The Changing Hitting Performance Profile In the Major League, 1996-2006

by

Paul M. Sommers

September 2007

MIDDLEBURY COLLEGE ECONOMICS DISCUSSION PAPER NO. 07-15

DEPARTMENT OF ECONOMICS
MIDDLEBURY COLLEGE
MIDDLEBURY, VERMONT 05753

http://www.middlebury.edu/~econ

THE CHANGING HITTING PERFORMANCE PROFILE IN MAJOR LEAGUE BASEBALL, 1966-2006

by

Paul M. Sommers

Department of Economics
Middlebury College
Middlebury, Vermont 05753
psommers@middlebury.edu

Abstract

The career batting profile of a regular starting major league ballplayer typically rises, at least up to a point, and then falls as skills diminish with age. The career batting profiles are derived for all regular starting players in the National and American Leagues for each of five different years: 1966, 1976, 1986, 1996, and 2006. The profiles have changed dramatically since the 1960s, with arguably stronger ballplayers reaching a higher peak several years after the batting average reached a peak for regulars in 1966. The profiles for 2006 show what might be early manifestations of baseball's tougher steroids policy.

Keywords: Major League Baseball, career profile, batting average, steroids

THE CHANGING HITTING PERFORMANCE PROFILE IN MAJOR LEAGUE BASEBALL, 1966-2006

For most professional athletes, productivity increases rather quickly, tops off, and then declines as skills diminish or health declines with age. Career length and the point where productivity "tops off" varies with one's sport. And, even within a given sport, the point where one reaches his or her prime may not be well-defined.

In 1989, Gerald W. Scully [1, p. 47] endeavored to show how a representative ballplayer's batting average varied with years in the major leagues. To do this, Scully [1, p. 200] regressed the batting average of a regular starting player in the National League (hereafter NL) against his career batting average (as of 1986), years in the majors, and years squared for all 96 players (eight regulars on each of the NL's then twelve teams) with a total of "564 years of playing experience." For a hitter (like Steve Garvey) with, say, 16 years of playing experience (as of 1986), there would be 16 observations (of the 564 total) for that particular player. Scully showed, not surprisingly, that for a regular player with a career batting average of .267, "[t]he batting average rises at a diminishing rate for several seasons and reaches a peak between the sixth and seventh year. Thereafter, the batting average declines." [1, p. 47] Scully did not indicate the minimum number of at-bats that were required for a "career year". Nor did Scully show the corresponding profile for players in the American League (AL) in 1986. Moreover, one might wonder whether the profile for players in the AL changes when designated hitters (typically players with many career years) are excluded. Did the performance profile in 1986 change from what it was one or two decades earlier? And, how has this profile changed (if at all) since 1986, during the so-called "enhancement era"? Finally, what effect (if any) has baseball's tougher steroids policy (beginning in 2004) had on performance profiles? This brief research

note examines the hitting performance profile in each league for five selected years, one each in five different decades. The results show a dramatic profile change for both leagues since the 1960s and, for one league, a sharp departure in 2006 that looks more like the profile in 1966.

The Model

Following the Scully approach, a ballplayer's batting average in year t [BA_t] for each of his n years in the majors with a minimum of 100 at bats per season was regressed against career year (which for a representative player varies from 1 to n) [$Year_t$], career years squared [$Years_t^2$], and his lifetime batting average [*Lifetime_BA*] as of the selected year (1966, 1976, 1986, 1996, or 2006), as follows:

$$(1) \qquad BA_t = \beta_0 + \beta_1 Year_t + \beta_2 Year_t^2 + \beta_3 Lifetime\_BA + \varepsilon_t$$

where ε_t denotes a stochastic disturbance (or error) term which may take on positive or negative values. As Scully notes, the inclusion of *Lifetime_BA* adjusts the profile for individual differences in hitting performance among the players. If a ballplayer's batting average rises with career year and then falls after a point, then b_1, the least squares estimate for β_1, should be positive and b_2, the least squares estimate for β_2, should be negative.[1] The peak point is found by taking the partial derivative of BA_t with respect to $Year_t$, setting this derivative equal to zero, and solving for $Year_t$ in terms of b_1 and b_2. That is,

$$(2) \qquad \frac{\partial BA_t}{\partial Year} = b_1 + 2b_2 Year_t = 0$$

or $Year^* = -b_1/2b_2$, where $Year^*$ denotes the career year where batting average peaks. All performance data on regulars in the years 1966, 1976, and 1986 are from *The Baseball*

Encyclopedia [2]; all corresponding data on regulars in 1996 and 2006 are from http://www.baseball-reference.com/teams/ .

