The Impact of Free Agency and the Salary Cap on Competitive Balance in the National Football League

ANDREW LARSEN
Iowa State University

AJU J. FENN
ERIN LEANNE SPENNER
Colorado College

This article examines the competitive balance of the National Football League (NFL) using Gini coefficients and the deviations of the Herfindahl-Hirschman Index. The authors present upper bounds for both measures that are constructed using actual playing schedules. They model competitive balance as a function of player talent, the incidence of strikes, the expansion of the NFL, change in the number of playoff teams, schedule length, team relocation, the number of new stadiums, and the introduction of free agency and the salary cap in the NFL using data from the 1970 to 2002 seasons. The authors find that free agency and salary cap restrictions tend to promote competitive balance, whereas a concentration of player talent reduces competitiveness among teams. Player strikes and the construction of new stadiums also affect competitive balance.

Keywords: *competitive balance; Herfindahl-Hirschman Index; National Football League; Gini coefficients; free agency; salary cap*

The competitive balance of a sports league can be described as the distribution of wins in that league. Leagues that have a roughly even distribution of wins among teams are said to have better competitive balance than those that do not. A chief determinant of the gate revenue earned by a franchise is the quality of its games. According to a study by El-Hodiri and Quirk (1971), predictable outcomes depress

AUTHORS' NOTE: Erin LeAnne Spenner thanks Colorado College for reseach support for the project. Andrew Larsen expresses gratitude to the University of St. Thomas for research support for this project.
Correspondence concerning this article should be addressed to Aju J. Fenn, Colorado College, Department of Economics and Business, 14 E. Cache La Poudre Street, Colorado Springs, CO 80903; e-mail: afenn@coloradocollege.edu.

JOURNAL OF SPORTS ECONOMICS, Vol. 7 No. 4, November 2006 374–390
DOI: 10.1177/1527002505279345

attendance. Also, Quirk and Fort (1992) find that the National Football League's (NFL) Cleveland Browns began losing attendance in the years that they dominated the league. These studies support the notion that sports fans lose interest when a league lacks parity. When fans lose interest, franchises lose money. Given that competitiveness influences profitability, there is a motive for leagues to promote competitive balance. Lowry (2003) states that NFL revenues totaled $4.3 billion in 2001. The NFL leads the professional sports industry in revenue. If competitive balance and attendance are connected, then a study of the determinants of competitive balance in the NFL is warranted.

The purpose of this article is to examine the determinants of competitive balance in the NFL. In particular, we wish to investigate how the introduction of a combination of free agency and a salary cap in 1993 has affected competitive balance in the NFL.[1] A casual inspection of Super Bowl participants suggests that free agency and the salary cap may have had an impact on competitive balance in the NFL. Since the introduction of free agency and the salary cap, 13 different teams have gone to the Super Bowl. During the 8 years before free agency and the salary cap, only 8 different teams reached the championship game. This article examines the evolution of competitive balance within seasons in the NFL over a time span of 30 years.

The article proceeds as follows. The first section reviews the relevant literature on competitive balance. The next section motivates the discussion of competitive balance using Gini coefficients for the NFL à la Schmidt and Berri (2001) and Utt and Fort (2002). The third section outlines the modifications to Depken's (1999) methodology to measure competitive balance in the NFL. The fourth section describes the data and the regression model. The next section discusses the empirical results. The article concludes with a discussion of the implications of our empirical results.

CURRENT RESEARCH ON COMPETITIVE BALANCE

There is an abundant literature on competitive balance. Summaries of the literature and important issues are discussed by Zimbalist (2002) and Fort and Maxcy (2003). As Fort and Maxcy point out, the empirical literature on competitive balance has bifurcated into two major streams: studies that investigate the relationship of competitive balance, attendance, and business practices in professional sports and those that concern themselves with the measurement and analysis of competitive balance.

A brief summary of the literature on competitive balance, attendance, and business practices follows. Leeds and Kowaleski (1999) measure free agency's impact on the distribution of salaries in the NFL. Berri and Vicente-Mayoral (2000) examine competitive balance and its impact on the failure of leagues. Richardson (2000) examines the National Hockey League and free agency. Drahozal (1986) and La Croix and Kawaura (1999) examine the distribution of league championships as a measure of competitive balance. McMillan (1997) examines player movement and its implications on the competitive balance of rugby in New Zealand. We turn next to the set of studies that concentrate on the analysis of competitive balance.

Rottenberg (1956) measured competitive balance in Major League Baseball (MLB) by counting the number of league pennants each team had won from 1920 to 1951. Measures of the dispersion of wins or championships are also common in the literature. One such measure was proposed by Noll (1988) and first applied by Scully (1989). This approach involves computing the actual standard deviation of the winning percentages of teams in a league and comparing it to the idealized standard deviation of winning percentages that would have resulted if all the teams were of equal strength.

Previous studies have focused their attention on baseball, especially the effect of the reserve clause on competitive balance. The reserve clause was removed in 1976, thus opening the door to free agency in MLB. Scully (1989) and Quirk and Fort (1992) use the standard deviation of winning percentages to estimate the impact of free agency on competitive balance in MLB. Their findings suggest that there has been no change in the competitiveness of the American League but some increase in competitiveness in the National League that can be attributed to the removal of the reserve clause.

