

The Champions League and the Coase Theorem

Stefan Szymanski¹

October 2006

Abstract

The Coase Theorem is both one of the simplest and most profound ideas in economics. Coase's insight was first expressed in print as a theorem by George Stigler, following the publication of the famous article "The Problem of Social Cost" by Nobel Laureate Ronald Coase (1960). Stigler stated it thus: "with zero transactions costs, private and social costs will be equal". In this paper the Coase Theorem is approached through the medium of a sports league. While Coase's article dates from 1960, a colleague at Chicago University published a discussion of the market for baseball players in 1956 which almost completely anticipates the more famous paper (Rottenberg (1956)).

JEL Classification Codes: L83

Keywords: sports, Coase Theorem

* Inaugural Lecture presented May 11, 2005. Draft: not for quotation without prior permission. Comments welcome.

¹Stefan Szymanski, Tanaka Business School, Imperial College London, South Kensington campus, SW7 2AZ, UK. Tel : (44) 20 7594 9107, Fax: (44) 20 7823 7685, e-mail: szy@imperial.ac.uk.

1. Introduction

The Coase Theorem is both one of the simplest and most profound ideas in economics. Coase's insight was first expressed in print as a theorem by George Stigler, following the publication of the famous article "The Problem of Social Cost" by Nobel Laureate Ronald Coase (1960). Stigler stated it thus: "with zero transactions costs, private and social costs will be equal". The significance of this statement is that "if private cost is equal to social cost, it follows that producers will only engage in an activity if the value of the product of the factors employed is greater than the value which they would yield in their best alternative use" (Coase (1988), p158). In other words, bargaining in an unrestricted market will produce full economic efficiency (assuming zero transactions costs), obviating the need to invoke government intervention in the form of Pigouvian taxes and subsidies to correct externalities. The implications for social policy are profound- simply by establishing property rights all externalities will be internalised and private transactions will be publicly optimal. Most notably, the lesson of Coase Theorem for environmental economics is that we need only establish property rights over the quantities of greenhouse gases in the atmosphere and toxins in the oceans, and pollution will be controlled at socially optimal, sustainable levels. Many economists would echo the words of Avinash Dixit and Mancur Olson (2000) "In his article 'The Problem of Social Cost', Ronald Coase introduced a very powerful idea of great importance. Coase's article has been arguably the single largest influence on thinking about economic policy for the last three decades. It is one of the most – if not the most – widely cited economics article in recent times".

Big theorems are notoriously difficult to test. Darwinism remains resolutely untestable, no one is holding out much hope of testing the theories of Freud or Marx and even in physics developments such as string theory remain testable only in principle. One problem with big theories is that they require big experiments- experimental frameworks that are capable of capturing a substantial degree of the complexity that a big theory addresses. In the world of economics big theorems such as the fundamental theorems of welfare economics or the law of demand are generally approached through specific examples. In this paper the Coase Theorem is

approached through the medium of a sports league. While Coase's article dates from 1960, a colleague at Chicago University published a discussion of the market for baseball players in 1956 which almost completely anticipates the more famous paper (Rottenberg (1956)). As in any team sport, the players are the principal asset, and teams historically have traded these assets, frequently for cash. In baseball a rule enforced by the owners, known as the Reserve Clause, prohibited players from moving teams without the permission of their current employer, effectively endowing the employer a monopsony right over the income stream of the player. As this restraint came under pressure from the players and their union, the owners sought to defend their rule by arguing that if players were free to move they would quickly migrate to the wealthiest teams, disturbing the essential element of "competitive balance" allegedly fostered by the Reserve Clause. Rottenberg argued, in the manner of Coase, that ownership rules would make no difference to the distribution of talent in a league. If owners controlled the movement of players, trade between clubs would cause each player to move to the location where his (marginal revenue) product is greatest. If players were free to move, bidding by the clubs to hire players would produce the same distribution (the only difference being that any economic rents would now accrue to the player, not the owner).

This paper re-examines the application of the Coase Theorem to the market for players in a sports league. It shows that plausible trading mechanisms will not achieve Coasian efficiency. This result is demonstrated using data from English football. The implications of these results for the development of the UEFA Champions League is then discussed.

2. The Coase Theorem and its discontents

The Coase Theorem has been subject to significant scrutiny in the economics literature and has been widely challenged (see e.g. Ellickson (1991) and Samuelson (1995)). Three examples of academic critiques are discussed here:

- (i) Practicality (e.g. Canterbury and Marvasti (1992)). Even if it is true that costless bargaining with full property rights produces efficiency, many

economists have argued that this is of little practical value, since most market failures which the Coase Theorem addresses refer to situations where property rights are very difficult to define precisely or in a way that that is legally enforceable (e.g. rights over the ocean fisheries- even if territorial waters are assigned, the fish often fail to respect the boundaries so that enforcing rights over fish that temporarily stray into another's jurisdiction is likely to be difficult). Additionally, the relevance of the costless bargaining paradigm is questionable since most of the difficult and important problems arise where bargaining costs are very high (e.g. pollution rights).

- (ii) Tautology (e.g. Usher (1998)). He argues that in a zero transaction cost world efficiency must be guaranteed among maximising agents, regardless of whether property rights exist at all- since otherwise there will exist unrealised gains from trade. Hence while it is strictly true that any allocation of property will produce efficiency in such a world, the existence of property rights is not necessary. The Coase Theorem, stripped of the necessity of property, merely becomes the statement that in a world where agents are willing and able to bargain until all potential gains from trade are realised the outcome will be economically efficient, which, as stated, appears tautological, since economic efficiency is defined as the realisation of all potential gains from trade.
- (iii) Falsity (e.g. Aivazian and Callen (2003)). These authors relate the Coase Theorem to Edgeworth's notion of the core. The core is defined as the set of efficient equilibrium bargains among parties. The perfectly competitive equilibrium of neoclassical economics belongs the set of resource allocations that are in the core, but others may exist as well. Clearly the outcome of Coasian bargaining must be in the core in as well. Different initial allocations of rights might product different allocations of resources, but all outcomes should be in the core (e.g. the allocation of resources will be different if the polluter has the absolute right to pollute compared to a situation where citizens have an absolute right to protection from the effects of pollution, but the Coase Theorem says that the amount of pollution should be fixed at the efficient level in both cases). However, if there is an allocation of property rights exists for which the core is empty-

i.e. there is no equilibrium bargain that is efficient- then the Coase Theorem fails. Aivazian and Callen provide a simple example of just such a case.

