Articles

The Economics of Professional Sports Leagues

Some Insights on the Reform of Transfer Markets

SPIROS BOUGHEAS
Nottingham University

PAUL DOWNWARD
Staffordshire University

Over the years, sporting leagues around the world have implemented a number of cross-subsidization policies aiming to preserve their competitive balance. The main target of these policies has been the sporting labor market. Some of these policies aim to directly affect players' salaries, whereas others attempt to reform the transfer market. In this article, after a brief review of cross-subsidization policies, the authors develop a simple model of the transfer market where players' salaries are determined by a bargaining process. Within this framework, the authors reexamine the effect of such policies for competitive balance, taking explicitly into account their effects on transactions in the transfer market.

Keywords: *professional sports leagues; bargaining; uncertainty of outcome; transfer system*

Walter C. Neale's (1964) exposition of "the peculiar economics of sports" helped to focus attention on two important and closely related themes in the economic analysis of professional team sports. The first theme concerns an adequate economic conception of sporting leagues, and the second concerns a clear statement of the "uncertainty of outcome hypothesis." More specifically, Neale made

AUTHORS' NOTE: *We would like to thank Alistair Dawson, Richard Disney, Stefan Kesenne, Tim Worrall, and two anonymous referees for helpful comments and suggestions. We are responsible for all remaining errors.*

JOURNAL OF SPORTS ECONOMICS, Vol. 4 No. 2, May 2003 87–107
DOI: 10.1177/1527002503251638

the first attempt to define a sporting league in economic terms and argued that leagues were examples of multiplant monopolies being singly responsible for the administration of sports. For example, a team cannot determine its own output level (as measured by the number of games). That depends on how many teams the league admits and on how many times the league decrees teams play one another during the season. For Neale, a central reason why leagues emerged as monopolies was the Louis-Schmelling Paradox. Along with Rottenberg (1956), this was one of the first formal statements of the uncertainty of outcome hypothesis, which states that sports fan interest is greatest when sporting competition is at its most intense. While each team strives to outperform its competitors, both on the field and financially, the league's survival depends on the maintenance of a competitive balance. In sports, every act of production requires the presence of an opponent—hence the paradox.

In contrast to Neale (1964), Sloane (1971), in an analysis of European football, argued that a sporting league and its constituent teams may be more accurately viewed as a cartel than as a multiplant firm. Teams in most sports take decisions about investment, about whether to produce at all, and they usually have substantial control over ticket prices. They also control merchandising, which is now a major source of revenue. One of the most important aspects of Sloane's definition of sporting leagues is that it provides a more plausible framework within which to explore the potential problems associated with managing the mutual interdependence between teams. More recently, Flynn and Gilbert (2001) reinforce this perspective by examining sports leagues as joint ventures. Neale's definition could imply that centralized control coordinates passive teams.

The problem of mutual interdependence, reinforced by these latter perspectives, can be illustrated in the context of league management policies. Many policies of cross-subsidization have been justified on the basis of the uncertainty of outcome hypothesis. Leagues have attempted to transfer resources from stronger, more successful teams to smaller, less successful teams to try to initiate a virtuous circle of overall benefit to the league that can be actively promoted and sustained by cross-subsidization policies. The effect of these policies has been addressed, with some debate, in a growing literature. Moreover, a common characteristic of the modeling approach in this literature is the assumption that competitive imbalance rests primarily in differences in market size and that structural changes in labor market conditions will not affect competitive balance (Fort, 2000). Key to this latter argument is that talent is homogeneous and perfectly divisible and, therefore, well suited for marginal analysis. An immediate consequence of this approach is that a team is identified, apart from some exogenous characteristics, by the total amount of talent it possesses. Furthermore, the price of talent (salaries) is determined in a competitive market. In this article, such key themes are readdressed using an alternative theoretical framework which captures the bargaining process of sporting labor market negotiations that we argue exists in soccer transfer markets. The object is to identify

whether such key themes stand up to scrutiny from such a different set of assumptions.

In the following section, the variety of cross-subsidization policies implemented by sporting leagues are reviewed. Some distinctions and similarities between U.S. and European professional team sports are also noted together with differences of opinion on the effectiveness of cross-subsidization policy initiatives drawn between, broadly speaking, European and U.S. sports economists. The third section outlines a benchmark model based on the alternative labor market assumptions as a foundation on which critical discussion can proceed. The fourth section relaxes some of the assumptions of the model to explore more realistic scenarios as a means of understanding the competitive imbalance in sports leagues. In particular, it is shown that in conditions corresponding to the labor market in soccer, policies proposing to make labor markets more competitive, such as those associated with current transfer system reform, could worsen competitive balance. In this respect, the article adds further support to the European literature that argues that competitive balance is not independent of policy initiatives.

CROSS-SUBSIDIZATION POLICIES

There have been a large number of ways in which leagues have intervened in the management of clubs' finances to provide cross subsidies between clubs. Although the particular administrative details may vary, however, they have had two major targets: the sporting labor market and revenue distribution.[1] Targeting sporting labor markets as a means of cross-subsidizing clubs is based on the idea that players' salaries and wages comprise a large proportion of sporting clubs' costs in the United States and Europe.[2] Policies that affect players will thus have a large direct financial effect on clubs. Second, and directly concerned with the uncertainty of outcome hypothesis, it is the players who ultimately affect a club's success or failure in matches through the production function. Consequently, both the resources of clubs as well as their results can, in principle, be affected through policies aimed at the labor market. Leagues have attempted to influence club financing and results through implementing three major types of labor market policy: drafting systems, salary caps, and reserve option arrangements.

