MANAGEMENT SCIENCE
Vol. 28, No. 12, December 1982
Printed in U.S.A.

DOES DIVISIONAL PLAY LEAD TO MORE PENNANT RACES?*

WAYNE WINSTON† AND ASHOK SONI†

A "pennant race" occurs in a sports league if the top two teams in a league finish within $\bar{m}$ games of each other. We develop an analytical model that can be used to predict the probability that a sports league will have a pennant race. The model is then tested on data from the last 50 major league baseball seasons and is shown to do a remarkable job of predicting the number of pennant races that actually occurred. Finally, we give conditions which ensure that dividing a league of $2n$ teams into 2 n-team divisions will increase the expected number of pennant races. As a corollary to this result we prove that the divisional structure which maximizes the expected number of pennant races must contain no divisions with 4 or more teams.
(RECREATION/SPORTS)

1. Introduction

When major league baseball split each twelve team league into two six team leagues during the 1969 season, a primary reason given by the baseball brass was as follows (see page 123 of [4]):

> The new divisional alignment was adopted with the primary hope that four six team divisions would provide closer pennant races and more excitement in the August and September stretch. There would now be four races instead of two, an arrangement employed with success by the major professional football leagues.

It is not obvious, however, that dividing a league into two divisions will always increase the expected number of pennant races. For example, during the 1972 National League season Pittsburgh won the Eastern Division by 10.5 games while Cincinnati won the Western Division by 10.5 games. If the National League had not adopted divisional play, however, there would have been an exciting pennant race in which Pittsburgh would have nosed Cincinnati out by 1 game. Similarly during the 1976 National League season Philadelphia won the Eastern Division by 9 games while Cincinnati won the Western Division by 10 games. In the absence of divisional play, Cincinnati would have nosed out Philadelphia by a single game.

These examples show that divisional play may lead to fewer, not more, pennant races. The purpose of this paper is to determine whether or not divisional play does increase the expected number of pennant races. In §2 we develop an analytical model that can be used to estimate the probability of a pennant race occurring in a sports league. In §3 we show that our model closely fits the actual number of pennant races occurring the last 50 major league baseball seasons. Finally, in §4 we give conditions which ensure that dividing an n-team league into any number of divisions will increase the expected number of pennant races (as long as each division contains at least two teams). As a corollary to this result we show that the divisional structure that maximizes the expected number of pennant races cannot have any division with 4 or more teams. Our major findings are summarized in §5.

* Accepted by Ambar G. Rao; received August 3, 1981. This paper has been with the authors 2 months for 1 revision.

† Indiana University.

0025-1909/82/2812/1432$01.25

2. A Model for Predicting the Probability of a Pennant Race

Let X_i $(i = 1, 2, \ldots, n)$ be a random variable representing the fraction of games won by team i during a season. Here we assume the X_i's are independent continuous random variables drawn from a distribution function $F(x)$ with a probability density function $f(x)$. However, there are two drawbacks in making this assumption. First of all X_i is clearly a discrete and not a continuous random variable. For a long season such as baseball, however, it does not appear that the assumption that each X_i is continuous seriously compromises reality. A more serious drawback is the fact that the derivation of our model critically depends on the independence of the X_i's. Since for every winner of a game there is also a loser, it is clear that if X_i is the fraction of games won during a season by team i, then $X_1 + X_2 + \ldots X_n = n/2$ must hold. Of course this implies that the X_i's cannot be independent. Since in major league baseball each team plays at most $1/9$ of its games against any single team, it seems that the assumption of independent X_i's may be fairly realistic. The only way to judge whether or not our two drawbacks are significant is to see if the model, despite its drawbacks, can be used to accurately predict the probability of a pennant race. This will be the subject of §3.

