Journal of Sports Economics
http://jse.sagepub.com

Concentration of Playing Talent: Evolution in Major League Baseball

Martin B. Schmidt and David J. Berri

Journal of Sports Economics 2005; 6; 412
DOI: 10.1177/1527002504268613

Concentration of Playing Talent

Evolution in Major League Baseball

MARTIN B. SCHMIDT
College of William and Mary

DAVID J. BERRI
California State University, Bakersfield

Several authors have recently suggested that an expanding labor pool has led to improvement in professional sports leagues' competitive balance. The basic premise is that a rise in team player options leads to less variability in player performance and therefore increased competition. The present work examines the initial step (i.e., the relationship between the influx of foreign-born players and various measures of talent compression). The results suggest that the geographic diversity of today's baseball players has reduced variability in individual player performance.

Keywords: *evolution; competitive balance; cotrending*

EVOLUTION AND MAJOR LEAGUE BASEBALL

In an effort to illustrate the nature of biological evolution, Stephen Jay Gould (1983) examined the disappearance of the .400-hitter in Major League Baseball (MLB). Gould began by noting that the distribution of athletic talent in a given population should follow a normal distribution. Furthermore, given the physical limitations of the human body, there exists a biomechanical limit to athletic ability. Consequently, athletes located at the far right tail of the distribution would tend to be fairly close to the biological limit and of relatively equal skill.

At the onset of the 20th century, participation in MLB was confined to White Americans, originating primarily in the Northeastern states. In which case, although a few athletes likely approached the limit of human ability, baseball employed a number of athletes whose talent was located much further from the right tail of the distribution. Given this diversity in ability, those of greater talent could offer performances far above the average player's. Hence, hitters could achieve the .400 plateau.

JOURNAL OF SPORTS ECONOMICS, Vol. 6 No. 4, November 2005 412–419
DOI: 10.1177/1527002504268613

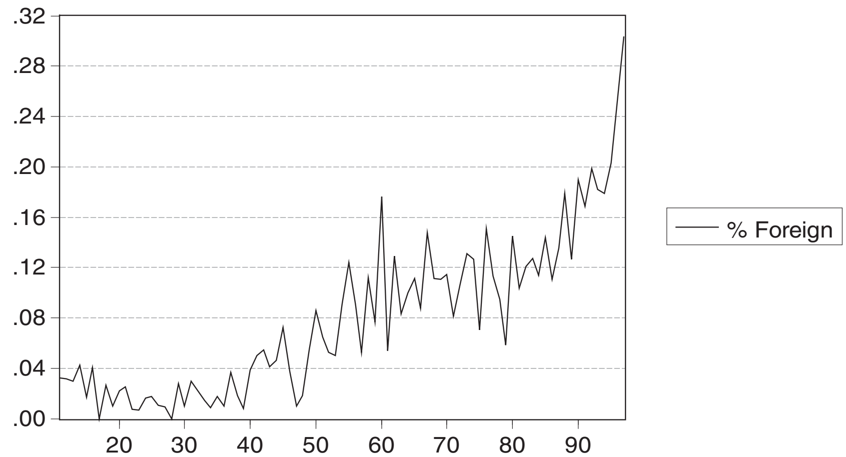

Figure 1: Measures of Player Diversification

Following the theme, as baseball's labor pool expanded, through migration and integration, diversity in skill level declined.[1] A number of recent empirical studies have noted that baseball became increasingly competitive toward the close of the 20th century (Butler, 1995; Horowitz, 1997; Schmidt & Berri, 2001). Such an increase seems a natural extension of Gould's (1996) hypothesis, as teams are stocked with players more similar in talent, the probability of a weak team bettering a strong team should rise (Chatterjee & Yilmaz, 1991).

The present work examines the intermediate step in the process (i.e., what is the relationship between the distribution of playing talent in MLB and expansions in the population of baseball talent?). Specifically, have racial integration and an expanding global search for talent led to less diversity in player performance within the MLB?

DATA AND EMPIRICAL METHODOLOGY

To further examine Gould's (1996) hypothesis, we require a measure of baseball's labor pool. As a proxy, Figure 1 reports the percentage of MLB players who were born outside the United States.[2] The assumption is that as MLB increased its search for players, or as it expanded its reach, geographical diversity should have been seen within the player pool. The figure highlights the fact that baseball has become more geographically diverse, particularly since the early 1950s and the Latin American explosion starting in the mid-1980s (see chapter 1 of Marcano Guevara & Fidler, 2003).

