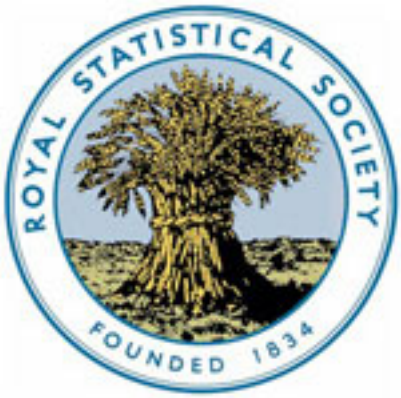

Appl. Statist. (1975),
24, *No.* 1, p. 17

A Statistical Analysis of Association Football Attendances

By R. A. HART J. HUTTON and T. SHAROT

University of Leeds *University of Aberdeen*

[Received April 1973. Final revision January 1974]

SUMMARY

Little quantitative work has so far appeared in the growing literature on professional sports. In this paper a model is constructed and estimated for the attendances of four English First Division clubs. The possibility of a common law-like relationship is examined and account is taken of common cross-sectional influences in order to raise the efficiency of the estimates.

Keywords: FOOTBALL ATTENDANCE: REGRESSION MODEL; SEEMINGLY UNRELATED LEAST SQUARES

1. INTRODUCTION

ALTHOUGH watching professional sports occupies an important position in the leisure-time activities of a large proportion of many nations' populations, it has received little serious attention from social scientists. In this paper we construct and estimate a model of the demand for the major British professional sport, Association Football.

The dependent variable in the model is the Saturday afternoon attendance record for four English First Division clubs, on their home ground. Three seasons (1969/70, 1970/71 and 1971/72) were pooled for each time series, but the parameters of the model were estimated for each of the clubs separately; a common law-like relationship is seen to be not altogether satisfactory.

The independent variables are what might be considered to be the major determinants of attendance and may be roughly divided into three groups.

The first are those of an economic nature, notably the entrance fee at the ground, the cost of alternative entertainment and the level of private incomes. The distinguishing feature of sporting industries, as shown by Neale (1964) and Sloane (1971), is that although clubs are competitive in the sporting sense, they co-operate closely in economic matters. For this reason the gate price is only considered important in relation to the prices of other leisure industries. During the sample period there was one small price rise (5p in August 1970). How this compares with price movements in competing industries, and with the rate of increase in money incomes is not the concern of a short-run model such as the one developed here. A simple exogenous trend was used to take account of any systematic movement over our period of analysis, which may have arisen owing to the above factors or other possible factors such as the apparent slow long-term decline from the euphoria created by England's World Cup success in 1966.

The second type of determinants are those of a demographic or geographic nature. Attendance figures clearly depend on the number of supporters each of the teams can claim, and this depends in turn on the population of each team's catchment area, the

local level of "enthusiasm" and the proximity of other clubs. Thus the estimation of the number of supporters behind a team is difficult and a certain amount of subjective judgment is necessary. The procedure used here is described in Section 3. Another influence is the distance which the away team's supporters have to travel. There are two separate disincentives at work here, one economic and the other psychological. We used road mileage as a proxy for both effects since it is easily measured and should be highly correlated with both cost and travelling time; refinements are of course possible. The significance of distance varies dramatically between the four clubs and it is the major component of the difference in structure referred to above.

The last group of determinants are those relating to the attractiveness of a match and of its rival attractions. The first we have proxied by the quality of the two teams, as measured by their League positions prior to the match. Possible alternative measures are the average number of points scored per match in the current season, or the total number of goals scored. However, none of these pick up key matches, such as "relegation battles", and an attempt was made to include these in the model on an *ad hoc* basis. Again, it is possible that the performance of teams in their most recent matches affects current attendance, and this hypothesis was also tested. The problem of rival football matches is diminished by the fact that towns with two or more football teams would be expected to have few floating supporters in comparison to the numbers which are faithful to one particular club and by the fact that fixture clashes are intentionally minimised by the Football League. Other attractions we have treated as random disturbances, and accordingly we have used an estimation technique to try to detect such disturbances as are common, or nearly so, for all four clubs; we cite, for example, the weather, Rugby Internationals on the television and seasonality.

