Second Draft: not for quotation without prior permission. Comments welcome.

The Economic Design of Sporting Contests: A Review

Stefan Szymanski[1]

October 2002

I am grateful to Jeff Borland, Braham Dabscheck, David Forrest, Bernd Frick, Brad Humphreys, Erik Lehmann, Stefan Kesenne, Gerd Muehlheusser, Steve Ross, Rob Simmons, 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.

[1] Imperial College Management School, 53 Prince's Gate, Exhibition Road, London SW7 2PG, UK.
Tel : (44) 20 7594 9107, Fax: (44) 20 7823 7685, e-mail: szy@ic.ac.uk

I. 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 of 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 the most significant branches of the entertainment industry today. The US Census Bureau estimated that in 1997 spectator sports generated a direct income of $14bn domestically (a mere 0.17% of GDP), but that annual attendance at spectator sports in the same year equaled 41% of the adult population, while annual household television viewing hours runs into the hundreds of millions (40 million American TV households tuned into the 2001 Superbowl alone). In all probability hours spent in discussion at the water cooler dwarfs both these figures. 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, lawyers and politicians also express a 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[2] is already well known (see e.g. Hillman and Riley (1979)). Given the objective function of the organizer and the technology of the auction/contest it is possible to design an optimal prize scheme contingent on the distribution of contestant abilities/willingness to pay. But

[2] The words “contest” and “tournament” are here used interchangeably, as they are in much of the theoretical literature.

while there have been a number of reviews of the economics of sports in recent years (e.g. Cairns et al (1986), Fort and Quirk (1995), Kahn (2000), Vrooman (2000)[3]) none of these have attempted to explore systematically the design of sporting contests. Since sporting contests might be thought of as the archetypal instance of an economic contest/tournament, there is a case for developing this relationship more fully, and that is the goal of this review.

This is perhaps an unfamiliar 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 the rules of the competition. In a recent review for this journal Fort and Quirk stated "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. However, many sporting contests are centrally co-ordinated 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 US Open Golf Championship. What all these contests have in common is a need to provide contestants with the appropriate incentives to participate and perform, and joint decision making through a cartel is simply one (possibly inefficient) mechanism among many to achieve this end.

This review attempts to systematize the contribution of economic thinking thus far to design issues in sports, and relate this research to the growing empirical literature on sports. However, this is an enterprise still in its infancy 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[4].

[3] See also Zimbalist (2001) for a useful collection of seminal articles in the sports literature.
[4] This paper can thus be distinguished from fields such as "sabermetrics"- the study of baseball statistics for their own sake. Admirable as such studies might be, they have no place in the economics journals.

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 bull fighting in Spain, the Palio in Siena, the precursors of Eton Fives, the original native American sport of lacrosse and various sports found in traditional cultures around the world (see e.g. Guttman (1998)). Not surprisingly, commercialization has largely involved the former rather than the latter. This paper deals primarily with 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), the Olympics (1896), motor sport (1895)[5].

Historians (see e.g. Mason (1980), 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 for whose interest the various sports and sports organizations were able to compete. It is probably for this reason that most modern sports were formalized either in Great Britain (the first industrialized nation) or the United States (the most rapidly industrializing nation of the late nineteenth century)[6].

In this paper we draw the distinction between *individualistic* sports (such as tennis, golf and boxing) and *team* sports[7], such as soccer and baseball[8]. The distinction rests on the

[5] All of these dates are subject to controversy but while earlier versions of these sports may have existed, few had proper rulebooks before this date. The only clear exceptions are golf, cricket and horse racing- where rules and clubs existed from the mid-eighteenth century.

[6] The other great industrial nation of the period, Germany, also developed its own sporting activity during this period, the gymnastic "Turnen" movement; but this movement eschewed competition and was ultimately ousted by the competitive Anglo-Saxon sports (see Guttman (1998), chapter 7).

[7] 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 co-operation between horse, trainer and jockey.

[8]One apparent difference between team sports and individualistic sports is that in the former the outcome of a match depends on the interaction of efforts among the players on the same team whereas in the latter

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 will 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. In this sense an individualistic sporting contest conforms naturally to the standard contest model, outlined in the next section, in which the organizer is principally concerned with designing a contest that will attract the best athletes who will then contribute maximum effort to winning the contest. Section III reviews the contribution of the empirical literature to testing the predictive power of contest models.

The demand for team sports is more complex. Firstly, it is useful to note that while the organizational structure of individualistic sports is fairly uniform (there is little difference between the organization of, say, the New York Marathon and the Berlin Marathon or the US and British Open Golf Championships), the organization of professional team

---

there are no team production issues. Thus in team activities output depends not only on the marginal product of each player, but also on the derivative of player marginal product with respect to the effort of other team members, and this adds an extra dimension to the analysis of marginal productivity. However, the significance of this may not be all that great. In many team sports such as baseball and cricket team member 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. There exists a substantial empirical literature concerning the estimation of sports team production functions (see e.g. Kahn (1993)), but this literature lies beyond the scope of the current review.

sports differs substantially on either side of the Atlantic. Section IV 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 see the best players, the nature of team sports "fandom" is that supporters tend to attach themselves to teams rather than the players, and teams identify themselves with particular locations[10]. In practice this 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[11]. To prevent this from happening, they argue, it is therefore necessary to design the contest in such a way that all teams have roughly equal chances of winning, or that at least that all teams win occasionally[12]. The competitive balance issue has tended to dominate the analysis of team sports and section V sets out some empirical evidence on competitive balance and related issues in North American and European team sports.

Section VI 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

[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.
[11] It is an interesting question to ask whether this happens in individualistic sports. For example, many fans were attracted to tennis when Bjorn Borg and John McEnroe were at their peak- but when they retired did the same fans continue to follow tennis? Tiger Woods has created many new golf fans- will they remain fans when his career ends?
[12] The Blue Ribbon Panel on Baseball Economics (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" (p1).

on competitive balance. Empirical evidence on the first of these propositions is discussed in section VIA while section VIB considers the theoretical basis of the second.

Section VII discusses other mechanisms that can be 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 this section. The underlying objectives of the organizers of team sports has been a consistent source of controversy over the years. Section VII 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 a economic or commercial dimension. Section VIII provides a brief review of antitrust issues on both sides of the Atlantic. Section IX concludes.

II. The Design of Individualistic Sporting Contests

It is relatively straightforward to apply contest theory to the design of an individualistic contest. Consider a simple foot race organized by a profit maximizing entrepreneur (e.g. the owner of a race track). The organizer may generate a profit by selling tickets, broadcast rights or selling refreshments and merchandise, or some combination of these. The organizer expects that spectators will be attracted by the quality of the field entering

the race and the effort they contribute. Thus the objective is to design a incentive mechanism to maximize the effort contribution of the selected entrants. The simplest kind of structure is a winner-take-all prize.

A. The symmetric winner-take-all contest

The winner-take-all contest has been applied to a number of economic problems and originates with Tullock's (1980) model of a rent-seeking contest. This literature has developed in the context of, inter alia, labor markets (e.g. Lazear and Rosen (1981)), competition for innovation through R&D (e.g. Loury (1979) and competition for research contracts (e.g. Taylor (1995)). There is also a substantial related literature on all-pay auctions, where effort contributed to winning the contest can be treated as a bid (see e.g. Hillman and Riley (1989))[13].

The organizer's program can be written as

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

subject to

(2) $p_i(e_i^*)\, V - e_i^* > p(e_i)\, V - e_i$, *for all* e_i (incentive compatibility)

$p_i(e_i^*)\, V - e_i^* > 0$ (individual rationality)

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. The cost of effort is assumed to be linear with marginal cost equal to unity. Equation (2) states that each contestant selects their optimal efforts (incentive compatibility) and that all contestants willingly participate (individual rationality). The total payoff to each contestant depends on the

[13] Theoretical research on the implications of rent seeking contests includes Baye et al (1997), Dasgupta and Nti (1998), Dixit (1987), Green and Stokey (1983), Higgins et al (1985), Nalebuff and Stiglitz (1983), Nitzan (1994), Nti (1997), Skaperdas (1996).

probability of success (p_i) multiplied by the value of the prize (V), less the cost of investment or effort. It is assumed that the contestants are risk neutral[14].

In order to identify a solution to this problem it is necessary to define the technology of winning. This is typically known as the Contest Success Function, 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 Contest Success Function is the logit function

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

where e is measure of investment or effort, and γ is a measure of the discriminatory power of the contest success function. 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, is widely recognized as one of principal factors behind the demand for sporting contests (e.g. Neale (1964)).

[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 in 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.

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

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

From which it is apparent that

(i) Individual and aggregate effort is increasing in the value of the prize

(ii) Individual and aggregate effort increases with the discriminatory power of the contest success function

(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 non-economists. 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. Fullerton and McAfee (1999) report a similar result for an asymmetric contest, as long as the best two contestants are selected (which they will be in their model). If the organizer

[15] The principal alternative to the logit function for imperfectly discriminating contests is the probit explored by, inter alia, Lazear and Rosen (1981) and Dixit (1987). As yet the probit model has not been applied to the analysis of a sporting contest.

[16] Here and in the following sub-sections we consider only pure strategy equilibria, although in sub-section F on dynamic contests mixed strategies are also considered. A mixed strategy may exist even if a pure strategy equilibrium does not, see e.g. Baye at al (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.

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.

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].

B. Multiple prizes in symmetric contests

In practice, most organizers of sporting contests do not offer a winner-take-all prize: as well as gold medals there are silver and bronze. Why should this be optimal? A recent paper by Moldovanu and Sela (2001) shows 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

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

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 winning the second

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

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

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

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

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

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 = ½ (the second prize is identical to the first prize) but will be positive for all values of $k \in [0, 1]$ for n > 2.

The intuition behind the negative incentive effect of second prizes in a symmetric contest is that the second prize has a smaller incentive effect that the first prize because the value of the second prize only kicks in when the first prize cannot be won, making it a more restricted opportunity. To see this we can write the first order condition of the contestant in equilibrium as

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

$$(8)\qquad \left[\frac{\partial p_{i1}}{\partial e_i}k+\left((1-p_{i1})\frac{\partial p_{i2}}{\partial e_i}-p_{i2}\frac{\partial p_{i1}}{\partial e_i}\right)(1-k)\right]V=$$
$$\left[\frac{n-1}{n^2}k+\left(\frac{n-2}{n(n-1)}-\frac{1}{n^2}\right)(1-k)\right]\gamma V=e_i$$

We can divide this expression into three parts. The first is the marginal incentive effect of the first prize, the second is the marginal incentive effect of the second prize, which is in fact larger than the first term when n > 2. However, the third term is the reduction in effort to win the first prize caused by shifting a unit of the prize fund from first to second prize, and this causes an increase in the share of the second prize to reduce total effort.

C. 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[20]. Secondly,

[19] Competitive balance is discussed in more detail in section V. But it is worth observing that the merits of a balanced contest are not always obvious: when it came to Christians and lions, the Roman public demonstrated little interest in competitive balance.

[20] So far we have assumed the overriding objective is profit maximization. But in some sports it is frequently argued that profit maximization is not the objective of the organizers (most notably see Sloane (1971)). As far as the design of competition is concerned this may not make much difference. For example, amateur sporting associations frequently use revenues from sporting competition to promote the sport to some target audience such as children. In that sense the objective may still be to maximize profit from a sporting event that spectators pay to watch and the assumption may not be problematic. However, in the case of team sports, the assumption of profit maximization among the contestants is a significant issue, when contestants can re-invest earnings into future success. For contestants in individualistic sports such as

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.