Let Y_1 be the largest of the X_i and Y_2 be the second largest of the X_i. A pennant race of closeness $\overline{m}$ is said to occur if the top two teams in the league finish within $\overline{m}$ games of each other. If each team in the league plays g games during a season, it is clear that a pennant race occurs if and only if $Y_1 - Y_2 \leqslant m$, where $m = \overline{m}/g$. Using familiar arguments from the theory of order statistics (see [2]), it is easy to determine $P(Y_1 - Y_2 \leqslant m)$. First we compute the joint density function $f(y_1, y_2)$ of Y_1 and Y_2. This is easily seen to be $f(y_1, y_2) = n(n-1)f(y_1)f(y_2)F(y_2)^{n-2}$. Then for $0 \leqslant m \leqslant 1$

$$P(Y_1 - Y_2 \leqslant m) = \int_m^1 \int_{y_l - m}^{y_1} f(y_1, y_2)\, dy_2\, dy_1 + P(Y_1 \leqslant m)$$

$$= n(n-1)\int_m^1 f(y_1)\left\{ \frac{F(y_1)^{n-1}}{n-1} - \frac{F(y_1 - m)^{n-1}}{n-1} \right\} dy_1 + P(y_1 \leqslant m).$$

Since $P(Y_1 \leqslant m) = F(m)^n$, the last equality reduces to

$$1 - F(m)^n - n(n-1)\int_m^1 f(y_1)\frac{F(y_1 - m)^{n-1}}{n-1}\, dy_1 + F(m)^n.$$

Thus

$$P(Y_1 - Y_2 \leqslant m) = 1 - n\int_m^1 f(y)f(y-m)^{n-1} dy. \tag{1}$$

In theory, given knowledge of $f(x)$ (1) enables us to estimate the probability of a pennant race occurring in a sports league.

3. Comparison of Model Predictions to Actual Number of Major League Baseball Pennant Races: 1931–1980

Before we use (1) to prove that divisional play does indeed lead to more pennant races, we will show that (1) accurately predicts the number of pennant races that have occurred during the last 50 major league baseball seasons. Since the probability of a pennant race as given in (1) depends on the number of teams we analyzed the validity of (1) with respect to the following 3 sets of data (obtained from [3]).

Data Set 1—National and American League Pennant Races 1931–1960 ($n = 8$);

Data Set 2—National League Pennant Races 1967–1980 and American League Pennant Races 1969–1977 ($n = 12$);

Data Set 3—National League Divisional Races 1969–1980 and American League Divisional Pennant Races 1969–1977 ($n = 6$).[1]

To determine if (1) accurately predicted the number of pennant races occurring in each of the above data sets, we first had to determine $f(x)$. Since league standings usually are symmetric about a .500 winning percentage, we hoped that a normal distribution would closely fit the observed data. To test whether or not each season's data could be assumed to have been drawn from a normal distribution, we used the standardized range test (see [1, p. 8]). That is, for each major league season we computed

$$SR = \frac{\max X_i - \min X_i}{s_x}, \qquad \text{where} \quad s_x = \frac{1}{n-1} \sum_{i=1}^{i=n} (X_i - 0.5)^2,$$

where X_i is the fraction of games won during the season by team i. For the critical region of 5%, the hypothesis that the X_i's for a given league season were drawn from a normal population was rejected for only 4 of the 84 seasons included in our data set. Thus we felt that the X_i's for a single season could be assumed to be observations from a normal distribution. For each data set we set $\mu = .50$ and estimated σ as the sample standard deviation for the performances of the teams in the data set. In other words, if X_{ij} is the fraction of games won by team i during season j of a data set consisting of y seasons, then we estimated σ by

$$\sigma = \frac{1}{ny-1} \sum_{i=1}^{i=n} \sum_{j=1}^{j=y} (X_{ij} - 0.5)^2.$$

TABLE 1

Eight Team Leagues 1931–1960

m	Observed Frequency of Pennant Race of Size m	Cumulative Probability of Pennant Race of Size m as predicted from (1)*
1/154	8/60 = 0.133	0.099
2/154	16/60 = 0.267	0.190
3/154	21/60 = 0.350	0.274
4/154	25/60 = 0.417	0.353
5/154	31/60 = 0.517	0.424
6/154	32/60 = 0.533	0.490
7/154	35/60 = 0.583	0.551
8/154	41/60 = 0.683	0.606
9/154	44/60 = 0.733	0.654
10/154	45/60 = 0.750	0.698
11/154	45/60 = 0.750	0.737
12/154	48/60 = 0.800	0.771
13/154	55/60 = 0.917	0.802
14/154	56/60 = 0.917	0.829
15/154	56/60 = 0.917	0.852
16/154	56/60 = 0.917	0.872
17/154	58/60 = 0.967	0.891
18/154	50/60 = 0.983	0.907
19/154	60/60 = 1	0.921