2. The Model

The following notation is employed in subsequent sections:

A_i	Total attendance at match *i*.
D_i	Distance travelled by away support.
$P_i^{\mathrm{H}}, P_i^{\mathrm{A}}$	Population of home and away teams' catchment area, respectively.
X_i, Y_i	League position of the home and away team prior to the match, respectively.
T_i	Trend; $= 1$ for the first match of 1969/70, with unit weekly increments thereafter.

Also let $f^{\mathrm{H}}(X, Y)$ and $f^{\mathrm{A}}(X, Y)$ be functions of X and Y which represent the "attractiveness" of the match to home and away supporters, respectively.

Our model can then be written

$$A_i = \{\alpha_1 P_i^{\mathrm{H}} f^{\mathrm{H}}(X_i, Y_i) + \alpha_2 P_i^{\mathrm{A}} D_i^{-\beta} f^{\mathrm{A}}(X_i, Y_i)\} T_i^{\gamma}. \tag{1}$$

The underlying assumptions are as follows.

Firstly, total attendance is assumed to be the sum of home support and away support, which are not jointly dependent except through X_i and Y_i. Secondly, the two factions are assumed to be linear in their respective populations. Ignoring non-local support, that is, considering P_i^{H} and P_i^{A} to be the populations of the two towns, this is tantamount to assuming that "enthusiasm" is not significantly correlated with town size or population density. Since large cities tend, for financial or other reasons,

to have the better teams, it is possible that some have built up a tradition of high enthusiasm which will appear in their attendance record even when adjusted for League position. In addition, those urban areas with two or more nearby clubs, such as Manchester, Birmingham and London, may experience higher enthusiasm due to the local rivalry. Conversely, sympathy for the underdog, or parochialism, may favour enthusiasm in smaller towns with poorer teams. The point is difficult to test quantitatively, especially without knowledge of the proportion of a crowd from each of the two towns. Similarly, little can be said about newly signed players and changes of manager which may have a substantial short-run bearing on the "glamour" and thus the attendance of a given club. Thirdly, it seems reasonable to assume that

$$f^{H}(X, Y) = f^{A}(Y, X), \tag{2}$$

which can be interpreted to mean that the relative response of supporters to their own and to the opposing team's League position is independent of whether the match is at home or away and further assumptions should not contradict this. Unless such complications as the "relegation battle" are thought important, it is also in order to ignore interaction effects, so that f^{H} and f^{A} become separable in their arguments, either additively:

$$f^{H}(X, Y) = f(X) + g(Y) \tag{3}$$

or multiplicatively:

$$f^{H}(X, Y) = f(X) \cdot g(Y) \tag{4}$$

and similarly for f^{A}.

It is necessary to be still more precise about these functions for empirical tests to be possible, but largely from ignorance we have taken a heuristic approach in first specifying the test procedure and then trying functional forms compatible with both it and the assumptions here.

One final point is that there is an upper limit on attendances, imposed by ground capacity. A few of our matches attracted near capacity crowds. We have not explicitly allowed for this, for it is not a major problem.

3. Data

The clubs chosen were Leeds United, Newcastle United, Nottingham Forest and Southampton. The principal reason for this choice was that each club is relatively isolated from other First Division clubs and thus complications in the definition of the home catchment area are reduced. In addition, the success record of these clubs has been very different in recent years, and their geographical situations are diverse. The four data series therefore contain as little duplication of information as possible.

Data were obtained from the following sources:

A_i *Rothman's Football Yearbooks* and club programmes.

D_i *A.A. Member's Handbook.*

X_i and Y_i *The Guardian* Matches at the beginning of the season for which either X or Y are not available were discounted.

P_i *Census Report* 1966 (H.M.S.O.) together with Boundaries Commission for England, *First Periodical Report* 1953–54 (Cmnd. 9311).