While theoretical objections abound, it is perhaps more important to understand whether the implications of the Coase Theorem are really relevant for economic policy. In other words, we need to understand whether, in a world where property rights are well defined and bargaining is not too costly, the outcomes of bargaining are plausibly close to efficiency. The team sports literature has been widely cited as an example of a situation where the Coase Theorem is put to a practical test.

3. The Coase Theorem in the sports literature

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))². This inevitably led to challenges in the courts by the players claiming the right to move freely between employers. Simon Rottenberg's celebrated (1956) article examined this issue and presented the team owner's rationale:

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

² 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.

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

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

(i) Free Agency

The advent of free agency in MLB in 1976 for six year veterans is a clear natural experiment³. 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.

³ In this case the change was exogenous- i.e. not itself motivated by a desire to affect competitive balance (see Bruce Meyer (1995) for a discussion of natural experiments).

Table 1: 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	(+)	(+)

Most of the studies simply look at the standard deviation of win percentages before and after 1976 (Scully (1989), Balfour and Porter (1991), Quirk and Fort (1995), Vrooman (1995), Michael Butler (1995)), while other measures include persistence in win percent (Balfour and Porter (1991), Vrooman (1996)), entropy (Horowitz (1997)), the Hirschman-Herfindahl index (Depken (1999)) and analysis of variance (Eckard (2001)). Most of these studies find either no change (seven cases) or an improvement in competitive balance (nine cases), contrary to the claim of the owners that free agency would reduce competitive balance (four cases only). However, this meta-data is hardly a ringing endorsement for the invariance principle, since “no effect” is reported in only seven out of twenty cases. Of course, it can be argued that many other factors have altered competitive balance (e.g. the increasing dispersion of local

TV revenues), but in that case the data, without controlling for these factors, can hardly be said to represent a test at all.

Some other studies have approached the invariance principle as a direct test of the Coase Theorem and tried to establish whether the distribution of talent in the league has been affected by the introduction of free agency. George Daly (1992) observes that under the Reserve Clause top line players were seldom traded, a situation that has been affected by free agency where the top stars have a choice after six years leading to increased mobility. Timothy Hylan, Maureen Lage and Michael Treglia (1996) in a study of pitcher movements finds that these players have become less mobile since free agency, a surprising result and one that they claim does not support the Coase Theorem. However, Donald Cymrot, James Dunley and William Even (2001) examine player mobility in 1980, controlling for possible selection bias and find that, for that season at least, there was no evidence that restricted players (with less than six years service) enjoyed more or less mobility than unrestricted free agents after controlling for player characteristics.

Daniel Marburger (2002) considers a different implication of the invariance principle. If trade is possible between two independent leagues then it should be more profitable to hire a player from the same league than the rival league. Intra-league trade raises the winning probability of the buying team by more than an inter-league trade, since in the former case not only does the buyer have a larger share of talent, but the seller now has a weaker team. Under the Reserve clause this effect will be built into the seller's price, but under free agency it will not, since the free agent is indifferent to the adverse effect on the team he is leaving. Thus with free agency the relative price of intraleague trades should fall and their share of total trades increase. Marburger found a statistically significant increase in the share of intraleague trades, from 60% to 73%, in MLB 1964 and 1992. This finding seems consistent with the invariance principle.

(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. The Japanese Professional Baseball League adopted a draft system at exactly the same time as MLB, and a study by La Croix and Kawaura (1999) also found that competitive balance improved over time (measured by the Gini coefficient for pennants) in both the Central and Pacific Leagues⁴. As they point out, these results are “virtually identical” to Fort and Quirk’s (1995) results for MLB. Kevin Grier and Robert Tollison (1994) examined the impact of the rookie draft in the NFL by running an autoregressive specification for win percentage together with the average draft order over the previous three to five seasons, and found that a low draft order significantly raises performance. These results seem to provide consistent evidence against the invariance principle and in support of the owners’ stated position.

4. Trading mechanisms and the allocation of talent

One difficulty with much of the preceding analysis is that the relationship between the trading mechanisms in sports leagues and the efficient distribution of talent is poorly defined. The approach followed here is to derive the distribution that teams will select when they maximise profits and to compare this with a plausible candidate for an efficient allocation of talent.

Szymanski (2004a) considered a simple model of competition in a league where demand depends on (a) the success of each team and (b) the degree of competitive

⁴ although the within season measure (standard deviation of win percent) was significant only for the Pacific League.

balance.⁵ The model assumes is a supply of talent to the market, which may be fixed or elastic, and team success, measured by the percentage of games won depends on the share of total talent hired. Teams are assumed to maximise profit. In a league of two teams this boils down a contest success function

$$(1) \quad w_1 = \frac{t_1}{t_1 + t_2}, \quad w_2 = 1 - w_1$$

Where w is win percentage and t is talent hired, which is assumed perfectly divisible, and a profit function, where is here given the simple form

$$(2) \quad \pi_1 = (\sigma - w_1) w_1 - ct_1, \quad \pi_2 = (1 - w_2) w_2 - ct_2, \quad \sigma > 1$$

where c is the (constant) marginal cost of talent and σ indicates that team 1 is capable of generating a larger revenue than team 2 from any given level of success. Note that the demand for competitive balance ensures that revenues are ultimately decreasing in success, but that given the adding up constraint in (1), there is no guarantee of an interior solution.