Kolmogorov-Smirnov statistic $= |0.517 - 0.424| = 0.093$.

Critical value for $\alpha = 0.20$ is 0.138.

*Letting $f(x)$ be normal with $\mu = 0.50$ and $\sigma = 0.091$.

[1] The 1961–1968 seasons were not used because expansion changed the number of teams in each league several times, thereby limiting the size of the data set.

TABLE 2
Twelve Team League 1969–1980

m	Observed Frequency of Pennant Race of Size m	Cumulative Probability of Pennant Race of Size m as predicted from (1)*
0/162	1/20 = 0.05	0
1/162	5/20 = 0.25	0.137
2/162	6/20 = 0.30	0.258
3/162	9/20 = 0.45	0.367
4/162	11/20 = 0.55	0.461
5/162	11/20 = 0.55	0.546
6/162	12/20 = 0.60	0.622
7/162	15/20 = 0.75	0.687
8/162	16/20 = 0.80	0.739
9/162	16/20 = 0.80	0.781
10/162	17/20 = 0.85	0.821
11/162	17/20 = 0.85	0.853
12/162	18/20 = 0.90	0.879
13/162	19/20 = 0.95	0.900
14/162	19/20 = 0.95	0.919
15/162	19/20 = 0.95	0.936
16/162	20/20 = 1	0.951

Kolmogorov-Smirnov statistic $= |0.25 = 0.137| = 0.113$.
Critical value for $\alpha = 0.20$ is 0.231.
*Letting $f(x)$ be normal with $\mu = 0.50$ and $\sigma = 0.071$.

For data set 1 this procedure yielded an estimate of $\sigma = 0.091$ while for data sets 2 and 3 this procedure yielded $\sigma = 0.071$. Numerical integration was then used on (1) to predict the probability of a pennant race. A comparison with the actual results follows in Tables 1, 2, and 3. To test the hypothesis that the probabilities of a pennant race predicted by our model are consistent with the observed frequencies we use the Kolmogorov-Smirnov Test (see [5, p. 817]). If we let H_0 be the hypothesis that the observed frequencies of pennant races are consistent with the probabilities predicted by our model, then we see that for $\alpha = 0.20$ then H_0 would be accepted for the data of Table 1, Table 2, or Table 3.

From Tables 1–3 it is clear that the Kolmogorov-Smirnov Test indicates that all 3 sets of data are consistent with the hypothesis that the distribution of the size of pennant races is given by (1). With this in mind, we proceed to use (1) to determine how modifications in a league's divisional structure changes the expected number of pennant races.

4. Analytical Results

Buoyed by the fact that the model of §2 appears to do a very good job of predicting the probability of a pennant race, we turn to the proof of our main analytical result. Our result requires that $f(x)$ satisfy the following conditions.

$$f(0.5 + c) = f(0.5 - c), \qquad 0 \leqslant c \leqslant 0.5, \tag{2}$$

$$f'(0.5 + c) \leqslant 0 \quad \text{for } 0 \leqslant c \leqslant 0.5 \text{ and } f'(0.5 + c) \geqslant 0 \text{ for } -0.5 \leqslant c \leqslant 0, \tag{3}$$

$$f(0) = f(1) = 0, \tag{4}$$

$$F(1) = 1 \quad \text{and} \quad F(0) = 0. \tag{5}$$

Condition (2) simply states that the density function for a team's performance is

