Journal of Economic Literature
Vol. XLI (December 2003) pp. 1137–1187

The Economic Design of Sporting Contests

Stefan Szymanski[1]

1. *Introduction*

What is the optimal number of entrants in a race, or the optimal number of teams in a baseball league? What is the optimal structure of prizes for a golf tournament, or degree of revenue sharing for a football championship? How evenly balanced should the competing teams be in the NASCAR or Formula One championships? What is the maximum number of entrants per nation to the Olympic Games that should be permitted? What quota of qualifying teams to the soccer World Cup should be allocated to the developing nations?

These are all examples of design issues in sports. Sporting contests are one of the most significant branches of the entertainment industry, measured by the amount of time that consumers devote to following them. According to the U.S. Census Bureau, annual attendance at spectator sports in 1997 totaled 110 million (equivalent to 41 percent of the population), while annual household television viewing of sports events is estimated to be 77 billion hours per year.[2] One might add to this several hundreds of millions of hours spent in discussion at the water cooler. Designing an optimal contest is both a matter of significant financial concern for the organizers, participating individuals, and teams, and a matter of consuming personal interest for millions of fans. Not surprisingly, many lawyers and politicians express close interest in the way that sporting contests are run.

Economists have something to offer as well. The design of a sporting contest bears a close relationship to the design of an auction. In both cases, the objective of the organizer is to elicit a contribution (a bid, an investment, or some effort) from contestants who may as a result win a prize. The analogy between an auction and a contest/tournament[3] is already well known (see e.g. Arye Hillman and John Riley 1979). Given the objective function of the organizer and the

[1] The Business School, Imperial College, London. I am grateful to Jeff Borland, Braham Dabscheck, David Forrest, Bernd Frick, Philippe Gagnepain, Brad Humphreys, Erik Lehmann, Stefan Késenne, Gerd Muehlheusser, Steve Ross, Rob Simmons, Peter Sloane, Paul Staudohar, Tommaso Valletti, Andy Zimbalist, the editor, and two anonymous referees for valuable comments. I also thank seminar participants at the ESRC study group on the Economics of Sport, Arts and Leisure, the Stockholm School of Economics, and Glasgow University for their observations. Errors are of course my own.

[2] Kagan Media estimates that sports accounts for 25 percent of all TV viewing, while Nielsen Media Research estimates the average U.S. household views 2738 hours of TV per year (7.5 hours per day). This significance to consumers is not reflected in dollar spending. The Census Bureau reported in 1997 that spectator sports generate a direct income of only $14 billion domestically (0.17 percent of GDP). The annual value of U.S. major league sports broadcast rights is in the region of $4 billion (Soonhwan Lee and Hyosung Chun 2001).

[3] The words "contest" and "tournament" are used interchangeably throughout.

technology of the auction/contest it is possible to design an optimal prize scheme contingent on the distribution of contestant abilities/willingness to pay. While there have been a number of reviews of the economics of sports in recent years (e.g. John Cairns, Nicholas Jennett, and Peter Sloane 1986; Rodney Fort and James Quirk 1995; Lawrence Kahn 2000; John Vrooman 2000),[4] none of these has attempted to explore systematically the design of sporting contests.

The contest design approach may seem an unusual way of thinking to those who use baseball or soccer as their sporting paradigm. In these and other team sports we are accustomed to thinking of teams as independent entities that come together to agree on rules of the competition. In their review for this journal, Fort and Quirk (1995) state, "Professional team sports leagues are classic, even textbook, examples of business cartels." Members of a sports league certainly have common interests and may benefit from a reduction of economic rivalry between the teams. Many sporting contests are centrally coordinated, however, with little or no input from the teams or individual contestants; examples include the Olympic Games, the soccer World Cup, the New York Marathon, and the U.S. Open Golf Championship. What all these contests have in common is the need to provide contestants with the appropriate incentives to participate and perform. Joint decision-making through a cartel is simply one (possibly inefficient) mechanism to achieve this end.

This review attempts to systematize the contribution of economic thinking to design issues in sports, and to relate this research to the growing empirical literature on sports. This is an enterprise still in its infancy, however, and much remains to be done to understand fully the interaction of contest design and outcomes. The review will suggest new directions in which the literature may develop. A unifying theme of the paper is that the empirical literature can do much to shed light on the issues raised by the theoretical literature.[5]

The classification of sports is a subject that has exercised the minds of sociologists and economists alike. One distinction that can be made is between modern sports that have been formalized, quantified, and regularized on the one hand, and traditional sports that are often informal and only semi-structured on the other hand. Examples of the latter might include medieval football in Europe or the Aztec Ball Game (see Allen Guttman 1998 for further examples). This paper deals primarily with the commercialized modern sports, almost all of which were formalized somewhere between 1840 and 1900—e.g., baseball (1846), soccer (1848), Australian football (1859), boxing (1865), cycling (1867), rugby union (1871), tennis (1874), American football (1874), ice hockey (1875), basketball (1891), rugby league (1895), motor sport (1895), and the Olympics (1896).[6]

Historians (see, e.g., Tony Mason 1980; Wray Vamplew 1988) have argued that the process of formalization of sports mimicked the formalization inherent in industrialization and urbanization (time-keeping, routinization). Indeed, the commercialization of sport was initially an urban phenomenon, since industrial towns and cities were capable of supplying large paying audiences. It is probably for this reason that most modern sports were formalized either in Great Britain (the first industrialized nation) or the

[4] See also Andrew Zimbalist (2001) for a useful collection of seminal articles in the sports literature.

[5] This paper can thus be distinguished from fields such as "sabermetrics"—the study of baseball statistics for their own sake—which has little to do with empirical testing of economic theory.

[6] All of these dates, associated with early rulebooks, are subject to controversy. By contrast, golf, cricket, and horse racing had established rules and clubs from the mid-eighteenth century.

United States (the most rapidly industrializing nation of the late nineteenth century).[7]

In this paper we draw the distinction between *individualistic* sports (such as tennis, golf, and boxing) and *team* sports, such as soccer and baseball.[8] The distinction rests on the unit of competition and the nature of the demand for the contest. In team sports, the players act as agents on behalf of the team—which may be an actual employer (e.g. a club) or some delegated authority (e.g. a national team).[9] In individualistic sports the player acts as a sole trader; typically in these sports, the athletes/players enter competition in order to establish who is the best, because this is what interests the spectators. The relationship between the tournament organizer and the players is relatively simple. Players perform and agree to abide by the tournament rules in order to compete for a prize which is usually measured in terms of both status and money. Players make little long-term commitment to the organizers, even if it is an annual event, and select among available competitions to maximize their own utility. Likewise, the organizers make few commitments to the athletes, and typically offer places to the best players they can attract. The demand for an individualistic contest depends to a significant degree on the quality of the contestants participating and the amount of effort they contribute to winning. Thus an individualistic sporting contest conforms naturally to the standard contest model, outlined in the next section. Section 3 reviews the contribution of the empirical literature to testing the predictive power of contest models.

The demand for team sports is more complex. Firstly, while the organizational structure of individualistic sports is fairly uniform (for example, there is little difference between the organization of the New York Marathon and the Berlin Marathon, or that of the U.S. and the British Open Golf Championships), the organization of professional team sports differs substantially on either side of the Atlantic. Section 4 discusses the major differences and considers how these differences emerged from the different institutional settings that ruled at the foundation of baseball (the archetypal North American team sport) and soccer (the archetypal European team sport) at the end of the nineteenth century.

Secondly, while consumers of team sports resemble those of individualistic sports in wanting to see the best players, the nature of team sports "fandom" is that supporters tend to attach themselves to teams rather than players, and teams identify themselves with particular locations.[10] In practice this

[7] The other great industrial nation of the period, Germany, also developed its own sporting activity during this period, the gymnastic "Turnen" movement. This movement eschewed competition between individuals in favor of the development of a disciplined athleticism with military purposes in mind, and was ultimately ousted by the Anglo-Saxon sports (see Guttman 1998, ch. 7).

[8] Like all classifications, this one is at best imperfect. For example, the competitors in motor racing are teams of mechanics, but much of the spectator interest focuses on the individual exploits of the drivers. Rowing involves teams of rowers competing in a format that is very similar to most individualistic athletic contests, and horse racing is based on a distinctive form of cooperation between horse, trainer, and jockey. One difference is that in individualistic sports each contestant's marginal productivity depends only on their own effort, while in team sports it also depends on the marginal productivity of other team members. While this makes individual productivity difficult to measure, its economic significance may not be all that great. In many team sports such as baseball and cricket, team members' marginal products are almost entirely independent. Even where interactions are more important, the economic implications are unclear and their importance unproved. For example, if interaction terms were truly of economic significance in some team sports, one might expect to see players offering themselves to the market as partnerships, as happens, for instance, with teams of bond traders or teams of consultants. Even in team sports where the labor market is open to such possibilities (e.g. soccer, rugby, or cricket), player partnerships are almost unknown. The substantial empirical literature concerning the estimation of sports team production functions (see e.g. Kahn 1993) lies beyond the scope of this review.

[9] In amateur sports the team is a kind of partnership, and early professional baseball and cricket teams were also organized on this basis.

[10] While it is possible to be a fan of an individualistic competition (e.g. Wimbledon tennis) or event (e.g. the Olympics), this tends to happen only in the case of a small number of elite contests.

can mean that fans attach themselves to perpetually weak teams that do not hire the best players, and maintain such attachments over an entire life. However, contest organizers often express the concern that fans will lose interest in perpetually weak teams and that when this happens they will desert the sport altogether. To prevent this from happening, they argue, it is necessary to design the contest in such a way that all teams have roughly equal chances of winning, or that at least all teams win occasionally.[11] The competitive-balance issue has tended to dominate the analysis of team sports, and section 5 sets out some empirical evidence on competitive balance and related issues in North American and European team sports.

Section 6 considers possibly the most important theoretical contribution to the analysis of team sports: the so-called invariance principle. This states that (a) changes in ownership rights over player services (such as the introduction of free agency) and (b) certain types of income redistribution (such as gate revenue sharing) will have no effect on competitive balance. Empirical evidence on the first of these propositions is discussed in section 6.1, while section 6.2 considers the theoretical basis of the second.

Section 7 discusses other mechanisms used to promote competitive balance, such as prizes, salary caps, luxury taxes, promotion and relegation. The role of exclusive territories and its implications for optimal league size are also discussed in that section. The underlying objectives of organizers of team sports have been a consistent source of controversy over the years. Section 7 discusses the implications of the most commonly proposed alternative to the profit-maximizing hypothesis, namely, win maximization. While the controversy over the proper specification of the objective function of privately owned clubs is unlikely to be settled in the near future, this section also highlights the parallel development of ostensibly not-for-profit international sporting organizations offering international contests based on national representative teams (e.g. the IOC and the Olympics; FIFA and the soccer World Cup). The section concludes with a discussion of the growing rivalry in the soccer world between club-based and national-team-based competition.

Most sports are governed hierarchically, with a committee or commissioner at the apex of a pyramid possessing the right to change rules and arbitrate disputes. As sporting governments, these have found their authority challenged by the courts when dealing with matters that have an economic or commercial dimension. Section 8 provides a brief review of antitrust issues on both sides of the Atlantic. Section 9 concludes.

2. *The Design of Individualistic Sporting Contests*

It is relatively straightforward to apply contest theory to the design of an individualistic contest. Consider a simple footrace, organized by a profit-maximizing entrepreneur (e.g., the owner of a racetrack).[12] The organizer may generate a profit by selling tickets, broadcast rights, refreshments, or merchandise, or some combination of these. The organizer expects that spectators will be

[11] The Blue Ribbon Panel on Baseball Economics (Richard Levin et al. 2000), which was formed by the commissioner to investigate whether revenue disparities among the teams in Major League Baseball were undermining competitive balance, defined a proper level of competitive balance as a state where "every well-run club has a regularly recurring hope of reaching postseason play" (p. 1).

[12] In some sports it is frequently argued that profit maximization is not the objective of the organizers (most notably, see Sloane 1971). This may not make much difference to the design of a competition. For example, amateur sporting associations frequently seek to maximize income from a popular sporting event, which is then used to develop the grass roots.

attracted by the quality of the field entering the race and the effort the entrants contribute. Thus the objective is to design an incentive mechanism to maximize the effort contribution of the selected entrants.

2.1 *The Symmetric Winner-Take-All Contest*

The winner-take-all contest has been applied to a number of economic problems and originates with Gordon Tullock's (1980) model of a rent-seeking contest.[13]

The organizer's program can be written as

$$\underset{V}{Max}\ \pi = R\left(\sum_{i=1}^{n} e_i\right) - V \qquad (1)$$

subject to

$$p_i(e_i^*)\,V - e_i^* > p(e_i)\,V - e_i\,, \textit{ for all } e_i \quad \text{(incentive compatibility)}$$
$$p_i(e_i^*)\,V - e_i^* > 0 \quad \text{(individual rationality)} \qquad (2)$$

where $R(.)$ is a strictly concave revenue function that depends upon the sum of contributions e_i of each contestant, which can be interpreted in a number of ways (e.g. effort, investment, bids, ability) dependent on the context—for the remainder of this section it is labeled "effort." The cost of effort is assumed to be linear with marginal cost equal to unity. Equation (2) states that each contestant selects their optimal effort (incentive compatibility) and that all contestants willingly participate (individual rationality). The total payoff to each contestant depends on the probability of success (p_i) multiplied by the value of the prize (V), less the cost of effort. It is assumed that the contestants are risk neutral.[14]

The probability of success is defined by the technology of winning—the Contest Success Function (CSF)—which depends on both the effort contribution of the athletes and their inherent abilities. For the time being, we assume that all contestants have equal ability (symmetry). A natural form for the CSF is the logit function

$$p_i = \frac{e_i^{\gamma}}{\sum_{j=1}^{n} e_j^{\gamma}}\ , \qquad (3)$$

where γ is a measure of the discriminatory power of the CSF. A high γ implies that even slightly higher effort than one's rivals ensures a high probability of winning the prize, while a low value of γ implies that differences in effort have little impact on outcomes.

This winning technology differs fundamentally from that assumed in an auction, where the highest bidder wins with probability one (the contest is perfectly discriminating). Here, the technology does not discriminate perfectly between effort levels, and the highest bidder can only be certain of winning if all other contestants contribute no effort at all, except in the limiting case as γ goes to infinity, when the logit contest becomes perfectly discriminating.[15] That contests are in fact imperfectly discriminating, yielding uncertainty of outcome, was recognized by Walter Neale (1964) in his seminal paper.

[13] The analysis of rent-seeking contests has been applied to, inter alia, labor markets (e.g. Edward Lazear and Sherwin Rosen 1981), competition for innovation (e.g. Glenn Loury 1979) and competition for research contracts (e.g. Curtis Taylor 1995). There is also a substantial related literature on all-pay auctions (see e.g. Michael Baye, Dan Kovenock, and Casper de Vries 1996). Theoretical research on the implications of rent-seeking contests includes Baye et al. (1999), Ani Dasgupta and Kofi Nti (1998), Avinash Dixit (1987), Jerry Green and Nancy Stokey (1983), Richard Higgins, William Shughart, and Robert Tollison (1985), Barry Nalebuff and Joseph Stiglitz (1983), Shmuel Nitzan (1994), Nti (1997), and Stergios Skaperdas (1996).

[14] Risk aversion is a natural assumption in many examples of labor-market contests, but in sporting contests involving professional athletes risk neutrality seems less objectionable. The very fact of investing the time and effort from an early age to become a professional athlete, when the probability of substantial earnings is very low, would seem to suggest selection in favor of those with negligible risk aversion.

[15] Lazear and Rosen (1981) and Dixit (1987) use the probit model, but as yet this has not been applied to the analysis of a sporting contest.

Solving the contestants' first-order conditions, we find the optimal effort level in the symmetric case[16]:

$$e_i^* = \frac{\gamma V(n-1)}{n^2}, \qquad (4)$$

from which it is apparent that

(i) individual and aggregate effort is increasing in the value of the prize;
(ii) individual and aggregate effort is increasing in the discriminatory power of the CSF;
(iii) individual effort decreases with the number of contestants;
(iv) aggregate effort increases with the number of contestants.

These results are intuitive, although perhaps the third might surprise noneconomists. Large fields of contestants are usually associated with highly prestigious contests such as the Olympics, so there may be a correlation between the value of a prize and the number of entrants, which obscures the discouragement effect of large fields on effort. However, organizers of individual race meetings typically do seek to limit the field so as not to dilute the incentives of the participants. The result is very similar to the standard Cournot-Nash oligopoly result that equilibrium output choices for individual firms decrease in the number of competitors but the aggregate output increases.[17] If the organizer is interested in obtaining the maximum winning effort then the optimal number of contestants is two (see e.g. Richard Fullerton and Preston McAfee 1999). If the organizer is interested in a specific level of performance then the reward function may look more complicated than a simple contest: e.g., a bonus based on the race time plus a prize for winning.

[16] Here we focus on pure strategy equilibria. A mixed strategy may exist even if a pure strategy equilibrium does not (see e.g. Baye, Kovenock, and de Vries 1994). Note that the equilibrium described here will not be symmetric if there are some contestants who decide not to enter the race; we ignore this possibility here.

[17] Nti (1997) shows that the result on aggregate effort is sensitive to the type of winning technology selected.

Having identified the incentive-compatible investment level, it is then trivial for the organizer to select the prize fund to maximize the difference between revenues and costs.[18]

2.2 *Multiple Prizes in Symmetric Contests*

In practice, most organizers of sporting contests do not offer a winner-take-all prize: in addition to gold medals, there are silver and bronze. Benny Moldovanu and Aner Sela (2001) show that multiple prizes can be optimal in a perfectly discriminating all-pay auction, depending on the cost structure of the bidding technology—if the cost of bidding is linear or concave, a single prize dominates any other prize structure. If costs are convex, however, a second prize can be optimal. Szymanski and Valletti (2002) extend the analysis of the problem to an imperfectly discriminating (logit) contest. They show that if contestants are symmetric, a first prize always dominates, while if contestants differ enough in ability then a second prize can be optimal. In an imperfectly discriminating contest offering a prize fund to be divided between the first and the second prize, the return to contestant i can be written as

$$\left(p_{i1}k + (1-p_{i1})p_{i2}(1-k)\right)V - e_i \qquad (5)$$

where k is the fraction of the prize fund allocated to the first prize, p_{i1} is the probability of contestant i winning the first prize, and p_{i2} is the probability of i winning the second prize (contingent on not having won the first prize). Note that when the contest is symmetric the probability of winning the second prize in equilibrium is the same whoever wins the first prize (other than contestant i). For a logit contest p_{i1} is still defined by (3), while p_{i2} is the equivalent expression for the probability of winning second prize, the only difference being that the contest for second prize involves $n-1$ contestants rather than n. Hence, in general:

[18] The first order condition is $R'\,\gamma(n-1)/n = 1$.

$$p_{i2} = \frac{e_i^{\gamma}}{\sum_{\substack{h=1 \\ h \neq j}}^{n-1} e_h^{\gamma}}. \quad (6)$$

In the symmetric case, the first-order condition for contestant i can be rearranged to show that

$$e^* = \gamma V\left(\frac{(n-1)}{n^2} - \frac{(1-k)}{n(n-1)}\right) \quad (7)$$

from which it follows that an increase in the weight attached to the second prize (reducing k) will lead to a reduction in effort. In a two-person contest, effort falls to zero when $k=\frac{1}{2}$ (the second prize is identical to the first prize) but will be positive for all values of $k \in [0, 1]$ for $n > 2$.

2.3. *Asymmetric Two-Person Contests*

Although symmetric contests should only ever have first prizes, most sporting contests are in practice asymmetric: there are favorites and long shots. This complicates the issue in two ways. First, in a symmetric contest there is no trade-off between winning effort, average effort, and the variance of effort. In an asymmetric contest the organizer must decide the appropriate objective. Maximizing winning effort is often important (e.g. breaking the world record). On the other hand, a close contest (competitive balance) may be valued if consumers like to see an even contest,[19] and the organizer may be keen to maintain the overall quality of the contest (average effort). Providing greater incentive for winning effort may reduce the effort of weaker contestants and so reduce average effort. Even if average effort does not decline, the variance of effort may increase. Secondly, in an asymmetric contest the existence of a second prize may not only increase the average and/or reduce the variance of effort, it may also increase the winning effort.

[19] Competitive balance is discussed in more detail in section 5.