Asymmetry has been little studied in the contest literature (two notable exceptions are Dixit (1987) and 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

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

(9)

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

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 (9), assuming the logit contest success function (3), imply that the effort ratio in equilibrium is

$$\frac{e_2}{e_1} = \frac{1-\beta}{1+\beta} \qquad (10)$$

---

golf, tennis or Olympic sprinting, this issue impinges only on the possibility that excessive effort is put into winning prizes by, saying, taking banned substances or cheating in some other way. Little has been written on the economics of cheating in this sense, although a recent paper by Duggan and Levitt (2000) illustrates the potential for research in this area. Yet another possibility, raised by Lazear (1989) is that tournaments increase the incentive to try and undermine the performance of rivals rather than improve one's own performance, i.e. sabotage. Garicano and Palacios-Huerta (2000) have examined this proposition for the case of soccer, where a change in points system appeared to lead both to more creative effort and more sabotage, meaning here rule infringements.

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 in a homogenous contest with fixed costs or a heterogeneous 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

$$(11) \qquad e_1 + e_2 = \frac{2\gamma V(2k-1)z^{\gamma}}{(1+z^{\gamma})^2(1-\beta^2)}$$

When $\gamma = 1$ it is clear (from (11)) that $e_1 + e_2 = \gamma V(2k-1)/2$, 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, p858) demonstrate similar results in a rank order labor tournament where the contest success function is asymmetric, but in their model the effect on total effort depends on the concavity or convexity of costs.

D. 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 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[21]. 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[22].

A second prize may also raise 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 the following combinations of effort contribution produce the same variance: (a) 1, 2, 3 and (b) 1, 1, 2.7. In case (a) the there is an equal gap between each contestant while in case (b) there is a larger gap between the strong player and two equally weak players. The race for first place may be more exciting in case (a), but even then the strong player has a big lead over the second best. By contrast case (b) will at least produce a close race for second place, which may compensate for a lack of tension

[21] The same effect would be observed if one contestant was given, say, a 50m headstart in a 100m dash.
[22] 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

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.

E. Matchplay

In many sporting contests the organizers must make a structural choice between matchplay 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 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 matchplay 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 non-decreasing.

Gradstein and Konrad (1999) compared simultaneous contests (which they labeled S-contests) and matchplay 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

---

context it might be viewed as simple recognition of the need for motivation for all contestants when abilities are heterogeneous.

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 multi-stage contest, in which contestants have to put in additional effort at each stage, can dissipate rents[23].

F. 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, pre-emption games associated with patent races (both these types of game are reviewed in Fudenberg and Tirole (1991)) and market share attraction games in the advertising literature (see Monahan and Sobel (1994)). These have some implications for contests which 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 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

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

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. Bulow and Klemperer (1999) show that if there are N+K contestants for N prizes and each contestant contributes effort only to the point at which she quits the game and each contestant values the prize differently (or equivalently has differing probability of success or cost of effort), then the unique equilibrium is for the K-1 contestants with the lowest valuation to exit immediately while the N+1 highest valuation players remain and play their mixed strategy. By contrast, if players must continue to contribute until the prize is won (regardless of whether or not they "give up"), then all players choose their mixed strategy, with the result that more players remain in the game at any point in time- in other words players are more committed. In a war of attrition all the rents associated with winning the prize are competed away by the expenditures of the contestants, and in the latter version the total value of effort expenditures can be far in excess of the expected value of the prize.

In the war of attrition contestants learn nothing from their continued participation in the game (the game is memoryless). In pre-emption games (e.g. Harris and 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 ε-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". However, this extreme result is sensitive to assumptions about information sets, and if there is some uncertainty about the state of play then the follower might have an incentive to "leap-frog" ahead of the leader (e.g. Harris and Vickers (1987)). Tirole (1988) uses the example of a foot-race to explain these results:

"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."

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[24]. 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.

III. Empirical Research on Individualistic Sports

The research agenda in individualistic sports has been principally driven by the positive research agenda of contest/tournament theory. Researchers in that literature have set out to explain a wider class of phenomena observed in labor and product markets, notably the widespread use of prizes as an incentive device. Thus the primary goal of the theoretical

[24] 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.

literature has been to develop models which can identify the optimal reward structure for given set of economic conditions[25].

The discussion of the previous section on contest design in theory can be summarized under the following points:

(i) In all contests effort (individual and total) is increasing in the prize fund

(ii) In a symmetric logit contest total effort is increasing in the number of contestants and the discriminatory power of the contest. Individual effort is decreasing in the number of contestants

(iii) Second prizes are never optimal in symmetric contests

(iv) In asymmetric contests pre-screening and handicapping are plausible methods for maintaining a competitive balance

(v) In an asymmetric two person contest (a) the effect of asymmetry on total effort depends on the discriminatory power of the contest (high power leads to lower effort as asymmetry increases).

(vi) In an asymmetric n-person contest (n>2) second prizes can increase the effort of all contestants if the asymmetry is great enough. Second prizes might also be desirable to enhance competitive balance even if effort falls.

(vii) Matchplay is preferred to simultaneous contests when the discriminatory power of the contest is low

(viii) In dynamic contests laggards are likely pull out, particularly if performance levels are common knowledge.

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 North American 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

[25] Normative issues have played a role, and theorists in this area have sought to apply their findings to operation of government sponsored contests with some success (e.g. in such areas as spectrum auctions or research contracts), but this normative work has been a consequence of a positive research agenda.

and relate that data to the prize structure offered in individual tournaments. Perhaps the best known results are those of Ehrenberg and Bognanno (1990a, b) who examined scores in American and European PGA golf tours. As Lazear (1995, p33) comments "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".

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. Major golf tournaments typically involve four rounds played over a number of days, and after two days the contestants who fail to reach a minimum standard are eliminated; the standard required being known as the "cut". The authors controlled for sample selection bias in the final round scores of contestants (only those who made the cut had a final round score), since they had scores for the first two rounds as well. This made no difference to their results. However, they did find that scores after only two rounds were not significantly reduced by the size of the prize fund (although the coefficient did have the right sign). This may suggest that in the early part of tournament the impact of the prize fund is relatively small because it is still a remote prospect, and the contestants may be more influenced by the status of making the cut. Finally, Ehrenberg and Bognanno 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 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

[26] Orszag (1994) was unable to replicate these findings using data on the 1992 US 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.

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 Lynch and Zax (2000) who examine data on nearly two thousand contestants covering 135 different road races in the US ranging between 5 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.

Maloney and McCormick (2000) use data on 115 foot races ranging between one mile and a full marathon involving nearly one thousand five hundred 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 the have the expected sign, even though in general the prizes seem quite small (a mean of around $400) compared, say, to golf. In terms of economic significance the impact might also be considered large, since doubling the prize spread reduces race times by 4%. However, one possible weakness of these foot race studies is that in general the contestants do not include a significant fraction of the world's best, which is reflected in the average times of the sample. Frick et al. (2001) consider sample of 57 marathons run world-wide and involving much larger prize money ($135,000 per race in 1993 dollars). They consider the impact of both the total prize fund and the distribution of the prize fund, as well as bonuses for achieving a fast time (rather than rank). They find that (a) doubling the average prize reduces average times by 1%, (b) doubling the spread improves average times by 2%, (c) doubling bonus payments improves average times by around ¾% (d)

increasing the prize fund, spread and bonuses also increases the closeness of the race, measured as the time difference between the winner and other finishers (e) race times are decreasing in the number of "in the money" ranks- i.e. the larger the number of prizes offered. The simultaneous use of both relative and absolute performance incentives in marathons suggest an interesting trade-off. One might expect that with risk neutral contestants and a objective of maximizing winning effort a winner take all contest would be the most efficient instrument. One explanation for specific performance incentives may be the threat of tacit collusion among the leading contestants. The effect of the number of prizes is also striking, suggesting that asymmetry may be significant enough to warrant more prizes to elicit more effort and a closer, more exciting race.

Apart from foot races and golf, almost the only other individualistic sport to have produced some empirical research is horse racing[27]. Fernie and Metcalf (1999) examined the effect of a change in the compensation of British jockeys which involved replacing performance related payments with non-contingent 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.

Finally, Maloney and 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

[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] Becker and 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. 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, but 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.

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 sample of 112 sponsors of motorcycle races.

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. That more work has not yet been done is perhaps surprising given the obvious points made above about the quality and lack of ambiguity in the data generated by sports events. 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 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.

Yet there is scope to consider more complex design issues in this area e.g. the value screening, the role of handicapping, contest structure (matchplay and simultaneous contests) and discouragement effects, the impact of penalties (e.g. failing to meet the cut), the impact of qualifying races, cheating, sabotage and possibly other issues. Moreover 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. This also appears to be a significant aspect of the development of the soccer World Cup run by FIFA, the governing body of the sport. Until the 1970s European and South American teams were awarded a share of qualifying places more than proportionate to their numerical strength. From that period on the newly elected President worked to restrict the number of qualifying places allocated to European teams, where soccer was strongest, and increase

the relative share of the African teams, among others (Sugden and Tomlinson (1999)). Since then African teams have been increasingly successful in the tournament, perhaps as a consequence of the President's design.

IV. The Comparative Economics of Team Sports

A. Peculiar Economics

The analysis of team sports has been primarily motivated by normative issues[29]. 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 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 seem to express opposition to player trading in principle[30].

[29] 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 Rosen and Sanderson (2001))

[30] Idson and Kahane (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.

It is possible that different attitudes may reflect broader cultural differences, but it may also be that historical accident and path dependency[31] may account for different practices. 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, which are themselves organized in quite different ways.

Cultural particularities may sometimes lead to the perception that there is only one right way to do things, but in fact different kinds of equilibria may be sustainable. 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.

B. Baseball

Seymour (1971), the authoritative historian of early baseball, makes 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 US, with teams and competitions proliferating and vying to attract spectators. The barn-storming 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.

[31] Clearly tradition and folk memory is 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?

The founders of the National League perceived this to be an unstable equilibrium that would result in the loss of spectator interest in the game, as well as being an unsatisfactory equilibrium for team owners. Thus the National League was a deliberately elitist affair, intended to invest the members with a stake in its long term success (to combat short run incentives for match fixing), to create exclusive territories (the incentive to invest in local markets) and to establish monopsony rights over the players (the Reserve Clause). The extraordinary success of this model made it not only the basis for the national sport of the US, but also for the other North American team sports (football, basketball and ice hockey), America exported this model to other countries (e.g. baseball in Japan and Mexico, basketball in Australia) and to sports in other countries influenced by the US (e.g. Australian Rules Football in the 1970s). In particular, while other team sports in the US 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
7. low player mobility and limited player trading, 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 have been adopted to a greater or lesser extent, but are present to some degree to in all the major leagues. Some other types of agreement, such

as gate revenue sharing (MLB[32] and NFL) and salary caps (NBA, NFL) have not been universally adopted, but are not be 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 US, most notably in the case of soccer, arguably the world's most popular team sport.

C. 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 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[33]. Unlike the founders of the National League, the founder of the Football League did not break away from the existing structures, but worked inside them. This meant (a) the Football League never attempted to become an exclusive institution, but set out to admit all the major team 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

[32] In the 1990s MLB ceased sharing gate revenues only in favor of local revenue sharing (including TV income).
[33] The first FA Cup final and the first international match (Scotland v. England) both took place in 1872.

authorities, (ii) a domestic league system incorporating promotion and relegation[34] and (iii) a system where star players are paid employees of clubs and compete for them primarily in league competition, but are also representatives on the national team whose success is usually seen as even more prestigious. This system applies not only to soccer world-wide, but has also been applied to a number of other team sports, mostly in Europe where the soccer system is dominant (e.g. rugby union and basketball in Europe). Comparing the distinctive institutional features of soccer with those of the North American sports, in each dimension soccer's structure shows material differences:

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 of the hierarchy, but promotion on sporting merit only
4. Non-exclusive 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[35]
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

[34] This is a structure in which clubs affiliated to the governing body are eligible for promotion 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.

