



The English Football Industry: profit, performance and industrial structure

Stefan Szymanski & Ron Smith

To cite this article: Stefan Szymanski & Ron Smith (1997) The English Football Industry: profit, performance and industrial structure, International Review of Applied Economics, 11:1, 135-153, DOI: [10.1080/02692179700000008](https://doi.org/10.1080/02692179700000008)

To link to this article: <http://dx.doi.org/10.1080/02692179700000008>



Published online: 28 Jul 2006.



Submit your article to this journal [↗](#)



Article views: 2837



View related articles [↗](#)



Citing articles: 78 View citing articles [↗](#)

The English Football Industry: profit, performance and industrial structure

STEFAN SZYMANSKI & RON SMITH

ABSTRACT *The English (Association) Football League is a long established industrial cartel selling a highly popular product with only imperfect substitutes. Despite that, the majority of its member clubs lose money and the industry has faced successive financial crises over the last decade. This paper develops an empirical model of the financial performance of English League clubs using a high quality dataset of 48 clubs over the period 1974–89. The underlying model explains how rents are competed away through the maximising behaviour of club owners subject to production constraints. This model is parameterised by a system of equations which describe the behaviour of a maximising owner subject to demand and production constraints. The model is then used to examine the coordination failure which lies at the heart of the English Football League's decline and to assess the prospects for the Premier League.*

1. Introduction

The idea of an industry, a set of competing specialist firms within a well-defined market, is little more than a metaphor in most of the economy, but in the area of professional team sports it represents a very precise description. This paper analyses the structure of the English Football industry. Despite what is generally perceived to be a very high and price inelastic demand for the services of professional football teams the clubs themselves seldom make money. In many other industries we see firms engaging in tacitly collusive behaviour in order to mitigate the impact of competition; in football we do not see such behaviour, at least in aggregate. In this paper we present a theoretical model of the behaviour of football club owners which rationalises the observed intensity of competition. We then fit that model to data for 48 English Football League Clubs over the period 1974–89. Thus we model the period before the changes during the 1990s, which we discuss below.

There is a large literature on the economics of professional team sports, which is surveyed in Cairns *et al.* (1986). Most of this literature has been concerned with one of two issues, first estimating the demand for professional team sports, and secondly trying to identify the objective function of professional clubs. However, few papers have attempted to marry the analysis of supply and demand conditions

Stefan Szymanski, Imperial College Management School, 53 Prince's Gate, Exhibition Road, London SW7 2PG, UK.

Ron Smith, Birkbeck College, London University, 7/15 Gresse Street, London W1P 2LL, UK.

in order to derive an equilibrium model of team performance. We attempt to develop such a model using accounting data as well as League performance data. The analysis highlights a coordination failure which encourages club owners to neglect the infrastructure of League Football whilst concentrating on short-term playing performance. Given that in most cases English clubs are neither the recipients of open-ended financial backing from industrialists nor receive significant support for the upkeep of stadia from municipal authorities, as is common in other countries, most clubs live on the brink of financial failure. Whilst these problems have been pointed out by previous research (see for example the Chester Reports of 1968 and 1983 (commissioned by the Football League); Jennett & Sloane, 1985; Arnold & Beneviste, 1987; and most recently the Taylor Report, 1989) there have been no previous attempts to model the economic problem systematically and confront it with the data.¹ The data set combines information on League performance and attendance with accounting data on revenue, wage and transfer expenditure.² The sample means are given in Tables 1a and 1b.

The broad structure of the model, which is set out in detail in Section 2, is as follow. Football skills are bought on a competitive market for players, with the quality adjusted wage being determined as a Nash equilibrium between the clubs. The amount of skill a club purchases determines its position in the League. This corresponds to the production function of the industry which we estimate. Position in the League determines the revenue a club earns from gate receipts, television rights, sponsorship, etc., and this corresponds to the demand function of the industry which we also estimate. Combining the estimated demand and production functions with the budget constraint we derive the empirical trade-off that the club faces between profit and position in the league. This has a negative slope which indicates that increased spending on players is not self-financing through higher performance and revenue. The objective function of the owner of the club depends on both profits and position (see for example, Sloane, 1971). Maximisation of the objective function, subject to the profit-position constraint, determines the optimal level of wages and thus the club's profits and position in the League. Whereas each individual club faces a negative trade-off between profit and position, between clubs profits and position are positively related. This is because clubs differ in their endowment, their inherent capability and revenue generating potential, and these differences in endowment generate the observed distribution of profit and position.

The estimates of the model are presented in Section 3, Section 4 deals with some econometric issues and Section 5 looks at some alternative specifications. Section 6 discusses some of the implications of the industrial structure and in particular how the interaction of non-profit objectives with the lack of an effective market for corporate control inhibits restructuring of the industry. This is discussed in the light of the changes introduced in the 1990s.

2. A Model of the English Football League

The English Football League contained 92 clubs divided into four divisions.³ We will consider a particular club, i , which can be characterised by the quality of its performance, Q_i , a function of its position in the League. Most managers and owners seem to regard this as the most significant indicator of the performance of a club. Whilst success in Cup competitions such as the FA Cup is highly prized,

Table 1a. Average annual performance by club, 1974-89

	League position	Turnover	Gate receipts	Wage spend	Net transfer spend	Pre-tax profits	Number of major cup games
Liverpool	2	3592	3208	1630	617	172	14
Everton	7	2943	2263	1486	416	25	10
Arsenal	7	921	2206	1493	307	205	10
Manchester United	8	4653	4340	1700	493	- 17	10
Tottenham Hotspur	11	3337	2868	1695	501	20	10
Aston Villa	14	2368	1593	1049	285	59	10
Coventry	14	1416	1050	963	- 152	- 33	8
Southampton	15	1470	1113	817	312	- 110	9
West Ham	16	2144	1980	1135	276	38	9
Leeds	18	1649	1604	1039	251	- 379	8
West Bromwich Albion	19	1560	1191	919	- 53	19	8
Newcastle	20	1810	1810	950	60	- 64	7
Luton	21	1546	1392	845	30	- 18	8
Leicester	21	1343	976	758	- 73	- 11	7
Chelsea	22	2530	1964	920	- 82	- 182	7
Birmingham	23	1167	947	886	- 68	- 96	7
Sheffield Wednesday	33	1271	1009	734	35	65	7
Oldham	35	638	368	412	- 93	- 28	6
Blackburn Rovers	35	586	421	465	- 85	7	6
Sheffield United	39	894	758	512	0	- 258	6
Bolton	42	803	510	566	41	- 317	5
Shrewsbury	44	353	251	315	- 31	- 55	5
Hull City	46	482	293	401	- 17	- 81	5
Bristol Rovers	47	523	394	310	- 64	- 126	5
Carlisle	49	436	308	342	- 48	- 13	5
Burnley	49	652	506	607	- 312	- 96	6
Plymouth	50	1026	411	425	- 38	18	5
Barnsley	50	445	387	391	- 2	- 82	5
Preston	54	565	331	395	- 101	- 34	4
Gillingham	55	494	473	336	- 4	- 153	4
Walsall	55	375	246	310	- 11	- 4	5
Huddersfield	56	519	380	438	- 91	24	5
Swindon	57	614	422	394	2	- 50	6
Rotherham	58	422	239	318	- 3	- 156	5
Cambridge	59	468	181	294	- 8	- 14	4
Wrexham	60	543	266	342	- 32	- 42	5
Reading	62	606	367	372	- 19	- 1	4
Brentford	63	592	351	343	17	- 32	4
Bury	66	325	184	301	- 57	- 19	4
Southend	66	436	245	258	- 43	- 70	4
Port Vale	67	425	186	322	9	9	4
Mansfield	68	352	216	317	- 59	- 67	4
Colchester	68	399	200	244	- 18	- 63	4
Lincoln	68	427	264	258	- 24	- 40	3
Peterborough	70	338	203	289	- 4	- 55	5
Northampton	77	317	181	255	- 47	- 38	3
Scunthorpe	81	348	201	259	- 13	- 67	4
Rochdale	86	205	152	171	- 24	- 126	3