Asymmetry has been little studied in the contest literature (two notable exceptions are Dixit 1987 and Kyung Baik 1994), even though this is a fundamental characteristic of many contests, not least in sport. Asymmetry can be modeled either as a difference in the cost of effort required to achieve a given winning probability or as a difference in the winning probability for any given level of effort. Taking the first of these approaches, the payoff functions in a two-person contest can be written as

$$\begin{aligned}\pi_1 &= p_{11}kV + (1-p_{11})\,p_{12}\,(1-k)V - (1-\beta)e_1 \\ &= (2k-1)\,p_{11}V + (1-k)V - (1-\beta)\,e_1\end{aligned}$$

$$\pi_2 = (2k-1)\,p_{21}V + (1-k)V - (1+\beta)\,e_2. \quad (8)$$

Asymmetry has two effects on the contest. Most obviously, it will create a competitive imbalance—the greater is β the larger the low-cost player's winning probability—and if asymmetry gets large enough the participation constraint of the weak contestant will be violated. Secondly, it can affect total effort. Faced with two asymmetric contestants, the usefulness of a second prize as an instrument of the contest organizer is relatively limited. Total effort increases in the size of the prize fund and the share awarded to the winner. The two first-order conditions for effort derived from (8), assuming the logit CSF (3), imply that the effort ratio in equilibrium is

$$\frac{e_2}{e_1} = \frac{1-\beta}{1+\beta}. \quad (9)$$

This tells us that while the contest becomes less balanced as the difference in the cost of effort increases, the prize structure has no effect on relative effort: a second prize does nothing to improve the balance of the contest. This suggests two policy options for the organizer if competitive balance matters: (a) screen for ability to ensure balanced contests and (b) handicap the strong player, i.e., increase the strong player's (marginal) cost or subsidize the weak player's (marginal) cost. Fullerton and McAfee (1999) consider

the case where ability is not observable and show that by both setting the prize and charging an entry fee the organizer can ensure that the best contestants enter and offer first-best effort either in a homogenous contest with fixed costs or a heterogenous contest. This may explain, for instance, why it is common to observe that entry to races with large financial prizes is by invitation only to an exclusive group of athletes.

Defining $z=(1-\beta)/(1+\beta)$, total effort is given by

$$e_1+e_2=\frac{2\gamma V(2k-1)z^{\gamma}}{(1+z^{\gamma})^2(1-\beta^2)}. \qquad (10)$$

When $\gamma=1$ it is clear that $e_1+e_2=V(2k-1)/2$, so that total effort is independent of β and any increase in asymmetry yields offsetting increases and decreases in effort from the strong and weak players respectively. When $\gamma>1$ (the contest is relatively discriminating) increasing asymmetry reduces total effort since the discouragement effect for the weaker player outweighs the encouragement effect for the stronger player. On the other hand, when $\gamma<1$ the reverse is true and increasing asymmetry increases effort (when the contest is not very discriminating no one is very motivated to supply effort, but asymmetry provides an encouragement to the strong player to secure the prize). Lazear and Rosen (1981, p. 858) demonstrate similar results in a rank order labor tournament where the CSF is asymmetric, but in their model the effect on total effort depends on the concavity or convexity of costs.

2.4 *Asymmetric Contests With More Than Two Players*

With more than two players a second prize can be a useful instrument for the organizer. For instance, a second prize can now be a motivational device. Szymanski and Valletti (2002) develop a formal model of a three-person contest to show that second prizes may not only improve competitive balance, but also increase total effort. The intuition is quite straightforward. Consider a three-person race with two weak contestants and one strong one. If the players are more or less evenly matched, then it pays to put all the weight on first prize as in a symmetric contest. But the motivation effect of the first prize is dulled if the two weak contestants are very weak, because however much effort they make they have little chance to win. It follows that if two out of three contestants give up then even the strong contestant is unlikely to make any effort. By introducing a second prize, the two weak contestants are given something to play for, and as a result of their effort even the strong contestant cannot coast along quite so easily and is provoked into supplying more effort. This observation suggests that large prize spreads should be observed when contestants are relatively evenly matched but narrower spreads should be offered when there are large differences in ability.[20]

A second prize may also improve competitive balance, but at this stage a problem of definition arises. It is natural to think of balance in terms of the variance of contributions, but with three or more contestants it is possible for different combinations of effort to produce the same variance, while in reality the organizer may not be indifferent among them. For example, consider a three-person contest where only effort matters. If contestant 1 contributes three units of effort, contestant 2 contributes two units, and contestant 1 a single unit, the variance of effort (equal to one) would be the same as an alternative case where the first contestant supplied 2.732 units and the other two supplied a single unit each. In the first case there is an equal gap between each contestant, while in the second case there is a larger gap

[20] The modern practice in schools and elsewhere of offering almost all competing students a prize of some sort for participating in sporting contests is often criticized as an excess of political correctness—but in this context it might be viewed as simple recognition of the need for motivation for all contestants when abilities are heterogenous.

between the strong player and two equally weak players. The race for first place may be more exciting in the first case, but even then the strong player has a big lead (in terms of effort). By contrast, the second case will at least produce a close race for second place, which may compensate for a lack of tension in the race for first place. An argument can be made for either case being more attractive. The problem is that there is no natural metric for competitive balance when $n > 2$, and thus it may be difficult to rank different incentive schemes.

2.5 *Match Play*

In many sporting contests the organizers must make a structural choice between match play and simultaneous play by many contestants. For example, a golf tournament could be organized by pairing contestants and allowing the winner from each pairing to enter the next stage until a winner emerges from the final pairing, or all players could play simultaneously and the player with the lowest score would be declared the winner. Some sports, such as tennis, cannot realistically be organized as simultaneous contests, while others, such as Olympic track and field, typically have elements of both (e.g., eight lanes of runners and the fastest go through to the next round).

Rosen (1986) specifically used the example of a tennis tournament to consider the optimal prize structure in order to maintain effort over a match play tournament. He showed that if the reward for winning increases linearly as the tournament progresses, then effort will decrease, since the added spur of reaching higher and higher prizes is diminished. This, he argued, rationalized the observation that rewards are often heavily skewed toward the top end of a contest, since this prize structure will ensure that effort is nondecreasing.

Mark Gradstein and Kai Konrad (1999) compared simultaneous contests (which they labeled S-contests) and match play contests (which they labeled T-contests) where a single prize is awarded to the ultimate winner of the contest. They showed that in a symmetric contest where the object is to ensure dissipation of all the rents (i.e. so that total effort expended equals the value of the prize), an S-contest is preferred for $\gamma > 1$ (high discriminatory power), while for $\gamma < 1$ a T-contest is preferred, and for $\gamma = 1$ the choice makes no difference. The intuition behind this result is that when discriminatory power is high a single simultaneous contest is enough to ensure that all rents are dissipated. But when the discriminatory power of each individual contest is low a single contest cannot dissipate all rents, whereas a multistage contest, in which contestants have to put in additional effort at each stage, can dissipate rents.[21]

2.6 *Dynamic Contests*

The contests described so far have been one-shot games or, in the case of sequential contests, it has been assumed that the contestants compete in every round until eliminated. However, if contestants acquire information about the state of play as the game progresses, they may decide to drop out altogether. There are a number of models in the economics literature that examine contests in a dynamic context, most notably the war of attrition and competition for monopoly, preemption games associated with patent races (both types of game are reviewed in Drew Fudenberg and Jean Tirole 1991) and market share attraction games in the advertising literature (see George Monahan and Matthew Sobel 1994). These have some implications for contests that involve a sequence of competitions such as the T-contests described above. Many of these types of contest are found in team sports, but individualistic contests can also involve a dynamic element, either because the contest itself

[21] See also Moldovanu and Sela (2002) for discussion of different contest architectures in all-pay perfectly discriminating auctions.

is drawn out (e.g., a marathon or a five-set tennis match) or because players compete throughout a season for rankings.

In the war of attrition, competitors supply effort in the expectation of winning a prize at some future date when all rivals have dropped out of the contest. If contestants are symmetric then a pure strategy equilibrium (in which each contestant is indifferent between staying in and dropping out of the game) does not exist. A mixed strategy equilibrium does exist where each player exits with some probability and the probability equates the expected value of remaining with the expected gain from quitting. However, asymmetric pure strategy equilibria also exist, and if the contestants have different abilities the game may be degenerate with weaker contestants withdrawing instantly (see e.g. Jeremy Bulow and Paul Klemperer 2001).

In the war of attrition, contestants learn nothing from their continued participation in the game (the game is memoryless). In preemption games (e.g. Christopher Harris and John Vickers 1985; Fudenberg and Tirole 1985) the players acquire experience (e.g. know-how in a patent race), and experience increases the probability of success, so that at any point the perception that one player has an established lead may cause all the other players to withdraw. In particular, if one player is known to enter the race with an established advantage, no other contestants will enter (or, if they enter, will supply zero effort), a result known as $\in$-preemption (see Fudenberg et al. 1983). This kind of first-mover advantage can thus undermine the incentive of contestants, especially weaker ones, to supply effort, effectively handing success to the dominant players "on a plate." This extreme result is sensitive to assumptions about information sets, however, and if there is some uncertainty about the state of play then the follower might have an incentive to "leapfrog" ahead of the leader (e.g. Harris and Vickers 1987).[22]

As far as a contest organizer is concerned, these types of games are degenerate, in the sense that spectators typically expect to watch a full contest and might ask for their money back if one of the contestants pulled out.[23] However, in contests where the cost of effort is extremely high (e.g. marathon running and heavyweight boxing) it is not uncommon for an out-of-contention player to pull out. Contest organizers may try to create some uncertainty about performance levels (perhaps even changing the rules) in order to prevent this from happening.

3. *Empirical Research on Individualistic Sports*

The research agenda discussed in the previous section can be summarized under four main headings:

(i) The impact of prizes on incentives to perform (depending on discriminatory power, effort functions, and the size of the prize fund);

(ii) The impact of the distribution, or spread, of the prize fund (second prizes, third prizes, and so on);

[22] "Consider a foot race between two athletes. Assume that it is common knowledge that the two athletes are equally good, and that they prefer to reserve themselves (run at a slow pace) rather than exhaust themselves by running at a fast pace. Suppose further that the leader has eyes in the back of his head and can monitor whether the follower is catching up. Because the leader can keep his lead by speeding up if the rival does so, there is no point for the rival in even engaging in the race. The leader can thus proceed at a slow pace without fear of being leapfrogged. But the picture changes dramatically if the two athletes run on tracks separated by a wall. Suppose that the wall has holes, so that from time to time each athlete can check his relative position. Now the leader can no longer run at the slow pace; if he did, the follower could run fast, leapfrogging the leader without his noticing it, and force him to drop out of the race at the next hole. Thus lags in information (or in reaction) engender competition" Tirole (1988).

[23] Of course, if a championship is decided as a "best of n matches" like the seven-match World Series, the organizers are keen to see the contest go to the wire. This is yet another reason for wanting competitive balance.

(iii) The impact of the structure of the contest (number of contestants, simultaneous or sequential contests and so on);
(iv) The impact of pre-screening and handicapping.

Researchers in the field of contest theory have set out to explain the widespread use of prizes as an incentive device in labor and product markets.[24] The claim that sports provides a natural laboratory for testing hypotheses from the economics literature is widely made (e.g. Kahn 2000). While that paper focused primarily on team sports, it pointed out that "some of the most intriguing evidence on the links from incentives to performance comes from sports … like golf and marathon running." In these sports it is possible to gather data on individual performance and relate that data to the prize structure offered in individual tournaments. Perhaps the best-known results are those of Ronald Ehrenberg and Michael Bognanno (1990a,b) who examined scores in American and European PGA golf tours.[25] Their principal finding is that scores tend to be lower (so performance is better) when the prize fund is larger, which seems to be a striking endorsement of tournament theory. They also considered the effect on the final round score of an individual's current position in the contest. Since the prize spread decreases with rank (the difference between the first and second prize is much larger than the gap between the tenth and eleventh prize) it is predicted that effort will be higher and scores lower in the final round when a player has a higher placing at the beginning of the round (this hypothesis presumably reflects the notion that a laggard will be discouraged as in a war of attrition). This prediction is also strongly confirmed by the data.[26]

Another important issue that Ehrenberg and Bognanno address is the relationship between performance in a given tournament and entry. If larger prizes attract better contestants then the observed improvement in scores may be attributable to the "sorting" effect rather than the tournament incentive effect. In fact, they found no evidence that their prize results were due to sample selection bias. This issue has also been addressed in James Lynch and Jeffrey Zax (2000), who examine data on nearly two thousand contestants covering 135 different road races in the United States ranging between five kilometers and a full marathon (42 km). They were able to construct a measure of pre-race expected rank, based on an athlete's previous history, and then to construct a measure of the incentive to supply effort based on the difference between the prize for achieving his or her pre-race rank and one rank lower than this (presumably the asymmetry of the race is thought to be large enough that multiple prizes are required to increase total effort). They find on this basis that recorded times are decreasing in the prize difference, apparently suggesting higher effort in response to larger prize spreads. However, once the pre-race ranking variable is included in the regression, to account for the quality of the field entering the race, the impact of the prize spread disappears. The authors thus attribute the impact of prize spreads to the sorting effect rather than the tournament incentive effect.

Michael Maloney and Robert McCormick (2000) use data on 115 footraces ranging between one mile and a full marathon

[24] This research agenda is therefore primarily positive rather than normative. However, the adoption of procurement auctions by governments has introduced a normative element to this literature.

[25] "The Ehrenberg and Bognanno work is perhaps the best test of tournament theory, not because it is easily generalizable to the corporation but rather because the data are so well suited to testing the model" (Lazear 1995, p. 33).

[26] Michael Orszag (1994) was unable to replicate these findings using data on the 1992 U.S. PGA tour. He argues that this may be due to increased media pressure since the 1980s causing more randomness (e.g. nerves) in the relationship between effort and performance.

involving nearly 1500 athletes. They identify the sorting effect with the total size of the prize fund and the incentive effect with the prize spread, and find that both effects are statistically significant and have the expected sign. Although on average the prizes seem quite small (about $400), their impact is significant since doubling the prize spread reduces race times by 4 percent. One weakness of these footrace studies is that the contestants do not include a significant fraction of the world's best, which is reflected in the average times of the sample. Bernd Frick, Joachim Prinz, and Alexander Dilger (2001) consider a sample of 57 marathons run worldwide and involving much larger prize money ($135,000 per race in 1993 dollars). They examine the impact of the total prize fund, its distribution, and bonuses paid for achieving a fast time. They find that (a) doubling the average prize reduces average times by 1 percent; (b) doubling the spread improves average times by 2 percent; (c) doubling bonus payments improves average times by around .75 percent; (d) increasing the prize fund, spread, and bonuses increases the closeness of the race, measured as the time difference between the winner and other finishers; and (e) race times are decreasing in the number of "in the money" ranks (i.e. the number of prizes).

Apart from footraces and golf, almost the only other individualistic sport to have produced some empirical research is horse racing.[27] Susan Fernie and David Metcalf (1999) examined the effect of a change in the compensation of British jockeys which involved replacing performance-related payments with noncontingent retainers. Their evidence shows that individual performance deteriorated.[28] Higgins and Tollison (1990) examine the impact of the number of contestants on the average distance of contestants behind the winner in the Kentucky Derby and find that larger fields tend to fall further behind the winner, which they equate with a slower race, consistent with contest theory. However, they also find that larger prizes do not appear to produce systematically faster times.

Michael Maloney and Kristina Terkun (2002) address an issue that has generally been neglected in the literature, notably the competition between prize-givers and the impact of this competition on prize spreads. They point out that if prize-givers compete to attract contestants, as is the case with motorcycle racing sponsors, who are the subject of their study, then if the prize fund offered by rival sponsors increases, all else equal, a given sponsor must reduce the prize spread in order to attract the same contestants. They find that this prediction, which they derive from Lazear and Rosen, is indeed supported by the data on prize funds and spreads in a sample of 112 sponsors of motorcycle races.

One concluding comment on individual contests concerns cheating. Thus far we have assumed that all efforts contributed are equally valid, while in reality certain kinds are proscribed (e.g. bribery and performance-

[27] Ignoring the possibility that horse and jockey operate as a team. Team elements might also be identified in golf (player and caddie) and foot races (e.g. runner and trainer).

[28] Brian Becker and Mark Huselid (1992) analyzed driver performance in NASCAR races and found that prize spread improved race times. While much of the interest of the fans is focused on the drivers in this sport, there is clearly a very strong team element in the preparation of the car. Rafael Tenorio (2000) considers the practice in boxing of providing a "purse" for title fights that depends not on current but rather on past performance. He points out that this may lead to inadequate effort supply in these matches. However, this phenomenon has much to do with the risk attached to boxing. Because of the fragmentation of governing bodies in boxing, promoters compete to offer boxers the best terms to stage a fight. A similar situation applies in the world of chess, where payments for the appearance of champions also tend to be high and independent of performance, but in this case the personal risks are not so great and so the temptation to "take the money and run" (or rather, fall over) is not so great.

enhancing drugs). Little has been written on the economics of cheating in this sense, although a recent paper by Mark Duggan and Steven Levitt (2002) illustrates the potential for research in this area. A related point, raised by Lazear (1989), is that tournaments create an incentive to undermine the performance of rivals in order to increase one's own probability of winning—i.e. sabotage. Luis Garicano and Ignacio Palacios-Huerta (2000) have examined this proposition for the case of soccer, where a change in the points system appeared to lead both to more creative effort and more sabotage (fouls, in the case of soccer).

Despite the enthusiasm of theorists for sports as a laboratory for testing contest theory, it is apparent that there remains a great deal more work to be done in this field. Almost the only issue considered thus far has been the impact of the size and spread of the prize fund. While most research seems to confirm the most basic economic proposition that bigger prizes produce more effort, even this result is subject to dispute due to the simultaneity of sorting and incentive effects. Larger prize spreads seem to elicit more effort, but the pure winner-take-all contest appears to be a purely theoretical possibility.

Issues deserving further attention include the value of screening, the role of handicapping, contest structure (match play and simultaneous contests) and discouragement effects, the impact of penalties (e.g. failing to make the cut in golf), the impact of qualifying races, cheating, sabotage, and possibly other issues. None of the papers discussed examined in any detail the objectives of the organizer, which are clearly critical in determining the optimal design. For example, rules on qualification for the Olympic Games reflect the values of the founders of the Olympic movement, and are not simply intended to find the fastest runner or swimmer. Discrimination against stronger nations by restricting the number of athletes per nation has a significant influence on the outcome of competition.[29]

4. *The Comparative Economics of Team Sports*

4.1 *Peculiar Economics*

The analysis of team sports has been primarily motivated by normative issues.[30] Economic analysis has been used to advise team owners and player unions when negotiating wage deals, as testimony in antitrust cases, as testimony in congressional hearings on legislation, and other proposed public interventions in the organization of sporting contests. Economists and lawyers have also used economic analysis to propose alterations to the design of sporting contests (see e.g. Fort and Quirk 1999; Zimbalist 2003; and Stephen F. Ross 1989).

The analysis of normative problems in sports, as in many activities, is often made more difficult by the role of culture. A contest design that is optimal for a particular group of consumers may not be to the taste of another. A good example is the attitude toward player trading in team sports. In North America most fans seem to frown upon player mobility and place the greatest value on players who remain loyal to the same team over their entire career. In Europe, however, player trading has always been an accepted part of the soccer system. While most fans would prefer that good players remain on the team, mobility is accepted as a fact of life and fans do not

[29] This is also an important issue in team sports. For example, until the 1970s, European and South American teams were awarded a disproportionate share of qualifying places in the soccer World Cup, while after that period the policy was reversed by the governing body (FIFA). Since the 1970s the African teams were given an increased share and have (therefore?) been increasingly successful in the tournament (John Sugden and Alan Tomlinson 1999).

[30] Perhaps the main exception, to this has been in the field of labor economics where data on earnings in team sports has been used to develop tests of discrimination (reviewed in Kahn 2000 and Sherwin Rosen and Allen Sanderson 2001).

seem to express opposition to player trading in principle.[31]

It is possible that different attitudes may reflect broader cultural differences, while historical accident and path dependency may also account for different practices. Clearly, tradition and folk memory are an important aspect of sports fandom—but are all traditions equally likely to stick, or are some more likely to hold in some cultures than in others? For example, Americans and Europeans seem to enjoy the same kinds of individualistic sports (Olympic sports, golf, tennis, boxing, etc.) but most are attracted to quite different team sports. Moreover, as pointed out in the introduction, while the design of individualistic contests seems to be relatively similar throughout the world, there are some substantial organizational differences between North American and European team sports. It is useful therefore to begin the analysis of team sports by some comparisons in the development of the archetypal American team sport, baseball, and the archetypal European team sport, soccer.

4.2 *Baseball*

Harold Seymour (1960), the authoritative historian of early baseball, made it clear that the structure of the National League, created in 1876, and the foundation of organized baseball emerged as a consequence of the free-for-all that was undermining interest in the new national sport. From the end of the Civil War, interest in the game spread rapidly across the United States, with teams and competitions proliferating and vying to attract spectators. The barnstorming teams of this era crossed the country in search of opponents, relying on reputations driven by winning records to generate income. The natural equilibrium of this free-entry dynamic game is (a) barnstorming teams attract support as long as they are winning and then collapse when they lose (a rational bubble); (b) team owners dissipate all the rents in competing to hire the best talent; and (c) the opportunities for gambling on the records of individual teams generate match fixing.

