Applied Economics, 2004, **36**, 2007–2014

Routledge
Taylor & Francis Group

Convergence and clustering in Major League Baseball: the haves and have nots?

MARTIN B. SCHMIDT* and DAVID J. BERRI‡

Department of Economics, The College of William and Mary, PO Box 8795, Williamsburg, VA 23187-8795, USA and ‡*Department of Economics, California State University-Bakersfield, 9001 Stockdale Highway, Bakersfield, CA 93311, USA*

There appear to be two distinct views on the level of competitive balance within Major League Baseball. One view, mostly associated with academics, is that competition is more equal today than it ever has been. The other view, mostly associated with the media and the industry, is that competition is far worse today. The present paper, borrowing from the literature on economic convergence, finds that both views are valid. More specifically, while competitive balance has continued to improve, the improvement has been such as to create distinct convergence clusters. A discussion of the composition of these clusters is offered in the text.

I. INTRODUCTION

A central tenet of economic thought is that competition between firms produces substantial social benefits. Perhaps equally central is the observation that firms prosper when competition is eliminated. An exception to this latter view is the professional sports industry, where the elimination of competition effectively eliminates the industry.[1] Consequently, economic actors within this industry have often proposed and enacted institutions designed, at least according to industry insiders, to promote competitive balance.

In spite of these efforts, both industry insiders and various members of the media believe the last decade of the 20th century was marked by a decided lack of competitive balance.[2] This argument was clearly stated by Major League Baseball's (MLB) Blue Ribbon Panel (Levin *et al.*, 2000).[3] The report of the Blue Ribbon Panel suggested that there were significant differences in the revenues earned by MLB teams and that these differences have resulted in declines in the level of competitive balance. To illustrate their point, the authors of this report point to the performance of teams with the lowest payrolls. From 1995–1999, the only years the panel considered, no team with a payroll in the bottom 50% of the payroll rankings appeared in MLB's annual post-season competition. Such a result suggests that teams located in smaller markets know the outcome of the season (i.e. these teams will not appear in the post-season) before the season begins.[4]

The academic literature, in contrast, has reached a decidedly different conclusion regarding the level of competitive balance in Major League Baseball. These investigations typically examine much longer periods of data and suggest that the playing field is much more competitive today than

* Corresponding author. E-mail: mbschm@wm.edu

[1] This point has been made in the work of Rottenberg (1956), Neale (1964), and El-Hodiri and Quirk (1971).

[2] For example within the American League, the New York Yankees won four World Series titles and in the National League, the Atlanta Braves won nine Division titles.

[3] The Commissioner's Blue Ribbon Panel on Baseball Economics was convened by MLB to investigate the issues of competitive balance and economic health. Specifically, the panel's stated purpose was to 'examine the question of whether Baseball's current economic system has created a problem of competitive imbalance in the game' (Levin *et al.*, 2000, p. 59).

[4] It is the case, that the BRP completely sidesteps the unavoidable question of what MLB teams maximize. It is possible, perhaps likely, that these teams maximize profits rather than wins. However, even in this case, competitive balance is important as attendance and, therefore, revenues are sensitive to competitive balance (Schmidt and Berri, 2001).

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684042000261851

Table 1. *League competitive balance measures by decade*

Period	AL-GINI	NL-GINI	AL-SD	NL-SD
1901–1910	0.107	0.135	0.097	0.121
1911–1920	0.112	0.095	0.103	0.085
1921–1930	0.099	0.098	0.090	0.090
1931–1940	0.115	0.098	0.103	0.089
1941–1950	0.099	0.103	0.090	0.093
1951–1960	0.101	0.087	0.091	0.078
1961–1970	0.085	0.086	0.076	0.079
1971–1980	0.081	0.078	0.072	0.069
1981–1990	0.071	0.070	0.064	0.063
1991–2000	0.071	0.077	0.065	0.068

Notes: The values are simple averages of yearly values over the indicated period.

it ever has been. Table 1, for example, highlights decade averages for both. Similar evidence exists elsewhere; Whether the standard deviation of team wins (Quirk and Fort, 1992), the dispersion and season-to-season correlation of team winning percentages (Balfour and Porter, 1991; Quirk and Fort, 1992; Butler, 1995), the relative entropy approach (Horowitz, 1997), the Gini coefficient (Schmidt, 2001; Schmidt and Berri, 2001, 2002), or the Herfindahl–Hirschman index (Depken II, 1999) are examined the findings from each of these studies have been remarkably consistent.[5]

