1-1-2002

Competitive Balance in Sports Leagues: An Introduction

Andrew S. Zimbalist
Smith College, azimbali@smith.edu

Competitive Balance in Sports Leagues

An Introduction

ANDREW S. ZIMBALIST
Smith College

Competitive balance is like wealth. Everyone agrees it is a good thing to have, but no one knows how much one needs.

Economic theory tells us that the optimal level of balance in a sports league is a function of the distribution of fan preferences, fan population base, and fan income across host cities. Profit maximizing teams will accumulate units of talent until the marginal revenue per win is equalized across all teams. This implies that in leagues with a fixed supply of teams (and monopoly or duopoly team rights to a territory), the league will maximize revenues when teams from large, rich, and fan-intense cities win more often.

In his seminal 1956 article, Rottenberg, among other things, anticipated the relevance of the Coase Theorem (Coase, 1960) in understanding talent distribution across teams and argued that the profit motive would limit the accumulation of player talent on any single team. El-Hodiri and Quirk (1971), in the first formal modeling of a professional sport league, find that individual team profit maximization is inconsistent with equal playing strengths among the teams except in the special case of identical team revenue functions.

Fort and Quirk (1995) focus on the problem of competitive balance in sports leagues and assess the degree to which different mechanisms create greater balance. Based on the Coase Theorem, they conclude that neither the reserve clause nor the reverse-order amateur draft aid balance. Explicitly assuming that gate revenue grows in proportion to the home team's playing strength, that each team faces the same cost per unit of playing strength, and implicitly assuming that owner risk aversion is invariant to team revenue and that cross subsidies are proportional to team revenue and not team rent, Fort and Quirk also conclude that increased revenue sharing will not improve competitive balance. The only mechanism in their

AUTHOR'S NOTE: *I would like to thank Lois Joy, Shreya Jain, and Bhavani Harimohan for research assistance, and Allen Sanderson, Brad Humphreys, Roger Noll, and Stefan Szymanski for comments on an earlier draft of this article.*

JOURNAL OF SPORTS ECONOMICS, Vol. 3 No. 2, May 2002 111–121

analysis that fosters balance is a salary cap, but a cap interferes with league-wide revenue maximization. Vrooman (1995), Marburger (1997), and Kesenne (2000) each vary the assumptions in the Fort and Quirk model and conclude that increased revenue sharing may indeed improve competitive balance.[1]

Needless to say, this conclusion will depend on the design of the revenue sharing system. If the system is similar to the one instituted by Major League Baseball (MLB) in 1996, providing an incentive for roughly the bottom half of teams to lowball payroll and reduce talent accumulation, then it is likely to exacerbate competitive balance, as the MLB system has (see Zimbalist, 2001a).

Daly and Moore (1981) challenge the conventional conclusion that the reserve clause and player draft do not affect the distribution of player talent. They argue that the combination of externalities and transactions costs limiting player sales meant, contrary to Rottenberg and the Coase Theorem, that the player draft and the reserve clause were effective equalizers of team playing strength.

In the present compendium, Marburger introduces another wrinkle into the Coase Theorem's applicability to sports leagues. He distinguishes between inter- and intraconference transactions for leagues with at least two conferences. The revenue impact of a player's movement is greater when he moves within a conference because it raises the strength of one team and lowers the strength of a rival team simultaneously. Marburger theorizes that when property rights are transferred to the player through free agency, there will be a higher share of intraleague transfers than previously because the player does not internalize the loss to the original club. Using probit estimation of intraleague transfer probabilities before and after free agency's onset in 1976 in MLB, Marburger generates empirical support for his theory.

MEASURING COMPETITIVE BALANCE

There are almost as many ways to measure competitive balance as there are to quantify the money supply. Among the more frequently used metrics are the standard deviation of win percentages (in a given year for a league or over time for a team), the ratio of the actual to the idealized standard deviation of win percentages, the ratio of the top to bottom win percentages, the range of win percentages, the gini coefficient of win percentages or of the concentration of championships, the Herfindahl-Hirschman Index of first-place finishes or championships, and excess tail frequencies.

