Association Internationale des Economistes du Sport
AIES IASE
International Association of Sports Economists

Working Paper Series, Paper No. 06-20

Tilting the Playing Field: Why a sports league planner would choose *less*, not more, competitive balance

Stefan Szymanski[1]

October 2006

Abstract

It has traditionally been argued that the organizer of a sports league would prefer more competitive balance to the level that emerges in a noncooperative equilibrium. This argument has been used to justify restraints on competition between teams, which also tend to raise profits at the expense of players and consumers. This paper shows that in theory a planner would prefer *less*, not more, competitive balance. The paper uses data from Major League Baseball to show just how unbalanced a league planner would choose.

JEL Classification Codes: D71, L25, L41, L83

Keywords: competitive balance, sports leagues, baseball

* Original version completed October 2004. Draft: Comments welcome.

---

[1] Stefan Szymanski, Tanaka Business School, Imperial College London, South Kensington campus, SW7 2AZ, UK. Tel : (44) 20 7594 9107, Fax: (44) 20 7823 7685, e-mail: szy@imperial.ac.uk. This research was initiated while the author was on sabbatical at Smith College, and the author gratefully acknowledges the financial support of the Provost's Office Visiting Scholars Fund. I have benefited from useful discussions with Andy Zimbalist and Leo Kahane. I also thank Dave Berri, Eberhard Fees, Tom Hoehn, Gerd Muehlheusser and Stephen F. Ross for useful comments on an earlier draft.

1. Introduction

The claim that sports economics should be treated as a distinct field of research rests largely on a single issue: the "competitive balance problem".[2] Each team in a league contributes to a series of sporting contests (games) that by the end of a season produces a champion. In any given contest the home team must not only supply their own athletes, but also obtain the services of the visiting team. Demand for the product depends on (ticket) prices, quality (which we might measure by the aggregate of talent fielded by each team during a game) and crucially, the distribution of talent between the home team and the visitors. The more talent fielded by the home team the more likely is the victory of the home team, and the utility of consumers (fans) is typically increasing in home team success. However, it is widely believed that diminishing returns will set in when the success of the home team increases beyond a certain point. Too much success means too much certainty in relation to the outcome of a contest (the match or the championship), and without uncertainty it will be relatively uninteresting (consider, for example, the excitement generated by watching a re-run of last year's World Series). This point raises a number of interesting theoretical questions. First, what will be the Nash equilibrium investment in the talent by each team in a league? Second, what is the socially optimal distribution of talent in a league (and will this coincide with the joint profit maximizing distribution)? Third, how is the competitive equilibrium affected by restraints on competition in the labor market (e.g. the reserve clause) or the product market (e.g. gate revenue sharing)?

This paper suggests that in theory, with only mild assumptions, the distribution of wins that would be chosen by a league planner will be *less* balanced than the noncooperative Nash equilibrium. The reason is simple. Competition imposes an externality- each team in a noncooperative game wants to choose more wins than is collectively desirable. However, at the Nash equilibrium, the externality imposed by teams with a small marginal capacity to attract fans is greater than the externality imposed by ones with a

---

[2] There has also been much empirical research in sports, motivated by issues of team work and incentives from labor economics, and public policy analysis related to stadium issues.

large marginal capacity to attract fans, because the un-internalized loss is greater at large potential teams than at small potential teams. Thus, contrary to received wisdom (and a good many economics papers), a planner interested in maximizing aggregate attendance would choose to skew results even more unevenly than at the Nash equilibrium.

The paper provides empirical evidence to support this theory. A number of papers in the past have estimated the relationship between competitive balance and attendance, but none have jointly taken account of the fact that each team wants more wins while in aggregate wins must satisfy an adding up constraint. This paper reports some estimates of the relationship between winning and match attendance, in order to simulate the effect of changing competitive balance in Major League Baseball (MLB). Consistent with economic theory, the evidence shows that increasing competitive balance would be more likely to *reduce* than to increase attendance. The paper also simulates the planner's optimum for MLB over the last three years and shows that the attendance maximizing distribution would have been much *less* balanced than the actual distribution. Also, if the theory is correct, the planner would make teams which are dominant at the noncooperative equilibrium even more dominant. The results show that the estimated planner's distribution is indeed positively correlated with the current distribution as well as being less balanced, implying that already dominant teams would tend to dominate more. However, in some cases, where successful teams are close to capacity or have attendance that is relatively insensitive to success, the planner's choice would be for these teams to win less often.

These findings have important policy implications. Sports leagues, in particular MLB, have demanded and won significant derogations from the application of antitrust law based on the claim that collusive devices are required to *increase* competitive balance. The list of restraints employed in North American major leagues, which would be deemed unacceptable in any other industry, is a long one. It includes not only the reserve clause and gate revenue sharing but also roster limits, salary caps, luxury taxes, restrictions on cash trades, collective selling of broadcasting and merchandising rights.

Without the competitive balance justification, most of these restraints are simply mechanisms to raise profits at the expense of fans and/or players.

The paper is set out as follows. The next section reviews briefly the history of the competitive balance argument as presented by the leagues and in the courts. Section 3 reviews the theory, and section 4 reviews the previous empirical work. Section 5 reports the data and results of the regressions, section 6 presents the estimates of attendance in a perfectly balanced league and section 7 calculates the distribution of win percentages that would maximize attendance. Section 8 concludes.

2. The history of the competitive balance problem

The competitive balance issue has been debated in team sports almost from the inception of sports leagues. The first and oldest surviving team sports league in the world is baseball's National League, founded in 1876. In 1879 the league created the Reserve Rule, which prevented players from moving to other teams in the league without consent. While clearly reducing the bargaining power of players, the owners were keen to justify it in terms of the benefits to the health of the game as a whole. In 1889 the League issued a statement claiming that the Reserve Clause was instituted to protect the weaker teams and that the need to equalize playing strengths of teams was widely recognized.[3]

The need to maintain competitive balance was also cited in the notorious Federal Baseball case which handed organized baseball its antitrust exemption. According to the submission of the owners, "If the reserve clause did not exist, the highly skillful players would be absorbed by the more wealthy clubs, and thus some clubs in the league would

---

[3] "As a check on competition, the weaker clubs demanded the privilege of reserving five players…", The statement was issued in connection with a dispute over the control player contracts with the rival Players' League created in the same year. The statement went on to say "the necessity for such power of preserving the circuit of a league, by approximately equalizing its playing strength, is recognized by the League". Spalding (1911). The early development of what might be called the "competitive balance defense" is discussed in Eckard (2001). He points out that the competitive balance defense started to be used only in the late 1880s by which time it had become apparent that Congress would pass an antitrust law; prior to that the reserve clause was justified simply in terms of reducing player wages.

so far outstrip others in playing ability that the contests between the superior and inferior clubs would be uninteresting, and the public would refuse to patronize them." (Court of Appeals, 1920).

In 1951, the congressional hearings on baseball organized by Congressman Emmanuel Celler concluded "Baseball's history allows that chaotic conditions prevailed when there was no reserve clause. Experience points to no feasible substitute to protect the integrity of the game or to guarantee a comparatively even competitive struggle. The evidence adduced at the hearings would clearly not justify the enactment of legislation flatly condemning the reserve clause."[4]

In 1999, the Commissioner for baseball established a Blue Ribbon Panel to consider "whether revenue disparities among clubs are seriously damaging competitive balance". In 2000 the panel (Levin et al.) reported in its conclusions that "large and growing disparities exist and are causing problems of chronic competitive imbalance" (p1). They also stated that "the goal of a well designed league is to produce adequate competitive balance…MLB must improve its competitive balance if it is to remain competitive with other sports attractions" (p5). In other words, decreasing (increasing) competitive balance causes decreasing (increasing) demand. In this paper we provide some empirical evidence on this proposition both at the level of the club and at the level of the league.[5]

3. Theory

The seminal paper in the sports economics literature is Rottenberg (1956). Rottenberg's paper anticipated the publication of the Coase's famous theorem (1960) by explaining that regardless of the ownership rights over playing talent (reserve clause or free agency) talent would migrate to where it is most valued, and therefore that competitive balance

---

[4] This and the previous reference cited in Szymanski and Zimbalist, 2005, chapter 7.

[5] The MLB constitution was also recently amended to include the maintenance of an "appropriate" level of competitive balance in the definition of the best interests baseball of baseball which the commissioner is empowered to defend.

would be unaffected by ownership rights, a result which became known as the *invariance principle* in the sports literature. The advent of free agency for six year veterans in 1976 created the opportunity to test this proposition. Szymanski (2003) reviewed 20 papers that have tried to determine whether competitive balance had been affected by the introduction of free agency: 7 found no change, 9 found that competitive balance had improved and only 4 that it had deteriorated. These results therefore provided little support either for the position of the owners or for the invariance principle. On the face of it, the reserve clause created *less*, not more, competitive balance. However, given that over the period under investigation there were so many other changes in the structure of MLB, the power of these tests is open to question.

