The Economic Journal, **111** (*February*), F69–F84. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

INCOME INEQUALITY, COMPETITIVE BALANCE AND THE ATTRACTIVENESS OF TEAM SPORTS: SOME EVIDENCE AND A NATURAL EXPERIMENT FROM ENGLISH SOCCER*

Stefan Szymanski

This paper examines the relationship between financial inequality, competitive balance and attendance at English professional league soccer. It shows that while financial inequality among the clubs has increased, competitive balance has remained relatively stable and match attendance appears unrelated to competitive balance. A clearer test of the relationship is suggested by comparison with FA Cup matches. Because income inequality is primarily driven by inter- rather than intra-divisional inequality, the FA Cup has been a much more unbalanced competition than the divisional championships. Attendance at FA Cup matches relative to the corresponding league matches has fallen over the last twenty years.

It is widely accepted that a degree of competitive balance is an essential feature of attractive team sports.[1] Sporting competition is a process that establishes a hierarchy among the participants – winners and losers. Competitive balance refers to the rational expectations of fans about who will be the winners. In a perfectly balanced contest, each participant starts with an equal chance of winning, so that the outcome will be completely uncertain. If there is no competitive balance then the exact outcome can be predicted with probability one. Without at least a degree of competitive balance, fans will lose interest in a competition. However, it is less clear that every decline in competitive balance will lead to a falling off of fan interest.

This is not merely a matter of academic concern. In the recent Premier League Broadcasting case, heard in the UK Restrictive Practices Court, the court decided that selling broadcast rights collectively (and preventing clubs from selling any broadcast rights individually) was in the public interest, in part because the collective sale promoted financial equality, which in turn promoted competitive balance/uncertainty of outcome. Similar views underlie the US Sports Broadcasting Act of 1961 (which exempts the collective sale of broadcast rights on national TV from antitrust prosecution), and the comments of the Advocate General of the European Court of Justice in the Bosman case.

Thus the received opinion contains two logical steps: (i) increasing income

* I would like to thank Syariza Kamsan for valuable research assistance. I also thank Steve Ross, Chris Walters, seminar participants at Edinburgh University, Salford University, London Economics and two anonymous referees for valuable comments.

[1] This may be a peculiarly modern phenomenon reflecting ethical sensitivities – the Romans, for instance, appear to have enjoyed the unbalanced contest between lions and Christians.

inequality tends to reduce competitive balance and (ii) competitive imbalance tends to reduce fan interest.[2]

This paper does two things. First it develops a simple theoretical model of league competition to show that increasing competitive balance is not always desirable. Fan interest depends on several factors, and while competitive balance is one, an equally important consideration is the success of each of the teams that fans support. If fan support is unequally distributed between teams (e.g. for demographic reasons) then a utilitarian welfare function is likely to suggest that imbalance in favour of more strongly supported teams is optimal. For example, it is currently said that 50% of committed football fans in England support Manchester United – if this is so it is difficult to argue that aggregate welfare is not enhanced by the relative success of this team.

The second contribution of this paper is to suggest a natural experiment to test for the relationship between income inequality, competitive balance and fan interest. Testing for the existence of the hypothesised relationships is fraught with difficulties because of the many factors that affect fan interest from season to season. While it may be possible to test whether or not the competitive balance of a match affects interest (e.g. attendance) it is more difficult to test for the competitive balance of an entire competition. Yet in terms of the theory, it is presumably the competitive balance of a competition which is most important in the long run.

English soccer provides a natural experiment that overcomes these problems. Because teams compete simultaneously in both league (Football League and Premier League) competition and cup (FA Cup) competition, we can compare the trend in support for each of these competitions. In the league competitions teams are segregated into divisions, while in the Cup competition teams from different divisions can be drawn against each other. Over recent years income inequality has grown, most noticeably between rather than within the divisions. This implies that the FA Cup has become a much more unbalanced competition relative to the league division championships. We can thus ask whether attendance at FA Cup matches has declined relative to league matches. We do this by creating a matched sample of same-division matches played in the league and the Cup (these constitute around half of all matches in most seasons). The matching controls for many of the possible differences between matches (e.g. team strength, local interest, demographic and economic factors). What remains can be attributed to the intrinsic imbalance of the FA Cup relative to league championships.

[2] '... An important element in the maintenance of the quality of the Premier League competition is competitive balance, that is to say the unpredictability of the outcome of a high proportion of the matches played within the competition and thus uncertainty about which club will win the championship ... we accept that an increase in financial inequality will tend to result in a reduction of competitive balance' *RPC Court Judgment, Premier League.*

'[A] professional league can flourish only if there is no too glaring imbalance between the clubs taking part. If the league is clearly dominated by one team, the necessary tension is absent and the interest of the spectators will probably lapse within a foreseeable period ... it is of fundamental importance to share income out between the clubs in a reasonable manner ...' *Advocate-General Lenz, Bosman.*

The results do indeed show a relative decline in attendance at same-division matches. The rest of this paper is set out as follows. The next section develops a simple theoretical analysis of league structure and competitive balance. The following section develops the natural experiment and discusses some robustness issues. The final section draws some conclusions.