A drafting system rations the order in which professional teams can sign new talent—rookies. Although drafting systems, being targeted at the physical reallocation of sporting talent, have an indirect financial implication for clubs, salary caps are targeted at the financial cost of players directly. Salary caps imply a maximum amount that clubs can spend on players. The intended implication and justification for these policies is that, in principle, it makes the best talent affordable to all teams, thus preserving a league's competitive balance.[3] The final form of labor market policy employed by sporting leagues has been reserve option clauses. These clauses essentially tied players to clubs for their life by giving the club the option to renew

the player's contract when it expired. The retain-and-transfer system in European soccer is another example. It is well known in the sports economics literature that these policies introduced monopsonistic power to sports clubs (Downward & Dawson, 2000).

The other main form of cross-subsidization policy adopted in sporting leagues has been to force clubs to redistribute some of their revenues. Such a policy has been implemented through a variety of forms such as gate-sharing arrangements and rules for television revenue redistribution. These developments are an unusual phenomenon in economic terms. In any other industry, the direct regulation of the terms under which firms compete in markets has been ruled as acting against the public interest. Despite the various legal frameworks for competition policy, in general, government intervention in sporting markets has been minimal (for a recent discussion, see Noll, 1999).

However, one notable exception to this attitude has been the introduction of free agency in the United States during the 1970s as a response to the monopsony power of clubs. Similarly, more recently, in Europe the retain-and-transfer system has been amended by the Bosman ruling. The important outcome of the Bosman Ruling was that no fee could be expected by clubs on the transfer of an out-of-contract player. It is interesting to note that the Advocate General of the European Court of Appeal accepted, in principle, the need for leagues to maintain competitive balance and uncertainty of outcome in making his ruling. He also accepted that smaller clubs often covered financial losses through transfer fee income. However, he argued that using the means of the transfer system to achieve these ends was not justifiable. This was because there were other methods of achieving competitive balance that did not restrict player mobility.

Prima facie, such proposals could be argued to reinforce the similarities with U.S. professional team sports. The Bosman ruling could be seen as a further move toward free agency. Compensation systems for the loss of, say, young players and periods of contract stability would echo the structure of U.S. labor markets reflecting years of experience. Yet despite the apparent similarity of developments in European and U.S. sporting labor markets, it remains that different specific contractual structures exist. In the United Kingdom, contract duration remains potentially more flexibly determined compared to the United States because of the latter threefold partition of the labor market.[4] In addition, cash-only player trades is a common feature of European pro team sports but not North American pro team sports.

More generally, the similarities and differences between U.S and European professional team sports has been debated. Fort (2000), for example, stresses the similarities. Primault and Rouger (1999) argue of the need to be aware of the differences. Likewise, in assessing the effect of cross-subsidization policies, there is some debate. This typically involves differences between U.S. and European sports economists. Most theoretical attempts to examine the effectiveness of cross-subsi-

dization policies in improving competitive balance have their origin in a cartel model of sporting leagues developed by El Hodiri and Quirk (1971). Essentially, they take as their starting point a situation of competitive imbalance and examine the consequences of changes in league policies. Different market sizes or catchment areas account for the competitive imbalances (Fort, 2000). Yet recent U.S. versions, which include Fort and Quirk (1995) and Vrooman (1995, 2000), and European versions, including Kesenne (2000a, 2000b) and Dobson and Goddard (2001), do indicate differences in outcomes.

The key differences lie with the findings presented by both Fort and Quirk (1995) and Vrooman (1995, 2000) that cross-subsidization policies have been ineffective in the United States. Quirk and Fort (1992) also provide detailed descriptive historical statistics to suggest that U.S. leagues tend to conform to long-run domination. This evidence has been presented as offering support for the invariance proposition. This states that a league's competitive balance (talent distribution) is independent of the allocation of the ownership right to sell players' services. It is a direct application of Coase's theorem. The effect of cross-subsidization polices has thus been to redirect economic rents from athletes to their clubs. Fort and Quirk (1995) do indicate that salary caps could help to promote competitive balance if they were made enforceable. However, Vrooman (1995, 2000) argues that, adjusting Fort and Quirk's assumption that marginal costs of talent are constant to reflect increasing marginal costs, a payroll cap set proportional to revenues will reduce to essentially the same problem as revenue sharing. Hence, the policy will be ineffective. In both of these cases, it is assumed that teams' profits maximize and that the labor market is competitive.

In contrast, from a European perspective, Kesenne (2000b) has argued that revenue sharing will affect competitive balance once the absolute quality of the teams is accounted for. Significantly, this result applies whether the teams seek to maximize profits or win percentages. Since Sloane (1971), it has long been argued that, in contrast to U.S. sports leagues, utility maximizing behavior, evidenced by the pursuit of sporting success, characterizes soccer. Dobson and Goddard (2001) provide similar conclusions to Kesenne (2000b) but suggest that greater competitive balance will be likely if teams maximize profits. Significantly, while retaining the assumption of a competitive labor market, they recognize that the labor market in soccer will be more open to international trade than in U.S. sports leagues. This relaxes the constraint on available talent. Finally, Kesenne (2000a) argues in a profit maximizing model that a salary cap can improve competitive balance once one allows for different qualities of players. Kesenne (2000a) distinguishes between top players whose salary is set by a competitive market and regular players who receive a minimum wage. On balance, thus it seems that the character of the labor market is crucial in understanding the differences between the conclusions drawn in the literature.

Finally, as far as the transfer system is concerned, once again some debate exists. Antonioni and Cubbin (2000) argue that the Bosman ruling will not affect clubs' investments in players. Moreover Szymanski (1999) argues that the Bosman ruling will help to weed out inefficient labor contracts and promote increased contract duration. In contrast, however, Ericson (2000), although assuming a competitive labor market, argues that free agency, although not harming competitive balance, will reduce the quality of talent by not compensating clubs for the development of players. By implication, the transfer system represents efficiency gains.