Despite such evidence, a problem does exist with the academic analysis. Whether standard deviations, Gini coefficients or the Herfindahl–Hirschman Index are employed, the approach is still the same. What is being measured is simply the distribution of wins in a given season. Although these measures all suggest that the dispersion of wins is declining over time, none of these measures explain how this decline is taking place. For example, it could be the case that all teams are converging upon the mean winning percentage, i.e., 0.5. Although this is one possibility, improvements in the level of competitive balance could also be seen if only a subset of teams were converging upon a common winning percentage.

In order to understand the nature of this convergence, the literature examining convergence in per capita productivity levels between nations is considered. Specifically, focus is upon the work of Baumol (1986) and Hobijn and Franses (2000). The latter begin their work by noting 'that no matter what definition of convergence is used, unconditional convergence of per captia productivity levels for all economies in the world does not exist' (Hobijn and Franses, 2000, p. 59). Baumol's prior work suggested that convergence might occur within a subset of nations. Although convergence might be observed within these subsets, labelled by Baumol as 'convergence clubs', convergence might not be observed between each individual grouping of nations.

This approach is applied to examine the nature of convergence in professional baseball. Specifically, convergence is examined in terms of individual team lifetime winning percentage. Rolling estimates of the number of convergence groups are used and whether the number and composition of these groupings changes across time is examined. In general the results suggest that convergence between MLB teams has occurred. However, the convergence is not universal, but rather confined within two distinct convergence clusters. Such a result can serve as a bridge between the view of improving competitive balance offered in the academic literature, and the perspective of the industry, where the prospects for success are not the same for all teams.

The paper has three additional sections. Section II describes the convergence algorithm first utilized by Hobijn and Franses. Section III reports the results of its application to MLB team winning percentages. Finally, Section IV provides concluding observations.

II. THE CONVERGENCE ALGORITHM

The question of competitive balance is closely aligned with that of convergence. Increased competitive balance may be defined as a situation where teams are more equal or perhaps more similar than they previously were. In fact, perfect competitive balance would be the situation where all teams have reached a level of similarity so that each has exactly the same probability of winning.

Along similar lines, the question of convergence asks whether two or more variables are moving towards one another, i.e., whether the variables are likely, independent of their current position, to converge to identical values in the future, i.e., the same probability of winning. Formally, Bernard and Durlauf (1995) define *stochastic convergence* as the situation where p variables, i.e., $p = 1, 2, \ldots, n$, have equal long-term forecasts at fixed time t.[6] As described by

[5] Quirk and Fort (1992) also measured competitive balance by comparing the standard deviation of each season to an idealized standard deviation which would exist if the league had maximized competitive balance.

[6] In several previous studies it was found that the time series data are stationary around a deterministic trend and therefore suggest a degree of convergence. An example in macroeconomics concerns divergences between real output for pairs of countries (see Hobijn and Franses, 2000) or pairs of regions within the USA (see Carlino and Mills, 1993; Loewy and Papell, 1995 and Tomljanovich and Vogelsang, 2001). Other examples can be found in disciplines such as tourism and marketing, where tourist arrivals and sales often display upward trending patterns. Finally, environmental data like temperatures may also display trends, and if these are upward moving this can be taken as evidence of global warming (see Bloomfield, 1992; Woodward and Gray, 1993; Zheng and Basher, 1999 and Fomby and Vogelsang, 2000).

these authors, testing whether two or more series are converging may be accomplished by examining whether the difference between the variables is expected to be eliminated over time.

Specifically, then, let y_{it} represent yearly winning percentage of team i in period t. In layman's terms, convergence, or increased competitive balance, is defined as the condition where team winning percentages are moving towards one another. Further $d_{(i,j)t}$ is defined as the difference between y_{it} and y_{jt} for teams i and j in period t. Convergence is therefore said to occur whenever the value of $d_{(i,j)t} \Rightarrow 0$. Hobijn and Franses (2000) define *asymptotically perfect convergence* as the condition where the value of $d_{(ij)t}$ is found to be zero mean stationary. The definition follows Bernard and Durlauf (1995) and implies that at any moment it may be expected that winning percentages for team i and j will converge to one another, regardless of their respective starting points.