Rottenberg also argued that "in baseball no team can be successful unless its competitors also survive and prosper sufficiently so that the differences in the quality of play among teams are not 'too great.'" This led him to conclude that "if they [the wealthy teams] behave like rational maximizers playing talent will be more or less equally distributed among teams". This proposition was then examined in a theoretical context in a series of papers by Mohamed El-Hodiri and James Quirk. El-Hodiri and Quirk (1971) showed that equal playing strength was consistent with profit maximization only if the marginal revenue of a unit of talent was equal across all teams (proposition 2, p1310). This could be true if all teams had identical revenue functions (proposition 3, p1312). However, a rule banning cash sales will lead to equal playing strength of teams over time (proposition 4, p1314).[6] Quirk and El-Hodiri (1974) further developed their model to claim that in a league "with franchise owners motivated solely by profit…the distribution of playing strengths is independent of the gate-sharing arrangements" and "is the same as it would be if the league were operated as a syndicate with central control over the allocation of players among teams" (pp36-7). This is a striking claim, since it implies that in the context of a sports league noncooperative profit maximization by competing firms generates the same distribution of resources as joint profit maximization.

---

[6] This result emerged on the assumption that over time any team will be equally capable of drafting new talent as any other team.

This position was restated by Fort and Quirk (1995) and Vrooman (1995). Atkinson et al. (1988) made the rather more modest claim that "the noncooperative solution under equal revenue sharing maximizes league revenues by yielding the optimal distribution of talent" (p.32). In that paper the authors set out to test whether marginal revenue products in the NFL are equal to wages, which they claim is a test of the profit maximization hypothesis. In passing they argue that the rules of the league have promoted competitive balance and that this is the optimal solution, but they do not present any empirical evidence in support of this claim.

Szymanski and Kesenne (2004) prove that, contrary to the claims of Quirk and El-Hodiri, increased gate revenue sharing will in fact lead a *reduction* in competitive balance. The reason for this is the same as the rationale advanced in this paper. As gate sharing increases teams get closer to adopting the planner's solution, and the planner's solution entails less competitive balance than the noncooperative Nash equilibrium. In a noncooperative model, teams impose externalities on each other through their choices. Both teams overinvest, but relatively speaking it is the weaker drawing teams that overinvest most[7]; revenue sharing eliminates the externality, and this therefore produces a more uneven talent distribution. Full revenue sharing is equivalent to joint profit maximization, so it follows that a league planner will choose a *less* balanced distribution of results than at the noncooperative equilibrium.

This point is easily illustrated using a simple model. The structure of the model is close to that of the earlier sports literature, e.g. Atkinson et al (1988), and Fort and Quirk (1995). Suppose that each team generate attendance according to the number of wins, and that this is a concave function $Q_i(w_i)$ with $Q_i' > 0$ and $Q_i'' < 0$; beyond some critical value it is possible that $Q_i' < 0$. The most important feature of a sports league model is the technology that determines wins. It is standard in the literature that there is a fixed pool of talent.[8] The share of wins (winning percentage) is supposed to be a function of the

---

[7] In passing it is worth nothing that in practice, weaker teams tend to be the ones that find themselves in financial difficulties

[8] in reality there is some elasticity, but at least in the short term it may be reasonably considered fixed for a league such as MLB

share of total talent. Each team can win between 0 and 100% of its games, and aggregate winning percentage for the league is n/2, where n is the number of teams.[9] Here we use a simple logit common in most of the literature, so that

$$(1) \qquad w_i = \frac{n}{2}\left(\frac{t_i}{\sum_{j=1}^{n} t_j} \right) \quad \textit{subject to } 0 \le w_i \le 1$$

Note that since talent shares must sum to unity, this formulation guarantees the adding up constraint for wins is met, i.e.

$$\sum_{i=1}^{n} w_i = \frac{n}{2}$$

Suppose that profits depend on attendance, so that we can write the profit for each team as

$$(2) \qquad \pi_i = p_i Q_i(w_i) - ct_i$$

where p is the price of a ticket and c is the marginal cost per unit of talent. We compare two cases, one where each team maximizes profit independently, and the other where a league planner maximizes joint profits. We assume that both the competitive and planner's equilibrium involves an interior solution. For simplicity we also assume that

---

[9] In this paper win percentage for a team refers to its performance over the entire season against all teams, rather than the sum of bilateral win percentages from which the season's win percentage must be derived. i.e. $w_i = \frac{1}{M_i}\sum_{j \neq i} m_{ij} w_{ij}$, where M_i is the total number of games played by a team in the season, m_{ij} is the number of games played between teams i and j in a season and w_{ij} is the percentage of these games won by team i. As long as every team plays the same number of games then the sum of win percentages must add to n/2. There is an interesting scheduling problem when, as in MLB, each team is not required to play every other team or to play an equal number of games against every other team. The scheduler can arrange any number of match-ups between any pair of teams except for 3, whose schedule must be determined in order to meet the constraint that each team plays the same number of games.

ticket prices are identical.[10] Given the technology of winning, the noncooperative Nash equilibrium is characterized by the set of first order conditions for the choice of talent[11]

$$(3) \qquad \frac{Q_i'}{Q_j'} = \frac{\dfrac{\partial w_j}{\partial t_j}}{\dfrac{\partial w_i}{\partial t_i}} = \frac{\sum\limits_{k \neq j} t_k}{\sum\limits_{k \neq i} t_k} \text{ for all i and j.}$$

It is easy to show that the second order conditions are satisfied given our assumptions.[12] In line with model of El-Hodiri and Quirk, a perfectly balanced league will only arise in the case where (marginal) revenue functions are identical. With asymmetric marginal revenue functions, $\Sigma_{k\#j}t_k > \Sigma_{k\#i}t_k$ implies *both* that $t_i > t_j$ *and* that $Q'_i > Q'_j$. Therefore even at equilibrium the marginal revenue of a win for a dominant team is greater than the marginal revenue of a win for a weak team. For the league planner, however, the equilibrium condition is simply

$$(4) \qquad \frac{Q_i'}{Q_j'} = 1 \text{ for all i and j}$$

Intuitively, the planner does not need to account for the relative cost of talent/wins when allocating talent, therefore only the marginal revenue of winning matters. To meet the planner's objective (4) requires that the marginal revenue of team i to fall relative to the equilibrium described in (3), and the marginal revenue of team j to rise. Given concavity, this can only happen if wins increase for team i and fall for team j, implying increased dominance of team i at the planner's equilibrium.

---

[10] In general terms the attendance decision depends on the total cost of attending a match, rather than the simple ticket price. See e.g. Forrest et al. (2002)

[11] Strictly speaking, if the supply of talent is fixed, teams should not choose talent but a budget which then determines the share of talent of each team. However, if we assume that the allocation of talent as a function of budgets is same as the allocation of wins as a function of talent, then the budget and talent choice problems are identical.

[12] An additional condition, ignored here, is that marginal revenues must equal the marginal cost of talent. When supply is fixed the price of talent must be bid up to satisfy this condition. If marginal cost is not bid up to meet the equilibrium condition there will be pressure either to expand the league or for a new league to enter the market.

In the sports literature there has been some confusion over the derivation of the first order condition (3). It has generally been assumed that because supply is fixed, it is necessary to adopt a "fixed supply conjecture".[13] This generated the strange conclusion that the noncooperative allocation of talent would always be the same as the planner's allocation (see e.g. Fort and Quirk (1995) and Vrooman (1995)). It turns out that the fixed supply conjecture also eliminates the usual competition externality. In an asymmetric model, the externality imposed by one competitor on another is of unequal size, and therefore creates a distortion relative to the planner's ideal. The only situation where this externality will be eliminated with be in the case of equal revenue sharing (as Atkinson et al. stated) since in this case each team's objective is just 1/n of the planner's objective.

It is worth briefly commenting on the assumption of an interior solution. The model is not inconsistent with the existence of corner solutions and these do not pose a problem.[14] For every team at a corner solution in the noncooperative case, the planner would like to increase (if w = 1) or decrease (if w = 0) their share of wins but cannot, and therefore there is no change in their winning percentage. For every team at an interior solution in the noncooperative game, the planner will both want to and be able to change its winning percentage. For such teams the difference in win percentage between any two pairs will increase.[15] In model here only the optimal allocation of win percentage, n/2 is considered

---

[13] $\frac{\partial w_i}{\partial t_i} = \frac{\sum_{i=1}^{n} t_i - t_i\left(1+\sum_{j\neq i}\frac{dt_j}{dt_i}\right)}{\left(\sum_{i=1}^{n} t_i\right)^2}$. Under the usual Nash conjecture ($dt_j/dt_i = 0$) we obtain (3). However, In the literature the conjecture is usually assumed to be $dt_j/dt_i = -1$, on the argument that a one unit increase in talent for team i implies a unit fall for all other teams. However, as explained in Szymanski and Kesenne (2004) and Szymanski (2004), it is a mistake to confuse a situation of a fixed supply with a fixed supply *conjecture*. Although the game is zero-sum in talent, this does not mean that it is wrong to search for a Nash equilibrium where the choice of each team is a best response to the fixed choice of the other teams. By contrast, a fixed supply conjecture implies that teams specifically choose wins rather than talent, and therefore, because of the adding up constraint, at least one team has no choice about their inputs.