Which is the right index or the right combination of metrics? The notion that competitive balance is important to sports leagues derives from an assumption that fans have a strong preference for uncertainty of outcomes or, as MLB commissioner Allan "Bud" Selig has put it, fans want to begin each season with hope and expectation. It follows that the best measure of competitive balance is the one to which the consumers show greatest sensitivity.

Many sports economists have used the concept of an idealized standard deviation of win percentages. This measure is based on the assumption that all teams are of equal strength. In a league of perfect parity, the standard deviation of win percentages would be

$$\sigma = 0.5/\sqrt{N},$$

where N is the number of games played during the season by each team. Statistically, it makes sense to compare this number to the actual standard deviation of win percentages because it standardizes for the length of a sport's season.

The problem with this index of competitive balance, however, is that fans do not experience it. If a fan of the Carolina Panthers is told that he or she should not be distressed that the team went 1 and 15 in 2001-2002 because the sample of games was too small to reveal the true strength of the team, it is likely that he or she would be unimpressed.

One meaningful way to distinguish among balance measures is to test how effective they are as predictors of attendance, controlling for relevant factors. In his piece in this volume, Brad Humphreys suggests a new index of competitive balance and then tests its explanatory power. Humphreys reasons that fans care as much about performance mobility of teams over time as they do about the imbalance across teams in a particular year. He devises a metric that reflects both these criteria and finds that it explains more of the variation in baseball attendance during the past 100 years than alternative measures of balance.

SPORT BY SPORT

Baseball

In its study on the state of baseball's economics (MLB, 2000), the Blue Ribbon Panel suggested several qualitative or normative definitions of competitive balance. The clearest statement appears on page 5 of the report:

> In the context of baseball, proper competitive balance should be understood to exist when there are no clubs chronically weak because of MLB's financial structural features. Proper competitive balance will not exist until every well-run club has a regularly recurring hope of reaching postseason play.

More recently, the commissioner's office has attempted to measure competitive imbalance in a different way. In December 2001, the office conducted a poll of 1,000 fans and discovered that 75% agreed that there is a lack of competitive balance in the game and 42% say that they will be less interested in baseball if it is not improved. Even if the survey was conducted properly, it is not clear what "less inter-

ested" means in terms of fan behavior. Nor is it clear how much improvement fans want to see or what changes would denote improvement.

The commissioner's office has also issued reports detailing the lack of balance since 1995. The numbers prima facie are compelling. Consider, for instance, this one: During the 7 years from 1995 through 2001, only four teams from the bottom half of payrolls reached the postseason, and combined they won a total of 5 of 224 games (MLB, 2001).

The article by Hall, Szymanski, and Zimbalist in this issue explores the statistical correlation between payrolls and performance in baseball and English soccer during the past 20 years. The yearly correlation in soccer is statistically significant and robust, whereas in baseball it is only strong since the mid-1990s. Because correlation does not denote causality, the authors perform Granger causality tests for each sport. During 1980-1994, baseball shows no evidence of Granger causality from payroll to performance. Since 1995 there is evidence that causality runs in both directions. In English soccer, the evidence suggests that payroll Granger causes performance during 1974-1999. The authors attribute the differing results to the more open players' market in English and world soccer.

Many observers, although acknowledging that baseball has been imbalanced since 1995, do not see a problem. They argue that baseball has had imbalance and dynasties in the past, yet it has continued to thrive. The present Yankee dynasty, for instance, is compared with the one between 1949 and 1965.

The last period of Yankee dominance in the 1950s and early 1960s, however, was not a time that baseball should seek to emulate. Between 1950 and 1965, average attendance at games grew by a total of less than 3%, even though real ticket prices remained virtually flat and real disposable incomes in the United States grew by 74%. Moreover, four and five decades ago baseball stood alone on the pedestal of popular team sports. Today it is challenged by the National Basketball Association (NBA) and the National Football League (NFL), as well as by the growing list of new professional sports and entertainment options ranging from the National Association for Stock Car Auto Racing (NASCAR), to the World Wrestling Federation, to the Internet.