How do teams choose talent? We can imagine this as a quantity setting (Cournot-type) or price-setting (Bertrand-type) game. As a quantity game, teams allocate a budget to hiring talent and the talent they can hire is proportional to their share of total budgets. The first order conditions for talent choice are therefore

$$(3) \quad \frac{\partial \pi_1}{\partial t_1} = (\sigma - 2w_1) w_2 - cT = 0, \quad \frac{\partial \pi_2}{\partial t_2} = (1 - 2w_2) w_1 - cT = 0$$

where $T = t_1 + t_2$, so that at the Nash equilibrium

$$(4) \quad w_1^* = \sigma / (1 + \sigma)$$

Team 1 dominates ($w_1 > 1/2$) in equilibrium because it has the larger drawing power ($\sigma > 1$). However at equilibrium the marginal revenue of a win for team 1 exceeds the marginal revenue of team 2

⁵ A more general model that is used to analyse the impact of gate revenue sharing is to be found in Szymanski and Kesenne (2004).

$$(5) \quad \frac{\partial R_1}{\partial w_1} = \sigma - 2w_1^* = \frac{\sigma(\sigma - 1)}{(\sigma + 1)} > \frac{\sigma - 1}{(\sigma + 1)} = 1 - 2w_2^* = \frac{\partial R_2}{\partial w_2}$$

This implies a distribution of talent in the league that is not jointly efficient. To see this note that joint profits are

$$(6) \quad \pi_1 + \pi_2 = (1 + \sigma)w_1 - 2w_1^2 - cT,$$

which is maximized when

$$(7) \quad w_1^M = (1 + \sigma)/4 > w_1^*$$

Hence the quantity bidding mechanism entails “too much” competitive balance at the Nash equilibrium. Intuitively, this result is a consequence of asymmetry. Competition always involves an externality- each team’s actions under competition fails to account for the negative effect that actions have on rivals’ profits. The externality imposed by the team with the lower win percentage in equilibrium is bigger precisely because the big team loses more than the small team when its rival wins more.

Dakhliya and Pecorino (2004) consider a rent-seeking model where teams not only bid for a quantity of talent but also submit a bid for the wage rate per unit of talent. If each team offers the same wage rate then the Nash equilibrium distribution of talent will be the same as above. However, if one team bids higher than the other it can attract all the talent, generating a corner solution. In their model, where teams only have a demand for winning and there is no value in competitive balance, they show that the dominant team will be willing to pre-empt all of the talent by offering a bid at which its rival’s demand for talent is zero, as long as the quantity of talent is not too great. However, if the supply of talent is large enough, pre-emption is not profitable, given that the team would have to hire all of the talent in order to pre-empt the market.⁶

⁶ Efficiency in their model of pure rent seeking (à la Tullock) is slightly peculiar, in that the most efficient result is for team 1 to win all the time since it values the payoff more. Moreover, even if team 1 pre-empts all the talent, it only need to employ ε of it to win with certainty, since team 2 hire zero in equilibrium. The point here is that the simple rent seeking game requires more structure in order for an interior solution to be efficient. If, for example, there is a demand for competitive balance, then an interior solution can be efficient.

The incentive to pre-empt can be identified by comparing the profit level at an interior equilibrium for a given marginal cost of talent with the profit made by one team raising price by ε above marginal cost, hiring all the talent and winning all the time. If this deviation can be shown to be profitable then a form of Bertrand competition will ensue.

To derive the condition for a profitable deviation first consider the demand for talent at the interior equilibrium. First note that $w_1 / w_2 = t_1 / t_2 = \sigma$. Writing (3) in terms of t_1 and t_2 , substituting for t_2 we obtain

$$(8) \quad t_1^* = \frac{\sigma^2(\sigma-1)}{(1+\sigma)^3 c}, \quad t_2^* = \frac{\sigma(\sigma-1)}{(1+\sigma)^3 c}$$

This implies that team 1 makes profit equal to

$$(9) \quad \pi_1^* = \frac{\sigma^4 + \sigma^2}{(1+\sigma)^3}$$

It can now be shown that team 1 would want to pre-empt by offering a wage rate “ $c + \varepsilon$ ” if the total supply to the market $T^S = t_1^* + t_2^*$. For ε small enough, the profits from pre-emption are

$$(9) \quad \sigma - 1 - cT^S = \frac{\sigma^3 - 1}{(1+\sigma)^2}$$

Pre-emption can therefore be profitable if

$$(10) \quad \sigma^3 > \sigma^2 + \sigma + 1 \Rightarrow \sigma > 1.84 \quad \text{approx}$$

Thus pre-emption can be a profitable strategy if the dominant team is sufficiently large. As Dakhli and Pecorino show, pre-emption additionally requires that the marginal profit of team 2 is negative when it hires zero units of talent, requiring that $bT^S > 1$, where b is the pre-emptive bid of team 1 and T^S is the total supply of talent. Additionally, however, if $T^S > t_1^* + t_2^*$ then it becomes less and less likely that pre-emption is profitable.

Thus, when there is bidding for talent, there is a possibility that we shift from an inefficient interior solution to a pre-emptive corner solution, which is also inefficient. In both cases it is assumed that any market clearing mechanism must involve identical

treatment for each unit of talent. If talent were sold for different prices to different teams then there would be no equilibrium among the players, since low paid players would with identical skills be willing to move to high paying teams.