Table 1b. Summary statistics by year (1985 Prices £000)

Year	Position	Turnover	Gate receipts	Wage bill	Net transfer spend	Pre-tax profits	Average weekly League gate (000)
1974	44	872	729	500	- 31	- 90	14.5
1975	43	796	665	480	0	- 87	15.1
1976	43	876	739	467	4	1	14.5
1977	41	912	751	454	20	31	15.1
1978	42	1075	862	508	53	33	15.0
1979	42	1059	834	564	103	- 41	14.1
1980	44	1062	815	621	80	9	13.7
1981	43	1075	804	657	61	- 114	12.4
1982	43	1064	779	686	104	- 274	11.6
1983	42	1090	817	641	30	- 54	10.8
1984	43	1206	924	639	25	9	10.9
1985	43	1265	965	684	21	- 17	10.8
1986	43	1169	913	700	55	- 123	10.2
1987	42	1314	1007	718	71	- 9	10.6
1988	43	1378	1053	782	23	6	10.6
1989	44	1562	1147	892	42	- 52	11.0

Net transfers means total expenditure minus total income from transfers. A negative figure means a net income for the club.

League position represents the outcome of sustained effort over an eight-month period, in which each club plays every other club in its division twice, without the random element introduced by the draw in a knock-out competition. In the empirical applications, in order to obtain linearity of the estimated relationships we used the negative of the log-odds (logit) of position: so quality is measured⁴ as $-\log(\text{position}/93 - \text{position})$.

In our model, clubs purchase players in a competitive labour market. In reality, prior to 1963 the clubs had an effective monopoly over the players on their own books, since a transfer required the consent of both clubs involved. Whilst a market for players existed in the sense that transfers did occur, the market was limited and likened to slavery (Sloane, 1969). Until 1977 the 'retain and transfer' system imposed some restrictions on the freedom of players to move between clubs but the market for players became much more active in the late 1960s and early 1970s. In 1977, the Professional Footballers' Association succeeded in establishing freedom of contract, and players are now almost completely free to accept the best contract on offer.⁵ Each year several hundred players move clubs, although only a small fraction are transferred for the large sums advertised in the newspapers. The market leads clubs to bid against each other competitively in order to secure the services of the best players. The effort and performance of players is easily observed and monitored, thus skill is readily identifiable.⁶

One important piece of evidence which we believe supports our assertion that the market is competitive is the fact that the performance and the wage bill of a club are very highly correlated. This suggests that the payment to teams is closely related to output. Figure 1 plots the average value of our measure of quality over

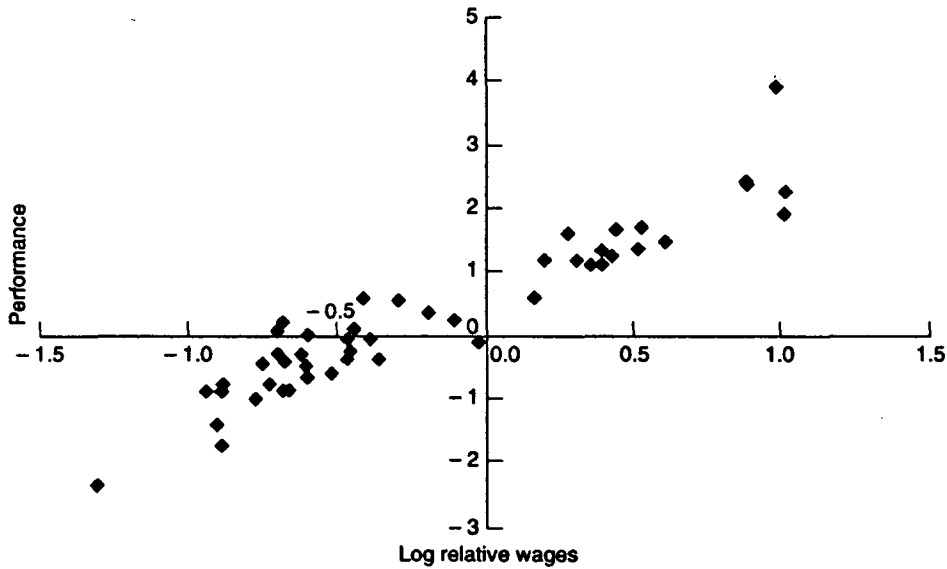


Fig. 1. Wages and performance: averages 1974-89.

1974-89 against the average of the logarithm of the wages paid by the club relative to average wages in the year over all clubs. The transformation seems to have induced linearity, the relationship is positive and fairly close.⁷ The most noticeable outlier is Liverpool, the highest performing club, which spends considerably less on wages than one would expect from the average relationship.⁸

Whilst there have been some papers on the operation of the English soccer labour market (e.g. Ruddock, 1979; Carmichael & Thomas, 1992) we know of no studies which have tested for the existence of a competitive labour market along the lines of US studies of the baseball labour market. This literature is surveyed in Cairns *et al.* (1986). These studies use data on individual salary levels and estimated marginal revenue product, which should be equal in a competitive market. They find that the system of labour contracts which prevailed prior to 1976 (which resembled English soccer's 'retain and transfer' system) was associated with salaries well below estimated marginal revenue product, but after 1976 when freedom of contract was established salaries rose considerably and were closer, if not identical to marginal revenue product. Separate tests for racial discrimination tended to suggest that major league baseball was not a perfectly competitive labour market. Replication of this type of work is not easy for soccer, or most team sports, since disentangling the contribution of the individual from the rest of the team is problematic, whilst in baseball the contribution of the individual in terms of hitting and pitching is much easier to identify. However, if the baseball literature is relevant at all, it does suggest that freedom of contract enables one to treat labour markets as approximately competitive, as we do here.

