of players increased their power, but not their batting average, as their careers progressed. Turning to the quadratic component, both batting and slugging averages boasted a far greater number of inverted-U patterns than by chance, indicating the expected career trajectory of peak performance somewhere in the middle of a career. There were fewer U-shaped trajectories than would be expected by chance, and many of these were artifacts resulting from players having uncharacteristically productive 2001 seasons, skewing the last available data points upward (e.g., Rich Aurilia).

Now we need to address the bias toward consistency mentioned earlier. Given that there were a disproportionate number of inverted-U patterns, and a disproportionate amount of consistency, for slugging average across careers, then we run into the problem that both the runs and the Durbin-Watson tests might have picked up the relatively poor early and late years as unusually consistent stretches of poor performance and the relatively productive middle years as an unusually consistent stretch of good performance. The same implication follows from U-shaped patterns, except that here the beginning and end of careers would appear to be unusually good and the middle unusually poor. If so, the findings for nonrandom consistency would be a methodological artifact. There is evidence suggesting that this may be the case. Of the 69 careers with an inverted-U pattern significant at .10, 18 (26.1 percent) were more consistent than chance. Of the 15 careers with a U-shaped pattern significant at .10, there were 4 (26.7 percent) that were more consistent than chance. In contrast, of the 213 careers with no curvilinear pattern significant at .10, only 22 (10.3 percent) were more consistent than chance, which is exactly the amount one would expect given random processes. A chi-square test of the entire data sets matching quadratic components with Durbin-Watson indices found a significant relationship between the two with a probability of .013 of occurring by chance. Thus, I must conclude that there is no good evidence for consistency across weeks that cannot be interpreted independently of the ups and downs across seasons in normal career patterns.

Table 3 – Overall Performance Change

	Batting Average		Slugging Average	
Significance Level	**.05**	**.10**	**.05**	**.10**
Number of regressions	297	297	297	297
Significant linear regressions indicating improvement	17	24	34	50
Significant linear regressions indicating decline	12	20	7	16
Significant curvilinear regressions indicating U-shaped relationship	11	15	12	15
Significant curvilinear regressions inverted-U relationship	33	52	50	69
Significant regressions for each by expected by chance	15	30	15	30

In conclusion, there is no evidence in this data that batters are less consistent than chance would allow in their week-to-week performance across multiple seasons. Although there is evidence that batters are more consistent than would be expected by chance across weeks, it is highly possible that this evidence is an artifact of the tendency for many batters to have non-linear, usually inverted-U-shaped career trajectories. Consistently with all previous research in this area of which I am aware, there is no reason to believe from this data that changes in offensive performance across weeks, independently of normal career patterns, are anything more than random fluctuations.

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

Get Your Own Copy

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

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

The easist 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.

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

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.

Quantifying Persistence in Run Model Errors Across Team-Seasons

Ted Turocy

It has been noted that teams that beat their runs scored estimates (of statistics such as Runs Created) have no increased probability of repeating that feat the next season. This implies that the statistics capture all important persistent factors that go into runs scoring – otherwise, a tendency to repeat would exist. But the issue has not been studied in full – and it is still possible that there are some small persistent characteristics, such as team speed, that would lead to a small but significant chance of a repeat in outperforming or underperforming an estimate in subsequent seasons. Here, the author checks if that is indeed the case.

The baseball analysis community has estimated many different specifications for "production functions" for runs scored, both linear and nonlinear, over the last few decades. In general, it has been observed that teams that overperform or underperform relative to their predicted run production in a season will tend to regress to the mean in the following season. This suggests that there are generally no important team-specific features which persist from season to season that are not captured as inputs into these functions. This note seeks to quantitatively verify this folk wisdom in the context of specifications estimated by linear regression.

I use team-season-level data from 1974 through 1992, inclusive, and denominate all quantities in per-game-played terms. I investigate two specifications. The first specification is a basic linear estimator, where the vector of regressors is (1B,2B,3B,HR,BB,HP). The second specification is an extended specification designed in Turocy (2003) to correctly account for team speed in the estimation; the estimated specification is called Model 4 in that paper. Essentially, this specification adds data on stolen bases and caught stealing (separated by attempts of second base and third base), grounded into double plays, and advancement on errors.

The ordinary least squares fit of the basic model to team offensive data gives a standard error of .1484 runs per game, and an autocorrelation of the residuals of .1204. There are 488 team-seasons in the sample; simulation of a like number of independent normal random variables with standard deviation .1484 resulted in an empirical autocorrelation of the residuals greater than .1204 in only 4 out of 1000 simulations, implying that the autocorrelation is significant. The full model fit to the offensive data drops the standard error to .1382, and gives residuals with an autocorrelation of .0598; in 1000 simulations, the autocorrelation exceeded .0598 83 times, putting this autocorrelation near the boundary of significance at the standard levels.

The difference between the specifications is the inclusion of factors related to speed in the full model. To see if this is in fact what is being controlled, the same specifications are estimated using data for teams on defense. In this case, the basic specification results in a standard error of .1499 with an autocorrelation of .0932 (simulated p-value .027) and the full specification a standard error of .1445 (simulated p-value .067). Inclusion of the additional regressors does not reduce the standard error and observed autocorrelation of residuals as much as when using offensive data, suggesting that these additional regressors are in fact picking up primarily differences in speed.

To correctly account for this autocorrelation, I assume that the error term in the regression, instead of being uncorrelated, follows a first-order autoregressive process given by $\varepsilon_{ti}=\rho\varepsilon_{t,i-1}+u_{ti}$, where u_{ti} is independent across teams and seasons (where i indexes teams and t indexes time). The value of ρ is estimated by the Cochrane-Orcutt procedure, which chooses ρ such that the sum of squares of residuals of the corresponding ordinary-least-squares estimators is minimized. The resulting estimates of ρ closely correspond to the autocorrelations observed above.

Specification	Estimated ρ
Basic, offense	.124
Full, offense	.063
Basic, defense	.096
Full, defense	.076

Importantly, however, the coefficient estimates obtained from the regression model augmented with autocorrelated errors do not differ in any significant fashion from those in the regressions ignoring the autocorrelation. Therefore, neglecting the autocorrelation in estimating the parameters of these type of models is likely not a problem.

Even though the autocorrelation does appear to be statistically significant, it is not substantial in magnitude. The autocorrelation of about .12 in the basic offense model implies that a team that outperforms the model's prediction by 30 runs in a season would expect, on average, to outperform the model's prediction by only about 4 runs the following campaign.

An interesting question is whether adding further factors to the model could further reduce the autocorrelation. Some candidates for sources of the autocorrelation of the residuals include strategy and managing styles, and team composition (balanced versus unbalanced lineups, for example). Also, note that while this estimates a linear approximation to the production of runs, the true function is nonlinear, but with a modest curvature. This alone could account for the autocorrelation, since teams that score many runs in one season are likely to do so in the next as well (the autocorrelation of runs per game in the sample is .42).

References

Turocy, T. L. (2003) *Offensive Performance, Omitted Variables, and the Value of Speed in Baseball*, http://econweb.tamu.edu/turocy/papers/runest.htm

--

Theodore L. Turocy, Department of Economics, Texas A&M University, College Station TX 77843, turocy@econmail.tamu.edu ♦