

Why It's So Hard to Hit .400

New Insights into an Old Statistic

Batting average measures the batter's ability to avoid striking out and his ability to "hit 'em where they ain't." However, the relative importance of these two skills is blurred by the traditional way in which we write AVG (H/AB). By rewriting AVG as a function of strikeouts per at-bat and hits per ball put in play, we gain greater insight into what has driven AVG over the last 100 years, why hitting .400 is so difficult, and whether or not batters can really place base hits.

To see the impact of strikeouts and putting the ball in play, let's write AVG in terms of these effects, as shown in Equations 1 and 2. BIP is the number of times the batter put the ball in play—not including sacrifices

$$AVG = \frac{H}{AB} = \frac{H}{AB} \times \frac{BIP}{BIP} = \frac{BIP}{AB} \times \frac{H}{BIP} \quad (1)$$

From Equation 1 we see that AVG is equal to the fraction of at-bats where the ball was put in play (BIP/AB) multiplied by the fraction of balls put in play that went for hits (H/BIP). This is intuitive—first the batter must put the ball in play and then it has to "become" a hit.

Since there are only two possible ways to have an at-bat: (1) the ball is put in play or (2) the batter strikes out, $BIP/AB = 1 - K/AB$, where K/AB is the fraction of at-bats where the batter struck out. Let's call the term K/AB the batter's "strike out average" or KAVG and H/BIP his "in-play average" or IPAVG. We can then write Equation 1 as

$$AVG = (1 - KAVG) \times IPAVG \quad (2)$$

Several other authors have referred to IPAVG as hits per ball in play (HPBP) or batting average per batted ball.¹ I prefer IPAVG because it more closely ties with

"AVG", where it plays a major role, and yet is not really a "batting average" because it rules out the possibility of a strikeout.²

To prove that Equation 2 works, we can calculate Ted Williams' 1941 batting average, which was .406. In 1941, Ted Williams had 456 at-bats, struck out 27 times, and tallied 185 hits. He put the ball in play 429 times ($456 - 27$), not including sacrifices. Therefore, his $KAVG = K/AB = 27/456 = .059$ and his $IPAVG = H/BIP = 185/429 = .431$. According to Equation 2, his AVG was $(1 - .059) \times .431 = .941 \times .431 = .406$.

Before moving on, we should note that Equation 2 makes it clear why batting average by count is misleading. If batting average is calculated for all counts with less than two strikes, then $KAVG = 0$ because no at-bat ends in a strikeout with less than two strikes. Therefore, AVG with less than two strikes is simply equal to IPAVG. Consider the following two situations: (1) Barry Bonds had an at-bat with less than two strikes and (2) Barry Bonds had an at-bat with two strikes. In which case is it more likely Bonds got a hit? When he had less than two strikes because we are sure he did not strike out. Comparing AVG across two-strike and non-two strike counts is meaningless! Yet, STATs Inc continues to publish AVG by count.³ For more detail, please see Bickel and Stotz (2002).⁴

The same reasoning that we used to develop Equation 2 can be applied to SLG. In this case, SLG measures the batter's ability to avoid striking out and the number of bases he obtains when he puts the ball in play. This is shown in Equation 3.

$$SLG = (1 - KAVG) \times IPSLG \quad (3)$$

In Equation 3 we have called Bases/BIP the "In-Play Slugging Average" or IPSLG.

Equations 2 and 3 suggest many interesting questions. What are the average KAVG, IPAVG, and IPSLG for major league hitters? Has this changed over time? Has the change in AVG over the last 100 years been driven by changes in KAVG, IPAVG, or both? What

J. ERIC BICKEL is President of Competitive Edge Decision Systems, which provides pitch/hit-charting software to amateur and professional baseball teams.

