

Seminar: Modern Law Review Seminar
The Regulation of Sport, May 11th 2006
CELI, Faculty of Law, University of Leicester

Remuneration of Sports Stars: Implications for Regulation

Richard Disney*

Abstract

The paper analyses remuneration in sports labour markets, ranging from semi-professional sportsman and women through to superstars. It describes how the competitive model of sports remuneration, which suggests that pay is largely determined by preferences, abilities, and the 'personal scale of operations' of sports stars, has implications for the allocative efficiency of the sports labour market in the presence of alternative regulatory structures and distributions of property rights, and for the incidence of 'punishments' levied on players, clubs or leagues. The implications of market power and collusion for pay and regulatory outcomes are explained. Finally, the paper examines the issue of 'competitive balance' and how the organisation of sports leagues and tournaments affects remuneration of sportsmen and women.

* Professor of Labour Economics University of Nottingham; Research Fellow Institute for Fiscal Studies; Member of NOHPRB; and club member (erstwhile season ticket holder) Chelsea FC.

Vignettes

“In 1929, Babe Ruth, the greatest American baseball player, earned \$70,000... An indignant reporter asked Ruth to justify how he could be worth substantially more than what the President of the United States earned. Ruth is reported to have said: ‘I had a better year’.”

(Quoted by Rosen and Sanderson, 2001)

“Do I get value for money? What do you think? I personally don’t get value for money from the players. If you see a Premiership player on, say, £5000 a week, do I get 20 times more pleasure than [from] a player in the Championship earnings £250 a week? Of course I don’t.

Paying players £100,000-£150,000 is just insane. I earned all my money the old-fashioned way and you would have to send me back to school for me to understand why they get paid so much for doing so little.

We are the ones stupid enough to pay them, but they are overpaid for what they contribute.”

Milan Mandaric, Portsmouth FC Chairman, quoted by the Daily Mail, 24-03-2006.

“Pat Burrell is in a profession where failure is routine. The best players in the game only get a hit one out of every three times at bat. But when Burrell inked a six year, \$50 million contract with the Philadelphia Phillies during the offseason, the lefthander’s room for failure decreased exponentially.

‘We play a game for a living, but this is a business,’ the Phillies leftfielder says. ‘There’s a reason for the big contracts.’

As the boos rained down on Donovan McNabb a week ago Sunday, heckling the same player they cheered during player introductions in Week 1, Burrell flinched. Burrell knows the \$115 contract McNabb signed last September not only is an endorsement of McNabb’s abilities and a security blanket for his future, it is an albatross.

‘They make all that stuff so public’, Burrell says, ‘Everyone knows what you are making so they decide if you’re worth it.’”

Dana Pennett O’Neil, Philadelphia Daily News, posted on 24-09-2003 on website: <http://www.psychologyofsports.com>

Remuneration of Sports Stars: Implications for Regulation

1. Background: Superstars, Tournaments and Professional Leagues

There has long been an interest among economists in the labour market for sportsmen and sportswomen. The early literature that emerged in the United States (US) stemmed from influential papers such as Rottenberg (1956) and Rosen (1981, 1986) sought to explain the high level of pay of sports 'stars', especially in baseball, and how this level was related to the organisation of sport. Interestingly, this now-extensive literature has developed models of remuneration (pay) that have been applied not only to professional sport leagues but also to explain remuneration in other labour markets which share the same characteristics, such as the pay of CEOs of large corporations and rewards to City traders. These underlying characteristics are, broadly, a limited pool of exceptional talent and a large 'personal scale of operations' (Mayer, 1960): that is where the volume of transactions affected by the star player's activities is very large (the volume of spectators, and the number of player-specific sporting items bought and sold – such as named shirts, equipment, products endorsements etc.) despite a low mark-up on each transaction. In such circumstances, in tournaments where there can only be one winner drawn from the ranks of only a few potential winners, remuneration levels for individual sports stars are potentially huge.

Whilst such an explanation for individual remuneration levels can easily be applied to sports that reward individual achievement such as golf and tennis (and to a lesser extent, athletics), many sports are essentially team sports in which performance is in part team-oriented and not simply a product of individual talents. Moreover, professional sport in the US is largely organised into 'closed' leagues with various explicit measures designed to ensure 'competitive balance' among teams (such as 'draft picks' inversely related to past team performance and salary bill caps). Here, arguably, the marketed product is not the performance of an individual sportsperson – or even an individual team – but is provided by the league as a whole. Individual remuneration can still be related to individual productivity because of US sport's love affair with statistics (yardage gained, completed passes, home runs, RBI, fielding errors etc.) and because salary contracts are generally made public. However, in looking at contractual negotiations between leagues and, on the one hand, their

employees and, on the other hand, outside bodies such as broadcasting companies, the argument has often been made that the activities of these professional sports leagues should be treated as ‘joint ventures’ (as in Flynn and Gilbert, 2001) and that the product should be treated as the provision of a competitive tournament rather than success in a tournament *per se*.

The view that the ‘product’ is the competition itself rather than the outcome, and that this outcome can be enhanced by explicit collusion, has not found favour with all regulators and nor indeed with all economists (e.g. Noll, 1974, 2002, 2003). It seems at first sight paradoxical that competition must be internally regulated by professional leagues in order to generate competition (in this context, a ‘competitive balance’ between teams) and that this may permit various non-competitive practises such as restrictions on new entrants, joint bargaining with purchasers of the product and various restrictions on duration, remuneration and other terms of employment contracts.

Traditionally, European sporting governing bodies are looser umbrella organisations that cover a hierarchy of leagues with promotion and relegation, such as the Football Association in England. Even before the Bosman ruling and other similar rulings, it has been harder for the organisers of such leagues (and of internal groups of clubs such as the Premiership or the G14 group) to form ‘closed leagues’ along US lines. Such an outcome – for example a European ‘superleague’ in football, might occur in the future (Hoehn and Szymanski, 1999) if existing ‘competitive balance’ in domestic leagues is eroded by the concentration of talent and success among a small sub-set of clubs, but long run trends are notoriously hard to predict.

