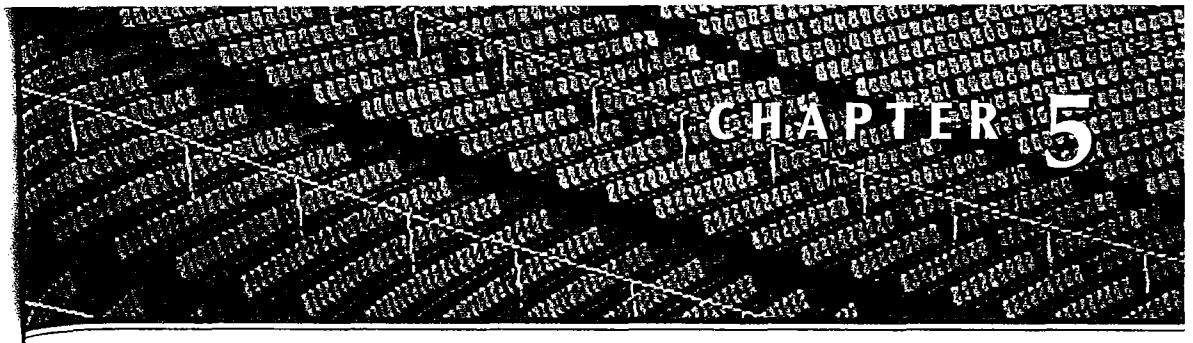
Your optimal strategy should be to serve to her forehand just often enough that Serena does not gain an advantage by guessing one way or the other. In other words, your best strategy is to choose p so that $\text{Prob}_{GF} = \text{Prob}_{GB}$. Setting these equations equal and solving for p, one finds that your optimal strategy would be to serve to Serena's forehand one-third of the time ($p = 1/3$) and to her backhand two-thirds of the time. Then, no matter how she guesses, you will win 53.3 percent of the points.[65]

[65]Similar reasoning will tell Serena her optimal guessing strategy—that is, what percentage of the time (q) she should prepare for a serve to her forehand. If either you or Serena deviates from your optimal strategy, the other will see this (i.e., you can tell how often Serena prepares for a forehand serve [q] , and she can tell how often you serve to her forehand [p]). Then if, say, Serena sees you are not following the best strategy of serving to her forehand one-third of the time, she can adjust her strategy to reduce your winning percentage below .533.



CHAPTER 5

Competitive Balance

When you lose a couple of times, it makes you realize how difficult it is to win.
— STEFFI GRAF (GERMAN TENNIS PLAYER)[1]

One of the oldest adages in professional football is that on any given Sunday, each team has a chance to beat the other. But what if, year after year, some teams almost always lost while other teams almost always won? No doubt sports would be less interesting. As early as 1956, economists noted that successful leagues must be based on relatively even competition.[2] The degree of parity in a league is known as competitive balance. This chapter discusses competitive balance from the perspective of the fan and the owner. In addition, it explores the various ways in which economists measure competitive balance, how leagues may try to alter the competitive balance in a league, and why such efforts may not be successful. As we explain competitive balance, we will see the following:

- Concerns over competitive balance are not new
- How competitive balance in U.S. sports leagues compares to leagues in other nations
- Why owners want competitive balance but not complete parity
- Why any single measure of competitive balance cannot capture all of its important dimensions
- Why the draft does not necessarily equalize talent across teams

[1]www.famous-quotes-and-quotations.com/sports_quotes.html

[2]Simon Rottenberg, "The Baseball Players Labor Market," *Journal of Political Economy*, v. 64, no. 3 (June 1956), pp. 242–258.

1 The Fan's Perspective

Suppose you are an exchange student to the United States, and your host family takes you to see your first baseball game; it's between the Tampa Bay Devil Rays and the New York Yankees. The game quickly gets boring because the talent on the two teams is so uneven. The Yankees score on the hapless Devil Rays over and over again, and you notice that most fans leave well before the game is even close to being over. The final score is 11–0. "That's OK," your host says, "the Devil Rays lose all the time, and the Yankees always win." If that were your only exposure to baseball, you would probably leave America thinking baseball was a waste of time. If instead you had seen the New York Mets beat the St. Louis Cardinals in a wild 5–4 game after the catcher hit a home run in the bottom of the ninth, and you learned that almost all games are like this, you might become a lifelong fan.

From the fan's perspective, an uncertain outcome is much more interesting than a foregone conclusion. While some fans may argue that the degree of parity in the current NFL is too great, historically, American fans have shown their displeasure with unbalanced competition, even when their own team did most of the winning. An often-cited example is the Cleveland Browns of the late 1940s; their continued dominance of the All-American Football Conference caused them to become less popular with their home fans. In baseball, the Yankees may have had the same kind of negative effect on attendance at their own games and across the American League when they won eight League pennants and six World Series between 1950 and 1958. Table 5.1 shows that between 1950 and 1958, a period generally marked by prosperity and growth, attendance at both Yankee games and those of the entire American League either stagnated or fell as the Yankees completely dominated

TABLE 5.1

New York Yankees' Success and American League and National League Attendance, 1950–1958

Year	*AL Champion*	*World Series Champion*	*Yankees Attendance*	*AL Attendance*	*NL Attendance*
1950	Yankees	Yankees	2,081,380	9,142,361	8,320,616
1951	Yankees	Yankees	1,950,107	8,888,614	7,244,002
1952	Yankees	Yankees	1,629,665	8,293,896	6,339,148
1953	Yankees	Yankees	1,531,811	6,964,076	7,419,721
1954	Cleveland	NY Giants	1,475,171	7,922,364	8,013,519
1955	Yankees	Brooklyn	1,490,138	8,942,971	7,674,412
1956	Yankees	Yankees	1,491,784	7,893,683	8,649,567
1957	Yankees	Milwaukee	1,497,134	8,169,218	8,819,601
1958	Yankees	Yankees	1,428,438	7,296,034	10,164,596

Source: Attendance data are from Rodney Fort and James Quirk, *Pay Dirt* (1992). Performance data is from the official MLB Web site http://www.MLB.com.

professional baseball. The effect was especially pronounced in the late 1950s, as National League attendance grew substantially while American League attendance fell.

In fact, fans enjoy a contest with an uncertain outcome even though they root for their team to win every game. Recent research shows that fans are most interested in games when the home team has a 60 to 70 percent chance of winning.[3] This is not to say that fans want their teams to lose; they want them to have *a chance of losing*. If fans were certain that their team would win every week, it would take away a major source of excitement from the game. Economists call this the **uncertainty of outcome hypothesis (UOH).**

There is currently a great deal of research on UOH in sports economics, much of which is centered on baseball and the ongoing concern among both fans and team owners, who recently sponsored a blue ribbon commission to study the subject. A prominent fear is that the Yankees are so wealthy relative to other teams that they can "buy" championships year after year by offering salaries high enough to attract all of the best players. These concerns have been heightened by the fact that the Yankees now have their own television network, YES (Yankees Entertainment & Sports). If the Yankees can make tens of millions of dollars from selling YES broadcast rights to cable companies, they might use that revenue to purchase even more top players. A review of team revenues seems to confirm these fears. Based on data from *Forbes*, the Yankees and Red Sox had revenues of $264 and $201 million in 2005, while the Pirates, Devil Rays, Nationals, Blue Jays, Marlins, Royals and Twins all had revenues of $110 million or less.[4] Although we know from previous chapters that high revenues do not necessarily mean large profits, the on-field success of the Yankees and Red Sox in recent years fuels the fear that championship-caliber teams require revenues at or near the top of the league.

A similar concern was expressed during the Chicago Bulls' dominance of the NBA in the Michael Jordan era, when the Bulls won six out of eight championships between 1991 and 1998. More recently the Los Angeles Lakers and San Antonio Spurs have combined to win six of the last eight championships. Perhaps the same concerns are now warranted in the NHL, as just three teams—the New Jersey Devils, Colorado Avalanche, and Detroit Red Wings—have won eight of the last eleven Stanley Cups.

Part of the debate over competitive balance focuses on whether dynasties are good or bad for a sport. Do fans like dynasties? It depends on whether they support the team that is consistently winning championships or support the team that has little chance of winning one. Dynasties of the past do tend to be

[3]Daniel Rascher, "A Test of Optimal Positive Production Network Externality in Major League Baseball," in *Sports Economics: Current Research*, ed. by John Fizel, Elizabeth Gustafson, and Larry Hadley (Westport, Conn.: Praeger Publishers, 1999), pp. 27–45.

[4]See Michael K. Ozanian (ed.), "The Business of Baseball," at www.forbes.com/2005/04/06/05/mlbland.html.

San Antonio Spurs celebrate winning the NBA championship.

remembered more romantically by fans in all cities, in part because of the legendary players who made the teams so dominant.

Despite the flurry of recent research on the topic and activity at league meetings to "fix the problem," unbalanced competition is not a new issue. The dominance of the Yankees extends back to the 1920s, when the Yankees won six American League championships between 1921 (the year after they acquired Babe Ruth) and 1928. It was even more pronounced in the late 1940s and 1950s, when they won five straight World Series, eight in total between 1947 and 1958. Although there were fewer teams at the time—a fact that makes the current Yankees' success more impressive—the long history of championship dominance indicates that the current Yankees teams are no more, and may be less, dominant than past Yankees teams.