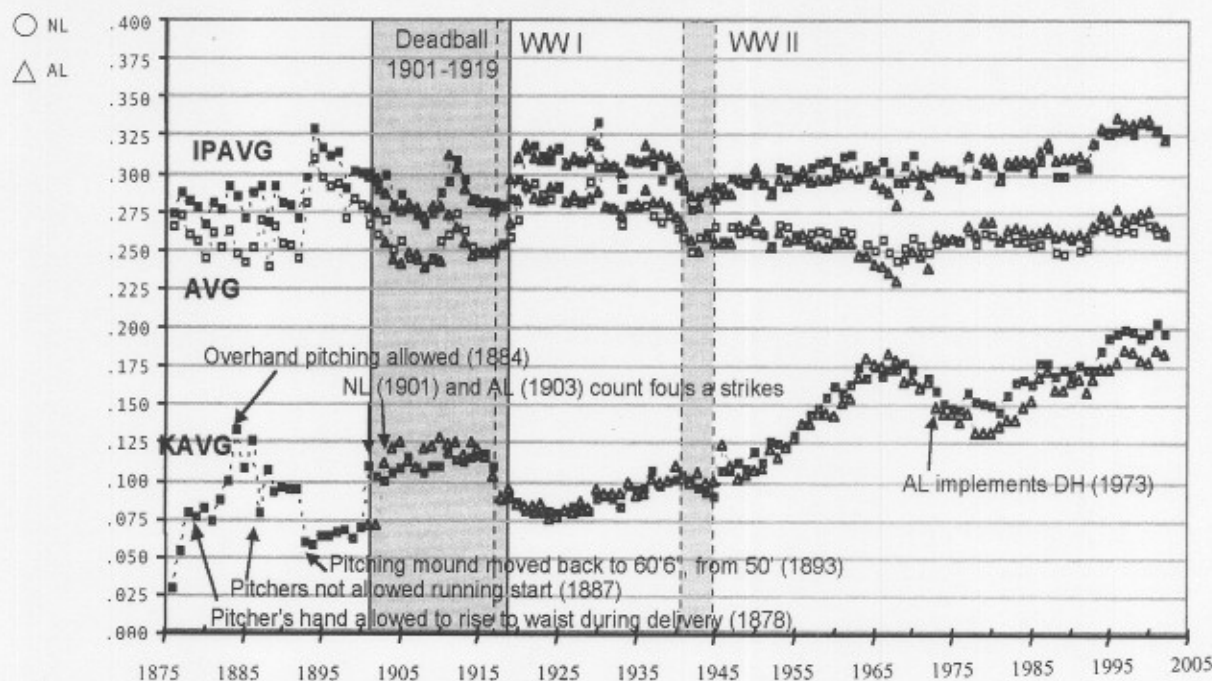


Figure 1. KAVG, IPAVG and AVG for the AL and NL since 1876.⁵

does it take to hit .400? Did Ted Williams hit .406 by avoiding strikeouts, or was he better at hitting them where they ain't? KAVG seems to be more under the control of the batter than IPAVG, in that IPAVG is driven by the structure of the game (e.g., the number and location of fielders, field dimensions, etc.) and, to some degree, just plain luck. Do some batters really have the ability to "place" the ball so that they get more hits when they put the ball in play? Can a major league hitter improve his AVG more by improving his KAVG or IPAVG? Do pitchers differ in the IPAVG and IPSLG that they allow?

In baseballprospectus.com, Voros McCracken reported that pitchers differ very little in their ability to prevent hits once the ball is put in play. In our words, they differ little in the IPAVG that they allow. Many other researchers have confirmed this effect.

KAVG AND IPAVG: A HISTORICAL PERSPECTIVE

Figure 1 displays KAVG, IPAVG, and AVG for the AL and NL since 1876. The lowest data set is KAVG. The circles represent the NL and the diamonds the AL. The middle data set is AVG for the NL and AL. And, finally, the top dataset is IPAVG for the NL and AL.

The Deadball Era (1901-1919) is clearly visible in the IPAVG data, with two dramatic declines—which would not be matched until World War II (WW II). It is interesting to note that strikeouts were relatively

high during this era. The KAVGs recorded during this time, would not be seen again until the mid-1950s. It is generally agreed that Deadball Era ended in 1920 with the banishment of the spitball and other "freak" pitches.⁶ However, KAVGs began to collapse as early as 1917, and IPAVGs began their dramatic increase in 1919 (mostly in the AL). These years coincide with America's involvement in WW I, but we are unsure as to the relative significance of WW I (e.g., loss of key players), the banishment of the spitball, and other factors.