In order to test for the stationarity of $d_{(i,j)t}$, Hobijn and Franses extend the stationary tests suggested by Kwiatkowski *et al.* (KPSS) (1992). More specifically, Hobijn and Franses propose the following multivariate version of the KPSS test statistic:

$$\eta_0 = T^{-2} \sum_{i=1}^{T} S_t'[G^{-1}]S_t \tag{1}$$

where S_t is the partial sum series of x_t, a multivariate version of $d_{(i,j)t}$. Specifically, x_t is defined as $x_t \equiv M_k^* y_t^*$ where:

$$y_t^* = [y_{1t}, \ldots, y_{kt}] \quad M_k = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \tag{2}$$

Consistent with the KPSS approach, x_t is assumed to satisfy the general stationarity conditions under the H_0.[7]

The second element, G_t, represents the Newey–West (1987) estimator of the covariance matrix under the null hypothesis of k clusters. Following KPSS, the variance measure is estimated from the following equation:

$$x_t = \alpha + \beta * t + u_t \qquad t = 1, 2, \ldots, T \tag{3}$$

Equation 3 provides estimates of u_i and, therefore, G_t. As is the case with the KPSS test, G_t requires a bandwidth choice for the Bartlett kernal estimator of the long run variance, i.e., l. Finally, Hobijn and Franses (2000) provide asymptotic critical values for the test with a null hypothesis.

In addition, Hobijn and Franses offer an algorithm to estimate the number and composition of the clusters. While the reader is referred to Hobijn and Franses (2000) for the more technical details and proofs, the clustering algorithm is briefly explained.[8] Given the n different series, there exist n possible clusters. These n clusters may be comprised of $n!$ separate combinations. The Hobijn and Franses algorithm, therefore, begins by initializing these possible different clusters, i.e., $k(i) = \{i\}$ for all $i = 1, \ldots, n$. The probability of stationarity between clusters is then examined for all possible clusters, i.e., $k(i)$ and $k(j)$. If there exists no combination of clusters i and j for which the probability is $p_0 > p_{\min}$, asymptotically perfect convergence is rejected for all clusters and no clusters are found to converge. If, however, there is such a combination, the algorithm chooses the combination which is associated with the highest p_0. These two clusters are assumed to be most likely to converge. The algorithm then repeats itself once the two clusters are combined, i.e., redefining $k(i)$ as the union of $k(i)$ and $k(j)$, and discarding cluster j, i.e., $k(j) = \phi$. The algorithm ends once no further combination has $p_0 > p_{\min}$. The clusters obtained then represent the 'asymptotically perfectly convergence' clusters.

An important limitation of the Hobijn and Franses approach is its applicability within small samples. Specifically, the concern is within panel data where the time series observations are less than the cross-sectional observations, i.e., $n > T$. In such situations, the estimated variance–covariance matrix may be nonsingular and the algorithm is less reliable. In general then, Hobijn and Franses argue against interpreting these results. Because of this concern, the present paper attempts to use as large a time series sample as possible. Consequently only the behaviour of Major League Baseball's original 16 teams, which are listed in Table 2, is examined. By examining this sample of teams it is possible to study the behaviour of team winning percentage from 1901 to 2001. A final additional reason for limiting the analysis to these teams is that these organizations were the only ones who experienced each wave of league expansion.[9]

III. CONVERGENCE CLUSTERS IN MAJOR LEAGUE BASEBALL

Aside from the choice of data, the Hobijn and Franses algorithm requires the researcher to make two additional decisions: (1) the choice of (l) for the Bartlett kernel

[7] In contrast to other tests of stationarity, the KPSS' tests have as their null (H_0) that the series is $I(0)$.

[8] The Gauss program to run the Hobijn and Franses algorithm is available from Bart Hobijn's homepage, [http://www.newyorkfed.org/rmaghome/economist/hobijn/data.html].

[9] MLB has expanded six times from the original 16 teams. In 1961, 1962, 1977, 1993 and 1998, the league expanded by two teams, while four teams were added in 1969.