These are among the key issues that underpin the economic analysis of remuneration in sports. The remainder of this essay is structured as follows. In the next section, a basic structure of pay levels in sports, which I term the ‘remuneration pyramid’, is outlined. Pay levels are determined by preferences and productivity, but with a sub-set of ‘superstars’ who obtain incomes well beyond those suggested merely by considerations of individual or team production. Section 3 considers the point that an essentially competitive model of pay may be inapplicable in an economic activity characterised by market imperfections: both in selling of the sport itself and in the purchase and payment of players (monopsony). It considers the implications of this for total sports revenue, and the implications of the Coase theorem for both total

revenues and the allocation of revenues among participants (players, owners, agents etc). It also examines the implications of our remuneration model for ‘crime and punishment’ – mostly notably the incidence of financial and non-financial penalties levied on clubs and players. Section 4 considers the issue of ‘competitive balance’, how to achieve it, and the implications of competitive balance for the remuneration of professional sportsmen and sportswomen. A brief concluding section follows.

2. Sports remuneration

Most people are involved in sport because they enjoy it. Like musicians, sportsmen and sportswomen do not set off on a sporting career with the idea of making money, with the exception of a few well-publicised parents of tennis players. Foregone earnings early in life, and uncertainty both over talent and the prospect of injury or loss of form bringing an early end to a sporting career, limit pecuniary objectives at the outset. Given the uncertainties of a sporting career, however, it is also unsurprising that sportsmen and women are anxious to maximise the returns from whatever success luck, or their talent, bring them.

Pay to sportsmen and women can be thought of as defined by a ‘remuneration pyramid’ of the stylised type depicted in Figure 1. Much of the interest of economists in sports remuneration concerns the apex of the pyramid, but at the base, large numbers of people engage in sport with no remuneration at all. In part this stems from a lack of any exceptional talent, but also from the fact that their activity generates no revenue from which any type of remuneration can be exacted.

At the next level of the remuneration pyramid are the sports, such as the lower leagues of football and, until recently, rugby union, where there is some talent and where there is some revenue to sporting events, largely derived from gate revenue and limited sponsorship. Here there is the potential to pay some or all participants part-time, or expenses. Historically perhaps cricket is one sport where paid and unpaid players might both participate in the same team – a tradition kept alive by having one or more ‘professionals’ attached to an otherwise amateur (i.e unpaid) team. As outside revenue expands, notably with the sale of television rights and associated revenues, ‘expenses’ may effectively cover full-time payments so that ultimately the sport is transformed into a professional structure.

2.1. *The paid professional*

Further up the diminishing hierarchy are sports or leagues with full time professional players. At the lowest rung of the professional level, participants are effectively paid their opportunity cost – that is pay is equal to outside earnings plus foregone earnings in training and any premium arising from the greater career uncertainty associated with professional sports. Teams in leagues such as the lower professional leagues of football (e.g. Divisions 1 and 2 in England) pay a wage that, while often exceeding outside earnings, does not represent a large rent to ability. Players are potentially substitutable by others (albeit not without the club incurring training costs or transfer fees) and revenue sources depend on, and are dominated by, the loyalty of the fan base rather than outside sources such as revenue from broadcasting rights.

Still further up the pay pyramid are sportsmen and women who are demonstrably more productive than any potential replacement, and where team revenue sources can be augmented by outside revenues such as sponsorship, merchandising and broadcasting rights. These additional sources of revenue have grown recently in many sports – notably broadcasting revenues.¹ Such players earn a ‘rent to ability’ and numerous studies have attempted to link individual remuneration to measures of this ‘rent to ability’ derived from indicators of productivity such as win percentages over the season and other measures of on-field performance. An early and influential study of baseball by Scully (1974) calculated these measures of success as ‘outputs’ in a production function conditioned on inputs such as batting and fielding averages, runs scored and defended, and so on. Such methods have been applied to other similar sports such as cricket (Schofield, 1988; Bairam, Howells and Turner, 1990). Interestingly, Scully showed that the application of the ‘reserve clause’ in baseball, which allowed professional teams to make exclusive long term contracts (see later), discounted individual salaries well below marginal products. Abolition of this clause raised baseball salaries significantly (MacDonald and Reynolds, 1994; Rosen and Sanderson, 2001) as have, unsurprisingly, the removal of

¹ Commercial receipts and gate receipts in the English Premier League between 1992 and 2003 both rose by around 300% in nominal terms and account for roughly equal sources of revenue - about 30% each. Revenues from broadcasting rights rose by over 3000% in the same period and accounted for over 40% of total revenues by that last year (cited by Michie and Oughton, 2004).

salary caps such as the abolition of the maximum wage in professional football in Britain.

Whilst pay may be broadly competitively determined by productivity in such settings, however, the simple paradigm of the atomistic competitive market does not completely hold. Team success is a joint, rather than an individual output. A revenue maximising owner can hire a manager and set of players that optimises expected team production given the raw materials and budget constraint, and both individual and joint remuneration depends on team performance (Rosen and Sanderson, 2001). Greater success allows the owner of the team to increase overall remuneration and perhaps to supplement the existing squad by new and better players. However, although pay may be differentiated among members of the squad, individually differentiated remuneration based on joint production has its limits: higher pay to the better players in the squad may induce claims for comparability amongst other squad players, and risk averse players might prefer longer and more stable contracts to pay systems which 'reward' exceptional performance in particular matches. Incentives contracts (such as 'win bonuses') are therefore generally applied at a team level rather than an individual level – players receive extra remuneration for winning an important cup tie, but a bonus is not usually awarded to the 'man of the match'.

With some pooling of rewards and incomplete matching of individual pay to productivity, the standard competitive model of individual remuneration predicts that a worker paid less than his or her marginal product is likely to quit (or ask to be released from contract). In turn, a club should be prepared to release or sell a player if it believes that that player's productivity would be enhanced by moving to another club since the present value of retaining the player is less than the future earnings and transfer fee, if any, that could be obtained by trading the player. This is an implication of the 'Coase theorem' discussed shortly. However the difficulties of disentangling individual productivity in a team setting, the uncertainty of measuring future productivity flows in alternative teams, and any residual loyalty to fans or to the club (or in the opposite case, disenchantment) may induce frictions into the labour market. Clubs retain players even when they appear to have no use for them (this is one sense in which the facts of the Bosman case, as opposed to the outcome, might puzzle the economist). Players may move club (or ask for a transfer) even when their club is

doing well, or not leave it even though they might do better elsewhere (less commonly).