IPAVG has generally been above .300 since the 1950s. NL and AL IPAVG data sets are similar because the rules governing balls put in play have been consistent between the two leagues. The AL gained a slight advantage over the NL with the introduction of the DH in 1973.

The trend of IPAVG has been generally upward since 1977 and is currently around .322 or almost 33%, with a dramatic increase starting in 1993. An IPAVG of about 33% means that the average MLB batter has about a one-third chance of getting a hit if he puts the ball in play. In a statistical sense, about two-thirds of the field is covered by the defense. Interestingly, 0.333 is very close to the IPAVG that Dean Stotz and I found in college baseball.⁷ Clearly, IPAVG is driven to a very large degree by the structure of the game (e.g., the shape of the field, the number of fielders, the distance to the outfield wall).

KAVG has been more variable and swung wildly between 1876 and 1901, as the NL tinkered with the number of balls allowed before a walk, the distance from the pitcher's mound to home, and whether or not

Table 1. Impact of some rule changes since 1876

SEASON	RULE CHANGE	KAVG	IPAVG	AVG
1876	Initially a strikeout effectively requires 4 strikes Initially a base on balls effectively requires 9 balls			
1878	Pitcher's hand must stay below waist during delivery, instead of below hip	+49%	-2%	-4%
1880	Number of balls required for a walk decreased to 8	+7%	-4%	-4%
1881	Pitcher's mound moved back from 45' to 50' Number of balls required for a walk decreased to 7 Number of strikes required for a strikeout reduced to 3 on a called third strike	-11%	+5%	+6%
1883	Pitcher's hand must stay below shoulder during delivery, instead of below waist Foul balls caught on one bounce no longer count as outs	+14%	+6%	+4%
1884	Overhand pitching allowed Number of balls required for a walk reduced to 6	+34%	-2%	-6%
1886	Number of balls required for a walk increased to 7	+17%	+6%	+4%
1887	Pitchers no longer allowed a running start, may only take one step Number of balls required for a walk reduced to 5	-37%	+2%	+7%
1888	Number of strikes required for a strikeout reduced to 3	+36%	-8%	-11%
1889	Number of balls required for a walk reduced to 4	-13%	9%	11%
1892	Outfield fence must be at least 235' from home plate	--	-3%	-3%
1893	Only round bats may be used Pitcher's mound moved back from 50' to 60'6", however, mound could be raised	-37%	+10%	+14%
1894	Foul bunts count as strikes for the first time	-2%	+10%	+11%
1895	Foul tips count as strikes if caught within the 10' catcher's box	+10%	-4%	-4%
1901	NL counts foul balls as strikes for the first time	+58%	--	-4%
1903	Mound height limited to 15" AL counts foul balls as strikes for the first time	-3%	+4%	+4%
		+55%	-3%	-7%
1911	Introduction of cork-centered baseball	+1%	+7%	+7%
1920	Foreign substances may not be applied to ball. Ball may not be deliberately scuffed	-6%	+5%	+5%
1926	Pitchers may use rosin bags Outfield fence must be at least 250' from home plate	+4%	-3%	-4%
1969	Mound height limited to 10" Strike zone reduced (armpits to top of knees)	-3%	+4%	+5%
1973	AL introduces the designated hitter	-10%	+6%	+9%

foul balls counted as strikes. Figure 1 highlights some of the rule changes. Table 1 presents a more exhaustive list of rule changes and what happened to KAVG, IPAVG and AVG.⁸

Changes to the pitcher's delivery, reducing the number of strikes required for a strikeout and counting foul balls as strikes had dramatic effects on strikeouts. We also see that IPAVG increased by 7% after the introduction of the cork-centered baseball in 1911. However, the Deadball Era resumed after this increase.