Table 2. *The original 16 teams in Major League Baseball*

Altlanta Braves/Milwaukee Braves/Boston Braves – (ATL)	Los Angles Dodgers/Brooklyn Dodgers – (LA)
Baltimore Orioles/St. Louis Browns – (BAL)	Minnesota Twins/Washington Senators – (MIN)
Boston Red Sox – (BOS)	New York Yankees – (NYY)
Chicago Cubs – (CHC)	Oakland A's/Kansas City A's/Philadelphia A's – (OAK)
Chicago White Sox – (CWS)	Philadelphia Phillies – (PHI)
Cincinnati Reds – (CIN)	Pittsburgh Pirates – (PIT)
Cleveland Indians – (CLE)	St. Louis Cardinals – (STL)
Detroit Tigers – (DET)	San Francisco Gaints/New York Giants – (SF)

estimator of the long run variance, (2) the choice of $p_{\min}$. Following the '$l(k)$ rule', where $l(k) = \mathrm{int}\,[k(T/100)^{1/4}]$ and k represents the number of hypothesized clusters, $l = k$ for much of the sample. Therefore the estimated results for bandwidth of 2, 3 and 4 are examined. As for the choice of $p_{\min}$, Hobijn and Francis is followed and $p_{\min} = 0.01$ chosen. While the choice of bandwidth does not affect the finding of convergence in any systematic way, the choice of $p_{\min}$ does impact the estimated results. As is described in Hobijn and Franses (2000), the smaller the chosen value for $p_{\min}$, the less likely convergence will be rejected. The results, however, were robust to higher values of $p_{\min}$, i.e., (0.05, 0.10).

Estimated number of clusters

Figure 1(a) reports rolling estimates of the number of clusters for the sample of teams considered. In order to show the robustness of the results, the figure reports the estimates for several different bandwidths. In general, these suggest that regardless of the choice of bandwidth a similar result appears: the number of clusters has fallen since 1960. Similar evidence is highlighted by the rolling cluster averages reported in Table 3.

In order to remove some of the short-run variations which may be due to excessive noise, Fig. 1(b) reports smoothed estimates. The smoothed estimates were computed by averaging the two previous estimates with the two future estimates. Overall these results suggest a marked reduction in the number of clusters from the early 1960s until the mid-1970s.

Composition of convergence clusters

Given these findings, two specific periods are examined more closely: 1901–1960 and 1901–2000. The make-up of the individual clusters for these two periods is reported in Table 4. The first time period covers the history of Major League Baseball prior to the first wave of expansion in 1960, while the second covers the entire data set. While the estimated clusters from a particular bandwidth (3) are reported, similar results were obtained from the alternative choices.

Specifically for the 1901–1960 subsample, five convergence clusters are estimated. Of these, two clusters have six members, while of the remaining three clusters, two are formed from two teams with the San Francisco/New York Gaints standing alone. In contrast, the number of clusters declines from five to three for the full sample.

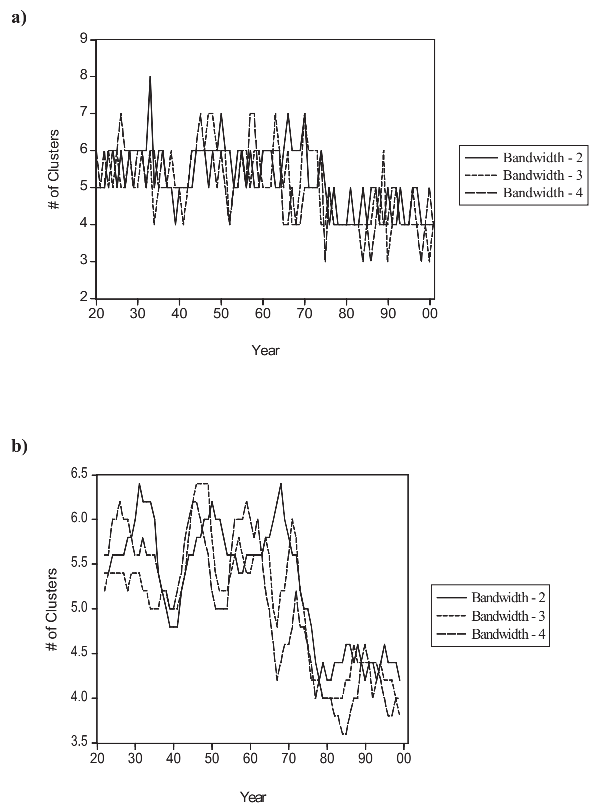

Fig. 1. *Number of estimated clusters w/ given bandwidth – MLB's original 16 teams. (a) Actual estimates; (b) Smoothed estimates. (The smoothed estimates were computed by averaging the two previous estimates with the two future estimates. In addition $\alpha = 0.01$.)*

Table 3. *Number of convergence clusters – original 16 teams*

Period	No. of clusters			Period	No. of clusters		
Bandwidth (l)	(2)	(3)	(4)	Bandwidth (l)	(2)	(3)	(4)
1901–1930	6	5	6	1901–1970	7	5	6
1901–1940	5	5	5	1901–1980	7	5	6
1901–1950	7	6	6	1901–1990	4	7	5
1901–1960	6	5	6	1901–2000	5	4	4

Notes: $\alpha = 0.01$.