The founders of the National League set out to create a new kind of equilibrium, more satisfactory for team owners. The National League was a deliberately elitist affair. Its exclusivity invested members with a stake in its long-term success (to combat short-run incentives for match fixing); its granting of exclusive territories guaranteed a local monopoly (providing an incentive to invest in the local market); and its reserve clause established monopsony rights over the players (ensuring that the income stream from matches accrued principally to the owners). The extraordinary success of this model made it not only the basis for the national sport of the United States, but also for the other North American team sports (football, basketball, and ice hockey). American sports played in other countries adopted this model (e.g. baseball in Japan and Mexico, basketball in Australia), as have some other sports in other countries influenced by the United States. (e.g. Australian Rules Football in the 1970s). While other team sports in the United States developed new organizing principles (e.g. the draft in football or the salary cap in basketball) these principles were largely integrated into a common framework that characterizes each of the major sports. These common elements include:

1. organizational independence of the domestic major leagues;
2. a fixed number of teams;
3. entry through the sale of expansion franchises;
4. exclusive territories and franchise mobility;
5. draft rules giving teams monopsony rights in player acquisition;
6. roster limits;

[31] Leo Kahane and Stephen Shmanske (1997) found that teams with more stable team rosters enjoyed higher attendance, all else equal. Fans appear genuinely to prefer team stability. There is no evidence of any such preference among European soccer fans.

7. low player mobility and limited player trading for cash, especially for top stars;
8. collective bargaining over player conditions;
9. collective sale of national broadcast rights (exempted from antitrust);
10. collective sale of merchandising;
11. restrictions preventing the stock market flotation of clubs.

Each of these arrangements has been adopted to a greater or lesser extent, but is present in all the major leagues.[32] Some other types of agreement, such as gate-revenue sharing (MLB[33] and NFL) and salary caps (NBA, NFL) have not been universally adopted, but are not inconsistent with the structure of the non-adopting leagues (which have considered adoption and may yet adopt). These structures are quite distinct from those found in sports leagues outside of the United States, most notably in the case of soccer, arguably the world's most popular team sport.

4.3 *Soccer*

The creation of the Football League in England in 1888 had similarly momentous implications for the national pastime of nations that adopted the British model of league organization (see Simon Inglis 1988 for full details). The Football League was formed by a group of teams that belonged to an all-encompassing governing body, the Football Association (FA), founded in 1863. As well as laying down the rules, the FA administered its own successful club competition, the FA Cup, and organized international representative matches against other countries using club players.[34] Unlike the founders of the National League, the founders of the Football League did not break away from the existing structures, but worked inside them. This meant that (a) the Football League never attempted to become an exclusive institution, but intended from the start to admit, eventually, all the major teams into its ranks, and (b) League teams accepted from the beginning the practice of releasing star players to represent their country in international competition without compensation (although this has become increasingly controversial).

As soccer spread rapidly around the globe and other nations adopted the British system, there evolved a distinctive organizational structure involving (i) an overarching governing body responsible for the rules and organizing highly successful competitions (e.g. the World Cup, the European Championship) independently of domestic league authorities; (ii) a domestic league system incorporating promotion and relegation[35]; and (iii) a system where star players are paid employees of clubs and play for them (primarily) in league competition, and are also representatives on the national team, whose success is usually seen as even more prestigious. This system has also been applied to a number of other team sports, usually in countries where the soccer system is dominant (e.g. rugby union and basketball in Europe[36]). Common elements of the "soccer system" include:

1. integrated governance structure within a global hierarchy and national leagues subordinate to national associations that participate in international competition using league players;
2. mobility of teams through the system of promotion and relegation;
3. free entry for new teams at the bottom

[32] Here meaning Major League Baseball (MLB), the National Football League (NFL), the National Basketball Association (NBA), and the National Hockey League (NHL).

[33] In the 1990s MLB ceased sharing gate revenues only in favor of local revenue sharing (including TV income).

[34] The first FA Cup final and the first international match (Scotland v. England) both took place in 1872.

[35] This is a structure in which clubs affiliated with the governing body are promoted from a given league division to its immediately senior division on the basis of league ranking at the end of each season, and subject to relegation to the immediately junior division on the same grounds.

[36] There are exceptions: in the United Kingdom, Rugby League has adopted many American-style restrictions. The case of Australian team sports is interesting, since these had structures resembling European sports until the 1980s but since then a number of American institutions have been adopted (see e.g. Braham Dabscheck 1989; Rob Hess and Bob Stewart 1998).

of the hierarchy, but promotion on sporting merit only;
4. nonexclusive territories;
5. competitive labor markets at the entry stage, no draft;
6. no roster limits;
7. high player mobility and trading for cash, especially for top stars;
8. limited unionization or collective bargaining over player conditions;[37]
9. limited collective sale of national broadcast rights (no antitrust exemption);
10. no collective sale of merchandising;
11. limited restrictions on the stock market flotation of clubs.

These are material differences from the "baseball system" described above. A further institutional difference lies in the plurality of major soccer leagues compared to the North American major leagues. While competition among rival leagues has characterized part of the history of North American sports, in most cases competition at the level of the league has not survived long. Fans are drawn to the best competition; competing head to head to attract talent drives down profits to the point where either leagues have folded or the incentive to reestablish monopsony has led to mergers. The close substitutability of rival major leagues in the eyes of consumers has thus been the driving factor toward establishing dominant major leagues in each of the North American team sports, particularly in the television age. In European soccer, however, the more rigidly defined regional loyalties associated with national territories has meant that the national leagues of Italy, Spain, Germany, and England have been seen as only imperfect substitutes, and while competition for player services is intense, it has not brought about league bankruptcy or mergers (even for relatively small European nations such as Belgium, Denmark, or Greece). This issue is discussed in more detail below.

Some commentators, most notably Fort (2000), have argued that these institutional differences have given rise to structural differences that are more apparent than real. For example, he argues that the difference between the closed, North American leagues and the open soccer leagues of Europe (i.e. open to new entry through promotion and relegation) has little practical effect, since both systems ensure that the best teams and talents migrate to where they are most valued, whether it be through franchise expansion or promotion. The proposition that institutional differences have no implications for the attractiveness of sporting contests is a natural starting point for both theoretical and empirical analysis of team sports, as has been shown by much of the comparative analysis of team sports inside the United States (e.g. Quirk and Fort 1992; Gerald Scully 1995).

Moreover, some proposals for the reform of North American leagues have a distinctly European flavor. For example, the proposal to break up the major leagues into competing entities (Ross 1989; Quirk and Fort 1999) would create a structure in which independent leagues competed among themselves in the regular season and came together for the play-offs. This is similar to the European model where teams compete in national leagues as well as a pan-European Champions' League. Roger Noll (2002) and Ross and Szymanski (2002) have proposed the adoption by the major leagues of the European promotion and relegation system (see section 7.5). Extending the analysis of team sports to assess the effect of the strikingly different institutions of soccer offers a rich laboratory for researchers.

[37] It is perhaps more historically accurate to say that unions were relatively weak both in North America and in Europe until the 1950s. On both continents, union power started to grow at this time, and had some notable successes in Europe (e.g. the abolition of the maximum wage and the retain-and-transfer system in England; Szymanski and Kuypers 1999, ch. 4). However, in North America the role of the unions has grown significantly over the past forty years, while in Europe they remain relatively weak to this day.

5. *Team Sports, Uncertainty of Outcome, and Competitive Balance*

The justification for the striking range of restrictions utilized in the baseball system (fixed number of teams, exclusive territories, draft rules, roster limits, limited player trading, especially in relation to cash sales, collective selling of national broadcast rights and merchandising, restrictions on ownership) has been based on the nature of competitive team sports. The argument, which has formed the basis of numerous antitrust defences in the U.S. courts, can be reduced to three core claims:

1. Inequality of resources leads to unequal competition.
2. Fan interest declines when outcomes become less uncertain.
3. specific redistribution mechanisms produce more outcome uncertainty.

These propositions have defined both the empirical and theoretical research agenda of team sports economics. This section reviews the empirical literature on the first two of these propositions. Section 6 will consider individual measures to improve competitive balance.

5.1 *Inequality and the Sensitivity of Success to Resources*

The starting point for empirical analysis is that better players produce more success, and acquiring better players costs more money. In other words, we can substitute "cash" for "talent," and talent plays the same role as "effort" in the CSF. Implicit in this notion is a functioning labor market, notwithstanding any constraints upon initial endowments or trading rights within that market. Direct testing of this hypothesis is relatively sparse in the literature. One implicit test is contained in the literature on monopsonistic exploitation, following the methodology of Scully (1974). Even if players do not receive their full marginal revenue products, in an efficient market the rate of exploitation per unit of talent should be the same—otherwise an arbitrage opportunity exists. If the rate of exploitation is common across players, then at the level of the team, contest success should be closely correlated with player salaries.

Aggregate data for total player wage bill per team provides a more direct test of the hypothesis. Table 1 reports a simple regression of regular season winning percentage (wpc) upon team wage bill, expressed relative to the average of all teams' wage spending in the season (RW), for the four North American major leagues and the four leading soccer leagues in Europe. These results suggest a fairly close correlation between success and relative wage spending. Since the average of RW is unity, by construction the coefficients α and β must sum to 0.5 for a representative sample (i.e. average wpc). A larger estimate of β implies a larger pay-performance sensitivity. Thus the pay-performance sensitivity of the two baseball leagues is much smaller than that of the NFL. However, this does not make baseball more balanced, since the variance of relative wage spending is much greater. Moreover, the explanatory power of the regression, as measured by the R^2, is also larger, most notably in the American League (home of the Yankees).[38]

The apparent explanatory power of the regression for the European soccer leagues of England, Italy, Germany, and Spain is greater than for the North American leagues, even though the pay-performance sensitivity is not significantly larger. Given a much larger variation in wage payments, the same pay-performance sensitivity can account for much more of the variation of win percentages. In that sense European leagues appear more predictable. It is striking, given the widespread concern in the United States about growing imbalance in

[38] The degree of sensitivity reported here seems much greater than that reported by other authors (e.g. James Quirk and Mohamed El-Hodiri 1974; Fort and Quirk 1999); this may be in part a consequence of choice of specification and using a larger and longer panel of data. Zimbalist (1992) reports a similar R^2 for baseball and concludes that "average team salary has been related only tenuously to team performance."

TABLE 1
PAY-PERFORMANCE SENSITIVITY ESTIMATES

League	α	β	σ_{wpc}	σ^*_{wpc}	σ_{RW}	R^2	Period	Obs
North America								
Baseball NL	.42	.08	.07	.04	.27	.11	80–96	208
Baseball AL	.40	.10	.07	.04	.33	.26	80–96	238
NFL	.19	.31	.19	.13	.13	.05	89–00	350
NBA	.21	.29	.16	.06	.22	.16	86–00	351
NHL	.35	.15	.10	.06	.23	.11	90–98	218
European soccer								
Premier League (England)	.33	.19	.11	.08	.34	.34	74–99	339
Serie A (Italy)	.34	.15	.13	.11	.63	.56	88–99	214
Bundesliga (Germany)	.39	.12	.11	.09	.47	.28	82–96	244
Primera Liga (Spain)	.43	.07	.11	.08	.87	.32	97–02	111

Notes: Estimated equation: $wpc_{it} = \alpha + \beta\, RW_{it} + \varepsilon$. *RW* is wage bill of a team relative to average wage bill for the league in that year. All estimates significant at the 1% level. σ^*_{wpc} is the idealized standard deviation if teams had an equal chance of winning each match they played ($= .5/\sqrt{m}$, where m is the number of matches played by each team). European data refers to the top division only.

baseball, that the variation of wages and the R^2 of the regression are only noticeably larger in the American League compared to the other North American sports and even then these do not reach the levels found in the European leagues. In more recent years, however, there may have been a trend toward increasing predictability (see Stephen Hall, Szymanski, and Zimbalist 2002).

Correlation does not imply causation. An implicit assumption in the regression specification is that wages cause performance—but it could be argued that causality runs in the opposite direction, from performance to wages. For example, it is usual for winning teams to be paid bonuses, and it is sometimes said that team owners would rather come second than win a championship in order to avoid excessive bonus payments (an example of the limited role of prizes in rewarding team, as opposed to player, performance).

Testing for the direction of causality is feasible. Hall, Szymanski, and Zimbalist (2002) tested for Granger causality from wages to performance and from performance to wages, and found that they could reject the latter direction of causality for English soccer but not for major league baseball (MLB). One interpretation of this result is that in English soccer there is an unrestrained market for players so that there is no barrier to the operation of an efficient market (for details of its operation, see Szymanski and Kuypers 1999). In MLB, player contracts are much more restrictive, both for players and owners, and this gives rise to bargaining over team rents, the outcome of which is likely to depend on past performance. Testing this hypothesis, which requires the collection of a wider range of potential explanatory variables for MLB, is an important subject for future research, as is the nature of causality in other leagues.[39]

One feature of table 1 that might strike a North American reader is the combination of relatively low standard deviation of winning percentages, often considered an indicator of competitive balance, in the European leagues, combined with relatively high standard deviation of wage payments (see also Ingo Kipker 2000, and David

[39] There have been relatively few attempts to analyze causality empirically in the sports literature. Brian Davies, Paul Downward, and Ian Jackson (1995) and Stephen Dobson and John Goddard (1998) look at the relationship between income variables (attendance and revenues) and success in English rugby league and soccer.

Forrest and Robert Simmons 2002b for detailed comparison). Given a reasonable degree of sensitivity of performance to wages (which does appear causal, at least in the English case) one might have expected a relatively high standard deviation of win percentage reflecting a high degree of competitive imbalance.[40]

The standard deviation of winning percentage, however, may be a relatively poor measure of competitive balance, largely because it only considers performance within a season. Performance in the open European leagues tends to be relatively bunched together, since teams near the bottom keep competing right to the end in order to avoid relegation (see section 7.5). Yet over a number of seasons the same big teams tend to dominate European competition, so there is little turnover at the top. Relatively little attention has been paid to measuring this notion of competitive balance, although this is clearly the aspect that figured heavily in the Blue Ribbon Panel's investigation into baseball (notably the dominance of the Yankees) and has been raised by some critics of static measures (e.g., Ross and Robert Lucke 1997; Woodrow Eckard 1998).[41]

[40] A number of authors have used the standard deviation of winning percentage relative to the idealized standard deviation (assuming winning probabilities) as an alternative measure (see e.g. Scully 1989; Quirk and Fort 1992; Vrooman 1995). Other static measures include the Gini coefficient (Quirk and Fort 1992), relative entropy (Ira Horowitz 1997), and the Hirschman-Herfindahl index (Craig Depken 1999).

[41] Eckard (1998) proposes a decomposition of the variance of winning percentages into a cumulative and time-varying component. For a given total variation a decrease in the variation through time implies greater cumulative variation; in other words, from season to season there is less turnover in team standings (competitive imbalance). Brad Humphreys (2002) proposes a similar measure. Alan Balfour and Philip Porter (1991) and Vrooman (1996) have estimated first-order autoregressive processes for win percent as a way to search for possible structural breaks associated with free agency (see below). In other words, they consider the degree of persistence, which might be thought a natural measure of dynamic competitive balance. Szymanski and Ron Smith (2002) adopt this approach to compare persistence across North American and European leagues.

Luigi Buzzacchi, Szymanski, and Valletti (2003) develop a dynamic measure based on estimating the number of teams entering the top k ranks of a league competition over T years (they look at the top rank and the top five ranks over the ten-year intervals from ten to fifty) relative to the idealized number of teams that would have entered these ranks under an equally balanced contest. Note that in an open system where the probability of success is identical for each team, there will be a very high turnover at the top over a twenty-year period, since so many more teams have access compared to a closed league. They compare three North American leagues (MLB, NFL, and NHL) with three national soccer leagues (Italy, England, and Belgium) and find that the number of entrants to the top ranks is slightly higher in North America, but that relative to potential entrants the number of actual entrants is very small in the European leagues. They suggest that an open system can be characterized as one that produces equality of opportunity, while closed leagues are more successful at producing equality of outcome. More research is required into the causes of these differences.

5.2 *Demand and Uncertainty of Outcome*

Whatever the causes of inequality, the lynchpin of team sports organizers' defense of restrictive agreements has been the claim that such measures are required to combat the threat of uneven contests that will reduce the interest of the fans. This proposition was first fully enunciated in the economics literature in a celebrated paper by Neale (1964). As a testable hypothesis it has now generated a substantial literature of its own. To begin with, it is useful to differentiate three types of uncertainty:

1. match uncertainty,
2. seasonal uncertainty,
3. championship uncertainty.

The meaning of match uncertainty is obvious. Seasonal uncertainty means a close championship race within a season, while

championship uncertainty means there is a variety of champions over a period of years, rather than domination by one or two teams.

Table 2 summarizes the research in this area. In recent years, research on match uncertainty has focused on the use of pre-match betting odds as a means of measuring uncertainty. There seems to be an emerging consensus that demand for match tickets peaks at the point where a home team's probability of winning is about twice that of the visiting team, i.e., a probability of around 0.66. (See e.g. Glenn Knowles, Keith Sherony, and Michael Haupert 1992; Forrest and Simmons 2002a summarizing the work of David Peel and David Thomas (1988, 1992, 1997), and Dan Rascher 1999). Several reviewers have commented upon just how *unbalanced* a contest characterized by this probability would be, and in most datasets there are relatively few observations involving such extremely unbalanced contests. Whether this imbalance is optimal from the point of view of the league is not something that these studies address, but it seems reasonable that the optimal balance for the league may be greater than that for the home team.

Less work has been done on the issue of seasonal uncertainty. The key problem in this area is controlling for all the other relevant factors that might influence demand. For example, Martin Schmidt and David Berri (2001) find that attendance is positively affected by uncertainty, using nearly a century of MLB data, but with no other explanatory variables. When they examine a shorter panel including influences such as price data, they find that, for the National League, attendance is significantly *decreasing* in uncertainty. While it is plausible that fans prefer a close championship race, a run of success by a single team may itself spark interest (like the old barnstorming teams). It may be that the causal relationships are too complex to isolate a single influence such as uncertainty of outcome.

Finally, championship uncertainty has hardly ever been tested, although the evidence comparing the relative long-run imbalance of European soccer to the North American leagues suggests that this is an issue worthy of investigation.[42] On the face of it, European soccer is every bit as popular with Europeans as the North American leagues are with Americans, despite long-run domination by a much smaller subset of teams.

Overall, of the 22 cases cited here, ten offer clear support for the uncertainty of outcome hypothesis, seven offer weak support, and five contradict it. Given that even supportive studies on the issue of match uncertainty seem to imply that attendance is maximized when the home team is about twice as likely to win as the visiting team, the empirical evidence in this area seems far from unambiguous. This is remarkable given the weight that is placed on this argument in policy making and in antitrust cases. Given that even quite unbalanced matches, championships, and leagues can be attractive to consumers, a more nuanced approach is called for.[43]

6. *The Invariance Principle*

In this section we turn to the consideration of specific rules and restrictions that might be designed to increase uncertainty of outcome and enhance competitive balance. Because of the cartel-like organizational structure of most team sports leagues, these rules and restrictions have often been debated in the antitrust courts. On the one hand,

[42] One exception is Szymanski (2001), who exploits the fact that, in soccer, teams participate in two national competitions at once, one of which contains a much less balanced selection of contestants than the other. By pairing the subset of matches in each tournament that involve the same teams he is able to infer the effect of the balance of each tournament taken as a whole.

[43] Paul Downward and Alistair Dawson (2000) reach a similar conclusion: "the evidence suggests that uncertainty of outcome has been an overworked hypothesis in explaining the demand for professional sports."

TABLE 2
OUTCOME UNCERTAINTY IN THE LITERATURE

Authors	Testing	Uncertainty measure	Data	Result
Noll (1974)	seasonal	- whether team in contention for playoff	ice hockey	weak support
		- whether championship race close	baseball	weak support
Hart et al. (1975)	match	- log difference in league positions	4 English football clubs 1969/70–1970/71	weak support
Jennett (1984)	seasonal	- championship/relegation significance of each game	Scottish League Football 1975–81	Support
Borland (1987)	seasonal	-diff. in games won between first and last	Victorian Football League (Australian Rules) 1950–86	weak support
		-sum of coefficients of variation of game won		
		-average no of games behind the leader		
	championship	-number of teams in contention		no support
Cairns (1987)	seasonal	dummy of contention in championship	4 Scottish football clubs 1969/70–1979/80	support
Jones and Ferguson(1988)	match	dummy for top of the table and bottom of the table matches	NHL Season 1977/78	no support
Whitney (1988)	seasonal	average expected probability of winning	baseball 1970–84	weak support
Peel and Thomas (1988)	match	betting odds (probability of home win)	1981/82 English football league matches	weak support
Knowles et al. (1992)	match	betting odds (probability of home win)	MLB 1988	support
Peel and Thomas (1992)	match	betting odds (probability of home win)	English Football League matches	weak support
Borland and Lye (1992)	seasonal	sum of matches required to qualify for the finals	Australian Rules	no support
Kuypers (1996)	match	betting odds (difference in max and min)	1993/94 individual English Premier League matches	no support
	seasonal	points and games left		support
Peel and Thomas (1997)	match	betting odds (points spread)	Rugby League 1994/95	support

TABLE 2 *(cont.)*

Authors	Testing	Uncertainty measure	Data	Result
Baimbridge et al. (1996)	seasonal	dummy when both teams in top (bottom) four positions	1993/94 Individual English Premier League Matches	no support
Rascher (1999)	match	betting odds (probability of home win)	MLB 1996	support
Szymanski (2001)	championship	competition type (with identical contestants)	English League and FA Cup matches 1977–98	support
Schmidt and Berri (2001)	seasonal	gini coefficient	MLB 1903–98 (Gini only) MLB 1975–88 (Gini plus other variables)	support support for AL no support for NL
Forrest and Simmons (2002a)	match	odds ratio (accounting for favorite-longshot bias)	Football League matches 1997/98	support

Source: Adapted from Szymanski and Kuypers (1999).

economists can try to shed light on whether specific restrictions achieve their stated aim (and whether they were strictly necessary to achieve it); on the other hand, they can also identify other consequences arising from a given restriction. These may be consequences for profits (the owners' interest); prices, quality, and choice (the consumers' interests); and employment conditions and remuneration (the players' interests). Economic analysis of these issues is usually both theoretical and empirical, and the balance between the two often depends on the nature of the restriction and the availability of data.