There are two possible forces at work here. One is uncertainty – player valuations, whether pay or transfer fees, are not fully certain in alternative settings and may only be revealed after contracts have been negotiated or renegotiated. Indeed it is sometimes argued as a case for maintaining some control of clubs over contracts – beyond the standard argument that clubs should recouping training costs – that owners bear the costs of the initial uncertainty as to player performance, which subsequently becomes transparent. Transfer fees permit clubs to extract some of the rent derived from efficient reallocation of players based on revealed productivity. This is especially true for young players, for whom ultimate performance is less certain.

Second, rather like homeowners who seek to move during house price downturns but refuse to lower their reservation price, owners and indeed the players themselves may be reluctant to believe that past expenditures on players are essentially ‘sunk costs’ and should be ignored in current valuations. Clubs and players’ representatives with under-performing players will typically argue that current productivity, or lack of it, arises from other aspects of the team rather than their own player’s worth. The fact that there may indeed be some truth to this assertion further complicates the issue.

2.2. Superstars

Towards the apex of the pyramid are sportsmen and women who have special and demonstrably irreplaceable talents, and where there is a sufficient revenue base to permit remuneration to far exceed outside earnings. Standard empirical models of remuneration and sporting productivity generally find a very small sub-set of professional whose remuneration far exceeds any measurable performance-related measure of productivity. The distinction between simple rents to ability and sports ‘superstars’ lies in the ability of the latter to construct a unique brand affiliation in addition to an observable contribution to team productivity. This affiliation, based on both perceived sporting achievement and high visibility, allows that sportsman or woman to benefit from the ‘personal scale of operations’ described earlier. In individual sports such as golf and tennis where there is huge potential for consumer

spending on sport-related products derived from brand identification with certain individuals (Tiger Woods, Andre Agassi), and where there are tournaments with explicit individual ‘winners’, the potential for superstar status is obvious.

In team sports, it is a little harder to identify which individuals will obtain superstar status. It is well documented in cricket, for example, that there are a few individuals who will induce extra crowds to attend sessions of cricket matches or to switch on their televisions – the arrival at the crease of a Don Bradman, Lara, Tendulkar, Viv Richards or, these days, of Freddie Flintoff being illustrations (Blackham and Chapman, 2004). But, as these cases illustrate, there are also indefinable aspects to ‘superstar’ status arising from character, media presence (either ‘genuine’ or carefully manipulated) and other facets that cannot so easily be measured.

3. Extensions of the pay & remuneration model

3.1. Market imperfections and the competitive model

Underlying much of the preceding discussion was a competitive paradigm of the determination of pay of sportsmen and sportswomen. The focus was on individual preferences, differences in measured productivity and the capacity of top stars to exploit their ‘personal scale of operations’. True, there are problems of measuring individual productivity in a team setting, and complications to the ‘story’ induced by various uncertainties, but the basic message was that participants’ pay was related in some way to their marginal revenue product.

However, towards the apex of the remuneration pyramid, sport is not organised as a competitive market. Professional sport is organised in a leagues, which rarely compete directly with each other. Leagues may be ‘closed’ (that is, a limited set of franchises compete in the league and the league decides whether to permit additional franchises to enter the sport, as in the United States), or ‘open’, in the sense that promotion and relegation is permitted from a hierarchy of leagues down to semi-professional lower leagues, as in Europe. Such leagues have joint market power as national purveyors of the sport, and joint monopsonistic power as sole national buyers of talent in that sport. Nevertheless despite the sports market not being structured competitively, leagues have an interest in maintaining ‘competitive balance’ among participating clubs as a way of maximising revenue to the clubs within the league.

Which of these structures, the open or closed structure, enhances ‘competitive balance’, and how this affects pay, is an issue to which I return shortly. For the moment, however, it is sufficient to note that professional leagues do not compete with one another except indirectly (e.g. in football the Premier League and the European Champions League), and that players who aspire to reach the highest echelons in their sport will typically have to enter a single league that is regulated in various ways by its governing body and that may include limits on the structure of remuneration and contracts.

The first limit on the competitive model therefore is the recognition in law that, in certain circumstances and on specific issues, collusive agreements among franchises or clubs within a league are not necessarily anti-competitive, since the ‘product’ that is being provided is a joint product (or ‘joint venture’ – see Flynn and Gilbert, 2001): namely a competitive sports league. Examples of such collusion could include decisions about who has first choice of newly trained players – as in the National Football League (NFL) where the worst performing teams have first pick of the college draft. Another issue of joint production is sharing of gate revenue, which is pervasive in US sport but less so in Europe (with exceptions such as the FA Cup) – see Hoehn and Szymanski (1999) – and the sharing of broadcasting revenues, which is much more pervasive globally since negotiations are often carried on at a league rather than a club level (see, *inter alia*, Cave and Crandall, 2001).

Other regulations include caps on salary bills, either as an absolute amount or as a proportion of turnover. In the NFL, for example, each team had a basic salary bill cap of \$85 million in 2005. In NBA (basketball) the figure was \$47 million with some exceptions. In baseball, teams with a high salary bill pay a ‘luxury tax’ that is redistributed to other teams. In Britain, rugby league teams had a salary bill cap of £1.7 million in 2005 and rugby union a cap of £2 million. Caps on salary bills have been suggested in order to improve ‘competitive balance’ in professional football in the UK, and also caps on the wage bill as a fraction of revenue as a means of keeping football clubs solvent.²

² Dave Whelan, Chairman of Wigan FC, has argued for salary bill caps in the English Premier League of £25-30 million to restore competitive balance as otherwise “...it’s Chelsea on their own”. See the interview on http://news.bbc.co.uk/sport1/hi/football/eng_prem on 22-09-2005. Michie and Oughton (2004) argue that revenues among Premier League clubs (and therefore, presumably, salary bills) have

For reasons that have been extensively discussed, regulations, whether designed to enhance ‘competitive balance’ or simply reflecting an older form of labour contract, have also permitted individual clubs within the league to place various restrictions on pay and contractual conditions of individual employees. As is well known, professional football in Britain enforced a maximum wage of £20 per week, which was not abolished until 1961. More pertinent has been the system of player registration that has operated in several sports and which allows clubs to determine whether or not a player can move between clubs irrespective of whether the player has an agreed contract with that club.

The principle of whether an out-of-contract player was free to move clubs without the consent of the club with which the player was registered was of course fundamental to the Bosman case. The ‘Reserve Clause’ in baseball which effectively ended in 1975-76 had a similar effect insofar as contracts with baseball players stipulated that at the end of a (typically annual) contract, the players could not simply negotiate with another team (‘free agency’) but had either to negotiate a new contract with the same team or ask to be released, which could be refused. The Reserve Clause was specifically designed to halt a perceived threat of salaries escalating out of control of the owners. The 2004-05 lockout in the National Hockey League in North America also in part stemmed from a disagreement between owners and players’ negotiators about a contractual clause that limited ‘free agency’ until the player was aged 31 or over.