Schmidt (2001) takes a different approach to measuring competitive balance. He uses Gini coefficients to measure changes in competitiveness rather than a traditional standard deviation measurement. His findings suggest that baseball league expansion promotes competitive balance. However, his research examines only the impact of expansion on competitive balance. Gini coefficients as a measure of competitive balance in baseball are revisited in Schmidt and Berri (2001). Fort and Quirk (1995) also use Gini coefficients to measure the impact of free agency on competitive balance in various professional sports leagues. The next section of this article examines competitive balance in the NFL using Gini coefficients.

Eckard (2001) examines the impact of free agency on competitive balance in MLB using an analysis of variance–based measure of competitive balance. This measure is useful when studying leagues in which the rankings of teams (in terms of winning percentages) change but the overall dispersion of the winning percentages stays the same. This new measure captures fluctuations in team performance from year to year. Eckard finds that the induction of free agency did indeed increase competitiveness in MLB. Humphreys (2002) presents yet another measure of competitive balance, the competitive balance ratio (CBR). He argues that the CBR is a more accurate indicator of competitive balance over time than other measures. The CBR is a ratio of the average of a given team's standard deviations of win/loss ratio across seasons divided by the average of the standard deviation of winning percentages for the league during the same number of seasons. The ratio is a number between 0 and 1, with 1 being perfect competitive balance over time and 0 being no competitive balance over time. Both Eckard and Humphreys provide measures that are useful when it comes to studying competitive balance in leagues in which the variation in winning percentages stays the same but the teams that do the winning change from year to year. This is not the case in the NFL. The standard deviation of winning percentages (across teams) for a given season declined steadily from a value of .26 in 1970 to a value of .16 in 2002.

Depken (1999) presents a new measure of competitive balance that allows us to test for the impact of various business practices, such as free agency, racial integration, and so on, on the competitive balance of MLB. His study controls for other factors that previous studies could not by using the Herfindahl-Hirschman Index (HHI), a measure most often used to examine market concentration in industrial organization. In his model, Depken controls for factors such as talent distribution, the integration of African American players into the league, the expansion of new teams, and the free agency of players. His stochastic framework proves beneficial in analyzing the impact of free agency on parity in MLB. According to Depken, prior studies that used frequencies of winning percentages, postseason appearances, or standard deviation of wins could not control for these exogenous factors that influence the competitive nature of baseball.

The present study modifies Depken's (1999) framework to examine competition in the NFL. To date, there have been relatively few economic studies conducted on competitive balance in the NFL. No studies have been done on the impact of free agency, expansion, and strikes on competitive balance in the NFL. This study investigates the impact of such factors on competitive balance in the NFL by adapting Depken's methodology. The next section presents a summary of competitive balance in the NFL using Gini and adjusted Gini coefficients.

COMPETITIVE BALANCE VIA ADJUSTED GINI COEFFICIENTS

Schmidt (2001) and Schmidt and Berri (2001) use Gini coefficients as a measure of competitive balance in sports leagues. Although the Gini coefficient is commonly used as a method to examine income or wealth distributions, this conventional economic measure is also useful for examining the distribution of wins in a sports league. The Gini coefficient always lies between 0 and 1, where 0 indicates a perfectly equal distribution of wins among teams and 1 corresponds to a single team winning all the games that are played. A league is said to have perfect competitive balance when every team wins exactly half of its games. A Gini coefficient closer to 1 is indicative of a league in which a few teams win most of the games that are played. Following Lambert (1993), Schmidt and Berri (2001) define the Gini coefficient as follows:

$$G_i = \left(1 + \frac{1}{N_i}\right) - \frac{2}{N_i^2 \mu_{xi}} \times \left(x_{N,i} + 2 \times x_{N-1,i} + 3 \times x_{N-2,i} + \ldots + N \times x_{1,i}\right), \tag{1}$$

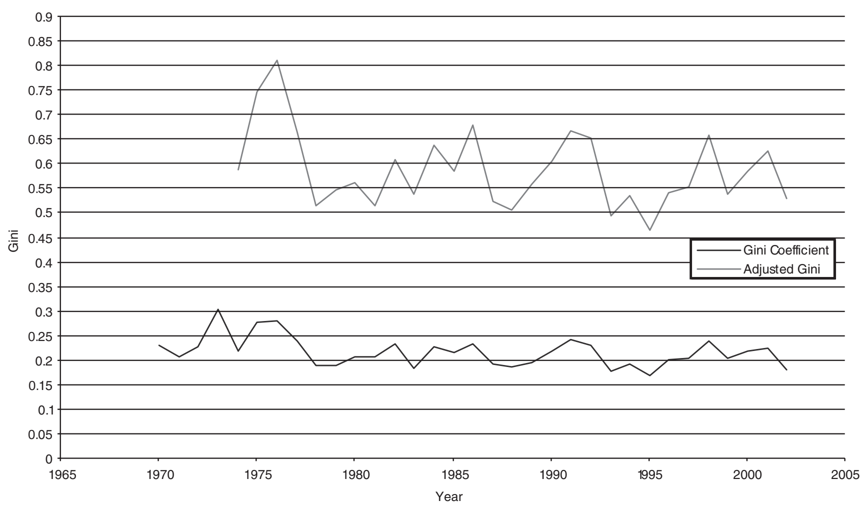

Figure 1: Gini and Adjusted Gini Coefficients

where N_i represents the number of teams in year i, x_N is the winning percentage of the Nth team, and μ_{xi} is the average value of x_i. The present analysis adopts the same measure used by Schmidt and Berri (2001), following Lambert (1993), to calculate the Gini coefficients for the NFL. Statistics for winning percentage are compiled from Carroll, Gershman, Neft, and Thorn (1999) and the NFL's Web site (http://www.nfl.com) for each year from 1970 to 2002. A graph of the Gini and adjusted Gini coefficients for the NFL is displayed in Figure 1.