The basic procedure adopted in defining P_i was to take the total male population of the urban parliamentary constituencies surrounding the club grounds. In the case

of most clubs the total population thus defined was used for P_i. In London and Birmingham, constituencies were allocated to clubs according to geographical proximity. Occasional allowance was made for clubs in lower Divisions. For instance, the population of the Birmingham area was divided between Aston Villa, West Bromwich Albion, Birmingham City and Wolverhampton F.C.'s. Similarly, some allowance was made for the existence of Notts County which, although no longer in the First Division, would be expected to have retained a significant number of Nottingham's football supporters. No allowance was made for non-League football, Rugby or other attractions.

These measures are necessarily arbitrary. For instance, no attempt was made to allow for ease of communication in defining the home catchment area, nor was any allowance made for possible variations in the enthusiasm of support in areas of equal population.† In order to take account of these factors a detailed study of the social and geographic characteristics of each area would be required.

4. Estimation

Multiple linear regression is an appropriate method of estimation, but this requires that the model be transformed so that it is linear in its parameters. Initial computer runs with various forms suggested that the data were fitted best by a log-linear relationship and this is convenient because the parameters are in this case free of the scale of the data, facilitating meaningful comparison both within and between equations. It also follows that equation (4) is preferable to equation (3).

Since the four teams are estimated separately, P^{H} is effectively constant and will henceforth be absorbed into α_1 and for clarity P will be written for P^{A}. Thus

$$A = \{\alpha_1 f(X)g(Y) + \alpha_2 P D^{-\beta} f(Y)g(X)\} T^{\gamma}. \tag{5}$$

Expanding each function as a Taylor series about the mean and retaining first order terms only gives

$$A + \delta A = \left\{\alpha_1 f(X)g(Y)\left(1 + \frac{f'}{f}\delta X + \frac{g'}{g}\delta Y\right)\right.$$
$$\left. + \alpha_2 P D^{-\beta} f(Y)g(X)\left(1 + \frac{\delta P}{P} - \beta\frac{\delta D}{D} + \frac{f'}{f}\delta Y + \frac{g'}{g}\delta X\right)\right\} T^{\gamma} + A\gamma\frac{\delta T}{T}, \tag{6}$$

where f, f', g and g' are evaluated at the appropriate points. It follows that

$$\frac{\delta A}{A} = a_1\left(\frac{f'}{f}\delta X + \frac{g'}{g}\delta Y\right) + (1 - a_1)\left(\frac{\delta P}{P} - \beta\frac{\delta D}{D} + \frac{f'}{f}\delta Y + \frac{g'}{g}\delta X\right) + \gamma\frac{\delta T}{T}, \tag{7}$$

where

$$a_1 = \frac{\alpha_1 f(X)g(Y)T^{\gamma}}{A} \tag{8}$$

is the proportion of the total attendance which is home support. The obvious choices for f and g are

$$f(u) = u^{\theta} \tag{9}$$

† The Chester Report (1968) asserts without supporting evidence, "Some areas are more passionate in support than others in the sense that an above average percentage of the population turn out" (Para. 144).

and

$$g(u) = u^{\phi}, \tag{10}$$

where

$$u = X \quad \text{or} \quad Y \tag{11}$$

and to try alternative forms if these fail. Then

$$\frac{\delta A}{A} = a_1\left(\theta\frac{\delta X}{X}+\phi\frac{\delta Y}{Y}\right)+(1-a_1)\left(\frac{\delta P}{P}-\beta\frac{\delta D}{D}+\theta\frac{\delta Y}{Y}+\phi\frac{\delta X}{X}\right)+\gamma\frac{\delta T}{T}. \tag{12}$$

But

$$\frac{\delta A}{A} = \delta(\log A) \tag{13}$$

and so on, so the appropriate regression equation is, including the stochastic error term,

$$\log A_i = b_0+b_1\log X_i+b_2\log Y_i+b_3\log P_i+b_4\log D_i+b_5\log T_i+u_i, \tag{14}$$

where

$$b_1 = a_1\theta+(1-a_1)\phi, \tag{15}$$

$$b_2 = a_1\phi+(1-a_1)\theta, \tag{16}$$

$$b_3 = 1-a_1, \tag{17}$$

$$b_4 = -(1-a_1)\beta, \tag{18}$$

$$b_5 = \gamma. \tag{19}$$

It will be seen that the restrictions which have been imposed are sufficient to identify the parameters of the model from the regression coefficients.