A SIMPLE MODEL OF THE TRANSFER MARKET

To contribute to this literature, the rest of this article reexamines the economics of sporting leagues based on the assumptions that clubs maximize profits and that labor markets are characterized by bargaining. In the former case, this is for two reasons. First, as discussed above, qualitative differences in results derived from models of sporting leagues seem to hinge less on the assumed objectives of clubs than the labor market. Second, this allows for direct comparison with the U.S. literature that assumes profit maximizing behavior. In the latter case, there is some strong evidence to support the presence of bargaining in the soccer literature (Carmichael & Thomas, 1993; Dobson & Gerrard, 1999).[5] The objective of this article is thus to produce some further insight into the dynamics of the labor market transactions of sporting leagues and pay particular attention to the likely effect of the reform of the transfer system.

We begin with some general assumptions that are broadly consistent with the literature. A team in our model is defined as a collection of tasks. Moreover, the productivity of players is dependent on other players. However, following Kesenne (2000a), players are characterized by their task-specific skills and the level of their skills. Thus, an important feature of the model is that even if the level of a player's skills is team independent, the player's expected productivity differs across teams. The reason is that teams, in contrast to typical firms, hire more than one player for each task. Although a team's performance depends on the talent it uses on the playing field, each player's contribution depends on the time he is expected to participate. As new players enter the league and older players retire, the league's distribution of talent changes. This creates opportunities for the transfer of players to teams where they are expected to make greater contributions to a team's success. Thus, a transfer market, which does not have a counterpart in other labor markets but is an implied and ubiquitous component of the market for professional players, finds a place within our framework.[6] Finally, salaries and transfer fees are determined through bargaining between the interested parties.

More specifically, consider a league consisting of N profit maximizing teams. The economic success of each team depends on their relative performance against the rest of their league and the success of the league as a unit. Relative performance, which itself depends on some measure of relative talent, determines the team's

share of league revenues. The level of league revenues, which provides a measure for the success of the league as a unit, depends not only on the aggregate level of talent but also on its distribution. For example, if a single team dominates the league for a long period of time then it will reduce competitive balance and outcome uncertainty, thus reducing the league's aggregate revenues.

Each team allocates talent to M tasks (playing positions) by hiring players. We assume that each player's talent is suitable for only one task and it takes one of three possible values.[7] There is an abundant supply of players with only basic skills, t. The rest of the players are either low, t_l, or high, t_h, talented; where $t_l < t_h$. Because the players who are either low or high-talented are in short supply, only the basic skill level is guaranteed for each position. Moreover, even if a team hires players who belong to the top two talent groups, it is assumed that these players will be on the playing field with probability p. Because of this restriction, teams might find it profitable to hire more than one talented player for each task. For simplicity, we restrict the number of talented players allocated to each position to be less than or equal to two. Of course, the team's performance depends on the amount of talent used and not on the total amount of talent it possesses.

A Benchmark Case for Comparison

It is instructive to begin the analysis with a benchmark case where both team and league revenues are proportional to the talent in use. This exercise will be useful in our attempt to understand under what kind of conditions transactions in the market for professional players do not have any effects on the league's competitive balance. Later, we will examine the implications of relaxing the main restrictions of the benchmark case. It should be noted that there is no necessary symmetry implied in team sizes.

In the benchmark case, if a team hires a talented player, then both its own and the league's expected revenues are raised by the same amount. After an appropriate normalization of talent units, we can set revenues identically equal to talent in use. Let t_t^j denote the talent of team i at task j. Then the total talent (revenues) of team i, t_i, is given by:

$$t_i = \sum_{j=1}^{M} t_i^j \tag{1}$$

and the league's total talent (revenues) is given by:

$$t = \sum_{i=1}^{N} t_i = \sum_{i=1}^{N} \sum_{j=1}^{M} t_t^j \tag{2}$$

Team costs are equal to the amount paid for players' salaries. It is assumed that wages are proportional to talent and after a normalization are set equal to t. In contrast, the salaries of talented players are determined by a bargaining process between them and their employers. Contracts are legally binding and can only be broken under mutual consent. Recently, Stole and Zwiebel (1996) developed a framework for analyzing intrafirm bargaining when contracts are nonbinding. In their setup, firms and workers commit only to a wage conditional on employment. Before production begins, any party can initiate a contract renegotiation. As new opportunities arrive, they alter the outside options of the two parties and can have significant effects on the employment relationship. In contrast, in the present model, because contracts are binding, such considerations are irrelevant. Nevertheless, their method of deriving the surplus also applies to the present framework.

Derivation of Surplus

When a team hires a second talented player, the surplus generated by that player depends not only on his or her talent but also on the talent of the player hired before him or her. There are seven cases to be considered that depend not only on the type of players hired by the team but also on the order of hiring. The first column of Table 1 describes these seven cases. For example, row 5 corresponds to the case where at task j, team i has one low-talented player and a one high-talented player, and the low-talented player was hired first. The second column shows the talent of team i at task j, and the third column the surplus generated by the last player hired. Because in the benchmark model total team talent is equal to the sum of talents at each task, the analysis will focus on a single task.

The derivation of the expressions in the second column is based on the following two considerations. First, what matters for team performance is the talent used in competition and not the available talent. Second, even if a player is needed, he or she will be on the playing field only with probability p.[8] The first row refers to the case where the team does not have any talented player available at this particular position. Given that there is an unlimited supply of players with basic skills, the team's talent at this task is t. The following two rows refer to the case where the team has only one talented player. This player is expected to be playing with probability p and, therefore, with probability $1 - p$ the team will have to replace him or her with a player with only basic skills. The next four rows refer to the case where the team has two talented players. If there is one low-talented player and one high-talented player, the team's first choice is the high-talented player. In the case described in row 5, the team first hired a low-talented player and then a high-talented player. The high-talented player is expected to play a fraction p of total playing time and the low-talented player a fraction p of the remaining time. Of course, for the calculation of total talent the sequence of hiring is irrelevant; hence, rows 5 and 6 show the same expected talent in use.