Utt and Fort (2002) discuss certain problems associated with measuring competitive balance in sports using Gini coefficients. The conventional calculation of Gini coefficients does not accurately specify a completely unbalanced distribution of winning percentages across a league. In the traditional calculation, perfect inequality among teams would be indicated by a Gini coefficient of 1, meaning that one team wins all of the games, and the rest of the teams win no games. Even in the most unbalanced league, one team cannot win all of the league's games. Rather, one team can at most win all of its games, which are only a fraction of the total number of league games. This means that the upper bound of the Gini coefficient for a sports league is much less than 1. Failing to recognize this leads to an overstatement of the competitive balance of a sports league. Utt and Fort suggest that team schedules must be taken into account when calculating Gini coefficients for sports leagues.

To address the problems discussed by Utt and Fort (2002), we calculate a hypothetical upper bound for a "most unequal distribution" of wins as described by Fort and Quirk (1997). The most unequal distribution of wins is created by looking at team schedules for each year and assuming that one team wins all of its games, the next team wins all of its games except for the ones it lost to the first team, and so on until the last team, which loses all of its games. The most unequal distribution in this study is created by assuming that wins are distributed in alphabetical order by teams' geographic locations, with the Arizona Cardinals going undefeated and the Washington Redskins being the hopeless team that never wins. It is important to note that the NFL's schedule allows for more than one team to go undefeated and, likewise, more than one team to go winless, depending on their schedules. A new adjusted Gini coefficient for each season was calculated using the most unequal distribution of wins as the upper bound. This upper bound is much closer to the Lorenz curve than the axes, which serve as the upper bound for the traditional Gini calculation.[2] Figure 1 displays the traditional Gini coefficients calculated for the NFL for every year from 1970 to 2002 and the adjusted Gini coefficients from 1974 to 2002.

As suggested by Utt and Fort (2002), the traditional Gini coefficients do appear to understate the level of competitive balance in the NFL. When the Gini coefficients are adjusted by constructing the upper bound according to the actual NFL schedule, the values more than double. It is interesting to note that according to both the adjusted and unadjusted coefficients, competitive balance appears to be cyclical, fluctuating up and down from year to year. This would suggest that a year in which wins are more evenly distributed is often followed by a year in which wins are more concentrated. Also, there are several years in which dramatic shifts in the adjusted Gini coefficient can be seen. For example, between 1976 and 1978, there was a relatively large decrease in the adjusted Gini coefficient from .81 to approximately .51. This could be because two expansion teams joined the league in 1976, and the league implemented a 16-game schedule in 1978.[3] Figure 1 displays drops in the value of both indices in 1987, when there was a strike, and in 1993, which corresponds to the introduction of free agency and the salary cap. These issues are formally introduced as testable hypotheses in the regression model.

MEASURING COMPETITIVE BALANCE WITH THE DEVIATION OF THE HHI

To measure competitive balance, Depken (1999) uses the deviation of the HHI (dHHI) from the "most equal distribution" of wins. He shows that the HHI is mathematically related to the standard deviation of wins. The HHI has been applied to many different industries to examine market structure. Among others, Borenstein (1989) and Evans and Kessides (1993) use it to examine market structure in the airline industry, and Gilbert (1984) applies the HHI to banking. Sullivan (1985) and Sumner (1981) examine market structure in the cigarette industry.

The HHI index is a quadratic summation of all firm market shares in an industry. It is defined as

$$\mathrm{HHI} = \sum_{i=1}^{N} \mathrm{MS}_i^2, \tag{2}$$

where MS_i is the market share of the ith firm, and N is the number of firms. For certain industries, acquiring output data for all of the firms in the market may be problematic, making it difficult to accurately measure the HHI. However, in the NFL, output is measured in total wins, and such data are readily available. A team's market share is its percentage of wins in the league in a given season.[4] Following Depken (1999), the definition of HHI is

$$\mathrm{HHI} = \sum_{i=1}^{N} \left[\frac{2\,\mathrm{Wins}_i}{NG} \right]^2, \tag{3}$$

where Wins_i is the number of wins for the ith team, N is the number of teams in the league, and $G_i = G$ is the equal number of games played by each team during a given season. Depken, in keeping with Quirk and Fort (1992), assumes that perfect parity is a situation in which every team wins half of its games. The value of the HHI for the most equal distribution of wins, as implied by Equation 3, is $1/N$.