That is, even if baseball was able to stabilize its attendance during the past Yankee dynasty, it is not clear that the sport can afford another protracted dynasty today. I tested the following model with panel data separately for the periods 1950-1965 and 1985-2000:

$$ATT = f(WPC, WPC_{t-1}, PRICE, STAD, TEAM\ DUMMIES, TREND),$$

where *WPC* is win percentage, WPC_{t-1} is win percentage lagged 1 year, *PRICE* is each team's yearly average ticket price, *STAD* is a dummy variable equaling 1 if the team was playing in a stadium built within the previous 3 years, *TEAM DUMMIES* control for team fixed effects, and *TREND* is a variable equaling 1 for the 1st year, 2 for the 2nd year, and so forth. The Cochrane-Orcutt iterative technique was used to

TABLE 1

	1950-1965	*1985-2000*
R^2	.507	.544
Estimated coefficient on *WPC*	2.40	3.33
T-statistic on *WPC*	9.21	12.07
Estimated coefficient on WPC_{t-1}	0.27	1.75
T-statistic on WPC_{t-1}	1.06	6.34

NOTE: *WPC* = win percentage.

correct for autocorrelation. The key results, which are robust for various specifications (including logs), are given in Table 1.

According to these results, during 1950-1965 one additional win increased the average team's yearly attendance by 14,819. During 1985-2000 one additional win increased the average team's yearly attendance by 31,325.[2] Alternatively, using the average ticket prices, one additional win was worth an average of $28,452 to annual team revenue during 1950-1965, whereas it was worth an average of $315,443 during 1985-2000 (both figures in current dollars). These results support the view that attendance (and revenues) today is more sensitive to team performance than it was 40 and 50 years ago.

Ice Hockey

Both in terms of the standard deviation of win percentages and the ratio of actual to idealized standard deviations, the National Hockey League (NHL) stands in the middle of the four major team sports leagues in the United States. For the 1990s, the pay-performance link is significant at the 5% level, but it is weaker ($R^2 = .11$) than in the English Premier League, MLB, or the NBA.

An interesting pattern emerges in hockey's pay-performance link. I obtained data for opening day and trade deadline payroll (around two thirds of the way through the season) for each team during 1995-2000 from the hockey players' association. The correlation coefficients between team win percentage and team payroll are shown in Table 2.[3]

The pattern is clear: Mid-season payrolls are consistently more closely correlated with win percentages than beginning payrolls. This result suggests that causality runs in both directions between pay and performance. Higher pay creates better teams, but better teams create higher pay. That is, teams performing well in mid-season make an effort (by acquiring new players) to bolster their rosters to increase their chances for regular and postseason success. Of course, the causality from performance to pay may also appear between seasons as winning teams have to spend more to retain their successful players.

TABLE 2

	Correlation Coefficient	
	Opening Day Payroll	*Trade Deadline Payroll*
1995	.465	.583
1996	.341	.489
1997	.338	.431
1998	.237	.238
1999	.499	.593
2000	.471	.632

Interestingly, another pattern has emerged in the NHL since 1990: The dynasties that dominated the sport for decades have not returned. Consider the following record of dynasties according to who won the Stanley Cup: Montreal, 1956-1960; Toronto, 1962-1964; Montreal, 1976-1979; New York Islanders, 1980-1983; and Edmonton, 1984-1985, 1987-1988, and 1990. Since 1990, no team has won the Cup more than 2 years in a row, and eight different teams have won the Cup.

By the metric of concentration of championships, then, hockey appears to be more balanced since 1990. The traditional policies related to competitive balance—revenue sharing or salary cap/luxury tax—however, have not been implemented in hockey. What might account for this change?

Since the advent of Bob Goodenow as executive director of the NHL Players' Association in 1991, players have followed the policy of salary disclosure. Together with their salary arbitration system, this disclosure policy has helped boost player salaries appreciably. Furthermore, the new collective bargaining agreement in 1995 brought the NHL its first system of unrestricted free agency for players older than 31 years of age. This lifted the compensation of veteran star players still more. It seems plausible to hypothesize that the resulting rapid growth in hockey player salaries in the 1990s, in the context of veteran free agency and arbitration, has made it more and more expensive to hold together winning teams—and easier to resurrect poor-performing teams.