TABLE 1: Talent in Use and Expected Surplus in the Benchmark Model

Available Players	*Expected Talent in Use*	*Expected Surplus*
$\{\underline{t}\}$	$\underline{t}$	
$\{t_l\}$	$pt_l + (1-p)t$	$p(t_l - t)$
$\{t_h\}$	$pt_h + (1-p)t$	$p(t_h - \underline{t})$
$\{t_l, t_l\}$	$p(2-p)t_l + (1-p)^2\underline{t}$	$p(1-p)(t_l - \underline{t})$
$\{t_l, t_h\}$	$pt_h + p(1-p)t_l + (1-p)^2\underline{t}$	$p(t_h - pt_l) - (1-p)\underline{t}$
$\{t_h, t_l\}$	$pt_h + p(1-p)t_l + (1-p)^2\underline{t}$	$p(1-p)(t_l - \underline{t})$
$\{t_h, t_h\}$	$p(2-p)t_h + (1-p)^2\underline{t}$	$p(1-p)(t_h - \underline{t})$

For the calculation of expected surplus, the sequence of hiring is important. The expected surplus generated by a newly hired player depends on players hired before him or her. This player might also affect the expected surplus of players hired before him or her. However, because contracts are binding, the team cannot renegotiate existing contracts. Therefore, the new player's salary will be based on the surplus he or she is expected to generate and not necessarily on his or her absolute talent. For example, to calculate the expected surplus in row 5, subtract the talent in use shown in row 2 from the one shown in row 5. The difference captures the additional expected talent in use which a high-talented players offers to a team that already possesses a low-talented player in the same position. On the contrary, in the framework of Stole and Zwiebel (1996), the team would renegotiate the low-talented player's contract so that the new contract would reflect his or her diminished contribution caused by hiring the high-talented player. Notice that the expected surpluses shown in rows 4 and 6 are equal even if the expected talents in use are not. This is because the team hires a second player who is low-talented and, consequently, he or she will be on the playing field with probability $p(1-p)$.[9]

Bargaining, Salaries, and Profits

The salary of a talented player is determined by a bargaining process between him or her and the team owner. Both parties understand that any agreement signed is final and can only be amended under mutual consent. The bargaining outcome is a split of the expected surplus. Our results do not depend on the bargaining power of the two parties; it is only the size of the surplus that matters. Despite its simplicity, the above model is sufficiently rich and captures these characteristics of sports leagues that set them apart from other sectors of the economy. Our next step is to describe the labor market dynamics of the benchmark model, that is, under what conditions new players enter the market and old players leave it.

Dynamic Considerations

Most of the important issues related to professional team sports in the literature are dynamic in nature, ostensibly related to the uncertainty of outcome hypothesis. The conflict between intraleague competition and the survival of the league itself is, essentially, a conflict between the short-run interests of individual teams and the long-run concern of all teams to survive as a unit. We assume that time is divided in discrete periods. Each period represents a full season and is divided into two subperiods. The first subperiod corresponds to preseason, where teams make their personnel decisions. During the second subperiod, teams compete during the season. For simplicity, acquisition of talented players is not allowed during the season.[10]

At the beginning of each period, each team inherits the talent they finished with one period earlier. This defines a beginning of period distribution of talent for each team at each task. To complete the dynamics, the effects of the preseason transactions on the above distribution need to be specified. It is assumed that a small number of talented players, relative to the total number of existing players, enter the league during each preseason. They are free to negotiate a contract with any team, and there is sufficient time to contact all teams. If two teams make the same contract offer, then the employment decision is decided randomly. For simplicity, it is assumed that new talented players enter the market sequentially and only after the new players before them have completed their negotiations. It is further assumed that the inflow of players is quite uncertain.[11] Of course, with no additional restrictions, no matter how slow the process is, after a sufficient number of periods all teams will be able to fill all positions with two high-talented players. Therefore, it is further assumed that a small number of players leave the league at the end of each season. Both the talents and the positions of incoming and outgoing players are random. At this point, the reader might feel uncomfortable about the lack of a rigorous specification of the model's dynamics. However, as it will be demonstrated below, none of the results of the benchmark model depend on any particular specification of the initial talent distribution or on the distributions that specify the entry of the new players and exit of those leaving the league.

New Talent and Competitive Balance

As new players enter the league and sign contracts, they alter the talent distribution. If, as a result of these new contracts, the distribution gets less uniform over time, then the league's competitive balance will be destroyed. In contrast, for the benchmark case, the following proposition holds:

Proposition 1: In the benchmark model, entry of new players, on average, increases the league's competitive balance.

Proof. It is demonstrated that, on average, new players sign contracts with teams which are relatively weaker at their task of specialization. There are three cases we need to consider for each type of talent. A new player will either sign a contract with a team that does not have any other talented player in the same position or with a team that has one more talented player who, in turn, is either high- or low-talented.

For new low-talented players, the relevant cases are shown in rows 2, 4, and 6 of Table 1. The expected surplus in cases 4 and 6 are equal and lower than the expected surplus in case 2. A low-talented player prefers to sign a contract with a team that has only players with basic skills available at the same position. In such a team, he or she is expected to play a fraction p of the time; in contrast, in the other two cases he or she is expected only to play as a substitute, that is, a fraction $p(1 - p)$ of the time. If the player's only choice is teams that have another talented player in the same position, then the player is indifferent between the two alternatives and the decision is determined randomly.

For high-talented players, the corresponding cases are shown in rows 3, 5, and 7 of table 1. Subtracting the expected surplus shown in row 5 from the one shown in row 3, we find that the difference is equal to $p^2(t_l - \underline{t}) > 0$. Therefore, a new high-talented player prefers to sign a contract with a team that has only players with basic skills available at the same position rather than with a team that has another low-talented player. Similarly, we find that a high-talented player prefers to sign a contract with a team that has another low-talented player rather than with one that has another high-talented player. Subtracting the expected surplus shown in row 7 from the one in row 5, we find that the difference is equal to $p(1 - p)(t_h - t_l) > 0$. A new high-talented player is expected to play a fraction p of the time when the other player is low talented and only as a substitute if the other player is also high talented.