In our model the wage bill of the club is $W = wL$, where w is the quality adjusted wage rate and L the quantity of playing skills the club employs. There is a limited supply of playing skills and the labour market clears when total

expenditure of all clubs is equal to the total wage bill corresponding to this supply, giving a wage rate of

$$w_t = \frac{\sum_{i=1}^N W_{it}}{\sum_{i=1}^N L_{it}} \quad (1)$$

Thus, the expenditure of each club depends on the proportion of the total playing skill they demand and the total expenditure by all clubs. This formulation assumes the existence of a Nash equilibrium in wages, i.e. each club bids for players in such a way as to maximise its objective function (discussed below), assuming that all other clubs do likewise. Thus, the share of the available playing skill that a club commands is equal to its share of the total wage-bill of all clubs

$$\frac{L_{it}}{\sum_{i=1}^N L_{it}} = \frac{w_t L_{it}}{w_t \sum_{i=1}^N L_{it}} = \frac{W_{it}}{\sum_{i=1}^N W_{it}} \quad (2)$$

We will characterise the technology of the club by two equations, corresponding to a production function and a demand function, the parameters of which will, in general, differ between clubs. The production function describes how performance is produced by the quality adjusted inputs of skilled players, purchased on a competitive market, relative to the inputs of other teams. Clubs differ in their inherent performance potential as a result of such factors as history and intrinsic management differences.⁹ Relative inputs are measured by relative wages. In the empirical implementation we allow for variations in total expenditure on wages by all clubs, but treating this as fixed for the moment, we can write the production function in terms of a clubs' expenditure on wages

$$Q_{it} = c_i + dW_{it} \quad (3)$$

Each club faces a budget constraint, which generates profits as a function of the set of decision variables that we model

$$P_{it} = p_{it}G_{it} + O_{it} - T_{it} - N_{it} - W_{it} \quad (4)$$

where upper case P refers to profit and lower case p refers to ticket price, G to gate attendance, O is other income sources, such as TV and sponsorship (discussed further in Section 6), T is transfers and N is non-wage expenditure. Each club faces a demand function from fans which specifies how Gate attendances, for a given quality of team, depend on ticket prices

$$G_{it} = f(p_{it}, Q_{it}) \quad (5)$$

In order to estimate these demand functions we assume that they are separable and linear in prices and quality. They differ between clubs because of the size of their catchment area, local population density, traditions, etc.¹⁰ Each club is assumed to charge a price which maximises net gate revenue, so that price can be substituted out to give a reduced form function in which gate revenue depends on performance.¹¹ If, in addition, we assume that other revenue, from TV etc, also depends on performance, then total revenue becomes a function of the club specific intercept, the performance of the team, which depends on its quality, and chance, captured by a stochastic error term. This ignores general shifts in attendance and revenue which affect all clubs, for example the weather and the general decline in

attendances but we shall return to these secular trends in Section 6.¹² The systematic component of the revenue equation (conditional on gate prices) is thus

$$R_{it} = a_i + bQ_{it} \quad (6)$$

We shall assume that non-wage expenditures, N , and net transfers T are linked to wage costs

$$N_{it} + T_{it} = \tau W_{it} \quad (7)$$

As will be seen below, empirically this works quite well. Proper treatment of transfers raises inter-temporal aspects which are a matter for further research.¹³ Clubs acquire playing skill in one of two ways. Though recruitment and training programmes directed at the young people in their locality they develop their own talent. Secondly they purchase already developed talent through the transfer market. Overall this is a market for human capital. Some of this human capital is firm specific, paid for by the club, and so the club requires compensation when a player leaves to join another club. We can think of each club's investment programme generating a team with value (wage-cost) W^A . The optimisation, described below generates a desired wage-bill W^* . Transfers are determined as

$$T = v(W^* - W^A) \quad (8)$$

a linear function of wages as we assumed above. Clubs with a high desired wage bill, will buy players (i.e. $T > 0$) and clubs with a low desired wage bill will sell players ($T < 0$). This is broadly consistent with the data. Clubs lower down the League, whose training and recruitment may be very good, earn significant income by selling players to clubs at the top of the League whose demand for playing skills is so great that it exceeds their ability to generate them by their own local investment.

Profits are the difference between revenue and total expenditure, thus at any moment in time

$$P_i = R_i - (1 + \tau)W_i \quad (9)$$

Thus the profit-quality trade-off the owner faces is

$$P_i = \pi_i + \phi Q_i \quad (10)$$

where

$$\pi_i = (a_i d + (1 + \tau)c_i)/d$$

and

$$\phi = b - (1 + \tau)/d$$

In this simple model profits are just a linear function of expenditure. If they were a positive function, then the owner faces no conflict, he can have both more profits and more quality by spending more. We shall assume that this is not the case (the estimates below confirm this) and that the owner faces a trade-off, summarised by (10). All clubs face the same price, given by the slope, but differ in their endowment, determined by the intercept. Intercepts differ, because clubs differ in their inherent revenue potential and in their efficiency in converting player inputs into performance. In the empirical applications, this intercept will also be shifted over time by the total revenue and the total expenditure on wages of all clubs, and we shall assume an intertemporal objective function with quadratic costs of adjustment, leading to a partial adjustment model.

We assume that the owner cares about both quality and profits so that the owner's objective function can be written as

$$W = P_i - (\alpha/2) P_i^2 + \beta Q_i - (\gamma/2) Q_i^2 \quad (11)$$

The correct way to identify the objectives of a football club has always been a matter of some controversy. Since it is the owners who make all the key financial decisions we have focused on owners. Sloane (1971) discusses this issue in some detail and argues that the objective of the football club is essentially to achieve playing success whilst remaining solvent. One reason why profit maximisation on its own is not a very plausible objective to assume is that unlike most large corporations whose ownership is embodied in shares which are actively traded on a stock market, there is a very limited market for corporate control. Normally the market for corporate control imposes a constraint on those who control a company to maximise profits. The absence of a market in football club ownership is in striking contrast to the Football, Baseball and Hockey franchises of North America. English clubs are usually tightly controlled by a small number of Board directors (see for example, Jennett & Sloane, 1985, p. 47). However, whilst most football clubs repeatedly report financial losses, being subsidised by their owners; these owners must place some limit on the extent of the subsidy. Our assumption is consistent with the idea that owners have some limit in mind, but that this differs between club. Welfare is maximised by quality of

$$Q_i^* = (\beta + \phi(1 + \alpha\pi_i))/(\gamma + \alpha\phi^2) \quad (12)$$

and the team will make profits (losses) of

$$P_i^* = (\pi_i\gamma + \phi(1 + \phi\beta))/(\gamma + \alpha\phi^2) \quad (13)$$

The team is more likely to make a profit if the owner puts less weight on quality relative to profits or if the team has a higher inherent endowment.