Our measurement of talent compression follows the work of Depken (2000, 2002), who offers idealized Herfindahl-Hirschman Index (HHI_t) scores for home

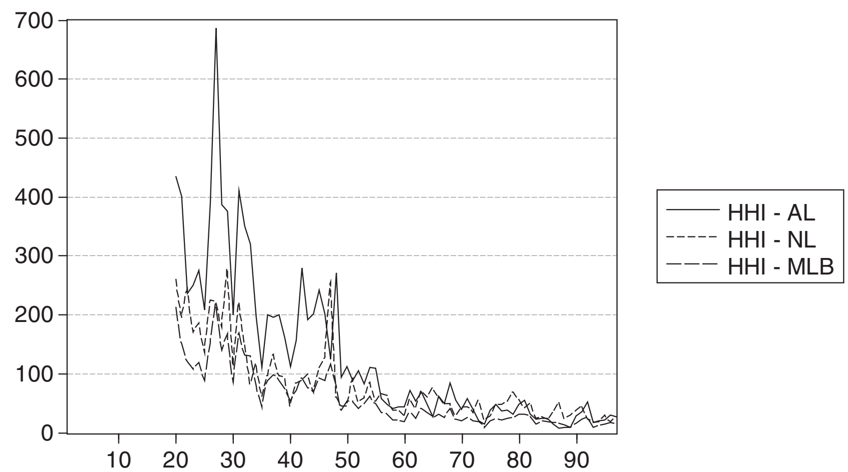

Figure 2a: HHI_t–Measure of Concentration (*Home Runs HHI-HR*$_t$)

NOTE: HHI = Herfindahl-Hirschman Index; AL = American League; NL = National League; MLB = Major League Baseball.

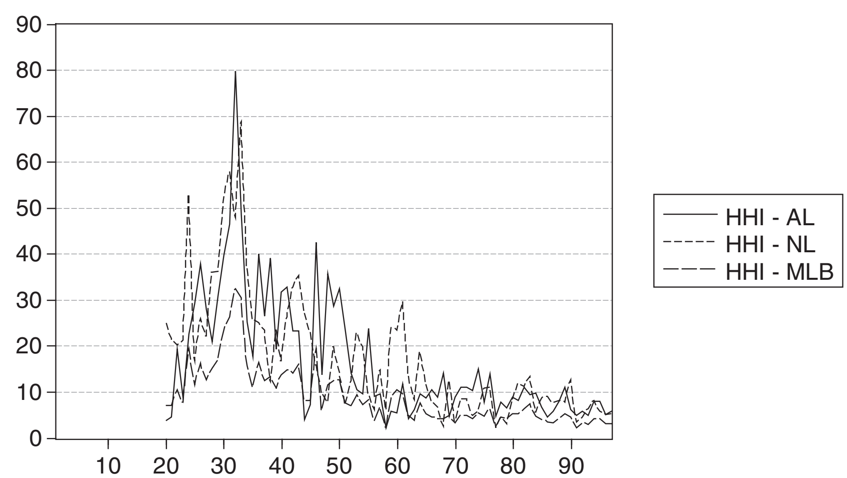

Figure 2b: HHI_t—Measure of Concentration (*Strike Outs HHI-K*$_t$)

NOTE = HHI = Herfindahl-Hirschman Index; AL = American League; NL = National League; MLB = Major League Baseball.

runs and strikeouts as measures of talent concentration within both the American League (AL) and National League (NL), as well as MLB. These are reported in Figures 2a and 2b.

A Nonlinear Explanation

Cursory examination of these figures suggests that each contains a nonlinear trend. On a theoretical level, such a trend is suggested by Gould's (1996) hypothesis. Specifically, Gould argues that the distribution of athletic talent in a population should be normally distributed. At the right tail of the distribution would lie those with the highest level of athletic ability. Assuming that there is a biomechanical limit, or a slowly moving limit, to potential ability or talent, the athletes in the far right tail tend to be relatively equal.[3] At the beginning of the 20th century, when the people playing MLB were only White, Northeastern American males, the population baseball could draw on was relatively small, and correspondingly, there existed a large degree of heterogeneity between players. As the probability of winning is closely aligned with playing talent, such diversity may lead to low levels of competitive balance.

As the population of players that MLB has to choose from rises, new players are added to the population in much the same way (i.e., normally). In which case, the absolute number of players close to the limit would rise and, given player demand, so would the average player's talent level or ability. The Gould hypothesis, therefore, argues that as the talent pool rises, greater player homogeneity should be observed.

The Gould (1996) hypothesis however makes a much stronger prediction: It predicts that the trend in competitive balance should be nonlinear. Specifically, the nonlinear trend should follow because although population increases raise the talent levels of both below average and above average players, below average players are farther away from the biomechanical limit and would approach the limit at a faster pace than above average players. In which case, given player demand, below average players would approach above average players asymptotically.