5. Some empirical evidence

The theory described in the previous section implies a simple empirical experiment. Since data exists for attendance at league matches and the success rates of teams, it is feasible to identify a statistical relationship for each team in a league. From this relationship it is possible to estimate a distribution of wins that would maximise attendance at home matches, and to compare this distribution with the actual distribution, which may be imagined as an approximation of the competitive Nash equilibrium. Szymanski (2004b) estimates this relationship for Major League Baseball, while Leach and Szymanski (2005) do the same for the lower divisions of English soccer. The first of these papers shows quite clearly that the attendance maximising distribution of wins will be much more unbalanced than what we observe in practice, and that the optimal distribution gives more weight to already dominant teams.

The key to the analysis is to estimate separate slope parameters for a quadratic equation relating home attendance to wins. While this has seldom been done before (a notable exception is Dobson and Goddard (2001)), the regression fits the data well in the case of English football, although the analysis must be restricted to the lower three divisions because such a large fraction of Premier League matches are sold out, and therefore attendance in these cases is largely insensitive to success. But in general, attendance is strongly dependent on success in winning matches- this should not be surprising given that in any team sports league the majority of those attending a game are fans of the home team.

Leach and Szymanski used attendance and wins data for all professional league clubs in England over the last thirty years and produced club by club estimates, a sample of which is shown in Table 2.

Table 2: Estimated sensitivity of attendance to wins for a sample of clubs.

<u>Club</u>	<u>Level Term (WPct)</u>		<u>Squared Term (WPct²)</u>	
	<i>Coefficient</i>	<i>Std Error</i>	<i>Coefficient</i>	<i>Std Error</i>
<i>Aldershot Town</i>	21450	4627	-15159	6752
<i>Aston Villa</i>	141207	80300	-103652	115000
<i>Barnsley</i>	40018	22100	-44124	35200
<i>Birmingham City</i>	73007	18000	-60110	27200
<i>Blackburn Rovers</i>	36529	21900	-24005	34200
<i>Blackpool</i>	25535	6764	-17645	8523
<i>Bolton Wanderers</i>	74357	9735	-53469	16100
<i>Bournemouth</i>	21700	3694	-11255	6926
<i>Bradford City</i>	59411	17100	-51505	23700
<i>Brentford</i>	25103	7938	-10149	11300
<i>Brighton & Hove</i>	92932	9154	-84248	12000
<i>Bristol City</i>	72672	33500	-84299	65000
<i>Bristol Rovers</i>	29861	1951	-22318	3664
<i>Burnley</i>	21804	11800	-6079	21400
<i>Bury</i>	21555	1724	-13862	2604

Source: Leach and Szymanski (2005)

The size of these sensitivities varies considerably by club and the restriction that coefficients are identical across clubs is easily rejected. The regressions seem well defined, with level and quadratic terms having the right signs in all cases and about three quarters of the estimates are significant at the conventional level. Using these estimates we can proceed to estimate the attendance maximising distribution of wins for any team in any particular season. By way of example, consider the estimates for two teams, Sheffield Wednesday and Watford. These were selected for simple arithmetical expedient that the quadratic terms for these two clubs are more or less identical. The two equations are:

Sheffield Wednesday: $Attendance = 75000(\% \text{ wins}) - 44000(\% \text{ wins})^2$

Watford: $Attendance = 59000(\% \text{ wins}) - 44000(\% \text{ wins})^2$

From these estimates we can calculate the attendance maximising percentage of wins for each team- 85% for Sheffield Wednesday and 67% for Watford. In each case attendance does start to decline when the team wins too big a share of its matches. Imagine a league which contained only these two teams. Clearly they cannot jointly reach their individual optimum, since the total number of wins must sum to 100%. A

central planner would choose an allocation of wins between the two teams which equalised the marginal attendance from a win for each team- in other words where

$$75000 - 88000w_1 = 59000 - 88000w_2$$

which implies $w_1^M = 59\%$ is the optimal win percentage for Sheffield Wednesday. However, if we now consider the Nash equilibrium of a talent hiring contest between the two teams, the marginal benefit of unit of talent for each team is

$$(75000 - 88000w_1)w_2 = (59000 - 88000w_2)w_1$$

which implies that at the Nash equilibrium Sheffield Wednesday's win percentage would be $w_1^* = 56\%$, less than the attendance maximising optimum, as illustrated in figure 1.