[35] It is perhaps more historically accurate to say that unions were relative 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) chapter 4). However, in North America the role of the unions has grown significantly over the past 40 years, while in Europe they remain relatively weak to this day.

But perhaps an even greater 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 has driven down profits to the point where either leagues have folded or mergers have been agreed. 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 US (e.g. Quirk and Fort (1992), 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. Noll (2002) and Ross and Szymanski (2002) have proposed the adoption by the major leagues of the promotion and relegation system which is commonly practiced in Europe. Extending the analysis of team sports to assess the real effect of the strikingly different institutions of soccer offers a rich laboratory for researchers.

V. Team sports, uncertainty of outcome and competitive balance

One noticeable feature of the difference between the organization of soccer and baseball is the range of restrictions on behavior in baseball (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) that either do not exist in soccer or exist only in an attenuated form. Moreover, North American sports have copied baseball's restrictions to a greater or lesser extent[36] while most European sports have a format not too dissimilar from soccer[37]. It is clear from the antitrust record that the North American sports have justified these restrictions on the basis of a series of propositions that relate demand, uncertainty of outcome and competitive balance. These can be stated as follows:

1. Inequality of resources lead to unequal competition
2. Fan interest decline when outcomes become less uncertain
3. Specific redistribution mechanisms produce more outcome uncertainty

These propositions have defined the empirical research agenda of team sports economics, and much of the theoretical research has also concentrated on related issues. This section reviews the empirical literature on the first two these the propositions. The third will be

[36] In some cases the restrictions have been adopted by baseball from other North American sports.
[37] There are exceptions: in the UK Rugby League has adopted many American style restrictions. The case of Australian team sports in interesting, since these had structures resembling European sports until 1980s but since then a number of American institutions have been adopted (see e.g. Dabscheck (1989), Hess and Stewart (1998)).

considered in section VI where individual measures to improve competitive balance are considered.

A. 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 contest success function. 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)[38]. 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

[38] See also Forrest and Simmons (2001b) for a similar analysis.

measured by the R^2 is also larger, most notably in the American League (home of the Yankees)[39].

Table 1: Pay-performance sensitivity estimates

League	α	β	σ_{wpc}	σ^*_{wpc}	σ_{RW}	R^2	Period	Obs
North American								
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	.09	.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

Estimated equation: $wpc_{it} = \alpha + \beta\ RW_{it} + \varepsilon$. RW is wage spend of a team relative to average wage spend 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/√m, where m is the number of matches played by each team). European data refers to the top division only. Premier League was called the Football League Division One until 1992.

The apparent explanatory power of the regression for the European soccer leagues of England, Italy and Germany 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 US about growing imbalance in baseball, that the variation of wages and the R^2 of the regression are only noticeably

[39] The degree of sensitivity reported here seems much greater than that reported by other authors e.g. Quirk and 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 as that

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. However, in more recent years there may have been a trend toward increasing predictability (see Hall et al. (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 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. Firstly, to justify running a simple levels regression of this type it is necessary to establish that the variables are stationary. Adopting a panel approach to the estimation (estimating a separate intercept for each team), Hall et al (2002) found that win percentage and relative spending were stationary for English league soccer and MLB[40]. They then 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 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

described above and concludes that "average team salary has been related only tenuously to team performance."

[40] One difference between the data presented here and Hall et al is that in the latter the English data covered all four divisions and the dependent variable was league rank (1 to 92). Szymanski and Smith (1997) discuss the validity of this approach. Hall et al. also used panel data methods for their estimation (implying a separate intercept for each team).

requires the collection of a wider range of potential explanatory variable for MLB, is an important subject for future research, as is the nature of causality in other leagues[41].

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 Kipker (2000) and Forrest and Simmons (2002b) for a more 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[42].

However, the standard deviation of winning percentage may be a relatively poor measure of competitive balance, largely because it only considers performance within a season. Performance in the open European leagues tend to be relatively bunched together, since teams near the bottom keep competing right to the end in order to avoid relegation. Over a number of seasons the same big teams tend to dominate European competition. 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 Lucke (1997))[43].

[41] There have been relatively few attempts to analyze causality empirically in the sports literature. Two exceptions are Davies et al (1995) and Dobson and Goddard (1998), who look at the relationship between income variables (attendance and revenues) and success in English rugby league and soccer.

[42] 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 (Horowitz (1997)) and the Hirschman-Herfindahl index (Depken (1999)). Humphreys (2002) proposes a measure of competitive balance based on the ratio of the sum of standard deviations of team performance through time to the sum of within season standard deviations of win percentage.

[43] Balfour and 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 Smith (2002) adopt this approach to compare persistence across North American and European leagues.

Buzzachi et al. (2001) 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 are 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.

B. Demand and uncertainty of outcome

Whatever the causes of inequality, the lynch-pin 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, much of which has been discussed in earlier surveys (e.g. Cairns et al (1986)). 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. Knowles at al (1992), Forrest and Simmons (2002a) summarizing the work of Peel and Thomas (1988, 1992 and 1997) and 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. Indeed, since this is only the turning point, it seems plausible to suppose that a match could become even more unbalanced without significantly reducing attendance. 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.

Evidence on seasonal uncertainty is by and large much less conclusive, and less work has been done in this area. The key problem in this area is controlling for the all the other relevant factors that might influence demand. For example, Schmidt and 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 barn-storming teams). It may be that the causal relationships are too complex to isolate a single influence such as uncertainty of outcome.

Table 2: Outcome Uncertainty In The Literature

Authors	Testing	Uncertainty measure	Data	Result
Noll (1974)	Seasonal	- Whether team in contention for playoff -Whether championship race close	Ice Hockey Baseball	Weak support Weak support
Hart et al. (1975)	Match	-Log difference in league positions	4 English Football Clubs 1969/70-1970/1971	Weak support
Jennett (1984)	Seasonal	- Championship/relegation significance of each game	Scottish League Football 1975-1981	support
Borland (1987)	Seasonal Championship	-Diff. in games won between first and last -Sum of coefficients of variation of game won -Average no of games behind the leader -Number of teams in contention	Victorian Football League (Australian Rules) 1950-1986	Weak support No support
Cairns (1987)	Seasonal	-Dummy of contention in championship	4 Scottish football clubs 1969/70-1979/80	support
Jones & 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 1970-1984	Weak support
Peel & 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

Authors	Testing	Variable	Data	Result
Borland and Lye (1992)	Seasonal	Sum of matches required to qualify for the finals	Australian Rules	No support
Kuypers (1996)	Match Seasonal	Betting Odds (difference in max & min) points & games left	1993/1994 Individual English Premier League Matches	no support support
Peel and Thomas (1997)	Match	-Betting odds (points spread)	Rugby League 1994/95	Support
Baimbridge et al (1996)	Seasonal	Dummy when both teams in top (bottom) four positions	1993/1994 Individual English Premier League Matches	no support
Rascher (1999)	Match	-Betting odds (probability of home win)	MLB 1996	Support
Szymanski (2001a)	Championship	Competition type (with identical contestants)	English League and FA Cup matches 1977-1998	Support
Schmidt and Berri (2001)	Seasonal	Gini coefficient	MLB 1903-1998 (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

Adapted from Szymanski and Kuypers (1999).

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[44]. 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 twenty-two 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.

VI. 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 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' interest) and employment conditions and remuneration (the players' interest). 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.

[44] One exception is Szymanski (2001a) 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.

A. 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))[45]. This inevitably led to challenges in the courts by the players claiming the right to move freely between employers. 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 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 (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)

[45] 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.

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*. In it original context Rottenberg argued that player trading rules such as the Reserve Clause would not affect the distribution of talent, as long as trading was permitted. El-Hodiri and Quirk extended the invariance principle to another policy instrument, gate revenue sharing.

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 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 Staudohar (1996) for more details on both of these innovations). These changes represent a natural experiment that can be studied to identify the impact of changes in talent allocation rules on competitive balance.

(i) Free Agency

The advent of free agency in MLB in 1976 for six year veterans is a clear natural experiment. 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), 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.

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	(+)	(+)

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. Daly (1992) observes that under the Reserve Clause top line players were seldom traded, a situation that has been affected by free agency where the top stars have a choice after six years leading to increased mobility. Hylan et al (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, Cymrot et al (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 service) enjoyed more or less mobility than unrestricted free agents after controlling for player characteristics.

Marburger (2002) considers a different implication of the invariance principle. If two leagues exist in parallel and players can be traded across 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 examines the distribution of trades between the National League and American League between 1964 and 1992 and does indeed find a statistically significant increase in the share of intraleague trades, from 60% to 73%. This finding then, seems consistent with the invariance principle.

An interesting difference between European soccer and North American sports is that player trading for cash has always played an important role in the development of the market, and there have been no restrictions on trading such as those which have 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 so that all players had the right to move club once their contract ended (contracts rarely being for longer than three years), but with the selling team entitled to substantial compensation (i.e. well in excess of any likely compensation 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 March 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".

(ii) 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. Interestingly, 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

[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, who were 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 so 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).

improved over time (e.g. measured by the Gini coefficient for pennants) in both in the Central and Pacific Leagues (although the within season measure (standard deviation of win percent) was significant only for the Pacific League). As they point out, these results are "virtually identical" to Fort and Quirk's (1995) results for MLB. Grier and 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. This, then, seems to suggest fairly strong and 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 (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)[47]. Fees and 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.

[47] Szymanski and Kuypers (1999) chapter 4 discusses the evolution of the player market in English soccer in some detail.

B. The Invariance Principle and gate revenue sharing

As mentioned above El-Hodiri and Quirk (1974) extended the invariance principle to another policy instrument, gate revenue sharing, i.e. they claimed that a change in the percentage of gate revenues allocated to the visiting team (between 100% and 50%) 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 sharing.

The empirical evidence in section V supports to some extent to conventional approach to the modeling of league competition. Firstly, it is normally assumed that teams choose investment in playing talent which 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. Playing success, usually represented in terms of win percentage, is generally assumed to follow a logit specification. The main difference, then, between a conventional contest model and a team sports model is that instead of each receiving a fixed prize with some probability determined by relative investment, each teams generates a revenue dependent of the share of matches won, where that revenue also varies according to the revenue generating capacities of the teams. However, the award of prizes is also possible in team sports, and this is considered below. Asymmetry in team sports arises not from differences in the cost of acquiring talent, but in differences in the level of revenue generated by any given quantity of talent hired. This asymmetry, it is argued, justifies intervention to restrict the choices of individual teams.

This raises the question as to how decision makers interact. Fort and Quirk, among others, support the cartel interpretation, while some maintain that leagues should be

considered (at least for antitrust purposes) as single economic entities (e.g. Roberts (1984)). Gilbert and Flynn (2001) suggest that the most natural treatment of a league is as a joint venture. From the point of view of contest design, it is reasonable to think of teams as contestants competing for a prize. The structure of league competition is, obviously, based around matchplay, and contestants generate their own income from matches they play. However, what distinguishes league competition from the kind of barn-storming matchplay 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. This suggests that the contest/tournament analogy is indeed appropriate for the modeling of league competition, where the league commissioner or governing committee of the league is the principal and the teams are the agents. The precise legal format adopted, however, may vary. Conventionally teams are joint owners of the league and delegate an official to manage collective negotiations. However, the example of Major League Soccer has demonstrated that it is possible to create a successful team tournament based on a single entity.