[14] The fact that teams in leagues seldom achieve 0% or 100% wins may be taken as weak support for the competitive balance hypothesis.

[15] This means that there could exist a metric for competitive balance according to which the observed variation of wins would fall under the planner's regime (for example, in a three team league where one team has w = 1 at the noncooperative equilibrium, while the other two are ranked second and third, the

(a short term decision); over the longer run the planner would presumably optimize over n, the number of teams in the league.

In practice, leagues such as MLB have created a number of mechanisms which, it is claimed, will improve competitive balance, such as roster limits, the draft, and more recently the luxury tax. It is not always obvious that these rules will in fact achieve their stated aims. For example, Zimbalist (2004), pp99-104, shows that the new revenue sharing agreement negotiated by MLB in 2002 would create higher marginal tax rates on player spending at weaker teams than at the already strong teams, a seemingly perverse decision for a league committed to creating a more balanced competition.

In terms of the model, if a set of rules exist which improves competitive balance, it remains the case that the league planner would prefer a less balanced distribution. If, however, the league rules already worsen competitive balance, it might be that the actual distribution met or even exceeded the planner's target. In our empirical section we will examine whether the actual distribution of wins under the current set of restraints is more or less balanced than the distribution which would maximize attendance.

From a theoretical perspective, therefore, we reach the rather striking proposition that the distribution of wins in a noncooperative environment will be more equal than under the planner's solution. This might suggest that concerns such as those suggested by the Blue Ribbon Panel are overstated. In the next section we examine what the empirical literature has had to say about the relation between competitive balance and demand.

4. The empirical literature

Both the extent to which competitive balance has changed over time and the effect that this has had on attendance have been the subject of some dispute in the literature. For a

---

team in the middle will clearly get closer to the top team while getting further away from the bottom team). However, this seems no more than a theoretical possibility, given that in practice perfect winning or losing survey of the literature across a range of sports, see Szymanski (2003) and Borland and McDonald (2003).

A measure of competitive balance is a measure of inequality, and there is no uniquely ideal measure for such a concept. Even in a two team league there is no unambiguous measure of balance.[16] With three or more teams a weighting must be given to the significance of different rivalries in order to construct any measure of balance. Consider, for example, a three-team league involving the New York Yankees and the Boston Red Sox, two historic rivals, and the Tampa Bay Devil Rays, an expansion franchise. Would a season in which Boston and New York each had winning records of 53% and Tampa 44% be more balanced than one in which Tampa's record was 55%, Boston's 50% and New York's 45%? The closer pennant race of the first distribution would certainly make it seem more balanced in the eyes of a majority of fans (especially given the levels of support for these teams). But in fact the standard deviations are almost identical (the second being slightly smaller).

The within-season standard deviation of team win percentages is perhaps the most popular measure of competitive balance in the literature. This is a convenient summary statistic for the degree of balance of league taken as a whole, and facilitates between season comparisons. Where the number of games played in the season varies, this can be normalized against the "idealized" standard deviation, which is expected standard deviation over M games when each team's expected winning percentage is 0.5 (see e.g. Fort and Quirk (1995)). A variety of other metrics have been used- e.g. the Gini coefficient (Schmidt and Berri (2001)), The Hirschman-Herfindahl index (Depken (1999)), entropy (Horowitz (1997), the average number of games behind the winner (Knowles et al (1992)). No doubt there are many others that could be constructed.

However, most studies concerning competitive balance and attendance have focused not on the league championship as a whole but at individual matches. In these studies

---

records are hardly ever observed.

attendance at a match is modeled as a function of a measure of the game's expected uncertainty of outcome. Studies have shown that MLB match attendance is related to pre-match betting odds (e.g. Knowles et al, (1992)), the relative standings of the teams (Butler (2002)), and the importance of the match for play-off participation (Rascher (1999)). What they show, in common with literature in other sports,[17] is that match attendance is maximized when the home team's probability of winning is substantially greater than the visitor's, usually in the range of 0.6 to 0.7, a very unbalanced contest by any standards. This finding should not be surprising given that it is the fans of the home team that fill the stadium. These findings have done little to shed light on the optimal distribution of wins across teams. While in principle it might be possible to construct such an estimate from the data, it has not been done.

Relatively few studies have looked at the relationship between competitive balance across a championship and attendance. Two which stand out are Schmidt and Berri (2001) and Humphreys (2002). In the first study the authors relate attendance over the past 100 years in the National League and American League to a Gini coefficient based on wins. They find over this period that demand increases significantly with increasing balance. However, when they estimate a panel regression over shorter, more recent, period, in order to include other economic variables of interest such as price, they find that the Gini coefficient has the opposite effect- reducing competitive balance increases attendance.[18] This result flagrantly contradicts the conventional wisdom.

Schmidt and Berri restrict the coefficient on win percentage to be common across all teams, a restriction which seems unlikely to be supported by the data. Humphreys (2002), proposes a measure of competitive balance which combines both the within-season standard deviation of win percentages and the between-season variation, capturing the idea that dominance between seasons is also likely to affect attendance. He also examines aggregate attendance over a 100-year period, pooling National League and

---

[16] Some measure of difference in win percentages seems natural, but should it be the absolute difference or the ratio?

[17] See Forrest and Simmons (2002) for a discussion of the substantial literature in the soccer world

[18] This is no longer true when they average over a three to five year period.

American League data, and finds that attendance is significantly affected by his measure, even when conventional measures of within competitive balance are insignificant. Both these studies suggest that competitive balance affects attendance in aggregate, but they give little indication as to how the distribution of wins between individual teams will affect attendance. Indeed, these studies imply that winning impacts each team in the same way.

5. Data and estimation

The panel data approach adopted here has much in common with the specification of Schmidt and Berri (2001). Data on win percentage and annual aggregate attendance for individual teams is easily available.[19] The estimation period selected was between 1978 and 2003. During this period both the American League (AL) and National League (NL) were relatively stable- the number of teams in the American League did not change (but in 1998 the Tampa Bay Devil Rays replaced the Milwaukee Brewers) while the National League expanded by two teams in 1994 (the Florida Marlins and the Colorado Rockies) and two teams in 1998 (The Arizona Diamondbacks and the Milwaukee Brewers). The period covered is essentially the post free agency era.

Win percentage in aggregate must always sum to n/2, hence it has remained constant at 7 in the AL and expanded from 6 to 8 in the NL. Individual team win percentages show no obvious trends over time. While teams may have periods of relative success or relative failure, there are no teams that had a sub .500 or above .500 record in every season. Attendance data, by contrast, showed a consistent upward trend over the entire period. Schmidt and Berri (2001) address the issue of cointegration and stationarity and conclude that attendance follows an I(1) process. It seems natural therefore to estimate a regression which relates innovations in attendance to innovations in win percentage. Thus the general model proposed here is

---

[19] The data used in this study was downloaded from www.baseball1.com.

$$(5) \qquad \Delta\,(\text{Attendance})_{it} = \alpha + \alpha_i + \alpha_t + \beta_i\,\Delta\,(\text{winpercentage})_{it} + \gamma X_{it} + \varepsilon$$

where X is a matrix of variables that also affect attendance. The panel structure of the data makes it possible to estimate a sensitivity of attendance to wins that is team specific, which can then be used to simulate the choices of a league planner. Schmidt and Berri (2001) and Humphreys (2002) use dummy variables to capture the impact of specific events such as the labor disputes in 1981, 1994 and 1995.[20] Here the period effects capture not only these events but any other year specific factors that contributed to the change in attendance. Likewise, the team specific fixed effects capture any distinctive features of individual teams that varied over the sample period, such as management and ownership effects or city size. The other factors included in the regression are dummies for the effects of new ballpark and success in the previous season. There is growing evidence in the literature of honeymoon effects (see e.g. Poitras and Hadley (2003)), and dummies were included for the last two years of an old park and each of the first four years of a new one. The effect of the previous season's performance on the propensity to attend is likely to vary significantly between individuals- some fans are diehards, some start to attend when they sense an upturn, while others require concrete evidence of success before investing financially and emotionally. Here dummies were included for clear indicators of success in the previous season: winning the World Series, winning the Pennant, winning the Division or winning a wild card spot.