Overall, these deviations from a competitive market structure would be expected to have an implication of the allocation of resources in sport (that is, the pricing of inputs and outputs relative to marginal cost) and the distribution of resources among participants (notably, as to who receives the rents from the departure of prices of inputs and outputs from marginal cost). In particular, the particular structure of the market in a sport would be expected to affect pay; indeed some measures such as limits on ‘free agency’ were explicitly designed to limit ‘salary escalation’. Enhancing ‘competitive balance’ by other measures such as salary bill caps might or might not increase revenues and remuneration.

become more unequal over time and suggest that there is a correlation between team revenues and league success (although of course the causation can run both ways).

At first sight, *any* restriction on levels of remuneration might be expected to affect pay levels adversely, especially of those at the top of the remuneration pyramid. A more subtle argument, described shortly, is that measures to improve ‘competitive balance’ may also affect player incentives. On the other side of the argument, however, advocates of regulation would make two points: first that greater competition might enhance revenues and thereby indirectly affect the size of the ‘cake’ that can be shared; secondly that there is a risk of an ‘arms race’ in remuneration levels in a ‘winner takes all’ setting which ultimately reduces the financial viability of the sport in question.

Most fundamentally, the elements of collusion, oligopoly and monopsony in the economics of sport economics can all be linked to the less attractive features of remuneration bargaining and resource allocation in sport: the ‘backroom deals’ and settlements of remuneration issues outside the public gaze, including cash transfers at motorway service stations (Bower, 2003). In European football, some club chairman like to point to the activities of sports agents, to the explosion in pay expectations and greed, and to the threat of financial ruin as a rationale for tightly regulating players’ contracts and remuneration levels despite the rapid rise in revenues in recent years. All the same arguments were made in baseball and US sports in the 1920s, and in British football in that same period, to justify forms of contractual servitude that are correctly deemed no longer to be acceptable. Moreover, many other industries that have seen large increases in revenues in short time periods in recent years due to deregulation, privatisation and technological changes have also been characterised by rapid increases in remuneration levels at the apex of the pyramid whilst leaving pay at the base of the period largely untouched. Indeed, it might be argued that those with experience of large revenue windfalls arising from mass privatisations taking place under arbitrary and dictatorial regimes are best qualified to run football clubs, which now inhabit a similar environment.

3.2. The implications of the Coase theorem, free agency and Bosman

There is one element to economic theory that has implications for resource allocation even where markets have high levels of imperfection, such as monopsony in the labour market. It implies that rational or profit maximisers may provide the ‘correct’ resource allocation decisions independently of who *owns* the rights (for

example, labour contracts) in the market. This is a version of the Coase theorem applied to sports economics (see Rottenberg, 1956; El-Hodiri and Quirk, 1971; Rosen and Sanderson, 2001; Schmidt and Berri, 2003). It is a necessary antidote to the commonly-held view that the 'laws of economics' do not apply to sport given the amount of collusion and market power held by certain participants.

Coase's theorem relates to the allocation of (and more pertinently, the presence or absence of) property rights. It is commonly applied to environmental issues – for example the allocation of rights to pollute or not to pollute. In the standard example, if an industrial firm pollutes a river downstream, the marginal cost of clearing up the pollution can be equated to the marginal social cost of the pollution itself so long as property rights exist. If the polluter has the right to pollute, the unfortunate downstream owner of the river pays. If the downstream owner has the right to unpolluted water, the firm pays. The optimum resource allocation should be unaffected by the distribution of property rights, although of course the *ex post* distribution of income will be affected by the allocation of property rights. It is the *absence* of property rights in the context of pollution that causes the environmental problem.

As hinted in the previous section, this theorem can be applied to the question of 'free agency' in sport. With free agency, the individual sportsman or sportswoman can choose to play for whatever team maximises his or her productivity and remuneration without hindrance. Without free agency, a revenue-maximising owner of the contract should either offer a contract to the player and use the player, or sell the player to the another club that can maximise the player's potential. Failing to release the player and not playing the player is not rational in a revenue-maximising sense.

Bosman (and moves to free agency in other sports) therefore have implications for the structure of remuneration of players relative to the profits of owners (redistribution), especially in the short run, but not necessarily for the allocation of players or for remuneration in the long run. In particular, application of an efficient bargaining model (which really requires a Coase-type theorem to hold) suggests that the post-Bosman framework will lead to a lengthening of contracts and a remuneration package at the existing club that will lessen the pay-off to a potential new club. Out-of-contract players will obviously benefit from free agency, and

players offered new contracts will have also benefited immediately since the longer and, most likely, more generous contract offer from the incumbent club will produce increased benefits over the duration of the contract. However these new terms essentially raise the threshold for any new club bidding for that player and, by reducing the pay-off to the outside offer, ultimately reduce the renegotiation pay-off to the player (Feess and Muehleusser, 2003). For short time horizons (career spans), it is not obvious that this intertemporal trade-off has much impact, so Bosman over the relevant time horizon benefits players. For the longer term, the results are less clear although most economic analyses seem to agree unequivocally that the incentives for clubs to engage in training young players, where the uncertainty of ultimate ability mitigates against any attempts to increase contract durations, is considerably reduced by Bosman.

3.3. Crime and punishment

The organisation of sport into professional associations and leagues inevitably raises the question of how its regulatory structures are able to discipline participants who, in some way or not, fail to obey the rules of that association. There are of course large issues here, such as whether there is a public interest in how professional bodies regulate themselves and if so, whether external regulation should supersede self-regulation. As noted earlier, the market structure, organisation and remuneration of professional sports have much in common with other markets characterised by high pay-offs to ‘winners’ and in which many of the transactions (e.g. large scale financial sales and purchases such as transfers of players, the intervention of third parties such as agents, takeovers etc.) take place ‘behind the scenes’, such as corporate finance and equity markets. In general, however, professional sports, at least in Europe, have managed to avoid such interventions, relying on their own investigations into malpractice such as ‘bungs’ in football (Bower, 2003).