6.1 *The Invariance Principle and Talent Allocation Rules*

One common characteristic of team sports as they developed on both sides of the Atlantic has been the desire of the owners of teams belonging to professional leagues to control the market for players, in particular to establish monopsony rights. Thus the reserve clause of baseball (see e.g. Quirk and Fort 1992 for an explanation) functioned in much the same way as the Retain and Transfer System of English soccer (see e.g. Sloane 1969).[44] This inevitably led to challenges in the courts by the players claiming the right to move freely between employers. Simon Rottenberg's celebrated (1956) article examined this issue and presented the team owner's rationale:

> "The defense most commonly heard is that the reserve rule is necessary to assure an equal distribution of playing talent among opposing teams; that a more or less equal distribution of talent is necessary if there is to be uncertainty of outcome; and that uncertainty of outcome is necessary if the consumer is to be willing to pay admission to the

[44] In fact, the two systems were so similar that it is hard to believe that the Football League did not copy the National League. However, no evidence to this effect has ever been produced.

> game. This defense is founded on the premise that there are rich baseball clubs and poor ones and that, if the players' market were free, the rich clubs would outbid the poor for talent, taking all competent players for themselves and leaving only the incompetent for other teams." (p. 246)

Rottenberg argued that (a) the reserve clause did nothing to prevent the migration of talent to the big city teams and so would not affect the distribution of talent, and that (b) by establishing monopsony power over a player throughout his career the team owners were able to hold down wages and raise profitability. Point (a) has since been identified as an example of the Coase Theorem at work: the initial distribution of ownership rights should have no impact on the efficient (here profit-maximizing) distribution of resources. El-Hodiri and Quirk (1971) and Quirk and El-Hodiri (1974) took this analysis one stage further in a formal dynamic model showing that, if teams have differing revenue generating potential, (i) profit maximizing behavior will not lead to an equal distribution of resources (playing talent) and (ii) revenue redistribution on the basis of gate sharing will have no impact on the distribution of playing talent. Points (a) and (ii) are both examples of the well-known *invariance principle*.

There have been two significant changes in talent-allocation rules in North American sports over recent years. Firstly, in 1976, major-league baseball players won the right of free agency after completing six years of service, and this practice rapidly spread to the other sports. Secondly, the draft rules of the NFL, which allocated the right to hire new talent entering the league on the basis of the reverse order of finish of the previous season's competition, were adopted by the other sports (see Paul Staudohar 1996 for more details on both of these innovations). These changes can be studied to identify the impact of changes in talent allocation rules on competitive balance.

Free Agency. The advent of free agency in MLB in 1976 for six-year veterans is a clear natural experiment.[45] The owners claimed that as a result of this limited free agency, the best veterans would migrate to the big city teams and competitive balance would be undermined. A number of studies have attempted to use this rule change to test the invariance hypothesis, and the findings from these studies are reported in table 3. Most of the studies simply look at the standard deviation of win percentages before and after 1976 (Scully 1989; Balfour and Porter 1991; Quirk and Fort 1995; Vrooman 1995; Michael Butler 1995), while other measures include persistence in win percent (Balfour and Porter 1991; Vrooman 1996); entropy (Horowitz 1997); the Hirschman-Herfindahl index (Depken 1999); and analysis of variance (Eckard 2001). Most of these studies find either no change (seven cases) or an improvement in competitive balance (nine cases), contrary to the claim of the owners that free agency would reduce competitive balance (four cases only). However, this meta-data is hardly a ringing endorsement for the invariance principle, since "no effect" is reported in only seven out of twenty cases. Of course, it can be argued that many other factors have altered competitive balance (e.g. the increasing dispersion of local TV revenues), but in that case the data, without controlling for these factors, can hardly be said to represent a test at all.

Some other studies have approached the invariance principle as a direct test of the Coase Theorem and tried to establish whether the distribution of talent in the league has been affected by the introduction of free agency. George Daly (1992) observes that under the reserve clause, top line players were seldom traded, a situation that has been affected by free agency, where after six years the top stars have a choice, leading to increased mobility. Timothy Hylan,

[45] In this case the change was exogenous—i.e. not itself motivated by a desire to affect competitive balance (see Bruce Meyer 1995 for a discussion of natural experiments).

TABLE 3
THE IMPACT OF FREE AGENCY ON COMPETITIVE BALANCE IN MLB

Study	Measure of Competitive Balance	Impact on Competitive Balance in NL	Impact on Competitive Balance in AL
Daly and Moore (1981)	Movement of free agents to large market teams	(-)	(-)
Scully (1989)	Standard deviation of win percent and Gini coefficient of pennant wins	(+)	(0)
Balfour and Porter (1991)	Standard deviation of win percent, persistence of win percent	(+)	(+)
Fort and Quirk (1995)	Standard deviation of win percent and Gini coefficient of pennant wins	(0)	(0)
Vrooman (1995)	Standard deviation of win percent relative to idealized standard deviation	(+)	(+)
Vrooman (1996)	Persistence of win percent	(+)	(+)
Butler (1995)	Standard deviation of win percent and serial correlation of win percent	(0)	(0)
Horowitz (1997)	Entropy	(-)	(0)
Depken (1999)	Hirschman-Herfindahl index of wins relative to ideal	(0)	(-)
Eckard (2001)	Analysis of variance of win percent	(+)	(+)

Maureen Lage, and Michael Treglia (1996) in a study of pitcher movements finds that these players have become less mobile since free agency, a surprising result and one that they claim does not support the Coase Theorem. However, Donald Cymrot, James Dunley, and William Even (2001) examine player mobility in 1980, controlling for possible selection bias and find that, for that season at least, there was no evidence that restricted players (with less than six years of service) enjoyed more or less mobility than unrestricted free agents after controlling for player characteristics.

Daniel Marburger (2002) considers a different implication of the invariance principle. If trade is possible between two independent leagues then it should be more profitable to hire a player from the same league than the rival league. Intra-league trade raises the winning probability of the buying team by more than an inter-league trade, since in the former case not only does the buyer have a larger share of talent, but the seller now has a weaker team. Under the reserve clause this effect will be built into the seller's price, but under free agency it will not, since the free agent is indifferent to the adverse effect on the team he is leaving. Thus with free agency the relative price of intraleague trades should fall and their share of total trades increase. Marburger found a statistically significant increase in the share of intraleague trades, from 60

percent to 73 percent, in MLB 1964 and 1992. This finding seems consistent with the invariance principle.

In European soccer, trading players for cash has always been an accepted part of the sport, and there have been no restrictions on trading such as those that emerged in North America in 1970s (see Daly 1992). In England a system akin to the reserve clause operated until 1963. Restrictions remained until 1978, when a form of free agency was introduced that gave players the right to move club once their contract ended (typically contracts lasted three years), but allowed the selling team to demand substantial compensation (i.e. well in excess of any damages that would be paid for breach of contract). In 1995 the European Court of Justice, in what is known as the Bosman judgment,[46] outlawed all such compensation payments for out-of-contract players and effectively established universal free agency. In 2001 FIFA reached agreement with the European Commission on a new set of transfer rules. These laid down that compensation was only payable to clubs for players under the age of twenty-three and only as a reflection of training costs. Beyond that age no transfer fee is to be paid for players out of contract and players can move clubs during one of two prescribed "transfer windows."

The Rookie Draft. The stated intention of the rookie draft system is to provide weaker teams with opportunities to acquire talented players by awarding them first pick. Of course, an additional consequence of this system is the creation of monopsony power. The draft system was instituted by the NFL in 1936 as a way of strengthening weak performing teams to maintain competitive balance, and has since been adopted by all the other major leagues (Fort and Quirk 1995, and Staudohar 1996 provide details).

Daly and Moore (1981) first analyzed whether the draft achieved its stated intention by examining competitive balance before and after the introduction of the MLB draft in 1965. They found a significant improvement in the balance of the National League and a smaller improvement in the balance of the American League. The Japanese Professional Baseball League adopted a draft system at exactly the same time as MLB, and a study by La Croix and Kawaura (1999) also found that competitive balance improved over time (measured by the Gini coefficient for pennants) in both the Central and Pacific Leagues.[47] As they point out, these results are "virtually identical" to Fort and Quirk's (1995) results for MLB. Kevin Grier and Robert Tollison (1994) examined the impact of the rookie draft in the NFL by running an autoregressive specification for win percentage together with the average draft order over the previous three to five seasons, and found that a low draft order significantly raises performance. These results seem to provide consistent evidence against the invariance principle and in support of the owners' stated position.

Neither with free agency nor with the rookie draft is there much convincing evidence on profits and consumer welfare. It is clear that free agency has increased the earning power of free agents, but it is not clear what the distributional effects have been on the player market as a whole. For example, it might be that increased expenditure on free agents caused by competition for their services has led to a reduction of investment in the development of rookie talent or lower salaries on average for players with less than six years service. Zimbalist

[46] Bosman was a Belgian playing for a Belgian team who refused a new contract and decided he wanted to transfer to a French club, that was willing to hire him and pay a transfer fee. Under the rules of the Belgian Football Association, the Belgian club had the right to veto the transaction without appeal (and so retain Bosman's services), which it did, on the grounds that it thought the buying club could not really afford the fee. This system was outlawed by the judgment (Court of Justice of the European Communities, Case C-415/93).

[47] However, the within-season measure (standard deviation of win percent) was significant only for the Pacific League.

(1992) reports significant differences in the rate of monopsonistic exploitation for players at different stages of their careers after the introduction of free agency. In Europe, where there are no roster limits, it does appear that the number of professional soccer players has been falling over time, and this could be associated with the trend toward free agency that was visible in England even before the Bosman judgement (i.e., teams substituting quality for quantity). Eberhard Fees and Gerd Muehlheusser (2002) compare the welfare implications of the pre- and post-Bosman transfer regimes and argue that while the new regime may increase player effort (since they can secure a larger share of the returns) investment in player development is likely to fall. These issues deserve empirical investigation.

6.2 *The Invariance Principle and Gate-Revenue Sharing*

El-Hodiri and Quirk (1974) extended the invariance principle to gate revenue sharing, i.e. they claimed that a change in the percentage of gate revenues allocated to the visiting team (between 100 percent and 50 percent) would have no affect on competitive balance. Empirical testing of this proposition is made difficult by the fact that revenue-sharing rules change infrequently within a single league, while the comparison of revenue sharing across different leagues is clouded by the interference of so many other league-specific factors. An alternative approach is to examine the theoretical basis for this proposition. This section develops a simple contest model that illustrates the basis of the invariance principle for gate-revenue sharing.

The conventional approach to the modeling of league competition (as in e.g. Fort and Quirk 1995) is to some extent supported by the empirical evidence in section 5. Firstly, it is normally assumed that teams choose investment in playing talent that is homogeneous and perfectly divisible, so that a given level of investment translates into a predictable level of playing success. Secondly, it is assumed that excessive dominance by one team will lead to a fall in revenue generation by that team, although at low levels of success revenues are increasing in team performance. The main difference between the team sports model and a conventional contest model is that instead of competing for a fixed prize with some probability determined by relative investment, each team generates a revenue dependent on the share of matches won, where that revenue also varies according to the revenue generating capacities of the teams. Thus asymmetry in team sports is not typically modeled as a difference in the cost of effort (talent investment), but as a difference in the value of the prize (revenue generating capacity).

The nature of the prize in team sports is somewhat different than in an individualistic contest. Success is usually equated with winning percentage, which in turn depends on the outcome of a sequence of bilateral contests. However, what distinguishes league competition from the kind of barnstorming match play observed prior to the creation of the National League is that fan interest is drawn to the progress of their team in the tournament as a whole, not just the individual matches. In other words, there is also a prize for success over the competition as a whole (the league championship) rather than simply collecting income from a series of events.

A further modeling issue concerns the way that decision makers interact. Fort and Quirk, among others, support the cartel interpretation, suggesting that clubs make independent decisions subject to cartel rules (i.e. a noncooperative game), and we follow this approach below.[48] The precise legal format adopted, however, may vary. Conventionally, teams are joint owners of the league and delegate an official to manage collective negotiations.

[48] Some maintain that leagues should be considered (at least for antitrust purposes) as single economic entities (e.g. Gary Roberts 1984), which could imply centralized decision making.

Analysis of the invariance principle is only relevant when there are asymmetries among the teams. If teams are symmetric, competitive balance cannot be an issue if, as here, we concentrate only on pure strategy equilibria. To concentrate on asymmetry we narrow our focus to a two-team model, as has been usual in most of the literature. Assuming the CSF takes the same logit form as in an individualistic contest (3) and that $\gamma=1$, we can write

$$p_1 = \frac{e_1}{e_1 + e_2}, \quad p_2 = 1 - p_1 \qquad (3')$$

where p_i can be thought of as the expected percentage of matches won by team i, which is increasing in the relative share of investment in talent, which is how e_i is now interpreted.[49] In a standard contest model the "adding-up constraint" requires that the probabilities sum to unity, while in a league context the constraint is that the sum of win percentages equals $n/2$. Obviously this condition is satisfied by (3'). Another way of expressing the adding-up constraint is

$$\frac{\partial p_1}{\partial e_1} = -\frac{\partial p_2}{\partial e_1} \quad and \quad \frac{\partial p_2}{\partial e_2} = -\frac{\partial p_1}{\partial e_2}. \qquad (11)$$

Note that the CSF (3) is identical to win percentage for a two-team model, but not with three or more teams, since expected win percentage then depends on the sum of bilateral investment shares (3') rather than simply investment divided by the sum of investments. Both functions will be increasing and concave in investment, and bounded by zero when investment is zero.

It is sometimes argued that a two-team model fails to capture some central features of a league championship. If $n > 2$ it is possible to specify each team's revenue function as a function of rival teams' win percentage, introducing the possibility of complementarities. Although this suggests a more complex set of interactions than is modeled here, the existence of production externalities (the success of my team increases or decreases your team's revenues) does not fundamentally change the decision problem, since even in the two-team case each team's investment produces a negative externality (my success reduces your income). The important economic issue is that private decision making will not necessarily be socially efficient when externalities, negative or positive, exist.[50]

In general, demand for attendance at or viewing of matches could be thought to depend on three main factors:

- the suspense associated with a close contest (uncertainty of outcome);
- the likelihood of the home team's success;
- the quality of the match, including the aggregate of player talent on show.[51]

The interaction of these three factors will give rise to some general revenue-generating function $R(.)$. The requirement of tractability demands some simplification and so for the moment we will ignore the impact of the

[49] Baik (1994) models asymmetry by assuming that the sensitivity of CSF to effort differs among contestants, an assumption that implies that all teams do not have access to the same technology for transforming talent into success. The assumption of symmetry effectively implies that all teams adopt best practice. The literature on team production functions sheds some light on this issue (see fn. 8).

[50] For $n > 3$ the CSF can be thought of as a championship success function (e.g. James Whitney 1988) In practice, the difference between the share of total matches won in a season and win percentage is small and the two measures are highly correlated. For example, in English soccer the correlation coefficient between league rank and win percent is about 0.9.

[51] Following most of the literature, we abstract from price issues. In North America, monopoly pricing is plausible due to distance and territorial exclusivity (see e.g. Donald Alexander 2001). Greater urban density and the promotion and relegation system in Europe make this less likely. For example, New York has two major-league baseball teams (population 20 million) while London (13 million) hosts six teams currently in the top division of English soccer, plus another six eligible to enter if promoted on merit. In Australian Rules Football and Australian Rugby League most of the teams are located around a single city (Melbourne and Sydney respectively). The implications of population density for revenue generation remain to be explored (but see Forrest et al. 2002 on the spatial pattern of demand for English soccer).

demand for quality.[52] We therefore focus on the impact of success and competitive balance probabilities. In most of the literature these two aspects of demand are captured by a revenue function that comprises a CSF and the assumption that team revenues have a unique maximum (e.g., at a winning record that lies between 0 percent and 100 percent). Here we assume that revenues are simple linear functions of these variables:

$$R_{11}=[1-\lambda(1-\mu)]\,p_1-(1-\lambda)\,p_1^2=\lambda\mu\,p_1+(1-\lambda)\,p_1(1-p_1) \quad (12)$$

$$R_{22}=p_2-(1-\lambda)\,p_2^2=\lambda p_2+(1-\lambda)\,p_2(1-p_2)$$

where R_{ii} is either the revenue generated by team i from matches played at the ground of team i or the revenue generated by championship success. $\mu \geq 1$ reflects the possibility that team 1 may be able to generate a higher revenue from a given level of success. Competitive balance can be measured by $p_1(1-p_1)=p_2(1-p_2)$ and λ is a parameter capturing the degree to which competitive balance matters in determining team revenues; if $\lambda=1$ only winning matters, while if $\lambda=0$ interest in a balanced contest dominates. Each firm's profit function is simply $\pi_1=R_{11}-ce_1$ and $\pi_2=R_{22}-ce_2$, where c is the constant marginal cost of talent, which is treated parametrically by the teams, but adjusts to ensure that the supply of talent equals demand. Note that if $\lambda=\mu=1$ the problem is isomorphic to the symmetric winner-take-all contest of section 2.1.[53] The owners of each team are assumed to be profit maximizers. Under these assumptions the first-order conditions are:

$$\frac{d\pi_1}{de_1}=\frac{\partial R_{11}}{\partial p_1}\frac{\partial p_1}{\partial e_1}-c=[1-\lambda(1-\mu)-2(1-\lambda)p_1]\frac{\partial p_1}{\partial e_1}-c=0 \quad (13)$$

$$\frac{d\pi_2}{de_2}=\frac{\partial R_{22}}{\partial p_2}\frac{\partial p_2}{\partial e_2}-c=[1-2(1-\lambda)p_2]\frac{\partial p_2}{\partial e_2}-c=0$$

These expressions state that owners invest in talent to the point where the marginal revenue from a unit of talent equals its marginal cost. For example, for team 1 the marginal revenue of a unit of talent equals the marginal revenue of a win $(1-\lambda(1-\mu)-2(1-\lambda)\,p_1)$ multiplied by the marginal impact on win percentage of a unit of talent $(\partial p_1/\partial e_1)$.

The standard assumption in the North American team sports literature has been that this latter quantity is equal to unity. Thus Fort and Quirk (1995, p. 1271) assume "a one unit increase in t_i yields the same increase in win-percent for any level of win-percent" and Vrooman (1995, p. 973) uses a model where teams directly choose win percent, whose marginal cost is assumed to be a constant, so that a unit of talent in the present model is equivalent to a unit of win percentage. Given identical marginal costs this implies that the marginal revenue of a win is equalized across teams. This seemingly innocuous assumption has important implications about the behavior of owners. From (3')

$$\frac{\partial p_1}{\partial e_1}=\frac{(e_1+e_2)-e_1\left(1+\frac{de_2}{de_1}\right)}{(e_1+e_2)^2},\qquad \frac{\partial p_2}{\partial e_2}=\frac{(e_1+e_2)-e_2\left(1+\frac{de_1}{de_2}\right)}{(e_1+e_2)^2}. \quad (14)$$

[52] Intuitively, if this enters the revenue function of each team symmetrically then it will shift out the demand for talent. Some consequences of including the interaction of quality in more complex cases are considered below.

[53] In the one-shot winner-take-all model, the payoff to the contestant is an expectation of the prize dependent on relative effort but only one contestant receives the prize ex post, while in the one-shot team sports version each contestant generates an income based on the share of success so that expected income equals ex post income (there is no stochastic element in the CSF). In an infinitely repeated game with no discounting, the value of the expected and actual payoffs are identical in both cases.

If we assume $de_1/de_2 = de_2/de_1 = -1$ then, normalizing the total supply of talent to unity, it will indeed be the case that $\partial p_1/\partial e_1 = \partial p_2/\partial e_2 = 1$. It should be obvious that this assumption is not the same as adding-up constraint (11). Since the expression (14) appears in the objective function of the teams de_2/de_1 is a conjectural variation, i.e. the expectation of team 1 (resp. 2) of the response of team 2 (resp. 1) to a unit increase in talent by team 1 (resp. 2). If we assume that this conjecture equal -1, then each team is assumed to suppose that whenever they increase their investment in talent by one unit, their rival will decrease their investment in talent by one unit.

The rationale for this assumption is that the total supply of talent is fixed, which is often thought a distinctive feature of the major leagues. It is probably true that all the best baseball players, wherever they are in the world, would prefer to play in MLB, and that all the best basketball players in the world would prefer to play in the NBA and so on. If the talent supply for each league is fixed (at least in the short term) then if one team hires an additional unit of talent there is one less unit for all other teams to hire.[54] But modeling a fixed talent supply by assuming non-zero conjectural variations has significant implications for the nature of the model's equilibrium. The normal approach to identifying a Nash equilibrium is to assume Nash conjectures, namely $de_1/de_2 = de_2/de_1 = 0$. Without Nash conjectures peculiar results may follow.