1. Competitive Balance in Theory

The relationship between income distribution, competitive balance and the attractiveness of sporting competition has received a limited amount of attention in the theoretical literature. This literature has been primarily been concerned with the proposition that income redistribution will lead to greater equality of outcomes. Quirk and Fort (1992) and Vrooman (1995) analysed this question and concluded that competitive balance would be unaffected by redistributive mechanisms such as gate sharing. Under gate sharing, the visiting team receives a fixed percentage of the home team gate (e.g. in the US NFL 40% of gate income is allocated to the visitors). The basis of their argument is as follows. If teams earn more from home matches when they are expected to win (i.e. winning teams attract more support) and the visiting team share the gate revenue, then the visiting team would prefer to be less successful compared to the case where they do not share in the gate receipts. This will lead both teams to invest less in winning (i.e. to invest less in playing talent) compared to the case where there is no gate sharing, but in their models this effect impacts equally on both teams so that competitive balance is unaffected.

Szymanski (1998) argues that gate sharing may even have the perverse effect of reducing competitive balance. In a model where there is constant marginal cost of talent but revenue is a strictly convex function of playing success (measured by win percentage), gate sharing will diminish the investment incentives of small market teams by more than that of large market teams. Intuitively, small teams stand to gain more from the success of the big teams than the big teams stand to gain from the success of the small teams. Hence small teams reduce investment in talent by more than the big teams. However, that paper also shows that the impact of redistribution on competitive balance depends critically on the objectives of the teams and on the type of redistribution scheme. The standard models of team sports in the United States assume clubs are profit maximisers, while researchers in Europe (e.g. Sloane (1971)) have presented evidence that, at least traditionally, clubs have been 'win maximisers' (or some variant of this). Win maximisation implies all surplus income is reinvested in talent. Under these circumstances, income redistribution from large to small clubs will tend to improve competitive balance.

Even if clubs are profit maximisers, schemes that raise income independently of playing success (e.g. collective selling of TV income), and then redistribute that income on the basis of playing success, will tend to improve competitive balance. This is because what matters for competitive balance is the investment decision of the teams, which in turn depends on the access of

teams to the income pool *ex ante*, rather than the *ex post* share-out. Income raised through lump sum taxes will not distort incentives while redistribution on the basis of performance will give equal incentives to all. Teams with a small drawing power are no longer deterred from investing in talent because of the limits imposed by their local market.

TV income is an interesting case, not least because of the interest expressed by competition authorities in the desirability of centralised collective selling. In theory, collective selling and the distribution of TV income purely on merit (e.g. on the basis of league ranking) will enhance competitive balance by giving small market teams equal access to the TV market. A club with a small local market can finance a successful team if TV income is the dominant source of finance. In practice collective TV revenues tend to be distributed equally (as in the US NFL) or only partly on the basis of merit. In the English Premier League only 25% is distributed on the basis of league ranking, while 50% is allocated as an equal share and the remaining 25% is awarded on the basis of the number of TV appearances.

However, these researchers have also questioned the value of competitive balance. If some teams draw on larger (or more devoted) fan bases, then the success of these teams will yield greater total utility than the success of teams with small fan bases. The important theoretical issue is whether unfettered competition or a regulated market will deliver the socially optimal outcome. A simple model may help to illustrate this point. Suppose that there is a sports competition consisting of two teams, one of which enjoys a larger domestic market than the other, in the sense that it will generate a higher level of fan utility for a given level of playing success. Fan utility depends on playing success, which in turn depends on the proportion of playing talent hired by each team. Hence the fan utility for each team can be written as follows:

$$U_1 = \mu_1 w_1 = \frac{\mu_1 t_1}{t_1 + t_2}, \qquad U_2 = \mu_2 w_2 = \frac{\mu_2 t_2}{t_1 + t_2} \tag{1}$$

where $\mu_1 > \mu_2$ reflects the intensity of support, w is the win percentage of each team and t is the quantity of playing talent hired by each team. Total utility will depend not only on the utility of committed team fans, but also on spectators with no particular loyalty to a team. These supporters might be labelled 'uncommitted', or, more pejoratively, 'couch potatoes', watching matches on TV and motivated only by an attractive spectacle. The utility of these spectators is thus dependent on competitive balance.[3] We adopt here a simple cardinal representation of total utility:

$$U = U_1 + U_2 + \theta w_1 w_2 \tag{2}$$

where θ represents the weight of couch potatoes relative to committed team fans in total utility. Maximising total utility with respect to the win percentage yields following social optimum:

[3] Most fans are likely to value the total quality of the playing talent involved in a match as well as competitive balance. The addition of total playing talent as an argument in the social welfare function will not affect the qualitative results.

$$w_1^* = \frac{1}{2} + \frac{\mu_1 - \mu_2}{2\theta}. \tag{3}$$

As θ becomes very large, only competitive balance matters. However, depending on the difference between the intensity of support for each team, as θ diminishes the social optimum implies higher degrees of inequality. Unless the two teams are equally well supported, there will exist a critical value of θ which implies that total utility would be maximised even if the more popular team never lost. The model implies a trade-off between the interests of the committed and uncommitted fans. The optimal balance depends on the relative weights placed on each of these populations.

It is clear from this that if redistribution of income led to an equal distribution of resources the outcome would be an equally balanced contest. This would be socially optimal only if either the intensity of support for each team were equal or the weight attached to utility of the uncommitted fans dominated completely. Perfect balance is not generally desirable.