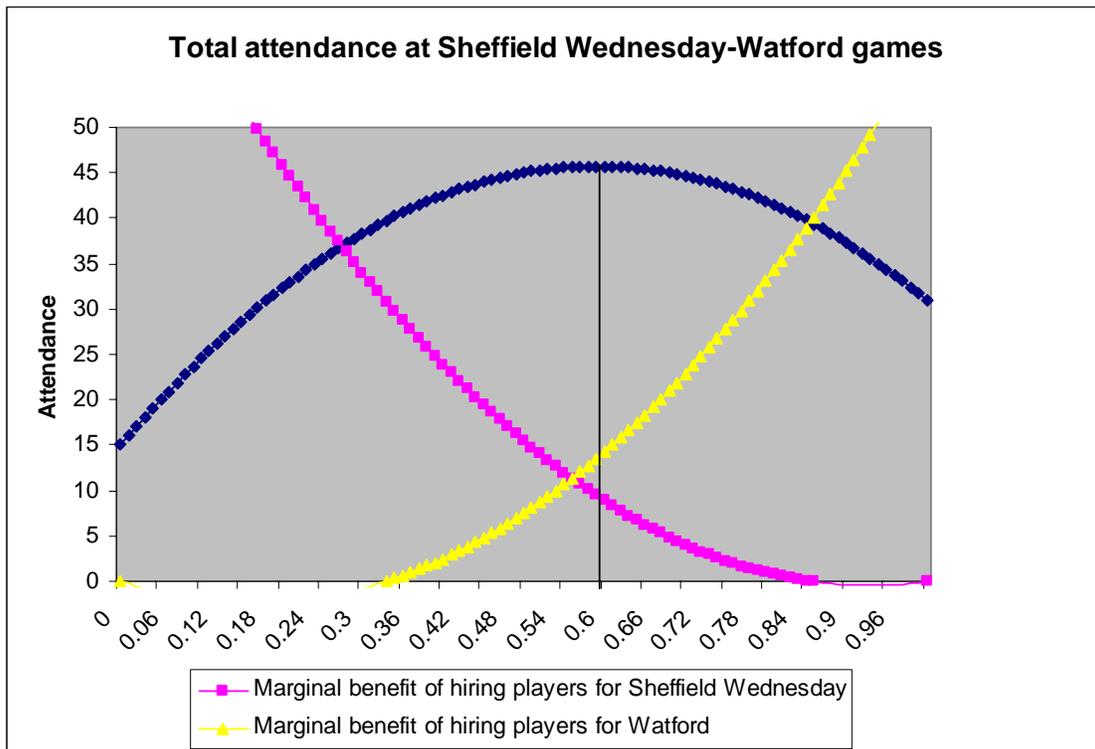


Figure 1

The calculations are somewhat more tedious, but it is straightforward to reproduce this result for any league division and any particular season, and the result always

appears to be the same- the attendance maximising distribution of wins is much more unequal than that which we observe in practice (which is presumed to be a Nash equilibrium). For example, table 3 below shows the actual standard deviation of wins for the second tier of English football over the last decade and the attendance maximising distribution.

Table 3: Standard Deviation of Actual and Optimal Win Percentages for the second Tier of English Football 1994-2003

Season	Second Tier Actual Standard Deviation of Win Percentage	Second Tier Constrained Standard Deviation of Optimal Win Percentage
1994	0.086	0.241
1995	0.081	0.187
1996	0.072	0.232
1997	0.094	0.232
1998	0.115	0.235
1999	0.116	0.242
2000	0.111	0.237
2001	0.115	0.246
2002	0.112	0.227
2003	0.106	0.203

Source: Leach and Szymanski (2005)

Thus the empirical evidence on attendance seems to support the theory. Several objections have been raised to these results.⁷ Firstly, it might be said that attendance is only one source of revenue, and that for many leagues TV income is as important. As a matter of fact, this may be true of the Premier League but is certainly not true for the lower divisions that attract only small TV revenues. If TV income did matter, and TV viewers wanted balanced contests (being less partisan than the typical supporter attending a match) then these results would be to a degree mitigated. However, even TV fans can be partisan, and even in the most TV oriented leagues clubs still obtain a large fraction of their income from gate money. Secondly, it might be said that a very unbalanced contest is all well and good in the short run, but that a league that was persistently unbalanced would be likely to lose its supporter base. This point is in fact

⁷ I am grateful, inter alia, to Robert Sandy, Peter Sloane and Andy Zimbalist for their insights.

addressed in the estimation technique, since the reported estimates are derived from a regression which includes lagged variables on the right hand side, and these are used to estimate a long run relationship. Critics might say that these estimates are unlikely to give a reliable indication about a much more unbalanced world since the optimal levels of balance estimated have never been observed in practice. However, it has been shown by Buzzacchi et al (2003) that, despite the fact that European football is much less balanced than the US major leagues (measured by the likelihood of teams entering the highest ranked positions), it has not been the case that attendance at European football has been declining relative to attendance at the US major leagues. What the data certainly does not suggest, contrary to the claims of many federations and sports economists, is that increasing competitive balance will lead to an increase in attendance in the short to medium run.

6. The future of the Champions League

In what sense does any of this contradict the Coase Theorem? What has been shown in the model is that different trading mechanisms generate different allocations, none of which are efficient. However, the model does not exhaust the potential range of mechanisms nor does it show that other forms of bargaining between teams would not achieve the productively efficient allocation. Indeed, owners of US major league teams might argue that they adopt the panoply of restrictive devices for which they are famous precisely to combat these potential inefficiencies (roster limits, salary caps, etc, etc). However, perhaps the model does shed some light on what a Coasian mechanism must look like. The analysis suggests that simple rules often fail to reach optimality. It suggests agents need to have some idea of what constitutes a globally efficient outcome, and that then they may need to adopt quite complex bilateral bargains to achieve the efficient outcome. Reliance on conventional market signals (e.g. prices, quantities) may not be enough. Significant planning and coordination among firms may be required to reach a Coasian bargain- often in ways that might, when viewed in a mundane fashion, violate antitrust laws. The most obvious mechanism for dealing with the inefficiency of specific trading mechanisms is to integrate the teams in the league into a single enterprise. In sport, it has been argued that a league should indeed be viewed as a single entity (see e.g. Roberts (1984))

since each the “product” is the league competition, rather than any particular games played in the league. This view is much debated in the sports literature, but the more general implication that the solution to the problem of Coasian bargaining is monopoly is surely not a very attractive one.

In the context of a sports league, such mechanisms are generally accepted on the grounds that teams must cooperate in order to produce “the product” defined as the league competition, even if the single entity defence is rejected. In this view, leagues are treated as a joint venture (Flynn and Gilbert (2001))). In the US, redistributive mechanisms are supported in the name of competitive balance, and are generally established in order to redistribute resources from strong teams to weaker teams. It could be argued that even if redistribution in this direction does not maximise attendance in the short run, it will serve to maximise interest in the long run, since excessively unbalanced competitions may cause loss of interest in the long run.⁸ However, such effects are difficult to prove, and the evidence that many European soccer leagues have remained highly popular over many years, despite being consistently and significantly unbalanced, suggests that this link may be weak or even non-existent. Moreover, even in the US redistributive measures have not been deemed adequate to maintain balance within a league. Team relocation is a controversial issue, with many motivations being attached to desire of team owners to move from one city to another (e.g. expansion into larger markets, extracting stadium subsidies, pre-empting the entry of rival leagues⁹). However, it is clear that if a team owner relocates from a weak drawing city to a strong drawing city competitive balance is likely to improve. Carlton et al (2004) go so far as to argue that league restrictions on the right of teams to relocate are an efficient mechanism for internalising the externality caused by the loss of rivalry when a franchise moves (initially, any traditional rivalry that many have existed between a team at its old location and nearby cities will be diluted or lost completely imposing an economic cost on these other franchises). They supply

⁸ Humphreys (2002), who shows that a combination of within and between season competitive balance affects attendance. The relative decline of interest in the English FA Cup as the financial gap between the different divisions of English football has grown, might also be interpreted in this light (Szymanski (2001).