To see the implications of this, combine the two expressions in (13) to obtain

$$\frac{\dfrac{\partial p_2}{\partial e_2}}{\dfrac{\partial p_1}{\partial e_1}} = \frac{\dfrac{\partial R_{11}}{\partial p_1}}{\dfrac{\partial R_{22}}{\partial p_2}} = \frac{1-\lambda(1-\mu)-2(1-\lambda)p_1}{1-2(1-\lambda)(1-p_1)}. \tag{15}$$

[54] This assumes the supply is not so great that the demand curve intersects the horizontal axis at a point to the left of the fixed supply, implying that there is more talent than MLB or the NBA require.

Note that the left-hand side of (15) is the ratio of the marginal impacts on win percentage of a unit of talent and the right-hand side is the ratio of marginal revenues of a win. Under the "fixed supply conjectural variation" the LHS is unity and so the marginal revenue of a win is equalized across teams. This is not true using the Nash conjectural variation, where it is only the marginal revenue from hiring a unit of talent that is always equalized in equilibrium, while the marginal revenue of a win will only be equalized at the equilibrium of a symmetric contest ($\mu = 1$).[55] At the asymmetric Nash equilibrium the marginal revenue of a win will be greater for the strong drawing team ($\mu > 1$) because this team hires a larger share of talent available and therefore has a lower marginal impact on win percentage from an extra unit of talent.

Nash conjectures and fixed-supply conjectures produce very different results when it comes to the impact of gate-revenue sharing. In the standard model it is assumed that each team retains a fraction α of revenues generated by home matches and pays the remainder $1-\alpha$ to the visiting team so that profits are now $\pi_1 = \alpha R_{11} + (1-\alpha)R_{22} - ce_1$ and $\pi_2 = \alpha R_{22} + (1-\alpha)R_{11} - ce_2$ and the first-order conditions are

$$\begin{aligned} \frac{\partial \pi_1}{\partial e_1} &= \alpha \frac{\partial R_{11}}{\partial p_1}\frac{\partial p_1}{\partial e_1} + (1-\alpha)\frac{\partial R_{22}}{\partial p_2}\frac{\partial p_2}{\partial e_1} - c = 0 \\ \frac{\partial \pi_2}{\partial e_2} &= \alpha \frac{\partial R_{22}}{\partial p_2}\frac{\partial p_2}{\partial e_2} + (1-\alpha)\frac{\partial R_{11}}{\partial p_1}\frac{\partial p_1}{\partial e_2} - c = 0 \end{aligned}, \tag{16}$$

which, using the adding-up constraint (11) can be rearranged to obtain

$$\left[\alpha \frac{\partial R_{11}}{\partial p_1} - (1-\alpha)\frac{\partial R_{22}}{\partial p_2}\right]\frac{\partial p_1}{\partial e_1} = \left[\alpha \frac{\partial R_{22}}{\partial p_2} - (1-\alpha)\frac{\partial R_{11}}{\partial p_1}\right]\frac{\partial p_2}{\partial e_2}. \tag{17}$$

If we now further assume fixed-supply conjectures it should be clear that since

[55] With Nash conjectures the LHS of (16) equals e_1/e_2.

$\partial p_1/\partial e_1 = \partial p_2/\partial e_2$ (17) collapses to the equality

$$\frac{\partial R_{11}}{\partial p_1} = \frac{\partial R_{22}}{\partial p_2}, \qquad (18)$$

which is clearly independent of α, hence the conclusion that the distribution of talent and success is independent of the revenue-sharing formula. However, once we introduce Nash conjectures this result will no longer hold, and instead we obtain

$$\frac{\frac{\partial p_2}{\partial e_2}}{\frac{\partial p_1}{\partial e_1}} = \frac{e_1}{e_2} = \frac{\left[\alpha \frac{\partial R_{11}}{\partial p_1} - (1-\alpha)\frac{\partial R_{22}}{\partial p_2}\right]}{\left[\alpha \frac{\partial R_{22}}{\partial p_2} + (1-\alpha)\frac{\partial R_{11}}{\partial p_1}\right]}. \qquad (15')$$

It should be clear that the LHS of (15') is identical to that of (15), but when $\alpha < 1$ the RHS of (15') and the middle term of (15) are not equal (unless revenue functions are symmetric), suggesting that the invariance principle does not hold under Nash conjectures. Using the expressions for marginal revenue in (15) after some manipulation it can be shown that

$$\frac{e_1}{e_2} = \frac{(1-2\alpha)(1-\lambda) + \lambda\left[\alpha + \mu(1-\alpha)\right]}{1-2\alpha+\alpha\lambda(1+\mu)}. \qquad (19)$$

Differentiating, we obtain

$$\frac{\partial\left(\frac{e_1}{e_2}\right)}{\partial\alpha} = \frac{\lambda^2(1-\mu^2)}{\left[1-2\alpha+\alpha\lambda(1+\mu)\right]^2} < 0. \qquad (20)$$

Thus under Nash conjectures, revenue sharing will in fact make competitive balance worse. Szymanski and Késenne (2004) show that this is in fact true for any concave revenue function. The intuition is that revenue sharing discourages both teams from investing, but since the weak drawing team has more to gain from a share of the strong drawing team's revenues than the strong drawing team does from a share of the weak drawing team's revenues, the weak drawing team cuts investment by more.[56]

Because revenue sharing diminishes the incentive of both teams to invest in talent, the demand for talent must fall. If the supply of talent is fixed then the wage rate per unit of talent (i.e. the marginal cost c) will fall to restore labor-market equilibrium. However, if competitive balance is to deteriorate then it must be that the strong drawing team will in fact increase its share of total talent while the weak drawing team reduces its share. If the supply of talent were elastic, however, this result need not necessarily hold, even though competitive balance must still be reduced. The assumption of elastic supply seems more reasonable in the case of European soccer where no national league is dominant and players move freely between leagues. Whether supply is fixed or not, total expenditure on talent will fall with gate-revenue sharing and total profits will increase.

There is a fundamental problem with the assumption of fixed-supply conjectures. If teams attempt to select win percentage, only one team can be decisive, since the other team's choice is thereby fixed in a two-team model. It is like a model of market share where each firm tries to choose market share—at most one firm can succeed. More generally, in an n team model with fixed supply conjectures only $n-1$ teams can be decisive, and the n^{th} team must accept the allocation of talent implied by the profit maximizing choices of all the other teams. In the two-team model with fixed-supply conjectures, every choice of winning percentage is a Nash equilibrium, since there is only one feasible response to this choice and so it is trivially the best response (see Szymanski 2004 for more detail). The way around this absurdity is to

[56] Scott Atkinson, Linda Stanley, and John Tschirhart (1988) also state that they do not obtain the invariance result (p. 33, fn. 14) but attribute this to the assumption of a more general revenue function. The key difference, however, is that they do not assume fixed-supply conjectures.

allow owners to select some variable that affects the share of total talent, such as investment, without constraining the choice of rivals by so doing. This approach will result in the Nash equilibrium described above.

It seems widely accepted in the broader economic literature that, in a static game of this type, only Nash conjectures make sense (see e.g. Xavier Vives 1999, pp 185–87) but alternative conjectural variations are sometimes defended as reduced forms of an underlying dynamic model. The original model of Quirk and El-Hodiri (1974) is indeed a dynamic model. The authors do not explain in detail the source of the invariance result but it appears to be a consequence of looking for an equilibrium where not only the profit of each team is maximized with respect to talent hired at that team, but also with respect to talent hired by every other team.[57] This kind of joint profit-maximizing program is likely to produce an optimal allocation of talent regardless of the distribution of revenues. It seems more natural, however, to examine revenue-sharing rules in the context of a noncooperative game. Fixed-supply conjectures reproduce the results of a cooperative game between the teams,[58] and therefore it is perhaps not surprising that a model based on these conjectures appears to support the Coase Theorem.

The fact that almost all models of sports leagues in the literature have been based on the assumption that the total supply of talent is fixed may be associated with the fact that most of the models have been written in the context of the North American major leagues, where arguably, at any point in time supply is fixed.[59] However, even in the relatively short term it may be possible to draft in talent from outside the league, effectively increasing total supply. The increasingly global search of the major leagues for talent suggests that in the longer term supply is elastic. It would be interesting to see some empirical attempts to measure the elasticity of supply.

Frederic Palomino and Joszef Sakovics (2000) develop a model based on competition for scarce talent to account for the common observation that revenue sharing seems more prevalent in North America than in Europe.[60] In addition to the demand for success and competitive balance, they introduce the demand for the quality of the contest (i.e. the talent of the players). Regardless of the supply elasticity, revenue sharing reduces the demand for talent, since own marginal revenue from success is reduced and marginal revenue from rival success (i.e. own failure at away matches) is increased. If the market for talent ensures that marginal revenue equals marginal cost, then revenue sharing in the fixed-supply model simply drives down total cost and so raises profits (see Quirk and Fort 1995). However, with elastic supply and competition between rival leagues for players, any reduction in the willingness to pay for players by the members of a league will reduce the quality of that league (measured by total units of talent employed) relative to its rivals, and therefore undermine its relative attractiveness.

Thomas Hoehn and Szymanski (1999) develop an elastic model of European league competition that presents a related reason why revenue sharing may adversely affect competitive balance. In European sports the leading teams typically compete in more than one championship in a

[57] See Noll (1974) p. 63, equation (ii) in particular.

[58] Just as a conjectural variation of -1 produces the joint profit-maximizing solution in a quantity setting oligopoly.

[59] Scully (1989), referred to in Vrooman (2000), has dissented from the mainstream view on revenue sharing, and this could be interpreted as the holding of the contrary view, that supply is elastic. Scully (1989, 1995) discusses the elasticity of supply and cites as evidence the large salary gap between the stars and lesser players to support the proposition that supply is relatively inelastic.

[60] Thomas Ericson (2000) also points out that in a European context the supply elasticity facing each league is non-zero, and he applies this to analyzing the impact of transfer rules on the distribution of talent across large and small market leagues.

season—the domestic league and European-wide league (e.g. The Champions' League[61])—and typically these competitions run concurrently. Thus the top teams have a revenue function that depends on success in both competitions, and the weaker teams have a revenue function depending only on domestic competition. Under domestic league revenue sharing, the weaker team will be more willing to reduce investment in talent to take advantage of the strong team's success than the strong team will be to reduce its own investment, since by doing so the latter reduces its expected revenue from the European-wide competition.

7. *Other Design Issues in Team Sports*

7.1 *Prizes and Lump-Sum Revenue Sharing*

Fort and Quirk (1995) observe that sharing of local TV revenues will tend to improve competitive balance, so that the invariance principle need not hold even with fixed supply conjectures. This finding arises out of the independence of local TV revenue generating functions: no adding up constraints are involved and hence the problem resembles more closely a standard Cournot-Nash model where (a) noncooperative behavior does not yield joint profit maximization, and (b) revenue sharing causes each firm to internalize the effects of its decisions on its rival and therefore leads to joint profit maximization. For example, suppose that in the two-team model each generated income only from local TV revenues, labeled L, and that these revenues are increasing in the success of the home team. With revenue sharing we can write the profit function for each team as

$$\pi_i = p_i(e_i)[\alpha L_i + (1-\alpha) L_j] - ce_i, \qquad i = 1,2. \tag{21}$$

[61] The "Champions" in this title being the domestic league champions of the previous season.

The first-order conditions are then

$$\frac{\partial \pi_i}{\partial e_i} = \frac{\partial p_i}{\partial e_i}\left(\alpha \frac{\partial L_i}{\partial e_i} + (1-\alpha)\frac{\partial L_j}{\partial e_i}\right) = c, \qquad i = 1,2. \tag{22}$$

Taking the ratio of the two first-order conditions, we can obtain

$$\frac{\dfrac{\partial p_1}{\partial e_1}}{\dfrac{\partial p_2}{\partial e_2}} = \left[\frac{\alpha \dfrac{\partial L_2}{\partial e_2} + (1-\alpha)\dfrac{\partial L_1}{\partial e_2}}{\alpha \dfrac{\partial L_1}{\partial e_1} + (1-\alpha)\dfrac{\partial L_2}{\partial e_1}}\right]. \tag{23}$$

If we suppose that $\dfrac{\partial L_j}{\partial e_i} = -\dfrac{\partial L_j}{\partial e_j}$ then for fixed labor supply the LHS of (23) equals unity and hence local TV revenue sharing has no impact on competitive balance. However, from the point of view of TV demand, there is no reason to suppose that the marginal revenue from a unit increase in the quality of the opposition is the same as the marginal revenue from a unit decrease in the quality of the home team (because in the former case the total quantity of talent on show increases while in the latter case it decreases). In general we suppose increasing the quality of the opposition will have a higher value than reducing the quality of the home team. In the absence of symmetry, revenue sharing will reduce the marginal revenue of the large market team more than the marginal revenue of the small market team and therefore revenue sharing will improve competitive balance. Marburger (1997) suggests that this kind of asymmetry might be true for gate revenues as well, where demand for absolute quality may be important.[62]

[62] He applies his model to the case of a luxury tax (see below). Stefan Késenne (2000a) shows that if team revenues depend on the quality of visitors, proxied by their winning percentage, and that the marginal revenue from visitor quality differs across teams, then revenue sharing improves competitive balance. This is essentially the same argument as that concerning local TV revenues. See also Philip Cyrenne (2001).

However, revenue sharing reduces the marginal revenue to each team from hiring an additional unit of talent, driving down the wage rate per unit of talent and increasing profits in equilibrium. Revenue sharing works in the opposite way to a prize because it diminishes effort incentives. This naturally raises the question of how prizes would affect competitive balance in a team sports context. While most individualistic sports offer substantial financial prizes to the winners, this is usually not the case with team sports. The team that wins a league championship may receive a cup, and team members may receive substantial bonuses, but the owners of the team in general stand to gain little or no direct monetary gain (i.e. prize money) from winning a championship. It is true that participation in the playoff or finals stage can be extremely valuable, and also that sponsorship income and merchandising are likely to be substantially increased by winning a championship,[63] and that these factors will impact on decision making in much the same way as an explicit prize. One might hope to see future research attempt to quantify the value of prize like elements in the different team sports.

Suppose that each team in the league were to contribute some fixed sum to a prize fund awarded to the winning team. In the two team case, where gate revenue depends only on success, team 1 has a greater revenue generating potential from success than team 2 ($\mu > 1$), and there is no local TV income, we can write the objective functions for each team as

$$\begin{aligned} \pi_1 &= p_1(e_1)[\mu + V] - V/2 - ce_1, \\ \pi_2 &= p_2(e_2)[1 + V] - V/2 - ce_2, \end{aligned} \tag{24}$$

where $V/2$ is the lump tax on each team used to create the prize fund V. Taking the ratio of first order conditions we obtain

$$\frac{\dfrac{\partial p_1}{\partial e_1}}{\dfrac{\partial p_2}{\partial e_2}} = \frac{1+V}{\mu+V} \tag{25}$$

from inspection the RHS of (25) converges to unity as V increases, implying that, for any elasticity of supply, a team funded prize will increase competitive balance.[64] Since a prize also increases aggregate effort (as in an individualistic contest), a contest designer could maximize both competitive balance and effort incentives through the use of such prizes.[65] The intuition seems quite straightforward: when teams have differing revenue generating potential then the large (marginal) revenue generating team dominates. The creation of a prize fund equalizes incentives, so that small (marginal) revenue generating teams have as much to gain from winning as their larger rivals.

While direct financial prizes are rare in team sports, European soccer leagues have adopted revenue sharing formulas for collectively negotiated TV income on a basis that introduces the flavor of a prize, in contrast to North America where all the major leagues distribute this income on the basis of strictly equal shares. For example, in the English Premier League 25 percent of annual TV income is awarded on the basis of League rank, with the League champions receiving twenty times as much (of the 25 percent) as the team ranked last in the League.[66]

[63] The difference between first and second is likely to be much greater than the difference between second and third, a superstar effect of the kind identified by Rosen (1981). Unlike a prize, the value of merchandising and related opportunities tends to differ between teams (e.g. because market sizes differ) and hence this kind of incentive promotes asymmetry.

[64] It should be obvious that this argument will not be affected if we introduce demand for competitive balance or team quality into the revenue functions.

[65] See Szymanski (2003) for a more detailed analysis of the implications of prizes in a model of team sport contests.

[66] The precise formula is $V_R = \dfrac{n+1-R}{\sum_{i=1}^{n} R_i}$ where V_R is the prize awarded to the R_{th} ranked team and n is the number of teams in the league.

Palomino and Sakovics (2001) develop a model of TV revenue sharing to show that for a joint profit maximizing league (a) full revenue sharing is optimal when it has monopsony power in the talent market, and (b) performance based rewards (prizes) are optimal when rival leagues compete for talent. With profit maximizing owners, equal sharing of income from collectively sold broadcasting rights will have no effect on competitive balance, and will just feed through directly to the profits of the owners. A sharing rule that equalizes ex ante incentives (equality of opportunity) but leads to inequality ex post (rewards winners) will, in the absence of capital market imperfections (e.g. credit constraints) generate a more balanced contest. This proposition, though well founded in economic theory, attracts considerable skepticism from noneconomists. This may have something to do with beliefs about the operation of capital markets or about the true objective function of team owners.

7.2 *Win Maximization and Ownership Rules*

So far we have assumed that all teams are profit maximizers, an assumption with which sports economists have been quite comfortable in the United States,[67] but which often seems less appropriate in the case of European soccer.[68] This has to do with both cultural and institutional factors. Culturally, the men who set up soccer clubs were by and large amateurs who looked down on the pursuit of profit, just as their counterparts did in aristocratic cricket.[69] While in many cases there may have existed a gap between stated objectives and reality, real constraints on behavior existed and continue to exist in many cases. Many clubs in Europe are also "clubs" in the legal sense—operating under a club committee who are volunteers and have no powers of borrowing and no shareholders to whom to distribute surplus. At the very least, the taking of profits in these situations is likely to be discouraged. Furthermore, institutional rules often favor nonprofit objectives. In England the governing body still retains a maximum dividend rule, currently set at 15 percent of paid up share capital.[70] In France the government has legislated favorable tax treatment for clubs established as "companies with a sporting objective," on condition that profit taking is restricted.[71]

If teams have objectives other than profit maximization then the outcome of competition and the implications of adopting specific incentive structures may be quite different than under profit maximization. Vrooman (1997a) shows that, inter alia, player costs (effort) will be higher and competitive balance greater in an asymmetric league of win maximizers compared to profit maximizers. Késenne (2000a) addressed the question of gate sharing in the context of a league composed of win maximizers and shows that in general it will lead

[67] Although Vrooman (1997a) considers seriously the implications of alternative objectives on the part of owners. One aspect of the North American situation that has not been considered in the economics literature is the predominance of ownership of sports teams as part of a larger business empire, e.g. Ted Turner and the Atlanta Braves, Rupert Murdoch and the Dodgers. The idea that teams might be operated as part of a wider business strategy deserves some attention.

[68] Dabscheck (1975) considered Australian sports teams to be revenue maximizers.

[69] In English cricket, amateurs and "players" (i.e. paid professionals) were segregated, changing in different rooms even when they were on the same team as recently as 1962. However, appearances can be deceptive: as far back as the 1880s the greed of many amateur cricketers in demanding "expenses" led to the coining of the word "shamateurism", to describe ostensibly amateur players who demand kickbacks of one form or another.

[70] Public corporations have managed to evade this rule by establishing the football club as a subsidiary of a holding company, which faces no such restrictions.

[71] Further discussion of this is to be found in Sloane (1971), Késenne (1996), and Jean-François Bourg and Jean-Jacques Gouguet (2001). Discussion of changing behavior patterns in recent years can be found in Wladimir Andreff and Paul Staudohar (2000).

to greater competitive balance. Intuitively, if teams spend all available income on hiring talent (i.e. they face a zero-profit budget constraint), then redistributing income from wealthy teams to poor teams will tend to equalize levels of talent and thus improve competitive balance[72].

Given that different types of owners may embrace different objective functions, and that these objectives yield different outcomes, it is open to contest designers to favor particular types of owner whose equilibrium behavior is expected to produce the desired outcome. This idea is reminiscent of the "strategic delegation" literature, where a profit maximizing owner might choose to appoint a sales maximizing manager in an oligopoly (Chaim Fershtman and Kenneth Judd 1987). Rules in North America that prohibit stock flotation might be deemed to encourage "sportsmen owners" whose association with success might lead them to behave more like win maximizers than profit maximizers.[73] Similarly, restrictions in Europe that have until recently limited the spread of ownership to the stock markets may have been intended to create the same effect. Whether the ends of league organizers can be achieved by means of this kind of social engineering must remain open to doubt.

7.3 *Salary Caps, Luxury Taxes, and the Unions*

Since the 1970s, wage negotiations in the North American major leagues have been characterized by collective bargaining. Among the successes of the unions have been the introduction of veteran free agency, minimum wages, and improved pension provisions. The invariance proposition suggests that the unions would have limited impact on competitive balance but reduce the rents extracted by owners. Support for the first of these propositions was considered in section 6.1, while Zimbalist (1992) presents evidence on the second.