Basketball

The NBA has been on the high end of most measures of imbalance. The league practices no revenue sharing other than income from the national television contract and NBA properties, and it has had a salary cap since 1983, albeit a porous one, at least until 1999. Indeed, the standard deviation of win percentages actually rose by 14.5% during the 1980s, the decade when the cap was introduced, and has continued to increase since 1990. Furthermore, by a considerable margin, the NBA has the highest ratio of actual to idealized standard deviation of win percentages among

U.S. leagues. Finally, the NBA has an intermediate link between payroll and performance. In a simple regression of win percentage on payroll during 1986-2000, the coefficient on payroll is significant at the .01 level ($R^2 = .16$), and in yearly regressions the closeness of fit displays no particular trend (with R^2s varying between .07 and .40).

The concentration of championships is high, especially since 1990.[4] The Chicago Bulls were crowned champions from 1991 through 1993; when Michael Jordan tried his hand at baseball, the Houston Rockets won in 1994 and 1995; and with Jordan back, the Bulls won from 1996 through 1998, followed by the San Antonio Spurs in 1999 and the Los Angeles Lakers in 2000 and 2001.

Given the strong statistical evidence of imbalance in the NBA, one might expect to see some effect in the form of diminished fan support. Some might argue that such a diminution has, in fact, transpired. The difficulty, of course, is attributing the reduced attendance and television ratings of recent years to competitive imbalance, as opposed to the effect of the 1998-1999 lockout, the temporary retirement of Michael Jordan, team and league pricing policies, the emergence of new entertainment alternatives, or general economic conditions.

To be sure, it might rather be argued that, despite the concentration of championships, strong consumer support during the years has been maintained because fans perceive fairness in the way that basketball's competition is structured. In particular, few would contest the proposition that the major reason for the Bulls dominance during the 1990s had to do with their good fortune in drafting Michael Jordan. Because of the Larry Bird exception (which allows a team to resign one of its own players without regard to the salary cap), teams tend to be able to keep their top players. In 1997-1998, for instance, 7 of the 11 top-paid players in the NBA still played for their original teams. Thus, if Michael Jordan had originally signed with Milwaukee or Utah instead of Chicago, he probably would have played in those cities throughout his career.

Football

As a function of playing only 16 regular-season games per season, the NFL has consistently had the highest actual standard deviation of win percentages among U.S. team sports leagues; yet it has also had the lowest ratio of actual to idealized standard deviations.[5] It has also had a remarkably low concentration of championships: No team has ever won three consecutive Super Bowls.

The NFL employs several policies to promote competitive balance: extensive revenue sharing (around 70% of all league revenues is divided equally among the teams), a relatively hard salary cap, a reverse-order draft,[6] and an unbalanced schedule. The consequence of these policies is that the ratio of top to bottom team revenue does not exceed 1.9 to 1, and the ratio of top to bottom team payroll does not exceed 1.5 to 1.

Revenue sharing in the NFL goes back to the league's inception in 1920 and the reinforcing policies of the commissioner, Pete Rozelle, in the early 1960s. It is ingrained in the NFL's contracts and its culture; hence, however desirable it might seem to some, it cannot be readily adopted by other leagues.

To be sure, many believe that the NFL has gone too far with its revenue sharing. Econometric work has made it clear that there is little, if any, profit incentive to win in the NFL. It remains to be seen whether this feature will eventually strain the fan's perception of the legitimacy of the league's competition.

English Soccer

The article by Noll in this compendium represents an important effort to analyze the economics of leagues characterized by promotion/relegation systems. The basic characteristic of these leagues is that successful teams rise to higher leagues and unsuccessful teams fall to lower leagues. From this, as Noll analyzes, many tendencies ensue: stronger demand, higher salaries, better players, less expansion, and more competitive balance, among others.