Table 4. *Composition of convergence clusters*

	Convergence cluster no.			Convergence cluster no.	
Team (Sample)	(1901–1960)	(1901–2000)	Team (Sample)	(1901–1960)	(1901–2000)
Cubs	1	2	Orioles/Browns	2	1
Indians	1	1	White Sox	2	1
Tigers	1	1	Dodgers	2	2
Twins/Senators	1	1	Yankees	2	2
A's	1	1	Cardinals	2	2
Phillies	1	1	Red Sox	4	2
Braves	3	1	Pirates	4	1
Reds	3	2	Giants	5	3

Notes: Bandwidth $-l=3$. In addition, $\alpha=0.01$.

This reduction in clusters was accomplished by the members of clusters 3 and 4 from the 1960 subsample being evenly divided into clusters 1 and 2. Also, the Chicago White Sox and Baltimore/St. Louis were moved from cluster 2 to 1, while the Chicago Cubs made the opposite move. The movement from cluster 2 to 1 may be seen as a negative move for the individual team as cluster 1 has lower winning percentage for the entire sample. Once again, though, the Giants remain in a cluster by themselves. While the composition of these clusters is discussed next, it should be noted that the reduction in the number of clusters does suggest convergence in the aggregate. The reduction in clusters suggests that groups of teams are becoming more balanced. However, the reduction in the number of clusters does not necessarily indicate that the number of convergence clusters will eventually fall to one, nor does it necessarily indicate a lack of disparity between the estimated clusters.

Examining the composition of convergence clusters

What is clear is that the number of convergence clusters has declined when the entire 20th century is considered. However, the above results suggest that the convergence has not been universal. Such a finding is consistent with the arguments offered by Major League Baseball, specifically the contention that professional baseball is segmented into the 'Haves' and the 'Have Nots'. However, the results above are consistent with a number of convergence patterns. Therefore, a next step is to identify which clusters, if any, can be identified with either of these a priori groupings.

As an initial exercise, Table 5 reports the lifetime winning percentage of the original 16 teams, organized by the 1901–2000 clusters. The average winning percentage of every team in clusters 2 and 3 is above 0.500. In contrast, cluster 1 is populated with a number of below average teams and furthermore, the average winning percentage of the cluster lies below 0.500. In addition, revenues were much larger for cluster 1. Not surprisingly, these results suggest that the membership in each convergence cluster is driven by on-field performance.

Table 5. *Lifetime winning percentage of original 16 teams: 1901–2001 clusters*

Cluster 1	Winning PCT	Cluster 2	Winning PCT
Pirates	0.517	Yankees	0.565
Indians	0.513	Dodgers	0.523
Tigers	0.511	Cardinals	0.516
White Sox	0.504	Red Sox	0.512
A's	0.479	Cubs	0.505
Braves	0.479	Reds	0.505
Orioles/Browns	0.478	Average	0.512
Twins/Senators	0.477	Average w/o NY	0.512
Phillies	0.458		
		Cluster 3	Winning PCT
Average	**0.490**	Giants	0.538

Certainly it could be the case that these average wining percentages benefited from each teams' early experience and therefore do not reflect day performance. For example, a team winning ten additional games will only change winning percentage by (0.06). Such a result, being averaged over an increasingly larger sample, would take a team a lengthy period to significantly alter their average lifetime winning percentage. In order to examine the more recent experience of each cluster's winning percentages, a rolling 5-year average was computed for clusters 1 and 2 from the full sample. These are plotted in Fig. 2. The results indicate that the two clusters diverged for much of baseball's early history, but were much closer during the early 1960s. However, they moved apart for much of the remaining periods. In particular, the latter 1990s, primarily because of the performance of the New York Yankees, saw cluster 2 gaining the upper hand.