The Results

Table 1 reports the average number of (minimum 100 at-bat) seasons for all regular starting players in both leagues in each of the five selected years. The 2006 average is discernibly different from the 1966 average for both leagues [$p = .023$ for the NL; $p = .002$ for the AL (with designated hitters)]. The 1966 AL average is also significantly smaller than the AL averages (with the DH) in 1986 ($p < .001$) and 1996 ($p = .011$). In each of the five selected years, there were no discernible differences between league averages. AL averages, however, are uniformly smaller when designated hitters (typically, veterans with many years of playing experience) are excluded.

For each selected year in each of the last five decades, Table 2 gives the estimated coefficient on $Year_t$, $Year_t^2$, and the derived value for $Year^*$, that is, the point in a regular starting player's career (with a .270 lifetime BA) where batting average peaks. For each of the five regressions in each league (including designated hitters in the AL as of the regression for 1976), the coefficient of determination (R^2) and the total number of years of playing experience or that cohort's sample size (N) are also reported in the last two columns of the table. In each regression, the coefficients on all variables (including *Lifetime_BA*) were significant at better than the .01 level (with the exception of the regression for 1966 in the AL, where all coefficients are significant at better than the .05 level). The peak year values tend to run longer than Scully's estimate for NL ballplayers in 1986 of between six and seven years.

A more revealing presentation of the summary results in Table 2 are shown in Figures 1 and 2 for the NL and AL, respectively. These two figures show the evolution of the batting average of a representative major league player (with a lifetime .270 BA) in each of five years, one each in the last five decades. In both leagues (during the 1970s, '80s, and '90s), the negative

quadratic term in the regression that forces the inverted U-shaped profile down in the twilight of a ballplayer's career has diminished since 1966. Why? Off-season training and better conditioning, not to mention the improvements in sports medicine and physical therapy, could explain the flatter profile beyond the peak point. But, there are more sinister explanations, like the (alleged widespread) use of performance-enhancing substances. In both leagues? Figure 2 for the AL (in sharp contrast to Figure 1 for the NL) shows a dramatic precipitous decline from the peak point for regular starting players in 2006 (compared to earlier years).

Concluding Remarks

The batting average profile of a representative regular starting player is derived for all major leaguers in five selected years over the last five decades. In the 1970s, '80s, and '90s, regulars from both leagues peaked later in their career than did their counterparts in the 1960s. Moreover, the rate of decline after the peak point diminished for major leaguers after the 1960s, with one notable exception: American League regulars in 2006. While there may be many plausible reasons why batting averages continue to rise longer and why performance profiles are flatter (and, in some instances, even higher) beyond the peak point than they were in the 1960s, one explanation might be baseball's lax (that is, no official) drug policy before 2004. The performance profile for American Leaguers in 2006 might be an early manifestation of baseball's now tougher steroids policy.

Table 1. Average Number of (Minimum 100 At-Bat) Seasons for All Regular Starting Players, 1966-2006 (Selected Years)

	National League	*American League*	
Year		With DH	Without DH
1966	5.238	*	4.838
1976	6.125	5.741	5.260
1986	5.948	7.000	6.446
1996	6.009	6.191	5.625
2006	6.555	6.405	5.964

Table 2. Regression Results and Regular Starting Players' Peak Year, by League, 1966-2006 (Selected Years)

National League

Season	Coefficient on Year	Coefficient on $Year^2$	Peak Year*	R^2	N
1966	.00642	-.000497	6.46	.54	419
1976	.00411	-.000215	9.62	.49	587
1986	.00338	-.000232	7.29	.44	571
1996	.00590	-.000330	8.93	.42	674
2006	.00223	-.000142	7.87	.39	839

American League

Season	Coefficient on Year	Coefficient on $Year^2$	Peak Year*	R^2	N
1966	.00218	-.000237	4.61	.49	387
1976	.00450	-.000301	7.49	.50	620
1986	.00451	-.000307	7.34	.45	881
1996	.00431	-.000223	9.68	.45	779
2006	.00698	-.000526	6.63	.41	807