TABLE 3
Six Team League 1969–1980

m	Observed Frequency of Pennant Race of Size m	Cumulative Probability of Pennant Race of Size m as predicted from (1)*
1/162	6/40 = 0.150	0.108
2/162	11/40 = 0.275	0.208
3/162	13/40 = 0.325	0.301
4/162	16/40 = 0.400	0.385
5/162	19/40 = 0.475	0.459
6/162	22/40 = 0.550	0.529
7/162	24/40 = 0.660	0.592
8/162	26/40 = 0.650	0.648
9/162	29/40 = 0.725	0.695
10/162	33/40 = 0.825	0.737
11/162	34/40 = 0.850	0.777
12/162	35/40 = 0.875	0.811
13/162	35/40 = 0.875	0.841
14/162	36/40 = 0.900	0.866
15/162	37/40 = 0.925	0.889
16/162	38/40 = 0.950	0.907
17/162	38/40 = 0.950	0.923
18/162	38/40 = 0.950	0.936
19/162	39/40 = 0.975	0.947
20/162	40/40 = 1	0.957

Kolmogorov-Smirnov statistic $= 0.825 - 0.737 \leqslant = 0.088$.
Critical value for $\alpha = 0.20$ is 0.169.
* Letting $f(x)$ be normal with $\mu = 0.50$ and $\sigma = 0.071$.

symmetric about 0.5. In light of the studentized range test results in §3, an assumption that $f(x)$ is symmetric about 0.50 seems consistent with past data. Condition (3) simply states that $f(x)$ decreases as you move away from 0.5, an assumption that again seems reasonable in light of the studentized range test results of §3. Conditions (4) and (5) simply state that no team can win all (or none) of its games. Since no team since 1900 has won more than 0.763 of its games (1906 Chicago Cubs) or fewer than 0.249 (1935 Boston Braves) of its games during a single season, this seems reasonable.

Let the expected number of pennant races of closeness $\overline{m}$ in a league whose teams are divided into k divisions of size $n_1, n_2, \ldots, n_k$ be denoted by $E_{\overline{m}}(n_1, n_2, \ldots, n_k)$. Before stating and proving our main result we state a lemma which is needed in the proof of our main result.

LEMMA 1. *For* $2 \leqslant n_i \leqslant n-2$ *the function* $g_i(u) = u^{n_i - 1} - u^{n-1}$ *has the following properties*:

$$\max_{0 \leqslant u \leqslant 1} g_i(u) = g_i(\bar{u}_i) \quad \textit{where } \bar{u}_1 = \left(\frac{n_i - 1}{n-1}\right)^{1/(n-n_i)}; \tag{6}$$

$$\bar{u}_i > \tfrac{1}{2}; \tag{7}$$

$$\textit{for } 0 \leqslant u < \bar{u}, \quad g_i'(u) > 0 \quad \textit{and} \quad \textit{for } \bar{u} < u \leqslant 1, \quad g_i'(u) < 0. \tag{8}$$

PROOF. See Appendix.
We can now state and prove our main result.

THEOREM 1. *If* $n \geqslant 4$ *and* $f(x)$ *and* $F(x)$ *satisfy* (2)–(5), *then*

$$E_{\overline{m}}(n) \leqslant E_{\overline{m}}(n_1, n_2, \ldots, n_k) \tag{9}$$

TABLE 4

m	Number of Pennant Races in 12 Team Leagues Without Divisional Play 1969–1980	Number of Pennant Races in 12 Team Leagues With Divisional Play 1969–1980
3/162	9	13
6/162	12	22
9/162	16	29

holds for $0 \leqslant \overline{m} \leqslant g$ *and all* $(n_1, n_2, \ldots, n_k)$ *satisfying* $n_i \geqslant 2$ $(i = 1, 2, \ldots, k)$ *and* $n_1 + n_2 + \cdots + n_k = n$. *Furthermore the inequality is strict unless* $\overline{m} = 0$.

PROOF. See Appendix.

The following corollary to Theorem 1 follows almost immediately.

COROLLARY 1. *For any league the divisional structure yielding the maximum expected number of pennant races must have no divisions with four or more teams.*

PROOF. See Appendix.