On an *ad hoc* basis, we have reason to believe that the above choice of f and g is sensible. A unit difference in League position near the top of the Division (for example, between 2nd and 3rd place) is represented by a change in $\log u$ of

$$\log_e 3-\log_e 2 = 0{\cdot}1761. \tag{20}$$

Further down the Division such a change is reflected less in $\log u$, e.g.

$$\log_e 13-\log_e 12 = 0{\cdot}0348. \tag{21}$$

The fractional change in A is thus approximately five times greater in the former case than in the latter, which might be considered a reasonable ratio. In the event we experimented with functions for which this ratio was both greater and smaller than 5. None fitted the data as well as the original choice.

5. Results

The ordinary least squares (OLS) estimates for equation (14) are given in Table 1. The figures in brackets below the estimates are the t-ratios.

The corresponding solutions of equations (15)–(19) for the parameters of the model are given in Table 2.

It can be seen that the estimates of b_1 are very imprecise. This is partly due to the fact that X displays little variation in comparison with Y, since it applies to a single team rather than to a cross-section, and hence is highly serially correlated. This is especially noticeable with Leeds, who were stationary at the top of the Division for almost an entire season, and Newcastle, at the lower end, for which the coefficient has the wrong sign.

TABLE 1

OLS estimates

	b_0	$b_1(X)$	$b_2(Y)$	$b_3(P)$	$b_4(D)$	$b_5(T)$	F	R^2
Leeds	9·397	−0·009	−0·092	0·128	−0·009	0·042	4·86	0·43
	(20·67)	(−0·21)	(−2·53)	(3·08)	(−0·24)	(1·83)	(5, 33)	
Newcastle	8·855	0·058	−0·080	0·307	−0·221	−0·079	10·09	0·63
	(10·57)	(0·51)	(−1·83)	(4·13)	(−3·69)	(−2·41)	(5, 29)	
Notts	10·393	−0·205	−0·108	0·189	−0·288	−0·033	9·45)	0·59
	(12·06)	(−0·65)	(−1·82)	(2·76)	(−4·68)	(−0·52)	(5, 33)	
Southampton	9·641	−0·091	−0·101	0·129	−0·075	−0·019	9·24	0·58
	(17·76)	(−1·84)	(−4·03)	(2·87)	(−1·67)	(−1·01)	(5, 33)	

(The 5 per cent points are $t_{33} = 2{\cdot}03$, $t_{29} = 2{\cdot}05$, $F_{5,33} = 2{\cdot}64$, $F_{5,29} = 2{\cdot}70$.)

TABLE 2

OLS values for model parameters

	a_1	β	γ	θ	ϕ
Leeds	0·872	0·070	0·042	0·005	−0·106
Newcastle	0·693	0·719	−0·079	0·168	−0·190
Notts	0·811	1·519	−0·033	−0·235	−0·078
Southampton	0·871	0·586	−0·019	−0·089	−0·103

The estimates of b_2 are agreeably uniform. In the same way, the values of ϕ are much better than those of θ, though it is interesting to note that the one estimate of θ derived from a significant estimate of b_1 (Southampton) is, at −0·089, of the same order as the estimates of ϕ. The large value of ϕ for Newcastle is due to the wrong sign on b_1.

It would seem, then, that −0·1 is a reasonable figure for ϕ and possibly also for θ. This value can be illustrated as follows. If a team is imagined to play, at home and on successive weeks, opponents who are respectively 1st, 5th, 10th, 15th and 20th in the Division, the drops in attendance, *ceteris paribus*, between successive matches are respectively 15 per cent, then a further 6·5, 4 and 2·8 per cent.

A striking example of the limitations of League position as a determinant of attendance was revealed by plotting the residuals from these regressions. The four histograms are satisfactorily fitted by Normal curves except for one observation, Leeds' last match of 1969/70, which lies out nearly four standard errors. The predicted

attendance for this match, against Manchester City, is 39,300; the actual figure was 22,932. This is simply a result of Leeds having finally lost, in the previous fortnight, all chance of finishing top of the League, being beaten to the post by Everton.