⁹ See e.g. Zimbalist (2003), pp29-33, Quirk and Fort (1992) pp298-302.

some evidence from ice hockey (the NHL) to show that franchise movements have only been permitted where they have been expected to add to league-wide profits.¹⁰

In the context of European football the system of promotion and relegation should automatically ensure that teams with roughly equal levels of support should reach similar competitive levels (i.e. divisions) without the need for team relocation. Tied to the very strong evidence that (a) team expenditure generates success on the pitch and (b) team success generates revenue (e.g. Szymanski and Smith (1997), Szymanski and Kuypers (1999), Forrest and Simmons (2002)), it is clear that both intra- and inter-divisional league rankings in European football should accurately reflect revenue-generating potential. The difficulty in Europe, however, is that this system still fails to produce competitive balance in domestic leagues, which tend to be dominated by a very small number of teams.

Szymanski and Zimbalist (2005) discuss the extent to which the national leagues of countries such as Portugal, Norway, Belgium, Greece, Turkey, among others, have been historically been dominated by between one and three teams. Smaller European countries have tended to be unbalanced because teams from only one or two dominant cities, usually the capital, have resources that far exceed those of teams representing smaller towns. Even in larger countries where it might be expected that a larger national population will generate a more diverse set of teams (e.g. Italy, Germany, England), competitive balance, measured by the number of teams effectively contesting the championship, is much weaker than in the US major leagues.

One plausible reason for this historical lack of balance is the balkanised nature of European football. If almost the entire world supply of baseball talent is insufficient to maintain competitive balance among 30 teams, as Major League Baseball appears to believe (see e.g. Levin et al.), then what chance is there that the world's supply of football talent, divided between 50 or so European domestic leagues, will stretch far enough to maintain competitive balance in each? Historically, the talent has been spread thinly across the nations, particularly since mobility was limited in the pre-

¹⁰ This is not quite the same as a competitive balance argument, but the general proposition is consistent with a competitive balance justification for relocation if increasing competitive balance increases league-wide profits.

Bosman era by various national restrictions. Post- Bosman we have seen a dramatic increase in player mobility combined with even greater concern about competitive. This concern has taken two forms:

- the fear that teams from larger nations dominate Europe at the expense of teams from smaller countries
- the fear that strong clubs in larger countries increasingly dominate national competition

The common factor linking these two fears is the Champions League. UEFA have expressed concern about the effect that the Champions League is having on the distribution of resources in Europe and the level of competitive balance:

“The Champions League has created, in almost every country in Europe, an elite of rich clubs whose increasing dominance is turning their domestic title races into predictable turn-offs for fans, UEFA are warning” -Observer, November 7, 2004

“UEFA created this fantastic competition in 1992, but that it has now become a monster that has produced this unequal struggle between haves and have-nots in countries across Europe” - Lennart Johansson, UEFA President

“It's a serious concern for us that in many European countries only a small number of teams can win the domestic league title...This competitive imbalance is not unique in England, but it is quite pronounced there.” -William Gaillard, UEFA Director of Communications.

It is not hard to see how the Champions League could have added to the level of competitive imbalance. Champions League revenues derive principally from gate receipts and from the distribution of TV revenues. Each team retains the gate revenues from its home matches, while in the 2003/04 season around £280 million was distributed between the 32 teams that participated in the group stages. Exactly half of these teams came from the big five TV markets (England, France, Germany, Italy and Spain) and these teams received over 70% of the money distributed. Partly this

reflected the greater success of clubs from these countries, since the payments are partly based on results, but in part it reflected pure pulling power. For example, the eventual winners, Porto, received a smaller distribution than Arsenal, Monaco, Manchester United or Chelsea.

The resources that are distributed to these teams are likely to be used to invest in talent in order to dominate the competition in the future. For example, a modest Champions League run to the quarter finals for Manchester United is likely to generate income in the region of around £30 million (about £20 million in broadcast fees and £10 million in ticket sales), while more than half of the Premier League has a total annual income of less than £50 million. It is hard to see how the smaller teams could generate the financial resources to compete effectively in the domestic league.

It would appear that the Champions League has created chronic imbalance both inside the competition itself and outside in the domestic league championships. UEFA's response has been to introduce restriction on the size of squads fielded by teams in the Champions League, and to oblige teams to field at least four players that were trained by the club, and additionally four players that were trained inside the national association. These rules were agreed at a UEFA Congress in April 2005 and their introduction was explained by UEFA thus: "restoring football's competitive balance is one of the reasons why UEFA is proposing new rules on locally trained players"¹¹. However, when introducing these proposals, UEFA Chief Executive Lars-Christer Olsson "admitted that there had been negative responses to the proposals from some major leagues and their larger clubs".¹² In fact, these proposals seem quite mild compared to some of the restrictions to be found in North America, and yet even these seem to meet resistance from the dominant teams.