A Nonlinear Unit Root Test

On an empirical level, Bierens (1997) proposes a test for nonlinear trends.[4] Specifically, Bierens extended the nonlinear unit root tests of Ouliaris, Park, Phillips, and Peter (1989) to derive tests for nonlinear trend stationary. Bierens's test considers nonlinear trend stationary as the alternative to a unit root (with a possible drift). Under the null hypothesis the time series, X, may be written as follows:

$$Ho: X_t = \mu + X_{t-1} + \varepsilon_t \qquad (1)$$

where μ is a constant and u is a stationary autoregressive process. The alternative is of the form

$$Ha: X_t = \mu + g(t) + \varepsilon_t \qquad (2)$$

TABLE 1: Bierens's Nonlinear Unit Root Tests

	t(m)	SE	A(m)	SE	F(m)	SE	*Unit Root*
$HHI\text{-}HR_t$							
American League	–4.148*	.082	–35.725*	.025	5.895*	.910	No
National League	–3.820*	.095	–28.995*	.089	5.024*	.910	No
Major League Baseball	–3.810*	.089	–28.631*	.064	5.076*	.887	No
$HHI\text{-}K_t$							
American League	–4.352*	.046	–37.801*	.016	7.001*	.966	No
National League	–3.585	.344	–34.282*	.069	4.700*	.629	No
Major League Baseball	–3.686*	.083	–28.374*	.059	5.180*	.923	No
% foreign born	–2.884	.258	–28.177*	.071	4.823*	.918	No

NOTE: p values are based on $m = 2$ and 5,000 simulations. $t(m)$ are variations of the test reported in Equation 3. Values with an asterisk are significant at the 5% critical level.

where $g(t)$ is a possibly nonlinear trend function. The test statistic is based on the following Augmented Dickey-Fuller–type regression:

$$\Delta X_t = \alpha * X_{t-1} + \sum_{i=1}^{m} \lambda_i * P_{i,t} + \sum_{j=1}^{p} \theta_j * \Delta X_{t-j} + \varepsilon_t \tag{3}$$

where $P_{i,t} = (P_{0,t}, P_{1,t}, \ldots P_{m,t})$ and the individual $P_{i,T}$ are Chebyshev time polynomial orthoganlized on t.[5] In which case, although $P_{0,t}$ captures the mean and equals 1, $P_{1,t}$ is equivalent to the linear trend, and $P_{2,t}$ through $P_{m,t}$ are cosine functions (Bierens, 1997, pp. 31-32). Under the null hypothesis, α and all but the first element of λ are 0. In which case, either the t statistic of α in an OLS regression of Equation 3 or

$$A(m) = \frac{T * \alpha}{\left| \left(1 - \sum_{j=1}^{p} \theta_j \right) \right|} \tag{4}$$

can be used as test statistics. Although these two tests provide unit root tests based on $\alpha = 0$, neither test takes into account that of the possibility, under stationarity, that all the elements save one of γ are 0. Bierens (1997) also provides an F-test version that takes this into account.

The results of these tests are reported in Table 1. In general, these tests reject the unit root null for all series and, therefore, suggest that these are stationary around a nonlinear trend. Although this result is consistent with Gould's (1996) hypothesis, it is also consistent, as Bierens (1997) points out, with the series having a number of trend breaks. We therefore move next to examine whether the trends within the HHI indexes and MLB's labor pool are related (i.e., whether they, in the sense of Bierens, 2000, cotrend).

TABLE 2: Bierens's Nonlinear Cotrending Test

	r	*Test Statistic*	*5% Critical Region*	*10% Critical Region*	*Cotrending Vector*
American League ($HHI\text{-}HR_t$)	1	.2481*	> .4658	> .3518	–2,030.25 ×
and % foreign born	2	1.2172	> .6742	> .5356	% foreign born
National League ($HHI\text{-}HR_t$)	1	0.2534*	> .4658	> .3518	–944.689 ×
and % foreign born	2	1.2214	> .6742	> .5356	% foreign born
Major League ($HHI\text{-}HR_t$)	1	0.2855*	> .4658	> .3518	–778.852 ×
and % foreign born	2	1.2137	> .6742	> .5356	% foreign born
American League ($HHI\text{-}K_t$)	1	0.1221*	> .4658	> .3518	–164.556 ×
and % foreign born	2	1.3796	> .6742	> .5356	% foreign born
National League ($HHI\text{-}K_t$)	1	0.1227*	> .4658	> .3518	–182.185 ×
and % foreign born	2	1.2315	> .6742	> .5356	% foreign born
Major League ($HHI\text{-}K_t$)	1	0.1319*	> .4658	> .3518	–86.185 ×
and % foreign born	2	1.3104	> .6742	> .5356	% foreign born

NOTE: HHI = Herfindahl-Hirschman Index; *HR* = home runs; *K* = strikeouts.
* indicates significance at the 5% level.