The antitrust exemption for collective bargaining agreements has bolstered the power of the unions by (a) enforcing exclusive bargaining rights and (b) enabling owners to enter into restrictive agreements that might not be permitted in the absence of the exemption. The value of the exemption to the owners has at times appeared so great that some union members have attempted to decertify the union in order to bring an antitrust suit against the league, most notably the NFL players' union at the time of the McNeil case (in 1989) and the NBA players' union following the expiry of the 1988 collective bargaining agreement (in 1994, for details see Staudohar 1996). In that case the union was aiming to get rid of the salary cap (introduced in basketball in 1984) which specified a maximum payroll equal to 53 percent of defined gross revenues, in exchange for a complex set of arrangements specifying minimum player payments and subsidies to weaker teams.

It is clear in theory that a salary cap should improve competitive balance,[74] and equally clear that making a salary cap effective has proved elusive. The NBA cap is perceived to have been ineffective because of the significant exemptions permitted (see Staudohar 1999) and Fort and Quirk (pp. 1277–82) find that the standard deviation of win percent

[72] As Quirk and Fort (2000) point out, this does not necessarily imply more competitive balance in a win-maximizing league for a given level of redistribution. Absent revenue sharing a win maximizing league could be less balanced than a profit-maximizing league and a given degree of revenue sharing might be inadequate to reverse the result.

[73] Brian Cheffins (1998) provides an interesting legal perspective on the different approaches in North America and Europe.

[74] Pace Vrooman (1995) makes the Coasian argument that even if teams are constrained to pay identical salaries, they still have incentives to ensure that talent gravitates to its most profitable location. A team could evade the effect of the cap through the promise of endorsements and non-pecuniary benefits.

has increased since its introduction (see also Késenne 2000b). From the point of view of contest design, a salary-cap system should have an effect similar to revenue sharing when teams are win maximizers. Under win maximization an increase in revenue sharing reduces the expenditure of the large revenue-generating teams, but also increases the spending of the small revenue-generating teams, and both effects enhance competitive balance. To be fully effective a salary-cap system also needs to ensure that the small revenue generating teams raise their spending to the level of the cap.[75]

A luxury tax works in a similar way to a salary cap, but instead of imposing a fixed limit (like a quota) it discourages acquisition of playing talent by taxing expenditure over a fixed limit (a tariff). The theoretical implications are discussed by Marburger (1997).[76] The only instance of this system in the major leagues has been the agreement between MLB and the MLBPA following the 232-day strike in 1994–95. When the two parties agreed on a settlement, it included a complex arrangement to tax expenditures of the top five payrolls on expenditures over fixed limits. The tax operated between 1997 and 1999 at a rate of 35 percent in the first two years and 34 percent in the third year. This system raised $30.6 million over the three years for redistribution to the weaker teams, compared to total MLB payroll spending of $3877 million over the same period. Not surprisingly the luxury tax was deemed to have little effect.[77] In 2002, MLB agreed on a new luxury tax after narrowly avoiding a strike.[78]

The roster limit, through which the number of players permitted on the payroll is fixed, is a much more venerable institution in North American sports, intended to prevent the stockpiling of top players, although there is surprisingly little academic research on its impact. In baseball it is commonly argued that the farm system has been the method by which teams have evaded the roster-limit rules, but there is a complex interaction between the rules and player contracts. The existence of roster limits is itself evidence that one of the most widely adopted assumptions in modeling team sports contests (and one adopted in this paper), namely that talent is perfectly divisible, does not hold. This is an issue clearly meriting further research.

Schemes such as salary caps, luxury taxes, and roster limits have not been introduced into the European soccer system. One reason is that there is no collective bargaining over salaries at the European Union level, another is that such bargaining would not, even if it existed, enjoy an equivalent antitrust exemption. Nor is it likely that such agreements could be agreed among the clubs in a system of multiple leagues. A salary cap tailored to the average team in the top division of a national league would seriously handicap a leading team in that league which was also competing at the European level. Moreover, a salary cap applied only in one national league would cause the most talented players in that league to move to rival national leagues which did not operate a cap. Any European-wide system would face the obstacle of significant international differences in standards of living, tax rates

[75] Arie Gavious, Benny Moldovanu, and Aner Sela (2002) show that imposing a bid cap in the context of an all pay auction reduces the bid of low cost (high revenue) types and increases the bids of high cost (low revenue) types, suggesting that even without imposing constraints there will be a tendency for competitive balance to improve.

[76] See also Gustafson and Hadley (1996).

[77] Somewhat oddly, the Blue Ribbon Panel (Levin et al. 2000) attributed its failure to the fact that the tax threshold was a floating one (p. 39), rather than the fact that the tax threshold was simply set too high.

[78] The tax regime was set for a four-year period, the tax thresholds being $117 million in 2003, $121 million in 2004, $128 million in 2005, and $137 million in 2006. Tax rates were 17.5 percent in 2003, rising to 22.5 percent for first-time offenders and 30 percent for repeat offenders in 2004 and 2005, with third-time offenders paying 40 percent in the latter year, and then 40 percent in 2006 except for first-time offenders.

and administrative systems. Only if a closed superleague system emerged in Europe, constructed on similar lines to the major leagues, is it likely that such arrangements would become feasible (Hoehn and Szymanski 1999 explore this possibility).

7.4 *Optimal Number of Teams in the League*

An obvious puzzle for the design of a sport's league is its optimal size. This issue has been a constant concern of league authorities in North America over time, and is also associated with the public policy concern over the relocation of franchises (or the threat of relocation) to extract subsidies from local government (Noll and Zimbalist 1997). Vrooman (1997b) addresses the issue of optimal league size directly and draws the analogy with James Buchanan's (1965) theory of clubs. If members have a joint interest in total revenues generated by the club, then the individual optimum is to agree to expansion to the point where average revenue per member is maximized, which in general involves a smaller number of members than the social welfare optimum (that maximizes total member revenues).[79]

The issue can be illustrated using a simple version of the contest model. Suppose that teams in a league compete in a symmetric contest with a CSF as defined by (3) and a payoff function that depends on the expected value of the prize, the cost of effort/talent, and some fixed "locational" rent or utility (U) of local citizens derived from the presence of a team.[80] To avoid underinvestment issues we assume this rent can be fully appropriated by the local team. Further we assume that some fraction of this locational rent is allocated to a prize fund V awarded to the league champion and that $(1-\phi)$ is retained by the owner. Thus team profits are equal to $(1-\phi)\,U+p_iV-e_i$ (the marginal cost of effort is normalized to unity). Maximizing with respect to e_i yields (and assuming the supply of talent is elastic[81]) we can find the equilibrium profit of each team to be:

$$\pi = U\left[1-\gamma\phi\frac{(n-1)}{n}\right]. \qquad (26)$$

Since all consumer surplus is appropriated aggregate welfare is simply the sum of profits:

$$W = n\pi = U[n-\gamma\phi(n-1)]. \qquad (27)$$

The derivatives of welfare and of profits with respect to the number of teams are:

$$\frac{\partial\pi}{\partial n} = -U\frac{\gamma\phi}{n^2} < 0 \quad \textit{and} \quad \frac{\partial W}{\partial n} = U(1-\gamma\phi)\,. \qquad (29)$$

Since the derivative of profits with respect to n is negative, teams will prefer smaller leagues, all else equal, while as long as either γ (the discriminatory power of the contest) or ϕ (the amount of locational utility allocated to the prize) are not too large, the derivative of welfare is positive and so expansion raises welfare. In the absence of side-payments the members of a league will expand to the point where the marginal profit from expansion equals the average profit per team, rather than where the marginal profit is zero. This problem is exacerbated further if teams cannot fully appropriate locational rents. Teams oppose expansion to optimal levels in the contest model partly because this reduces their own probability of winning the prize, even though this matters little from the social planner's perspective in the symmetric case.

[79] This same argument has been applied to the inefficiency of a labor-managed firm (e.g. Benjamin Ward 1958; James Meade 1974), which might be thought an appropriate analogy for a sports league.

[80] See John Siegfried and T. Petersen (2000) for an interesting analysis of locational rents.

[81] An assumption that can be justified here since the optimal league size is a long-run decision, and in the long run talent supply is elastic (e.g. talent can be attracted away from other sports).

In a contest model where teams value championship success, there will typically be less expansion than in the "win percent" model where teams generate revenue from their success probability against each visiting team. In the contest model teams oppose expansion since it reduces their own probability of success in the contest. In the symmetric win percent model, absent capacity constraints, the teams would favor unlimited expansion since this would imply unlimited additional revenues. With a fixed talent supply teams would only wish to expand to the point where all talent resources are fully utilized.

Fort and Quirk (1992, 1995) provide a good deal of evidence to show that in fact expansion generally occurs to meet the threat of entry of a new league. Since the expected profit required to facilitate entry by an entire league is much greater than that required for a single team, underexpansion seems inevitable. In a contest, model, efficiency requires side payments (as in the standard model of a cartel; see e.g. Kevin Roberts 1985) and in practice, new entrants do make side payments in the form of expansion fees. If all the locational rents are appropriable (and municipal subsidies are often substantial) then efficient expansion should occur. However, this is tantamount to assuming that leagues are capable of operating as efficient cartels. Efficient side-payments would in principle be tailored to the opportunity costs of each incumbent team but the information requirements for this procedure would be both significant and subject to moral hazard and adverse selection. With large numbers cartel agreements may become unenforceable (Peter Cramton and Thomas Palfrey 1990).[82] In Europe these issues have never arisen. The hierarchy established by the promotion and relegation system ensures that all locations have a right to enter the league structure at some level, and after a period of years reach the highest level if the local willingness to pay is adequate.

7.5 *Promotion, Relegation, and Exclusive Territories*

The European Commission (1998) described promotion and relegation as "one of the key features of the European model of sport." It is the rule whereby the worst-performing teams at a given level of league competition are demoted at the end of the season to play in the immediately junior league and are replaced by the best performing teams from that league. For example, at the end of each season the three teams in the English Premier League with the lowest number of points won are demoted to the Football League Division One and are replaced in the Premier League by the three best performing teams from Division One. There is promotion and relegation at every level of English soccer, from the Premier League right down to the lowest level of amateur competition, so that in theory any English soccer team might one day reach the Premier League. This system is operated in all the major soccer nations and applies to most other team sports played in Europe (e.g. rugby union, basketball, ice hockey).[83] In economic terms, promotion and relegation represents an opportunity for teams to enter the market at every level of competition. Applied to baseball in the United States, for example, it would mean that AAA teams could one day play in the majors (and conversely, that the Yankees might one day play AAA baseball). The economic consequences seem to be fairly similar to the effects of open entry in any market.

First, there is no credible threat of franchise relocation in Europe, since every city

[82] Cyrenne (2001) considers a related issue, the optimal number of games in a season, and contrasts the choice of a cartel to that of a social planner.

[83] Noll (2002) and Ross and Szymanski (2002) analyze the system in more detail.

has at least one team with the potential to enter the major league (as long as it is prepared to invest in player talent) without needing to attract someone else's team. As a result teams are unable to extract large subsidies from local government in the manner so familiar in the United States (see e.g. Siegfried and Zimbalist 2000).[84] Second, teams are motivated not only to win, but also to avoid the punishment of relegation. It was noted above that the variance of seasonal win percentages is smaller in European soccer than the North American major leagues even though the variance of team expenditure is greater. That is because teams must fight to the end of the season even if they are out of contention for the championship.[85] Promotion and relegation also undermines the value of territorial exclusivity, and while it is not theoretically inconsistent for the two to co-exist, in practice open entry in Europe has meant freedom to establish a team wherever one wishes.

Promotion and relegation also has some advantages from the perspective of contest design. Authorities in a league system with promotion and relegation can optimize the number of teams eligible for the championship each season without simultaneously having to determine the size of the league. One consequence of this is that the top divisions of European soccer leagues are in fact smaller (typically with fewer than twenty teams) than the North American major leagues have become, and this can mean a less extreme difference between the best and the worst.

On the face of it this might suggest that promotion and relegation is a superior system from the point of view of consumers, although clearly inferior for the profitability of teams. However, the welfare questions are not so clear-cut. While promotion and relegation affords an opportunity for more cities to participate in the major league, it might be argued that the relegation of the Yankees to be replaced by the home team of Boise, Idaho would not represent a net increase in welfare. This is a fine judgement, even if in practice the major teams are almost never relegated.

A more subtle problem concerns the distribution of talent. If this is fixed, and the promotion and relegation system leads to a more even spread of talent across teams (because the incentive for the smaller teams to compete is greater) then the average quality of teams at the highest level (e.g. the thirty best teams) may fall, reducing the quality of individual matches.[86] Finally, as Szymanski and Valletti (2003) show, promotion and relegation may undermine the incentive to share revenues. The cost of revenue sharing to large drawing teams is the foregone income from current success, while the benefit is their share in a more valuable (because more balanced) contest. In a closed league every team is guaranteed to participate in that contest, while in an open league any team might be relegated in the future. This may be one factor contributing to the observation that leagues in Europe have adopted many fewer mechanisms to promote competitive balance than the North American majors.

If an open system obliges teams to supply more effort and reduces profits, why would the leading teams simply not secede from

[84] Relocation is in general prohibited by the governing bodies in Europe. Recently, Wimbledon, a team playing in the second tier of English soccer, was permitted to relocate, after lengthy debate, but only because the team's stadium had been closed and the local government did not allow them to build a new one (even at their own expense).

[85] In the closed North American leagues low ranked teams may prefer to lose toward the end of the season if this gives them a better draft pick. Beck Taylor and Justin Trogdon (2002) find empirical support for this proposition in the NBA.

[86] It is a mistake to argue that there is not enough talent to support a promotion and relegation system because talent will be spread too thinly. Rather, an efficient promotion and relegation system requires player mobility, since the best talent will always migrate to the top division. In practice this often happens with extraordinary speed. Promotion and relegation is a discipline on the owners rather than the players.

the League and set up on their own? The answer to this in practice is the fear of expulsion from the national Association and the international network. Indeed, it is this fear that inhibits the clubs from demanding compensation for release of contracted players to represent their national team (for as many as twenty matches in a season). FIFA pays no compensation to the clubs who continue to pay the full salaries of their players during international tournaments, and while players receive some appearance money, this is generally a tiny fraction of their total remuneration. This makes the World Cup Finals not only the world's most popular sporting event (33 billion viewers for a total of 64 matches[87]), but also, with turnover of $4 billion, one of the world's most profitable team sports events. Clubs fear expulsion from the Association since they know that most of the players are willing to play for their country for almost nothing either because of patriotism or because of the reputation effects and its impacts on endorsement income.[88] Thus any breakaway league would find it hard to retain players.

7.6 *Club versus Country*

National teams have been unimportant in the development of the major team sports in the United States, but in other sports national teams and international representative sport has been the driving force in developing the popularity of the game and providing some of the most attractive events within the sport. In individualistic sports it is clear that the Olympics has provided a showcase for development of traditional events (e.g. athletics and swimming) as well as the development of new events (e.g. Taekwondo). In team sports the soccer World Cup has been a significant contributor to the development of the sport in countries with limited professional leagues. The competition itself has helped to bring players from particular countries to international recognition while the profits generated by the competition have been used in part to fund the development of the sport (notably, on both counts, in the case of the African countries). Most ostentatiously, the decision of FIFA to locate the 1994 World Cup in the United States was seen by many as an blatant attempt to promote the game in that country given its revenue generating potential (see e.g. Sugden and Tomlinson 1999). The North American major league sports have pursued their own development activities abroad. In Europe the NFL has established its own league, with moderate success in Germany and Spain, MLB has made more than one attempt to enter the European market on a modest scale, and in China the NBA has established a subsidiary to develop the league in that market. However, they are all to a degree hampered by their own commercial objectives, given that they are ultimately responsible to profit-oriented team owners.

Soccer is simply one example of international representative competition dominating domestic league competition. Other examples include cricket (the dominant sport in India, as well as a major sport in nations of the British Commonwealth including England, Australia, South Africa, Pakistan, New Zealand, Sri Lanka, and the Caribbean islands that play collectively as the "West Indies"), and Rugby Union, a sport similarly found in most Commonwealth countries and historically dominated by New Zealand. What is striking about these examples is that (a) competitive balance plays no obvious role in the popularity of these sports; (b) the dominant teams are seldom drawn from the larger or richer

[87] This is FIFA's claimed viewership for the "France '98" World Cup. This implies everyone on the planet could have watched five games, around 50 billion viewing hours. The IOC claimed 36 billion viewing hours for Sydney 2000.

[88] In fact top players from weak countries with little chance of winning the World Cup are sometimes reluctant to appear. In the 2002 World Cup the captain of the Republic of Ireland walked out on his team claiming that the national Association was not prepared to spend enough money on training facilities for the players.

nations; and (c) international representative competition is used to subsidize domestic league competition.

On the first two points, consider the New Zealand rugby union team known as the "All Blacks." They have been playing in international competition since 1903 and have an all-time winning record of 72 percent, despite being dwarfed in terms of population size by many of their larger rivals. For example, the All Blacks currently have a winning record of 78 percent against England, with only 8 percent of the latter's population. Similarly, the Australians in cricket have a winning record of 56 percent against England in 209 matches over the period 1877 to 2001 (ignoring ties), despite a much smaller population.[89] The West Indies, drawing on the smallest population[90] of the ten Test Match Cricket playing nations, have the second highest all-time winning record (57 percent).[91] While this phenomenon is not unknown in the individualistic sports, where small and/or poor nations seem able to produce a disproportionate number of winners, it is easily exaggerated. Andrew Bernard and Meghan Busse (2000) show that population and GDP are remarkably reliable predictors of Olympic medal success.

In soccer, the dominant countries in the World Cup (played every four years) have been Brazil (five victories), Germany and Italy (three victories), Argentina and Uruguay (two victories). These five teams account for fifteen of the seventeen World Cup wins (88 percent), despite entry being open to the entire planet. Brazil has a 76 percent winning record in all World Cup matches played. Yet this dominance does not seem to have undermined interest in these competitions. One reason may be that these international competitions bring together the best players in the world and when combined with national fervor these factors outweigh a rational concern with competitive balance. The aspect of quality may also explain why such competitions have, at least as yet, limited appeal for the North American major league sports- all of the best players are already on show in the major leagues, so in that sense an international representative competition would not offer a higher level of competition. If baseball, for example, were played to a higher level in other national leagues then international competition would become attractive. In other words, the existence of a dominant national league in team sports seems to undermine the demand for international representative competition. By contrast, where there is no dominant national league, international competition becomes attractive.

The dominance of international competition creates some interesting problems in team sports. Most notably international cricket has been seriously undermined by the revelation that many of the top players have been accepting substantial bribes to fix matches for gambling purposes (see Sir Paul Condon 2001). Ian Preston, Ross, and Szymanski (2001) suggest that corruption stems not merely from moral frailty but also from the remarkably low salaries paid to the players who were induced to accept as bribes what were, for world class athletes, remarkably small sums of money (e.g. as little as $10,000). Low salaries in cricket are due not to the lack of popularity of the game (an international series of five matches can generate an income of $30 million) but to the use of these funds to subsidize domestic leagues which attract no interest from paying fans due the focus on international matches. Without the subsidy there would

[89] Over the last twenty years Australia's dominance has become embarrassing, with a 66 percent winning record in decisive matches.

[90] The other ten are Australia, Pakistan, England, South Africa, India, Sri Lanka, New Zealand, Zimbabwe, and Bangladesh, with a combined population of 1.4 billion, compared to the Island population of around 4 million. Even excluding India, this would amount to no more than 1 percent of the population of the cricketing nations.

[91] If baseball were regularly played at the international representative level, such phenomena might also emerge. For example, it is well recognized that the tiny Dominican Republic would be a competitive nation, not to mention Cuba.

be no competitive environment in which to raise players to the necessary international standard. The case of cricket contrasts with soccer where there is a balance of interest in club competition (with healthy finances) and international representative competition, which means that the former can afford to supply talent at no cost to the latter. In theory this can be seen as a kind of league tax to fund the development of the sport.

In the case of rugby union the international representative game traditionally dominated, but in recent years a successful international club competition has emerged in the southern hemisphere (played between teams from New Zealand, Australia, and South Africa) and may be emerging in Europe (where the dominant teams are located in England and France).[92] This suggests three models of sporting development: a dominant national league (North America) with limited international competition, a dominant international competition and weak national leagues (cricket, Rugby Union) and a combination of powerful national leagues with strong international representative competition (soccer). Given that talent is to a degree substitutable between sports in its developmental years (i.e. early to late teens) and sports increasingly compete to find the best talents worldwide, it is tempting to suggest that only sports with a strong financial structure based on a viable model of league competition will survive as major sports. (see Ross and Szymanski 2003 for an analysis of optimal league design.) Already cricket is suffering from a loss of interest in some of its traditional centers (e.g. the West Indies). Culture may defend other sports more robustly, but the notion that structure may influence long term popularity may be worthy of further research.

[92] John McMillan (1997) provides an interesting discussion on the balance between centralized coordination and decentralized decision making in the case of New Zealand rugby union. But see Dorian Owen and Clayton Weatherston (2002) for analysis of how provincial competition in that country has been subordinated to the needs of the national team.