One explanation for this apparent difference in attitude is the following. All teams in the US major leagues are recognised to be the dominant teams and have little or no fear of being disenfranchised, thanks to the "closed" league system. In Europe, however, presently dominant teams that make concessions to weaker teams have

¹¹ "UEFA out to get the balance right", Thursday, 3 February 2005, Mark Chaplin. <http://www.uefa.com/uefa/news/Kind=128/newsId=277348.html>, last accessed 8/5/05

¹² Ibid

every reason to fear that in the future they will sink into obscurity- and one need only consider the present status of former European Champions such as Nottingham Forest (outside of the Premier League for the last six seasons and relegated to the third tier this year) to see how severe the penalty for poor performance can be. In the US, concessions by currently strong teams are expected to be repaid in future when the tables are turned. Hence all can be persuaded to make agreements that are collectively optimal. Under the promotion and relegation system, however, concessions that might make sense collectively do not make much sense individually, and certainly make much less sense than they do in the context of a closed league (see e.g. Szymanski and Valletti (2003)). This is not to say that the promotion and relegation system is itself a bad thing- indeed Ross and Szymanski (2002) argue that many of the inefficiencies observed in the US closed league system would be remedied by the application of promotion and relegation. The point is that reaching agreement on decisions that are collectively efficient in sports leagues is extremely difficult and the actual decisions made often seem to fall well short of Coasian efficiency.¹³

Strong teams that compete inside the Champions League also perceive themselves to face an alternative to sharing income and resources with a large number of weaker teams. Since the early 1980s there has been more or less persistent discussion about the creation of a European Superleague. Such a league could be formed by an elite group of European teams on a closed or near-closed basis. There are clearly many formats that such a league might take- Hoehn and Szymanski (1999) discussed one such format- but it is clear that such a league would be likely to be built more around TV income. For example, if the 16 large market teams competing in the Champions League can generate an income of around £200 million from playing around 40 matches, a full scale European Superleague of twenty teams might be capable of generating ten times this amount from a schedule of 380 games. £2000 million, however, is more than the combined total turnover of Europe's twenty richest teams, according to this year's Deloitte and Touche rich list.¹⁴ A figure of £2000 million for

¹³ Ross and Szymanski (2005) argue that part of the problem lies in the vertical integration of championship organising and club management functions that prevails in most team sports leagues. Efficient structures are more likely to emerge when organisers are distinct from competitors, as is the case with many individualistic sports contests such as marathon running or golf.

¹⁴ "World's top 20 football clubs set to break €3 billion income mark in 2005" Deloitte & Touche, 2/17/05.

European Superleague TV rights would be in line with the six-eight year deals recently signed by the NFL in the US for about \$3000 million per year, and indeed such thoughts may well be tied up with the continuing interest of the Michael Glazer, owner of the NFL's Tampa Bay Buccaneers, in acquiring Manchester United. Any such income, of course, would be before considering potential income from ticket sales, merchandising, sponsorship and so on.

The fact that a breakaway Superleague has been considered many times before is no secret. Most recently, plans for a mid-week league developed by Media Partners, associated with Silvio Berlusconi, owner of AC Milan, were eventually shelved in 1998 following reform of the Champions League that favoured the largest teams in Europe. In recent years there have continuous rumours that G-14, a lobby group of the top European clubs formed in 2000, is preparing its own plans for the formation of a Superleague. Of course, critics would accuse any clubs that were part of such a plan of financial greed and a betrayal of the traditions of the sport. However, it should also be observed that a Superleague of 30 clubs organised on lines similar to the US major leagues would be likely to be more balanced than the present Champions League or any domestic league, even absent redistributive measures. This is simply because the distribution of talent among the 20 or 30 largest European teams is already more balanced than in existing competition.¹⁵ By eliminating competition from the weakest teams, the competitive externality identified in this paper would be mitigated, if not eliminated. This would seem clearly to involve the kind of Coasian bargaining described in this paper. Whether any such bargains come to pass remains to be seen.

http://www.deloitte.com/dtt/press_release/0,1014,sid%253D1017%2526cid%253D74209,00.html, last accessed 8/5/05

¹⁵ Note that the Champions League does not include all of the strongest teams since each country is limited by the number of teams that can compete, a restriction which works against some of the largest clubs.

References

- Aivazian, Varouj and Callen, Jeffrey (2003) "The Core, Transactions Costs, and the Coase Theorem", *Constitutional Political Economy*, 14, 287-299.
- Balfour, Alan and Philip Porter. 1991. "The Reserve Clause in professional sports: legality and effect on competitive balance" *Labor Law Journal*, 42, 8-18.
- Butler, Michael. (1995) "Competitive Balance in Major League Baseball" *The American Economist*, 39, 2, 46-52.
- Buzzachi, Luigi, Stefan Szymanski and Tommaso Valletti. 2003. "Static versus Dynamic Competitive Balance: Do teams win more in Europe or in the US?" *Journal of Industry, Competition and Trade*, 3, 3, 2003, 167-86
- Canterbery, E. and Marvasti, A. (1992) "The Coase Theorem as a negative externality", *Journal of Economic Issues*, XXVI, 4, 1179-1189
- Carlton, Dennis, Alan Frankel and Elisabeth Landes, 2004, "The Control of Externalities in Sports Leagues: An Analysis of Restrictions in the National Hockey League", *Journal of Political Economy*, 2004, vol. 112, issue S1, pages S268-S288
- Coase, Ronald (1988) "Notes on The Problem of Social Cost" in *The Firm and the Market*, University of Chicago Press.
- Coase, Ronald (1960) "The Problem of Social Cost", *The Journal of Law and Economics*, 3, 1-44
- Cymrot, Donald, James Dunlevy and William Even. 2001. " 'Who's on first' : an empirical test of the Coase Theorem in baseball" *Applied Economics*, 33, 593-603.
- Dakhliia, Sami and Paul Pecorino, 2004, "Rent-Seeking with Scarce Talent: A Model of Preemptive Hiring" University of Alabama Economics, Finance and Legal Studies Working Paper No. 04-09-01
- Daly, George and William Moore. 1981. "Externalities, property rights, and the allocation of resources in Major League Baseball" *Economic Inquiry*, 29, 77-95.
- Depken, Craig. 1999. "Free-Agency and the Competitiveness of Major League Baseball" *Review of Industrial Organization*, 14, 205-217.
- Dixit, Avinash and Mancur Olson (2000) "Does voluntary participation undermine the Coase Theorem" *Journal of Public Economics*, 76, 309-335
- Dobson, Stephen and John Goddard. 2001. *The Economics of Football*. Cambridge University Press.
- Eckard, Woodrow. 2001. "Free Agency, competitive balance and diminishing returns to pennant contention" *Economic Inquiry*, 39, 3, 430-443.