Finally, changes in ticket prices were included for some versions of the model. Ticket prices are problematic, both Humphreys (2002) and Schmidt and Berri (2001) mention a number of the problems. First, given that seat prices vary, there is no obvious summary statistic. Second, the data that is available covers two different time periods (up to 1988 and after 1991) with the intervening years being absent. These series are quite different, and rather than attempt to merge them the regressions including price were run on only the more recent data.[21] Finally, since teams are generally local monopolists and fix prices before the start of the season, it seems plausible that observed price is the supply price,

---

[20] See also Schmidt and Berri (2004) for an analysis of the impact of labor disputes on baseball demand.
[21] The ticket price data can be found on Rod Fort's excellent website, http://users.pullman.com/rodfort/

that this is a function of expected demand, which is itself predicted on the basis of historic demand. Clearly, if price is endogenous, then including ticket prices risks incorporating a function of the LHS variable on the RHS. It turns out that the price variable is positive and significant when included in the regression, but does not materially affect the estimated coefficients. Given its potential endogeneity, and the absence of plausible instruments, the estimates that exclude the price variable are preferred.

Several alternative functional forms were tested. Squared terms were used but these were almost always insignificant both individually and jointly. Moreover, in several cases the linear term was negative and the squared term positive, implying increasing returns. Not only is there reason to doubt this would be true on theoretical grounds, but it is also the case that almost no teams had winning percentages in the sample period outside the 0.333 to 0.666, which suggests caution when extrapolating the effect of very high win percentages. Using the linear estimates to construct a quadratic approximation will be considered in the next section. Estimates including the lagged value of win percentage were also considered, but were only significant in two cases in the American League and four cases for the National League. While previous work has found lagged win percentage significant, in many cases this may have been due to estimating the relationship in levels rather than differences.

Table 1 shows the results of the two regressions, one for the American League and the other for the National League. Overall the regressions seem well specified; for a difference model the explanatory power of the right hand side variables is quite high. The t-statistics are adjusted for heteroscedasticity and asympototic normality cannot be rejected (the critical value for the Jarque Bera test statistic at the 5% level is 5.99).

The upper segment of the Table shows the sensitivity of attendance to winning percentage for each team. The Minnesota Twins have the highest estimated sensitivity in the American League, generating an additional 47,233 spectators for each additional win

---

percent, while the Anaheim Angels have the lowest, with only 14,653 extra fans for every additional win percent. These estimates are significant at the 5% level for all but three teams, the Boston Red Sox, where the p-value is very close to 5%, the Texas Rangers, where the estimate is significant as the 10% level, and Tampa Bay, an expansion team for whom there are only 4 observations. In the National League the highest estimated sensitivity is for the Atlanta Braves, with 56,670 additional spectators for every addition win percent. At the bottom end, two expansion teams, The Colorado Rockies and the Milwaukee Brewers (who moved from the AL), have negative point estimates that are insignificantly different from zero. The lowest significant estimate is for the Chicago Cubs, with 15,826 additional spectators for each additional win percent. The estimate for the Arizona Diamondbacks is also insignificant. Thus, of the 5 the franchises that entered or changed leagues, all but one have estimates that are insignificant. Of the remaining 26 cases, all but two are significant at conventional levels, and those two are borderline cases. These results can be compared with Schmidt and Berri (2001), who report a pooled estimate of sensitivity for the American League of 21,000 per win percent and 25,000 for the National League, both in 1988, and Scully (1989) who estimates a sensitivity of 21,500 across all MLB teams. The average sensitivity in this sample is 25,965 for the AL and 26,081 for the AL, suggesting that these estimates are broadly consistent with previous work.

The lower segment of Table 1 reports the estimated impact of new ballparks and success in the previous season. As far as the new ballparks are concerned, the estimates suggests a big step change associated with the first year of a new park (464,500 for the AL and 590,770 for the NL). The AL estimates show a significant positive effect associated with the last year of the old stadium (293,170), giving a total effect of 757,670. In the NL the anticipatory effect is insignificant at conventional levels, and there is also a negative impact in the second year after the opening of the park, which is close to conventional significance. If, for the sake of argument, these two borderline impacts are added to the first year effect, the aggregate impact in the NL is a shift of 532,660, somewhat smaller than the impact in the AL.

The impact of success in previous seasons is quite imprecisely estimated. Winning a wild card, a divisional title, a pennant, and the World Series all add to next season's attendance, but all estimates are insignificant. Perhaps oddly, the biggest effect seems to come from winning a wild card, especially in the American League. However, these estimates do not suggest that a team should place any great hope in expanding ticket sales on the back of last year's performance.

The team fixed effects and the year effects are not reported in full here. All of the team specific effects were insignificant indicating the absence of any team specific trend in attendance. The year effects are almost all insignificant apart from the years affected by the labor disputes. Thus 1981 registered a fall of 541,405 per team in the AL and 705,460 in the NL, while following year produced an increase of 619,598 per team in the AL and 681,860 in the NL. Thus the long term impact of the strike was insignificant. The strike of 1994/95 produced a fall of 757,721 in the AL and 703,230. However, the bounce of 1996 was not as big as in 1982, replacing only about one quarter to one half of the loss (250,403 in the AL and 331,690 in the NL). Only two other years had significant effects, both in the NL, and these were 1993 (+384,860) and 2002 (-235,960).

Table 1: Sensitivity of attendance to win percentage, stadium effects and previous year's success

American League			**National League**		
(a) The effect of a 1% increase in win percentage					
	coefficient	t stat		coefficient	t stat
Minnesota Twins	47233	5,171	Atlanta Braves	56670	5,654
Cleveland Indians	32370	4,171	Florida Marlins	40753	3,463
New York Yankees	31494	3,106	New York Mets	35628	3,816
Oakland Athletics	30744	4,170	San Diego Padres	34940	4,489
Baltimore Orioles	29290	3,472	Pittsburgh Pirates	34204	3,821
Detroit Tigers	28585	3,727	Los Angeles Dodgers	33750	4,162
Milwaukee Brewers	27914	2,370	Philadelphia Phillies	32657	3,521
Kansas City Royals	27378	2,990	San Francisco Giants	30462	3,900
Chicago White Sox	23037	2,579	Montreal Expos	28044	3,201
Seattle Mariners	22733	2,767	Houston Astros	27608	3,108
Toronto Blue Jays	21708	2,140	St. Louis Cardinals	26285	3,452
Boston Red Sox	19662	1,908	Arizona Diamondbacks	26168	1,121
Tampa Bay Devil Rays	17972	0,456	Cincinnati Reds	18458	2,692
Texas Rangers	14702	1,741	Chicago Cubs	15826	2,296
Anaheim Angels	14653	2,024	Milwaukee Brewers	-1728,1	-0,063
			Colorado Rockies	-22427	-0,938
(b) Change in attendance associated with					
2 years before new ballpark	87348	0,841		-147340	-1,123
Last year before new ballpark	293170	2,818		196610	1,514
1st year of new ballpark	464500	4,502		590770	4,790
2nd year of new ballpark	26759	0,260		-254720	-1,912
3rd year of new ballpark	61369	0,590		77405	0,646
4th year of new ballpark	146520	1,378		-84710	-0,638
Divisional champion t-1	101020	1,815		45169	0,836
Wild card t-1	238500	2,106		136560	1,146
Pennant win t-1	127270	1,177		151160	1,578
World Series win t-1	78879	0,609		144390	1,064
Constant	-34129	-1,792		-19046	-0,995
(c) diagnostics					
R^2	0.508			0.601	
Jarque Bera $\chi^2(2)$	0.125			0.182	

NB t-statistics are corrected for heteroscedasticity. The LHS variable is the change in attendance for team i in year t.

6. Simulating the impact on attendance of a perfectly balanced league

From the estimates of the previous section it is straightforward to calculate what the total attendance would have been in any given season on the assumption that each team had .500 winning record, holding all else equal. Given the size of the coefficients reported in Table 1, these changes have substantial effects on the attendance at individual teams. For example, in the AL, the 2003 New York Yankees enjoyed a dominant season with a winning record of .620 and had attendance of 3.41 million. With .500 record they would only have drawn only 3.04 million. By contrast the Detroit Tigers almost broke the record for the largest number of losses in a season and ended up with a .265 record, drawing 1.37 million. Had they been blessed with a .500 record, they would have drawn 2.04 million. Similarly, in the NL, the Atlanta Braves had the best winning record at .623 and drew 2.40 million, but would only have drawn 1.82 million with a .500 record, while the San Diego Padres attracted 2.03 million fans with a winning record of .395, when they would have attracted 2.49 million with a .500 record. Clearly, the data shows that team attendance figures can be highly sensitive to changes in win percentage.