Up to this point, it has been shown that new players reduce team differences in talent at their playing position. It is still possible that a new player might decrease the league's competitive balance by signing with a team that has low talent in his or her position but high talent in other positions. However, the skill of each new player is randomly determined and, therefore, teams which have talent deficiency in many positions (i.e., low average talent) have a greater chance of matching these positions with the skills of new players. It is the last consideration that warrants the qualifier "on average" in the statement of the proposition.

The intuition is straightforward. The surplus that new talented players expect to generate, and hence the salaries they expect to receive, is higher the lower the talent their new team possesses at their task. Therefore, on average, low-talented teams have a better chance of signing new talented players.

The Transfer Market and Competitive Balance

The transfer market is a distinct characteristic of sporting labor markets. In contrast to firms in other labor markets, sports teams hire more than one player for

employment in the same task. It has been assumed that talent is exogenously given. However, because contracts are legally binding, each player's expected surplus depends on the talent of the player hired before him or her and, consequently, a player's expected surplus can be different across teams. Then, a transfer market can be beneficial to both teams and players if the surplus the players are expected to generate with their new teams is higher than the one they produce at their old teams. The surplus difference will be referred to as the net expected surplus of the transfer. The player receives a fraction of the surplus and the rest is divided between the two teams. The payment received by the selling team is called the transfer fee. Teams might also find it beneficial to exchange players if such an exchange generates a positive net expected surplus. The following proposition refers to the effects of the above transactions on the league's competitive balance.

Proposition 2: In the benchmark model, transfers and exchanges of players, on average, increase the league's competitive balance.

Proof. It must be shown that all transfers and exchanges of players that generate a positive net expected surplus, on average, increase the league's competitive balance.

Consider transfers. Notice that symmetric transfers, that is, transfers that leave the league's talent distribution intact, generate zero expected surplus and, therefore, can be ignored. For example, a transfer of a talented player from a team that has not another talented player in the same task to a team with no talented players (in the same task) generates zero expected surplus.

Any transfer of a talented player from a team with another talented player in the same task to a team with no talented players (in the same task) generates positive expected surplus and increases the league's competitive balance. This follows from Proposition 1. It has been previously established that the expected surplus of talented players in teams with another talented player in the same task is lower than their expected surplus in teams with no talented players (in the same task).

Finally, consider transfers of talented players from teams with two talented players to teams with only one talented player in the same task. The only such transfer generating a positive expected surplus is of a high-talented player from a team with two high-talented players to a team with only a low-talented player in the same task. The only such transfer that generates a positive expected surplus is when a high-talented player moves from a team with another high-talented player to a team with one low-talented player. To find the expected surplus of this transfer, subtract the expected surplus shown in row 7 of Table 1 from the one shown in row 5. The difference is equal to $p^2(t_h - t_l) > 0$. It must also be shown that the transfer improves competitive balance. Before the transfer, the difference in expected talent in use was equal to $p(t_h - t_l) + (1 - p)t_h - p(1 - p)\underline{t}$ (subtract the expected talent in use shown in row 2 from the one shown in row 7). After the transfer, the difference is equal to

$p(1-p)t_h - p(1-p)\underline{t}$ (subtract the expected talent in use shown in row 3 from the one shown in row 5), since the difference has been reduced.

Next, consider exchanges. There is one case that generates a positive expected surplus. A team with two high-talented players exchanges one of them for one of the two low-talented players of another team. The exchange reduces the expected talent in use of the team that originally had two high-talented players by $p(1-p)(t_h - t_l)$ (subtract the expected talent in use shown in row 6 from the one shown in row 7) while it increases the expected talent in use of the team which originally had two low-talented players by $p(t_h - t_l)$. It is clear that the overall expected change is positive. Furthermore, after the exchange, the two teams have equal talent in use in that task.

For the same reasons given in the proof of Proposition 1, the qualifier "on average" has been added in the statement of Proposition 2.

Again, the intuition is simple. These type of transactions involve talented players who move from teams with other talented players at their specialized task, and hence, their expected surplus is low, to teams with lower talent at the same task. On average, it is expected that low-talented teams have a better chance of benefiting from these transactions.

Propositions 1 and 2 are quite robust because their proof does not require any restrictions on the distribution functions or the values of the model's parameters. It is important that the propositions suggest that the increased mobility of players can have potentially positive effects on competitive balance. This might, of course, add some weight to the argument that movements toward free agency will be beneficial to soccer. Clearly, this is an important issue that we return to in the next section.

Free Agency and Competitive Balance

Does the invariance proposition hold in the benchmark case? To address this question, the assumption that at the beginning of each period teams inherit the talent they have finished with one period earlier must be relaxed. Suppose that at the end of each season players are free to move to another team. Such a position is consistent with free agency.[12] It is clear that a player will decide to move to another team only if the expected surplus at his or her new team is higher.[13] Therefore, transfers do not depend on who owns the right to sell the player's services. The only difference is that under free agency it is only the player and his or her new team's owners who share the surplus, that is, there is no transfer fee. The following proposition summarizes this argument.

Proposition 3: In the benchmark model, the league's competitive balance is independent of the ownership of the right to sell players' services.