There are broadly three forms of regulatory intervention within professional sports’ regulatory bodies that lead to the enforcement of penalties or disciplinary action: the requirement of a solvency condition as a prerequisite for a team’s participation in the league, issues surrounding registration and transfers of registrations of players, and issues concerning conduct of players such as public behaviour on and off the field. Inevitably there are several tensions in this essentially

self-regulated environment – between participants in the league and the regulatory body of the league as to whether alleged misdemeanours warrant punishment and whether such punishments are applied consistently, but also within the league as a whole as to whether ‘self-regulation’ is publicly seen to be carried out effectively (most notably on issues of ‘sleaze’) as a way of forestalling outside regulation. Again, sport is not unique in facing these dilemmas and in how it attempts to resolve them.

Solvency tests are generally applied to clubs participating in leagues as a means of ensuring the fulfilment of contracts, fixtures etc. Such conditions are sometimes underpinned by explicitly redistributive procedures operated at the league level (such as ‘parachute payments’, revenue-sharing etc.), which may be designed to protect clubs from adverse shocks to finances, such as relegation. It is often argued that such solvency tests are applied loosely and unequally – that is, more loosely to higher profile participants. The essential difference between sports and other markets in this context is that brand identification and brand loyalty is much stronger in the affiliation to clubs than to products in other markets. Consequently clubs have some market power in deterring competitors or creditors from driving the club to exit (bankruptcy) and it is unusual to see high profile professional clubs driven out of professional leagues. This leeway, plus the tournament nature of sport – have implications for the financial strategy of league participants. Whilst not justifying some of the more bizarre financial behaviour of football clubs in particular (such as debt overhangs on borrowing to finance the purchase of players that have long since been sold), nor contesting the view that normal solvency conditions should apply, the combination of a fall-back bargaining position and high pay-offs to success explains why professional sports clubs adopt more risky financial strategies than firms operating in other markets.

The issue of regulating registration and changes in registrations of players is at the heart of the organisation of professional sporting leagues, for the reasons described in the previous sub-sections. Since these rules are central to economic aspects of sport such as ‘competitive balance’ and thereby affect player remuneration indirectly, it is hard to justify the argument that they lie outside the sphere of economic activities, although professional associations sometimes seek to make that argument. Because of the history of protecting clubs through player registration

requirements, it is natural that leagues attempt to exact strong retribution on clubs that overtly challenge these rules. Relegation or points deduction are commonly applied to clubs that field non-registered players, and associations often fine or penalise clubs that transgress in, for example, talking to registered and contracted players with a view to encouraging them to switch clubs, sometimes known as ‘tapping up’.

Several points can be noted about punishments for such activities. First, it is obvious to outsiders that such meetings go on all the time – the threat or application of fines seems to be in proportion to the flagrancy of the breach of the rules and the importance (or self-importance) of the offended party (the incumbent club) in the league hierarchy. Second, it might be argued that the league, by enforcing such a rule, is exploiting its dominant market position since clubs ultimately have little choice but to comply with such rules where the league has a monopoly position as the provider of that professional sport in that country. Third, given the steady erosion of this type of restrictions with the shift to free agency, it is arguable that long term contracts which offer no scope for renegotiation will (and are) themselves ultimately under threat. For example, unlike contract in other labour markets, the absence of a clause in many sports contracts allowing a participant to give notice that he or she intends to terminate the contract prematurely in return for appropriate compensation to the other party, or even talk to a potential outside employer, seems strange from an economic point of view, since again the application of the Coase theorem would suggest that the incumbent club would be better-off selling the player to the suitor at a fee commensurate with the future expected revenue product of that player during the remainder of the contract. In the event of a dispute as to what that sum of compensation should be, binding arbitration (especially pendulum arbitration) would surely resolve the problem.

Finally, there is the range of punishments associated with other forms of conduct that the professional association deems improper – bad sportsmanship, criticising other participants, including opposing teams, referees and umpires and the league associations themselves. At face value, such punishments (which are usually short suspensions and/or financial penalties) seem to operate on a variety of principles, often simultaneously: the gravity of the offence, the extent of visibility of the offence, whether the offender has ‘form’, the ability to pay of the club or individual and the threat of legal action (whether retaliatory, or to forestall outside

intervention e.g. charges of assault or defamation). It is often argued, for example, that professional sportsmen and sportswomen at the apex of the remuneration pyramid have such large incomes that anything but an enormous fine or penalty will have no impact on behaviour – if this character-reforming perspective is indeed the appropriate interpretation of the intent behind punishing the individual.

From an economic point of view, the appropriate model here is the model of *tax incidence* – that is whether, the individual on which the tax is levied effectively ultimately bears the tax – in this case a fine (see Atkinson and Stiglitz, 1980). To illustrate the issue of tax incidence: VAT is levied on the seller of a good, but few believe that anyone other than the consumer bears the tax. Social security contributions are notionally levied jointly on the employer and the employee, and recovered from the employer, but again the notional and effective incidence of each part of the tax are likely to be either on the employee or (if either the firm or worker has some market power) on the consumer.

In the standard analysis of labour markets, where labour supply is not completely elastic because workers have ability to extract a rent over and above the opportunity cost, part of the tax (fine) should be borne by the person on whom the fine is levied. But again, this takes no account of the inelasticity of the demand for the product (loyalty to the club) and the personal scale of operations, which may allow the person on whom the fine is levied to renegotiate the incidence of the firm on to consumers, such as season ticket holders. To a large extent, the ultimate incidence of these fines in the long run depends on a perception by the most loyal supporters and purchasers of the club's products as to who is the wronged party in the particular setting – a perception that both the professional association and the person concerned will attempt to sway by sympathetic media coverage.

4. Competitive balance

As suggested at the start of this paper, one of the key issues in the economics of sport is the issue of 'competitive balance'. Revenue to professional leagues and thereby levels of remuneration, it is argued, depend in large part on the excitement of the competition generated by teams that are relatively even-matched. Since the league structure is providing that competition, it is commonly suggested that the league should be treated as a single entity in law rather than each individual club, with

implications for the applicability of anti-trust or anti-competitive laws concerning collusive behaviour designed to enhance 'competitive balance' (this position is stated by Flynn and Gilbert, 2001).