However, when all these changes are added up, in aggregate there is almost no difference between actual annual attendance and attendance at the hypothetical perfectly balanced league. Table 2 shows (i) the actual attendance in each season, (ii) the fitted value, based on the regression coefficients and the actual levels of win percent, and (iii) the fitted values assuming each has .500 record. In the AL the effect of introducing perfect balance would have been negative for 15 seasons and positive for 10, but in no case was the effect greater than 1.27%. In the NL the effect would have been negative in 12 seasons and positive in 13, with the largest effect being a reduction of 3% in the 2002 season. Summing across all seasons, the effect of perfect balance would have been astoundingly small.

One objection to this finding is that teams successful teams may be operating close to capacity, while unsuccessful ones have plenty of spare capacity. One of the main reasons for selecting baseball rather than any other major league sport for this study is that ballparks seldom operate close to capacity over an entire season, unlike the other major leagues. However, some teams sell-out for the more popular games, creating some capacity constraints.

Successful teams operating at capacity might not lose many fans if their record came closer to 0.500, while the weaker teams would still gain. Thus redistributing win percentage in favor of weaker teams might generate a larger net increase in attendance than these estimates suggest. Re-estimating attendance under the assumption that teams with win percentage above .500 did not lose any fans if their ballpark was operating at more than 80% of stated capacity, did not produce a large change in the estimates. Obviously, under this assumption perfect balance would lead to somewhat greater attendance than in the absence of the constraint, but in almost all years the effect was less than 1% and was still often negative (since not all ballparks were constrained under this measure). Over the 25 years, the aggregate effect was +0.87% in the AL and +0.32% for the NL- which are still negligible effects.

Advocates of competitive balance might argue that more balance would have created a more dynamic effect on attendance and this will be discussed in the conclusions, but it is worth noting at this stage that the sensitivities reported in Table 1 do not suggest that, as a general, rule, perpetually weak teams generate larger increases in attendance, when they become successful, than traditionally successful teams. Thus, for example, while Detroit have been relatively poor over the data period, their sensitivity to win percentage is lower than that of the relatively strong Yankees. Of course, given that there are so many teams in the data set no doubt there are comparisons which could be made to support almost any argument. A more reasonable way to proceed is to find a way to identify a distribution of win percentages that would maximize attendance, and to consider whether this would be more or less balanced than the existing distribution.

Table 2: Annual Attendance- actual, fitted, and under perfect competitive balance (millions)

American League				**National League**		
	attendance	*fitted*	*perfect balance*	*attendance*	*fitted*	*perfect balance*
1979	22.372	22.389	22.328	21.178	21.293	21.726
1980	21.890	21.907	21.629	21.124	21.238	21.198
1981	14.066	14.083	14.194	12.478	12.593	12.889
1982	23.080	22.633	22.921	21.507	21.622	21.319
1983	23.991	23.982	24.006	21.549	21.663	21.456
1984	23.961	23.979	23.914	20.781	20.896	20.931
1985	24.532	24.549	24.620	22.292	22.406	22.728
1986	25.173	25.190	25.469	22.333	22.448	22.535
1987	27.277	27.295	27.212	24.734	24.848	25.090
1988	28.500	28.517	28.242	24.499	24.613	24.984
1989	29.849	29.402	29.507	25.324	25.438	25.865
1990	30.332	30.323	30.503	24.492	24.606	24.817
1991	32.118	31.670	31.708	24.696	24.810	24.432
1992	31.759	31.285	31.040	24.111	24.225	24.047
1993	33.333	33.324	33.510	29.376	29.491	29.624
1994	24.202	23.290	23.067	20.589	20.703	20.427
1995	25.359	25.323	25.560	25.110	24.934	24.774
1996	29.718	29.735	29.726	30.379	30.203	30.406
1997	31.283	30.836	31.111	31.885	31.709	32.284
1998	29.668	29.637	29.517	33.005	32.829	33.000
1999	30.254	29.785	29.961	33.601	33.426	33.626
2000	31.676	31.105	30.972	39.683	39.447	40.555
2001	32.854	32.747	32.422	39.727	39.491	39.300
2002	30.918	30.839	30.682	37.026	36.789	36.564
2003	30.907	30.827	30.839	36.723	36.487	35.750
Total	689.074	684.652	684.658	668.207	668.207	670.327

7. Identifying the distribution of win percent that will maximize attendance

It was mentioned earlier that a quadratic specification for win percentage neither fitted the data well nor produced coefficients with a sensible economic interpretation. However, when thinking about the optimal distribution of win percent the linear model produces an implausible conclusion. Given that no team can have a win percent greater than 1, the optimal allocation would be a corner solution: give the n/2 teams with the highest estimated coefficients a win per cent equal to unity and the rest zero. This would of course, be a much more unbalanced league, and thus this would support the theoretical argument presented in section 3. However, to rest the argument that the planner will choose *less*, not more, competitive balance, relative to the noncooperative equilibrium, on the belief that attendance is linear in win percentage is not entirely convincing.

One reason for the failure of the quadratic regression model is that the data lies within a very restricted range. For example, in the NL, all win percentages in the sample lie between 0.333 and 0.666. In the AL there are a small number outside this range- the 2003 Detroit Tigers have already been discussed- and two teams endured seasons with 0.327 records, but all the remainder were above 0.333. Three teams in the AL during the sample also had seasons above 0.666- Seattle in 2001 (0.716) the Yankees in 1998 (0.704) and Cleveland in 1995 (0.694). Even if winning percentages of 0.900 or 0.950 might cause a relative decline in attendance, there is no data on which to test this proposition. Thus it is possible that the true underlying relationship is quadratic, but that is well approximated by a linear relationship in the range of the sample data.

Suppose that in reality the linear relationship that was almost identical to a quadratic relationship, and that we can recover the quadratic coefficients from the linear estimates. On these assumptions we can find an interior solution for the optimal degree of competitive balance from the recovered quadratic coefficients. Note that such a procedure is biased in favor of finding a more balanced optimum than the one identified by the data, and is therefore biased against this paper's thesis (that a planner would choose a less balanced outcome than that found at the noncooperative equilibrium).

The approach adopted here is to assume that the true relationship is represented by a quadratic equation, and that two points of this equation pass through the fitted level of attendance associated with a .333 and a .666 record. The third point is then related to the capacity of the ballpark. For most teams, if win percentage approached 1 in any year then the expected level of attendance from the linear regression would be at or beyond capacity in that year. To ensure sufficient curvature in the relationship it was assumed that maximum effective capacity for most teams was equal to 75% of the official stated capacity.[22] In a small number of cases, where the stadium capacity is much larger than attendance, it was necessary to assume that the club would be constrained at even lower levels of capacity utilization. For teams already operating close to capacity, a more realistic figure was assumed.

The approach is illustrated in figure 1. For each team the fitted relationship from the linear regression provides the value of A_1 and A_2. The value of A_3 is then derived as a function of the capacity of each ballpark. Given these three coordinates the coefficients for a quadratic equation can be derived (see appendix 1). This is done for each year for which the optimal distribution is to be estimated.

Note that the quadratic parameters differ for each team, reflecting the same idiosyncrasies as the linear model. Of course, the quadratic estimates will be less efficient in relation to the data than the linear estimates, but we can check this by comparing the correlation between attendance and the fitted values of the linear regression with the correlation between attendance and the fitted values of the quadratic approximation. These are in fact very close. For the NL the correlation for the linear regression is 0.931, while the quadratic it is 0.924. In the AL the correlations are .942 and .931 respectively.

---

[22] stadium capacities, past and present can be obtained from http://www.ballparksofbaseball.com/

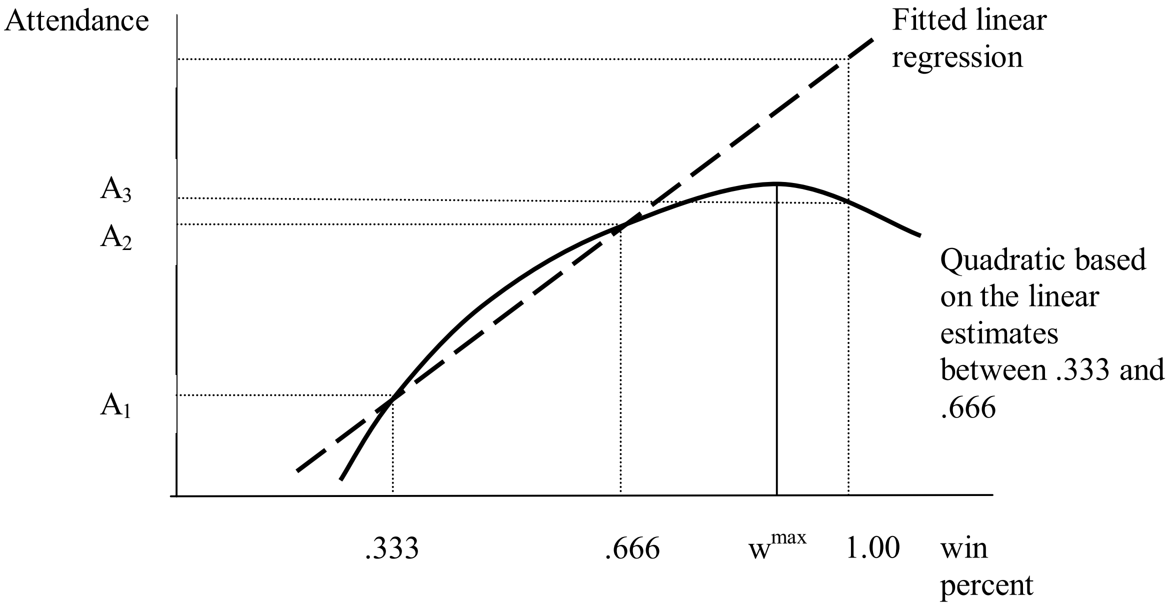

Figure 1: Deriving a quadratic relationship from the linear regression assuming capacity constraints

It is now possible to identify the attendance maximizing win percentage for each team taken on its own, as shown by w^{max} in figure 1. Figures 2 and 3 show the estimated quadratic relationship for teams in the AL and NL. The way figure 1 is drawn implies that this value is at an interior solution, but it is plain from inspection of figures 2 and 3 that in many cases the maximum lies beyond 1.