Analysis of the invariance principle is only relevant when there are asymmetries between 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 contest success function takes the same form as in an individualistic contest (3) and that $\gamma = 1$ we can write

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

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[48]. Team sports models frequently cite the "adding up constraint" that the sum of win percentages in a league must equal n/2. This is equivalent to the constraint in a standard contest model that the probabilities sum to unity. Obviously this condition is satisfied by (3'). Another way of expressing the adding up constraint is

$$\frac{\partial p_1}{\partial e_1} = -\frac{\partial p_2}{\partial e_1} \quad and \quad \frac{\partial p_2}{\partial e_2} = -\frac{\partial p_1}{\partial e_2} \tag{12}$$

Note that the contest success function (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. The principal difference when n > 2 is that it is possible to specify each team's revenue function not merely as a function of its own win percentage but also as a function of rival team's win percentage, which may be thought of as introducing the possibility of complementarities between the teams. Although this would suggest a more complex set of interactions between teams than is modeled here, the existence of production externalities does not imply that the team's decision problem will be fundamentally altered, since even in the simple function considered here each team's investment produces a negative externality. In particular, the existence of externalities, negative or positive, implies that private decision making will not necessarily be socially efficient, and this conclusion is not altered by the existence of positive externalities[49].

[48] Baik (1994) models asymmetry by assuming that the sensitivity of contest success function to effort differs among contestants, an assumption which 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).

[49] For n > 3 the contest success function can be thought of as a championship success function (e.g. Whitney (1988)) In practice, the difference between the share of totals 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.

Revenue generated by each team is assumed to be a function of contest success. 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[50]

The interaction of these three factors will give rise to some general revenue generating function R(.). The requirement of tractability demands some simplification. Firstly we will for the moment ignore the impact of the demand for quality. Intuitively, if this enters the revenue function of each team symmetrically then it will shift out the demand for talent without any distributional consequences. However, some consequences of including the interaction of quality in more complex cases are considered below. Secondly, we 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 contest success function and the assumption that team revenues have a unique maximum (e.g. at a winning record that lies between 0% and 100%). To focus on the trade-off between the two elements we assume that revenues are simple linear functions of these variables.

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

(13)

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

[50] Following most of the literature, we abstract from price issues. In North America teams are located at such great distances from each other, principally due to territorial exclusivity rules, that monopoly pricing is a plausible assumption (see e.g. Alexander (2002)). However, in Europe where there is no territorial exclusivity monopoly pricing seems less likely. The New York SMSA has two major league baseball teams for a population of 20 million. London's population of 13 million houses 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). Density clearly has important implications for revenue generation and competitive

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 (e.g. win percentage). Note that $p_1(1-p_1) = p_2 (1 - p_2)$ is a measure of competitive balance and that λ is a parameter intended to capture 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 $\lambda = \mu = 1$ then problem is isomorphic to the symmetric winner-take-all contest of section IIA[51]. The owners of each team are assumed to be profit maximizers. Under these assumptions the first order conditions are

$$\begin{aligned} \frac{d\pi_1}{de_1} &= \frac{\partial R_{11}}{\partial p_1}\frac{\partial p_1}{\partial e_1} - c = [1 - \lambda(1-\mu) - 2(1-\lambda)p_1]\frac{\partial p_1}{\partial e_1} - c = 0 \\ \frac{d\pi_2}{de_2} &= \frac{\partial R_{22}}{\partial p_2}\frac{\partial p_2}{\partial e_2} - c = [1 - 2(1-\lambda)p_2]\frac{\partial p_2}{\partial e_1} - c = 0 \end{aligned} \tag{14}$$

These expressions state that owners invest in talent to the point where the marginal revenue from a unit of talent equals its marginal cost (= c). 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, p1271) assume "a one unit increase in t_i yields the same increase in win-percent for any level of win-percent" and

balance which do not seem to have been fully explored in the literature. Forrest et al (2002) explore the spatial pattern of demand for English soccer.

[51] In the one-shot winner-take-all model the payoff to the contestant is an expectation of the prize dependent on relative effort while only one contestant receives the prize ex post, while in the one-shot team sports version the each contestant generates an income based on the share of success so that expected

Vrooman (1995, p973) 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')

$$(15) \qquad \frac{\partial p_1}{\partial e_1} = \frac{(e_1 + e_2) - e_1\left(1 + \frac{de_2}{de_1}\right)}{(e_1 + e_2)^2}, \quad \frac{\partial p_2}{\partial e_2} = \frac{(e_1 + e_2) - e_2\left(1 + \frac{de_1}{de_2}\right)}{(e_1 + e_2)^2}$$

If we assume $de_1/de_2 = de_2/de_1 = -1$ then it will indeed be possible to normalize the total supply of talent to unity so 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 (12). Since the expression (15) 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. This implies that the talent supply for each league is fixed (at least in the short term), and therefore if one team hires an additional unit of talent there is one less unit for all other teams to hire[52]. But modeling a fixed talent supply by assuming non-zero conjectural variations has significant implications for the nature of the model's equilibrium. To find a

income equals ex post income (there is no stochastic element in the contest success function). In an infinitely repeated game with no discounting the expected and actual payoffs are identical in both cases.

[52] Although note that 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. This must be a possibility given the fixed rosters employed by most of these leagues. In long term

Nash equilibrium would require the assumption of Nash conjectures, namely $de_1/de_2 = de_2/de_1 = 0$. Without Nash conjectures the equilibrium will not be Nash.

To see the implications of this combine the two expressions in (14) to obtain

$$(16) \qquad \frac{\frac{\partial p_2}{\partial e_2}}{\frac{\partial p_1}{\partial e_1}} = \frac{\frac{\partial R_{11}}{\partial p_1}}{\frac{\partial R_{22}}{\partial p_2}} = \frac{1-\lambda(1-\mu)-2(1-\lambda)p_1}{1-2(1-\lambda)(1-p_1)}$$

Note that the left hand side of (16) 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 ($\mu = 1$)[53]. 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 - α to the visiting 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

supply must be at least to a degree elastic, as athletes are attracted from other sports by the salaries and talented athletes move into sport from alternative activities.

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

$$\frac{\partial \pi_1}{\partial e_1} = \alpha \frac{\partial R_{11}}{\partial p_1}\frac{\partial p_1}{\partial e_1} + (1-\alpha)\frac{\partial R_{22}}{\partial p_2}\frac{\partial p_2}{\partial e_1} - c = 0$$

(17)
$$\frac{\partial \pi_2}{\partial e_2} = \alpha \frac{\partial R_{22}}{\partial p_2}\frac{\partial p_2}{\partial e_2} + (1-\alpha)\frac{\partial R_{11}}{\partial p_1}\frac{\partial p_1}{\partial e_2} - c = 0$$

Which, using the adding up constraint (12) can be rearranged to obtain

(18)
$$\left[\alpha \frac{\partial R_{11}}{\partial p_1} - (1-\alpha)\frac{\partial R_{22}}{\partial p_2}\right]\frac{\partial p_1}{\partial e_1} = \left[\alpha \frac{\partial R_{22}}{\partial p_2} - (1-\alpha)\frac{\partial R_{11}}{\partial p_1}\right]\frac{\partial p_2}{\partial e_2}$$

If we now further assume fixed supply conjectures it should be clear that since $\partial p_1/\partial e_1 = \partial p_2/\partial e_2$ (18) collapses to the equality

(19)
$$\frac{\partial R_{11}}{\partial p_1} = \frac{\partial R_{22}}{\partial p_2}$$

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

(16')
$$\frac{\frac{\partial p_2}{\partial e_2}}{\frac{\partial p_1}{\partial e_1}} = \frac{e_1}{e_2} = \frac{\left[\alpha \frac{\partial R_{11}}{\partial p_1} - (1-\alpha)\frac{\partial R_{22}}{\partial p_2}\right]}{\left[\alpha \frac{\partial R_{22}}{\partial p_2} + (1-\alpha)\frac{\partial R_{11}}{\partial p_1}\right]}$$

It should be clear that the LHS of (16') is identical to that of (16) but when $\alpha < 1$ the RHS of (16) and (16') are not equal in an asymmetric equilibrium, 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

(20)
$$\frac{e_1}{e_2} = \frac{(1-2\alpha)(1-\lambda) + \lambda\left[\alpha + \mu(1-\alpha)\right]}{1-2\alpha+\alpha\lambda(1+\mu)}$$

Differentiating we obtain

$$\frac{\partial\left(\frac{e_1}{e_2}\right)}{\partial\alpha} = \frac{\lambda^2(1-\mu^2)}{[1-2\alpha+\alpha\lambda(1+\mu)]^2} < 0 \tag{21}$$

Thus under Nash conjectures revenue sharing will in fact make competitive balance worse. Szymanski and Kesenne (2002) 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[54].

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 (it is clear from (20) and (21) that the effect on competitive balance is independent of c) then it must be that strong drawing team will in fact increase it share of total talent while the weak drawing team reduces it share. If the supply of talent were elastic, however, this result need not necessarily hold, even though competitive balance must still be reduced. Whether supply is fixed or not, total expenditure on talent will fall with gate revenue sharing and total profits will increase.

This assumption of elastic supply seems more reasonable in the case of European soccer, where no national league is dominant in Europe as a whole and players move freely between leagues. In any case, the notion of a fixed talent supply can be modeled by looking for equilibrium in the labor market without requiring the fixed supply conjecture to be incorporated into the objective function of the owners. It seems widely accepted in

[54] Atkinson et al (1988) also state that they do not obtain the invariance result (p33, 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.

the broader economic literature that in a static game of this type only Nash conjectures make sense (see e.g. Vives (1999), pp 185-187) but alternative conjectural variations are sometimes defended as reduced forms of an underlying dynamic model. The original model of El-Hodiri and Quirk (1974) is indeed a dynamic model. The authors do not explain in detail the source of the invariance result but it appears to follow as 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[55]. In essence this seems to be a joint profit maximizing program, and it does indeed seem plausible that in such circumstances the optimal allocation of talent would be invariant to the way in which total revenues were distributed. However, it seems more natural 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[56], and therefore it is perhaps not surprising that a model based on these conjectures appears to support the Coase Theorem.

Finally, there is a further fundamental problem with 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, with fixed supply conjectures only n-1 teams can be decisive in an n team model, and the n^{th} team must accept the allocation of talent implied by the profit maximizing choices of all the other teams. With fixed supply conjectures the equilibrium cannot be Nash, since at least one team has no opportunity to select its best response. The solution to this conundrum, as in the case of market share models, it to 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.

[55] See Noll (ed, 1974) p63, equation (ii) in particular.
[56] Just as a conjectural variation of –1 produces the joint profit maximizing solution in a quantity setting oligopoly.

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[57]. 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.

Palomino and 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[58]. 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 downs 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.

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

[57] 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.

[58] 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. Dobson and Goddard (2001) talk of a “closed” North American model and an “open” European model.

championship in a season- the domestic league and European-wide league (e.g. The Champions' League[59])- 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.

VII. Other Design Issues in Team Sports

A. Prizes and lump sum revenue sharing

Fort and Quirk (1995) observe that while gate revenue sharing will have no impact on competitive balance, sharing of local TV revenues will do so, and will tend to improve competitive balance. Thus even with fixed supply conjectures, the invariance principle need not hold. 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) non-cooperative 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 our 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

$$(22) \qquad \pi_i = p_i(e_i)[\alpha L_i + (1-\alpha) L_j] - ce_i, \qquad i = 1,2$$

The first order conditions are then

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

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

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

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

If we suppose that $\frac{\partial L_j}{\partial e_i} = -\frac{\partial L_j}{\partial e_j}$ then for fixed labor supply the LHS of (24) 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. This result would hold even if we introduced demand for success and competitive balance into the model.