As mentioned previously, there are two broad 'models' of the operation of a professional league. The first is the North American variant of a closed league (new entrants or 'franchises' can only be jointly approved in the event of an agreed expansion or the end of an existing franchise) with a self-contained league management structure which negotiates joint broadcasting rights and sometimes shares gate revenue. There are often 'artificial' measures to promote greater competition or to restrict the financial and performance discrepancies between the previous best and worst teams, such as a 'draft' system to feed the best new players into the worst performing teams in the previous session and caps on salary bills (and, in the past, salaries) that may bind on the more successful teams. Since franchises are not infrequently shifted between geographical locations, this structure puts a premium on the national attractiveness of the product as a competitive sport (in terms of television revenues and national merchandising), with less reliance on local loyalty as a driving force (e.g. gate receipts and local merchandising).

The second structure of professional sports organisation is where there is an open hierarchy of leagues with competition induced not just by the possibility of winning but also from promotion/relegation between leagues. Typically, league organisation is through a looser umbrella federation. There are typically no salary bill caps or limits on which teams can recruit which players (although there are restrictions on salary bills in some sports, and both restrictions of salaries and on free agency previously existed). Broadcasting rights are often negotiated jointly but sharing of gate revenues is less pervasive than in North America. In this setting, local loyalties are important and it is unusual, and usually disastrous, for clubs to move geographically. Whilst outside sources of revenue, notably from television rights, have increased dramatically, local affiliation and loyalties are still an important factor in determining club policy. Inevitably, too, governing bodies wield less effective power in such circumstances.

Within the traditional league federation model, there have of course been attempts to break-up existing structures and to introduce competitions of top players along more North American-oriented lines. The 'Packer revolution' in cricket was

ultimately incorporated into the existing structure of cricket organisation but had major impacts on remuneration levels, the structure of domestic cricket, and so on (for example, the introduction of ‘central contracts’ into English cricket). The professionalisation of rugby union led some clubs to argue for a closed rather than an open structure to the professional league with no promotion/relegation although this was ultimately defeated. The ‘G-14’ group in European football (actually 18) and threats to introduce a ‘super league’, perhaps developing out of the European Champions League, may yet induce changes to the structure of European football (Hoehn and Szymanski, 1999).

There are several issues that stem from this typology of league structures. Where can the line be drawn as to when interventions by sports’ governing bodies are promoting ‘competitive balance’ or are simply flagrantly anti-competitive? Does greater ‘competitive balance’ increase revenue and does that get translated into higher remuneration of professional sportsmen and sportswomen? What are the incentive effects of different league structures and how do these affect remuneration?

These issues concerning competitive balance have been considered at length in the literature on the economics of sports. To simplify and condense this discussion, consider the reaction of individual spectators to sporting outcomes in a team sport such as football (or indeed US pro-football). Some spectators may be genuinely impartial over outcomes and would be happy to see a high-scoring game without caring which side wins. At the other extreme, fanatical loyalists would like to see their team trounce every other team by a large score, irrespective of whether the other team make a fight of it or simply ‘throw in the towel’. Ideally, perhaps the ‘average’ spectator would like to see a close high-scoring contest ultimately decided in favour of the team to which that supporter has a perceived affiliation (perhaps bar any particular fierce local rivalries where total humiliation is the perceived objective).

Since contests take place within the league against known rivals with perceived strengths and weaknesses, decisions such as investments in new players have strategic implications. In the ‘competitive outcome’ scenario, joint investments in better players by a number of teams increase competition, excitement, revenue and potential player remuneration – we can term these investments as ‘strategic complements’ (Hoehn and Szymanski, 1999, using the terminology and model structure of Bulow, Genakopolos and Klemperer, 1985). In the ‘fanatical loyalist’

scenario, teams will maximise their revenue by investments at the expense of other teams and this scenario sees joint investments as ‘strategic substitutes’. Clearly different weightings of ‘competitive balance’ and ‘win success’ in the revenue pay-offs to teams and leagues have implications for league structure – for example, over the value of interventions explicitly intended to maintain ‘competitive balance’ such as the ‘draft’ and caps on salary bills.³

Do ‘open’ or ‘closed’ leagues increase ‘competitive balance’? At first sight, the analysis would suggest that competition is enhanced by explicit measures to equalise teams, such as a ‘draft’, salary bill caps, and revenue sharing such as occur in ‘closed’ leagues. In contrast, in ‘open’ leagues, the richest clubs might be expected to perform better, year-on-year, so that competition throughout the whole league is reduced. Michie and Oughton (2004) argue that, during the years in which the English Premiership has existed, and in which broadcasting revenues have accelerated rapidly, the share of tournaments won, and revenues, have tended to become increasingly concentrated among the ‘top five’ teams – the implication being that, had there been restrictions such as a salary bill cap in existence, a greater degree of competition would have been experienced.⁴ However, there have been periods in the past, such as the early 1980s, when one team dominated the then championship in England even though broadcasting and merchandising revenues were much less significant.⁵ It is equally true that measures to equalise teams in ‘closed leagues’ do not guarantee an equal likelihood of wins across clubs.⁶ Repeated winning of championships in successive seasons is relatively unusual, whether in an open or closed league.

³ Hoehn and Szymanski (1999) argue that entry into multiple leagues with different potential revenue streams and organisational rules (such as domestic football leagues and the European Champions League) may induce clubs to switch strategies. For example, they argue that growing revenue streams attached to Europe-wide competitions in football may ultimately induce clubs to switch from domestic competition into a closed European-wide Super League. It is far from clear that this will happen; a better illustration of the process outlined by the authors may however be in cricket, where international contests between the top teams increasingly dominate revenue receipts at the expense of domestic cricket leagues, and where consequently top cricketers may essentially be exempted from playing in domestic competitions through ‘central contracts’.

⁴ For a somewhat different ‘take’ on the data and an alternative test, see Szymanski (2001).

⁵ Between 1975-76 and 1989-90 (15 seasons), Liverpool FC won the English Championship ten times. Between 1992-93 and 2005-06 (14 seasons), Manchester United FC won the English Premiership eight times.

⁶ For example, in the NFL, three teams (the Raiders, the Redskins and the 49ers) won 8 of the Superbowls contested in the 1980s, two teams won 5 contested in the 1990s (the Cowboys and Broncos, with 2 Superbowls also won by the Redskins and 49ers), and the Patriots have won 3 of the 5 Superbowls since 2001.