To model the outcome of a competitive market some assumptions are required about the objectives of clubs and the form of their objective functions. Here it will be assumed that clubs are profit maximisers, in line with the US literature, and an increasingly plausible assumption in English soccer now that the largest clubs are quoted on the stock market. It is assumed that the clubs are able to appropriate a fraction of the utility of fans through the sale of tickets and related products, while they can hire playing talent in the market at a constant marginal cost. Thus

$$\pi_i = \phi_i \mu_i w_i - c t_i \tag{4}$$

where ϕ is the fraction of winning utility that the clubs can appropriate. It is also assumed that clubs are unable to appropriate any of the utility derived by the uncommitted fans. This may be somewhat extreme, but in general one might expect that it is relatively difficult to generate income from this group. It is easy to show that profit maximisation implies the ratio of talent at each club will equal the ratio of intensity of support (μ_1/μ_2). Comparing this with the socially optimal level of talent at each club implied by (3), it is clear that the socially optimal level of competitive balance would arise only by chance. If $\mu_1 = \mu_2$ the market outcome is socially optimal. However if $\mu_1 > \mu_2$ then the social optimum would only be achieved for a particular value of θ. This critical value is increasing in μ_1. In other words, the greater the weight attached to the utility of uncommitted fans the stronger the intensity of support for team 1 would have to be to achieve the social optimum. If intensity of support for team 1 is too large, the contest will be less balanced than the social optimum, if it is too small the contest will be more balanced than is socially optimal.

A simple model such as this captures some basic ideas about the relationship between inequality in the distribution of resources, competitive balance and social optimality. The basic insight is that while perfect competitive balance is not desirable, the market equilibrium is unlikely to achieve the social optimum. In particular, intensely supported teams are likely to create excessively

unbalanced competitions. This might be taken as grounds for limited redistribution. Whether members of a league or the league authorities themselves will be able to impose such redistribution depends largely on their ability to appropriate the surplus of uncommitted fans – otherwise they have no incentive to act and no basis for an agreement. In such cases, intervention by an independent regulator committed to the best interest of the sport will be desirable, at least in theory.

2. Competitive Balance: A Natural Experiment

2.1. *The Trend in English League Soccer*

Before developing the natural experiment it is useful to review the data from the league alone to indicate the difficulty in analysing competitive balance. This paper deals with competitive balance in a sense which has not generally been examined in the earlier literature. Kuypers (1997)[4] defines competitive balance in three senses: the balance of attractiveness of a match, the closeness of a championship race and the absence of long run domination. Most previous studies have concentrated on competitive balance in the first two senses. Thus Jennett (1984), Peel and Thomas (1988), Cairns (1988), Jones and Ferguson (1988) and Kuypers (1997) concentrate on the match uncertainty. They hypothesise that uncertain matches will attract greater support and focus on finding suitable proxies for match uncertainty. Demmert (1973), Noll (1974), Whitney (1988) and Kuypers concentrate on the closeness of specific championship races and examine whether this increases attendance at matches. In the present study the focus of interest is the balance of the entire championship over a period of time. In league competition this can be measured by the variance of team winning (win percentage) over time or the dominance of high ranks by particular teams. For Cup competition, since the teams also participate in hierarchical leagues, competitive balance can be analysed by looking at the success of teams from different divisions.

The growing financial inequality in English is widely remarked upon. However, most of this growth in inequality is inter- rather than intra-divisional, as is shown in Fig. 1. This graph shows the coefficient of variation of income for a sample of 39 clubs over the 22-year period 1976/7 to 1997/8. This makes it difficult to test for the impact of growing inequality in league soccer, since there are no inter-divisional matches. Moreover, it is not evident that there has been any significant increase in intra-divisional competitive balance either within or between seasons. Table 1 illustrates the absence of any significant trend in dominance over time, measured by the number of teams accounting for the top positions over different time periods (three and seven years). While there is some slight evidence of increasing dominance in the Premier League over the last three years in the sample, there is no clear trend.

It is useful to consider the mobility of teams between the divisions. In any

[4] Kuypers (1997) and Szymanski and Kuypers (1999) provide a useful survey; for an earlier survey see Cairns *et al.* (1986).