8. *Antitrust and Public Policy*

In the words of Michael Flynn and Richard Gilbert (2001), "One is struck by the frequency with which the structure and rules of professional sports leagues have been the subject of antitrust challenges in recent decades." It is not intended to provide an exhaustive review of these issues, which can be found elsewhere.[93] However, given the abiding interest of the courts and legislators in the fortunes of sports leagues, the implications of both the theory and the empirical research reviewed here are worthy of brief discussion. Broadly speaking, the legal issues associated with individualistic sports have been far less numerous and weighty than those of team sports. For example, in Weiler and Roberts' exhaustive textbook, out of 1007 pages only 69 are devoted to individual sports, while most of the remainder is focused on team sports. This is perhaps because the object of competition—to find the best players/athletes—is clear-cut, and the appropriate mechanism to achieve this—contests with very large prizes and spreads—is not in question. Any restriction intended to prevent these mechanisms from working while raising profitability (e.g. excluding athletes from competition without due cause) would be unlikely to stand up in court.[94]

The focus of dispute, and in some cases legislative intervention, in team sports has been the contention of team owners and league authorities that economic restraints of one form or another are required to maintain a competitive balance which is in the

[93] The U.S. literature is particularly rich—e.g. Flynn and Gilbert (2001), Thomas Piraino (1999), Roberts (1984), Ross (1989, 1997, 1998, 2001), Ross and Lucke (1997), Paul Weiler and Gary Roberts (1998). There are also a number of European texts, e.g. Simon Gardiner et al.(1998), Alexandre Husting (1998). For an Australian perspective see Dabscheck (2000).

[94] Some issues remain, such as rules relating to eligibility, and in particular eligibility and disabilities.

interest of consumers. A natural starting point therefore is the nature of the relationship between the teams and the league. As Gilbert and Flynn observe, the antitrust analysis of agreements among business units depends to a significant degree on their ownership—subsidiaries of a holding company cannot collude among themselves, while independent entities may. In the four major leagues (MLB, NFL, NBA, and NHL) the teams are independent business entities which associate as a league to agree the rules of competition and so on. In Major League Soccer, however, the team owners have a stake in the MLS entity itself, which in turn owns all the player contracts. Moreover, it seems clear that this business structure was selected specifically to avoid the attention of the antitrust authorities.[95]

Gilbert and Flynn suggest that a natural interpretation of the economic structure of the major leagues is as a joint venture. Recognizing the "peculiar economics" of team sports—Neale's (1964) famous phrase—that production requires the cooperation of rivals, so that each team has a vested interest in the existence, and even the success, of its competitors, it is reasonable to suppose that some kinds of agreements can be legally entered into. Most obviously these include agreements on the rules of the game. This is no different from the antitrust treatment that would be accorded an agreement between two competitors entering into an agreement to bring a product to market that would not exist in the absence of the joint-venture agreement. Facilitating the joint venture may in all likelihood require the agreement of restraints among the partners. The essential legal issue is whether such ancillary restraints have the effect of significantly limiting competition, and whether such restraints are proportional to their intended benefit (see also Herbert Hovenkamp 1995 and Piraino 1999 for the legal perspective on these issues).

The types of restraints that might fall under this analysis include both labor market restraints (e.g. reserve clause, draft, salary cap, roster limits, restrictions on player trading), product market restraints (e.g. revenue sharing, collective selling, exclusive territories) and capital market restraints (e.g. restrictions on ownership). Most of these issues have been the subject of litigation. The most famous litigation in sport is *Federal Baseball v. National League*, (259, U.S. 200 (1922)) that reached the now widely condemned conclusion that baseball was exempt from the federal antitrust laws since it did not involve interstate commerce. See Zimbalist (2003) for an interesting analysis of the exemption. Since then the courts have set out to interpret this exemption for sporting leagues as narrowly as possible, and where possible to conduct a rule of reason analysis of challenged restraints.

In the labor market, *Flood v. Kuhn* (107, U.S. 258 (1972)), examined the reserve clause in baseball but refused to prohibit it on the grounds that it is for Congress to overturn the now venerable antitrust exemption of baseball. *Smith v. Pro Football, Inc.* (593 U.S. F.2d 1173 (1978)) considered the NFL draft and declared it an unreasonable restraint of trade. Writing contracts intended to evade salary cap restrictions was considered (*Bridgeman v. NBA* (re: Chris Dudley), 838 F. Supp. 172 (D.N.J. 1993)) and upheld in this limited context. *Mackey v. NFL*, 543 F.2d 606 (8th Cir.1976) rejected the "Rozelle Rule" that required teams signing a free agent in the NFL to compensate the player's previous team with a draft pick, and *McNeil et al. v. NFL* (70, F. Supp. 871 8th Circ. 1992) rejected the NFL's subsequent plan (Plan B) to allow teams to protect up to 37 players on their roster. *Finley v. Kuhn* (569, F. 2d 1193, 6th Circuit 1978) upheld the right of the commissioner of baseball to

[95] In *Fraser v. MLS* the Appeal Court cast doubt on the credibility of the single entity claim of MLS (LLC 284 F 3d.47 (1st Circ. 2002)), describing it as "somewhere between a single company (with or without wholly owned subsidiaries) and a cooperative arrangement between existing competitors."

penalize teams selling players for cash on the grounds that it might weaken the selling team and reduce competitive balance.

The relationship between collective selling of TV rights, competitive balance and revenue sharing was considered in *United States v. NFL*, 116 F. Supp. 319 (E.D. Pa. 1953) and *NCAA v. Board of Regents*, 468 U.S. 85, 107 (1984) and in both cases competitive balance justifications were considered potentially valid reasons for the maintenance of the challenged restraints (on individual selling) and so were not *per se* illegal, but in both cases on a *rule of reason* the restraints were deemed either excessive or not tailored to achieve the stated aim.[96] In the Raiders' case (*Los Angeles Memorial Coliseum Commission v. NFL*, 726 F.2d 1381 (ninth Cir. 1984) the court upheld a jury verdict that the league's application of the NFL rule requiring a majority of three-quarters of member teams to permit a relocation (thus protecting exclusive territories) restrained competition. It rejected the claim that the rule was justified by any legitimate interest of the NFL, including maintaining competitive balance. In *Sullivan v. NFL* (U.S. Court of Appeals, First Circuit, 34 F.3d 1994) the court allowed that motives such as competitive balance might on a rule of reason justify prohibiting public ownership of a franchise.

On balance it might be argued that the courts have demonstrated some skepticism about competitive balance as a justification for restraints, although they have accepted it as a possible justification under a rule of reason. However, this state of affairs has been complicated by the nonstatutory exemption for collective bargaining agreements, which has rendered the unions in North American sports so much more powerful than their European counterparts. As discussed in the case of salary caps, above, the exemption has enabled unions to bargain away rights won in the courts and to facilitate the maintenance of labor market restraints. Moreover, Congress has intervened through the 1961 Sports Broadcasting Act to exempt collectively negotiated national sponsored broadcasting agreements from antitrust scrutiny. As a result, in practice the major leagues operate a wide range of restraints, adumbrated in much of the foregoing discussion.

In European sports the power of the courts is supplemented less by the role of the legislature, which has not interfered significantly in the operation of team sports, than by the European Commission, which acts as an executive body representing the member states (who hold a power of veto over many of its activities). The competition directorate (DG IV) of the Commission wields considerable power and has intervened to challenge various restraints in recent years, and has in most cases reached agreement with the leagues prior to going to court. In European competition law the Commission in general only acts on the complaint of parties deeming themselves to be harmed by a challenged restraint, but in recent years, as the value of TV contracts has escalated so has the number of complaints received.

In the Bosman case (see above) the complaint was taken to the European Court of Justice. The court held that competitive balance was not a valid defence of the old transfer system, even though in other cases it could justify a restraint (such as revenue sharing). Moreover, the free movement of labor, a principle enshrined in the Treaty of Rome, overrode any specific consideration of the interests of the league. In 2000 the Commission went further and challenged the economic basis of the transfer fees being paid for players within contract on the grounds that they restricted the free movement of labor within the European Union. In 2001 it was announced that the Commission had reached agreement with the football governing bodies (FIFA and UEFA, the European governing body to which all the national governing bodies belong) on a com-

[96] The NCAA case also considered in detail the effect of broadcasting on live gate.

pensation system that would allow clubs to claim significant fees for players under age 23 on the grounds of investment in training costs. Players over 23 would have the right to move clubs annually, even if employed under a long-term contract, subject to an economically justifiable (presumably moderate) compensation payment.[97]

Later in 2001 the Commission issued a statement of objections[98] to the collective sale of broadcasting rights to the lucrative Champions' League competition for the top European clubs, run by UEFA. Agreement was later reached over UEFA's right to market the Championship as a whole subject to some significant restrictions. Collective selling of broadcasting rights has been challenged at the national level in a number of European countries, notably Germany (ruled illegal and then given an antitrust exemption by parliament), the United Kingdom (upheld), Denmark (upheld), the Netherlands (no decision), Italy (ruled illegal), and Spain (prohibited); see Szymanski (2002) for details.

It seems that the soccer leagues of Europe have received much less favorable antitrust treatment than the North American leagues. Given that the European leagues have maintained a high degree of public interest and structural stability over the last half century despite having fewer restraints and less competitive balance than the North American leagues, would it be correct, as the European Commission (1998) has done, to speak of a European model of sport?[99] Currently the main issue is whether the existing structures are stable or whether the growing commercialization of the sport will lead to restructuring. Hoehn and Szymanski (1999) suggest one kind of restructuring in which the dominant clubs of Europe (who are already organized in an exclusive bargaining group called G14) break away to form their own closed superleague along North American lines. If competitive balance really matters then we should expect the European system to collapse.

9. *Conclusions*

It is a commonplace among economists to hold up sports as an example of contest/tournament theory in action, but in practice a lot remains to be done both to understand the relationship between tournament structures and incentives in theory, and to test theories against the data. One objective of this review has been to discuss the contest theory literature in the context of sports. While there been a good deal of research that has direct implications for the design of individualistic contests, empirical testing remains limited despite widespread agreement that this would be a very fruitful area in which to conduct testing. Moreover, there are many aspects of the organization of individualistic sports that could be modeled more fully with a view to establishing an optimal design: e.g. optimal prize spreads in asymmetric contests, competition between rival contest organizers, the entry rules for contestants and optimal handicapping, to select just a few.

The relationship between team sports and contest theory seems even less well developed. The role of prizes in providing incentives has been largely ignored in the team sports literature, where much of the policy oriented research has focused on redistribution mechanisms such as revenue sharing, and has been preoccupied with the proposition that such sharing is likely to have a neutral impact. In this paper that claim is shown to depend on the assumption that an inelastic supply of talent is incorporated into the conjectural variations of the owners generating an equilibrium that is not Nash. This seems a relatively unfruitful avenue for research. An alternative way forward is the analysis of incentive structures. That prizes

[97] See *FIFA Regulations for the Status and Transfer of Players*, July 2001 (see also the comments in section 6.1 above).

[98] *European Union Official Journal*, C 169, 13.06.2001, p. 5.

[99] See Didier Primault and Arnaud Rouger (1999) for a trenchant assertion of difference.

enhance incentives is surely a fundamental proposition of economic theory, but one that has been little studied in the team sports literature. The analysis of revenue sharing has paid little attention to the different ways that revenues for sharing can be collected or the basis of their allocation. For example, even if TV rights are sold collectively, different rules for distributing that income have quite different implications for incentives (and profits). The impact of prize funds also depends on the organizational structure of a sport (e.g. with or without inter-league economic rivalry).

One weakness of much of the existing literature is that the appropriate definition of a welfare function against which the optimality of contest can be measured is not carefully specified. This paper has not touched in detail on this issue, but it is clearly critical. A conventional IO approach would be to focus on consumer surplus, but the complex specification of consumer demand, given the role of team loyalty, competitive balance and team quality, as well as the more mundane issue of price, makes this approach problematic. In the contest literature the convention has been to focus on the issue of rent dissipation—but is this an appropriate yardstick for sporting contests? More work remains to be done to settle this crucial issue.

Comparative institutional analysis has much to offer for our understanding of organizational issues in team sports, not just between North America and Europe, but with other countries such as Australia with developed national sports and with other multinational sports such as cricket. Rosen and Sanderson (2001) reflected on the difference between North American and European leagues thus:

> All schemes used in the United States punish excellence in one way or another. The European football approach punishes failure by promoting excellent minor league teams to the majors and demoting (relegating) poor performing major league teams back down to the minors. The revenue loss from a potential demotion to a lower class of play is severe punishment for low quality—severe enough that salary treaties, league sharing arrangements, and unified player drafts are so far thought to be unnecessary, even though star salaries are enormous. It is an interesting economic question as to which system achieves better results.

Careful consideration of the impact of institutional differences may eventually lead to a better understanding of the incentive effects of contest design.

Empirically, some fundamental issues remain unresolved. For example, the central claims of sports economists, that uncertainty of outcome boosts demand for sporting contests and that inequality of economic resources leads to more certainty of outcome obtain only weak support in the literature. Given that many successful team sports are characterized by highly unbalanced competition (e.g. soccer) and that proposed balance enhancing measures are almost always profit enhancing, there are grounds for caution. From a policy point of view it may be that the invariance principle has been unhelpful in encouraging the view that restrictive measures would at least do no harm, even if they do no good. Given the role that economists frequently play in antitrust analysis, these theoretical and empirical perspectives have important policy implications.

We are still some way from being able to fully model and test an optimal design of a sporting contest. Such a project, however, is not beyond the capabilities of the economics profession.

References

Alexander, Donald. 2001. "Major League Baseball: Monopoly Pricing and Profit-Maximizing Behavior," *J. Sports Econ.* 2:4, pp. 341–55.

Andreff, Wladimir and Paul Staudohar. 2000. "The Evolving Model of European Sports Finance," *J. Sports Econ.* 1:3, pp. 257–76.

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 J. Econ.* 19:1, pp. 27–43

Baik, Kyung. 1994. "Effort Levels in Contests with Two Asymmetric Players," *South. Econ. J.* 61, pp. 367–78.

Baimbridge, Mark; Samuel Cameron and Peter Dawson. 1996. "Satellite Television and the Demand

for Football: A Whole New Ball Game?" *Scot. J. Polit. Econ.* 43:3, pp. 317–33.

Balfour, Alan and Philip Porter. 1991. "The Reserve Clause in Professional Sports: Legality and Effect on Competitive Balance," *Lab. Law J.* 42, pp. 8–18.

Baye Michael; Dan Kovenock and Casper de Vries. 1994. "The Solution to the Tullock Rent-Seeking Game When R>2: Mixed-Strategy Equilibria and Mean Dissipation Rates," *Public Choice* 81:3–4, pp. 363–80.

———. 1996. "The All-Pay Auction with Complete Information," *Econ. Theory* 8:2, pp. 291–305

———. 1999. "The Incidence of Overdissipation in Rent-Seeking Contests," *Public Choice* 99, pp. 439–54.

Becker, Brian and Mark Huselid. 1992. "The Incentive Effects of Tournament Compensation Systems," *Admin. Sci. Quart.* 37, pp. 336–50.

Bernard, Andrew and Meghan Busse. 2000. "Who Wins the Olympic Games: Economic Development and Medal Totals," NBER work. paper #7998.

Borland, Jeff. 1987. "The Demand for Australian Rules Football," *Econ. Record* 63:182, pp. 220–30.

Borland, Jeff and Jenny Lye. 1992. "Attendance at Australian Rules Football: A Panel Study," *Appl. Econ.* 24, pp. 1053–58.

Bourg, Jean-François and Jean-Jacques Gouguet. 2001. *Economie du Sport*. Paris: Editions La Decouverte.

Buchanan, James. 1965. "An Economic Theory of Clubs," *Economica* (Feb.) pp. 1–14.

Bulow, Jeremy and Paul Klemperer. 2001. "The Generalized War of Attrition," *Amer. Econ. Rev.* 89:1, pp. 175–89.

Butler, Michael. 1995 "Competitive Balance in Major League Baseball," *Amer. Economist* 39:2, pp. 46–52.

Buzzacchi, Luigi; Stefan Szymanski and Tommaso Valletti. 2003. "Equality of Opportunity and Equality of Outcome: Open Leagues, Closed Leagues and Competitive Balance," *J. Indust. Comp. Trade* 3:3, pp. 167–86.

Cairns, John. 1987. "Evaluating Changes in League Structure: The Organisation of the Scottish Football League," *Appl. Econ.* 19:2, pp. 259–75.

Cairns, John; Nicholas Jennett and Peter Sloane. 1986. "The Economics of Professional Team Sports: A Survey of Theory and Evidence," *J. Econ. Stud.* 13, pp. 3–80.

Cheffins, Brian. 1998. "Sports Teams and the Stock Market: A Winning Match?" *UBC Law Rev.* 32, pp. 271–91.

Condon, Sir Paul. 2001. *Report on Corruption in International Cricket*. London: Anti-Corruption Unit, Int. Cricket Council.

Cramton, Peter and Thomas Palfrey. 1990. "Cartel Enforcement with Uncertainty about Costs," *Int. Econ. Rev.* 31:1, pp. 17–47.

Cymrot, Donald; James Dunlevy and William Even. 2001. "'Who's on First': An Empirical Test of the Coase Theorem in Baseball," *Appl. Econ.* 33, pp. 593–603.

Cyrenne, Philip. 2001. "A quality-of-play model of a professional sports league" *Econ. Inquiry* 39, pp. 444–52.

Dabscheck, Braham. 1975. "Sporting Equality: Labour Market vs Product Market Control," *J. Ind. Relat.* 17:2, pp. 174–90.

———. 1989. "Abolishing transfer fees: the Victorian Football League's new employment rules" *Sporting Traditions* Nov. pp. 63–87.

———. 2000. "Sport, Human Rights and Industrial Relations," *Australian J. Human Rights*, 6:2, pp. 129–59.

Daly, George. 1992. "The Baseball Player's Market Revisited," in *Diamonds Are Forever: The Business of Baseball*. P. Sommers, ed. Brookings Institution.

Daly, George and William Moore. 1981. "Externalities, Property Rights, and the Allocation of Resources in Major League Baseball," *Econ. Inquiry* 29, pp. 77–95.

Dasgupta, Ani and Kofi Nti. 1998. "Designing an Optimal Contest," *Europ. J. Polit. Econ.* 14, pp. 587–603

Davies, Brian; Paul Downward and Ian Jackson. 1995. "The Demand for Rugby League: Evidence from Causality Tests," *Applied Econ.* 27, pp. 1003–1007.

Depken, Craig. 1999. "Free-Agency and the Competitiveness of Major League Baseball," *Rev. Ind. Org.* 14, pp. 205–17.

Dixit, Avinash. 1987. "Strategic Behavior in Contests," *Amer. Econ. Rev.* 77, pp. 891–98.

Dobson, Stephen and John Goddard. 1998. "Performance and Revenue in Professional League Football: Evidence from Granger Causality Tests," *Applied Econ.* 30, pp. 1641–51.

Downward, Paul and Alistair Dawson. 2000. *The Economics of Professional Team Sports*. London: Routledge.

Duggan, Mark and Steven Levitt. 2002. "Winning Isn't Everything: Corruption in Sumo Wrestling," *Amer. Econ. Rev.* 92:5, pp. 1594–605

Eckard, Woodrow. 1998. "The NCAA Cartel and Competitive Balance in College Football," *Rev. Ind. Org.* 13, pp. 347–69.

———. 2001. "Free Agency, Competitive Balance and Diminishing Returns to Pennant Contention," *Econ. Inquiry* 39:3, pp. 430–43.

Ehrenberg, Ronald and Michael Bognanno. 1990a. "Do Tournaments Have Incentive Effects?" *J. Polit. Econ.* 98:6, pp. 1307–24.

———. 1990b. "The Incentive Effects of Tournaments Revisited: Evidence from the European PGA Tour," *Ind. Lab. Relat. Rev.* 43, pp. 74S–88S.

El-Hodiri, Mohamed and James Quirk. 1971. "An Economic Model of a Professional Sports League," *J. Polit. Econ.* 79, pp. 1302–19

Ericson, Thomas. 2000. "The Bosman Case: Effects of the Abolition of the Transfer Fee," *J. Sports Econ.* 1:3, pp. 203–18.

European Commission. 1998. "The European Model of Sport," consultation paper of DGX.

Fees, Eberhard and Gerd Muehlheusser. 2002. "Transfer Fee Regulations in Europe," *Europ. Econ. Rev.* forthcoming.

Fernie, Susan and David Metcalf. 1999. "It's Not What You Pay, It's the Way that You Pay It, and That's What Gets Results: Jockeys' Pay and Performance," *Labor* 13, pp. 385–411.

Fershtman, Chaim and Kenneth Judd. 1987.

"Equilibrium Incentives in Oligopoly," *Amer. Econ. Rev.* 77, pp. 927–40.

Flynn, Michael and Richard Gilbert. 2001. "An Analysis of Professional Sports Leagues as Joint Ventures," *Econ. J.* 111, pp. F27–F46.

Forrest, David and Robert Simmons. 2002a. "Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer," *J. Royal Statist. Society, Series D* (*The Statistician*), 51:2, pp. 229–41.

———. 2002b. "Team Salaries and Playing Success in Sports: A Comparative Perspective," *Zeitschrift für Betriebswirtschaft* 72:4.