It is interesting that KAVG has doubled during the last 50 years; batters are twice as likely to have their AB end in a strikeout now than they were in 1945. This increase in KAVG has tended to hold batting averages in check.

HITTING .400

What does it take to hit .400? Equation 2 shows that at a minimum the batter must get a hit 40% of the time he puts the ball in play (IPAVG = .400). In this case, he will hit .400 if he never strikes out. If he strikes out in 10% of his ABs (KAVG = .100), he would need to have an IPAVG of .444 ($.4 \div [1 - .9]$)—or he would have to get a hit 44.4% of the time he put the ball in play.

A KAVG of .100 is quite low by today's standards, being about half of the current 0.190 average. Furthermore, an IPAVG of .444 is astronomical by any historical measure. Table 2 presents the top 20 IPAVGs

since 1913 (note: we have not included seasons prior to 1913 because individual strikeout data is unavailable in the AL until this time).⁹

Table 2 shows that IPAVG has been over .444 in only 11 player seasons since 1913 (about .12%). We also see that home run hitters have generated most of the top IPAVGs. This is understandable given that balls hit into the stands do not have the possibility of being caught. The record IPAVG of .478 belongs to Manny Ramirez. However, he "only" hit .351 because almost 27% of his ABs ended in strikeouts. Babe Ruth compiled five of the top 20 IPAVG seasons. Only three of the top 20 IPAVGs have resulted in AVGs over .400—all of them by Rogers Hornsby.

Table 3 runs down the top 20 batting averages since 1913. No batter has hit .400 or higher with an IPAVG below .420 (Ty Cobb, 1922). Furthermore, no .400 AVG has ever been accompanied by KAVG of greater than .080 (Rogers Hornsby, 1922). This is about one-third the current MLB average.

A COMPARISON OF WILLIAMS, BRETT AND GWYNN

Before leaving this historical look at IPAVGs and KAVGs, let's look at four more incredible player-seasons: Ted Williams 1941, Ted Williams 1957, George Brett 1980, and Tony Gwynn 1994.

In 1941, Ted Williams' KAVG was 42% lower and his IPAVG 45% greater than the league average (see

Table 2. Top 20 IPAVGs since 1913

PLAYER	LEAGUE	YEAR	AB	AVG	KAVG	IPAVG
Manny Ramirez	AL	2000	439	.351	.267	.478
Babe Ruth	AL	1923	522	.393	.178	.478
Babe Ruth	AL	1920	458	.376	.175	.455
Rogers Hornsby	NL	1924	536	.424	.060	.450
Jim Thome	AL	2001	526	.291	.352	.449
Jose Hernandez	NL	2002	525	.288	.358	.448
Babe Ruth	AL	1924	529	.378	.153	.446
Sammy Sosa	NL	2001	577	.328	.265	.446
Mo Vaughn	AL	1997	527	.315	.292	.445
Manny Ramirez	AL	1999	522	.333	.251	.445
Babe Ruth	AL	1921	540	.378	.150	.444
Sammy Sosa	NL	2000	604	.320	.278	.443
Larry Walker	NL	2001	497	.350	.207	.442
Mo Vaughn	AL	1998	609	.337	.236	.441
Babe Ruth	AL	1926	495	.372	.154	.439
Jeff Bagwell	NL	1994	400	.368	.163	.439
Andres Galarraga	NL	1993	470	.370	.155	.438
Rogers Hornsby	NL	1925	504	.403	.077	.437
Rogers Hornsby	NL	1922	623	.401	.080	.436
Lou Gehrig	AL	1927	584	.373	.144	.436