Two of the three other major sports have similar histories. The Boston Celtics won every NBA championship but one between 1959 and 1969. Between 1965 and 1979, the Montreal Canadiens won the NHL's Stanley Cup 10 times. The Canadiens' dynasty was followed by that of the New York Islanders, who won the Cup the next four years in a row. Only in the NFL has no team ever won the league championship—the Super Bowl—more than twice in a row, but even the NFL has several franchises that are historically uncompetitive, such as Cincinnati and Arizona.

Internationally, unbalanced competition in the elite European soccer leagues is even more skewed toward a few dominant teams. In the top Italian league, 13 of the 15 championships between 1992 and 2006 were won by just two teams: Juventus and AC Milan. In England, Manchester United has won

the Premier League eight times since 1992–1993, while Real Madrid and FC Barcelona have dominated La Liga, the top Spanish league, over the same time period.

Changes in the relative importance of the various revenue sources and the growth of the sports industry in general have increased the concerns that the financial consequences of unbalanced competition are becoming more severe. Accordingly, we must consider the owners' perspective on equalizing competition.

5.2 The Owners' Perspective

As discussed in Chapters 2 and 3, leagues have long been aware that their success depends on staging games with an uncertain outcome. Whether it is a great boxing match between two undefeated rivals, a tennis championship between the first and second-ranked players in the world, or a baseball game in which the league's top two pitchers square off, fans' demand typically increases with the intensity of the contest. The more appealing the contest, the more fans will attend or watch the game on TV, and the more revenue will be generated.

If it is in the best interest of leagues to have relatively close competition between their member teams, they have an incentive to promote competitive balance. Leagues do not need to take specific action if they tend naturally toward equal strength. If, however, a few teams flourish while most teams languish, the league has an incentive to act. For example, in auto racing, leagues such as NASCAR go to extensive lengths to promote equal competition between cars. Each car in a NASCAR race is measured using a series of 30 or more templates and scales for height, weight, length, width, air displacement, and overall shape. In addition, engines must meet an exacting set of criteria and are even restricted to reduce horsepower on larger tracks where speeds are greatest. To further ensure fair competition, some cars are retested at the conclusion of the race. By placing so many restrictions on the cars, racing leagues hope to ensure close competitions decided by the skills of the drivers and their teams. Chapters 3 and 4 explained how teams in larger, more populous markets generally have a larger fan base and higher gate and television revenues. They can thus afford to hire better players in a free market. In addition, successful teams are likely to have an advantage in attracting players, thereby creating self-perpetuating dynasties. Because all North American leagues now have free agency in one form or another, players can decide for whom they would most like to play. For example, veteran players near the end of their careers such as Karl Malone and Gary Payton in basketball, Dominik Hasek in hockey, or Roger Clemens in baseball, all made conscious decisions to play for top contending teams.

If dynasties are self-perpetuating, then leagues have an economic incentive to intervene in the market to ensure enough competitive balance to foster long-term profitability. Although a league's monopoly power affords it the ability to do so, a potential complication stems from the fact that, in the end, team owners control leagues. As a group, they would like to see balanced competition, but each individual owner would also like his or her team to be consistently successful, a point we return to later in the chapter.

One strictly economic force is present in all labor markets that mitigates the effects of unbalanced revenues across teams. The **law of diminishing marginal returns** (diminishing returns) reduces the incentive of any team to stockpile all of the top talent in a league, though this force may not be sufficient to prevent competitive imbalance. Diminishing returns to labor are found in every industry. In the short run, as a firm adds units of labor, the marginal product (the additional output) of the last unit of labor must eventually fall, even if the labor is homogeneous. The reason is straightforward: In the short run, capital is fixed. Thus, eventually, the additional workers have insufficient capital to work with, so they are not as productive.

In the context of sports, diminishing returns may set in very quickly, especially in basketball, where only five team members play at a time, substitution is relatively limited (i.e., the five starting players play the vast majority of the total minutes), and, as the saying goes, "there is only one ball." In most sports, players specialize in particular positions, and rules allow only a fixed number of players on the field at one time. For basketball, once a team has even two players who shoot frequently, adding a third shooter to the roster is likely to add very little to team quality, certainly less than the addition of the first two scorers. The Miami Heat would not dispute that Tim Duncan is a great player, but his value to the Heat is surely less than the price another team without a top center would pay, given that the Heat already have Shaquille O'Neal. Duncan's value will be greater to a team that does not already have a top scoring center or power forward. Therefore, while some baseball aficionados may claim that a team can never have too much pitching, a team with 15 pitchers certainly *would* have too much pitching, since rosters are fixed at 25 players, so that team would have almost no substitutes at other positions and some of those pitchers would rarely play.

Thus, teams have an incentive to allow talent to spread across their league—it simply doesn't make economic or strategic sense for a single team to have all or even most of the good players at any given position. Teams may benefit from stockpiling talent to prevent rivals from signing available stars, but with fixed roster sizes, the ability to do so is limited. The same type of restriction is imposed at the Division I level of collegiate football; the number of scholarships a team can offer is limited to 85.

The Effect of Market Size

Differences in market size across the league provide an additional challenge to team owners. If, as research shows, the dollar value of a win is greater to teams

in large cities than to teams in smaller ones, maximizing competitive balance and maximizing total league profits may not be consistent goals.[5] Even if fans desire some level of uncertainty, a profit-maximizing league may prefer to have the teams in the largest markets win more often than teams elsewhere. In a 30-team league, perfect parity would mean that the Yankees and Dodgers—teams in the two largest markets—would win the World Series only once every 30 years. If championships were allocated so that they were distributed equally on a per-fan basis, rather than a per-team basis, the large-market teams would win more frequently than once every 30 years because they have so many more fans than small-market teams. The tension here between individual team profits and overall league profits is similar to that of a cartel. Each individual team can increase its profits by improving relative to the rest of the league, but from the league perspective, it is better if some teams are more successful than others.

To see why big-market teams benefit more from winning than small-market teams do, assume that each team gets its revenue only from tickets and local television revenue. Assume further that teams benefit from having a higher winning percentage, but the additional benefits of increasing the winning percentage become smaller as it approaches 1.000 (note that in this two-team example, the winning percentages must sum to 1). The logic here is that increasing a team's winning percentage from .470 to .500 increases revenues more than increasing it from .870 to .900. Thus, the marginal revenue curve from additional wins is positive but downward sloping. Because teams in larger cities enjoy greater increases in fan support (more marginal revenue) from an additional win than teams in small cities, the marginal revenue for a large city is greater at any given winning percentage. In Figure 5.1, the marginal revenue curve for a large-market team (MR_L) lies above the marginal revenue curve for the small-market team (MR_S).[6] Teams maximize profits by setting the marginal revenue from an additional win equal to the marginal cost of creating that win. To keep the focus on revenue, assume that the marginal cost of a win is constant and equal for all teams. This assumption is based on the implicit assumption that each team has equal access to player talent. Figure 5.1 shows that in equilibrium, when the teams maximize profits—rather than wins or championships—the marginal revenue of each team is set to marginal cost and teams in large cities will have more talent and higher winning percentages than teams in small markets.[7]

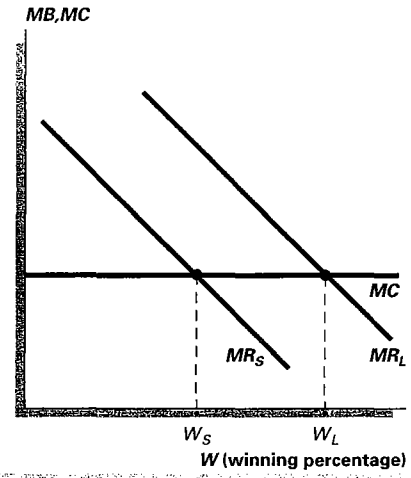
[5]For more on this topic, see Eric M. Leifer, *Making the Majors* (Cambridge, Mass.: Harvard University Press, 1995). See also John D. Burger and Stephen J.K. Walters, "Market Size, Pay, and Performance: A General Model and Application to Major League Baseball," *Journal of Sports Economics*, v. 4, no. 2 (May 2003), pp. 108–125, for research supporting the relationship between market size and performance.

[6]This model first appeared in Mohamed El-Hodiri and James Quirk, "An Economics Model of a Professional Sports League, "*Journal of Political Economy*, v. 79, no. 6 (November/December 1971), pp. 1302–1319.

[7]Gerald W. Scully developed this approach in *The Business of Major League Baseball* (Chicago: University of Chicago Press, 1989).

FIGURE 5.1

The Benefit of Improving the Team Depends on Market Size



3 How Competitive Balance Can Be Measured

This section and those that follow focus primarily on the **analysis of competitive balance (ACB),** which centers on how to measure competitive balance and its importance to leagues. We also address how competitive balance might be, and whether it should be, altered by league rules and regulations.

There are three general approaches to measuring competitive balance.[8] Two focus on the dispersion of winning percentages and one on the concentration of championships won. The debate as to which approach is best is ongoing and may not have a single answer. The measure that reflects a fan's perspective may or may not be the same as the one that reflects an owner's perspective, and the method best suited to study short-run effects on demand may not be best suited to study the long-run impact. Moreover, more than one approach may be required to capture the various aspects of competitive balance that are important to fans and leagues. Each approach is described separately in the following pages.