Although each individual owner faces a negative trade-off between performance and profits, over teams we will observe a positive association. As the break-even performance increases, an owner would choose higher levels of both profits and performance (assuming they are both normal goods) tracing out the equivalent of an income expansion locus. This is shown in Fig. 2. The more similar are the tastes of different owners, then the closer will be the cross-section relation between profits and performance. Figure 3 plots the observed relationship over the sample between average performance and average profit margin (logarithm of the expenditure/cost ratio). As the theory suggests, the relationship is significantly positive although there is a considerable scatter, which would be consistent with preferences between performance and profits differing between owners.

3. Empirical Results

This section presents basic estimates for simple versions of the revenue, performance and total cost functions. Extended versions are considered in the next two sections, but the basic estimates are robust to these extensions.

Above, we assumed that adjustment of revenue to quality and of quality to position was instantaneous, for empirical implementation, we allow for partial adjustment, because both success and failure generate considerable persistence. The data were transformed to produce linearity and the transformed variables will

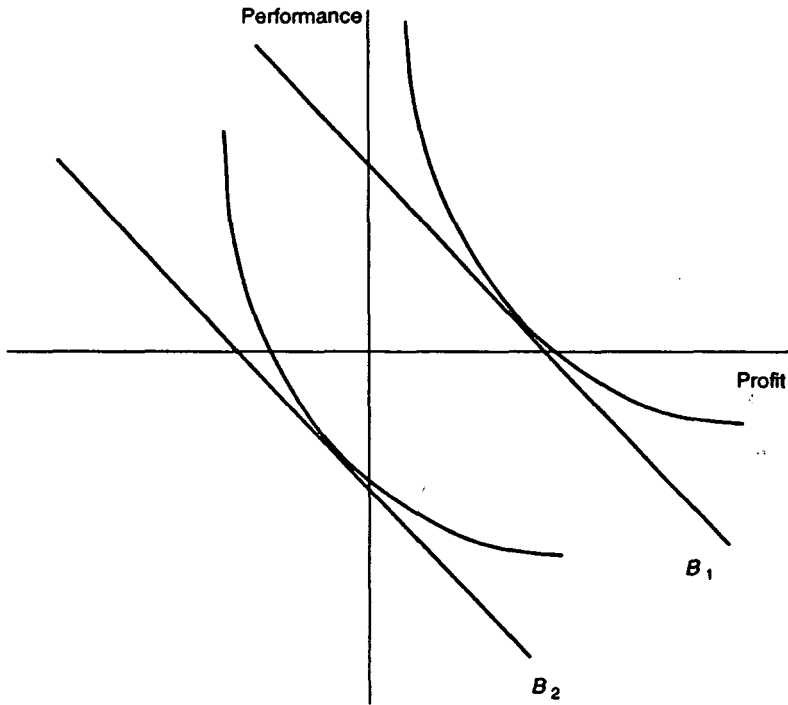


Fig. 2. The profit-performance trade-off for differently endowed clubs.

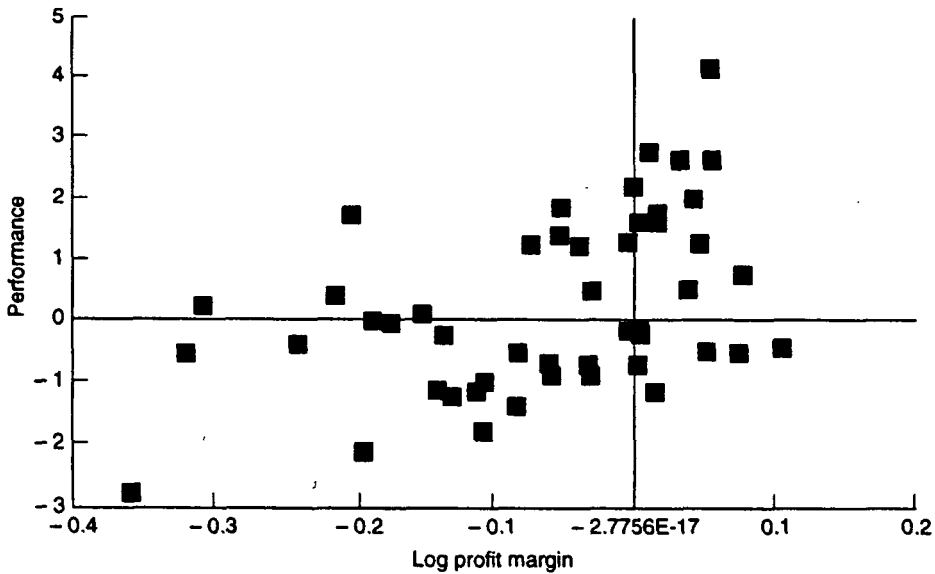


Fig. 3. Profits and performance. Averages 1974-89.

be denoted by lower case letters. As was explained in the last section, quality was measured by the negative of the log-odds of position in the league, N which can run from 1 (top) to 92 (bottom). Thus, in our empirical applications $q = -\ln\{N/(93 - N)\}$. Labour input, w , was measured as the deviation of the logarithms of expenditure on wages by the club from its mean for that year over all clubs. Revenue, r , was measured by the deviation of the logarithm of revenue from its mean over all clubs.¹⁴ These transformations induce a broadly linear relation between profit margin $\log(\text{revenue}/\text{costs})$ and performance (the logit of position), as can be seen from Fig. 3. They also imply a relationship between the absolute level of profits and position that fits the pattern observed. Clubs high in the League make profits, clubs in the middle make large absolute losses and clubs at the bottom of the League make small absolute losses.

The estimated model for the revenue function is

$$(r_{it} - \bar{r}_t) = \lambda a_i + \lambda b q_{it} + (1 - \lambda) (r_{it-1} - \bar{r}_{t-1}) + \varepsilon_{it} \quad (14)$$

and for the production function is

$$q_{it} = \mu c_i + \mu d(w_{it} - \bar{w}_t) + (1 - \mu)q_{it-1} + \varepsilon_{it-1} \quad (15)$$

Both functions were estimated by OLS allowing for club specific intercepts on pooled data (for the 48 clubs over 15 years (1974 was used for lags): 720 observations). Wu-Hausman tests (using lagged revenue, performance and wages as instruments) indicated that both performance and wages could be treated as exogenous ($t = 0.56$ and 1.75 respectively).