EXPLAINING COMPETITIVE IMBALANCE

The benchmark model has provided a simple analytical framework for addressing some of the issues concerning competitive team sports leagues. In comparison with the rest of the literature, the modeling approach followed in this article has offered a variety of novel features. It has completely dispensed with the assumptions that talent is homogeneous, divisible, and inelastic in supply, and therefore has allowed the introduction of a bargaining approach to salary determination which is in accord with experience. In addition, observation of common practice has dictated the postulate that contracts are legally binding which, as it was argued above, in conjunction with the fact that teams hire more than one player at each task, offers a rationale for the emergence and extensive use of the transfer market. It has been demonstrated that the benchmark model is sufficiently rich to yield predictions about the effects of transactions in the market for talent on the league's competitive balance and, furthermore, we have managed to derive a simple version of the invariance proposition.

Despite its merits, the benchmark model has a linear structure that is definitely restrictive. For both analytical and expositional simplicity, the following assumptions were made: (a) a team's expected talent in use is equal to the sum of the expected talents in use at each task, (b) each team's share of league revenue is proportional to its share of league talent, and (c) the league's total revenue is proportional to the league's total talent. In this section of the article, we relax these assumptions and discuss why there are strong reasons why competitive imbalance might be worsened in sporting leagues such as professional soccer if there is a more free exchange of players. Importantly, these discussions build on the bargaining framework as opposed to assuming competitive markets.

Information Asymmetries

Thus far, we have ignored the presence of costs other than salary payments. This is the case because implicitly it has been assumed that teams have access to perfect capital markets. In this case, even if a team's other costs set limits on its liquid assets, it could still finance the acquisition of new players either through bank lending or by issuing more shares. However, in reality, teams, especially those located at the bottom of the league tables, have limited access to capital markets. The reason is the uncertainty surrounding the talent of new players, which is partially captured by the parameter p in the benchmark model. It is hard to believe that team owners and potential lenders have the same beliefs about the talents of new players, and such informational asymmetries can limit the ability of teams to raise outside funds. If that is the case, then the presence of other costs can inhibit the access of poor teams either to the market for new players or to the transfer market. If players move only to teams that can afford them, then the league's competitive balance could be destroyed. It is important to note in this regard that Dobson and Goddard (1998)

show that the transfer system has channeled funds from larger to smaller teams. It is also clear, however, that following the Bosman ruling, aggregate revenue flows have begun to dwindle and, as Szymanski (1999) argues, powerful teams now increasingly look overseas for players.

Player Complementarities

The explanation for this can be further explored by relaxing the assumption that a team's expected talent in use is equal to the sum of the expected talents in use at each task. In contrast to the rest of the literature, the model has allowed for a multiplicity of tasks to capture the idea that players are not perfect substitutes for each other. This is a factor implied in the dynamics of the transfer market discussed above. However, the first assumption that a team's expected talent in use is equal to the sum of its corresponding talents at each task implies that tasks are perfect substitutes. Although perfect task substitutability is not crucial for the conclusion of Propositions 1 and 2 that, on average, transactions in the sporting labor market improve the league's competitive balance, there is no doubt that the higher the degree of substitutability the more robust the conclusion is. For example, consider the opposite extreme where tasks are perfect complements, which implies that a team's i talent is given by:

$$t_i = \min\{t_i^1, \ldots, t_i^j, \ldots, t_i^M\} \tag{3}$$

This corresponds to the case where teams need to be well balanced across all tasks to be competitive. Suppose that one team has filled all positions with high-talented players with only one exception, task j, where it has only a low-talented player. Another team has filled all positions with low-talented players except task j, where there is a high-talented player available. Then an exchange of players can improve the first team without affecting the absolute strength of the second team. Competitive balance in this case is harmed. Although the two teams have started with the same talent in use, after the transaction one of the teams has become considerably stronger. Thus, the recognition that player skills might be complementary indicates that a reform of the transfer system could adversely affect competitive balance.

None of this, of course, suggests that competitive imbalance is independent of team revenues, as emphasized in the literature. Examining the implications of this can be understood by relaxing the assumption that each team's share of league revenue is proportional to its share of league talent. This is a crucial supposition. Consider an equally plausible case where successful teams tend to receive a disproportionately high share of league revenues. Then, transactions in the market for new players and in the transfer market would decrease the league's competitive balance. At this point, it is important to distinguish between the surplus in talent that a new

player is expected to generate from the revenues that this surplus is expected to yield. A high-talented player is expected to generate a higher surplus of talent with a team that has, for example, only players with basic skills at his specialized task than with a team that has another talented player. However, his talent might be more profitable for the second team if the latter is, overall, more successful. Similarly, a relatively weak team might sell its high-talented players to stronger teams if such transfers generate positive expected revenues.[14]

Moreover, it is clear in the benchmark model that the assumption that total league revenues are proportional to total team talent is too strong. In fact, it can be argued that according to the uncertainty of outcome hypothesis that total league revenues should be higher the more uniform the distribution of talent across teams is. In other words, we need to consider moments of the distribution higher than the first. Nevertheless, doing so will only strengthen the observations made above with respect to the first two suppositions. Indeed, more generally, a monotonic relationship between competitive balance and league revenues will suffice. Consequently, the results above do not necessarily hinge on the uncertainty of outcome hypothesis.

The above discussion suggests the following important observation. By considering an extended version of the bargaining model analyzed above, it can be shown that transactions can affect the rest of the league, something that is impossible within the benchmark case where transactions only affect participating teams. These externalities imposed on other teams might be strong enough that the overall effect of transactions in the sporting labor market could be deterioration in the league's competitive balance—in contrast, to the conclusions drawn from Propositions 1 and 2. In this sense, reform of the transfer system to free up the movement of players would damage competitive balance. On the contrary, the invariance proposition (Proposition 3) still holds in the more general case. Once more, the allocation of the right to sell the players services affects only the distribution of the revenues they are expected to yield and not the distribution of talent in the league. Thus, it remains that who gets what share of the trades is not relevant to competitive balance.