Because of the nature of the index, the value of the HHI decreases as the number of teams in the league increase. This factor must be taken into account, because the NFL has undergone various periods of expansion. To correct the downward bias in the HHI due to expansion, in keeping with Depken (1999), we define dHHI as

$$\mathrm{dHHI} = \mathrm{HHI} - \frac{1}{N}. \tag{4}$$

According to Depken (1999), the standard deviation of winning percentages does not specifically account for expansion and may increase when new teams join the league. The most equal distribution of wins occurs when each team wins exactly half of the number of games it plays. This results in a dHHI value of 0.[5] Utt and Fort's (2002) critique may be extended to the upper bounds for the HHI and dHHI. The upper bound for the HHI in the NFL is not unity, as it is in other industries. This is because even if one team were to win all of its games, no single team could win all of the games played in the league. The most unequal distribution for the dHHI is created as in the case of the adjusted Gini coefficient by consulting actual playing schedules and by assuming that wins are distributed in alphabetical order. The Arizona Cardinals are assumed to be the undefeated team, and the Washington Redskins are assumed to be the hopeless team that never wins.[6] Given the most unequal distribution of wins, hypothetical upper bound values for the dHHI were calculated for the 1974 to 2002 seasons. These bounds are also shown in Figure 2.

Generally speaking, the dHHI, as the measure of parity, may be used as a dependent variable in a regression model to test specific hypotheses about the influ-

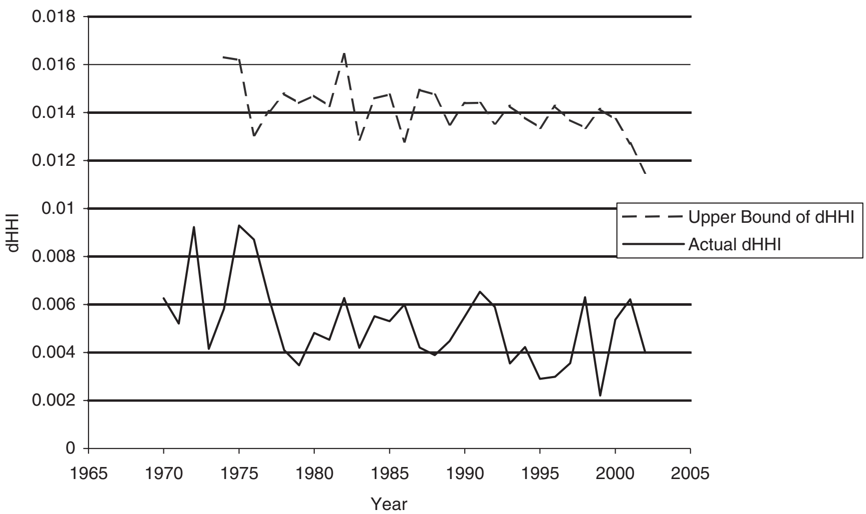

Figure 2: Actual Deviation of the Herfindahl-Hirschman Index (dHHI) and the Upper Bound of dHHI

ence of exogenous variables, such as the distribution of talent, league expansion, strike years, and free agency. The next section outlines this model.

DATA AND REGRESSION MODEL

The dependent variable, dHHI, measures the concentration of wins as described above. The data on wins, losses, points scored by a team, and points scored against a team are collected on an individual-game basis and then aggregated up to annual totals. The data used in this model come from NFL regular-season games played from 1970 to 2002. Data are compiled from Carroll et al. (1999) and the NFL's Web site.

The model used for this study is as follows:

$$\begin{gathered}\mathrm{dHHI}_t = f(\mathrm{FA/SAL},\ \mathrm{EXPAN}_t,\ \mathrm{STRIKE1},\ \mathrm{STRIKE2},\ \mathrm{STRIKE3},\\ \mathrm{PLAYER\ TALENT}_t,\ \mathrm{SCHEDULE\ LENGTH},\ \mathrm{PLAYOFF\ SPOTS},\\ \mathrm{RELOCATE},\ \mathrm{NEW\ STADIUMS}).\end{gathered} \tag{5}$$

The definitions and descriptive statistics of the variables in Equation 5 are displayed in Table 1. dHHI_t is the deviation of the HHI from the most equal distribution of wins in year t, as described in Equation 4.

TABLE 1: Variable Definitions and Descriptive Statistics

Variable	*Definition*	M	SD
$dHHI_t$	Deviation from ideal concentration of wins	0.0052	0.0017
FA/SAL	Dummy variable (0 = before free agency)	0.3030	0.4667
$EXPAN_t$	Dummy variable (1 = expansion year)	0.1212	0.3314
$HHIPA_t$	Concentration of points scored against the team	0.0368	0.0023
$HHIPF_t$	Concentration of points scored for the team	0.0370	0.0022
STRIKE1	Dummy variable (1 = strike year 1973)	0.0303	0.1741
STRIKE2	Dummy variable (1 = strike year 1982)	0.0303	0.1741
STRIKE3	Dummy variable (1 = strike year 1987)	0.0303	0.1741
SCHEDULE LENGTH	Dummy variable (1 = years 1970 to 1977)	0.2727	0.4523
PLAYOFF SPOTS	Change in number of playoff spots	1.3939	0.9334
NEW STADIUMS$_t$	Number of new stadiums in the past 5 years	0.6364	0.8223
$RELOCATE_t$	Number of teams that relocated in the past 5 years	0.1818	0.4647

The NFL instituted free agency along with a salary cap in 1993.[7] Because the addition of free agency to the league and the institution of a salary cap occurred at the same time, it is difficult to separate the two for the purpose of measuring their individual contributions to the competitive balance. We attempt to measure the impact of this joint policy using a dummy variable, FA/SAL, that takes on a value of 0 before 1993 and a value of 1 for 1993 and subsequent years. The NFL's salary cap in combination with free agency reduces the possibility that one team is able to sign all of the premier talent in the free agent pool, because each team has a fixed amount that it may spend on players' salaries in a given year. The combination of free agency and the salary cap may enhance competitive balance in the NFL. The predicted sign of the coefficient of the dummy variable FA/SAL is negative because the introduction of this policy should lower the value of dHHI.