The greater expected degree of competitive balance in promotion/relegation leagues emerges from several factors. Perhaps most intriguing among them is that because league membership is not fixed, teams do not have monopoly control over their territories. Any individual with sufficient capital and drive can create a lower division team, invest in top players and coaches, and the team can potentially rise up the divisional ladder to the top level.

As a result of this characteristic, promotion/relegation leagues tend not to have monopolies or even duopolies in large cities. London has had 10 different teams playing in the top-level English Premier League (PL) since 1990.[7] If a city has too few teams relative to its economic potential to support teams, then entrepreneurs are free to increase supply without paying the existing league for the permission to become an expansion team. Thus, London PL teams do not enjoy the rent from the large market that is bestowed, say, on baseball's Yankees.

Since 1992-1993, Manchester United (MU) has dominated the English PL. MU won the league championship 7 of the 9 years between 1992 and 2000. During the 26-year stretch from 1974 to 1999, the pay-performance link in the PL was stronger (R^2 = .34) than it was for any U.S. league.[8]

Prima facie, one might expect there to be a negative fan reaction to this concentration at the top along with the apparent ability for a team to buy success. Yet the era of MU dominance has coincided with a growth spurt in soccer's popularity in England. After going through four decades when average attendance per game fell by more than 50% in English soccer, attendance at PL games shot up 60.9% between 1992 and 2001 despite rapidly rising ticket prices.

England's improved economy, new policies regarding crowd control at stadiums, John Major's love of the game replacing Margaret Thatcher's disdain, and increased televising of games, among other factors, all contributed to soccer's ascen-

dancy in the 1990s, but it appears that the growing mystique of MU played a role as well. A dominant team is not necessarily bad for a sport if fans perceive the team's success to be legitimately gained. In the case of MU, the team did not benefit from the rent of being the only team in England's largest media market. On the contrary, with England's compressed geography, there are approximately 50 professional league soccer teams within 100 miles of the MU stadium. Rather, MU's success seems to have its roots in an aggressive, well-executed management strategy.[9] The knowledge that poorly run teams can face demotion irrespective of their home city must also contribute to the sense of fairness of outcome.

Of course, another part of the legitimacy function is the fans' confidence that the league is producing the highest possible quality of product. In the long run, the best guarantor of this is competitive labor markets. Promotion/relegation systems provide strong incentives for quality; purported single-entity structures, such as the U.S. Major League Soccer (MLS), do not. Ultimately, successful leagues must strive for both relative and absolute quality.

CONCLUSION

The complexity of factors affecting competitive balance is daunting. Sanderson's piece in this issue underscores this point. Sanderson discusses the host of nonpecuniary influences (e.g., technology, demography, artificial enhancement, playing rules, and field conditions) that affect the ability of sports to provide balanced competition.

Fan perceptions and behavior are at the core of the competitive balance problematic. Humphreys's article is a model for the research that needs to done. To wit, measures of balance need to be tested against fan behavior for different sports. Logically, fans respond not only to the statistical outcomes but also to the processes and institutional features of competition in the various sports. Moreover, the demography of fan bases (age, gender, race, and income) differ from sport to sport.

Matters are further complicated by the variety of ownership objective functions. Whereas some team owners seek simple profit maximization in the management of their franchises, others use their franchises to maximize returns on their entire portfolio of assets and/or to gain consumption benefits. When George Steinbrenner recently opened his wallet again to sign Jason Giambi, he was thinking not only about Giambi's expected contribution to revenue of the Yankees but also about the successful launch of his new YES sports network (estimated to be worth $900 million). Thus, Steinbrenner's evaluation of Giambi's marginal product is assessed much differently than that of many other owners. Similar situations involve the owners of the Dodgers, Rangers, Red Sox, Cubs, and Braves. Given this and other related party transaction issues in baseball and other sports, owner behavior in the players' market might be relatively unresponsive to certain revenue sharing schemes.

The need for competitive balance has been used as an all-purpose justification for competitive restraints in antitrust cases in the United States and Europe. Given

the apparent ambiguities in identifying the nature and scope of the problem, there is good reason for this justification to receive close scrutiny in the future.