Within-Season Variation

Undoubtedly, the **absolute quality** of play affects demand, as fans want to see the game played at the highest level possible. For example, the demand for major league games is greater than the demand for minor league games. Within a league of a given quality, however, demand is also determined by **relative quality,** the quality of each team relative to the others in the league. Within-season measurement focuses on the relative quality of teams over a single season.

Measuring competitive balance for a given season is complicated by the fact that we must consider each team's winning percentage, not just an overall average. We cannot use the average winning percentage for the league because each game has one winner and one loser (ignoring ties). This means that the league-wide average winning percentage always equals .500. Thus we need to measure the dispersion of winning percentages—the variation around the average to measure the balance of competition.

To compute the dispersion of winning percentages (or of any variable), economists rely on the **standard deviation.** Standard deviation is a statistic that describes the average distance that observations lie from the mean of the observations in the data set. The formula for the standard deviation of winning percentages within a single season is

$$\sigma_{w,t} = \sqrt{\frac{\sum_{i=1}^{N}(WPCT_{i,t} - .500)^2}{N}},$$

where $WPCT_{i,t}$ is the winning percentage of the ith team in the league in year t, .500 is the average winning percentage of all teams for the year, and N is the number of teams in the league. The larger the standard deviation, the greater is the dispersion of the winning percentages. For example, consider the final standings for the NBA's 2004–2005 season. Table 5.2 shows the final standing for the Atlantic Division of the Eastern Conference and the Southeast Division of the Eastern Conference. It is possible to get a first impression of how balanced the divisions were simply by looking at the winning percentages. In the Atlantic Division, the two best teams (Boston and Philadelphia) won between 52.4 and 54.9 percent of the time, and the worst teams (Toronto and New York) won only about 40 percent of the time. The Southeast Division seems less balanced, as the best team (Miami) won over 70 percent of its games, and the worst team (Atlanta) won only about 16 percent of its games.

TABLE 5.2

Winning Percentages for the NBA Atlantic and Southeast Divisions 2004–2005

Atlantic	*Winning Percentages*	*Southeast*	*Winning Percentages*
Boston	0.549	Miami	0.72
Philadelphia	0.524	Washington	0.549
New Jersey	0.512	Orlando	0.439
Toronto	0.402	Charlotte	0.22
New York	0.402	Atlanta	0.159
σ	0.0668	σ	0.223

Source: http://www.nba.com/standings/2004/team_record_comparison/conference New_Std_Div.html

[8]For an excellent discussion of the technical merits of these measures, and more sophisticated variations of them, see Brad Humphreys, "Alternative Measures of Competitive Balance, " *Journal of Sports Economics*, v. 3, no. 2 (May 2002), pp. 133–148.

Computing the standard deviations (σ) confirms our first impression about the competitive balance in the two divisions. The standard deviation was .0668 in the Atlantic Division and .223 in the Southeast Division. Thus, the average winning percentage in the Southeast Division was more than three times as far from the division's mean than in the Atlantic Division.[9] If all teams in the NBA for the 2004–2005 season are included, the standard deviation of winning percentage is 0.152. Thus, the Atlantic Division was more competitive than the league as a whole, and the Midwest Division was less competitive than the league as a whole.

Although the standard deviation provides a more rigorous method for measuring competitive balance, it will be more useful when comparing different leagues and eras if we can define a common standard against which to measure them. What standard deviation is expected in a league in which each team has an equal chance of winning every game that it plays? This is equivalent to flipping a coin to see if the result is heads or tails. Try flipping a coin 10 times; the odds are that you will not come up with exactly 5 heads and 5 tails—which would translate to 2 teams with .500 winning percentages. As you flip the coin 100, 1,000, or 10,000 times, however, you will see that eventually, the number of heads and tails even out. Thus, for short seasons, all else being equal, teams are likely to have a greater spread of winning percentages, with some teams getting lucky breaks and other teams getting unlucky breaks. Only over a longer season would such breaks even out, just like a run of heads will eventually be offset by a run of tails when flipping a fair coin. If economists use the standard deviation to compare competitive balance across different sports, they must adjust for differences caused by differences in league size. Note that in the equation for within-season standard deviation, the number of teams N appears in the denominator. As N increases, the standard deviation falls, all else being equal.

The standard deviation that corresponds to the "ideal" competitive balance in which each team has a 0.5 chance of winning each game is $\sigma_I = .5/\sqrt{G}$, where .5 indicates that each team has a 0.5 probability of winning, and G is the number of games each team plays. In baseball, each team plays 162 games per season, so the ideal standard deviation is .039. In the NFL, the ideal standard deviation is much larger, .125, because teams play only 16 games, and a randomly occurring string of wins or losses has a greater impact on a team's final winning percentage. In the NHL and NBA, where teams play 80 and 82 games, the standard deviations are about .056.

To study competitive balance within a single season, we take the ratio (which we call R) of the actual standard deviation to the ideal standard deviation.[10]

$$R = \sigma_w/\sigma_I$$

[9]Because of interdivisional and interconference play, the mean winning percentages are not exactly .500 in this example.

[10]See Scully (1989) for an early use of this method as applied to professional baseball. We can also use the same idea to evaluate competitive balance over many seasons by calculating the average value of the standard deviation for a given year and using that value to create the ratio of actual to ideal standard deviation.

Thus, for the NBA in 2004–2005,

$$R = 0.152/0.056 = 2.71.$$

Based on this result, we see that the standard deviation of winning percentages in the NBA is more than twice what it would be in a world with absolutely balanced teams. Again, this result is consistent with our casual observation that competition appears unbalanced in the NBA, as three teams (San Antonio, Dallas, Phoenix) had winning percentages of over .700, while five teams had winning percentages of less than .350.

Based on the actual to ideal ratio R, the NBA is not the only seriously unbalanced league. The dispersion in winning percentages is greater than the "ideal" distribution in every sport. Table 5.3 shows the actual and ideal standard deviations for six major leagues, as well as the results for a recent season and long term historical average. Among North American leagues, the NFL is typically the most balanced, followed by the NHL and MLB. In both the American and National baseball leagues, where competitive balance is a continual topic of discussion and concern in the popular press as well as the league offices, R has historically exceeded 2.0. Although as we discuss later, the English Premier League and the Bundesliga are highly unbalanced leagues by other measures, based on the ratio of standard deviations, they are similar to the NFL. We return to this point later in the chapter.

The data in the table clearly show the NBA to be the least balanced league in North America. In *The Wages of Wins*, Dave Berri, Martin Schmidt, and Stacey Brook discuss why this might be so.[11] Their theory is based on the old adage

TABLE 5.3

Dispersion of Winning Percentages

League	*2005 Actual*	*Ideal*	*2005 Ratio*	*Historical Ratio*[a]
MLB	.066	.039	1.69	2.07 (AL) 2.13 (NL)
NFL	.208	.125	1.66	1.56
NBA	.152	.056	2.71	2.55
NHL	.099	.056	1.76	1.85
English Premier League	18.61	8.62	2.16	1.61
Bundesliga	12.20	8.20	1.49	1.45

Sources: 2005 data are generated from official league Web sites. Soccer and NHL data are from 2005–2006. Because teams receive one point for overtime losses, NHL value is computed based on percentage of possible points. English Premier and Bundesliga statistics are standard deviations of points rather than winning percentage.

[a]Historical data are from *The Wages of Wins* (2006), by David Berri, Martin B. Schmidt, and Stacey Brook, p. 61.

[11]Dave Berri, Martin Schmidt, and Stacey Brook, *The Wages of Wins* (Stanford, Calif.: Stanford University Press, 2006). The arguments presented here are based on a previous paper by Berri et al. titled "The Short Supply of Tall People: Explaining Competitive Imbalance in the National Basketball Association," *Journal of Economic Issues*, vol. 39, no. 4 (December), pp. 1029–1041.

"you can't teach height." In basketball, taller players have a distinct advantage over shorter ones. There are good players who are not tall, but given that the basket is 10 feet above the floor, if we compare two players of equal skill but substantially different heights, the taller player will be more effective. The number of very tall people who are also very gifted athletes—and whose athletic skills are well suited for basketball—is extremely small. Thus, when extraordinary players such as Shaquille O'Neal do appear, they can only play on one team. That scarcity creates competitive imbalance—not every team gets a Shaq.

Between-Season Variation

For baseball fans everywhere, spring is a special time of year that brings with it the promise of a new baseball season and the chance that "this could be the year" that their team wins it all. Across seasons, competitive balance implies that each team has the opportunity to move up in the standings each year and compete for playoff berths. This type of competitive balance is called **turnover,** or team-specific variation. It is quite distinct from within-season variation in that it considers the change in the relative positions of the teams in the standings each year rather than the distance between teams in a given season. Brad Humphreys (2002) defines team-specific variation for a team as

$$\sigma_{i,T} = \sqrt{\frac{\sum_{t=1}^{T}(WPCT_i - \overline{WPCT})^2}{T}},$$

where T is the number of seasons, and $\overline{WPCT}$ is the team's average winning percentage over the T seasons.[12] The larger σ_T becomes, the more a team's fortunes change from year to year. If a team always finished with the same record, σ_T would be zero. The more a team's fortunes change from year to year, the greater the standard deviation. If fans support a team only if it has a reasonable chance of winning its division or conference, variation across seasons is vital to maintaining fan interest over long stretches of time. If σ_T was zero for all teams, you would know how all teams would finish before the season even started. Such a situation would surely reduce demand for the weaker teams with below-average winning percentages, and over time it would probably hurt the stronger teams as well.