The NFL has seen the addition of several new teams since 1970. NFL expansion teams are constructed through an expansion draft. All else equal, it could be the case that the addition of a team may promote competitive balance because of talent dispersion among a larger number of teams. In an expansion draft, each team in the league must allow for the expansion team to draft certain players. However, the existing teams are able to protect most of their top talent and leave very little talent for the new teams, making it difficult for them to field teams that are up to par with the rest of the league. Using $dHHI_t$ as the dependent variable eliminates the downward bias that is prevalent in the HHI when leagues expand. However, as Depken (1999) points out, league expansion needs to be included as an exogenous variable in the model. We capture the effects of expansion through a dummy variable called $EXPAN_t$ that takes on a value of 1 for the years during which new teams join the league and a value of 0 otherwise. The maintained hypothesis is that expansion will lead to competitive balance by dispersing existing talent among more teams. Thus, the predicted sign of the coefficient of $EXPAN_t$ is negative.

Another variable of interest is the effect of a strike on competitive balance. During the 1973 strike, teams lost preseason play because of the ongoing negotiations between owners and players. The strike may have affected competition during the year because teams had less time to assess their talent during the preseason. Strikes by players led to shortened seasons in 1982 and 1987, with replacement players being employed in 1987. This may have promoted competitive balance or hindered it, because each team was forced to find replacement players. The talent levels of these replacement players may be homogeneous or heterogeneous among teams depending on the comparative ability of each team to find replacement players. We treat the three strikes as individual occurrences by creating three dummy variables, STRIKE1, STRIKE2, and STRIKE3.[8] Each strike dummy takes on a value of 1 in the strike year (STRIKE1 = 1 for 1973, STRIKE2 = 1 for 1982, and STRIKE3 = 1 for 1987) and a value of 0 otherwise. Although we control for the impact of strikes on competitive balance, we do not posit specific signs for these variables.

Depken (1999) points out that the competitive balance of a league is affected by the distribution of player talent. Thus, the dispersion of talent among teams must be accounted for in the model. Depken included HHI measures of runs scored and runs allowed as explanatory variables to account for the impact of offensive and defensive talent concentration on competitive balance. We adapt his method to account for the nuances of the NFL. Equation 6 presents our attempt at quantifying the dispersion of player talent in the NFL. Player talent in year t is measured by the sum of HHIPF_t and HHIPA_t. HHIPF_t is the HHI of points scored by a team in year t, and HHIPA_t is the HHI of points scored against a team in year t.

$$\text{PLAYER TALENT}_t = \text{HHIPF}_t + \text{HHIPA}_t. \tag{6}$$

HHIPF_t is defined as

$$\text{HHIPF}_t = \sum_{i=1}^{N} \left[\frac{\text{Points Scored}_i}{\text{Total Points Scored in the League}} \right]^2, \tag{7}$$

where Points Scored_i is the total number of points scored by the ith team in year t. All else equal, if offensive talent is concentrated on a few teams, these teams will score most of the points in the league. Thus, higher values of HHIPF_t correspond to a decrease in competitive balance, which implies higher values of the dependent variable (dHHI_t). In other words, higher concentrations of offensive talent go hand in hand with lower levels of competitive balance. However, some of the points scored by a team may be due to defensive talent or special teams talent. Defenses may score points either on fumble recoveries or on interceptions that are returned for touchdowns, or they may put the offenses of their teams in favorable positions, resulting in scores. Similarly, a good running offense may help a team keep the opposing offense out of the game, thereby limiting the number of points given up.

Special teams may aid offense or defense or both by establishing favorable field positions or returning punts or kickoffs for touchdowns. It is for these reasons that we choose to combine the measures HHIPF_t and HHIPA_t into a single measure called PLAYER TALENT.[9]

We use a similar measure (HHIPA_t) to loosely proxy the distribution of defensive player talent among teams, because offenses and special teams can do their parts to keep other teams from scoring:

$$\text{HHIPA}_t = \sum_{i=1}^{N} \left[\frac{\text{Points Scored Against}_i}{\text{Total Points Scored in the League}} \right]^2, \tag{8}$$

where $\text{Points Scored Against}_i$ is the total number of points that the ith team gives up in year t. All else equal, the concentration of defensive talent on a few teams would lead to fewer points scored against those teams and large numbers of points scored against the teams with weaker defenses. Thus, larger values of HHIPA_t go hand in hand with lower levels of competitive balance and higher values of dHHI_t. Because both components of player talent are hypothesized to vary positively with competitive balance, the expected sign of the coefficient of PLAYER TALENT in the regression equation is positive. The concentration measures of player points for (HHIPF_t) and points against (HHIPA_t) certainly depend on player talent but may also incorporate the effects of other variables, such as coaching, the training staff, and so on. The purpose of this variable is to control for these factors to isolate the impact of the FA/SAL variable on competitive balance. In addition, the concentration of player talent on a team is related to the status of free agency in the league. We use an instrumental-variables approach to tackle this problem in the "Estimation and Results" section.[10]