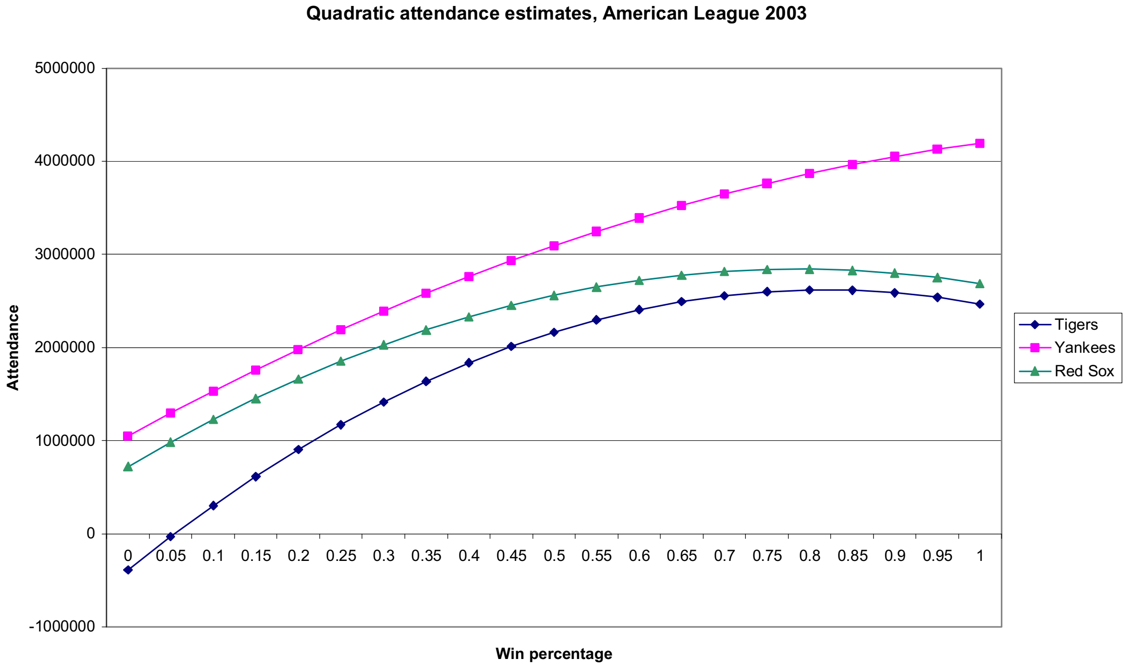

Figure 2

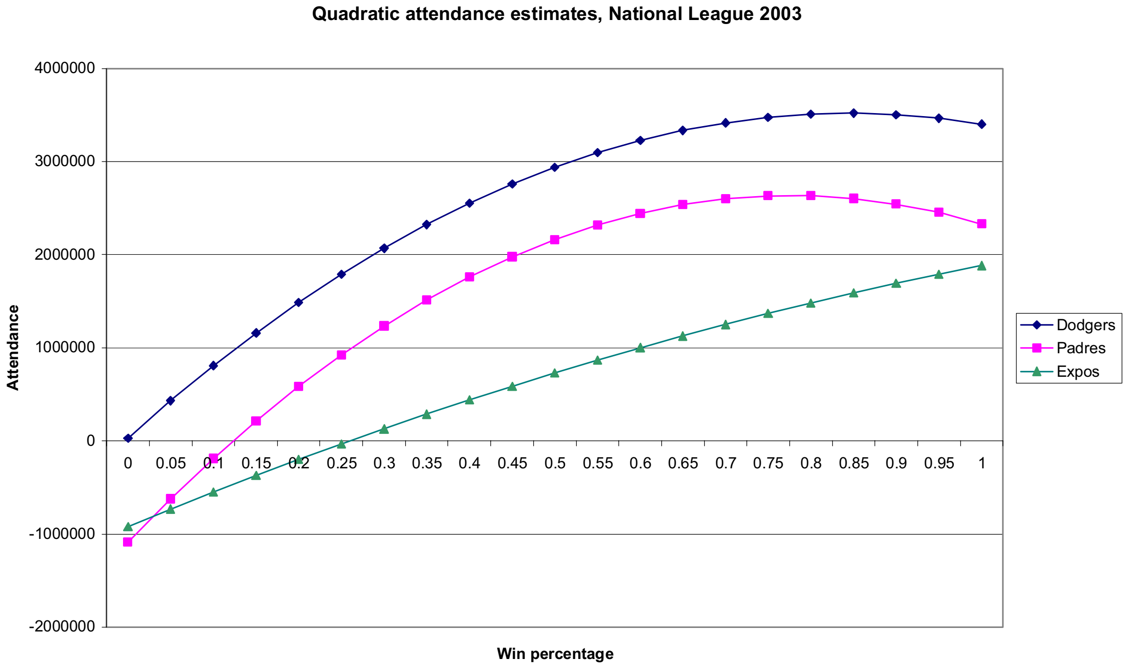

Figure 3

Clearly the win percentage that maximizes attendance for an individual team will be much higher than that which contributes to maximizing league attendance (this is the cause of the competitive externality). The next step, therefore is to calculate the distribution of wins that maximizes league attendance. This can be calculated as follows. Given that there exists for each team a quadratic attendance function

$$\text{Attendance}_{it} = a_i + b_i w_{it} + c_i w_{it}^2$$

With $b > 0$ and $c < 0$ for all i, the league planner will maximize total attendance when the marginal attendance for each team with respect to wins is equalized; in other words

(6) $b_i + 2c_i w_i = b_j + 2c_j w_j$ for all i and j.

Thus

(7) $w_i = (b_j - b_i)/2c_i + (c_j/c_i) w_j$

If we sum over all w_i not including w_j, then

(8) $\sum_{(i\neq j)} w_i = \sum_{(i\neq j)} [(b_j - b_i)/ 2c_i] + c_j w_j \sum_{(i\neq j)} (1/c_i)$

but also

(9) $\sum_{(i\neq j)} w_i = n/2 - w_j$.

(this is the adding up constraint introduced in (1)), and hence

(10) $w_j = \{n/2 - \sum_{(i\neq j)} [(b_j - b_i)/2c_i]\}/[1 + c_j \sum_{(i\neq j)}(1/c_i)]$

Given that we have estimates of the parameters a, b and c for each team in each season, we can now calculate the attendance maximizing win percentage for each team in any given season.

Table 3 shows the comparisons between the attendance maximizing win percent derived by the method above to the actual win percent for each team over the last three seasons in the dataset, 2001-2003. The fitted and actual levels of attendance are also provided. In each league teams are ranked by win percent. Not surprisingly, the ranking of teams conforms closely to the ranking of linear regression estimates in Table 1. However, in some cases, the fact that a club has substantial spare capacity may push it up the table (Philadelphia in the NL is one example), and conversely, teams close to their capacity limit sometimes have a lower ranking than from the regression. Note that in the NL all the expansion teams were excluded from the exercise. Three of these teams (Colorado, Milwaukee and Arizona had statistically insignificant results from the regression (see Table 1), and two of these were negative. This reflected the relatively small number of observations (4 each for Tampa Bay and Arizona, 9 for Colorado and Florida, after allowing for differencing).