One frustrating aspect of using the variation between seasons is that, unlike the within-season standard deviation, there is no obvious standard of comparison. It is not possible to say whether fans or owners care more about how much their team's winning percentage varies across the years or how their team's position changes relative to other teams. For example, would Philadelphia Flyers' fans feel better if the Flyers had a very good record instead of a mediocre record but always finished second to the New Jersey Devils anyway?

[12]If you are interested in reading about the debate over which measures are most appropriate, see the articles by Brad R. Humphreys and E. W. Eckard in *Journal of Sports Economics*, v. 4, no. 1 (February 2003).

Though turnover is certainly important, the absence of an absolute standard means that σ_T is useful only as a relative measure of dispersion (when comparing one time period with another or one sport with another).

The Hirfindahl-Hirschman Index Another measure that economists have used to measure turnover is the Hirfindahl-Hirschman Index (*HHI*), which was originally developed to measure the concentration of firms in an industry. In our case, we calculate the *HHI* by taking the number of times each team finished first, squaring it, adding these numbers together, and then dividing them by the number of years under consideration. A small *HHI* means that a large number of teams finish first, while a large *HHI* means that a small number of teams dominated the league.

$$HHI = \frac{\Sigma f^2}{T}.$$

To see why, consider the championships won in a league with 5 teams over three 10-year periods. Suppose that in the 1960s, each team in the league finished first, twice. In the 1970s, competition was less balanced, and two teams each finished first, five times. Finally, in the 1980s, 1 team placed first all 10 times. The *HHI* for each decade would be

$$HHI_{60} = (2^2 + 2^2 + 2^2 + 2^2 + 2^2)/10 = 20/10 = 2.$$

$$HHI_{70} = (5^2 + 5^2 + 0^2 + 0^2 + 0^2)/10 = 50/10 = 5.$$

$$HHI_{80} = (10^2 + 0^2 + 0^2 + 0^2 + 0^2)/10 = 100/10 = 10.$$

As competition becomes less balanced, the value of *HHI* rises. An advantage of using the *HHI* instead of the standard deviation is that, for any given league, the *HHI* allows us to compute a benchmark against which we can compare results (as in the 1960s example above). A disadvantage, as with the other measures, is that the *HHI* still does not address the issue of optimal balance. A second disadvantage of this measure is the interpretation of the standard itself. According to this "ideal," fans in a league with N "perfectly balanced" teams will wait an average of N years for their team's "turn" to finish first to come around, after which they will wait another N years.

Frequency of Championships It is also possible to evaluate competitive balance by looking at the frequency with which teams win successive championships. On the one hand, if the Yankees win the World Series every year, then the winning percentages of the teams in the league do not matter as much, since the league is clearly unbalanced. On the other hand, if different teams win the American League and National League pennants every year, then it's possible to argue that competition in each league is balanced, regardless of how bad the worst teams are relative to the best teams. This criterion is similar to the turnover criterion discussed above, but it relates to championships rather than regular

season standings. To focus strictly on championships, we can measure interseason balance by applying the *HHI* methodology to championships won instead of to first-place finishes. Another, more direct approach is to simply count the championships won by a team for a specific period of time.

If we consider only how often teams win championships, Table 5.4 again shows that the NBA is the least balanced league in North America. Just two teams out of a 30-team league won 50 percent of the championships between 1980 and 2006. Of these, many of the wins were in consecutive years. The Chicago Bulls won their six titles in just eight years. More recently, the Lakers won three straight. Over that same time period, 70 percent of teams did not win once. In contrast, repeat champions in the NFL are almost unheard of. Evidence on competitive balance outside North America is mixed. In two of the top European soccer leagues—The English Premier League and the German Bundesliga—competition is very unbalanced. Based on frequency of titles, Manchester United and Bayern Munich are two of the most dominant organizations in all of team sports. In contrast, eleven different teams have won the title in the Australian Rules Football League (AFL), which since 1995 has consisted of 16 teams and had only 12 teams until 1986. In the last 17 years, there have only been three instances of champions successfully defending their titles.

In summary, there are many ways to measure competitive balance, and no single method should be regarded as most appropriate. To fully grasp the state of competitive balance in a league requires consideration of intraseason balance—the spread of winning percentages across teams—as well as interseason balance, the turnover of teams in the standings, and the frequency of championships. Leagues must be concerned about all these forms of competitive balance because fan interest—and correspondingly attendance, television ratings, and league profits—is likely to be affected by each. The next section addresses how leagues may attempt to alter competitive balance.

Attempts to Alter Competitive Balance

Currently, both the popular press and the professional literature are discussing competitive balance in every major sport in the United States as well as in the European soccer leagues. The greatest concern in the United States is in baseball, in which fans are concerned that beginning in the 1990s, competitive balance has become substantially worse. Because a degree of parity is so important to the success of any league, all the major sports have developed processes designed to promote competitive balance. The most important of these are revenue sharing, salary caps, luxury taxes, and the draft. Leagues that have implemented these policies claim that they equalize access to talented players so that no one team or small number of teams can hoard an excessive number of talented players. The success of these policies depends to a large extent on two factors: Whether a team's performance is strongly related to its payroll and whether making rules that limit payroll (such as salary caps and luxury taxes) increases competitive balance. The basics of each technique are described in following paragraphs, followed by a discussion of whether or not they are effective.

TABLE 5.4

Distribution of Championships: 1980 to 2005–2006

NBA	*NHL*[a]	*NFL*	*MLB AL*[b]	*MLB NL*[b]	*English Premier*[c, d]	*Bundesliga (Germany)*[c]	*AFL—Australian*[c] *Rules Football*
Lakers—7	Oilers—5	49ers—5	Yankees—8	Braves—5	Manchester United—8	Bayern Munich—14	Hawthorne—5
Bulls—6	Islanders—4	Cowboys—3	Athletics—3	Cardinals—4	Arsenal—3	Werder Bremen—3	Adelaide—4
Celtics—3	Devils—3	Redskins—3	Blue Jays—2	Phillies—3	Blackburn Rovers—1	BV 09 Borussia Dortmund—2	Essendon—4
Spurs—3	Red Wings—3	Patriots—3	Red Sox—2	Marlins—2	Chelsea—1	FC Kaiserslautern—2	Carlton—4
Pistons—3	Avalanche—2	Broncos—2	Indians—2	Giants—2		VfB Stuttgart—2	Brisbaine Lions—3
Rockets—2	Canadiens—2	Giants—2	Twins—2	Mets—2		Hamburger SV—2	West Coast Eagles—2
76ers—1	Penguins—2	Raiders—2	Royals—2	Padres—2			Richmond—1
Heat—1	Lightning—1	Steelers—2	Angels—1	Dodgers—2			Collingwood—1
	Hurricanes—1	Bears—1	Tigers—1	Reds—1			North Melbourne—1
	Rangers—1	Rams—1	Orioles—1	Diamondbacks—1			Kangaroos—1
	Stars—1	Ravens—1	Brewers—1	Astros—1			Sydney—1
	Flames—1	Packers—1					

[a] No NHL champion 2004–2005.
[b] No Postseason play 1994.
[c] Results through 2005.
[d] English Premier data begin 1992–1993.

Revenue Sharing

As was discussed in Chapter 3, a primary outcome of revenue sharing is more equal profits. To the extent that revenue sharing also increases a financially weak team's ability to sign and retain players, it may also improve competitive balance. It is, however, only an indirect method of redistributing players and may not equalize talent. Two conditions must hold for revenue sharing to result in a more equal distribution of talent and thus improve competitive balance. First, if it is assumed that teams attempt to maximize profits, the teams that receive revenue must benefit financially from improving their performance. If these teams do not benefit sufficiently from spending the money on better players, they may simply keep the payment, and revenue sharing will have no effect on competitive balance. Second, revenue sharing can help a team acquire better players only if players can move—or be moved—easily from team to team. This presupposes the existence of either free agency or the ability to buy and sell the rights to players. Unfortunately, data from Commissioner Selig's blue ribbon commission show that some of the worst on-field teams in baseball were more profitable than other teams with much better records *and* were net receivers of shared revenue, a fact that seems to indicate that those teams that received revenue were inclined to keep it rather than spend it on better talent.

Salary Caps and Luxury Taxes

Salary caps stipulate the maximum that a team may spend on player salaries in a given year. Such a system is currently in place in the NBA, NHL, and NFL. In 2006 (2005–2006), each NFL team was not supposed to spend more than $102 million on player salaries. The NHL team cap was $39 million, and the NBA cap was $49.5 million. Teams that violate the cap are subject to fines from the league. Such stipulations have two primary effects. First, overall spending on players will decline, a topic covered further in Chapter 8. Second, and more importantly for competitive balance, no team will be able to pay to hire all of the best players, which should equalize talent across teams.