From 1970 to 2002, there were changes in the length of the season, the number of playoff spots, and the locations of some teams, and new stadiums were constructed. To control for these variables, we use the following variables: SCHEDULE LENGTH, PLAYOFF SPOTS, NEW STADIUMS, and RELOCATE.[11]

SCHEDULE LENGTH is a dummy variable that takes on a value of 1 for the years 1970 to 1977 and value of 0 otherwise. During this time, the season was 14 games long. After this period, the season was extended to 16 games. Football is a physical game, and a longer season may lead to more injuries. Injuries to key players on teams may affect the competitive balance of the league.

PLAYOFF SPOTS is the change in the number of teams that were allowed to enter the playoffs. From 1970 to 1979, there were 8 playoff spots. From 1980 to 1989, there were 10 playoff spots. Since 1990, there have been 12 playoff spots.[12] Teams that may not better their positions in the playoff race (i.e., make the playoffs or gain home-field advantage) are likely to rest their starters. Similarly, teams that are mathematically eliminated from the playoffs are likely to use the rest of the season to evaluate younger players. Thus, the number of teams that are allowed to enter the playoffs may affect competitive balance during the regular season and should be controlled for in the model.

NEW STADIUMS$_t$ is the cumulative sum of the number of new stadiums that opened across the league in the past 5 years over the span of the data set. This variable is included to control for any impact that a new stadium may have on competitive balance on the field. The impact may be in the form of larger home crowds who have turned out to see the new stadium or new luxury box revenues that are exempt from revenue sharing.

Over the time frame of the study, several teams relocated from one city to another. The Oakland Raiders relocated to Los Angeles and back, the Cleveland Browns moved to Baltimore and became the Ravens, the Baltimore Colts moved to Indianapolis, the St. Louis Cardinals moved to Arizona, and so on. Relocation may affect competitive balance through changes in the marginal revenue product of the team that may be the result of a change in market size and/or increased attendance from the novelty of being a new team in town. RELOCATE$_t$ is the cumulative sum of the number of teams that have relocated in the past 5 seasons over the span of the data set.[13]

ESTIMATION AND RESULTS

To estimate the model, an instrumental-variables approach is used. As stated earlier, the concentration of playing talent on a team may be connected to the status of free agency and the salary cap in the NFL. With the introduction of free agency and the salary cap, the top teams were no longer able to hoard the best players in the league. This in turn should lead to a more equal distribution of playing talent and consequently of wins among teams. The potential endogeneity of playing talent and the free agency dummy variable is resolved by using an instrumental-variables approach. We use the number of free agents (who change teams in a given year in the NFL) along with other truly exogenous right-hand-side variables as instruments for PLAYER TALENT.[14] Data on the total number of free agents are not available for our entire time frame. Thus, we are limited to using the number of free agents who actually change teams. The instrument set explains about 85% of the variation in player talent, and the number of free agents who change teams in a given year is statistically significant in explaining the PLAYER TALENT variable.[15] A Sargan test confirms the validity of the instrument in that it is uncorrelated with the error term at the 1% significance level.[16] The results are reported in Table 2.[17]

The variable for free agency and the salary cap (FA/SAL) is negative and statistically significant at the 10% significance level. This implies that with the advent of free agency along with a salary cap, the competitiveness of the NFL has increased. This differs from the findings of Depken (1999), who discovered that the induction of free agency into MLB has no effect on parity in the National League but has reduced parity in the American League. Our results suggest that the NFL has

TABLE 2: Instrumental-Variables Estimation Results

Variable	*Coefficient* (t *statistic)*
FA/SAL	–.0012 (–1.95)**
EXPAN_t	–.0009 (–0.76)
PLAYER TALENT	.0350 (7.48)*
STRIKE1	–.0035 (–4.41)*
STRIKE2	.0014 (4.20)*
STRIKE3	–.0012 (–3.64)*
SCHEDULE_t	.0029 (4.12)*
PLAYOFF SPOTS	.0012 (5.2500)*
STADIUM_t	.0003 (2.6300)*
RELOCATE_t	–.0001 (–0.25)
R^2	.53

*Significant at the 5% level. **Significant at the 10% level.

become more competitive because of free agency and the salary cap. The expansion of the league does not have a statistically significant impact on competitive balance.

Player talent concentration is statistically significant and positive. This shows that a concentration in the offensive, defensive, or special teams talent in the league leads to reduced competitive balance. It makes sense that if one team possesses an extreme amount of talent, it will become dominant, and this will reduce the competitive balance in the league. On the other hand, a more equal dispersion of playing talent among teams promotes competition and leads to lower dHHI values.

The coefficient on the occurrence of strikes is negative and significant for STRIKE1 and STRIKE3 and positive and significant for STRIKE2. The changes in sign suggest that different strikes affect competitive balance differently. Alternatively, the change in sign may be due to the use of annual dummy variables that are used to model the three separate strikes. These coefficients pick up year-specific effects, including the impact of the strikes, and it may not be accurate to attribute the magnitude and significance of these coefficients to the strikes alone.