Table 3: The attendance maximizing distribution of win percent, 2001-2003

American League					**National League**				
Name	*quadratic fitted win percent*	*actual win percent*	*quadratic fitted attendance*	*actual attendance*	*team*	*quadratic fitted win percent*	*actual win percent*	*quadratic fitted attendance*	*actual attendance*
2003									
Minnesota	0.881	0.556	2.972	1.946	Atlanta	0.691	0.623	2.834	2.401
New York	0.758	0.620	3.780	3.466	Philadelphia	0.566	0.531	1.818	2.260
Oakland	0.731	0.593	2.486	2.217	New York	0.557	0.410	3.237	2.141
Kansas	0.634	0.512	1.966	1.780	San Diego	0.551	0.395	2.820	2.030
Detroit	0.576	0.265	2.358	1.368	Los Angeles	0.550	0.525	3.096	3.139
Cleveland	0.569	0.420	3.101	1.730	Pittsburgh	0.543	0.463	2.301	1.637
Baltimore	0.560	0.436	3.237	2.455	San Francisco	0.508	0.621	3.449	3.265
Chicago	0.554	0.531	1.764	1.940	Houston	0.489	0.537	2.471	2.454
Seattle	0.516	0.574	3.489	3.269	St. Louis	0.472	0.525	2.809	2.910
Boston	0.470	0.586	2.497	2.724	Montreal	0.460	0.512	0.615	1.026
Toronto	0.273	0.531	1.082	1.799	Cincinnati	0.348	0.426	2.124	2.355
Texas	0.266	0.438	2.003	2.094	Chicago	0.266	0.543	2.379	2.963
Tampa Bay	0.142	0.389	0.526	1.059					
Anaheim	0.071	0.475	1.795	3.061					
2002									
Minnesota	0.975	0.584	3.137	1.924	Atlanta	0.650	0.627	3.227	2.603
New York	0.770	0.640	3.764	3.466	Philadelphia	0.593	0.497	1.753	1.618
Oakland	0.726	0.636	2.461	2.170	New York	0.591	0.466	2.819	2.805
Kansas	0.622	0.383	1.967	1.323	San Diego	0.577	0.407	2.551	2.221
Cleveland	0.610	0.457	3.373	2.617	Los Angeles	0.571	0.568	2.961	3.131
Detroit	0.606	0.342	2.428	1.504	Pittsburgh	0.539	0.447	2.721	1.785
Baltimore	0.529	0.414	3.556	2.682	San Francisco	0.514	0.586	3.286	3.253
Seattle	0.506	0.574	3.187	3.543	Montreal	0.513	0.512	0.658	0.812
Chicago	0.462	0.500	1.468	1.677	Houston	0.498	0.519	2.782	2.517
Boston	0.445	0.574	2.402	2.651	St. Louis	0.484	0.599	2.970	3.012
Tampa Bay	0.318	0.342	0.909	1.066	Cincinnati	0.287	0.481	1.537	1.856
Texas	0.307	0.444	2.418	2.352	Chicago	0.183	0.414	1.879	2.693
Toronto	0.293	0.481	1.232	1.638					
Anaheim	0.133	0.611	1.227	2.306					
2001									
Minnesota	0.978	0.525	3.130	1.783	Atlanta	0.622	0.543	3.553	2.824
Oakland	0.969	0.630	2.853	2.133	New York	0.561	0.506	3.169	2.658
New York	0.631	0.590	3.654	3.265	Los Angeles	0.554	0.531	3.007	3.017
Kansas	0.561	0.401	1.854	1.536	San Diego	0.553	0.488	2.771	2.378
Cleveland	0.538	0.562	3.678	3.176	Philadelphia	0.543	0.531	2.131	1.782
Detroit	0.532	0.407	2.739	1.921	Pittsburgh	0.530	0.383	2.918	2.465
Baltimore	0.524	0.389	3.770	3.095	San Francisco	0.510	0.556	3.060	3.312
Chicago	0.468	0.512	1.831	1.766	Houston	0.492	0.574	3.165	2.904
Seattle	0.456	0.716	3.082	3.507	Montreal	0.485	0.420	1.169	0.643
Boston	0.439	0.509	2.533	2.625	St. Louis	0.479	0.574	3.265	3.110
Tampa Bay	0.377	0.383	1.325	1.298	Cincinnati	0.360	0.407	2.107	1.880
Anaheim	0.264	0.463	1.684	2.001	Chicago	0.311	0.543	2.650	2.779
Texas	0.254	0.451	2.269	2.831					
Toronto	0.019	0.494	0.537	1.915					

Table 4 summarizes the empirical findings for each league. Consistent with the theory, if the planner's objective were to maximize attendance she would choose a substantially more uneven distribution of win percentage than is observed in reality. In the AL, the planner's standard deviation is more than twice as large as the actual, in the NL it is around 50% larger. Recall that the theory also suggested that the reason the planner would demand less balance is that the relatively small teams, with low win percents impose a large externality on the large teams with larger win percents. By implication, the planner should exaggerate the existing distribution. It therefore seems supportive of the theory that there is a positive and often substantial correlation between the fitted and the actual distribution of win percents.[23]

Table 4: Attendance maximizing win percentage (quadratic fitted) compared to actual.

	quadratic fitted standard deviation of win percent	*actual standard deviation of win percent*	*Correlation coefficient of actual and fitted win percents*	*quadratic fitted attendance*	*actual attendance*	*% increase*
American League						
2003	0.235	0.097	0.356	33.06	30.907	7.0
2002	0.221	0.106	0.228	33.53	30.918	8.5
2001	0.254	0.098	0.318	34.94	32.854	6.4
National League						
2003	0.109	0.0741	0.172	29.95	28.58	4.8
2002	0.134	0.0724	0.397	29.14	28.307	3.0
2001	0.087	0.0663	0.139	32.96	29.752	10.8

Table 4 also provides evidence of the impact that the planner's choice would have on attendance. These range from a 3% increase to a 10.8% increase in total attendance, a not insubstantial figure. Given an average ticket price of $18.70 in 2003, this would in

---

[23] There are some outliers. Some ballparks are close to selling out regardless of the success of the team (e.g. Boston and the Chicago Cubs. Even without capacity constraints some teams have very loyal fans (e.g. St Louis). Thus the marginal returns for the planner are small at high levels of win percent.

aggregate have implied an extra $66 million of revenue for MLB, which compares to an average gate revenue of $45 million per team in 2003.[24]

8. Discussion and Conclusions

The conventional wisdom in the field of sports economics has been that league planners would choose more competitive balance than we observe at a noncooperative equilibrium, and hence that restrictive measures proposed by leagues themselves should be accepted as in the public interest. However, this proposition flies in the face of economic theory. While the result that a planner would choose *less*, not more, competitive balance may seem counterintuitive, it is in fact just one example of a standard result in the economic literature. Consider for example, a Cournot quantity setting game in which firms have asymmetric (constant) marginal costs. At an interior equilibrium the high marginal cost firms depress industry profits not only because of the pure competition effect, but also because the industry could profitably replace any unit of output from one of these firms with a unit of output from a low marginal cost firm. A planner would shut down all the high marginal cost firms. A similar story can be found in auction theory. In a first or second price auction with private values, a low valuation bidder will impose an externality on the high valuation bidder (who wins the auction), by forcing them to increase their bid. In general, the elimination of the low valuation bidder will create a larger welfare gain (for the bidders only) than eliminating the high valuation bidder.[25]

This argument may seem counterfactual: surely leagues in practice do try to improve competitive balance and this does increase attendance/revenues? Most frequently cited in

---

[24] Average taken from data on the Rod Fort website.

[25] Imagine bidder 1 values the object at V_1 and bidder 2 at V_2, with $V_1 > V_2$. In an English auction bidder 1 wins at a price equal to V_2, so bidder 1's net gain is $V_1 - V_2$ and bidder 2's is zero. Eliminate bidder 1 and bidder 2 wins the object and pays 0, giving a net gain of $V_2 – (V_1 - V_2) = 2V_2 – V_1$. Eliminate bidder 2 and bidder 1 wins at a price of 0 with a net gain of $V_1 – (V_1 - V_2) = V_2$. But clearly the elimination of bidder 2 produces the larger net gain, since $V_2 > 2V_2 – V_1 \Rightarrow V_1 > V_2$, which is true by assumption. This also assumes, of course, that if the low valuation bidder wins he cannot transfer the object of the auction to the high valuation bidder.

this respect is the NFL, with its near perfect redistribution of resources and financial success. However, there are some important differences between the NFL and MLB; first, all the NFL teams are capacity constrained and almost every match is a sell-out, second, the NFL is more oriented toward TV viewership, and maximizing viewership may require more balance than maximizing attendance when there is spare capacity.[26] It is also interesting to note that the NFL's well-known parity has started to create complaints about the general mediocrity of the teams, which might in the longer term undermine interest in the sport.[27] To take a different example, consider for a moment the world of soccer. In most European soccer leagues only a very small number of teams ever win the title. For example, in England over the past twelve seasons, only three teams have won the Premier League- Manchester United 8 times, Arsenal 3 times and Blackburn Rovers once.[28] During this period of seemingly excessive dominance, attendance increased from 9.8 million to 13.5 million. It is simply not the case that competitive balance is either necessary or sufficient increase the popularity of a sport. No doubt there are a variety of idiosyncratic factors at play which explain the specific histories of individual leagues.[29]

One final objection may be that the discussion in this paper deals entirely with the notion of competitive balance within a season, and says nothing about the effect of competitive balance between seasons. It may be that, regardless of within season balance, excessive dominance will cause fans to be entirely lost to the game, and that therefore an intervention to ensure that there is turnover between seasons might be desirable (see e.g.