Other analyses also question whether ‘competitive balance’ is enhanced by a closed internally regulated structure to leagues. Noll (2002) argues that attendances, and potentially therefore revenue, investment and remuneration, are higher in leagues that allow promotion and relegation; however promotion and relegation are likely to be associated with ‘strategic substitutability’ where newly promoted teams may be ‘out of their depth’ and have no incentive to make the investments to maintain their position in the league. Forrest and Simmons (2002) are among a number of authors that point out that as ‘win uncertainty’ increases in terms of equality of team abilities as a result of ‘competitive balance’, home advantage necessarily becomes an increasingly important factor in determining outcomes so that overall results may become more predictable, not less.

The implication of these various findings seems straightforward. Whilst there are basic arguments that suggest that more closely-regulated leagues designed to give greater competitive balance *should* produce more evenly-matched teams and greater competition, the empirical evidence is far from conclusive. But the rationale for many apparently anti-competitive measures associated with greater internal regulation of leagues is precisely that they foster increased competition. If the evidence suggests that ‘competitive balance’ is not particularly enhanced by control over players’ contracts, levels of remuneration (both in total and of individual players) and other league interventions, relative to the more ‘free for all’ approach of open leagues, outside bodies such as competition regulators and anti-trust legislators may revise their opinions of what constitutes anti-competitive conduct in professional sports.

4.1. Impact of competitive balance on sports remuneration

The economic analysis of measures that are sometimes proposed to increase competitive balance gives mixed results. At first sight, caps on salaries or on salary bills would restrict the distribution of remuneration among sportsmen or sportswomen – possibly both the inequality and the average level of salaries. However, if increased ‘competitive balance’ raises revenues through greater attendance or broadcasting rights, the total ‘pot’ of money available may increase so that the aggregate salary bill is higher than would be the case were there to be no restrictions on salary levels (Késenne, 2000). On the other hand, inequality will not lessen if, within the overall ceiling, owners choose teams of mixed superstars and mediocre players rather than

teams composed of players of roughly similar ability (a choice faced by any participant in a virtual ‘Fantasy Football’ league). This analysis is of course weakened insofar as ‘win success’ rather than competitive balance is the prime motivation behind league revenues, and in which a competition among rather similar and average teams generates lower revenues than the chance to see (and/or support) a sub-set of teams of superstars.

The economic analysis of sharing of gate receipts also suggests mixed conclusions. If teams have the same ‘technology’ of converting talent into results, and have a similar support base (revenue generating function), then gate sharing will have no impact on competitive balance (El-Hodiri and Quirk, 1971). This result essentially arises because teams jointly maximise profits and allocate talent accordingly (Szymanski and Késenne, 2004). However revenue sharing will reduce the incentive to win for any given team and therefore reduce the incentive to invest in talent so that the average quality of players, and the average level of revenue-maximising remuneration, will fall. Szymanski and Késenne also argue that gate-sharing may reduce competitive balance if clubs have unequal ‘drawing power’ or different intensities of ‘win support’. If revenue to any given team depends in large part on its own success, any form of revenue sharing that dilutes the quality of both teams and reduces the win probability of the more strongly supported team will have an adverse impact on gate receipts, which will have a disproportionately adverse effect on the weaker team. Indirectly, the same argument is implied by Szymanski (2001) in his evidence that football cup competitions (which do have gate sharing and the possibility of ‘upsets’ of strong teams) have lower gate revenues than league competitions, where such ‘redistributive’ rules are not in place.

The desire to enhance ‘competitive balance’ may therefore seem attractive but it does not necessarily maximise total revenues and therefore remuneration. Leagues composed on teams of roughly equal talent may produce contests of lower quality, with consequently lower joint revenues and remuneration of participants. In contrast, having some teams with a large revenue base and strong and passionate support may be of advantage to the league as a whole, although such teams may be disproportionately successful. However these types of results from economic analysis rely on assumptions concerning joint revenue maximisation in a league setting – it is hardly surprising that individual teams within a league that struggle to win trophies

are more vehement in their support of measures that are intended to reduce the inequality of success.

5. Conclusion

This paper has considered remuneration in sports labour markets, ranging from semi-professional sportsman and women through to superstars. It describes how the competitive model of sports remuneration, which suggests that pay is largely determined by preferences, abilities, and the ‘personal scale of operations’ of sports stars, has implications for the allocative efficiency of the sports labour market in the presence of alternative regulatory structures and distributions of property rights, and for the incidence of ‘punishments’ levied on players, clubs or leagues.

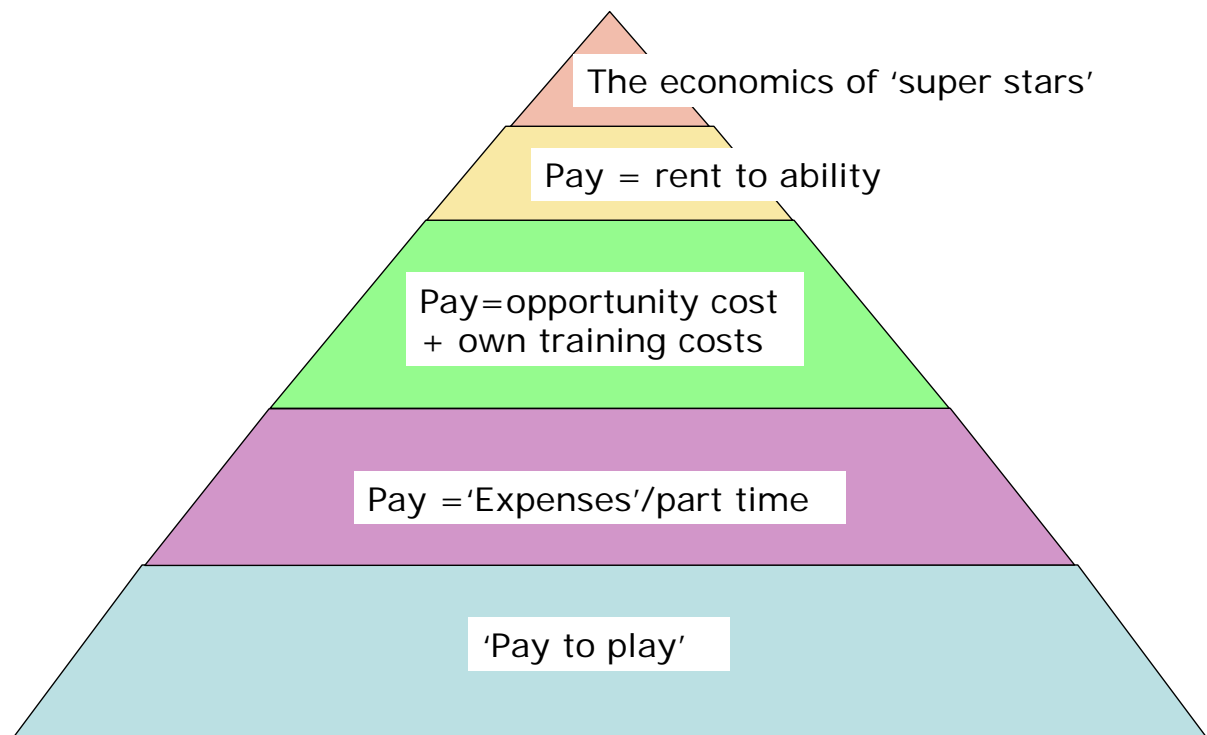
The economic analysis of remuneration in sports markets has a number of implications. First, departures from a competitive framework by sports clubs and leagues, especially in the treatment of labour contracts and remuneration have to be very carefully argued – many justifications seem rather spurious or not borne by empirical analysis of outcomes, especially surrounding the need for measures to enhance ‘competitive balance’ by limiting salaries, by contractual restrictions or by revenue-sharing. That many of these restrictions have been eroded and outlawed over time (notably in the general shift towards free agency) has not eroded ‘competitive balance’ despite dire warnings to the contrary from within the sport.