Thus no divisional structure in which some one division has at least four teams can maximize the expected number of pennant races. Of course, Theorem 1 implies that two six team divisions do indeed have a larger expected number of pennant races than a single twelve team league. As Table 4 indicates, the past 12 years has borne this out.

To see how the expected number of pennant races depended on the number of divisions, we numerically integrated (1) (assuming $f(x)$ normal with $\mu = 0.50$ and $\sigma = 0.071$ as determined from the 1969–1980 data) to obtain Table 5.

Thus we see that each successive creation of a new division generates fewer new pennant races, and six two team divisions generate more expected pennant races than four three team divisions.

TABLE 5

Number of Pennant Races per League Season Predicted by (1)
(*using* 1969–1980 *data for* $n = 12$)

Number of Divisions	$m = 3/162$	$m = 6/162$	$m = 9/162$
One Twelve Team Division	0.37	0.61	0.78
Two Six Team Divisions	0.60	1.06	1.39
Three Four Team Divisions	0.75	1.38	1.87
Four Three Team Divisions	0.85	1.59	2.20
Six Two Team Divisions	0.87	1.73	2.52

5. Summary

We have shown that the probability of a pennant race in an n-team major league baseball league can be accurately predicted by the simple order statistic model presented in §2. The model of §2 was then used to prove that divisional play actually does increase the expected number of pennant races.

Appendix

THEOREM 1. *If* $n \geqslant 4$ *and* $f(x)$ *and* $F(x)$ *satisfy* (2)–(5), *then*

$$E_{\overline{m}}(n) \leqslant E_{\overline{m}}(n_1, n_2, \ldots, n_k) \tag{9}$$

holds for $0 \leqslant \overline{m} \leqslant g$ *and all* $(n_1, n_2, \ldots, n_k)$ *satisfying* $n_i \geqslant 2$ $(i = 1, 2, \ldots, k)$ *and* $n_1 + n_2 + \cdots + n_k = n$. *Furthermore the inequality is strict unless* $\overline{m} = 0$.

PROOF. Letting $m = \overline{m}/g$ (1) implies that (9) is equivalent to

$$k - \sum_{i=1}^{i=k} n_k \int_m^1 f(x)F(x-m)^{n_i-1}dx \geqslant 1 - n\int_m^1 f(x)F(x-m)^{n-1}dx$$

or $k - 1 \geqslant h(m)$ where

$$h(m) = \sum_{i=1}^{i=k} n_i \int_m^1 f(x)\left\{F(x-m)^{n_i-1} - F(x-m)^{n-1}\right\}dx. \tag{10}$$

Since

$$h(0) = \sum_{i=1}^{i=k} n_i \left[\frac{F(x)^{n_i}}{n_i} - \frac{F(x)^n}{n}\right]_0^1 = \sum_{i=1}^{i=k} n_i\left(\frac{1}{n_i} - \frac{1}{n}\right) = k-1,$$

it suffices to show that $h'(m) < 0$ holds for $m > 0$. Differentiating (10) with respect to m by applying Leibnitz' Rule for differentiating under the integral sign yields:

$$h'(m) = \sum_{i=1}^{i=k} n_i \int_m^1 f(x)\{(n-1)f(x-m)F(x-m)^{n-2}$$
$$-(n_i-1)f(x-m)F(x-m)^{n_i-2}\}dx.$$

Integrating this expression by parts (letting $u = f(x)$) and

$$dv = \left\{(n-1)f(x-m)F(x-m)^{n-2} - (n_i-1)f(x-m)F(x-m)^{n_i-2}\right\}dx$$

and using (4) and (5) yields

$$h'(m) = \sum_{i=1}^{i=k} n_i \int_m^1 \left(F(x-m)^{n_i-1} - F(x-m)^{n-1}\right)F'(x)\,dx. \tag{11}$$