Luxury taxes, such as the one implemented in Major League Baseball (which they refer to as a competitive balance tax) force teams to pay an additional fee to the league on any payroll expenditure above a certain amount. In 2005, any MLB team that spent more than $128 million on salaries for the year had to pay a tax of 17.5 percent of any overage into a league fund, which was then distributed to low-revenue teams as a form of revenue sharing. In subsequent years, the tax rate rises to 30 percent for the second offense and 40 percent for the third offense.[13] The luxury tax can be considered a weaker form of a salary cap. With a hard salary cap, no team is permitted to exceed the cap, regardless of its willingness to pay a penalty. With a luxury tax, teams may choose to exceed the tax threshold as long as they are willing to pay the tax. For

[13]Bloom, Barry M., "Yanks, Red Sox Hit with Luxury Tax Bills." Online at http://mlb.mlb.com/NASApp/mlb/news/article.jsp?Ymd=200512218.content.id=12862258vkey=news_mlb&fext=.jsp&c_id=mlb accessed March 16, 2007.

example, in 2005, the Yankees' payroll of over $200 million exceeded the cap by over $70 million. Based on the formula designed by Major League Baseball, the tax owed by the Yankees amounted to about $34 million. The high cost of obtaining a substantial percentage of the top players and the fact that low-revenue teams receive payments from the very high salary teams allows teams to have different profit-maximizing strategies with respect to the tax. As explained in Chapter 3, the benefits of fielding a high-quality team may be enormous for a large-market team such as the Yankees. Thus, it may be worthwhile for it to exceed the cap and pay the tax, whereas the cost of doing so would surely exceed the possible benefits for a small-market team.

The Reverse-Order Entry Draft

The reverse-order entry draft allows teams to choose incoming players in reverse order from their finish of the previous season. The team with the worst record is given the first choice, the second-worst team chooses second, and so on until the team that won the previous season's championship chooses last. The same procedure is followed through subsequent rounds. When the last round is complete, all remaining players who have not been chosen are free to try out for and sign contracts with any team.

The origin of the draft can be traced to 1934, when two NFL teams—the old Brooklyn Dodgers and the Philadelphia Eagles—bid against each other for the services of Stan Kostka, an All-American player at the University of Minnesota. The resulting bidding war drove salary offers to the then unbelievable level of $5,000 (what Bronko Nagurski—the greatest player of the era—made).

At the next league meeting, Bert Bell, the Philadelphia Eagles owner, proposed a unique way to avoid bidding wars over unsigned players in the future. Teams would select the rights to sign unsigned players, with the order of selection determined by each team's performance in the previous season. Ironically, the fact that the worst teams choose first in the draft—a feature that has led all leagues to cite the draft as a key to maintaining competitive balance—stems from the fact that Bell's Eagles were a last-place team at the time.[14] The implications for the power that this gave teams over the players' ability to negotiate for higher salaries are further discussed in Chapter 8. For now, we emphasize the fact that if the worst teams are always able to select the best players, it may help to equalize the talent across teams, because it prevents the teams with the most revenue from dominating the bidding for the most talented players.

Schedule Adjustments in the NFL

The NFL has a unique method for introducing an additional element of parity across seasons that is unrelated to the movement or acquisition of players. By rule, each team's schedule for the following season is determined in part by the team's performance in the previous season. The formula for opponents scheduled requires that each team play 14 of its 16 games against opponents that are

[14]James Quirk and Rodney Fort, *Paydirt* (Princeton, N.J.: Princeton University Press, 1992), p. 187.

common to all members of a division. Each team plays the other three teams in its own division twice, plus all four teams in one other division within the conference, plus all four teams from one division in the other conference for a total of 14 games. The relevant portion of the schedule for this discussion is that for each team, the remaining two games are played against opponents determined based on performance in the previous season. The first-place team in each division plays the first-place teams in the two divisions that the team is not scheduled to play; the second-place team plays the other two second-place teams, and so on. As a result, stronger teams play stronger schedules the following year, and weaker teams play weaker schedules the following year, creating a natural tendency toward parity.

The Effects of Attempts to Alter Competitive Balance

Revenue sharing, drafts, salary caps, and luxury taxes can all help to increase the parity in a league. There are, however, a number of reasons why they might not. The most obvious is that player talent is intangible. Any given player may be dominant in one year and then perform very poorly in the next year. Similarly, no matter how highly regarded a player is in high school or college, his talent may not translate to the professional level. If players were robot-like in their consistency, it would be much easier to predict the best teams from season to season by comparing the quality of team rosters based on past performance. A logical extension of this argument is that if better players command higher salaries, whichever team spends the most on talent would always win.

Unfortunately for the owners, who must decide how much to pay each player, the relationship is not precise. Table 5.5 shows the correlation between payroll and winning percentage for each team in the four major sports in 2005. A correlation coefficient close to 1 would indicate that wins and payroll are very strongly related. A correlation coefficient close to zero would indicate that wins and payroll are not related. The table shows that although payroll and winning percentage are positively related, the relationship is far from perfect, especially in the NFL and NBA, where the correlation is less than 0.2.

Table 5.5 also shows that, even if owners are willing to spend more to hire better players, success is not assured. The level of uncertainty is even higher when trying to select players who are entering the league for the first time. There are countless stories of highly drafted players who end up never playing a down in the NFL, never swinging at a MLB pitch, never taking a shift in an NHL game, and never playing a minute in the NBA. Even for those who do make the teams that choose them, success is not precisely related to draft order.

There are many examples of the tenuous relationship between expenditure or draft order and performance. Michael Jordan, perhaps the greatest basketball player ever, was chosen third in the NBA draft by the Chicago Bulls, behind Hakeem Olajuwon and Sam Bowie. While Olajuwon went on to have an outstanding career, Bowie played just five lackluster seasons in the league before injuries forced him to retire. In the NHL, the Flyers traded away the rights to Peter

TABLE 5.5

The Correlation between Payroll and Winning Percentage

Sport	*Correlation Coefficient*	*Sport*	*Correlation Coefficient*
MLB	.492	NBA	.105
NFL	.156	NHL	.319

Sources: NHL data are for 2003–2004. All other sports are seasons ending in 2005. Performance data are from official league sites. Salary data are from *USAToday* at www.usatoday.com/sports/front.htm.

Forsberg, Ron Hextall, and 4 other players *plus* $15 million to acquire the rights to Eric Lindros. Forsberg went on to become one of the most dominant forwards in the game for the Colorado Avalanche (who were the Quebec Nordiques at the time of this deal), leading them to the Stanley Cup in 1996 and 2001. Unfortunately for the Flyers, Lindros was plagued by concussions and had a falling out with team management, leading to one of the unhappiest chapters in the team's history.[15]

In summary, the uncertainty that surrounds talent and the relatively weak relationship between expenditure and team quality cast doubt on the effectiveness of policies that equalize expenditure and limit opportunities to select new players. The next section discusses another powerful force that may serve to undo any efforts by the league to enhance competitive balance.

The Coase Theorem and Competitive Balance

In order for any attempt to equalize talent to be effective, there must be some institutional mechanism to move players from one team to another. Chapter 4 described how for much of the 20th century, professional sports leagues controlled player movement through a reserve system that essentially gave the teams the "rights" to a player for his entire career. More recently, players in baseball, basketball, and football have won the right to sell their services to the highest bidder through free agency. Owners have long argued that free agency is bad for competitive balance because it allows the teams with the most resources to purchase the best talent and dominate the league. As it turns out, economists argue that it does not matter which system is in place. Neither system affects competitive balance because the owners' reasoning flies in the face of a theory of resource markets known as the **Coase Theorem.** The central finding of the Coase Theorem is that the initial allocation of property rights (whether a player has the right to sell his services to any team or the owner has the right to hold a player's contract for his entire career) does not matter.[16] As long as property rights are clearly established (as long as someone owns the

[15]Ironically, in 2005, the Flyers reacquired Forsberg by signing him as a free agent and traded him away again in 2007.

[16]The theorem was largely responsible for Ronald Coase's receiving the 1991 Nobel Memorial Prize. See Ronald Coase, "The Problem of Social Cost," *Journal of Law and Economics*, v. 3 (October 1960), pp. 1–44.

resource in the first place) and bargaining costs are low, the resource will be put to use by the person or firm that benefits from it most. If a player owns the rights to his services, that is, if he has the right to sell his services to the highest bidder, the team that values him the most will offer the best contract and sign him. If profit-maximizing teams hold the property rights to player services and are able to buy and sell the rights to players, the best players again end up on the teams that value their services the most, because those teams will make the highest offers. Thus, the Coase Theorem predicts that free agency has no impact on the distribution of talent. Economists commonly refer to this as the **invariance hypothesis.**

The allocation of property rights should not affect the final distribution of talent. Regardless of who owns the rights, players end up on the teams that value their services the most. The only difference between the two systems is that if the players own the rights to their services (as in free agency), they keep the gains from their movement from one team to another, whereas if the owners reserve the players' rights, the benefits from the payments flow to the owner who sells the player. Chapter 8 covers this last point in more detail. This section focuses only on what the Coase Theorem predicts about the distribution of talent.