Second, there is the issue of joint bargaining and the role of league organisations as co-ordination mechanisms. It is perhaps oversimplistic to say that, because a sport requires tournaments to have competition, that therefore the only relevant legal jurisdiction is the tournament organiser itself (i.e. the league or federation). There is a large difference in organisational models between the North American closed leagues and the more loosely federated open leagues in Europe. It is a moot point as to whether the explicit reallocative interventions associated with closed leagues, which have engendered the close interest of antitrust authorities, really do generate greater ‘competitive balance’ than the looser regulatory framework of the open league. It may be that European Superleagues in sports such as football will develop along US lines, but such leagues have to overcome fierce local loyalties and rivalries, and the risk that ‘competitive balance’ alone is not sufficient to generate the revenues to support such undertakings.

Finally, as suggested at the beginning of this section there are issues concerning the regulation of financial viability, of penalties for approaching and/or recruiting players who are registered with other teams, and the methods of disciplining or punishing miscreants by league rules: whether officials, managers or players of clubs. At a straightforward level, economic analysis of the incidence of fines and taxes and restrictions on contracts would tell us to be cautious in confusing the notional incidence of these interventions (that is, on whom they are levied) with who actually bears the cost. It also appears hard to justify restrictions (for example on labour contracts) over and above those that would be applied to participants in other, similar, labour market activities.

Figure 1

A remuneration pyramid in sport



References

- Atkinson, A. and Stiglitz, J. (1980) *Lectures on Public Economics*, London: McGraw-Hill.
- Bairam, E., Howells, J. and Turner, G. (1990) 'Production functions in cricket: The Australian and New Zealand experience', *Applied Economics*, 22, 871-879.
- Blackham, J. and Chapman, B. (2004) 'The value of Don Bradman: additional revenue in Australian Ashes tests', *Economic Papers* (Economic Society of Australia), 23, 369-385.
- Bower, T. (2003) *Broken Dreams: Vanity, Greed and the Souring of British Football*, Pocket Books: London.
- Bulow, J., Geanakoplos, J. and Klemperer, P. (1985) 'Multimarket oligopoly: strategic substitutes and complements', *Journal of Political Economy*, 93, 488-511.
- Cave, R. and Crandall, R. (2001) 'Sports rights and the broadcast industry', *Economic Journal*, 111, F4-F26.
- El-Hodiri, M. and Quirk, J. (1971) 'An economic model of a professional sports league', *Journal of Political Economy*, 79, 1302-1319.
- Feess, E. and Muehlheusser, G. (2003) 'The impact of transfer fees on professional sports: An analysis of the new transfer system for European football', *Scandinavian Journal of Economics*, 105, 139-154.
- Flynn, M. and Gilbert, R. (2001) 'The analysis of professional sports leagues as joint ventures', *Economic Journal*, 111, F27-F46.
- Forrest, D. and Simmons, R. (2002) 'Outcome uncertainty and attendance demand in sport: the case of English soccer', *The Statistician (JRSS Series D)*, 51, 29-241.
- Hoehn, T. and Szymanski, S. (1999) 'The Americanisation of European football' *Economic Policy*, 14, 203-240.
- Késenne, S. (2000) 'The impact of salary caps in professional teams sports', *Scottish Journal of Political Economy*, 47, 422-430.
- MacDonald, D. and Reynolds, M. (1994) 'Are baseball players paid their marginal products?' *Managerial and Decision Economics*, 15, 443-457.
- Mayer, T. (1960) 'The distribution of ability and earnings' *Review of Economics and Statistics*, 42, 189-195.
- Michie, J. and Oughton, C. (2004) 'Competitive balance in football: trends and effects', *Football Governance Research Centre, Research Paper No. 2*, Birkbeck College, London.
- Noll, R. (1974) (ed.) *Government and the Sports Business*, Washington D.C.: The Brookings Institution.
- Noll, R. (2002) 'The economics of promotion and relegation in sports leagues: the case of English football', *Journal of Sports Economics*, 3, 169-203.
- Noll, R. (2003) 'The organization of sports leagues', *Oxford Review of Economic Policy*, 19, 530-551.

- Rosen, S. (1981) 'The economics of superstars', *American Economic Review*, 71, 845-898.
- Rosen, S. (1986) 'Prizes and incentives in elimination tournaments', *American Economic Review*, 76, 701-715.
- Rosen, S. and Sanderson, A. (2001) 'Labour markets in professional sports', *Economic Journal*, 111, F47-F68.
- Rottenberg, S. (1956) 'The baseball players' labor market', *Journal of Political Economy*, 64, 242-258.
- Schmidt, M. and Berri, D. (2003) 'On the evolution of competitive balance: The impact of an increasing global search', *Economic Inquiry*, 41,
- Schofield, J. (1988) 'Production functions in the sports industry: an empirical analysis of professional cricket', *Applied Economics*, 20, 177-193.
- Scully, G. (1974) 'Pay and performance in major league baseball', *American Economic Review*, 64, 915-930.
- Szymanski, S. (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, S. and Késenne, S. (2004) 'Competitive balance and gate revenue sharing in team sports', *Journal of Industrial Economics*, 52, 165-177.