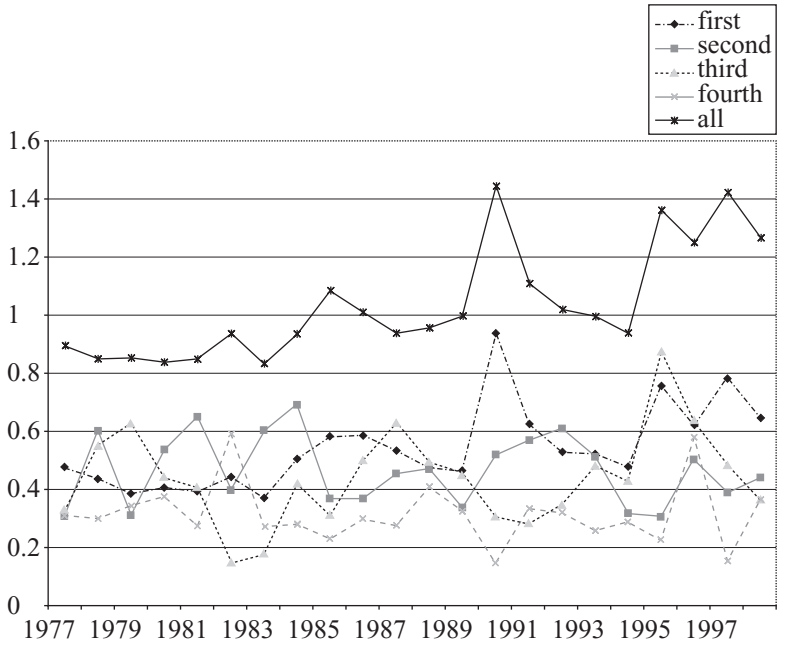

Fig. 1. *Coefficient of Variation of Sales by Division 1977–1998*

Table 1

Number of Teams in Each of the Top N Positions by 3 and 7 Year Intervals (1978–98)

(a) Three Year Intervals

Number of teams in top 3					Number of teams in top 5					Number of teams in top 10				
	division					division					division			
Period	1	2	3	4	Period	1	2	3	4	Period	1	2	3	4
1	6	9	9	6	1	9	12	15	11	1	13	18	18	16
2	6	8	9	9	2	9	11	12	13	2	13	15	19	22
3	6	9	7	8	3	8	13	11	13	3	11	20	16	22
4	6	9	8	8	4	8	14	14	12	4	12	18	18	21
5	8	7	9	7	5	10	11	13	11	5	15	19	20	21
6	6	7	6	8	6	10	11	12	11	6	15	18	19	19
7	4	8	9	8	7	7	13	13	13	7	13	19	18	22

(b) Seven Year Intervals

Number of teams in top 3					Number of teams in top 5					Number of teams in top 10				
Period	1	2	3	4	Period	1	2	3	4	Period	1	2	3	4
1	8	17	15	17	1	10	17	22	23	1	12	19	25	31
2	7	14	17	19	2	9	17	20	28	2	11	21	23	34
3	8	12	16	20	3	11	17	19	27	3	15	17	25	31

one year there are 92 league clubs, and over the seasons 1976/7 to 1997/8 there have been 99 teams participating in the four divisions, given that there have been a small number of demotions to the lower semi-professional divisions. Of these 99 teams, only 5 have never been relegated or promoted over the period, indicating that there is a fairly high degree of mobility between the divisions.[5] Furthermore, over the period more teams have ranged between three divisions (43) than have moved only between two (32), while 12 teams managed to visit all four divisions over the space of 22 years.

Dominance within seasons is considered in Fig. 2. This shows the standard deviation of win percentage over time. This measure is the mostly widely used indicator of competitive balance in the US literature and although there are a greater proportion of drawn games in soccer, win percentage is still a reliable indicator of success. It is closely correlated with the more usual measures of success such as league position (correlation coefficient 0.91) and points scored (0.95). Perhaps surprisingly, the charts show that there is no clear trend in win percentage, suggesting that divisional championships have not tended to become more one sided over time. Fig. 2 also illustrates the change in attendance at league matches over the 22 seasons, that may be taken as an indicator of fan interest. There have been two very distinct phases – a secular decline in attendance until 1985 and a consistent increase thereafter. This is in itself puzzling given that most pundits have generally dated the recovery of

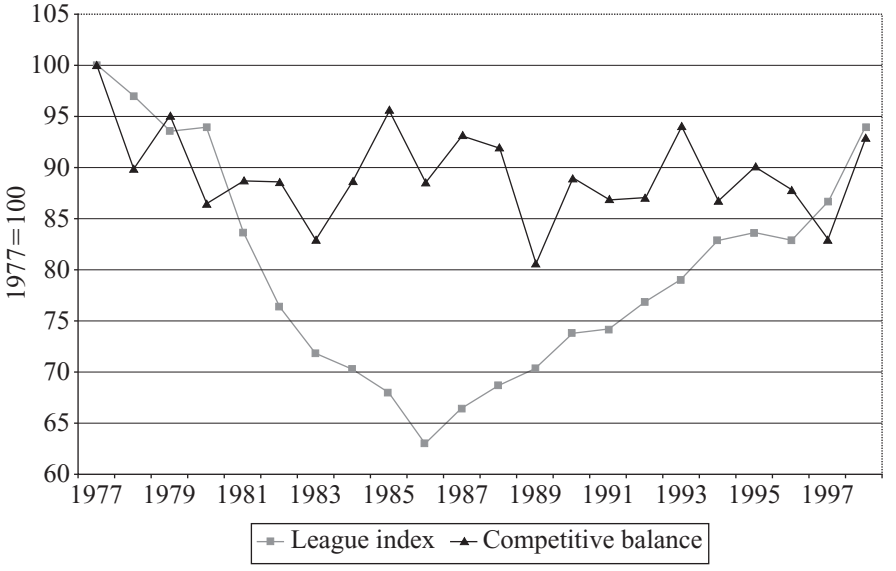

Fig. 2. *Competitive Balance and League Attendance Trends 1977–1998*

[5] Four of these teams have remained in the top division (Arsenal, Coventry, Everton and Liverpool) while one has remained in the lowest division (Rochdale).

interest in English football at 1990 (when England reached the semi-final of the World Cup) or even 1992 (the foundation of the Premier League). There were still problems in English football in the late 1980s (high levels of crowd violence, poor facilities at stadiums and high levels of policing. Worst of all was the Hillsborough stadium disaster of 1989 in which 95 fans were crushed to death).