Suppose that a star shooting guard adds $5 million per year to the revenue of the Indiana Pacers. Largely because of the difference in market size, the value of that same player would be $7 million if he played for the Los Angeles Lakers. Under complete free agency, the Lakers would outbid the Pacers, and the player would play in Los Angeles for between $5 million and $7 million. If instead there were no free agency but teams could "sell" players for cash or payment in kind, the Lakers would pay the Pacers more than the $5 million that the player contributes to the Pacers but less than the $7 million that the player is worth to the Lakers, and both teams are better off. In either case, the player ends up playing for the team that values his services the most. Although some leagues have placed restrictions on the outright sale of players, as long as teams can trade draft picks and players, the same result holds. The Coase Theorem thus predicts that free agency alone does not distribute playing talent less equally than a reserve system. A recent study of Major League Baseball shows that the advent of free agency in 1976 did not lessen competitive balance and may actually have improved it. The study found that the distribution of winning percentages did not change significantly after 1976 and that the correlation of winning percentages (turnover) from one year to the next actually fell after 1976.[17]

Salary Caps

Salary caps can even out the talent across a league, at least in principle. There are two types of salary caps, soft and hard. A soft cap, such as was in place from 1983 until 1999 in the NBA, does very little to promote balance. Under this system, NBA teams were allowed to sign new players up to the value of the cap and then re-sign their own free agents in excess of the cap figure, a provision

[17]Michael R. Butler, "Competitive Balance in Major League Baseball," *American Economist*, v. 39, no. 2 (Fall 1995), pp. 46–50.

that became known as the "Larry Bird exemption," as it was first used by the Boston Celtics in order to re-sign their star player in 1983. At that time, the Celtics were loaded with stars, and a hard cap would have meant that the Celtics would have had to break up this very popular team in order to re-sign Larry Bird. To avoid this, the league softened the cap, permitting the Celtics to keep their team intact. The result was that teams routinely spent much more than the salary cap, and competitive balance was not improved. In fact, the within-season standard deviation of winning percentages actually rose consistently through the soft-cap era, indicating a decrease in competitive balance.[18]

A hard salary cap, such as the ones currently in force in the NFL and NHL, limit any one team's ability to sign a large number of free agents because it provides only a few exceptions (such as for injuries) to the salary limit. Going back to the earlier example about the Pacers' star guard, if the Lakers are already at their salary cap limit, the value they place on the Pacers' guard is irrelevant. They cannot offer him a contract unless they drop other players from the roster. If strictly enforced, these payroll limits would prevent wealthy teams from stockpiling talent. Enforcement is not a trivial component here, as the players union reported that in 1995, as many as 26 teams used loopholes in the salary cap language to spend in excess of the cap in the NFL. Although some research supports the idea that a hard salary cap improves competitive balance, the lack of a hard salary cap in baseball and English soccer and strategic maneuvering to circumvent the letter or spirit of the salary cap make empirical testing difficult.

In addition to exceeding salary caps, teams frequently undermine their effectiveness by restructuring their players' contracts (e.g., deferring bonus payments) in order to obey the letter of the salary cap regulations while violating them in spirit. For example, some NBA teams that were close to the salary cap exploited the Larry Bird exemption to attract free agents by signing them to a low salary for one year and then re-signing them to a much higher salary a year later. The Portland Trailblazers were the first to use this provision—over the protests of the NBA commissioner's office—in 1993, when they signed Chris Dudley away from the New Jersey Nets. The Miami Heat later used the same tactic to lure Alonzo Mourning away from the Charlotte Hornets. The NBA owners tried unsuccessfully to curb this practice in the 1995 collective bargaining agreement. Although they failed to institute a hard salary cap covering all players in the 1995 agreement, the owners were able to institute a rookie salary cap. The larger problem of the loophole in the overall team cap remained (under which Michael Jordan alone virtually exhausted the Chicago Bulls' salary cap) and was a major factor in the hard line taken by the NBA in the 1998 contract negotiations. The impact of the more restrictive 1999 collective bargaining agreement on the NBA will not be fully felt for several years. Most players in the NBA have multiyear contracts that are "grandfathered" (exempt from) the new agreement. Until the new provisions that set individual salary caps cover all players, it will be impossible to measure their full impact.

[18]Andrew Zimbalist, "Competitive Balance in Sports Leagues: An Introduction," *Journal of Sports Economics*, v. 3, no. 2 (May 2002), pp. 111–121.

Anecdotally, there is evidence that it will cause salaries to fall. For example, Juwan Howard signed a contract with the Orlando Magic worth approximately $5.9 million per year, a dramatic decrease from his prior contract with the Washington Wizards, which was valued at approximately $15 million per year, despite no decrease in his productivity.

The Draft

Despite its curious origins, all leagues claim that the reverse-order draft is vital to preserve competitive balance. In theory, teams that finish last choose the best available players and have the best chance to improve. Unfortunately, several factors limit the effectiveness of drafts. The biggest obstacle is the uncertainty over whether and when a player will make significant contributions to the team. Thus, the draft highlights the managerial skill of the team. Some teams consistently finish last because their management and coaching staff do a poor job of identifying or developing talented players. High draft picks do such teams little good.[19] In some sports, such as baseball, players chosen in the draft are usually several years away from being ready to play in the major leagues. For some positions in football, the players in them undergo extensive training before it is apparent how well they will perform. Thus, even if they choose the right players, the weakest teams may see improvement only with a significant lag.

Managerial skill also plays a role in whether the team ever uses the draft picks to which it is entitled. Skillful trading of draft picks may garner teams more talent than it could have obtained through the draft, while missteps may doom it to years of mediocrity. The NBA has attempted to reduce the problems caused by bad managerial decision making by restricting the number of first-round picks that teams can trade. An added problem is that even if a team selects the "right" player, that player may be unwilling to play for the team that picked him. This happened when Steve Francis, a star guard for the University of Maryland, refused to play for the Vancouver Grizzlies, who had selected him in the 1999 NBA draft. Unable to force Francis to sign, the Grizzlies traded him to the Houston Rockets for players that most observers agree provided the Grizzlies with less than equal value.

In a related problem, draft rules may not square with league-wide profit maximization. If players are chosen based on their productivity so that the most productive players are chosen first, then players chosen later in the draft should be worth less to the teams that choose them. The problem with this reasoning is that value of a player to his team is the combination of marginal productivity and the value of that added output to the team. Because a player's marginal revenue product is determined by both the player's marginal product and the marginal revenue of his output, players are generally not assigned (drafted) in a profit-maximizing way. For example, a player who is drafted by a

[19]See, for example, R. Hoffer, "The Loss Generation," *Sports Illustrated*, April 17, 2000, pp. 56–59.

small-market team may have a lower market value than a player selected later in the draft by a large-market team even though the player drafted by the small-market team is more productive. If the league assigned players to teams in a profit-maximizing way instead of through a reverse-order draft, it would begin by matching the player–team combination that creates the highest marginal revenue product, and then the second highest, and so on. Although this may be profit-maximizing, it is likely that it would be so politically unpopular with fans and small-market owners that it could never become policy (even though according to the Coase Theorem, players will end up distributed this way anyway).

Finally, the reverse-order draft gives teams an incentive to lose late in the season with the hope of improving their draft position.[20] The NBA lottery system, wherein the last-place finishing team is no longer guaranteed the first choice, was created specifically to prevent teams from intentionally losing games near the end of the season to ensure a good draft position.

Revenue Sharing and Luxury Taxes

As was discussed in Chapters 3 and 4, professional sports share revenue in a variety of ways. To equalize teams' ability to pay for talent, leagues can stipulate that national television contract revenue, gate revenue (in some sports), and licensing revenue are shared. They can also tax teams that overspend and distribute those funds to low-revenue teams (as in baseball and basketball). Previous chapters have shown that the primary motivation for such revenue sharing is league-wide economic stability, not competitive balance. Revenue sharing can affect competitive balance if the teams that are net recipients of funds (that receive more funds than without revenue sharing) use the additional money to improve the quality of their teams. If those teams simply pocket the additional revenues they receive, revenue sharing cannot improve competitive balance. To date, there is no strong empirical evidence that revenue sharing does equalize competition. It remains to be seen whether the new, more generous sharing formulas in baseball will increase competitive balance.

Promotion and Relegation

The promotion and relegation system provides an additional incentive mechanism that may increase competitive balance. We saw that top European leagues such as the English Premier League and the Bundesliga have actual to ideal competitive balance ratios that are not far from the NFL, North America's most balanced league. Yet, if we consider frequency of championships as the measure of balance, Table 5.3 provides strong evidence that these same leagues are

[20]For more information on this phenomenon, see Beck A. Taylor and Justin G. Trogdon, "Losing to Win: Tournament Incentives in the National Basketball Association," *Journal of Labor Economics*, v. 20, no. 1 (January 2002), pp. 23–41.

among the least competitive. The promotion and relegation system may help to provide an explanation to this puzzle.

When a team in a North American league is having a bad season and stands to finish near the bottom of the standings, it may not have much incentive to win. Once eliminated from the playoffs, the team may instead use the remaining games to try out new players in its minor league system, or raise capital by selling off some of its top players to playoff-bound teams that are in search of that "missing piece" needed to make them a championship contender. Thus the winning percentage of the poor team may erode further, increasing the standard deviation of winning percentage. In a promotion and relegation league, teams near the bottom of the standings have no such luxury. If they allow their performance to continue to slide, they may end up being relegated to the next lower league. Teams that are near the bottom of the standings have an incentive to continue to play to win right to the end. Thus, promotion and relegation may not create turnover of the league champion, but it could well decrease the standard deviation of wins within the league.

BIOGRAPHICAL SKETCH

BUD SELIG

Selig listened and questioned and murmured empathetically, all of the things he did best.