Table 3. Top 20 AVGs since 1913

PLAYER	LEAGUE	YEAR	AB	AVG	KAVG	IPAVG
Rogers Hornsby	NL	1924	536	.424	.060	.450
George Sisler	AL	1922	586	.420	.024	.430
George Sisler	AL	1920	631	.407	.030	.420
Ted Williams	AL	1941	456	.406	.059	.431
Rogers Hornsby	NL	1925	504	.403	.077	.437
Harry Heilmann	AL	1923	524	.403	.076	.436
Rogers Hornsby	NL	1922	623	.401	.080	.436
Bill Terry	NL	1930	633	.401	.052	.423
Ty Cobb	AL	1922	526	.401	.046	.420
Lefty O'Doul	NL	1929	638	.398	.030	.410
Harry Heilmann	AL	1927	505	.398	.032	.411
Rogers Hornsby	NL	1921	592	.397	.081	.432
Tony Gwynn	NL	1994	419	.394	.045	.413
Harry Heilmann	AL	1921	602	.394	.061	.419
Babe Ruth	AL	1923	522	.393	.178	.478
Harry Heilmann	AL	1925	573	.393	.047	.412
Babe Herman	NL	1930	614	.393	.091	.432
Ty Cobb	AL	1913	428	.390	.072	.421
Al Simmons	AL	1931	513	.390	.088	.427
George Brett	AL	1980	449	.390	.049	.410

Table 4. Comparison of Williams, Brett, and Gwynn

YEAR	PLAYER	INDIVIDUAL			LEAGUE			DIFFERENCE		
		AVG	KAVG	IPAVG	AVG	KAVG	IPAVG	AVG	KAVG	IPAVG
1941	Ted Williams	.406	.059	.431	.266	.102	.297	52%	-42%	45%
1957	Ted Williams	.388	.102	.432	.255	.138	.296	52%	-26%	46%
1980	George Brett	.390	.049	.410	.269	.133	.310	45%	-63%	32%
1994	Tony Gwynn	.394	.045	.413	.270	.184	.327	46%	-75%	26%

Table 4). In 1957 his IPAVG was even higher than 1941, but he struck out almost twice as often as 1941. George Brett and Tony Gwynn had incredibly low KAVGs—less than .050, or about 70% lower than the league average (much better than Ted Williams). However, their IPAVGs were too low to break the .400 barrier.

AN EXPLOSION IN OFFENSE

IPAVG has increased dramatically since 1993. As we know, this increase has been driven by a major increase in home runs. Figure 2 makes this abundantly clear.

Several interesting trends appear in Figure 2. First, the fraction of balls put in play that have become singles (1BIP) has remained quite stable since 1946. Next, the likelihood of a triple (3BIP) has decreased substantially since 1876. This must be driven in large part to changes in field dimensions. And, most strikingly, the

likelihood that a ball put in play becomes a home run (HRIP) increased dramatically since the middle of the century and has exploded since 1993.

Many of these trends become apparent by looking at averages over several decades, as shown in Table 5. Table 6 presents the change in the average to 2002.

The chance of a single, given the ball has been put in play, has only increased by 1% from the 1876-1886 average (.217 vs .219). This should not be surprising given that the factors that determine whether a ball put in play becomes a single (field dimensions, location of fielders, speed of batted ball) have not changed appreciably, if at all, since 1876. On the other hand, the likelihood of doubles, triples, home runs, and all extra base hits (Extra H) has changed dramatically over the years. Likewise, IPSLG has seen significant increases.

Table 5. Hit type IPAVG and IPSLG by decade (NL and AL)

DECADE	1B	2B	3B	HR	EXTRA H	IPSLG
1876-1885	0.217	0.043	0.014	0.004	0.061	0.363
1886-1895	0.221	0.042	0.018	0.009	0.070	0.397
1896-1905	0.233	0.040	0.017	0.006	0.063	0.387
1906-1915	0.224	0.040	0.016	0.005	0.061	0.371
1916-1925	0.228	0.047	0.016	0.009	0.072	0.405
1926-1935	0.223	0.056	0.014	0.015	0.085	0.438
1936-1945	0.218	0.051	0.011	0.016	0.078	0.419
1946-1955	0.213	0.047	0.010	0.024	0.082	0.434
1956-1965	0.214	0.047	0.009	0.031	0.087	0.459
1966-1975	0.218	0.046	0.008	0.026	0.080	0.439
1976-1985	0.219	0.051	0.008	0.026	0.086	0.451
1986-1995	0.219	0.057	0.007	0.031	0.095	0.477
1996-2002	0.219	0.064	0.007	0.039	0.111	0.525