The analysis of winning percentage suggests the clusters can be defined in terms of on-field performance. The aforementioned Blue Ribbon Panel, though, did not simply argue that baseball is populated by teams that generally win and those that generally lose. Rather, what the panel argued is that the teams that win, or the 'Haves', reside

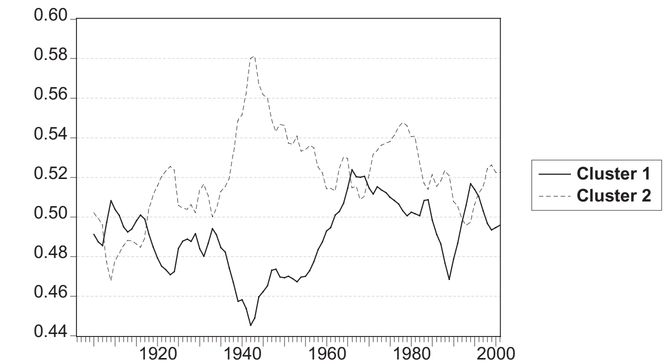

Fig. 2. *Average winning percentage by cluster. (Sample – 1901–2000; Bandwidth – (l = 3) and α = 0.01.)*

within larger markets and those that lose, or the 'Have Nots', are located in smaller markets. In other words, because of circumstances exogenous to the 'Have Nots', the small market teams are destined to lose more often than they win.

Therefore, it is necessary to define the terms 'large' and 'small' market. The Blue Ribbon Panel defined market size in terms of team payroll, which clearly is a function of team revenue. Unfortunately, given that Major League Baseball teams are private entities, revenue data is not frequently made available to the general public. *Financial World* and *Forbes* did report team revenue data for each Major League Baseball team for the years 1990–1997. When the average level of revenue in each cluster is examined, reported in Table 6, it is clear that cluster 2 is able to accumulate larger revenues than cluster 1. This advantage is maintained even when the New York Yankees, Major League Baseball's revenue leader, is eliminated from the sample.

Table 6 also reports each team's average winning percentage for the years team revenue data is available. The results indicate that the average winning percentage in cluster 1, or the supposed 'Have Nots', is actually higher than the average winning percentage of the 'Haves' over this time period. Much of this is driven by the recent performance of Cleveland and Atlanta, two historically 'small' market teams that were consistent contenders in the 1990s. The performance of these two teams cast doubt on Major League Baseball's contention that 'small' market teams cannot offer consistent above average performances.

Table 6. *Winning percentage and average revenue of original 16 teams: 1901–2001 clusters*

Cluster 1	Winning PCT	Average Revenue	Cluster 2	Winning PCT	Average Revenue
Pirates	0.507	40.0	Yankees	0.525	104.2
Indians	0.524	62.0	Dodgers	0.518	76.4
Tigers	0.461	45.4	Cardinals	0.487	58.9
White Sox	0.539	67.2	Red Sox	0.510	77.1
A's	0.495	53.9	Cubs	0.473	67.2
Braves	0.583	65.1	Reds	0.520	45.8
Drioles	0.520	78.9	Average	0.505	71.6
Twins	0.475	41.1	Average w/o NY	0.502	65.1
Phillies	0.471	50.3	Cluster 3	Winning PCT	Average Revenue
Average	0.508	56.0	Giants	0.499	53.3

Notes: Sample: 1990–1997.

Beyond the anomalous performance of Cleveland and Atlanta, whether much is learned by defining the size of a team's market in terms of payroll or revenue could also be questioned. Despite the evidence reported in Table 6, it would be expected that teams that perform well on the field to generally perform well in terms of revenue, and vice versa. Certainly numerous studies into the economic value of professional baseball players have demonstrated such a link.[10] However, if the 'Haves' are simply defined in terms of team revenue, a factor that is impacted by team wins, it is merely being stated that teams that perform well on the field also perform well off the field of play. For Major League Baseball to make its argument that teams are incapable of competing characteristic is needed, beyond the control of the team in questions, that allows why the teams in each cluster have accumulated their win and revenue totals to be understood.

One obvious characteristic is the population of each team's host city. Utilizing data from the US Census Bureau, the average population of each cluster was calculated.[11] The results are reported in Table 7.[12] If the size of the market drives the composition of the convergence clusters, then there should be differences in the average population of the teams located in each cluster.