There are no easy formulas for competitive balance success. Economic research needs to reflect the heterogeneity of sports environments. Hopefully, the articles in this collection will move us in the right direction.

NOTES

1. Other early contributions to the competitive balance debate appear and are discussed in Zimbalist (2001b).

2. Of course, the effect in percentage terms is smaller. Whereas one win during the first period raises attendance by 1.3%, one win during the second period raises attendance by 1.5%.

3. I also ran the regressions using points (based on two points per win and one point per tie) instead of win percentages. The correlations were a bit stronger but revealed an identical pattern.

4. The National Basketball Association (NBA) has experienced earlier periods of single-team dominance as well, but not since the 1960s did a team repeat 3 years in a row as champions. Between 1959 and 1966, the Boston Celtics won eight consecutive championships, then lost in 1967 but won again in 1968 and 1969. What is notable about the dominance of the Bulls during the 1990s is that it occurred during the salary cap era.

5. For a discussion of competitive balance measurements across U.S. team sports as well as English and Italian soccer, see Szymanski (2001).

6. The other U.S. team sport leagues also employ a reverse-order draft. However, the NBA, to avoid the possibility that poorly performing teams perceive an incentive to lose games during the latter part of the season to gain a higher draft pick, has instituted a modification of the standard reverse-order picking system. In the NBA system, all teams not qualifying for the postseason enter a lottery to determine the order of picks among them. In recent years, the NBA has further adjusted its system to give lower finishing teams a higher probability of success in the lottery.

7. The Premier League (PL) was so named in 1992-1993. Before that, it was simply called Division I, today's Division I was called Division II, and so forth.

8. This is in contrast with an $R^2 = .23$ in U.S. baseball during 1980-2000, the highest R^2 for any U.S. league by a significant margin. The PL is also characterized by a relatively low standard deviation of win percentages of .11, and with 38 games per regular season, an extremely low ratio of actual to idealized standard deviation of 1.22.

9. For an excellent discussion of Manchester United (MU) and other aspects of the English soccer industry, see Szymanski and Kuypers (2000).

REFERENCES

Coase, R. (1960, October). The problem of social cost. *Journal of Law and Economics*, *3*, 1-44.

Daly, G., & Moore, W. J. (1981, January). Externalities, property rights and the allocation of resources in Major League Baseball. *Economic Inquiry*, *19*(1), 77-95.

El-Hodiri, M., & Quirk, J. (1971, November/December). An economic model of a professional sports league. *Journal of Political Economy*, *79*(6), 1302-1319.

Fort, R., & Quirk, J. (1995, September). Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature*, *33*(3), 1265-1299.

Késenne, S. (2000, February). Revenue sharing and competitive balance in professional sports. *Journal of Sports Economics*, *1*(1), 56-65.

Major League Baseball. (2000, July). *Report of the independent members of the commissioner's Blue Ribbon Panel on Baseball Economics*. New York: Author.

Major League Baseball. (2001, December). *MLB updated supplement to the report of the independent members of the commissioner's Blue Ribbon Panel on Baseball Economics*. New York: Author.

Marburger, D. (1997). Gate revenue sharing and luxury taxes in professional sports. *Contemporary Economic Policy*, *15*(2), 114-123.

Rottenberg, S. (1956, June). The baseball players' labor market. *Journal of Political Economy*, *64*(3), 242-258.

Szymanski, S. (2001). *The economic design of sporting contests: A review* (Working Paper). London: Imperial College Management School.

Szymanski, S., & Kuypers, T. (2000). *Winners and losers: The business strategy of football*. London: Penguin.

Vrooman, J. (1995, April). A general theory of professional sports leagues. *Southern Economic Journal*, *61*(4), 971-990.

Zimbalist, A. (2001a). Competitive balance in Major League Baseball. *Milken Institute Review*, *3*(1), 54-64.

Zimbalist, A. (Ed.). (2001b). *The economics of sport*. London: Edward Elgar.

Andrew S. Zimbalist is a professor of economics at Smith College in Northhampton, Massachusetts.