The estimate of the revenue function is

$$(r_{it} - \bar{r}_t) = \lambda a_i + 0.134 q_{it} + 0.466 (r_{it-1} - \bar{r}_{t-1}) \quad R^2 = 0.94$$

(0.012) (0.031)

and for the production function is

$$q_{it} = \mu c_i + 1.076 (w_{it} - \bar{w}_t) + 0.427 q_{it-1} \quad R^2 = 0.85$$

(0.120) (0.057)

Standard errors are given in parentheses. Clearly there is substantial persistence, both in performance and revenue. The implied estimates of the long run revenue and performance parameters (with standard errors in parentheses) are

$$b = 0.252 \quad d = 1.894$$

(0.019) (0.178)

In the logarithmic formulation, the profit margin of the club can be written

$$(R/C) = ac(1 + \tau)^{-1} Q^{(b-1/d)}$$

where a and c are the (antilogarithms) of the intercepts in the revenue and performance functions. The slope of the budget constraint is $b - 1/d = -0.28$, which is negative as assumed.

The 48 club specific intercepts in the performance function are of some interest. When transformed, they indicate the notional position the club would have had in the League had it spent the average amount on wages. If we imagine notional Divisions divided at average positions 22, 44 and 68; our sample of 48 clubs would be divided into 14 First, 7 Second, 21 Third, and 6 Fourth Division Clubs on the basis of their average position over the 16 years. If we examine the intercepts, the notional average position they would have had, were they to spend

average amounts on wages, we find only 1 First (Liverpool), 31 Second, 16 Third and no Fourth Division clubs. On the basis of the intercepts, all but five clubs would have average positions between 25 and 55. The two below 25 are Liverpool and Shrewsbury; the three above 55 are Scunthorpe, Rochdale and Burnley.

When estimated by OLS on the whole panel, the relation between the logarithms of total costs and wages is given by

$$c_{it} = 0.272 + 1.038w_{it} \quad R^2 = 0.833 \\ (0.099) \quad (0.017)$$

The hypothesis that total costs are proportional to wages, over all clubs, would just be rejected at the 5% level, $t = 2.27$, but would be accepted if the significance level is adjusted to take account of sample size in the way explained below. If club specific intercepts are included, the result is

$$c_{it} = \alpha_i + 0.922w_{it} \quad R^2 = 0.88 \\ (0.020)$$

which suggests that the club-specific intercepts in the total cost function are positively correlated with the level of wages a club chooses.

4. Econometric Issues

The results above are based on the assumption that the slopes are constant across clubs. If the slopes are not the same, then pooled estimators of dynamic models can give very misleading estimates, Pesaran & Smith (1995) discuss the issues. In dynamic heterogeneous models the appropriate estimator is the Swamy (1971) Random Coefficient Model (RCM), which estimates the weighted average of the 48 individual time-series regressions for each club. In fact the RCM estimates of the long-run coefficients are almost identical to the panel estimates in this case: $d = 1.86$ (0.465) and $b = 0.28$ (0.048).

In testing for coefficient equality in large samples, it is inappropriate to keep the size of the test or significance level constant at the conventional 5% level. As the sample increases one would wish to use the extra information to reduce both type one and type two errors rather than just type two as would be implied by using a constant size. A consistent way of doing this is provided by the Posterior Odds Criterion,¹⁵ see Klein & Brown (1984). The F statistics for coefficient equality (i.e. testing 48 separate regressions against the fixed effect pooled model) are 1.748 for the revenue equation and 1.586 for the performance equation; both with 94 and 576 degrees of freedom. The null would be rejected at the 1% level for the first and accepted for the second. However, if one uses the Posterior Odds Criterion, which allows for the effect of sample size on model selection, the pooled model is clearly chosen in both cases. Lagged exogenous variables were not significant either with the fixed effect or Swamy estimator.

The fixed effect estimator is inconsistent (N going to infinity fixed T) when lagged dependent variables are included. But the bias declines quite rapidly with T and the fact that the fixed effect and RCM estimates are so similar confirms that this is not a problem. Neither random-effect nor cross-section estimators would be appropriate in this case, since the effects are almost certainly correlated with the regressors: for instance the intercept in the quality equation will determine the optimal expenditure on wages which is the regressor in that equation.

Table 2. Demand equation. Dependent variable: log gate attendances (*A*)

Year effects	Excluded		Included	
	Coeff.	<i>t</i> -statistic	Coeff.	<i>t</i> -statistic
Price	- 0.38	- 8.05	- 0.19	- 2.01
<i>Q</i>	0.20	13.46	0.20	14.86
<i>A</i> (- 1)	0.50	13.24	0.45	11.84
<i>D</i> 2	0.10	2.19	0.09	1.96
<i>D</i> 3	0.18	2.86	0.15	2.48
<i>D</i> 4	0.41	4.96	0.38	4.88
<i>P</i> 1	0.27	3.50	0.30	4.20
<i>P</i> 2	0.30	4.97	0.33	5.83
<i>P</i> 3	0.01	0.17	0.03	0.79
<i>R</i> 1	- 0.37	- 4.58	- 0.40	- 5.29
<i>R</i> 2	- 0.11	- 1.98	- 0.13	- 2.43
<i>R</i> 3	- 0.11	- 2.62	- 0.11	- 2.95
<i>CUP</i>	0.01	2.55	0.01	3.61
<i>R</i> ²	0.95		0.96	

5. Further Issues

In this section, we present estimates of the demand function (Table 2) for attendance and an extended revenue function (Table 3) which allows for a number of other influences. The estimates above assume a homogenous competition, in fact League competitions are divided into four Divisions.¹⁶ Each year some clubs in each division get relegated to the division below, and some clubs of the lower divisions get promoted to the division above, the number involved has varied over the sample. Thus, the division a club belongs to is a non-linear function of lagged position. If revenue depended on which division the club played in rather than position as such, this non-linearity would bias the results. To test for the effect of being in a particular division, three Division dummy variables, *D*2, *D*3 and *D*4, were added for being in divisions 2, 3 and 4 respectively. To test for the effect of promotion or relegation the changes in the Division dummies were also added: *P*1, *P*2, *P*3 and *R*1, *R*2, *R*3, where the former take a value of 1 if the club was promoted to the relevant division and 0 otherwise, and the latter take a value of 1 if the club was relegated from the relevant division and zero otherwise.

Revenue also depends on performance in Cup competitions since this generates extra games as well as (more speculatively) increased support for the club as a whole. We collected data on the number of rounds played in each year by each club in each of the FA Cup, The League Cup (variously known as the Milk Cup, the Rumbelows Cup and now the Coca-Cola Cup), and the various European club competitions, the European Cup, The UEFA Cup and the European Cup Winners' Cup. The variable *CUP* is the total number of games played in any of these cups. When entered separately, their effects were not significantly different, either in the demand function or the revenue function.