Figure 1

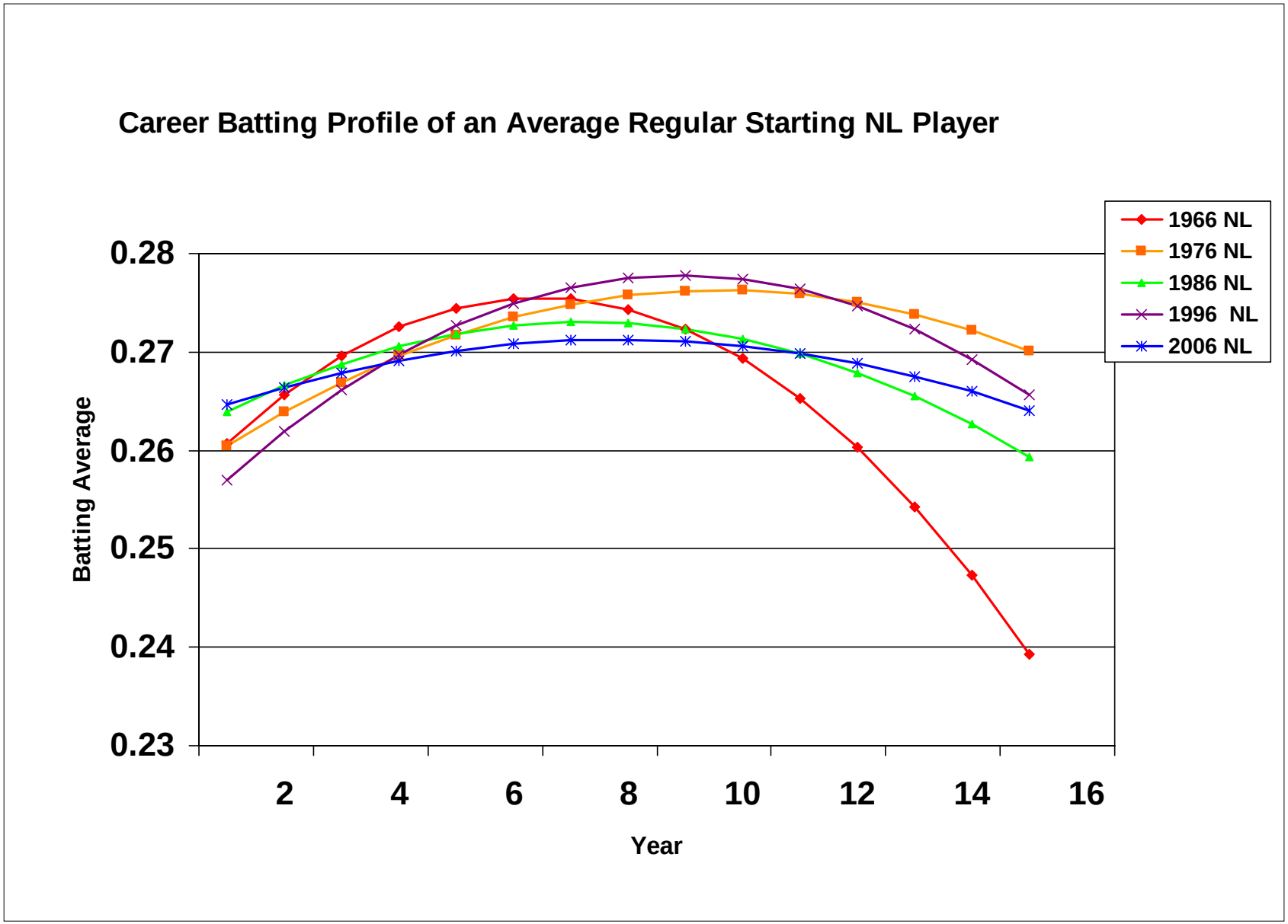
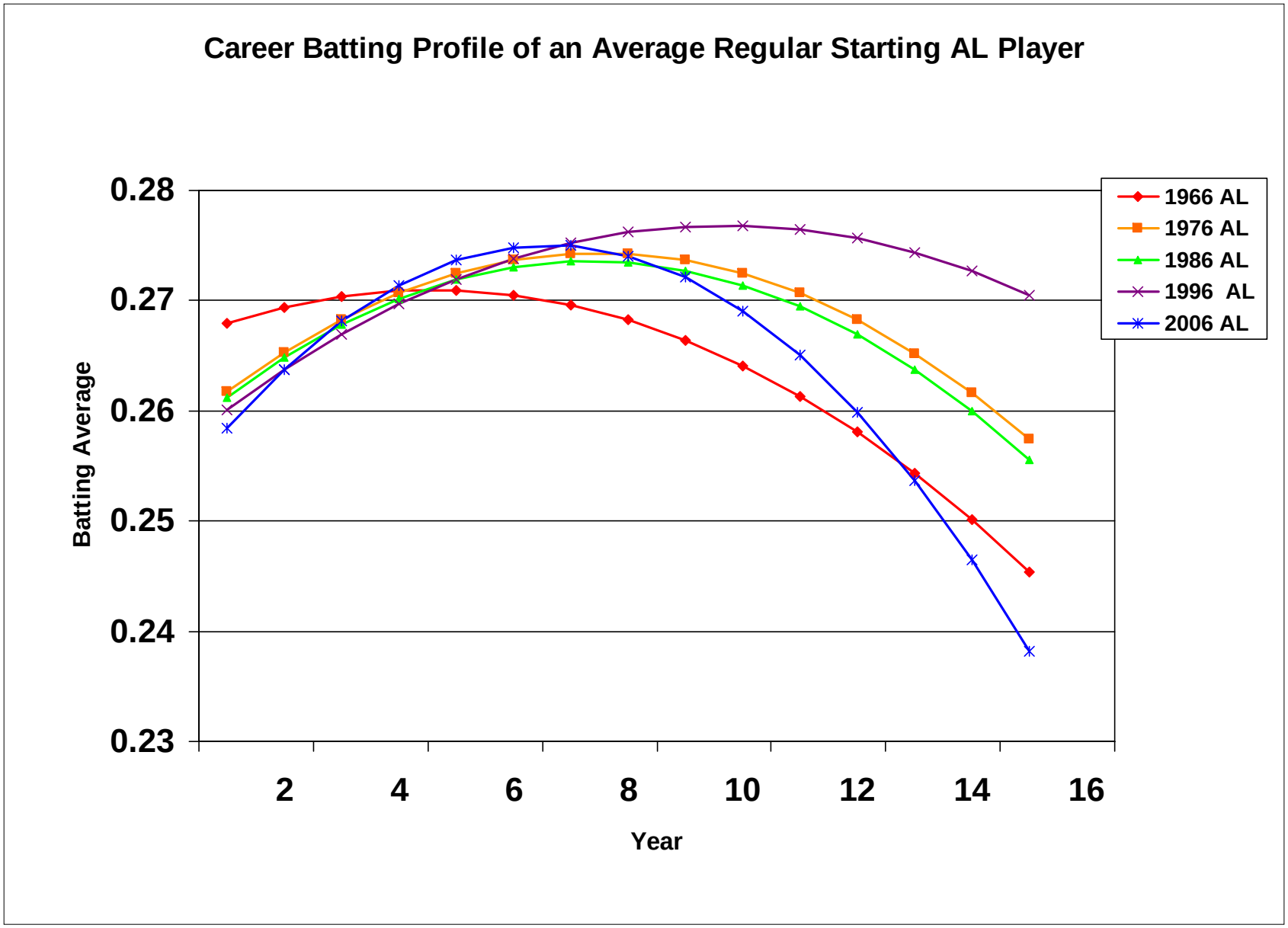

Career Batting Profile of an Average Regular Starting NL Player
Batting Average
0.28
0.27
0.26
0.25
0.24
0.23
2
4
6
8
10
12
14
16
Year
1966 NL
1976 NL
1986 NL
1996 NL
2006 NL

Figure 2


Career Batting Profile of an Average Regular Starting AL Player
0.28
0.27
0.26
0.25
0.24
0.23
Batting Average
2
4
6
8
10
12
14
16
Year
1966 AL
1976 AL
1986 AL
1996 AL
2006 AL

References

1. G. W. Scully, *The Business of Baseball* (Chicago, Illinois: The University of Chicago Press, 1989).

2. *The Baseball Encyclopedia* (New York: Macmillan Publishing Company, 1990).

Footnote

1. The coefficient on *Lifetime_BA* should be close to 1.0. More often than not, b_3, the least squares estimate for β_3, will be less than 1.0. But, since seasons with fewer than 100 at bats (usually in a rookie's first year or two in the majors or in years when the ballplayer is injured) tend to result in uncharacteristically low batting averages and hence lower the player's lifetime batting average, the estimated coefficient b_3 could be marginally greater than 1.0 (as it was in two of the ten regressions, both for the AL, in 1966 and 1976).