This point is articulated by Marburger (1997), who suggests that this asymmetry might be true for gate revenues as well, where demand for absolute quality may be important[60]. Note that the demand for quality will not influence our results if it enters the revenue functions symmetrically apart from ruling out symmetric cooperative equilibria where

[60] Marburger applies his model to the case of a luxury tax (see below). The same result goes through when the supply of talent is elastic. Kesenne (2000a) shows that if team revenues also depend on the quality of the opponent, which can be proxied by the opponent's winning percentage, and that the marginal revenue from opponent quality differs across teams, then revenue sharing improves competitive balance. This is essentially the same argument as that concerning local TV revenues.

neither team invests in talent, creating a perfectly balanced, but presumably uninteresting, contest. In general an increased demand for talent will simply increase the wage rate for the fixed supply case and also expand the aggregate quantity of talent employed in the elastic case (for more details see Marburger (1997), Cyrenne (2001) and Szymanski (2001b)).

It is equally important to note that revenue sharing of this kind will, as with gate revenue sharing, reduce the marginal revenue to each team from hiring an additional units of talent and hence drive down the wage rate per unit of talent and increase profits in equilibrium. More generally, revenue sharing works in the opposite way to a prize because it diminishes effort incentives. Falconieri et al (2002) introduce this consideration into a model of the market for TV sports rights to examine when collective selling raises welfare (compared to individual selling). If aggregate effort is deemed valuable by contest designers (as was generally assumed in the case of individualistic sports), then sharing of TV revenues is not an unmitigated blessing even if it improves competitive balance.

This naturally raises the question of how prizes would affect competitive balance in a team sports context. The first point to note is that, 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 sponsorship income and merchandising are likely to be substantially increased by winning a championship (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)), and this will impact on decision making in much the same way as an explicit prize. However, unlike a prize, the value of merchandising and related opportunities tend to differ between teams (e.g. because market sizes differ) and hence this kind of incentive promotes asymmetry.

Suppose instead that each team in the league were to contribute some fixed sum to a prize fund, and then that prize fund were 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

$$\pi_1 = p_1(e_1)[\mu + V] - V/2 - ce_1,$$

(25)

$$\pi_2 = p_2(e_2)[1 + V] - V/2 - ce_2,$$

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

(26)
$$\frac{\frac{\partial p_1}{\partial e_1}}{\frac{\partial p_2}{\partial e_2}} = \frac{1 + V}{\mu + V}$$

from inspection the RHS of (26) converges to unity as V increases, implying that, for any elasticity of supply, a team funded prize will increase competitive balance[61]. 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[62]. 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.

[61] 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.

[62] See Szymanski (2001b) for a more detailed analysis of the implications of prizes in a model of team sport contests.

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% of annual TV income is awarded on the basis of League rank, with the League champions receiving twenty times as much (of the 25%) as the team ranked last in the League[63]. 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 non-economists. This may have something to do with beliefs about the operation of capital markets or about the true objective function of team owners.

B. 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 US[64], but which often seems less

[63] The precise formula is $V_R = \frac{n+1-R}{\sum_{i=1}^{n} R_i}$ where V_R is the prize awarded to the Rth ranked team and n is the number of teams in the League.

[64] 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.

appropriate in the case of European soccer[65]. 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[66]. 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, 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 non-profit objectives. In England the governing body still retains a maximum dividend rule, currently set at 15% of paid up share capital; although 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. 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[67].

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 will be greater in an asymmetric league of win maximizers compared to profit maximizers. Kesenne (2000a) has addressed the question of gate sharing in the context of a league composed of win maximizers and shows that in general it will lead to greater competitive balance (even when the supply of talent is fixed). Intuitively, if teams spend all available income on hiring talent (i.e. they face a

[65] Dabscheck (1975) considered Australian sports teams to be revenue maximizes.
[66] 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.
[67] Further discussion of this is to be found in Sloane (1971), Kesenne (1996) and Bourg and Gouguet (2001). Discussion of changing behavior patterns in recent years can be found in Andreff and Staudohar (2000).

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[68].

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 (Fershtman and 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[69]. 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.

C. Salary caps, luxury taxes and the unions

The power of unions in North American sports came to the fore in the 1970s and since then wage negotiations in the 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. Given the invariance proposition, one might expect 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 VIA, while support for the second is discussed in Zimbalist (1992) presents evidence on the second.

[68] 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.

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

The power of the unions has been bolstered by the antitrust exemption for collective bargaining agreements, since this (a) ensures that unions have exclusive bargaining rights and (b) enables 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 as a bargaining ploy, most notably the NFL players' union at the time of the McNeil case (in 1987) and the NBA players' union following the expiry of the 1988 collective bargaining agreement (in 1994, for details see Staudohar (1996)).

One of the principal objectives of the union in the NBA case was to prevent the continued operation of the salary cap which had been introduced in basketball in 1984. This specified a maximum payroll equal to 53% of defined gross revenues, in exchange for a complex set of arrangements specifying minimum player payments and subsidies to weaker teams. It is widely accepted that the impact of an effective salary cap is to increase competitive balance, but it is also widely accepted 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 (pp1277-1282) find that the standard deviation of win percent has increased since its introduction (see also Kesenne (2000b)).

However, Vrooman (1995) points out that even if teams were constrained to pay identical salaries they would still have incentives to ensure that talent gravitated to its most valuable activity, which could well imply unbalanced competition if teams have unequal drawing power, rendering the salary cap ineffective. A variety of alternative mechanisms other than salaries might achieve this end, most notably the non-playing income that players can obtain from endorsements.

From the point of view of contest design, an effective salary cap system has similar effects as win maximization with revenue sharing. 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, as well cutting the expenditure of the large revenue generating teams.

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)[70]. 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 1994-5. When the two parties agreed a settlement it included an arrangement to tax expenditures in excess of fixed limits, rising from $51m in 1997 to $58.9m in 1999, at a rate of 34% in the first two years and 35% in the third year. This system raised $30.6m over the three years for redistribution to the weaker teams, compared to total MLB payroll spending of $3877m over the same period. Not surprisingly the luxury tax was deemed to have little effect[71]. In 2002 MLB agreed a new luxury tax after narrowly avoiding a strike[72].

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 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.

[70] See also Gustafson and Hadley (1996).
[71] Somewhat oddly, the Blue Ribbon Panel (Levin et al (2000)) attributed its failure to the fact that the tax threshold was a floating one (p39), rather than the fact that the tax threshold was simply set too high.
[72] The tax regime was set for a four year period, the tax thresholds being $117m in 2003, $121m in 2004, $128m in 2005 and $137m in 2006. Tax rates were 17.5% in 2003, rising to 22.5% for first time offenders

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 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).

D. 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 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

and 30% for repeat offenders in 2004 and 2005, with 3rd time offenders paying 40% in the latter year, and then 40% in 2006 except for first time offenders.

smaller number of members than that the social welfare optimum that maximizes total member revenues[73].

The issue can be illustrated using a simple version of the contest success model. Suppose that teams in a league compete in a symmetric contest with a contest success function 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[74]. To avoid underinvestment issues we assume this rent can be fully appropriated by the local team. Further we assume that some fraction ϕ of this locational rent is allocated to a prize fund V awarded to the winner of the league championship and (1-ϕ) retained by the owner. Thus team profits are equal to (1-ϕ) U + $p_i V$ - 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[75]) we can find the equilibrium profit of each team to be:

$$\pi = U\left[1 - \gamma\phi\,\frac{(n-1)}{n}\right] \tag{27}$$

If we take aggregate welfare to be the sum of profits (since all consumer surplus is appropriated), we obtain:

$$W = n\pi = U[n - \gamma\phi(n-1)] \tag{28}$$

If we now take the derivative of welfare and of profits with respect to the number of teams and subtract we obtain:

[73] This same argument has been applied to the inefficiency of a labor managed firm (e.g. Ward (1958), Meade (1974)), which might be thought an appropriate analogy for a sports league managed by the member teams.
[74] See Siegfried and Petersen (2000) for an interesting analysis of this issue.
[75] An assumption that can be justified in this case since the optimal league size is a long run decision, and in the long run talent supply is elastic, since even with a monopsony league talent can be attracted away from other sports.

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

Since the derivative of profits with respect to n is negative teams will prefer smaller leagues, all else equal, while as long 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.

A contest model such as that analyzed here where teams value championship success will typically imply a lower degree of expansion than in the "win percent" model, where teams are only interested in generating 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. However, with a fixed talent supply teams would only wish to expand to the point where all talent resources are fully utilized. Subject to this constraint, the implied extent of expansion should be efficient.

Fort and Quirk (1992 and 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. 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, as mentioned above, 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 (Cramton and Palfrey (1990))[76].

In the European context these issues have never arisen. The hierarchy established by the promotion and relegation system ensure 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.

E. Promotion, relegation, and exclusive territories

The number of teams competing in any given sport in any country is very large. For example, the Football Association of England has of the order of 44,000 registered clubs as members for a population of less than 50 million, or around one soccer club per 1000 inhabitants. Yet in any country the number of professional teams is relatively small. In North America the major leagues in each sport number around thirty, and yet there are many cities, some with populations in the region of one million, that would like, but cannot obtain, a major league franchise (q.v. Noll and Zimbalist (1997)). In European soccer cities do not face this constraint. Moreover, the relocation of a team from one city to another is almost unheard of[77]. The reason for this is the promotion and relegation

[76] 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.

[77] Although there is at least one exception. Wimbledon, the home of tennis, also had a football team from the early 20th century, but it was very unsuccessful until the 1970s, when it was promoted to the professional leagues and rose rapidly from the Fourth to the First Division (now the Premier League). Soon after this its tiny stadium was closed down and the club moved "temporarily" to another part of London to share its ground with another team. However, it appears the municipal authorities did not want them back and no new stadium was built. Exasperated with this state of affairs, the team owners sought permission to relocate to Dublin, in the Republic of Ireland, which has its own national association. This move was blocked by the governing bodies which wanted to keep national leagues segregated. They were then refused permission to move to another city in England that did not have a Premier League team. After

system, which requires the worst performing teams in any league division to drop down to the immediately inferior division, to be replaced by a successful team from the immediately junior division. This hierarchical relationship in theory passes right down from the top division to the very lowest divisions of amateur sport. In practice, what the system means is that there is a large pool, perhaps more than one hundred potential entrants in each national Association, competing to reach the highest level. In a US context, this would mean extending the opportunity to play major league sport from cities with populations in excess of 1.5 million (the 30 largest) to cities with populations as small as 400,000[78].

The main issue when comparing closed league systems with open ones (i.e. leagues operating promotion and relegation systems) is that while both systems permit entry by new teams, the rules of entry are different. In the closed system cities wanting to attract a franchise must bid. Existing members will only add a franchise if they perceive a gain in profit, so that if the league operates to maximize joint profits, entry will be permitted until joint profits are maximized. Even if teams do not operate an efficient cartel, they are unlikely to permit entry to the point where rents are completely dissipated. In an open system the league authority has to decide the optimal size of the division, but once this is fixed entry can be achieved by any team[79]. Any city in Europe wanting a top division team does not have to pay all the other teams for the right to compete, or pay to attract an existing team to relocate: it can simply hire players to make the local team competitive. Since anyone is free within the system to set up a team, the notion of territorial exclusivity does not arise[80]. And even if teams did relocate, the value to a city of

---

being relegated to the Football League First Division they were, after much hand wringing and the threat of legal action, given permission in 2002 to move to Milton Keynes, 70 miles from their old home.