SCHEDULE_t is positive and significant, which suggests that the shorter 14-game season was more competitively balanced than the current 16-game season. The change in the number of playoff spots (PLAYOFF SPOTS) and the emergence of new stadiums (STADIUM_t) are both positive and significant, which suggests that the increase in playoff spots and the occurrence of teams getting new stadiums reduces competitive balance. The number of teams that relocate in the past 5 years (RELOCATE_t) does not affect competitive balance significantly.

In keeping with Depken (1999), a second model is presented, with standard deviations replacing dHHI and HHI measures. Annual winning percentages are calculated for each team in a given year. The standard deviation of these winning percentages (STDWIN%) in a given year is substituted for dHHI_t. Larger variances

TABLE 3: Instrumental-Variables Estimation With Standard Deviations

Variable	*Coefficient (*t *statistic)*
SDPFPA	.0022 (16.39)*
EXPAN	–.0122 (–0.711)
FA/SAL	–.0025 (–0.23)
STRIKE1	–.0214 (–2.08)*
STRIKE2	.1149 (16.90)*
STRIKE3	.0398 (5.11)*
RELOCATE	–.0036 (–1.29)
STADIUM	–.0084 (–2.45)*
SCHEDULE	.0334 (2.56)*
PLAYOFF SPOTS	–.0134 (–1.83)
R^2	.61

*Significant at the 5% level.

in winning percentages among teams would indicate that a league has lesser competitive balance than a league with a smaller variance in winning percentages.

Similar measures of standard deviation are used to measure offensive and defensive talent distributions. The total number of points scored by a team is calculated for a given year. The standard deviation of the total points scored by teams across the league in a given year is used as a proxy for the distribution of offensive talent $HHIPF_t$. Similarly, the total number of points scored against a given team is calculated for a given year. The standard deviation of the points scored against the different teams across the league is used to replace $HHIPA_t$. PLAYER TALENT is now given by the sum of the standard deviation among points scored by different teams and the standard deviation of the points scored against different teams in a given year. By including these alternative measures of competitive balance and player talent distribution, we are able to see that the findings of the model do not depend on the use of the dHHI and the HHI to measure competitive balance and player talent distribution, respectively. The definitions of the other variables (i.e., FA/SAL, STRIKE1, STRIKE2, STRIKE3, $EXPAN_t$, $SCHEDULE_t$, PLAYOFF SPOTS, $STADIUM_t$, and $RELOCATE_t$) remain the same from the previous model. The results of the standard deviation model are reported in Table 3.[18]

The coefficient on the PLAYER TALENT variable (SDFPA) is positive and significant. STRIKE1, STRIKE2, and STRIKE3 are significant as well, but their signs do not always coincide with the results from the first model. The free agency dummy (FA/SAL) is not significant in this model. This may be due to the high multicollinearity that is detected between several of the right-hand-side variables, such as PLAYOFF SPOTS and SCHEDULE, EXPAN and STADIUM, and so on. When the model is estimated without these variables, the FA/SAL variable is negative and significant. The results here seem to support some of the testable hypotheses borne out by the results of the dHHI regression from Table 2. The increase in schedule length (SCHEDULE) reduces competitive balance in both regressions. This model seems to suggest that the cumulative number of stadiums over a 5-year period promotes competitive balance, whereas the previous model implies that the cumulative number of new stadiums reduces the level of competitive balance. This apparent contradiction may be due in part to the difference in the measures of competitive balance between the two models. In general, the standard deviation model does not perform as well as the dHHI model, as was the case in Depken (1999). We include the additional results for comparison purposes.

CONCLUSIONS

This research examines the impact of various determinants of competitive balance on the level of competitive balance in the NFL. Adapting the methodology developed by Depken (1999), competitive balance is measured by the concentration of wins in the NFL for each year from 1970 to 2000.

We find some evidence that the introduction of free agency and the salary cap has increased the competitive balance in the NFL. This study differs from those done on baseball because of the introduction of free agency in conjunction with a salary cap in the NFL. We also find evidence that the distribution of playing talent among teams influences competitive balance. Finally, our results support the notion that schedule length and the construction of new stadiums affect the competitive balance of the league. This study provides a starting point for further research on competitive balance in the most profitable sports league in the United States.

NOTES

1. A salary cap is a monetary constraint that limits how much a team may spend on players in a given year. Free agency may be loosely defined as the ability of a player to play for a team of his choice (presumably the highest bidder for his services) after his existing contract has expired.

2. For a detailed explanation with a diagram, see Utt and Fort (2002).

3. From 1970 to 1977, each NFL team played only 14 games.

4. The possibility of a tie does exist in the NFL. To account for this, each team that ties a game is given half of a win.

5. Given that each team wins half the games it plays (i.e., $\text{Wins}_i = G/2$), it follows from Equations 3 and 4 that dHHI = 0.

6. Shuffling the order in which teams finish does result in a different value for the upper bound of the dHHI, but the difference is detectable only after about seven decimal places.

7. Unrestricted free agents are players who are not contractually obligated to remain with their current teams. They may elect to join other teams. The salary cap is the maximum amount of money that an NFL team may spend on player salaries in a given year.