---

[26] This would be true, for example, if the distribution of team support were more balanced among TV viewers than among fans attending games.

[27] See e.g. Berger (2004). Interestingly, a google search for the terms "NFL mediocrity" produces a large number of articles by journalists on the subject.

[28] One American reader commented that this observation was inconclusive, since play-off results tell us little about regular season balance; however, there are no play-offs for the Premier League championship (or for any other major European domestic championship) and hence the winner of the championship is team which wins the highest number of points in the regular season, which is almost always the team with highest win percentage (but because ties are frequent in soccer, win percentage is not clearly defined- the convention in most soccer leagues is to award three points for a victory and one for tie).

[29] Some commentators argue that institutions in soccer such as promotion and relegation compensate for the lack of competitive balance, and that American sports, lacking these institutions, require more competitive balance to be successful. However, such arguments start from the presumption that competitive balance *does* increase aggregate attendance, a proposition which this paper shows is not true for within season competitive balance.

Ross (1997)). This seems inconsistent with a policy of maintaining the kind of within-season imbalance that will maximize annual attendance. However, the institution of play-offs, which ensures more randomness in the identity of the winner of the World Series than in the identity of the team with the highest regular season win percentage, might be the kind of institution that can reconcile some of these aims.

A regular rotation among dominant teams would also help to balance these conflicting objectives. It may be that the NFL's success could be attributed to such swings. It may also help to explain the popularity of the European Champions League, which has been dominated by large clubs but few have managed to maintain long-term dominance. However, it is also worth observing that (a) leagues seldom justify measures alleged to improve competitive balance on the grounds that they will increase between season uncertainty (justifications almost always relate to improving balance within a season) and (b) many people argue that dominance across seasons, in the form of dynasties, can be good for a league, either because of the sheer quality of dominant teams or the excitement associated with David v. Goliath contests. Establishing whether a planner would prefer more or less between-season competitive balance, and if so how that would integrate with within-season balance is both unclear and an issue worthy of further research.

Finally, it is worth re-iterating the significance of this paper for public policy in relation to sports leagues. Without a competitive balance justification, the restrictive practices operated by the major leagues lose all their public policy justification and should therefore be treated in the same way as restrictions imposed by any other cartel.[30] This paper therefore suggests that there is a need for a substantial re-evaluation of public policy in relation to sports leagues.

---

[30] In practice sports leagues in the US and Canada can use the labor exemption, which permits employers to enter restrictive agreements in exchange for concessions to the unions on wage levels, insurance provisions and so on. This defence would not work in the EU, which has no equivalent of the labor exemption.

Appendix 1

The quadratic equations to be derived are

(A1) $a + b\,w_1 + c\,w_1^2 = A_1$.

(A2) $a + b\,w_2 + c\,w_2^2 = A_2$.

(A3) $a + b\,w_3 + c\,w_3^2 = A_3$.

For any win percentages w_1, w_2, and w_3, we can derive the quadratic equation coefficients a, b and c if we know the associated attendances A_1, A_2, and A_3. A_1 and A_2 are taken from the linear regression estimates and A_3 is based upon the assumed capacity constraint. The greater the capacity constraint the greater the curvature of the quadratic equation.

It is straightforward to show that

(A4) $$c = [(w_2 - w_1)(A_3 - A_1) + (w_3 - w_1)(A_1 - A_2)] / [(w_1^2 - w_2^2)(w_3 - w_2) + (w_3^2 - w_1^2)(w_2 - w_1)]$$

Using this value for c we can write

(A5) $$b = [A_2 - A_1 + c\,(w_1^2 - w_2^2)] / (w_2 - w_1)$$

and finally, using (A5) and (A4)

(A6) $$a = A_1 + b\,w_1 + c\,w_1^2$$

References

Atkinson, Scott, Linda Stanley and John Tschirhart. 1988. “Revenue Sharing as an incentive in an agency problem: an example from the National Football League” *Rand Journal of Economics*, 19, 1, 27-43

Richard Berger, 2004. “NFL Mediocrity” *The Yale Herald*, XXXVII, 2, January.

Depken, Craig. 1999. "Free-Agency and the Competitiveness of Major League Baseball" *Review of Industrial Organization*, 14, 205-217.

Jeffery Borland and Robert McDonald (2003) “Demand for Sport”, *Oxford Review of Economic Policy*, 19, 4, 478-502

Michael Butler, 2002. “Interleague Play and Baseball Attendance”, *Journal of Sports Economics*, 3, 4, 320-34

Eckard, Woodrow. 2001. “The Origin of the Reserve Clause,” *Journal of Sports Economics*, 2, 2, 113-130.

El-Hodiri, Mohamed and James Quirk. 1971. “An Economic Model of a Professional Sports League” *Journal of Political Economy*, 79, 1302-19

Forrest, David and Robert Simmons. 2002. “Outcome uncertainty and attendance demand in sport: the case of English soccer,” *Journal of the Royal Statistical Society*, *Series D* (*The Statistician*), Vol. 51, No.2, 229-241"

Forrest, David, Robert Simmons and Patrick Feehan. 2002. “A spatial cross-sectional analysis of the elasticity of demand for soccer,” *Scottish Journal of Political Economy*, 49, 3, 336-356

Fort, Rodney and James Quirk. 1995. “Cross Subsidization, Incentives and Outcomes in Professional Team Sports Leagues” *Journal of Economic Literature*, XXXIII, 3, 1265-1299

Horowitz, Ira. 1997. “The increasing competitive balance in Major League Baseball” *Review of Industrial Organization*, 12, 373-387.

Humphreys, Brad. 2002. “Alternative measures of competitive balance in sports leagues” *Journal of Sports Economics*, 3, 2, 133-148.

Knowles Glenn, Keith Sherony and Michael Haupert. 1992. "The demand for major league baseball: a test of the uncertainty of outcome hypothesis" *The American Economist*, 36, 72-80.

Levin, Richard, George Mitchell, Paul Volcker and George Will. 2000. *The Report of the Independent Members of the Commissioner's Blue Ribbon Panel on Baseball Economics*. NY: Major League Baseball.

Marc Poitras and Lawrence Hadley, 2003. "Do new Major League ballparks pay for themselves?" University of Dayton School of Business Administration Working Paper Series, WP03-6E.

Quirk, James and Mohamed El Hodiri. 1974. "The economic theory of a professional sports league" in *Government and the Sports Business,* Roger Noll, ed. Washington: Brookings Institution.

Rascher, Dan. 1999. "A test of the optimal positive production network externality in Major League Baseball" in *Sports Economics: Current Research*. John Fizel, Elizabeth Gustafson and Larry Hadley, eds. Praeger, Westport CT.

Ross, Stephen F. 1997. "The Misunderstood Alliance Between Sports Fans, Players, and the Antitrust Laws" *University of Illinois Law Review*, 1997, 519

Rottenberg, Simon. 1956. "The baseball player's labor market" *Journal of Political Economy*, 64, 242-258

Schmidt, Martin and David Berri. 2001. "Competitive Balance and Attendance: the Case of Major League Baseball", *Journal of Sports Economics*, 2, 2, 145-167.

Schmidt, Martin and David Berri. 2004. "The impact of labor strikes on consumer demand: an application to professional sports", *American Economic Review,* 94, 1, 344-357.

Scully, Gerald. 1989. The Business of Major League Baseball. Chicago: University of Chicago Press.

Albert Spalding, 1911. America's National Game. New York: American Sports Publishing Company.

Stefan Szymanski, 2003. "The Economic Design of Sporting Contests" *Journal of Economic Literature*, XLI, 1137- 1187.

Stefan Szymanski, 2004. "Professional team sports are only a game: the Walrasian fixed supply conjecture model, Contest-Nash equilibrium and the Invariance Principle" *Journal of Sports Economics*, 5, 2, 111-126.

Stefan Szymanski and Stefan Késenne, 2004. "Competitive balance and gate revenue sharing in team sports" *Journal of Industrial Economics*, LII, 1, 165-177.

Stefan Szymanski and Andrew Zimbalist, 2005. National Pastime: How Americans play baseball and the rest of the world plays soccer. Washington, DC: Brookings Institution Press.

Vrooman, John. 1995. “A General Theory of Professional Sports Leagues” *Southern Economic Journal*, 61, 4, 971-90

Zimbalist, Andrew. 2004. May the Best Team Win: baseball economics and public policy. Washington, DC: Brookings Institution Press.