Youth Team Policies: A Challenge to the Invariance Proposition

This latter proposition is also subject to challenge, however. As discussed above, although there have been some similar developments in European and U.S. sporting labor markets, a key difference is the potential for European sporting contract length to vary more than in the United States. An expressed implication associated with the Bosman ruling was that contract duration would increase.[15] Indeed, since the ruling, a number of European football clubs have signed consistently long-term agreements with their young players.[16] In the context of the previous analysis,

although the above observation can be explained by the ability of team owners to extract a share of the surplus generated by players moving to other teams, the invariance proposition makes clear that such transfers do not affect the league's distribution of talent. In this final section, another plausible explanation that challenges the very premises of the invariance proposition is considered.

The Coase theorem that lies at the heart of the invariance proposition holds in environments where transaction costs are absent and, hence, parties have the ability to write complete contracts. In recent years, the fast developing literature on incomplete contracts has been addressing problems arising when parties are unable to write contracts contingent on all future events that might affect their relationship.[17] This inability might stem, for example, from the fact that certain actions taken by the two parties cannot be verified by a third party (e.g., court of law). If as a consequence of these actions there are rents to continuing the relationship, then, unless the two parties can make long-term commitments, these actions might be taken at sub-optimal levels.

These ideas have been extensively explored in the labor economics literature.[18] Firms often make investments that enhance the future productivity of their workers. In many cases it is not feasible to specify all the characteristics (e.g., quality) of these investments in sufficient detail so that they can be verified by a third party. In the absence of long-term agreements this might lead to underinvestment, especially if these skills are useful to other firms.[19] If workers are not financially constrained, then it is optimal for the workers to finance their training. However, because investments are not verifiable it is doubtful that workers can raise funds in the financial markets.

The market for professional footballers offers a straightforward application of these ideas. The aim of youth-team policies is to improve the skills of young players. The implementation of these policies requires costly investments in training grounds, other related facilities, and professional staffing. Teams expect that the benefits of their investment will be realized when the players mature and become regular first team members. However, at that time other teams might attempt to attract these players and capitalize on another team's investment. At this point, the issue of verifiability becomes crucial. It might be impossible for courts to verify that a player's talent is the result of the team's youth policy. As a consequence, team owners might be reluctant to invest in these policies if they believe that they might not be compensated.

A long-term binding contract, effectively, transfers to teams the ownership of the right to sell the players' services. Teams and players can agree on a salary profile that over time compensates teams, thus providing them with the right incentives to invest. Furthermore, in the case of a transfer, the team can capture part of the surplus, because the original contract can only be broken by mutual consent. In contrast to the invariance proposition (Proposition 3), the foregoing discussion clearly suggests that the league's talent distribution depends on the allocation of control

rights because the latter have a direct effect on teams' decisions to invest in youth training policies.

Thus, the foregoing analysis suggests that reforms that seek to promote more movement toward free agency may well indeed have negative effects on competitive balance. More generally, the discussion reinforces the view that the specific ways in which sporting leagues evolve does matter in understanding the development of competitive balance. Consequently, further detailed league-specific research that captures the institutional realities of particular sporting leagues may yield interesting insights into this development as opposed to research that may overstate similarities.

CONCLUSION

This article has presented a theoretical framework of sporting leagues that shares the objectives of clubs postulated in U.S. literature but reflects the bargaining process that can be argued characterizes soccer labor markets. Beginning the analysis from a simple benchmark case, it has been argued that, along with the rest of the literature, if an unrestricted transfer market exists and capital markets are perfect, then free agency does not threaten competitive balance. This is a restatement of the invariance proposition that is essentially drawn from U.S. literature. However, it was also stressed that the invariance proposition does not imply that transactions in the market for professional players cannot have destabilizing effects. The multiteam league model clearly suggests that from the moment there is a digression from the limiting benchmark case, then the above transactions can have significant effects on the distribution of talent across teams.

An interesting suggestion for future research would be to combine the benchmark model with features of the Quirk and Fort (1992) and Vrooman (1995) model. In these models, revenues depend on market size and relative team strengths. By abandoning the linear structure of the model transfers, acquisitions and exchanges of players can impose negative externalities on teams not participating in these transactions with potential adverse consequences for the league's competitive balance. Furthermore, imperfect capital markets can limit the ability of poor teams to finance the acquisition of players. Finally, it has been argued that to understand the recent proliferation of long-term contracts in European football, it is important to appreciate that the invariance proposition ceases to hold in an incomplete contacting environment. Thus, it is argued that the above theoretical discussion provides some basis for understanding the advocacy of the transfer system in the literature and by implication the wider public.

NOTES

1. For a more extensive discussion of the institutional details, see Downward and Dawson (2000), Szymanski and Kuypers (1999), and Fort and Quirk (1995).

2. See, for example, Scully (1989) and Szymanski and Kuypers (1999).

3. There are also salary caps that set a limit on salaries based on a fixed percentage of turnover and are imposed to ensure the financial stability of teams.

4. The implications of increasing contract length are discussed further below.

5. Hylan, Lage, and Treglia (1996); Kahn (1993); Krautmann and Oppenheimer (1994); MacDonald and Reynolds (1994); and Marburger (1994) provide examples of bargaining in U.S. sporting labor markets.

6. For example, in English football, transfer surpluses have recently amounted to more than 30% of gate revenues; Dobson and Goddard (1998).

7. What is important for our purposes is that there is a limited supply of high-skill talent. Our modeling restriction to three types of players is only introduced for analytical convenience.

8. We assume that the season is long enough so that this probability also represents the actual time that the player will be available, that is, expected talent in use is equal to actual talent in use.

9. It is assumed that if the newly hired player has the same talent as the player hired before him or her, then he or she will be asked to play only if the other player is not available.

10. Basic skill players might be needed for replacement. Clearly, the historical duration of these sequences is somewhat arbitrary. Consequently, one could argue that they might represent the idea of limited transfer windows currently in place in soccer, in which case comparing this scenario with that below when the assumption is relaxed indicates some likely implications of any move toward more free agency.