The demand function explains the logarithm of average annual gate attendance by the logarithm of price (gate revenue per person, deflated by the RPI). Our

data on gate attendances refer to League games whilst gate revenue covers all games played. In order to adjust for this in our price series we define this as gate revenue divided by average gate times the number of league games plus half the number of cup games all divided by the RPI.¹⁷ Gate revenue was not available for all clubs, so that for this exercise our sample was reduced to 659 observations. Also included were performance, lagged attendance, the divisional, promotion and relegation dummies and the number of Cup games played. Club specific intercepts were included and the equation was estimated with and without year specific intercepts. In our study we have not attempted to capture the effect of uncertainty of outcome, which has been shown to affect demand (see for example Peel & Thomas, 1988) since our data are annual rather than match by match. The club specific intercepts capture effects such as local population size, age and gender structure and wealth, which will influence demand for football. Given the severe difficulties in defining these variables properly,¹⁸ we do not believe that using club specific intercepts is in practice likely to involve much loss of accuracy.

The implied long run price elasticity from the equation without time dummies for each year is -0.76 but this falls to -0.34 when time dummies are allowed (with a t -statistic of -2.05). This suggests that the price elasticity picks up the effect of the secular decline in the demand for football unless time dummies are included. The finding that demand is price inelastic is in line with earlier studies looking at the demand for football over the longer term.¹⁹ If we suppose that football clubs are local monopolists this implies that there is considerable scope to increase profits through higher prices. However, we believe there is reason to be sceptical about this conclusion. There are (at least) three possible approaches to interpreting these estimates.

- (i) Misspecification induced by aggregating gate attendances. Suppose clubs set prices in such a way as to ensure that the ground is full for every game, taking into account the different attractiveness of each fixture. This would be consistent with a high elasticity of demand being rationally exploited by a price discriminating monopolist but would not show up in our estimates.
- (ii) Misspecification of the production function: if attendances are an input into the performance of a team, then this consideration will feed through the objective function of the owner into the price setting decision, making it rational to charge low prices even though a pure profit maximiser might set higher prices.
- (iii) Owners may not be monopolists at all. Even if other clubs are not good substitutes for the local product, then other leisure activities may be. If clubs are competing against these activities then their capacity to set higher prices may be limited.

It should be noted that the crucial parameter in our argument, the slope of the profit position trade-off does not depend on this elasticity. All the promotion and relegation dummies are significant except for promotion from the Fourth to the Third Division. They suggest that there is a one-off boost to revenue on promotion to the first and second divisions, which disappears in subsequent years, and also that there is a one-off dip in revenues after relegation from the first, second or third divisions which also disappears in later years. These seem to be picking factors such as supporter morale. Promotion generates an enthusiasm which is not sustained in later years, whilst relegation creates an initial despair and loss of support which is partly recaptured once the initial depression has worn off.

Table 3. Revenue equation

Variable	Coefficient	Absolute <i>t</i> -statistic
<i>Q</i>	0.11	7.76
<i>R</i> (-1)	0.46	14
<i>D2</i>	0.00	0.03
<i>D3</i>	0.02	0.41
<i>D4</i>	0.09	1.24
<i>P1</i>	0.21	2.82
<i>P2</i>	0.25	4.10
<i>P3</i>	0.06	1.47
<i>R1</i>	-0.42	-5.39
<i>R2</i>	-0.06	-1.14
<i>R3</i>	-0.08	-2.05
<i>CUP</i>	0.03	9.25
$R^2 = 0.95$		

The significant positive values of the lower division dummies, do not indicate that members of the lower divisions get larger attendances than first division clubs, but that they get more than would be expected from extrapolating the relationships between attendance and performance, price, etc of a first division club. This might reflect the possibility that first division clubs are capacity constrained.

As might be expected, cup performances increase gate attendances significantly, although it is perhaps surprising that the restriction that all cups increase attendances equiproportionately is accepted by the data. For the European club championships there are only a very small number of observations in our sample, but it is striking that the League Cup is apparently as popular as the FA Cup, since this competition was often criticised as being superfluous and uninteresting.

The revenue function was re-estimated including the additional variables (except relative price which was not significant in any case) and the results are given in Table 3. The long-run effect of performance on revenue is 0.21 (0.03) which is not significantly different from the earlier estimates without the additional variables. The divisional dummies are now insignificant. This means that for a given position in the League, being in a particular division exerts no additional effect on revenue, the positive relative effects of being in a lower division on attendance being offset by the negative effects on other revenues. By contrast the promotion dummies are significant for the first and second divisions and the relegation dummy is significant for the first and third divisions.

6. Market Structure

Conditional on the club-specific intercepts and aggregate wages and revenues, the model described above explains the main features of the market: only the top clubs make profits and there is a high degree of concentration. The top five clubs in terms of average League position account for some 20% of all income over our sample period. Were club owners profit maximisers, the loss makers would be

likely to exit, raising the profitability of those remaining and bringing the industry back toward an equilibrium where firms generate non-negative profits. But given the owners' objective functions this does not happen, and given the lack of a market in corporate control they are not constrained by threat of take-over. Jennett & Sloane (1985) also argue that the wider shareholding, which would be associated with a market for corporate control, would provide a source of capital for the improvement of facilities.

This leads one to ask if the observed market structure is efficient either in terms of the general interest or in terms of the joint interest of the owners. Football has aspects of a quasi-public good in that it provides entertainment to a fairly wide public of fans. This public good is produced as a joint product, with a private good, the personal interests of the club owner. The interests of the owner may differ from the general interest, particularly in relation to the setting of ticket prices. Clubs are generally thought of as local monopolists and therefore pricing is likely to involve a dead-weight loss. Perhaps more interesting than this fairly standard economic problem is the issue of whether the manner of competition between club owners promotes either the joint interests of the owners or the broader general interest.