One explanation may be derived from the model outlined in Section 1. The trend growth in financial inequality started from a point in the early post-war period when income was quite evenly distributed and clubs were restricted by a maximum wage rule that limited team expenditure. Thus until 1961 teams with large potential supporter bases were constrained to hire teams of roughly equal ability to those with small potential supporter bases. In terms of the model, the constrained equilibrium meant that the success of the larger teams was below the optimal level. Once the maximum wage was abolished clubs could utilise their greater resources to achieve a higher rate of success and this may have led both to less competitive balance and greater interest in the league football.

2.2. *The Natural Experiment*

The idea of a natural experiment is to identify two sets of data in which competitive balance differed significantly but all other relevant factors are the same. In US team sports where clubs compete only in a single tournament, such natural experiments are not available, but this is not true of soccer. Traditionally teams participate in two main competitions during the season, a league competition and a cup, or knock-out competition. The oldest such competition in the world is the FA Cup in which all 92 league clubs compete annually. The FA Cup is in fact open to all registered football clubs in England, and amateur teams compete in preliminary rounds. However, the first round of the Cup consists mainly of the teams in the two lowest divisions. The teams in the top two divisions do not enter the competition until the third round which consists of 64 teams.

In each round of the competition the matches are determined by a random draw. In many cases the opposing teams are from different divisions, but on average one third of matches from the third round onwards are contests between teams from the same division, and an even greater proportion of matches in rounds 1 and 2 are same-division matches. This is the basis for a natural experiment. Using a sample of about one thousand same-division FA Cup matches over the last 22 years we can compare attendance with attendance at the equivalent league fixture played in the same season (this includes equivalence in the sense that the same team has home advantage). As was pointed out in the previous section, the main source of the growth in inequality between league teams has been the growth of inter-divisional income inequality. If income inequality leads to a less balanced contest, we should expect to see a lower degree of fan interest in a Cup fixture, which forms part of a more unequal championship than the corresponding league fixture. The

test is therefore not a test of the attractiveness of a fixture in its own right, say as a function of the quality of the teams or the history of competition between the two teams, rather it is a test of the relative attractiveness of the championships in which the two teams are participating. In fact it is a very low powered test. It excludes from consideration matches between teams from different divisions which might be thought to be particularly unbalanced and therefore to attract fewer spectators. These contests are excluded because there are no equivalent league fixtures with which they can be compared.

Of course, the natural experiment cannot control for every possible source of difference between the two fixtures. Match attendance can be affected by the current form of the two teams, the day of the week on which a match is played (weekend matches tend to have higher attendance) and the point of the season at which the match is played (end of season matches tend to have higher attendance). These factors can be controlled for through a regression analysis, although one might expect that in a matched sample as large as one thousand these factors would not exert systematic effect.

While the data make clear that income inequality between the divisions has grown over recent years, it is not so easy to establish that competitive balance has in fact declined. The standard deviation of win percentage or other indicators of success make little sense in this context. One way to compare is to look at the survival of teams from different divisions. There is surprisingly little evidence of a trend toward domination by the larger clubs, perhaps because even in 1977 the large clubs dominated the FA Cup. Thus in that season 78% of all appearances in the FA Cup from the third round on were from teams in the first and second division. Given that 64 teams enter the third round of which 44 are from the top two divisions, the theoretical maximum share of the top two divisions is 84% (106 out of 126 appearances) while the theoretical minimum is 44% (56 appearances). Over the period the share of the top two divisions never fell below 72%, within 12% of the theoretical maximum (see Fig. 3). The top division clubs dominate the final rounds of the competition. There were only three cases of a club from outside the top division appearing in the final in the 22 years from 1977 to 1998, and only 15 cases out of a possible 88 of such a team playing in a semi-final. The lowest number of top division clubs to survive to the fourth round in any year during this period was 10, while on average 14 survived. Since the Premier League was reduced to 20 teams in 1996 at least 15 teams have survived the third round in each year.

There is some evidence of increasing dominance. The proportion of cases where a team from a lower division wins a match has declined. Between 1977 and 1987 on average just over 11 of the 63 matches played per season (from round 3 on) resulted in a win for the lower division team. From 1988 to 1998 the average fell to just under 10. The incidence of 'giantkilling', defined as a team beating an opponent placed at least two divisions higher, has also fallen. Such events are in any case rare, there having been only 67 cases in the 22 seasons covered by the data, an average of 3 per season. Between 1977 and 1987 there were 42 cases, an average of over four per season, while between 1988 and 1998 there were only 25 cases, an average of only just over 2 per