—*John Helyar*[1]

Perhaps no person symbolizes the struggle over competitive balance more than baseball Commissioner Alan H. ("Bud") Selig. To his supporters, Selig is a hero of the underdog. To his detractors, Selig is simply out of his depth. Whether admiring him or hating him, however, no one disagrees that Selig is motivated by a devotion to his home state of Wisconsin with a deep love of baseball.

Selig's Wisconsin roots run deep. Born in Milwaukee, he graduated from the University of Wisconsin at Madison in 1956 and, after serving in the military for two years, joined his father's automobile business. Business proved so good that, when major league baseball came to Milwaukee, Selig was able to act on his love of baseball by becoming a stockholder in the Milwaukee Braves. Selig's ties, however, were to the *Milwaukee* Braves. When the team moved to Atlanta in 1965, Selig promptly sold his stock and formed a group dedicated to bringing a new team to Milwaukee. His efforts bore fruit when the Seattle Pilots, a badly financed expansion team, went bankrupt after the 1970 season. Selig immediately bought the team for $10.8 million and moved it to Milwaukee.

With Selig as their president, the Brewers gained a reputation as an exemplary organization, and the team came within a game of winning the 1982 World Series. The Brewers' performance on and off the field led Selig to play a growing role in the governance of Major League Baseball's affairs. When the owners forced Fay Vincent to resign as commissioner in 1992, Selig, as chairman of the owners' executive council, effectively took over the duties of commissioner. For the next six years, Selig walked a tightrope, serving the interests of all of baseball while working to advance the interests of his own Milwaukee Brewers. Finally, in July 1998, Selig's fellow owners elected him as commissioner. Selig then put his holdings in the Brewers into a blind trust and turned operations of the Brewers to his daughter Wendy Selig-Prieb.

Selig's popularity with his fellow owners and his insistence on consensus among the team owners has brought unprecedented tranquility among the owners and a cohesion that has enabled him to introduce a variety of innovations designed to bring greater excitement to the game. Under his tenure, baseball raised the number of divisions per league from two to three, increasing the number of teams entering the postseason. The number was further increased by the introduction of a "wildcard" playoff team (which won both the 2002 and 2003 World Series). He also oversaw a greater consolidation of the American and National Leagues, whose war of the early 1900s did not fully end until Selig brought both leagues under the authority of the commissioner's office in 2000.

Most importantly, by bringing the often fractious owners together, Selig reversed a trend of over 20 years. Under previous commissioners, all labor stoppages had effectively ended with the owners' capitulating. The 1994–1995 strike effectively ended in a draw, with neither side achieving its aims. In the near strike of 2002, the ownership actually succeeded in forcing the players association to blink and to approve a revenue-sharing plan and luxury tax that it had bitterly opposed. This marked ownership's first outright victory in negotiations since the first dispute in 1972.

Despite these successes, Selig's tenure has not been without controversy. Some critics have accused him of unduly favoring small-market teams in general and his own Milwaukee Brewers in particular. They cite, for example, his refusal to enforce limits on the amount of debt a team could carry as long as the Brewers were deeply in debt (up to 97 percent of the team's value by some estimates) and reversing himself only after the Brewers had secured a new ballpark and significantly reduced their debt burden.

Critics also point to Selig's threat in the winter of 2001 to "contract" two teams, generally assumed to be the Montreal Expos and Minnesota Twins. While the Expos were in genuinely dire straits, many saw darker motives behind the threat against the Twins. Some claimed that contraction was nothing more than a heavy-handed attempt to secure public financing for a new ballpark in Minneapolis and to intimidate the players association while the owners and players were negotiating a new contract. Some also construed it as a way to improve the finances of the Milwaukee franchise by eliminating a significant geographic rival. Finally, some felt that if the league bought out the Twins at more than market value and then shut them down, it would represent a payback of a secret loan that the Twins' owner had made to the Brewers in 1995, a loan that apparently violated League Rule 20(c), which prohibits loans among owners without the express consent of the commissioner and the other owners. Former Commissioner Vincent has called the loan "treacherous," while Drayton McLane, the owner of the Houston Astros has claimed that it was "honest and ethical."

Even Selig's victories over the players association came at a large cost. In particular, the 1994–1995 stoppage caused the cancellation of the 1994 World Series, something that two world wars had failed to do. Selig's awkward declaration that the 2002 All-Star Game was a tie (due to both leagues' having exhausted their bullpens) created further embarrassment at what should have been a celebration of baseball.

Because of his controversial record, some see Bud Selig as a man who saved the game. Others feel that he was ill suited to be anything other than the owner of a small-market team. Whatever one's opinion of him, few can deny that he has had a greater impact on the game than almost all his predecessors as commissioner.

[1]John Helyar, *Lords of the Realm* (1994), p. 505.

Sources: Associated Press, "MLB Official Says 'Nothing Improper' About 1995 Loan," ESPN Baseball at http://espn.go.com/mlb/news/2002/0108/1307601.html, January 9, 2002; Anonymous, "Bud Selig," BaseballLibrary.com at http://www.baseballlibrary.com/baseballlibrary/ballplayers/S/Selig_Bud.stm; Anonymous, "Alan H. 'Bud' Selig" Commissioners at http:// mlb.mlb.com/NASApp/mlb/mlb/history/mlb_history_people_profile.jsp?section=bio&persontype=com&personid=9

Summary

In order for a league to be financially successful in the long run, there must be a semblance of even competition among teams. That said, it is unlikely, given that the value of a win is much greater in large cities, that leagues would maximize revenue from perfect parity across teams and would likely do better to have better teams in cities where demand for the sport is greatest.

There are a variety of methods to measure and analyze competitive balance both within a single season and across seasons. Within-season measures focus on the dispersion of winning percentages across the league between the best and worst teams. Across-season measures tend to focus on turnover of standings. Because fans are concerned about both types of balance, it is not possible to choose a single measure of competitive balance, and instead a measure based on the type of balance under study must be considered.

Attempts to alter competitive balance using systems such as revenue sharing, salary caps, and drafts are currently in place in all American professional sports. They do not appear to have a major impact in most cases, although the NFL is the most balanced league in terms of competition and shares revenue the most aggressively across teams.

Discussion Questions

1. Why might owners not want perfect parity in a league?
2. How would competitive balance in the American and National Leagues change if baseball owners forced the Yankees to move to Albuquerque, New Mexico?
3. What do you believe means more to fans, having a chance to win the championship once in a while or being competitive every year?
4. Could Major League Baseball successfully adopt a system of promotion and relegation?
5. Why might the players unions not want owners to enact competitive-balance enhancing measures?
6. Most leagues have about 30 teams. As a fan, would you be willing to endure 29 losing seasons if you were guaranteed a championship in the 30th?

Problems

5.1 Suppose, as an owner, you could leave the highly competitive league (in terms of closeness of contests) that you currently play in and enter a league that assured that your team would never lose again. Would you want to do so? Why or why not?

5.2 Explain how the law of diminishing returns provides a natural tendency toward competitive balance.

5.3 Suppose in a six-team league, the winning percentages were as follows at the end of the season. Team A: .750, Team B: .600, Team C: .500, Team D: .500, Team E: .400, Team F: .250. Compute the standard deviation of winning percentages.

5.4 In question 3, suppose each team plays a 50-game schedule. Compute the "ideal" benchmark standard deviation based on equal playing strength, and the ratio of the actual to the ideal.

5.5 If the NFL increased its schedule from 16 games to 30, what would the new benchmark ideal standard deviation be (assuming equal playing strength)?

5.6 What is the main prediction of the Coase Theorem with respect to free agency and competitive balance?

5.7 Which would be more effective for increasing the level of competitive balance in baseball, a hard salary cap, or a 50-50 gate revenue-sharing plan? Why?

5.8 Suppose that over five seasons, the order of finish for five teams in the West League and the East League are as follows. Use the HHI to determine which league has better competitive balance across seasons.

West League Season					*East League Season*				
1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
A	A	A	E	E	A	B	C	D	E
B	B	D	D	D	E	A	A	A	A
C	C	C	C	C	C	B	D	E	D
D	D	B	B	B	B	D	B	B	B
E	E	E	A	A	D	E	E	C	C

5.9 If you were a fan of team A, which set of distributions shown in the previous question (West or East) would you prefer? Why?

5.10 If, as commissioner of professional baseball, you could make one change in the current league rules with the goal of increasing competitive balance as much as possible, what change would you make and why?

5.11 Go to the official MLB Web site (http://www.mlb.com) and check the order of finish in the American League East for the 1997–2003 seasons. What is the between-season variation for this league over this time span?

APPENDIX 5A

Two Additional Ways to Measure Competitive Balance: The Lorenz Curve and the Markov Chain Method

In the body of the chapter, we introduced several ways to measure competitive balance. Depending on the literature that you read on competitive balance, you may encounter others. In this Appendix, we cover two additional methods for measuring competitive balance: the Lorenz curve and the Markov chain.

The Lorenz curve is most often found in studies of income inequality, as it was originally designed for this purpose and can be used to easily summarize how close a group of people (such as the citizens of a nation) are to perfect income equality. In Chapter 8, we will see an application of the Lorenz curve to athlete salaries in individual sports. Here, we apply the Lorenz curve concept to team success over a fixed time period. The Lorenz curve is a cumulative distribution of observations as measured against a variable of interest. Sports economists have adapted the Lorenz curve to show how near or far a league is from perfectly balanced competition over a given period of time by measuring the percentage of teams that have won championships over a given number of years.