By (11) the proof of the theorem will be complete if we show that for $2 \leqslant n_i \leqslant n-2$ and $m > 0$

$$\int_m^1 f'(x)\left\{F(x-m)^{n_i-1} - F(x-m)^{n-1}\right\}dx < 0. \tag{12}$$

For particular values of m and n_i we can demonstrate that (12) holds in the collectively exhaustive cases $F(1-m) < \bar{u}_i$ and $F(1-m) \geqslant \bar{u}_i$. To prove (12) in the case where $F(1-m) < \bar{u}_i$ note that (2) enables us to rewrite (12)

$$\int_0^{0.5} f'(0.5-z)\left\{g_i\left[F(0.5-z+m)\right] - g_i\left[F(0.5+z+m)\right]\right\}dz. \tag{13}$$

Since $F(1-m) \leqslant \bar{u}_i$, (8) implies that $g_i[F(0.5-z-m)] - g_i[F(0.5+z+m)] < 0$. Since (3) implies $f'(0.5-z) \geqslant 0$, it follows from (13) that $h'(m) < \int_0^{0.5} f'(0.5-z)(0)\,dz = 0$, which is the desired result.

To prove (12) in the case where $F(1-m) \geqslant \bar{u}$, define c by $F(0.5+c-m) = \bar{u}_i$. By (7) $m < c \leqslant 0.50$. Again utilizing (2), (12) may be rewritten as

$$\int_0^c f'(0.5-z)g_i\left[F(0.5-z-m)\right] - g_i\left[F(0.5+z-m)\right]dz$$
$$+\int_c^{0.5} f'(0.5-z)g_i\left[F(0.5-z-m)\right] - g_i\left[F(0.5+z-m)\right]dz. \tag{14}$$

By (8) and the definition of c it follows that for $0 \leqslant z \leqslant c$

$$g_i\left[F(0.5 - z - m)\right] - g_i\left[F(0.5 + z - m)\right] < 0.$$

Since $f'(0.5 - z) \geqslant 0$ follows from (3), it follows that the first term in (14) is strictly negative. To show that the second term in (14) is nonpositive, it suffices to show that for $c \leqslant z \leqslant 0.5$.

$$g_i\left[F(0.5 - z - m)\right] - g_i\left[F(0.5 + z - m)\right] \leqslant 0. \tag{15}$$

Since $(0.5 + z - m) \geqslant \bar{u}$, (8) implies that $g_i[F(0.5 + z + m)] \leqslant g_i[F(0.5 + z - m)]$. Thus (15) (and therefore (14)) will follow if

$$g_i\left[F(0.5 - z - m)\right] - g_i\left[F(0.5 + z + m)\right] \leqslant 0. \tag{16}$$

By the symmetry of $f(\cdot)$ (i.e. (2)) it follows that

$$1 - F(0.5 + z + m) = F(0.5 - z - m). \tag{17}$$

Letting $a = F(0.5 + z + m)$ (17) allows us to rewrite (16) as

$$g_i(1 - a) - g_i(a) \leqslant 0. \tag{18}$$

Since $F(1) = 1$ and $F(\frac{1}{2}) = \frac{1}{2}$, the proof of the theorem will be complete if we show that (18) is valid for $\frac{1}{2} \leqslant a \leqslant 1$. By the definition of $g_i(a)$ (18) reduces to

$$a^{n_i - 1} - a^{n-1} \geqslant (1 - a)^{n_i - 1} - (1 - a)^{n-1}.$$

Factoring both sides yields

$$a^{n_i - 1}(1 - a)(1 + a + a^2 + \cdots + a^{n - n_i - 1})$$
$$\geqslant (1 - a)^{n_i - 1}a\left(1 + (1 - a) + (1 - a)^2 + \cdots + (1 - a)^{n - n_i - 1}\right).$$

This inequality simplifies to

$$a^{n_i - 2}(1 + a + a^2 + \cdots + a^{n - n_i - 1})$$
$$\geqslant (1 - a)^{n_i - 2}\left(1 + (1 - a) + (1 - a)^2 + \cdots + (1 - a)^{n - n_i - 1}\right).$$

Since $n_i \geqslant 2$ and $a \geqslant 1 - a$, (18) is indeed valid and the proof of the theorem is complete.