Table 6. Change in IPAVG and IPSLG since 1876 (NL and AL)

YEARS	1B	2B	3B	HR	EXTRA H	IPSLG
1876-2002	1%	50%	-52%	824%	80%	45%
1886-2002	-1%	53%	-62%	320%	59%	32%
1896-2002	-6%	60%	-59%	570%	76%	36%
1906-2002	-2%	61%	-57%	678%	81%	41%
1916-2002	-4%	38%	-56%	323%	54%	30%
1926-2002	-2%	16%	-51%	156%	30%	20%
1936-2002	1%	27%	-39%	139%	41%	26%
1946-2002	3%	36%	-33%	63%	35%	21%
1956-2002	2%	37%	-24%	27%	27%	14%
1966-2002	0%	41%	-13%	49%	38%	20%
1976-2002	0%	26%	-19%	51%	29%	16%
1986-2002	0%	14%	-6%	27%	17%	10%

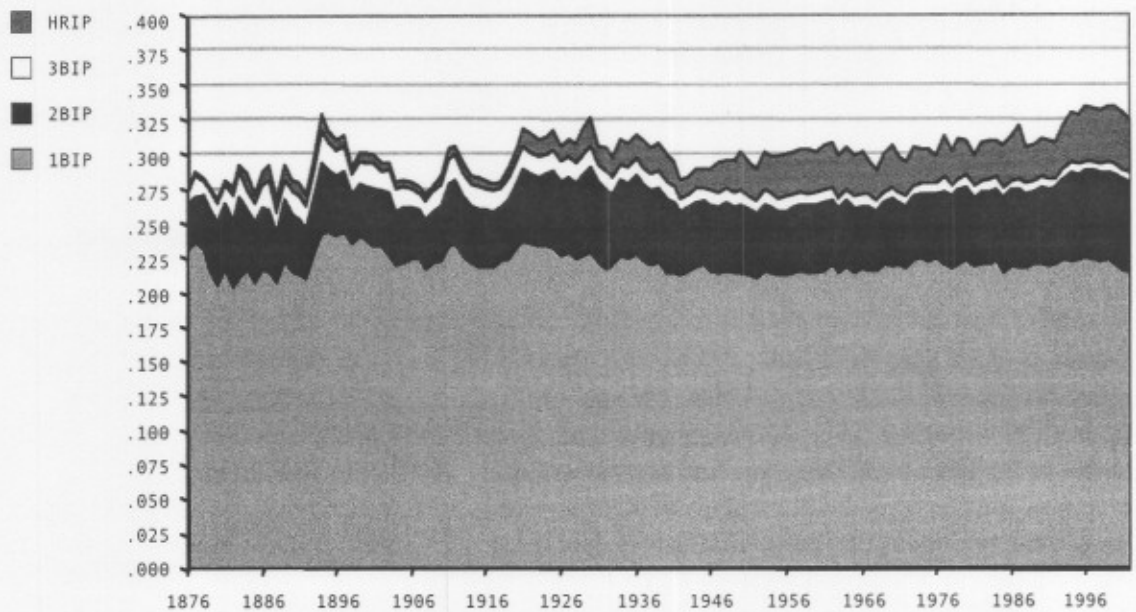


Figure 2. Hit type KAVG, and IPAVG by decade (NL and AL)

PLACING BASE HITS

Do batters have the ability to place base hits? This is a question worthy of more space than we can provide here. However, a quick look suggests that, yes, the ability to place base hits is a skill. Table 7 presents the highest and lowest IPAVGs (in total and by hit type) and KAVG for the 2002 season (only players eligible for a batting title were considered).

Jose Hernandez led the majors with a .448 IPAVG (1.33 times the league average), while Neifi Perez's IPAVG was only .261 (0.77 times the league average). Even if we exclude home runs, and thereby do not include power hitters, we find a significant difference between player performance—Jose Hernandez's .377 versus Jay Gibbons' .219. Not surprisingly, the difference between the best and the worst is even greater for other hit types.