A Nonlinear Cotrending Relationship

Bierens (2000) extended his earlier nonlinear tests to determine whether groups of nonlinear trend stationary variables possess similar nonlinear trends (i.e., whether they are cotrending variables). Cotrending, as defined by Bierens, is as follows:

> Given the assumption that the components of a vector time series are stationary about nonlinear deterministic time trends, nonlinear co-trending is the phenomenon that one or more linear combinations of the time series are stationary about a linear trend, hence the series have common nonlinear deterministic time trends. (p. 323)

The Bierens (2000) cotrending test is nonparametric, and therefore, there is no need to specify the nonlinear trends or any process for serial correlation. Thus, we need only create a vector, z, which contains one of the competitive balance measures and our labor pool measure. The corresponding cotrending test statistics examine the H_0: r nonlinear common trends against the alternative of $r - 1$ nonlinear common trends. The test statistics and their 5% and 10% critical levels are reported in Table 2.

The results are consistent with the Gould (1996) hypothesis, as each concentration measure and our labor pool variable are found to be cotrending (i.e., the tests suggest that one cotrending relationship exists between our competitive balance and labor pool measures). Furthermore, the estimated cotrending vectors, which are reported in the final column of Table 2, return the expected negative relationship, thereby suggesting that as geographical diversity rises, greater player homogeneity is obtained.

CONCLUDING OBSERVATIONS

The reported results indicate that the compression of talent across time responds to changes in the geographic diversity of players. Such results are consistent with prior work connecting changes in the population of players to the competitive balance of MLB. The present results coupled with these findings suggest that MLB should seek to promote the game of baseball both at home and abroad. The pursuit of this policy should further expand the population of available baseball players, hence, further improving the competitive balance of baseball in the future.

NOTES

1. Gould's (1996) argument was also largely echoed in Zimbalist (1992). Zimbalist noted that competitive balance had improved following the institution of free agency, but such improvement was most likely because of the compression of baseball talent.

2. The authors would like to thank Sean Lahmen (2002), author of the *Baseball Archive* (http://www.baseball1.com/), for the data on the number of foreign-born players in Major League Baseball.

3. The presumption here is not that these limits cannot be altered. They certainly may, through training, nutrition, and self-selection. Rather, the important assumption is that although the pool is growing relatively quickly, the limits evolve slowly through time and thereby lead to the nonlinearity.

4. The author gratefully acknowledges that the nonlinear unit root and cotrending tests were performed using Herman Bierens's econometric (free) package. It is available at http://econ.la.psu.edu/~hbierens/EASYREG.HTM.

5. The interested reader is referred to Bierens (1997, 2000) for the more technical aspects of the tests.

REFERENCES

Bierens, H. J. (1997). Testing the unit root hypothesis against nonlinear trend stationarity, with an application to the price level and interest rate in the U.S. *Journal of Econometrics*, *81*, 29-64.

Bierens, H. J. (2000). Nonparametric nonlinear co-trending analysis, with an application to interest and inflation in the U.S. *Journal of Business and Economic Statistics*, *18*, 323-337.

Butler, M. R. (1995). Competitive balance in Major League Baseball. *American Economist*, *39*(2), 46-52

Chatterjee, S., & Yilmaz, M. R. (1991). Parity in baseball: Stability of evolving systems? *Chance*, *4*, 37-42.

Depken, C. A. (2000). Wage disparity and team productivity: Evidence from Major League Baseball. *Economics Letters*, *67*, 87-92.

Depken, C. A. (2002). Free agency and the concentration of player talent in Major League Baseball. *Journal of Sports Economics*, *3*(4), 335-353.

Gould, S. J. (1983, March). Losing the edge: The extinction of the .400 hitter. *Vanity Fair*, *120*, 264-278.

Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin*. New York: Three Rivers.

Horowitz, I. (1997). The increasing competitive balance in Major League Baseball. *Review of Industrial Organization*, *12*, 373-387.

Lahmen, S. (2002). *The baseball archive*. Retrieved February 2003 from http://www.baseball1.com/

Marcano Guevara, A. J., and Fidler, D. P. (2003). *Stealing lives: The globalization of baseball and the tragic story of Alexis Quiroz*. Bloomington: Indiana University Press.

Ouliaris, S., Park, J. Y., Phillips, & Peter C. B. (1989). Testing for a unit root in the presence of a maintained trend. In B. Raj (Ed.), *Advances in econometrics and modeling* (pp. 7-28). Amsterdam: Kluwer.

Schmidt, M. B., & Berri, D. J. (2001). Competition and attendance: The case of Major League Baseball. *Journal of Sports Economics*, *2*(2), 147-167.
Zimbalist, A. (1992). *Baseball and billions*. New York: Basic Books.

Martin B. Schmidt is an associate professor in the Department of Economics at the College of William and Mary.

David J. Berri is an assistant professor in the Department of Economics at California State University, Bakersfield.