Corollary 1. *For any league the divisional structure yielding the maximum expected number of pennant races must have no divisions with four or more teams.*

Proof. Suppose the divisional structure that yields the maximum expected number of pennant races has at least one division consisting of $k \geqslant 4$ teams. Then by dividing that division into two divisions consisting of 2 and $k - 2$ teams, respectively we can (by Theorem 1) increase the expected number of pennant races. Thus the maximum number of expected pennant races cannot be obtained by a divisional setup in which at least one division consists of four or more teams.

Lemma 1. *For $2 \leqslant n_i \leqslant n - 2$ the function $g_i(u) = u^{n_i - 1} - u^{n-1}$ has the following properties*:

$$\max_{u \in [0,1]} g_i(u) = g_i(\bar{u}_i) \quad \textit{where } \bar{u}_i = \left(\frac{n_i - 1}{n - 1}\right)^{1/(n - n_i)}; \tag{6}$$

$$\bar{u}_i > \tfrac{1}{2}; \tag{7}$$

$$\textit{for } 0 \leqslant u < \bar{u}, \quad g_i'(u) > 0 \quad \textit{and} \quad \textit{for } \bar{u} < u \leqslant 1, \quad g_1'(u) < 0. \tag{8}$$

PROOF. Since

$$g_i'(u) = u^{n_i - 2}\{n_i - 1 - (n-1)u^{n-n_i}\} \tag{19}$$

it follows that $g_i'(\bar{u}_i) = 0$. It is also clear from (19) that (8) is valid, so $\bar{u}_i$ indeed maximizes $g_i(u)$ on $[0, 1]$. To prove (7) consider the following optimization problem.

$$\max_{\substack{w = n - n_i \\ 2 \leqslant w \leqslant n-2 \\ n = 4,5,\ldots}} h(w), \qquad \text{where} \quad h(w) = \left(1 - \frac{w}{n-1}\right)^{1/w}. \tag{20}$$

Thus (7) will follow if we prove that the optimum value of $h(\cdot)$ in (20) is more than $\frac{1}{2}$. To show this note that since for all $w(\partial h(w)/\partial n) \geqslant 0$, the minimum in (6) must occur at an integer pair (n, w) satisfying $n = w + 2$. This implies that (7) will follow if we show that for $n \geqslant 4$

$$h(n-2) = \left(\frac{1}{n-1}\right)^{1/(n-2)} > \frac{1}{2}.$$

To prove this define

$$v(n) = \left(\frac{1}{n-1}\right)^{1/(n-2)}.$$

Then

$$\log v(n) = -\frac{\log(n-1)}{n-1}$$

and

$$\frac{d[\log v(n)]}{dn} = \frac{1}{n-2}\left(\frac{\log(n-1)}{n-2} - \frac{1}{n-1}\right).$$

Since

$$\log(n-1) \geqslant \frac{n-2}{n-1}$$

holds for $n \geqslant 3$, it follows that $\log v(n)$, and therefore $v(n)$, is an increasing function of n. Thus for $n \geqslant 4$

$$\left(\frac{1}{n-1}\right)^{1/(n-2)} \geqslant \left(\frac{1}{3}\right)^{1/2} = \frac{\sqrt{3}}{3} > \frac{1}{2}$$

and the proof of (7) is complete.

References

1. FAMA, E., *Foundations of Finance*, Basic Books, New York, 1976.
2. DAVID, H. A., *Order Statistics*, Wiley, New York, 1970.
3. NEFT, D. S., COHEN, R. M. AND DEUTSH, J. A., *The Sports Encyclopedia Baseball*, Grosset and Dunlap, New York, 1981.
4. *The Sporting News*, *Official Baseball Guide*, St. Louis, Mo., 1969.
5. WINKLER, R. L. AND HAYS, W. L., *Statistics*: *Probability*, *Inference and Decision*, Holt Rinehart and Winston, New York, 1971.