Competition between clubs on ticket prices is not likely to be a general phenomenon. Competition takes place for position in the League, and the medium for this competition is the market for playing talents.²⁰ Clearly the Nash equilibrium will not be the joint profit maximising one, since increased competition is a negative sum game. Competing for position itself is a zero sum game, but an owner who wishes to improve position, spends money on players. But spending on players puts upward pressure on the quality adjusted wage rate, increases the rent extracted by players and puts downward pressure on profits. The removal of what was effectively an indenture system in the early 1960s has provoked a spiral in players' wages which can be seen in Table 1b.²¹ The clubs themselves are well aware of this phenomenon but the scope for outright collusion is (now) limited by the law while the possibility of sustaining tacit collusion is limited by the improbability of credible punishment strategies.²²

A consequence of this 'excessive' competition for players is that clubs have fewer resources available for the purchase of other inputs required to produce services consumed jointly with attendance at a football match: a safe, comfortable environment, refreshments and so on. During the 1980s a series of tragic accidents in football stadia highlighted the problem: English grounds were antiquated and dangerous, attractive to few other than hooligans, while the clubs had neither the means nor the will to change matters. Given that competition for players was also competing away all economic profit, the clubs had no incentive to improve the football environment since any extra profits generated would pass immediately to the players. Thus, not only was the structure of competition against the joint interest of the owners, but it also acted against the general interest.²³

This story characterises English football up to 1989, the end of the period covered by our data set. Since then, however, two critical developments have helped to transform the picture. First, the implementation of the Taylor Report commissioned by the government after the Hillsborough disaster has obliged clubs to invest in all-seater stadia which has had the side-effect of modernising the facilities to make them more attractive to spectators. By obliging clubs to do something collectively, which it would not have been individually rational to do (for standard prisoner's dilemma reasons) the government has helped to push the

clubs towards an alternative to the low quality equilibrium in which they were caught.²⁴

Secondly, increasing competition among the television companies has increased the available funds for football clubs. In particular, competition from satellite broadcasting upset the pre-existing terrestrial 'duopsony', and brought about a tenfold increase in broadcast revenues. The most notable effect of this change was to cause the leading clubs to break away from the rest of the Football League to form the Premier League and so keep almost all the broadcast revenues for themselves. However, the Premier League has so far been little different from the old First Division. In the short term the extra funding has helped to pay for the rebuilding programme for the larger clubs. In the long term it has created a powerful vested interest, not in the performance of individual clubs but in the attractiveness of football as a whole. This may act as a focus for coordination between clubs in the future.

Both of these developments have had the effect of providing a focus on the interests of the entire sport and so have helped to deal with the prisoner's dilemma element in the owner's problem. Fortuitously, these developments seem to have coincided with a more general increase of interest in football. The post-war trough in League attendances occurred in the season 1985/6. Since then there has been a steady increase in attendances despite rising real ticket prices. This may be connected with increased effectiveness of policing at Football grounds leading to a lower incidence of hooliganism.²⁵ Overall, the future of English football seems brighter than at any other time since the late 1940s.

7. Conclusions

Over the period covered by our data, English League Football is an interesting example of a mature industry with a set of very characteristic problems. Like many old industries it had to deal with declining demand and competition from more technologically advanced leisure products. Most firms made losses, its plant was antiquated and grossly under-utilised and heavy investment was required both to meet government mandated safety standards and to improve the quality of the product sufficiently to compete on modern markets. Firms in this industry have little control over their main input cost, players, which are traded on a competitive market. Of course, many mature industries are very profitable.

The interesting question is why market forces so often fail to restructure such industries. In this case, and perhaps others, the answer seems to lie in the failure of the market for corporate control. Obstacles to take-over and acquisition allow current owners to follow non-profit objectives, allow small groups to veto changes which might be in the general interest and inhibit the internalisation of the externalities inherent in adjustment. Only when external forces impose a coordinated strategy on the competing firms is there any prospect of change. In the case of English Football, the Taylor Report and the TV companies have provided a focus for coordination.

Our model has a number of potential applications. It can be used to derive the required expenditure to raise a club from a low position to the top of the League. It is also possible to analyse the effect of the new League structure in England and Wales, which combines a Premier League of top clubs (organised by the FA) and a rump of Football League clubs. Finally, the model can be used to analyse the impact of other reforms which may be proposed for the organisation of English

football in the future, and to understand their chances of success by understanding the likely impact on the distribution of profits throughout the industry.

Notes

We are grateful to Sean Wilkins for excellent research assistance in collecting the data and to two anonymous referees for constructive comments. The paper was written while the second author was visiting the Centre for Economic Forecasting, London Business School.

1. In fact the organisational structure of studying English League football is relatively easy to study since all clubs are independent limited companies whose annual accounts are lodged at Companies House. To the best of our knowledge most economic research on football emanates from the UK, although there are some exceptions such as Gartner & Pommerehne (1978), Janssens & Kesenne (1987).
2. We obtained all the accounting lodged at Companies House for the 92 clubs in the English League (in 1990). Since the accounts of some clubs were incomplete or did not break down the data in sufficient detail we focused on the 48 with complete records for the period 1974–89. To check the representativeness of our sample we calculated its average position in each sample year. The highest was 41st and the lowest 44th (against a population average of 46th). Whilst there is some bias toward the top of the League, we do not believe this is significant.
3. Until 1992, these were the First, Second, Third and Fourth divisions. Since the start of the 1992/93 season, they are the Premier, First, Second and Third. The old first division left the Football League, to form the Premier Division, organised by the Football Association, mainly in order to avoid having to share TV revenue with lower division clubs.
4. One advantage of this functional form is that it imposes reflecting barriers at positions 1 and 92. This is an approximation at the lower barrier since there have been a small number of exits over our sample period. Treating the ranking as a continuous variable also introduces a small approximation.
5. After our estimation period the European Court judgment on the Bosman case further changed the position.
6. This means that the 'agency' problem common to many markets for highly skilled labour is not as pressing in the market for football players. An alternative view of the market for players is provided by Carmichael & Thomas (1992) who present a bargaining model of transfers. In their approach the transfer price of each player is the outcome of a bargain between each club, the price being determined by the threat points and bargaining strength of each side. Where buyers and sellers have market power, transfers will not reflect the value of the individual's contribution to the team. Unfortunately measuring these individual contributions is problematic in football (unlike, for example, baseball, see Cairns *et al.*, 1986) and the two approaches are difficult to distinguish empirically.
7. In particular, the fit is much better than comparable regressions of measures of corporate performance on payments to top executives, e.g. Smith & Szymanski (1995).
8. Another outlier of the same sort is Nottingham Forest whose accounts are only available from 1982 onwards and so are not included in the data set. These cases, which involve non-appropriable sources of competitive advantage, are discussed in more detail in dell'Osso & Szymanski (1991).
9. Strictly, this is an ownership rather than management difference, since managers can be hired on a competitive market and their wages will be included in the total payment for skill. The intercept captures a separate club specific effect partly associated with the tendency of the owners to interfere. For instance, Brian Clough, probably the ablest English manager of his generation was sacked by Brighton and Leeds partly because the owners could not get on with him.
10. Hart *et al.* (1975), Smart & Goddard (1991), Cairns (1987), Bird (1982), Jennett (1984), Walker (1986) and Peel & Thomas (1988), present estimates of football demand functions. Other studies have included variables such as population within a local catchment area, seats as a proportion of ground capacity and local unemployment rates. In our analysis these effects are picked up by year and club-specific intercepts.
11. This follows conveniently from our separability assumption as well as the fact that we ignore possible feedback effects from attendances to performance quality. Were such mechanisms allowed then there could be a trade-off between ticket prices and performance. Whilst our estimates reported below lead us to believe that any such effects are likely to be small in practice, we discuss their possible impact in Section 5.