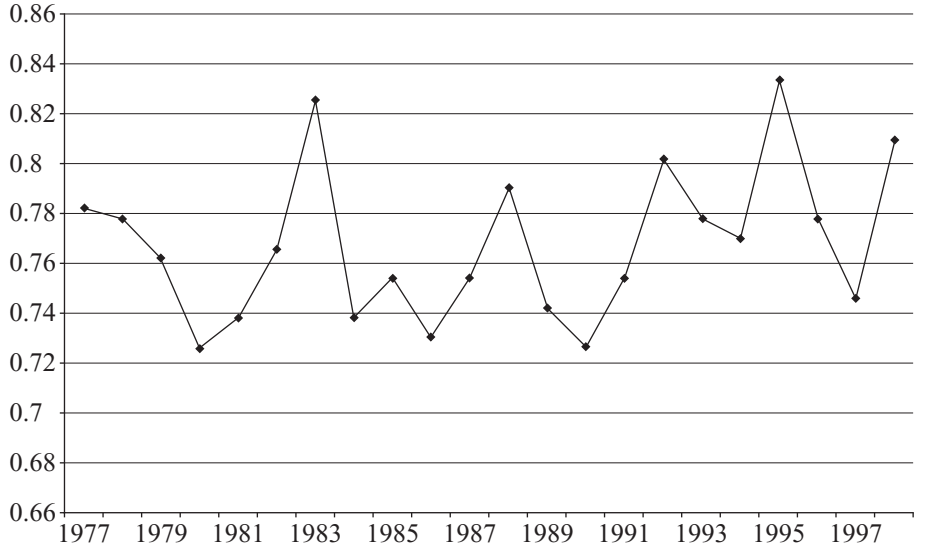

Fig. 3. *Share of Division 1 & 2 Teams in all FA Cup Matches Played from the Third Round Onwards*

season. This evidence seems to suggest that an already unbalanced contest has become yet more unbalanced.

The natural experiment suggested here is that if the competitive balance of a championship taken as a whole affects the attractiveness of individual matches, we should see a relative decline in attendance at matches in a championship whose balance is deteriorating faster. Since the inequality of income has grown faster as between participants in the FA Cup competition and participants in divisional championships, we should expect to see a relative decline in attendance at FA Cup matches. This does indeed appear to be the case, based on the sample of 997 same division matches played between 1977 and 1998.[6]

Fig. 4 shows a graph of the ratio of average attendance at FA Cup and league matches over time. FA Cup matches are traditionally better supported than league matches, and between 1977 and 1987 the average FA Cup fixture would attract an audience 43% larger than the equivalent league match. This difference declined to 25% in the second half of the data period, and declined almost continuously until 1998 in which year FA Cup matches attracted a slightly smaller audience on average. Thus even with this very low powered test, there appears to be have been a significant decline in the relative attractiveness of the FA Cup during a period when inter-divisional income inequality was

[6] The database includes replays of drawn matches. The inclusion of these games might be thought to bias the average FA Cup gate downwards since replays tend to be scheduled at times other than the weekend, making it more difficult for fans to attend. On the other hand, a replay may be a good indicator of an exciting match, since the previous draw already indicates a degree of competitive balance. In any case, a separate analysis of decisive matches only did not indicate any systematic difference in the underlying trend.

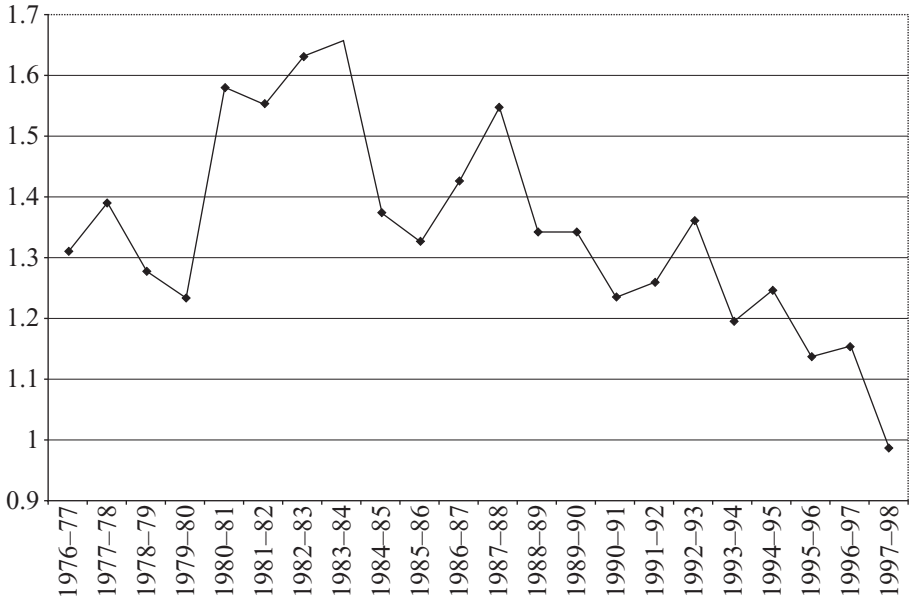

Fig. 4. *FA/League Attendance Ratio*

growing and there appears to have been some deterioration in the competitive balance of the FA Cup.

The database was compiled using matches from the first two rounds of the Cup, involving the lower divisions, as well as the later rounds. This ensures that the sample contains a large number of Cup matches between teams in the lower divisions. It might be suspected that the effect shown here was related primarily to the Premier League, where much of the media interest has been focused. It might also be suspected that the results were due to the fact that many top division cup matches tended to be played to capacity stadia, and that the falling ratio was an artefact of this constraint during a period when interest in football was growing. To deal with this issue and other potential factors that might influence attendance a regression approach was adopted.