8. We are grateful to an anonymous referee for suggesting this approach.

9. Including both HHIPF_t and HHIPA_t in the regression results in high multicollinearity.

10. We acknowledge helpful suggestions from both referees on this issue.

11. We gratefully acknowledge suggestions from an anonymous referee for these variables.

12. The playoff brackets are available at http://rjccourt85.tripod.com/playoffsystem.html.

13. The relocation data are taken from http://users.pullman.com/rodfort/SportsBusiness/BizFrame.htm.

14. Data on the number of free agents who changed teams come from http://www.skillpositions.com.

15. These results are available on request.

16. The Sargan test is outlined by Gujarati (2003). The results are available on request.

17. Serial correlation is not detected using the procedure outlined by Woolridge (2003) for instrumental variables. Heteroskedasticity is also detected by a White (1980) test. In keeping with Depken (1999), a White correction of the standard errors is used.

18. Heteroskedasticity is detected by a White test (1980). In keeping with Depken (1999) a White correction of the standard errors is used.

REFERENCES

Berri, D., & Vicente-Mayoral, R. (2000, March). *Competitive balance measures across successful and unsuccessful sports leagues: Does competitive balance play a role in the failures?* Paper presented at the annual meeting of the Midwest Economics Association, Chicago.

Borenstein, S. (1989). Hubs and high fares: Dominance and market power in the U.S. airline industry. *Rand Journal of Economics*, *20*(3), 344-365.

Carroll, B., Gershman, M, Neft, D., & Thorn, J. (1999). *Total football II: The official encyclopedia of the National Football League*. New York: HarperCollins.

Depken, C. (1999). Free-agency and the competitiveness of Major League Baseball. *Review of Industrial Organization*, *14*(3), 205-217.

Drahozal, C. (1986). The impact of free agency on the distribution of playing talent in Major League Baseball. *Journal of Economics and Business*, *38*(2), 113-121.

Eckard, E.W. (2001). Free agency, competitive balance, and diminishing returns to pennant competition. *Economic Inquiry*, *39*(3), 430-443.

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

Evans, W., & Kessides, I. (1993). Localized market power in the U.S. airline industry. *Review of Economics and Statistics*, *75*(1), 66-75.

Fort, R., & Maxcy, J. (2003). Comment: Competitive balance in sports leagues: An introduction. *Journal of Sports Economics*, *4*, 154-160.

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

Fort, R., & Quirk, J. (1997). Introducing a competitive environment into professional sports. In W. Hendricks (Ed.), *Advances in the economics of sports* (Vol. 2, pp. 3-26). Greenwich, CT: JAI.

Gilbert, A. (1984). Bank market structure and competition: A survey. *Journal of Money, Credit and Banking*, *16*(4), 617-645.

Gujarati, D. (2003). *Basic econometrics*. New York: McGraw-Hill.

Humphreys, B. (2002). Alternative measures of competitive balance in sports leagues. *Journal of Sports Economics*, *3*, 133-148.

La Croix, S., & Kawaura, A. (1999). Rule changes and competitive balance in Japanese professional baseball. *Economic Inquiry*, *37*(2), 353-368.

Lambert, P. J. (1993). *The distribution and redistribution of income: A mathematical analysis*. Manchester, UK: Manchester University Press.

Leeds, M., & Kowaleski, S. (1999). *Labor market issues in team sports* (J. Fizel, Ed.). Westport, CT: Praeger.

Lowry, T. (2003, January 27). The NFL machine. *Business Week*, p. 89.

McMillan, J. (1997). Rugby meets economics. *New Zealand Economic Papers*, *31*(1), 93-114.

Noll, R. (1988). *Professional basketball* (Working Paper No. 144). Palo Alto, CA: Stanford University.

Quirk, J., & Fort, R. (1992). *Pay dirt: The business of professional team sports*. Princeton, NJ: Princeton University Press.

Richardson, D. (2000). Pay, performance, and competitive balance in the national hockey league. *Eastern Economic Journal*, *26*(4), 393-417.

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

Schmidt, M. (2001). Competition in Major League Baseball: The impact expansion. *Applied Economic Letters*, *8*(1), 21-26.

Schmidt, M., & Berri, D. (2001). Competitive balance and attendance: The case of Major League Baseball. *Journal of Sports Economics*, *2*, 145-167.

Scully, G. W. (1989). *The business of Major League Baseball*. Chicago: University of Chicago Press.

Sullivan, D. (1985). Testing hypotheses about firm behavior in the cigarette industry. *Journal of Political Economy*, *93*(3), 586-598.

Sumner, D. (1981). Measurement of monopoly behavior: An application to the cigarette industry. *Journal of Political Economy*, *89*(5), 1010-1019.

Utt, J., & Fort, R. (2002). Pitfalls to measuring competitive balance with Gini coefficients. *Journal of Sports Economics*, *4*, 367-373.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

Woolridge, J. (2003). *Introductory econometrics: A modern approach*. Mason, OH: Thomson Southwestern.

Zimbalist, A. (2002). Competitive balance in sports leagues: An introduction. *Journal of Sports Economics*, *3*, 111-121.

Andrew Larsen is at Iowa State University.

Aju J. Fenn is in the Department of Economics and Business at the Colorado College.

Erin LeAnne Spenner is in the Department of Economics and Business at the Colorado College.