Forrest, David, Robert Simmons and Patrick Feehan. 2002. "A Spatial Cross-Sectional Analysis of the Elasticity of Demand for Soccer," *Scot. J. Polit. Econ.* 49:3, pp. 336–56

Fort, Rodney. 2000. "European and North American Sports Differences," *Scot. J. Polit. Econ.* 47:4, pp. 431–55.

Fort, Rodney and James Quirk. 1995. "Cross Subsidization, Incentives and Outcomes in Professional Team Sports Leagues," *J. Econ. Lit.* 33:3, pp. 1265–99.

———. 2000. "Owner Objectives and Competitive Balance," work. paper, econ. dept. Washington State U.

Frick, Bernd; Joachim Prinz and Alexander Dilger. 2001. "Pay and Performance in Professional Road Running: The Case of City Marathons," mimeo,U. Witten.

Fudenberg, Drew; Richard Gilbert, Joseph Stiglitz and Jean Tirole. 1983. "Pre-emption, Leapfrogging and Competition in Patent Races," *Europ. Econ. Rev.* 22, pp. 3–31.

Fudenberg, Drew and Jean Tirole. 1985. "Pre-Emption and Rent Equalization in the Adoption of New Technologies" *Rev. Econ. Stud.* 52, pp. 383–402.

———. 1991. *Game Theory*. Cambridge, MA and London: MIT Press.

Fullerton, Richard and Preston McAfee. 1999. "Auctioning Entry in Tournaments," *J. Polit. Econ.* 107:3, pp. 573–606

Gardiner Simon; Alexandra Felix, John O'Leary, Mark James and Roger Welch. 1998. *Sports Law*. London: Cavendish Pub.

Garicano, Luis and Ignacio Palacios-Huerta. 2000. "An Empirical Examination of Multidimensional Effort in Tournaments," mimeo, econ. dept. Brown U.

Gavious, Arie; Benny Moldovanu and Aner Sela. 2002. "Bid Costs and Endogenous Bid Caps," *Rand J. Econ.* 33:4, pp. 709–22.

Gradstein, Mark and Kai Konrad. 1999. "Orchestrating Rent-Seeking Contests," *Econ. J.* 109, pp. 536–45.

Green, Jerry and Nancy Stokey. 1983. "A Comparison of Tournaments and Contests," *J. Polit. Econ.* 91, pp. 349–64

Grier, Kevin and Robert Tollison. 1994. "The Rookie Draft and Competitive Balance: The Case of Professional Football," *J. Econ. Behav. Org.* 25, pp. 293–98.

Gustafson E. and L. Hadley. 1996. "The Luxury Tax Proposal for Major League Baseball: A Partial Equilibrium Analysis," in *Baseball Economics: Current Research*. Fizel, Gustafson, and Hadley, eds. Westport, CT: Praeger.

Guttman, Allen. 1994. *Games and Empires*. NY: Columbia U. Press

Hall, Stephen, Stefan Szymanski and Andrew Zimbalist. 2002. "Testing Causality between Team Performance and Payroll: The Cases of Major League Baseball and English Soccer," *J. Sports Econ.* 3:2, pp. 149–68

Harris, Christopher and John Vickers. 1985. "Perfect Equilibrium in a Model of a Race," *Rev. Econ. Stud.* 52:2, pp. 193–209.

———. 1987. "Racing with Uncertainty," *Rev. Econ. Stud.* 54, pp. 1–22.

Hart, R.A.; J. Hutton and T. Sharrot. 1975. "A Statistical Analysis of Association Football Attendances," *Applied Statist.* 24:1, pp. 17–27.

Hess, Rob and Bob Stewart. 1998. *More Than a Game: An Unauthorised History of Australian Rules Football.* Melbourne U. Press.

Higgins Richard; William Shughart and Robert Tollison. 1985. "Free Entry and Efficient Rent-Seeking," *Public Choice* 46, pp. 247–58.

Higgins, Richard and Robert Tollison. 1990. "Economics at the Track," in *Sportometrics*. Brian Goff and Robert Tollison, eds. Texas A&M U. Press, pp. 15-34.

Hillman, Arye and John Riley. 1989. "Politically Contestable Rents and Transfers," *Econ. Politics* 1:1, pp. 17–39.

Hoehn, Thomas and Stefan Szymanski. 1999. "The Americanization of European Football," *Econ. Policy* 28, pp. 205–40.

Horowitz, Ira. 1997. "The Increasing Competitive Balance in Major League Baseball," *Rev. Ind. Org.* 12, pp. 373–87.

Hovenkamp, Herbert. 1995. "Exclusive Joint Ventures and Antitrust Policy," *Columbia Bus. Law Rev.* 1.

Humphreys, Brad. 2002. "Alternative Measures of Competitive Balance in Sports Leagues," *J. Sports Econ.* 3:2, pp. 133–48.

Husting, Alexandre. 1998. *L'Union Europeene et le Sport*. Lyon: Les Editions Juris Service.

Hylan, Timothy; Maureen Lage and Michael Treglia. 1996. "The Coase Theorem, Free Agency, and Major League Baseball: A Panel Study of Pitcher Mobility from 1961 to 1992," *South. Econ. J.* 62, pp. 1029–42.

Inglis, Simon. 1988. *League Football and the Men Who Made It*. Willow Books. London.

Jennett, Nicholas. 1984. "Attendances, Uncertainty of Outcome and Policy in the Scottish Football League," *Scot. J. Polit. Econ.* 31:2, pp. 176–98.

Jones, J.C.H. and D.G. Ferguson. 1988. "Location and Survival in the National Hockey League," *J. Ind. Econ.* 36:4, pp. 443–57.

Kahane, Leo and Stephen Shmanske. 1997. "Team Roster Turnover and Attendance in Major League Baseball," *Applied Econ.* 29:4, pp. 425–31.

Kahn, Lawrence. 1993. "Managerial Quality, Team Success, and Individual Performance in Major League Baseball," *Ind. Lab. Relat. Rev.* 46, pp. 531–47.

———. 2000. "The Sports Business as a Labor Market Laboratory," *J. Econ. Perspect.* 14:3, pp. 75–94.

Késenne, Stefan. 1996. "League Management in Professional Team Sports with Win Maximizing

Clubs," *Europ. J. Sport Manage.* 2:2, pp. 14–22.
———. 2000a. "Revenue Sharing and Competitive Balance in Professional Team Sports," *J. Sports Econ.* 1:1, pp. 56–65.
———. 2000b. "The Impact of Salary Caps in Professional Team Sports," *Scot. J. Polit. Econ.* 47:4, pp. 422–30.
Késenne, Stefan and Claude Jeanrenaud, eds. 1999. *Competition Policy in Professional Sports: Europe after the Bosman Case*. Antwerp Standaard Eds.
Kipker, Ingo. 2000. "Determinanten der Zuschauernachfrage im professionellen Teamsport: Wie wichtig ist die sportliche Ausgeglichenheit?" PhD dissert. Heinrich Heine U., Dusseldorf.
Knowles Glenn, Keith Sherony and Michael Haupert. 1992. "The Demand for Major League Baseball: A Test of the Uncertainty of Outcome Hypothesis," *Amer. Economist* 36, pp. 72–80.
Kuypers, Timothy. 1996. "The Beautiful Game? An Econometric Study of Why People Watch English Football," *Discus. Papers Econ.* 96-01 U. College London.
La Croix S. and Kawaura A. 1999. "Rule Changes and Competitive Balance in Japanese Professional Baseball," *Econ. Inquiry* 37:2, pp. 353–68.
Lazear, Edward. 1989. "Pay Equality and Industrial Politics," *J. Polit. Econ.* 97:3, pp. 561–80.
———. 1995. *Personnel Economics*. MIT Press.
Lazear, Edward and Sherwin Rosen. 1981. "Rank Order Tournaments as Optimal Labor Contracts," *J. Polit. Econ.* 89, pp. 841–64.
Lee, Soonhwan and Hyosung Chun. 2001. "Economic Values of Professional Sport Franchises in the United States," *Sport J.* 5:3 (an electronic journal: http://www.thesportJ.org).
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.
Loury, Glenn. 1979. "Market Structure and Innovation," *Quart. J. Econ.* 93, pp. 385–410.
Lynch, James and Jeffrey Zax. 2000. "The Rewards to Running: Prize Structure and Performance in Professional Road Racing," *J. Sports Econ.* 1, pp. 323–40.
McMillan, John. 1997. "Rugby Meets Economics," *New Zealand Econ. Papers* 31:1, pp. 93–144.
Maloney, Michael and Robert McCormick. 2000. "The Response of Workers to Wages in Tournaments: Evidence ffrom Foot Races," *J. Sports Econ.* 1, pp. 99–123.
Maloney, Michael and Kristina Terkun. 2002. "Road Warrior Booty: Prize Structures in Motorcycle Racing," *Contrib. Econ. Anal. Policy* 1:1.
Marburger, Daniel. 1997. "Gate Revenue Sharing and Luxury Taxes in Professional Sports," *Contemp. Econ. Policy* 15, pp. 114–23.
———. 2002. "Property Rights and Unilateral Player Transfers in a Multiconference Sports League," *J. Sports Econ.* 3:2, pp. 122–32
Mason, Tony. 1980. *Association Football and English Society, 1863–1915*. Brighton, UK: Harvester Press.
Meade, James. 1974. "Labor-Managed Firms in Conditions of Imperfect Competition," *Econ. J.* 84, pp. 817–24
Meyer, Bruce. 1995. "Natural and Quasi-Experiments in Economics" *J. Bus. Econ. Statist.* 13, pp. 151–62.
Moldovanu, Benny and Aner Sela 2001. "The Optimal Allocation of Prizes in Contests," *Amer. Econ. Rev.* 91:3, pp. 542–58.
———. 2002. "Contest Architecture," discus. paper, U. Mannheim.
Monahan, George and Matthew Sobel. 1994. "Stochastic Dynamic Market Share Attraction Games," *Games Econ. Behav.* 6, pp. 130–49.
Nalebuff, Barry and Joseph Stiglitz. 1983. "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *Bell J. Econ.* 14, pp. 21–43
Neale, Walter. 1964. "The Peculiar Economics of Professional Sport," *Quart. J. Econ.* 78:1, pp. 1–14.
Nitzan, Shmuel. 1994. "Modelling Rent Seeking Contests," *Europ. J. Polit. Econ.* 10, pp. 41–60
Noll, Roger. 1974. "Attendance and Price Setting" in *Government and the Sports Business*. Roger Noll, ed. Washington: Brookings Inst.
———. 1999. "Competition Policy in European Sports After the Bosman Case," in *Competition Policy in Professional Sports: Europe after the Bosman Case*. Stefan Késenne and Claude Jeanrenaud, eds. Antwerp: Standaard Eds.
———. 2002. "The Economics of Promotion and Relegation in Sports Leagues: The Case of English Football," *J. Sports Econ.* 3:2, pp. 169–203.
Noll, Roger and Andrew Zimbalist. 1997. *Sports, Jobs and Taxes*. Washington: Brookings Inst.
Nti, Kofi. 1997. "Comparative Statics of Contests and Rent-Seeking Games," *Int. Econ. Rev.* 38:1, pp. 43–59.
Orszag, Michael. 1994. "A New Look at Incentive Effects and Golf Tournaments," *Econ. Letters* 46, pp. 77–88.
Owen, Dorian and Clayton Weatherston. 2002. "Professionalization and Competitive Balance in New Zealand Rugby Union," mimeo, U. Otago.
Palomino, Frederic and Joszef Sakovics. 2000. "Revenue Sharing in Professional Sports Leagues: For the Sake of Competitive Balance or as a Result of Monopsony Power?" mimeo, CentER discus. paper 2000-110, Tilburg U.
———. 2001. "Inter-League Competition for Talent vs. Competitive Balance," work. paper, CentER, Tilburg U.
Peel, David and David Thomas. 1988. "Outcome Uncertainty and the Demand for Football: An Analysis of Match Attendances in the English Football League," *Scot. J. Polit. Econ.* 35:3, pp. 242–49.
———. 1992. "The Demand for Football: Some Evidence on Outcome Uncertainty," *Empir. Econ.* 17, pp. 323–31.
———. 1997. "Handicaps, Outcome Uncertainty and Attendance Demand," *Applied Econ. Letters* 4, pp. 567–70.
Piraino, Thomas. 1999. "A Proposal for the Antitrust Regulations of Professional Sports," *Boston U. Law Rev.* 79:889.

Preston, Ian; Stephen F. Ross and Stefan Szymanski. 2001. "Seizing the Moment, a Blueprint for Reform of World Cricket," work. paper, University College, London.

Primault, Didier and Arnaud Rouger. 1999. "How Relevant is North American Experience for Professional Team Sports in Europe" in *Competition Policy in Professional Sports: Europe after the Bosman Case*. Stefan Késenne and Claude Jeanrenaud, eds. Antwerp Standaard Eds.

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 DC: Brookings Instit.

Quirk, James and Rodney Fort. 1992. *Pay Dirt: The Business of Professional Team Sports*. Princeton: Princeton U. Press.

———. 1999. Hard Ball: The Abuse of Power in Pro Team Sports. Princeton U. Press.

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.

Roberts, Gary. 1984. "Sports Leagues and the Sherman Act: The Use and Abuse of Section 1 to Regulate Restraints on Intraleague Rivalry," *UCLA Law Rev.* 32, 219, pp. 286–87

Roberts, Kevin. 1985. "Cartel Behaviour and Adverse Selection" *J. Ind. Econ.* 33, pp. 401–13.

Rosen, Sherwin. 1981. "The Economics of Superstars," *Amer. Econ. Rev.* pp. 845–58

———. 1986. "Prizes and Incentives in Elimination Tournaments," *Amer. Econ. Rev.* 76, pp. 701–15.

Rosen, Sherwin and Allen Sanderson. 2001. "Labor Markets in Professional Sports" *Econ. J.* 111:469, pp. F47–F68

Ross, Stephen F. 1989. "Monopoly Sports Leagues," *U. Minnesota Law Rev.* 73:643.

———. 1997. "The Misunderstood Alliance Between Sports Fans, Players, and the Antitrust Laws," *U. Illinois Law Rev.* 1997:519.

———. 1998. "Restraints on Player Competition that Facilitate Competitive Balance and Player Development and their Legality in the United States and in Europe," in *Competition Policy in Professional Sports: Europe after the Bosman Case*. Stefan Késenne and Claude Jeanrenaud, eds. Antwerp: Standaard.

———. 2001. "Antitrust Options To Redress Anticompetitive Restraints And Monopolistic Practices By Professional Sports Leagues," paper delivered to 2nd Annual Amer. Antitrust Inst. Conf.

Ross, Stephen F. and Robert Lucke. 1997. "Why Highly Paid Athletes Deserve More Antitrust Protection than Unionized Factory Workers," *Antitrust Bull.* 42:3, pp. 641–79.

Ross, Stephen F. and Stefan Szymanski. 2002. "Open Competition in League Sports," *Wisconsin Law Rev.* 2002:3, pp. 625–56.

———. 2003. "The Law and Economics of Optimal Sports League Design," work. paper, Business School, Imperial College, London.

Rottenberg, Simon. 1956. "The Baseball Player's Labor Market," *J. Polit. Econ.* 64, pp. 242–58.

Schmidt, Martin and David Berri. 2001. "Competitive Balance and Attendance: the Case of Major League Baseball," *J. Sports Econ.* 2:2, pp. 145–67.

Scully, Gerald. 1974. "Pay and Performance in Major League Baseball," *Amer. Econ. Rev.* 64, pp. 915–30.

———. 1989. *The Business of Major League Baseball*. Chicago: U. Chicago Press.

———. 1995. *The Market Structure of Sports*. U. Chicago Press.

Seymour, Harold. 1960. *Baseball: The Early Years*. NY: Oxford U. Press.

Siegfried, John and T. Petersen. 2000. "Who Is Sitting in the Stands? The Income Levels of Sports Fans," in *The Economics of Sports.* W. Kern, ed. Kalamazoo: Upjohn Inst.

Siegfried, John and Andrew Zimbalist. 2000. "The Economics of Sports Facilities and Their Communities," *J. Econ. Perspect.* 14, pp. 95–114.

Skaperdas, Stergios. 1996. "Contest Success Functions," *Econ. Theory* 7, pp. 283–90.

Sloane, Peter. 1969. "The Labor Market in Professional Football," *British J. Ind. Relat.* 7:2, pp. 181–99.

———. 1971. "The Economics of Professional Football: The Football Club as a Utility Maximizer," *Scot. J. Polit. Econ.* 17, 2, pp. 121–46.

Staudohar, Paul. 1996. *Playing for Dollars: Labor Relations and the Sports Business*. Cornell: ILR Press.

———. 1999. "Salary Caps in Professional Team Sports," in Késenne and Jeanrenaud, eds. op cit.

Sugden, John and Alan Tomlinson. 1999. *Great Balls of Fire*. Edinburgh: Mainstream Pub.

Szymanski, Stefan. 1999. "The Market for Soccer Players in England after Bosman: Winners and Losers," in Késenne and Jeanrenaud, eds. op cit.

———. 2001. "Income Inequality, Competitive Balance and the Attractiveness of Team Sports: Some Evidence and a Natural Experiment from English Soccer," *Econ. J.* 111, pp. F69–F84.

———. 2002. "Collective Selling of Broadcast Rights to Sporting Events," *Int. Sports Law Rev.* 2: 1, pp. 3–7.

———. 2003. "Incentives and Competitive Balance in Team Sports," *Europ. Sport Manage. Quart.* 3:1, pp. 11–30.

———. 2004. "Professional Team Sports Are Only a Game: The Walrasian Fixed Supply Conjecture Model, Contest Nash Equilibrium and the Invariance Principle," forthcoming, *J. Sports Econ.*

Szymanski, Stefan and Stefan Késenne. 2004. "Competitive Balance and Gate Revenue Sharing in Team Sports," forthcoming. *J. Indust. Econ.* 42:1, pp. 513–25.

Szymanski, Stefan and Tim Kuypers. 1999. *Winners and Losers: The Business Strategy of Football*. London: Viking Books.

Szymanski, Stefan and Ron Smith. 1997. "The English Football Industry, Profit, Performance and Industrial Structure," *Int. Rev. Appl. Econ.* 11:1, pp. 135–53.

———. 2002. "Equality of Opportunity and Equality of Outcome: Static and Dynamic Competitive Balance in European and North American Sports Leagues" in *Transatlantic Sport.* Carlos Barros,

Muradali Ibrahim and Stefan Szymanski, eds. Cheltenham, UK: Edward Elgar.

Szymanski, Stefan and Tommaso Valletti. 2002. "First and Second Prizes in Imperfectly Discriminating Contests," work. paper, Business School, Imperial College, London.

———. 2003. "Open and Closed Leagues," work. paper, Business School, Imperial College, London.

Taylor, Beck and Justin Trogdon. 2002. "Losing to Win: Tournament Incentives in the National Basketball Association," *J. Lab. Econ.* 20:1.

Taylor, Curtis. 1995. "Digging for Golden Carrots: An Analysis of Research Tournaments," *Amer. Econ. Rev.* 85, 4, pp. 873–90.

Tenorio, Rafael. 2000. "The Economics of Professional Boxing Contracts," *J. Sports Econ.* 1:4, pp. 363–84.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* MIT Press.

Tullock, Gordon. 1980. "Efficient Rent Seeking," in *Toward a Theory of Rent Seeking Society*. J. Buchanan, R Tollison and G. Tullock, eds. Texas A&M U. Press, pp. 97–112.

Vamplew, Wray. 1988. *Pay Up and Play the Game: Professional Sport in Britain 1875–1914.* Cambridge, UK: Cambridge U. Press

Vives, Xavier. 1999. *Oligopoly Pricing*. MIT Press

Vrooman, John. 1995. "A General Theory of Professional Sports Leagues," *South. Econ. J.* 61:4, pp. 971–90

———. 1996. "The Baseball Players Market Reconsidered," *South. Econ. J.* 62:3, pp. 339–60.

———. 1997a. "A Unified Theory of Capital and Labor Markets in Major League Baseball," *South. Econ. J.* 63, pp. 594–619.

———. 1997b. "Franchise Free Agency in Professional Sports Leagues," *South. Econ. J.* 63:July, pp. 191–219.

———. 2000. "The Economics of American Sports Leagues," *Scot. J. Polit. Econ.* 47:4, pp. 364–98.

Ward, Benjamin. 1958. "The Firm in Illyria: Market Syndicalism," *Amer. Econ. Rev.* 48:4, pp. 566–89.

Weiler, Paul and Gary Roberts. 1998. *Sports and the Law: Text, Cases, Problems.* 2nd ed. Amer. Casebook Series, West/Wadsworth.

Whitney, James. 1988. "Winning Games vs Winning Championships: Economics of Fan Interest and Team Performance," *Econ. Inquiry* 26:Oct., pp. 703–24.

Zimbalist, Andrew. 1992. *Baseball and Billions*. NY: Basic Books.

———, ed. 2001. *The Economics of Sport*. 2 vols. Cheltenham, UK: Edward Elgar.

———. 2003. *May the Best Team Win: Baseball Economics and Public Policy*. Washington: Brookings Institute.