The matched sample for the seasons 1982/3 to 1997/8 was used to analyse any trend in attendance at FA Cup matches relative to league matches. The results are reported in Table 2. Column 1 reports an OLS equation for the full data set. Column 2 reports a Tobit equation for the full sample, with upper censoring to account for the fact that about 10% of matches appear to have been played at capacity. The capacity figure was approximated as 95% of the figure reported in the Rothmans Football Yearbook. This Rothmans figure is likely to be an overstatement because capacity is limited for some matches by the requirement to have adequate segregation of fans, leaving many seats deliberately unoccupied.[7] Column 3 reports an OLS equation for the sample omitting top division teams, which account for 92% of sell-outs in the data.

[7] Capacity data is problematic. The Rothmans figure is also misleading for the 1990s when there was significant stadium rebuilding, often during the season. However, no other capacity figures are available.

Table 2

Attendance Regressions 1982/3–1997/8 Seasons

	OLS whole sample	Tobit whole sample	OLS excluding first division
Constant	28,323	27,774	12,867
	(23.095)	(38.317)	(13.814)
2nd division	−11,442	−11,181	
	(−19.592)	(−18.416)	
3rd division	−17,525	−17,012	−6,452.5
	(−35.905)	(−41.573)	(−15.001)
4th division	−20,043	−19,373	−8,402
	(−40.134)	(−48.418)	(−19.973)
Sum of team league positions	−262.38	−174.58	−138.34
	(−12.73)	(−12.812)	(−7.852)
Match played on Sunday	1,690.8	1,609.4	913.31
	(1.712)	(1.949)	(1.104)
Season month	129.73	−201.89	253.44
	(1.301)	(−3.149)	(3.397)
Replay	−1,145.1	−1,080.2	458.53
	(−1.812)	(−2.536)	(0.758)
FA Cup match 1983–86	3,617.5	3,519.9	1,998.8
	(4.322)	(6.706)	(2.815)
FA Cup match 1987–90	3,456.3	3,283.3	1,989.3
	(4.255)	(7.408)	(3.777)
FA Cup match 1991–94	912.61	926.63	543.19
	(1.410)	(1.803)	(1.302)
FA Cup match 1995–98	−138.29	−179.21	279.83
	(−0.166)	(−0.35)	(0.449)
Observations	1,286	1,286	772
R^2	0.637		0.479
Log L	−13,856	−12,713	−7,766

Heteroscedastic consistent t-statistics in parentheses. Time dummies included but not reported. Tobit coefficients are marginal effects.

The models were estimated with heteroscedastic consistent errors. In the case of the Tobit estimation a multiplicative model of heteroscedasticity was adopted (see Greene (1993)). In both the OLS and Tobit models there was evidence of non-normality using Pagan Vella tests. This is known to be a particular problem in the Tobit model since it renders the estimator inconsistent. However, the consonance of the Tobit and OLS results provides a little comfort.

The regressions account for the day of the week the game was played, the sum of league positions, the month and whether the match was an FA Cup replay as well as divisional and time dummies. Sunday matches tended to attract a higher attendance than Saturday matches, possibly because broadcast matches are usually scheduled for a Sunday and broadcasters select the most attractive games.

Most FA Cup replays are in midweek, and as a result these two variables tended to pick up the same effect. If anything, the replay variable worked better. The sum of league positions variable (positions at the date the game was played) picks up the quality of the teams on show. If competitive balance

mattered for attendance at individual matches, one might have expected that the absolute difference in league positions would be significant – but in fact it was not.

The FA Cup effects are represented as dummies for successive four year periods. The estimates suggest that between 1982/3 and 1989/90 an FA Cup match would attract about 3,000 more spectators than an league match. The OLS figure is somewhat higher, the figure for the lower three divisions is somewhat lower. For the seasons 1990/1 to 1993/4 FA Cup matches still attracted more bodies on average, but the size of this effect (between 500 and 1,000 people) is not statistically significant at the conventional 5% level.[8] The dummy variables for the last four-year period (1995–8) are all much smaller (two indicate a negative impact of FA Cup matches on attendance) and are all statistically insignificant. Thus the regression analysis appears to support the evidence of Fig. 4 – in the 1980s FA Cup matches would attract significantly higher attendance than equivalent league matches – in the 1990s this effect has disappeared, and FA Cup matches attract attendances that are no higher on average than equivalent League matches. The 'magic of the Cup' seems to be fading.

Before concluding there are two possible flaws in the natural experiment that should be considered. If other factors had altered the relative attractiveness of attending matches in league and Cup competitions then the trend in the ratio might be attributed to these factors rather than competitive balance. Firstly, if the price of tickets for Cup matches relative to league matches had risen, this might have caused the relative decline in attendance. Price data are not available for the entire period, but the annual Football Trust Digest of Football Statistics provides an analysis of FA Cup and League gate receipts between 1984/5 and 1994/5. The ratio of prices derived from these data shows no overall trend, during a period when the relative decline of FA Cup attendance was pronounced. A more serious question is raised by the increasing tendency to sell season tickets. This means that for an increasing proportion of fans the marginal cost of attendance at league matches is effectively zero. This could account for the relative decline of interest in the FA Cup. However, the fact that clubs can sell an increasing proportion of seats for league matches in advance suggests that the attractiveness of the league competition has increased. If interest in the FA Cup had grown at the same rate, we might observe FA Cup season tickets being sold or simply higher prices for FA Cup matches. The fact that we do not suggests that the relative attractiveness of the Cup has indeed declined.