11. The reason behind this assumption is to simplify the owner's decision process. More specifically, when the opportunity to sign a new low-talented player comes they will have the incentive to do so rather than wait for a better one.

12. It is implicitly assumed that players sign one-period contracts. The issue of optimal contract length is discussed below.

13. There is an implicit assumption behind this statement: The player's bargaining power is the same across teams.

14. The case of task complementarity, discussed above, is a good example.

15. Sir John Hall, Chairman of Newcastle United Football Club PLC, quoted in the *Times* (September 20, 1995).

16. For example, Manchester United. Simmons (1997) illustrates other examples.

17. See Hart (1995).

18. See Malcomson (1997) for a comprehensive review of this literature.

19. See Stevens (1996) and Adnett and Bougheas (1998) for the case of transferable training.

REFERENCES

Adnett, N., & Bougheas, S. (1998, September). *Transferable training and the hold-up problem: Policy implications*. Paper presented at the Meetings of the European Economic Association, Berlin.

Antonioni, P., & Cubbin, J. (2000). The Bosman ruling and the emergence of a single market in soccer talent. *European Journal of Law and Economics*, *9*, 157-173.

Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: Theory and evidence. *Applied Economics*, *25*, 1467-1476.

Dobson, S., & Gerrard, B. (1999). The determination of player transfer fees in English Professional Soccer. *Journal of Sport Management*, *13*(4), 259-274.

Dobson, S., & Goddard, J. (1998). Performance, revenue and cross subsidization in the football league 1927-1994. *Economic History Review*, *51*(4), 763-785.

Dobson, S., & Goddard, J. (2001). *The economics of football*. Cambridge, UK: Cambridge University Press.

Downward, P., & Dawson, A. (2000). *The Economics of Professional Team Sports*. Routledge: London.

El Hodiri, M., & Quirk, J. (1971). An economic model of a professional sports league. *Journal of Political Economy*, *79*, 1302-1319.

Ericson, T. (2000). The Bosman case: Effects of the abolition of the transfer fee. *Journal of Sports Economics*, *1*(3), 203-218.

Flynn, M. A., & Gilbert, R. J. (2001). The analysis of professional sports leagues as joint ventures. *Economic Journal*, *111*, 27-46.

Fort, R. (2000). European and North American sports differences. *Scottish Journal of Political Economy*, *47*(4), 431-455.

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

Hart, O. (1995). *Firms, contracts and financial structure*. Oxford, UK: Clarendon Press.

Hylan, T., Lage, M., & Treglia, M. (1996). The Coase theorem, free agency and Major League Baseball. *Southern Economic Journal*, *62*, 1029-1042.

Kahn, L. (1993). Free agency, long-term contracts and compensation in Major League Baseball: Estimates from panel data. *Review of Economics and Statistics*, *75*, 157-164.

Kesenne, S. (2000a). The impact of salary caps in professional team sports. *Scottish Journal of Political Economy*, *47*(4), 422-430.

Kesenne, S. (2000b). Revenue sharing and competitive balance in professional team sports. *Journal of Sports Economics*, *1*(1), 56-65.

Krautmann, A., & Oppenheimer, M. (1994). Free agency and the allocation of labor in Major League Baseball. *Managerial and Decision Economics*, *15*, 459-469.

MacDonald, D., & Reynolds, M. (1994). Are baseball players paid their marginal products? *Managerial and Decision Economics*, *15*, 443-457.

Malcomson, J. (1997). Contracts, hold-up, and labor markets. *Journal of Economic Literature*, *35*, 1916-1957.

Marburger, D. (1994). Bargaining power and the structure of salaries in Major League Baseball. *Managerial and Decision Economics*, *15*, 433-441.

Neale, W. (1964). The peculiar economics of professional sports. *Quarterly Journal of Economics*, *78*, 1-14.

Noll, R. (1999). Competition policy in European sports after the Bosman case. In S. Kesenne & C. Jeanrenaud (Eds.), *Competition policy in professional sports: Europe after the Bosman case*. Antwerp, Belgium: Standard Editions.

Primault, D., & Rouger, A. (1999). How relevant is North American experience for professional team sports in Europe. In S. Kesenne & C. Jeanrenaud (Eds.), *Competition policy in professional sports: Europe after the Bosman case*. Antwerp, Belgium: Standard Editions.

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

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

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

Simmons, R. (1997). Implications of the Bosman ruling for football transfer markets. *Institute of Economic Affairs*, *17*(3), 13-18.

Sloane, P. (1971). The economics of professional football: The football club as a utility maximizer. *Scottish Journal of Political Economy*, *18*, 121-146.

Stevens, M. (1996). Transferable training and poaching externalities. In A. Booth & D. Snower (Eds.), *Acquiring skills: Market failures, their symptoms and policy responses*. Cambridge, UK: Cambridge University Press.

Stole, L., & Zwiebel, J. (1996). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies*, *63*, 375-410.

Szymanski, S. (1999). The market for soccer players in England after Bosman: Winners and losers. In S. Kesenne & C. Jeanrenaud (Eds.), *Competition policy in professional sports: Europe after the Bosman case*. Antwerp, Belgium: Standard Editions.

Syzmanski, S., & Kuypers, T. (1999). *Winners and losers: The business strategy of football*. Harmondsworth, UK: Viking Press.

Vrooman, J. (1995). A general theory of professional sports leagues. *Southern Economic Journal*, *61*, 971-990.

Vrooman, J. (2000). The economics of American sports leagues. *Scottish Journal of Political Economy*, *47*(4), 364-398.

Spiros Bougheas has been a lecturer at Nottingham University since 1999. Before he moved to Nottingham, he held a lectureship at Staffordshire University from 1994-1999. His research interests include the application of contact theory to labor and financial markets.

Paul Downward is a reader in economics at Staffordshire University. His research interests include the economics of sports, leisure, and tourism.