[10] For a representative sample of this literature see Scully (1974), Medoff (1976) and Zimbalist (1992).

[11] The data ultilized to construct these averages was taken from the web site of the United States Census Bureau [http://www.census.gov]. The population of the cities that housed multiple teams was divided equally between the resident organizations. For example, the population of Philadelphia in 1910 was 1 549 008. Given that both the A's and Phillies played in Philadelphia for that season, the population total used for these cities was half this number, or 774 504.

[12] The data in Table 5 begins with the 1903 season. Although the American League began operations in 1901, for the 1901 season the St. Louis Browns were located in Milwaukee while the New York Yankees played in Baltimore. For the 1902 season, the Milwaukee franchise moved to St. Louis. After this season, the Baltimore franchise was moved to New York. Major League Baseball saw no movement of franchises until the Boston Braves moved to Milwaukee following the 1952 campaign.

Table 7. *Average population of clusters*

Year	Cluster 1	Cluster 2	Cluster 2 w/o NY	Cluster 3	Cluster 4	Cluster 5
(a) Sample 1901–1960						
1903	514 862	743 199	474 842	303 174	301 031	1 145 734
1910	666 525	991 519	593 224	349 442	434 599	1 588 961
1920	900 354	1 174 090	707 917	387 639	481 187	1 873 349
1930	1 099 190	1 426 095	836 726	420 877	530 206	2 310 149
1940	1 132 436	1 496 890	838 151	420 509	528 534	2 484 998
1950	1 241 440	1 585 716	889 092	452 360	538 764	2 630 652
1960	1 260 567	1 718 950	1 175 940	621 937	650 765	740 316

(b) Sample 1901–2000				
Year	Cluster 1	Cluster 2	Cluster 2 w/o NY	Cluster 3
1903	442 095	672 454	435 814	1 145 734
1910	579 096	885 494	533 760	1 588 961
1920	750 105	1 043 213	628 145	1 873 349
1930	896 281	1 260 208	735 238	2 310 149
1940	917 745	1 319 574	736 862	2 484 998
1950	994 952	1 400 817	785 900	2 630 652
1960	1 094 231	1 475 912	992 897	740 316
1970	957 031	1 459 129	961 468	715 674
1980	812 659	1 320 553	877 499	678 974
1990	750 174	1 355 142	893 914	723 959

For the first sample, 1901–1960, cluster 2, or the teams that on average win, appears to be the large market club, followed by clusters 1, 4 and then 3. If New York is removed from the sample, though, the average population of cluster 2 falls below cluster 1. Such a finding suggests that the large market status of cluster 2 is derived mostly by the presence of New York.

A similar story can be told when the larger sample 1901–2000 is considered. Again, cluster 2 appears as the large convergence cluster. With New York removed, though, the size disparity is reversed for 1901–1960. Not until 1990 does cluster 2 have 10% more than the average population enjoyed by cluster 1. This advantage is not evident when Los Angeles is also removed from cluster 2. Without New York and Los Angeles, the average population of cluster 2 is consistently below the average observed for cluster 1. Such a finding suggests that membership in the clusters is not necessarily driven by the population of the host cities.

IV. CONCLUDING OBSERVATIONS

Major League Baseball contends that baseball is segmented into the 'Have' and the 'Have Nots'. The 'Haves' generally expect to contend prior to the commencement of each season and the 'Have Nots' understand that contention is not a likely possibility. Consequently, following this argument, competitive balance is a significant problem for the sport. Prior research into the competitive nature of Major League Baseball, though, suggested that the league is more competitively balanced at the end of the 20th century than at any previous time. The evidence reported herein appears to reconcile these disparate perspectives.

Borrowing from the literature on economic convergence, an algorithm developed by Hobijn and Franses (2000) is applied to examine the level of convergence in Major League Baseball. A cursory examination of these results does appear to support the contention of the Blue Ribbon Panel. The league is indeed divided into two main groups. The characteristic of each group, though, is open to debate. Although one group clearly enjoys higher revenues and a higher lifetime winning percentage, its advantage in terms of population is not as clear. Furthermore, in terms of performance on the field, the advantage of the so-called 'Haves' disappears when the analysis is confined to the time period coinciding to the years revenue data was available. Such a result suggests that baseball is not clearly divided between 'large' and 'small' market teams.