[78] According to the 1998 estimates, Orlando was the 30th ranked SMSA, while Boise, Idaho was the 100th.

[79] This means that the authorities in a divisional system with promotion and relegation can optimize the number of teams eligible for the championship each season (which in this system means the size of the division), 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.

[80] In theory it might be possible to create territorial exclusivity within a promotion and relegation system, however most league authorities would argue that this is at odds with the principle of open entry on which the system is founded.

attracting an existing team would be smaller than in a closed system, since that team would not be immune from the possibility of relegation in the future.

One reason why expansion may go further than the joint profit maximum in a closed system is the threat of entry by rival leagues (see the discussion in the previous section). If the supply of sports leagues were a contestable market, every city capable of supporting a competitive franchise would be admitted to the major leagues. In practice, given the size of sunk costs involved in setting up a new league the assumption of contestability is not plausible. This implies that even with the entry threat there will be underexpansion. One way to think of the promotion and relegation system is that it is a rationing mechanism to ensure that all the teams capable of running a competitive team have access to the top division. It is clearly a sensitive welfare judgment as to whether sharing access is more desirable than offering permanent membership to a subset of cities. Promotion and relegation resembles a speeded up process of franchise relocation, the difference being that a city losing a top division team through relegation does not lose the team identity altogether, it is merely condemned to a lower level of competition for the current season.

If the main public policy benefit of an open system is that it avoids competition to provide public subsidies, the central question is whether an open system can offer an equal quality of competition. A problem that faces all leagues is the issue of meaningless matches toward the end of a season once teams are no longer in contention. In North America the development of the play-off system has been the main response, ensuring that a large fraction of teams are in contention to a late stage of the competition. However, very poor teams often fall out of contention by mid-season, and under the reverse order of finish draft system can enter into a competition to achieve the lowest possible performance[81]. The costs of this to the fans must be large, particularly since it is not only small city teams that fall into this trap. It is often said that the Chicago Cubs has no need to field a competitive team since fans will turn up to Wrigley Field come what may. The threat of relegation means that in European soccer a low effort strategy is not

[81] See Taylor and Trogdon (2002) for empirical support of this proposition from the NBA.

optimal, so long as the reduction in revenues from relegation is large enough, as it generally is (it can more than halve revenues). Poor performing teams compete right to the end of the season, and often the interest is as great in the outcome of the relegation battle as it is in the championship race.

However, the fact that relegation provides an additional incentive to supply effort does not imply that aggregate effort is higher. Firstly, if the aggregate supply of effort, or in this case talent, is fixed, then an open system can only affect the distribution of talent. In a symmetric contest where the prize is awarded to the championship victor all teams will hire an equal quantity of talent in a closed league, while in an open league teams in the higher division will have a larger share than teams in the lower division (since they are not eligible for the prize in the current year). In an asymmetric contest with fixed talent supply this effect can involve welfare losses: the relegation of a large market team which is replaced by a small market team will adversely affect more fans. But large market teams are less likely to be relegated in any case, since they hire stronger teams. In this case promotion and relegation creates "yo-yo" teams, i.e. teams strong enough to be promoted but not strong enough to avoid immediate relegation[82]. The threat to large market teams in a promotion and relegation system is that they face entry from rivals in the local market. Most major European cities would have at least two teams in the national leagues, and sometimes more, while in the US it is unusual for a city to have two major league teams in a single sport[83].

Szymanski and Valletti (2002) consider both symmetric and asymmetric tournaments for the case of elastic talent supply. They find that the effect on aggregate effort of promotion

[82] It is, of course, 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, it is clear that an efficient promotion and relegation system requires player mobility, since the best talent will always migrate to the top division, and in practice this often happens with extraordinary speed. The point of promotion and relegation is to discipline the owners rather than the players.

[83] Some critics have argued that the relevant comparison is the number of major league teams across all sports, and that on this basis US cities are as well served as European cities, which are the victims of a single dominant team sport, rather than the multiplicity of US team sports. This raises the intriguing possibility that the dominance of a single sport (soccer) in Europe compared to the diversity of major league sports in the US is in fact an endogenous response to the incentive structures. As yet there is no research on this issue.

and relegation depends on the discriminatory power of the contest success function. In a relatively discriminating contest (i.e. one where a small advantage in effort or talent has large effect on the probability of success, or in terms of the contest model γ is large), a closed system produces more effort. Intuitively, since effort increases with γ there is less likely to be an under-supply of effort, and placing some teams in a lower division simply reduces opportunities, and therefore effort, of lower division teams. But when γ is small, then eliciting effort is costly, and the relegation threat is a relatively cheap way to elicit effort increases. Obtaining general results for an asymmetric contest is more difficult since much depends on the type of asymmetry assumed. They consider the case of two strong (i.e. high revenue generating potential) and two weak teams in either a four team closed league or two two-team divisions with promotion and relegation. They find that when the difference in strength is small the open system elicits less effort (and thus raises team profitability), while when the difference in strength is large this result is reversed. However, there are still many unresolved issues and this is clearly a subject for further research.

If an open system obliged teams to supply more effort and reduces profits, why would the leading teams simply not secede from the League and set up on their own? The answer to this 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). Note that FIFA pays no compensation to the clubs who continue to pay the full salaries of their players during international tournaments. 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[84]), but also, with turnover of $4bn, 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

[84] 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.

either because of patriotism or because of the reputation effects and its impacts on endorsement income[85]. Thus any breakaway league would find it hard to retain players.

F. Club versus Country

National teams have been unimportant in the development of the major team sports in the US, but in other sports national teams and international representative sport has been the driving force in developing the popularity of the game and provide 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 (most 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.

[85] 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.

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 nations 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, and the dominant teams are seldom drawn from the larger or richer nations and (b) international representative competition is used to subsidize domestic league competition.

On the first point, 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 over 74%, 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% against England with only 8% of the latter's population. Similarly the Australians in cricket have a winning record of 56% against England in 209 matches over the period 1877 to 2001 (ignoring ties), despite a much smaller population[86]. The West Indies, drawing on the smallest population[87] of the ten Test Match Cricket playing nations have the second highest all-time winning record (57%)[88]. 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. Bernard and Busse (2000) show that population and GDP are remarkably reliable predictors of Olympic medal success.

[86] Over the last twenty years Australia's dominance has become embarrassing, with a 66% winning record in decisive matches.
[87] The other ten are Australia, Pakistan, England, South Africa, India, Sri Lanka, New Zealand, Zimbabwe and Bangladesh, with a combined population of 1.4bn, compared to the Island population of around 4m. Even excluding India, this would amount to no more than 1% of the population of the cricketing nations.
[88] If baseball were regularly played at 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.

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%), despite entry being open to the entire planet. Brazil has a 76% 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 Condon (2001)). Preston et al (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 $30m) but 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 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.

A similar trend seems to be emerging in Rugby Union, where traditionally the international representative game 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)[89]. 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. 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.

VIII. Antitrust and Public Policy

In the words of Gilbert and Flynn (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[90]. However, given the abiding interest of the courts and legislators in the fortunes of sports leagues, the implications of both the theory and

[89] Macmillan (1997) provides an interesting discussion on the balance between centralized co-ordination and decentralized decision making in the case of New Zealand rugby union.

the empirical research reviewed here are worthy of brief discussion. From the outset it is worth noting that the legal issues of 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[91].

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 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[92].

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 co-operation of rivals, so

[90] The US literature is particularly rich- e.g. Flynn and Gilbert (2001), Ross (1989, 1997, 1998, 2001), Ross and Lucke (1997), Weiler and Roberts (1998), and there are a number of European texts, e.g. Gardiner et al. (1998), Husting (1998). For an Australian perspective see Dabscheck (2000).
[91] Some issues remain, such as rules relating to eligibility, and in particular eligibility and disabilities.
[92] However, this does not mean that any restrictions entered into by MLS might not be subject to litigation under Section 2 of the Sherman Act.

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 then, 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 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 (e.g. reserve clause, draft, salary cap, roster limits, restrictions on player trading), product market (e.g. revenue sharing, collective selling, exclusive territories) or capital market (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 in fact 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 thirty seven players on their roster. Finley v. Kuhn (569, F. 2d 1193, 6th Circuit 1978) upheld the right of the Commissioner of baseball to 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[93]. In the Raiders' case (Los Angeles Memorial Coliseum Comm'n v. NFL, 726 F.2d 1381 (9th Cir. 1984)) the court found that 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 justification for restraints, although they have accepted them as possible justifications under a rule of reason. However, this state of affairs has been complicated by the non-statutory 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

[93] The NCAA case also considered in detail the effect on live gate.

agreements from antitrust scrutiny. As a result, in practice the major leagues operate a wide range of restraints, as 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, but 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 finally taken to the European Court of Justice. In this case the court held that competitive balance, although it could in principle justify a restraint, could not be held in this case to be a justification. Moreover, the free movement of labor, a principle enshrined in the Treaty of Rome that founded the Union, 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. In contrast to Finley v. Kuhn, the Commission claimed that the league authorities did not have the right to maintain a system of payments that had the effect of restricting the 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 compensation system that would allow clubs to claim significant fees for players under twenty three on the grounds of investment in training costs. Players over twenty three 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[94].

Later in 2001 the Commission issued a statement of objections[95] 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 UK (upheld), Denmark (upheld), the Netherlands (no decision), Italy (ruled illegal) and Spain (prohibited), see Szymanski (2002) for details.

Thus it is clear that the soccer leagues of Europe have received much less favorable antitrust treatment than the North American leagues. Given that the European leagues seem much less competitively balanced than the North American leagues, have fewer restraints but seem to have no problems surviving, and even seem to be thriving, would it be correct, as the European Commission (1998) has done, to speak of a European model of sport[96]? Certainly its distinctive antitrust regime is complemented by its distinctive competitive structure. However, the main issue must be whether the current 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. However, if the concessions made by the governing bodies to these strong clubs are adequate, a breakaway of this kind may not come to pass. One obvious conclusion is that if competitive balance really matters then we should expect the European system to collapse given the evidence of inequality and the startling absence of any significant measures to maintain balance. On the other hand, the evidence of the empirical section

[94] See FIFA Regulations for the Status and Transfer of Players, July 2001 (see also the comments in section VIA above).
[95] European Union Official Journal, C 169, 13.06.2001, p. 5.
[96] See Primault and Rouger (1999) for a trenchant assertion of difference.

provides very little support for the proposition that balance is really critical to the long term health of a league system.

IX. 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 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 may also be quite different depending 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. However, such a project is not beyond the capabilities of the economics profession.

References

Alexander D. (2002) “Major League Baseball: Monopoly Pricing and Profit-Maximizing Behavior” *Journal of Sports Economics*, 2, 4, 341-355

Andreff W. and Staudohar P. (2000) "The Evolving Model of European Sports Finance" *Journal of Sports Economics*, 1, 3, 257-276.

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

Baik K. (1994) “Effort levels in contests with two asymmetric players” *Southern Economic Journal*, 61, 367-378.

Baimbridge, M., Cameron, S. & Dawson, P. (1996). Satellite Television and the Demand For Football: A Whole New Ball Game? *Scottish Journal of Political Economy*, Vol 43, No.3.

Balfour A. and Porter P. (1991) “The Reserve Clause in professional sports: legality and effect on competitive balance” *Labor Law Journal*, 42, 8-18.

Baye M., Kovenock D. and de Vries C. (1994) "The Solution to the Tullock Rent-Seeking Game When R > 2: Mixed-Strategy Equilibria and Mean Dissipation Rates," *Public Choice*, 81, pp. 363-380.