We also see a significant difference in terms of KAVG. Jose Hernandez led the majors with a .358 KAVG in 2002. That's right, the same Jose Hernandez that led the league in IPAVG! 35.8% of Jose's at-bats resulted in strikeouts, but if he put the ball in play he got a hit 44.8% of the time. That is the sixth highest IPAVG since 1913 (see Table 2). Jason Kendall had the lowest KAVG with only 5.3% (.053) of his ABs ending in strikeouts.

Table 7. Highest and lowest IPAVGs and KAVG in 2002

STAT	PLAYER	INDIVIDUAL	RELATIVE TO MLB
IPAVG	Jose Hernandez	0.448	1.33
	Neifi Perez	0.261	0.77
IPAVG w/o HRs	Jose Hernandez	0.377	1.30
	Jay Gibbons	0.219	0.75
1BIP	Luis Castillo	0.302	1.40
	Jay Gibbons	0.149	0.69
2BIP	Bobby Abreu	0.110	1.61
	Luis Castillo	0.034	0.50
3BIP	Brad Wilkerson	0.023	3.50
	(20 Players)	0.000	0.00
HRIP	Jim Thome	0.152	3.23
	Fernando Vina	0.002	0.04
KAVG	Jose Hernandez	0.358	2.05
	Jason Kendall	0.053	0.30

As you can see, batters appear to differ to a significant degree in their ability to get a hit once they put the ball in play and their ability to avoid strikeouts. However, the spread among players is greater for KAVG than it is for IPAVG. That is, players differ to a greater degree in their ability to avoid striking out than in their ability to get a hit once they put the ball in play.

Further research should focus on whether or not a player's performance is consistent over time.

CONCLUSION

Batting average measures the batter's ability to avoid striking out and his ability to "hit 'em where they ain't." By rewriting AVG as a function of these two effects, we gain greater insight into what has driven changes in AVG over the last hundred years, why hitting .400 is so difficult, and whether batters can really place hits.

ACKNOWLEDGMENTS

The author thanks Jim Charlton and three anonymous reviewers for many helpful suggestions.

NOTES

1. Dean Stotz and I referred to this as IPA in our 2002 BRJ article (Bickel, J. Eric, and Dean Stotz, "Batting Average by Count and Pitch Type: Fact and Fallacy," *The Baseball Research Journal*, No. 31, 2003, pp. 29-34). In that same issue, Dick Cramer referred to this concept as "batting average per batted ball" (Cramer, Dick, "Preventing Base Hits: Evidence That Fielders Are More Important Than Pitchers," *The Baseball Research Journal*, No. 31, 2003, pp. 88-92). See also, Birnbaum, Phil, "Factors Affecting Pitcher Ball-in-Play Average," *By the Numbers*, May 2001, pp. 8-11.
2. John Burnson developed equation 1 independently in *Ron Shandler's 2003 Baseball Forecaster* (pp. 7-8). Stanford Baseball has been calculating IPAVG since 1998, and this stat has been publicly available since 2000.
3. In addition, Jim Albert and Jay Bennett analyze AVG by count in their book *Curve Ball* (see pp. 87-107). They note, "Players who have small pitch count effects are relatively unlikely to strike out." This is expected because AVG for batters that do not strike out a lot is almost equal to IPAVG, which is not as dependent on the count as KAVG.
4. Bickel, J. Eric and Dean Stotz, *op cit*.
5. This data was obtained from the "The Lahman Baseball Database" (www.baseball1.com). Individual player data on strikeouts is unavailable from 1897 to 1909 in the NL and 1901 to 1912 in the AL. However, league total strikeout data was obtained from *Total Baseball*.
6. Steinberg, Steve L., "The Spitball and the End of the Deadball Era," *The National Pastime*, No. 23, 2003, pp. 7-17.
7. Bickel, J. Eric, and Dean Stotz, *op cit*.
8. These rule changes were obtained from David Nemec's outstanding book *The Rules of Baseball* (1994, Lyons and Burford Publishers) and *Total Baseball*.
9. Only player seasons for which the batter had enough plate appearances to be considered for a batting championship were included.