To establish the claims of baseball it is necessary to find a characteristic, unique to 'large' market teams, that allows such teams to consistently succeed. Baseball has focused on differences in team payroll and team revenue. Such discrepancies, though, do not appear to be perfectly consistent with the population of a team's host city. If such is not the case, it can be simply argued that teams that are better managed will accumulate higher revenues, and such revenues can be used to produce more wins. Following this argument, baseball does not have a problem with differences in market size, but rather a shortage of quality managers capable of transforming an underperforming organization into a consistent winner on and off the field.

REFERENCES

Balfour, A. and Porter, P. K. (1991) The reserve clause and professional sports: legality and effect on competitive balance, *Labor Law Journal*, **42**(1), 8–18.

Baumol, W. J. (1986) Productivity growth, convergence and welfare: what the long-run data show, *American Economic Review*, **76**, 1072–85.

Bernard, A. and Durlauf, S. N. (1995) Convergence in international output, *Journal of Applied Econometrics*, **10**, 161–73.

Bloomfield, P. (1992) Trends in global temperatures, *Climate Change*, **21**, 275–87.

Butler, M. R. (1995) Competitive balance in Major League Baseball, *American Economist*, **39**(2), 6–52.

Carlino, G. A. and Mills, L. O. (1993) Are US regional incomes converging?, *Journal of Monetary Economics*, **32**, 335–46.

Depken II, C. A. (1999) Free-agency and the competitiveness of Major League Baseball, *Review of Industrial Organization*, **14**(3), 205–17.

El-Hodiri, M. and Quirk, J. (1971) The economic theory of a professional sports league, *Journal of Political Economy*, **79**, November–December, 1302–19.

Fomby, T. and Vogelsang, T. J. (2000) The application of size robust trend analysis to global warming temperature series, *mimeo*, Department of Economics, Cornell University.

Hobijn, B. and Franses, P. H. (2000) Asymptotically perfect and relative convergence of productivity, *Journal of Applied Econometrics*, **15**, 59–81.

Horowitz, I. (1997) The increasing competitive balance in Major League Baseball, *Review of Industrial Organization*, **12**, 373–87.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic series have a unit root?, *Journal of Econometrics*, **54**, 159–78.

Levin, R. C., Mitchell, G. J., Volcker, P. A. and Will, G. F. (2000) *Blue Ribbon Panel on Baseball Economics*, http://www.majorleaguebaseball.com/u/baseball/mlbcom/pressrelease/blueribbon_release.htm.

Loewy, M. B. and Papell, D. H. (1995) Are US regional incomes converging? Some further evidence, *Journal of Monetary Economics*, **38**, 587–98.

Medoff, M. H. (1976) On monopsonistic exploitation in professional baseball, *Quarterly Review of Economics and Business*, **16**(2), 113–21.

Neale, W. (1964) The peculiar economics of professional sports, *The Quarterly Journal of Economics*, **78**(1), 1–14.

Newey, W. K. and West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703–08.

Quirk, J. and Fort, R. (1992) *Pay Dirt: The Business of Professional Team Sports*, Princeton University Press, Princeton, NJ.

Rottenberg, S. (1956) The baseball player's labor market, *Journal of Political Economy*, **64**, 242–58.

Schmidt, M. B. (2001) Competition in Major League Baseball: the impact of expansion, *Applied Economics Letters*, **8**(1), 21–6.

Schmidt, M. B. and Berri, D. J. (2001) Competition and attendance: the case of Major League Baseball, *Journal of Sports Economics*, **2**(2), 147–67.

Schmidt, M. B. and Berri, D. J. (2002) Competitive balance and market size in Major League Baseball: a response to the MLB's Blue Ribbon Panel, *Review of Industrial Organization*, **21**(1), 41–54.

Scully, G. W. (1974) Pay and performance in Major League Baseball, *American Economic Review*, **64**, 917–30.

Tomljanovich, M. and Vogelsang, T. J. (2002) Are US regions converging? Using new econometric methods to examine old issues, *Empirical Economics*, **27**(1), 49–62.

Woodward, W. A. and Gray, H. L. (1993) Global warming and the problem of testing for trend in time series data, *Journal of Climate*, **6**, 953–62.

Zheng, X. and Basher, R. E. (1999) Structural time series models and trend detection in global and regional temperature series, *Journal of Climate*, **12**, 2347–58.

Zimbalist, A. (1992) *Baseball and Billions*, Basic Books, New York.