To see how the Lorenz curve works, suppose that a league has five teams. With perfectly equal competition, we would expect each team to win 5 championships each over a 25-year period. Put in cumulative terms, 20 percent of the teams (1 team in this case) would win 20 percent of the titles (5). Continuing, 40 percent of teams (2) would win 40 percent of the titles, and so on, until we reach 100 percent of the teams have won 100 percent of the titles. Such a distribution of winning would result in the Lorenz curve in Figure 5A.1. In this case, the curve is not a curve at all but a straight 45-degree line that now serves as the benchmark for perfect equality.

If we change the distribution of wins in our hypothetical league, we can see how the curve shifts. When all teams were equally successful, it did not matter in what order we recorded the teams' performance. When championship titles are not distributed equally, we form the cumulative distribution by arranging the observations such that we begin with the team that won the fewest and continue to the team that won the most. Suppose instead of perfect equality that the number of wins over the 25 years under consideration were as shown in

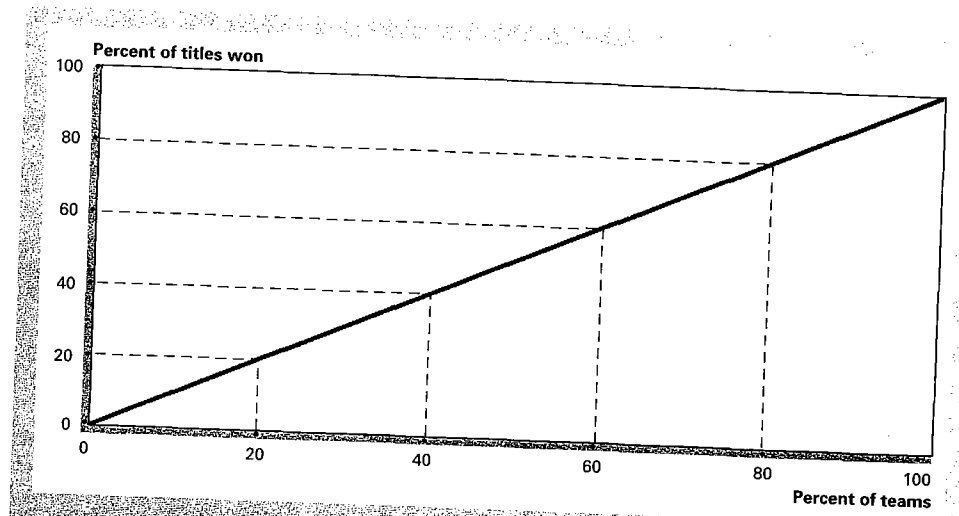


FIGURE 5A.1

The Lorenz Curve with Perfect Equality

TABLE 5A.1

Distribution of Championships over 25 Years

Team	*Number of Titles*	*Cumulative Percentage*	*Perfect Equality*
A	1	4.0 = (1/25)*100	20
B	2	12.0 = (3/25)*100	40
C	5	32.0 = (8/25)*100	60
D	7	60.0 =(15/25)*100	80
E	10	100 = (25/25)*100	100

Table 5A.1. The far-left column shows the cumulative distributions we would expect under perfect equality.

Figure 5A.2 shows the distribution of titles from Table 5A.1 as well as the line of perfect equality. As you can see, when competition is unbalanced, the line bows downwards, away from the line of perfect equality. The more unequal is competition, the more the line will be bowed downward. Thus, the

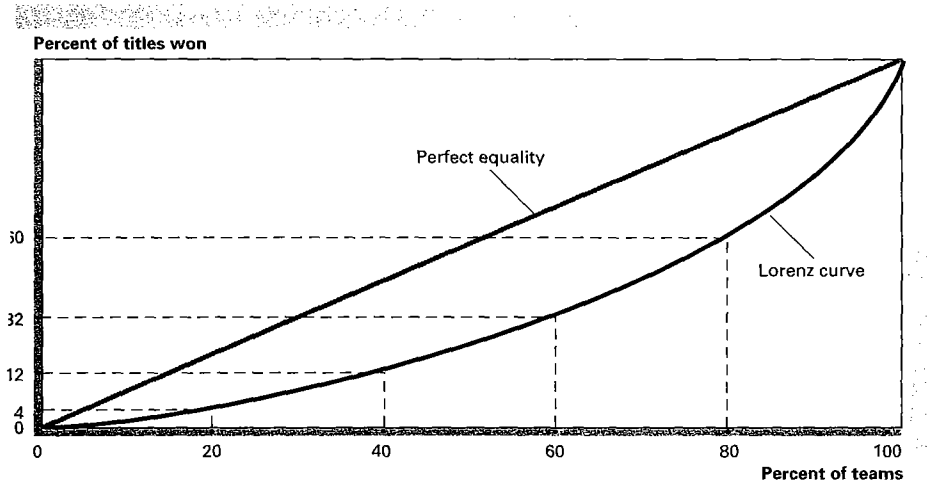


FIGURE 5A.2

The Lorenz Curve with an Unequal Distribution of Wins

Lorenz curve is useful to compare relative equality of competition from one time period to another or from one sport to another.[21]

A relatively new measure of interseason balance is a new application of an existing technique known as the **Markov process.** A Markov process shows the probability that a person or group changes in some particular way from one period to the next. Hadley, Ciecka, and Krautmann (2003) have applied this technique to baseball to evaluate competitive balance in the 1990s. They use the Markov process to measure the probability that a given team's performance in one season depends on its performance in the previous season. Their model is structured such that a team can be in one of three "states": winner, contender (within six games of qualifying for the playoffs), and loser (a team that finishes more than six games out of the playoffs).[22] At

[21]A more quantitative basis of comparison from the same data would be to construct a statistic known as the GINI coefficient, which measures the ratio of the area between the line of perfect equality and the Lorenz curve, and the area under the lower triangle. The closer the GINI coefficient is to zero, the closer the league is to perfect equality. A GINI coefficient of 1.0 would indicate perfect inequality (one team wins all of the championships).

[22]Larry Hadley, James Ciecka, and Anthony Krautmann, "Competitive Balance in the Aftermath of the 1994 Players' Strike," *Journal of Sports Economics*, v. 6, no. 4 (November, 2005), pp. 379–389. The authors acknowledge that the choice of six games back at the breakpoint between contender and loser is somewhat arbitrary, but it indicates that a contending team had a chance to be a winner up until the last week of the season.

the end of each season, every team must be in one of these three states. The appeal of this model is it is very simple to compute and interpret results, yet it is highly revealing.

From a competitive balance standpoint, we are most interested in the probability that a team's quality changes (i.e., it moves from one state to another over time). We can calculate a set of three "transitional probabilities" for each group (winners, contenders, and losers). For example, P_{WW} is the probability that a team will go from being a winner in one season to a winner in the next season, P_{WC} is the probability that a team will go from being a winner in one season to a contender in the next season, and P_{WL} is the probability that a team will go from being a winner in one season to a loser in the next season.

If team quality is highly variable from year to year, teams would change from one state to a different state more frequently than if some teams were usually good while others were usually bad. As a benchmark for competitive balance, the authors choose a world where a team's finish in one year has no impact on how it will do the next year. For example, a team that was a winner this year has a probability of one third, or .33, of being a winner, a contender, or a loser the next year. Table 5A.2 shows the states and probabilities that exist for the three-state baseball model. With perfect balance, each of these probabilities would equal .33.

To see how transitional probabilities are calculated, let's assume that we are interested in the competitive balance for a 3-team league over a 10-year period. If, in 5 of the 10 years, the winner repeated the following year, $P_{WW} = 5/10 = .50$. If in 3 of the 10 years, the winner dropped to being a contender, $P_{WC} = 3/10 = .30$. In the other two years, the contender transitioned to loser, $P_{WL} = 2/10 = .20$.

Table 5A.3 shows the changes in the transitional probabilities that occurred in the aftermath of the 1994 players' strike, a period widely cited as being characterized by a marked decline in competitive balance.

These probabilities shed light on the probability that teams can become winners at some time over a period of seasons and whether those probabilities have changed over time. The table shows that very few of the state probabilities are close to 0.33. It also shows that P_{WW} increases markedly in the post-1994

TABLE 5A.2

Conditional Probabilities for a Model with Three States

Winners	*Contenders*	*Losers*
P_{WW}	P_{CW}	P_{LW}
P_{WC}	P_{CC}	P_{LC}
P_{WL}	P_{CL}	P_{LL}

TABLE 5A.3

Markov Conditional Probabilities for Major League Baseball in the Pre-1994 and Post-1994 Eras

Transitional Probabilities	*Pre-1994*	*Post-1994*
P_{WW}	0.20	0.55
P_{WC}	0.18	0.11
P_{WL}	0.61	0.34
P_{CW}	0.27	0.33
P_{CC}	0.14	0.11
P_{CL}	0.59	0.56
P_{LW}	0.12	0.14
P_{LC}	0.12	0.08
P_{LL}	0.76	0.78

Source: Hadley, Ciecka, and Krautmann, "Competitive Balance in the Aftermath of the 1994 Players' Strike" (2005).

strike era, from 0.20 to 0.55 while P_{WL} decreases from 0.61 to 0.34. Both of these statistics indicate a decrease in interseason competitive balance in the post-strike era.