- El-Hodiri, Mohamed and James Quirk. 1971. "An Economic Model of a Professional Sports League" *Journal of Political Economy*, 79, 1302-19
- Ellickson, R. (1991) "The case of Coase against Coaseanism" *Yale Law Journal*, 99, 611-33
- Flynn, Michael and Richard Gilbert. 2001. 'An Analysis of Professional Sports Leagues as Joint Ventures' *Economic Journal*, 111, F27-F46.
- Forrest, David and Robert Simmons. 2002. "Team salaries and playing success in sports: a comparative perspective," *Zeitschrift für Betriebswirtschaft* Vol. 72, No. 4.
- Fort, Rodney and James Quirk. 1995. "Cross Subsidization, Incentives and Outcomes in Professional Team Sports Leagues" *Journal of Economic Literature*, XXXIII, 3, 1265-1299
- Grier, Kevin and Robert Tollison. 1994. "The rookie draft and competitive balance: the case of professional football" *Journal of Economic Behavior and Organization*, 25, 293-298.
- Hoehn, Thomas and Stefan Szymanski. 1999. "The Americanization of European Football" *Economic Policy* 28, 205-240.
- Horowitz, Ira. 1997. "The increasing competitive balance in Major League Baseball" *Review of Industrial Organization*, 12, 373-387.
- Humphreys, Brad. 2002. "Alternative measures of competitive balance in sports leagues" *Journal of Sports Economics*, 3, 2, 133-148.
- Hylan, Timothy, Maureen Lage and Michael Treglia. 1996. "The Coase Theorem, free agency, and Major League Baseball: a panel study of pitcher mobility from 1961 to 1992" *Southern Economic Journal* 62, 1029-1042.
- La Croix S. and Kawaura A. (1999) "Rule changes and competitive balance in Japanese professional baseball" *Economic Inquiry*, 37, 2, 353-368.
- Leach, Stephanie and Stefan Szymanski, 2005, "Tilting the Playing Field: Why a sports league planner would choose less, not more, competitive balance: The case of English football" mimeo, Tanaka Business School, Imperial College London.
- Levin, Richard, George Mitchell, Paul Volcker and George Will. 2000. *The Report of the Independent Members of the Commissioner's Blue Ribbon Panel on Baseball Economics*. NY: Major League Baseball.
- Marburger, Daniel. 2002. "Property Rights and Unilateral Player Transfers in a multiconference sports league", *Journal of Sports Economics*, 3, 2, 122-132

- Quirk, James and Mohamed El Hodiri. 1974. "The economic theory of a professional sports league" in *Government and the Sports Business*, Roger Noll, ed. Washington: Brookings Institution.
- Quirk, James and Rodney Fort. 1992. *Pay Dirt: The Business of Professional Team Sports*, Princeton N.J.: Princeton University Press.
- Roberts, Gary. 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
- Ross, Stephen F. and Stefan Szymanski. 2002. "Open competition in league sports" *Wisconsin Law Review*, vol 2002, 3, 625-656.
- Ross, Stephen F. and Stefan Szymanski. 2004, "Antitrust and Inefficient Joint Ventures: Why Sports Leagues Should Look More Like McDonald's and Less Like the United Nations", mimeo, College of Law, University of Illinois, Urbana-Champaign
- Rottenberg, Simon. 1956. "The baseball player's labor market" *Journal of Political Economy*, 64, 242-258
- Samuelson, Paul (1995) "Some uneasiness with the Coase Theorem" *Japan and the World Economy*, 7, 1-7.
- Scully, Gerald. 1989. *The Business of Major League Baseball*. Chicago: University of Chicago Press.
- Sloane, Peter. 1969. "The Labor Market in Professional Football", *British Journal of Industrial Relations*, 7, 2, pp. 181-199.
- Staudohar, Paul. 1996. *Playing for Dollars: Labor Relations and the Sports Business*. ILR Press, Cornell.
- Szymanski, Stefan. 2001, "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, Stefan, 2004a, "Professional team sports are only a game: The Walrasian fixed supply conjecture model, Contest-Nash equilibrium and the invariance principle" *Journal of Sports Economics*, 5, 2, 111-126
- Szymanski, Stefan, 2004b, "Tilting the Playing Field: Why a sports league planner would choose less, not more, competitive balance: The case of baseball", mimeo, Tanaka Business School, Imperial College London.
- Szymanski, Stefan and Stefan Késenne, 2004. "Competitive balance and gate revenue sharing in team sports" *Journal of Industrial Economics*, LII, 1, 165-177.

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

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

Szymanski, Stefan and Tommaso Valletti (2003) "Promotion and relegation in sporting contest", mimeo, Tanaka Business School, Imperial College London.

Szymanski, Stefan and Andrew Zimbalist, (2005) *National Pastime: How Americans play Baseball and the Rest of the World plays Soccer*. Brookings Institution Press.

Usher, Dan (1998) "The Coase Theorem is tautological, incoherent or wrong", *Economics Letters*, 61, 3-11.

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

Vrooman, John. 1996. "The baseball players market reconsidered" *Southern Economic Journal*, 62, 3, 339-60

Zimbalist, Andrew. 2003. *May the best team win: Baseball economics and public policy*. Washington: Brookings Institute.