A second weakness of the experiment might be that the structure of the competitions themselves had changed enhancing the relative attractiveness of the League. Since 1986/7 a system of playoffs for some promotion places was introduced into the lower three divisions. The effect of this was to give more teams at any given time an interest in the possibility of promotion, and to involve every team a longer fraction of the season in contention. This has

[8] In the Tobit model the marginal effect is significant at the 10% level.

almost certainly stimulated interest in league competition in the lower divisions. To see if this effect was driving the relative decline of FA Cup attendance, those matched pairs which involved league matches played from March onwards (at which point progressively more teams are ruled out of contention) were omitted from the sample. However, for the remaining matches the relative decline of FA Cup attendance appeared just as pronounced as for the full sample.

3. Conclusions

This paper has attempted to draw out the relationship between the unequal distribution of resources, competitive balance and the interest of the fans. The resurgence of interest in English league football, in particular the Premier League, has occurred at a time when the distribution of income has become much more unequal. Many commentators have bemoaned this fact, worrying that it will lead to a decline of interest in soccer. So far, there is only weak evidence that the concentration of income has been associated with a decline in competitive balance, and no evidence at all that it has reduced interest in league football. It may be that the polarisation of recent years has been an adjustment away from an excessively egalitarian distribution toward an unequal distribution that more accurately reflects the interest of the fans. It may also be the case that competitive balance is only sensitive to very large changes in the income distribution, and hence growth in inequality may only have caused small changes in competitive balance.

There are many factors that influence attendance, and isolating the effect of competitive balance using only a short time series is unlikely to reveal capture all the dynamics of the underlying relationship. However, by comparing same-division fixtures that occur in both the FA Cup and the league we can conduct a natural experiment on the effect of growing inequality. The only important difference between the matched pairs is the competition in which they are played. Other sources of difference such as home advantage, the quality of the teams, form over recent seasons and so on are filtered out by the matching. Since inter-divisional inequality has grown much faster than intra-divisional equality, the FA Cup is a competition where the resources of the participants have become more unevenly distributed (compared to the league) over time. The data show, just as one might have predicted, that this relative increase in inequality has led to a relative decline in attendance. In the 1970s it was not unusual for attendance at an FA Cup match to be 50% higher than the attendance at the equivalent league fixture. By 1998 the average attendance at FA Cup matches was lower than at the matched fixture. Thus the natural experiment appears to confirm the standard hypothesis about the impact of income inequality and competitive balance on the attractiveness of sporting competition.

Imperial College London

References

Cairns, J. A. (1988). 'Outcome uncertainty and the demand for football', University of Aberdeen Department of Economics Discussion paper 88-02.

Cairns, J. A., Jennett, N. and Sloane, P. J. (1986). 'The economics of professional team sports: a survey of theory and evidence'. *Journal of Economic Studies*, vol. 13, pp. 3–80.

Court of Justice of the European Communities (1995). Case C-415/93 (The Bosman Judgment).

Demmert, H. (1973). *The Economics of Professional Team Sports*, Massachusetts: Lexington Books.

Greene, W. (1993). *Econometric Analysis*. Second Edition. Basingstoke: Macmillan.

Fort, R. and Quirk, J. (1995). 'Cross-subsidisation, incentives and outcomes in professional team sports leagues'. *Journal of Economic Literature*, vol. 33, September, pp. 1265–99.

Jennett, N. (1984). 'Attendance, uncertainty of outcome and policy in Scottish league football'. *Scottish Journal of Political Economy*, vol. 31, pp. 176–98.

Jones, J. C. H. and Ferguson, D. G. (1988). 'Location and survival in the national hockey league'. *The Journal of Industrial Economics*, vol. 36, no. 4, pp. 443–57.

Kuypers, T. (1997). Unpublished PhD Thesis. University College London.

Noll, R. (ed) (1974). 'Government and the sports business'. Brookings Papers.

Peel, D. and Thomas, D. (1988). 'Outcome uncertainty and the demand for football'. *Scottish Journal of Political Economy*, vol. 35, pp. 242–9.

Restrictive Practices Court (1999). 'In the Matter of an Agreement between the Football Association Premier League Limited and the Football Association Limited and the Football League Limited and their Respective Member Clubs and in the Matter of an Agreement relating to the supply of services facilitating the broadcasting on television of Premier League football matches and the supply of services consisting in the broadcasting on television of such matches.' (28th July).

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

Sloane, P. J. (1971). 'The economics of professional football: the football club as a utility maximizer'. *Scottish Journal of Political Economy*, vol. 17, no. 2, pp. 121–46.

Szymanski, S. (1998). 'Hermetic leagues, open leagues and the impact of revenue sharing on competitive balance', paper submitted as evidence to the Restrictive Practices Court in the Premier League/Sky/BBC case.

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

Vrooman, J. (1995). 'A general theory of professional sports leagues'. *Southern Economic Journal*, vol. 61, no. 4, pp. 971–90.

Whitney, J. (1988). 'Winning games versus winning championships: the economics of fan interest and team performance'. *Economic Inquiry*, vol. 26, October, pp. 703–24.