12. For empirical implementation we shall proxy these by average revenue over all clubs, which will be added to the equation.
13. See also Carmichael & Thomas (1992).
14. Using deviations from year means is equivalent to including year specific intercepts, given that q also has mean zero.
15. This chooses the model which has the higher value of a criterion which is calculated as the maximised Log-Likelihood less the number of parameters times half the log of the sample size.
16. Up to 1987, the first two divisions had 22 clubs each, while the third and fourth divisions had 24. In 1988, the numbers were 21, 23, 24, 24; in 1989, 20 in the first, 24 in the other three.
17. This approximation is flawed in four possible ways. First, it assumes that half of all cup games are played at home, whilst the actual proportion depends on the draw. However, over time this should not induce any systematic bias in the estimates. Secondly it ignores replays which will lead to slight overestimate of prices. Thirdly it assumes that gate attendances at cup games equal attendances at League games on average. Whilst attendances at particular cup fixtures may be much larger than average, we do not believe that this bias will be very great on average (one Arsenal fan told us that the largest gates of the season are for games against the other leading clubs: Manchester United, Liverpool, Tottenham, regardless of whether they are cup or league matches). Fourthly the estimate does not allow for the possibility that clubs raise prices for cup games. In fact the brisk business of ticket touts at popular games suggest that clubs are not varying their prices by enough to account for the differing attractiveness of different games.
18. For example, it is hard to choose the precise catchment area for a club and this is certainly not likely to be stable over time given factors such as increased car ownership. Wealth effects are likely to be important, but it is again hard to define the wealth of a potential catchment area precisely.
19. Bird (1982) estimates a price elasticity of -0.2 using data for 1948–80. A recent study by Dobson & Goddard (1996) using data for 1925 to 1992 estimates an elasticity of -0.015 . Our much higher estimate probably arises from the fact that the period covered by our data set was one of rapidly increasing real prices and, generally, falling attendances. Earlier periods had relatively stable prices and attendances (see for example Dobson & Goddard, 1996, Figs 1, 2 and 3).
20. The most famous recent example is Jack Walker, the owner of Blackburn Rovers, who raised the club from the bottom of the (old) second division to Premier League Champions by spending about £50m over a five year period. However it must be admitted that attempts by sugar daddies to pull off similar feats at other clubs have often failed.
21. The average annual increase over the period was nearly 4% in real terms, which is a very high rate to sustain over such an extended period.
22. Tacit collusion would involve an unwritten agreement not to pay too much to players combined with the threat of some sort of punishment for clubs which cheated on the arrangement. The punishment would need to raise the costs of the cheater's strategy: it is not clear how this could work in practice.
23. More critical fans often blame the quality of the management that runs a club for failure to innovate, but a zero profit industry is unlikely to attract the most talented (commercial) managers.
24. The government has also helped to fund these changes, but most of the burden is borne by the clubs.
25. In and around the grounds at least.

References

- Arnold, A. J. & Beneviste, I. (1987) Wealth and poverty in the English football league, *Accounting and Business Research*, 17, pp. 195–203.
- Bird, P. J. (1982) The demand for league football, *Applied Economics*, 14, pp. 637–649.
- Cairns, J. A. (1987) Evaluating changes in league structure: the reorganisation of the Scottish Football League, *Applied Economics*, 19, pp. 259–275.
- Cairns, J., Jennett, N. & Sloane, P. J. (1986) The economics of professional team sports: a survey of theory and evidence, *Journal of Economic Studies*, 13, pp. 1–80.
- Carmichael, F. & Thomas, D. (1992) Bargaining in the transfer market: theory and evidence, mimeo, University of Aberystwyth.
- dell'Osso, F. & Szymanski, S. (1991) Who are the champions? An analysis of football architecture, *Business Strategy Review*, Summer 1991, pp. 113–130.
- Dobson, S. M. & Goddard, J. A. (1996) The demand for professional league football in England and Wales, 1925–1992, *Journal of the Royal Statistical Society*, forthcoming.

- Gartner, M. & Pommerehne, W. (1978) Der Fussballzuschauer—ein Homo Oeconomicus? *Jahrbuch für Sozial Wissenschaft*, 29, pp. 88–107.
- Hart, R. A., Hutton, J. & Sharot, T. (1975) A statistical analysis of association football attendances, *Applied Statistics*, 24, pp. 17–27.
- Janssens, P. & Kesenne, S. (1987) Belgian soccer attendances, *Tijdschrift voor Economie en Management*, 32, pp. 305–315.
- Jennett, N. (1984) Attendance, uncertainty of outcome and policy in Scottish League Football, *Scottish Journal of Political Economy*, 31, pp. 176–198.
- Jennett, N. & Sloane, P. J. (1985) The future of League football: a critique of the report of the Chester Committee of Enquiry, *Leisure Studies*, 4, pp. 39–56.
- Klein, R. W. & Brown, S. J. (1984) Model selection when there is 'minimal' prior information, *Econometrica*, 52, pp. 1291–1312.
- Peel, D. & Thomas, D. (1988) Outcome uncertainty and the demand for football, *Scottish Journal of Political Economy*, 35, pp. 242–249.
- Pesaran, M. H. & Smith, R. (1995) Estimating long-run relationships from dynamic heterogenous panels, *Journal of Econometrics*, 68, pp. 79–113.
- Ruddock, L. (1979) The market for professional footballers: an economic analysis, *Economics*, 15, pp. 70–72.
- Sloane, P. J. (1969) The labour market in professional football, *British Journal of Industrial Relations*, 7, pp. 181–199.
- Sloane, P. J. (1971) The economics of professional football: the football club as a utility maximiser, *Scottish Journal of Political Economy*, 8, pp. 121–146.
- Smart, R. A. & Goddard, J. A. (1991) The determinants of standing and seated football attendances: evidence from three Scottish League clubs, *Quarterly Economic Commentary*, 16, pp. 61–64.
- Smith, R. & Szymanski, S. (1995) Executive pay and performance, the empirical importance of the participation constraint, *International Journal of the Economics of Business*, 2, pp. 485–495.
- Swamy, P. A. V. (1971) *Statistical Inference in Random Coefficient Regression Models* (Berlin, Springer-Verlag).
- Taylor, Rt Hon Lord Justice (1989) *The Hillsborough Stadium Disaster* (HMSO) Cm962.
- Walker, B. (1986) The demand for professional League Football and the success of football league teams: some city size effects, *Urban Studies*, 23, pp. 209–220.