Baye M., Kovenock D. and de Vries C. (1997) “The incidence of overdissipation in rent-seeking contests”. Mimeo.

Becker B. and Huselid M. (1992) “The incentive effects of tournament compensation systems” *Administrative Science Quarterly*, 37, 336-350.

Bernard A. and Busse M. (2000) “Who Wins the Olympic Games: Economic Development and Medal Totals.” Mimeo.

Borland, J. (1987). The Demand For Australian Rules Football. *The Economic Record.* September pp. 221-230.

Borland J. and Lye, J. (1992) “Attendance at Australian Rules Football: A Panel Study” *Applied Economics*, 24, 1053-1058.

Bourg J. F. and Gouguet J.F. (2001) Economie du sport. Editions La Decouverte, Paris.

Buchanan J. (1965) "An Economic Theory of Clubs", *Economica*, February, 1-14.

Bulow J. and Klemperer P. (2001) "The Generalized War of Attrition" *American Economic Review*, 89, 1, 175-89.

Butler M. (1995) "Competitive Balance in Major League Baseball" *The American Economist*, 39, 2, 46-52.

Buzzachi, L. , Szymanski, S. and Valletti, T. (2001) "Static versus Dynamic Competitive Balance: Do teams win more in Europe or in the US?" Imperial College Management School Discussion paper.

Cairns J.A. (1987) "Evaluating changes in League Structure: The Organisation of The Scottish Football League." *Applied Economics*, 19, 2, 259-275.

Cairns, J.A., Jennett, N. and Sloane, P.J., (1986), "The Economics of Professional team Sports: A Survey of Theory and Evidence", *Journal of Economic Studies*, 13, pp.3-80.

Cheffins B. (1998) "Sports Teams and the Stock Market: A Winning Match?", *UBC Law Review* 32: 271-291.

Condon, Sir P. (2001) *Report on Corruption in International Cricket*, Anti-Corruption Unit, International Cricket Council, London.

Cramton P. and Palfrey T. (1990) "Cartel enforcement with uncertainty about costs" *International Economic Review*, 31, 1.

Cymrot D., Dunlevy J. and Even W. (2001) " 'Who's on first' : an empirical test of the Coase Theorem in baseball" *Applied Economics*, 33, 593-603.

Cyrenne P. (2001) "A quality-of-play model of a professional sports league" *Economic Inquiry*, 39, 444-452.

Dabscheck B. (1975) "Sporting equality: labour market versus product market control", *Journal of Industrial Relations*, 17, 2, 174-190.

Dabscheck B. (1989) "Abolishing transfer fees: the Victorian Football League's new employment rules" *Sporting Traditions*, November, pp. 63-87

Dabscheck B. (2000) "Sport, Human Rights and Industrial Relations" *The Australian Journal of Human Rights*, 6, 2, 129-159.

Daly G. (1992) "The baseball player's market revisited" in Diamonds are Forever: The Business of Baseball, Sommers P. (ed.) The Brookings Institute.

Daly G. and Moore W. (1981) "Externalities, property rights, and the allocation of resources in Major League Baseball" *Economic Inquiry*, 29, 77-95.

Dasgupta A and Nti K. (1998) “Designing an optimal contest” *European Journal of Political Economy*, 14, 587-603

Davies B., Downward P. and Jackson I. (1995) "The demand for rugby league: evidence from causality tests" *Applied Economics*, 27, 1003-1007.

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

Dixit, A. (1987) “Strategic behavior in contests” *American Economic Review*, 77, 891-898

Dobson S. and Goddard J. (1998) "Performance and revenue in professional league football: evidence from Granger causality tests" *Applied Economics*, 30, 1641-1651.

Dobson S. and Goddard J. (2001) The Economics of Football. Cambridge University Press. Cambridge: UK.

Duggan M. and Levitt S. (2000) Winning isn't everything: corruption in Sumo Wrestling" NBER Working Paper no. W7798.

Eckard E. (2001) “Free Agency, competitive balance and diminishing returns to pennant contention” *Economic Inquiry*, 39, 3, 430-443.

Ehrenberg R. and Bognanno M. (1990a) “Do Tournaments have incentive effects” *Journal of Political Economy*, 98, 6, 1307-1324

Ehrenberg R. and Bognanno M. (1990b) “The Incentive Effects of Tournaments Revisited: Evidence from the European PGA Tour.*”* *Industrial and Labor Relations Review*, 43, 74S-88S.

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

Ericson, T. (2000) “The Bosman Case: Effects of the abolition of the transfer fee” *Journal of Sports Economics*, Vol 1, No 3, pp203-218.

European Commission (1998) The European Model of Sport. Consultation paper of DGX.

Falconieri S., Palomino F. and Sakovics J. (2002) “Collective vs Individual Sale of Rights in League Sports”. Mimeo.

Fees E. and Muehlheusser G. (2002) “Transfer Fee Regulations in Europe”, *European Economic Review*, forthcoming.

Forrest D., Simmons R and Feehan P. (2002) “A spatial cross-sectional analysis of the elasticity of demand for soccer,” *Scottish Journal of Political Economy*, 49, 3, 336-356

Fernie S. and Metcalf D. (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, 385-411.

Fershtman C. and Judd K. (1987) "Equilibrium incentives in oligopoly" *American Economic Review*, 77, 927-40.

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

Forrest D. and Simmons R. (2002b) “Team salaries and playing success in sports: a comparative perspective,” *Zeitschrift für Betriebswirtschaft* Vol. 72, No. 4.

Fort, R. (2000), “European and North American Sports Differences (?)”, *Scottish Journal of Political Economy*, Vol 47, No 4, pp431-455.

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

Fort R. and Quirk J. (2000) “Sports team behavior and sports policy: the winning percent maximising league” mimeo.

Frick B., Prinz J. and Dilger A. (2001) “Pay and performance in professional road running: the case of city marathons” mimeo.

Fudenberg D., Gilbert R., Stiglitz J. and Tirole J. (1983) "Pre-emption, leapfrogging and competition in patent races" *European Economic Review*, 22, 3-31.

Fudenberg D. and Tirole J. (1985) "pre-emption and rent equalization in the adoption of new technologies" *Review of Economic Studies*, 52, 383-402.

Fudenberg D. and Tirole J. (1991) Game Theory. MIT Press.

Fullerton, R. and McAfee, R. (1999) “Auctioning entry in tournaments” *Journal of Political Economy*, 107, 3, 573-606

Gardiner S. , Felix, A., O’Leary J, James, M. and Welch, R. (1998) Sports Law. Cavendish Publishing.

Garicano L. and Palacios-Huerta I. (2000) “An empirical examination of multidimensional effort in tournaments” mimeo.

Gilbert R. and Flynn M. (2001) 'An Analysis of Professional Sports Leagues as Joint Ventures' *Economic Journal*, 111, F27-F46.

Gradstein M. and K. Konrad (1999) "Orchestrating Rent-Seeking Contests" *Economic Journal*, 109, 536-545.

Green J. and Stokey, N. (1983) "A comparison of tournaments and contests" *Journal of Political Economy*, 91, 349-64

Grier K. and Tollison R. (1994) "The rookie draft and competitive balance: the case of professional football" *Journal of Economic Behavior and Organization*, 25, 293-298.

Gustafson E. and Hadley L. (1996) "The luxury tax proposal for Major League Baseball: A partial equilibrium analysis" in Fizel, Gustafson, and Hadley (eds.) Baseball Economics: Current Research. Westport, CT: Praeger.

Guttman A. (1994) Games and Empires. Columbia University Press

Hall S., Szymanski S. and Zimbalist A. (2002) "Testing causality between team performance and payroll: the cases of Major League Baseball and English soccer." *Journal of Sports Economics*, 3, 2, 149-168

Harris C. and Vickers J. (1985) "Perfect equilibrium in a model of a race" *Review of Economic Studies*, April, 193-209.

Harris C. and Vickers J. (1987) "Racing with uncertainty" *Review of Economic Studies*, 54, 1-22.

Hart, R.A., Hutton, J. & Sharrot, T. (1975). "A Statistical Analysis of Association Football Attendances" *Applied Statistics.* 24, No. 1, pp. 17-27.

Hess R. and Stewart B. (1998) More Than A Game: an unauthorised history of Australian Rules Football. Melbourne University Press.

Higgins R., Shughart W. and Tollison R. (1985) "Free entry and efficient rent-seeking" *Public Choice*, 46, 247-258.

Higgins R. and Tollison R. (1990) "Economics at the Track" in Goff B., Tollison R. (eds.) Sportometrics. Texas A&M: Texas.

Hillman A. and Riley J. (1989) "Politically Contestable Rents and Transfers", *Economics and Politics*, 1, 1, 17-39.

Hoehn, T. and Szymanski, S. (1999) "The Americanization of European Football" *Economic Policy* 28, 205-240.

Holmstrom B. (1982) “Moral Hazard in Teams” *The Bell Journal of Economics*. Vol. 13 (2). p 324-40.

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

Hovenkamp, H. (1995) "Exclusive Joint Ventures and Antitrust Policy" *Columbia Business Law Review*, 1995, 1

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

Husting A. (1998) L’Union Europeene et le sport. Lyon: Les Editions Juris Service.

Hylan T., Lage M. and Treglia M. (1996) “The Coase Theorem, free agency, and Major League Baseball: a panel study of pitcher mobility from 1961 to 1992” *Southern Economic Journal* 62, 1029-1042.

Inglis S. (1988) League Football and the men who made it. Willow Books. London.

Jennett N. (1984) “Attendances, Uncertainty of Outcome and Policy in the Scottish Football League” *Scottish Journal of Political Economy*, 31, 2, 176-198.

Jones, J.C.H. & Ferguson, D.G. (1988). Location and Survival In The National Hockey League. *The Journal Of Industrial Economics*. Volume XXXVI, No. 4, pp. 443-457.

Kahane L. and Shmanske S. (1997) "Team Roster Turnover and Attendance in Major League Baseball," *Applied Economics*, Vol. 29, No. 4, 425-431.

Kahn L. (1993) “Managerial quality, team success, and individual performance in Major League Baseball” *Industrial and Labor Relations Review*, 46, 531-547.

Kahn L. (2000) "The sports business as a labor market laboratory" *Journal of Economic Perspectives*, 14, 3, 75-94.

Kesenne, S., (1996) “League Management in Professional Team Sports with Win Maximizing Clubs”, *European Journal for Sport Management*, vol. 2/ 2, pp. 14-22

Kesenne, S. (2000a), “Revenue Sharing and Competitive Balance in Professional Team Sports” *Journal of Sports Economics*, Vol 1, No 1, pp56-65.

Kesenne, S. (2000b), “The Impact of Salary Caps in Professional Team Sports”, *Scottish Journal of Political Economy*, Vol 47, No 4, pp422-430.

Kipker I. (2000) “Determinanten der Zuschauernachfrage im professionellen Teamsport: Wie wichtig ist die sportliche Ausgeglichenheit?” Unpublished chapter of PhD dissertation

Knowles G., Sherony K. and Haupert M. (1992) "The demand for major league baseball: a test of the uncertainty of outcome hypothesis" *The American Economist,* 36, 72-80.

Kuypers, T.J. (1996) “The Beautiful Game? An Econometric Study of Why People Watch English Football. *Discussion Papers in Economic, University College London*. 96-01

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

Lazear E. (1989) “Pay Equality and Industrial Politics” *Journal of Political Economy*, 97, 3, 561-80.

Lazear E. (1995) Personnel Economics. MIT Press.

Lazear, E. and Rosen, S. (1981) “Rank Order Tournaments as Optimal Labor Contracts”, *Journal of Political Economy,* Vol. 89, pp. 841-864.

Levin, R. C., Mitchell, G. J., Volcker, P. A., & Will, G. F. (2000). The Report of the Independent Members of the Commissioner’s Blue Ribbon Panel on Baseball Economics, Major League Baseball.

Loury G. (1979) “Market structure and innovation” *Quarterly Journal of Economics*, 93, 385-410.

Lynch J. and Zax J. (2000) “The rewards to running: prize structure and performance in professional road racing” *Journal of Sports Economics,* 1, 323-340.

MacMillan J. (1997) “Rugby Meets Economics” New Zealand Economic Papers, 31, 1, 93-144.

Maloney M. and McCormick R. (2000) “The response of workers to wages in tournaments: evidence from foot races” *Journal of Sports Economics*, 1, 99-123.

Maloney M. and Terkun K. (2002) “Road Warrior Booty: Prize Structures in Motorcycle Racing” *Contributions to Economic Analysis and Policy*, 1, 1, article 3.

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

Marburger D. (2002) “Property Rights and Unilateral Player Transfers in a multiconference sports league”, *Journal of Sports Economics,* 3, 2, 122-132

Mason T. (1980) Association Football and English Society, 1863-1915. Harvester Press.

Meade J. (1974) "Labor-Managed Firms in Conditions of Imperfect Competition", *Economic Journal*, 84, 817-824

Moldovanu B. and Sela A. (2001) "The optimal Allocation of Prizes in Contests" *American Economic Review*, 91, 3, 542-558.

Moldovanu B. and Sela A. (2002) "Contest Architecture" Mimeo.

Monahan G. and Sobel M. (1994) "Stochastic Dynamic Market Share Attraction Games" *Games and Economic Behavior*, 6, 130-149.

Nalebuff, B. and Stiglitz J. (1983) "Prizes and incentives: Towards a general theory of compensation and competition" *Bell Journal of Economics*, 14, 21-43

Neale W. (1964) "The Peculiar Economics of Professional Sport", *Quarterly Journal of Economics*, 78, 1, 1-14.

Nitzan, S. (1994) "Modelling rent seeking contests" *European Journal of Political Economy*, 10, 41-60

Noll R. (1974) "Attendance and price setting" in Noll (ed) Government and the Sports Business. Brookings Institute.

Noll, R. (1999), "Competition Policy in European Sports after the Bosman Case" in Kesenne and Jeanrenaud (eds). Competition Policy in Professional Sports: Europe after the Bosman Case, Standaard Editions Ltd: CIES.

Noll R. (2002) "The economics of promotion and relegation in sports leagues: the case of English football" *Journal of Sports Economics*, 3, 2, 169-203.

Noll R. and A. Zimbalist (1997) Sports, Jobs and Taxes. Brookings Institution Press.

Nti K. (1997) "Comparative statics of contests and rent-seeking games" *International Economic Review*, 38, 1, 43-59.

Orszag J. (1994) "A new look at incentive effects and golf tournaments" *Economic Letters*, 46, 77-88.

Palomino F. and Rigotti L. (2000) "The Sport League's dilemma: Competitive Balance versus Incentives to Win" mimeo.

Palomino F.and Sakovics J. (2000) "Revenue sharing in professional sports leagues: for the sake of competitive balance or as a result of monopsony power?" mimeo

Palomino F.and Sakovics J. (2001) “Inter-league competition for talent vs. competitive balance” mimeo

Peel D. and Thomas D. (1988) “Outcome Uncertainty and the Demand for Football: An Analysis of Match Attendances in the English Football League” *Scottish Journal of Political Economy*, 35, 3, 242-249.

Peel D. and Thomas D. (1992) "The Demand for Football: some evidence on outcome uncertainty" *Empirical Economics*, 17, 323-331.

Peel D. and Thomas D. (1997) ""Handicaps, outcome uncertainty and attendance demand", *Applied Economics Letters*, 4, 567-570.

Piraino, T.A. (1999) "A Proposal for the Antitrust Regulations of Professional Sports" *Boston University Law Review*, 79, 889

Preston I. , Ross S. and Szymanski S (2001) “Seizing the moment, a blueprint for reform of world cricket” mimeo.

Primault, D. and Rouger, A (1999), “How Relevant is North American Experience for Professional Team Sports in Europe” in Kesenne and Jeanrenaud (eds). Competition Policy in Professional Sports: Europe after the Bosman Case, Standaard Editions Ltd: CIES.

Quirk J. and El Hodiri M. (1974) “The economic theory of a professional sports league” in Noll (ed.) Government and the Sports Business. Brookings Institution.

Quirk J. and Fort, R., (1992) Pay Dirt: The Business of Professional Team Sports, Princeton N.J.: Princeton University Press.

Quirk J. and Fort R. (1999) Hard Ball: The Abuse of Power in Pro Team Sports. Princeton University Press.

Rascher D. (1999) "A test of the optimal positive production network externality in Major League Baseball" in Fizel, Gustafson and Hadley (eds) Sports Economics: Current Research. Praeger, Westport CT.

Roberts G. (1984) "Sports Leagues and the Sherman Act: the Use and Abuse of Section 1 to Regulate Restraints on Intraleague Rivalry," *U.C.L.A. Law Review*, 32, 219, 286-87 .

Roberts K. (1985) "Cartel behaviour and adverse selection" *Journal of Industrial Economics*, 33, 401-13.

Rosen S, (1981) “The economics of superstars” *American Economic Review* , 845-858

Rosen S. (1986) “Prizes and incentives in elimination tournaments” *American Economic Review*, 76, 701-15.

Rosen S. and Sanderson A. (2001) “Labor Markets in Professional Sports” *Economic Journal*, 111, 469, F47-F68

Ross, S. (1989) "Monopoly Sports Leagues" *University of Minnesota Law Review*, 73, 643

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

Ross S. (1998) “Restraints on Player Competition that Facilitate Competitive Balance and Player Development and their Legality in the United States and in Europe”, in Kesenne and Jeanrenaud (eds).

Ross S. (2001) “Antitrust Options To Redress Anticompetitive Restraints And Monopolistic Practices By Professional Sports Leagues” paper delivered to the 2nd Annual American Antitrust Institute Conference.

Ross S. and Lucke R. (1997) “Why Highly Paid Athletes Deserve More Antitrust Protection than Unionized Factory Workers”, *Antitrust Bulletin* 42, 3, 641-679

Ross S. and Szymanski S. (2002) “Open competition in league sports” *Wisconsin Law Review*, vol 2002, 3, 625-656.

Schmidt, M. and Berri, D. (2001) “Competitive Balance and Attendance: the Case of Major League Baseball”, *Journal of Sports Economics*, 2, 2, 145-167.

Scully G. (1974) "Pay and performance in Major League Baseball", *American Economic Review*, 64, 915-30.

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

Scully G. (1995) The Market Structure of Sports. The University of Chicago Press.

Seymour H 1971, Baseball: The Golden Age, Oxford University Press, New York

Siegfried and Petersen (2000) "Who is sitting in the stands? The income levels of sports fans" in The Economics of Sports, W. Kern (ed) Upjohn Institute, Kalamazoo.

Skaperdas S. (1996) “Contest Success Functions”, *Economic Theory*, 7, 283-90.

Sloane, P.J., (1969), "The Labor Market in Professional Football", *British Journal of Industrial Relations*, 7, 2, pp. 181-199.

Sloane, P.J., (1971), "The Economics of Professional Football: The Football Club as a Utility Maximizer", *Scottish Journal of Political Economy,* 17, 2, 121-146.

Staudohar P. (1996) Playing for Dollars: Labor Relations and the Sports Business. ILR Press, Cornell.

Staudohar P. (1999) "Salary Caps in Professional Team Sports" in Jeanrenaud C. and Kesenne S. (eds) Competition Policy in Professional Sports. Standaard Editions, Antwerp.

Sugden J. and Tomlinson A. (1999) Great Balls of Fire. Mainstream Publishing.

Szymanski, S. (1999), "The Market for Soccer Players in England after Bosman: Winners and Losers", in Kesenne and Jeanrenaud (eds). Competition Policy in Professional Sports. Standaard Editions, Antwerp.

Szymanski S. (2001a) "Income Inequality, Competitive Balance and the Attractiveness of Team Sports: Some Evidence and a Natural Experiment from English Soccer" *Economic Journal,* 111, F69-F84

Szymanski S. (2001b) "Competitive Balance and Income Redistribution in Team Sports." Imperial College Management School discussion paper.

Szymanski S. (2002) "Collective selling of broadcast rights to sporting events" *International Sports Law Review*, 2, 1, 3-7.

Szymanski S. and Kesenne S. (2002) "Competitive balance, the contest success function and gate revenue sharing in team sports" mimeo

Szymanski S. and Kuypers T. (1999) Winners and Losers: The Business Strategy of Football. Viking Books, London.

Szymanski S. and Smith R.(1997) "The English Football Industry, Profit, Performance and Industrial Structure" *International Review of Applied Economics*, 11, 1, 135-153

Szymanski S. and Smith R. (2002) "Equality of Opportunity and Equality of Outcome: Static and Dynamic Competitive Balance in European and North American Sports Leagues" in Barros C., Ibrahim, M. and Szymanski S. (eds), Transatlantic Sport. Edward Elgar: Cheltenham, UK.

Szymanski S. and Valletti T.(2001) "Open and Closed Leagues." Mimeo.

Szymanski S. and Valletti T. (2002) "First and second prizes in imperfectly discriminating contests" Mimeo.

Taylor B. and Trogdon J. (2002) “Losing to win: Tournament incentives in the National Basketball Association” *Journal of Labor Economics*, 20, 1,

Taylor C. (1995) "Digging for Golden Carrots: An Analysis of Research Tournaments", *American Economic Review*, 85, 4, 873-90.

Tenorio R. (2000) “The economics of professional boxing contracts” *Journal of Sports Economics*, 1, 4, 363-384.

Tirole J. (1988) The Theory of Industrial Organization. MIT Press.

Tullock (1980) "Efficient Rent Seeking" in J. Buchanan R Tollison and G. Tullock, eds. Toward a Theory of Rent Seeking Society. Texas A&M University Press, 97-112.

Vamplew W, (1988) Pay up and play the game: Professional sport in Britain 1875-1914. Cambridge University Press

Vives X. (1999) Oligopoly Pricing. MIT Press

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

Vrooman J. (1996) “The baseball players market reconsidered” *Southern Economic Journal*, 62, 3, 339-60

Vrooman J. (1997a) “A Unified Theory of Capital and Labor Markets in Major League Baseball” *Southern Economic Journal*, 63, 594-619

Vrooman J. (1997b) "Franchise Free Agency in Professional Sports Leagues" *Southern Economic Journal*, July, 191-219.

Vrooman J. (2000) "The Economics of American Sports Leagues" *Scottish Journal of Political Economy*, 47, 4, 364-98

Ward B. (1958) "The Firm in Illyria: Market Syndicalism" *American Economic Review*, 48, 4, 566-589

Weiler, P. and Roberts G., (1998) Sports and the Law : Text, Cases, Problems (2nd Edition), American Casebook Series. West/Wadsworth.

Whitney, J.D. (1988). Winning Games Versus Winning Championships: The Economics Of Fan Interest and Team Performance. *Economic Inquiry*. Vol.XXVI, October, pp. 703-724.

Zimbalist A. (1992) Baseball and Billions. Basic Books.

Zimbalist A. (2001) (ed) The Economics of Sport (2 volumes). Edward Elgar.

Zimbalist A. (2003) May the best team win: Baseball economics and public policy. Brookings Institute.