Both of the demographic variables, P and D, perform very strongly. The estimates of b_3 are also, through equation (17), estimates of the mean proportion of total support from away and appear to be eminently reasonable, except possibly for Newcastle. This latter estimate might be suspected since it has the largest variance of the four. In addition, we examined the residuals for consistent biasses which could be associated with particular towns. A few mistakes were revealed in this way; for example, we had divided Manchester's population equally between its two clubs, and consequently underestimated the support for United, and overestimated for City.

Distance is perhaps the most interesting variable. For Leeds, low values of $|\beta|$ and hence b_4 might be expected *a priori* owing to its great ease of access. Newcastle has a reasonable value of 0·719; the corresponding drop in away support for a doubling in distance is 29 per cent. Nottingham has produced a value which we suspect is rather high; a true value near unity is not inconsistent with this estimate. Finally, given its relative isolation, Southampton should have a larger coefficient than 0·568. In this case we suspect that the $D_i^{-\beta}$ dependence is not an accurate representation of reality, for there are effectively two populations here. London supporters are little over an hour away by train, but access from the North and Midlands is not particularly convenient owing to the need to cross London.

The time trend does not show much explanatory power, and for Leeds the estimate goes against long-run behaviour. However, its omission results in a considerable drop in significance for some of the other parameter estimates; on balance there is a net gain from its inclusion.

The variables are jointly significant in all four equations. Finally, there was no evidence of serial correlation in the residuals, but the usual measures (such as the Durbin–Watson statistic) are not presented because the series do not consist of equally spaced observations.

6. Ancillary Results

It seemed worth while to compute our model's predictions for attendances in the season 1972/73, which was not over at the time of writing. Nottingham was relegated from the First Division after 1971/72 and has not been included, nor have matches against teams newly promoted to the First Division.

We would not expect the predictions to be particularly accurate for the reason that there was a large ($33\frac{1}{3}$ per cent) football admission-price increase prior to 1972/73 which would be expected to have reduced attendances, though not, of course, gate receipts. This is borne out by the predictions, which are presented in Tables 3 (Leeds), 4 (Newcastle) and 5 (Southampton).

We suspect that a moderate amount of the remaining prediction errors is due to inaccuracies in the P series, though only for Manchester is it positively identifiable.

We turn now to the hypothesis that the recent past performance of teams may influence current attendance. It seemed sensible to include in equation (14), *ad hoc*, and entirely without prior restrictions, the variables $\log X_{i-1}$, $\log X_{i-2}$, $\log Y_{i-1}$ and $\log Y_{i-2}$ in order to obtain upper limits to the increase in association which would be obtained from the use of these variables within, for instance, an adaptive expectations framework. No results are presented from these regressions, because they were not at all successful, the significance of the lagged variables being low and the increase

TABLE 3

Leeds predictions 1972/73

Date	*Opponents*	*Attendance*	*Prediction*	*% Error*
16/9/72	Leicester	33,930	36,440	+7·4
30/9/72	Liverpool	46,468	50,540	+8·8
7/10/72	Derby	36,477	34,710	−4·9
21/10/72	Coventry	36,241	37,870	+4·5
11/11/72	Sheffield United	35,600	37,700	+5·9
25/11/72	Manchester City	39,879	41,710	+4·6
9/12/72	West Ham	30,270	40,720	+34·5
3/1/73	Tottenham	32,404	43,140	+33·1
27/1/73	Stoke	33,487	36,650	+9·4

Mean % error = +11·5. S.D. (% points) = 13·3.

TABLE 4

Newcastle predictions 1972/73

Date	*Opponents*	*Attendance*	*Prediction*	*% Error*
26/8/72	Ipswich	24,250	19,460	−19·8
9/9/72	Arsenal	23,849	35,060	+47·0
23/9/72	Leeds	40,127	40,790	+1·7
21/10/72	Manchester United	38,170	33,490	−12·3
9/12/72	Southampton	23,750	20,220	−15·0
23/12/72	Manchester City	28,240	33,940	+19·4
30/12/72	Sheffield United	28,610	25,760	−10·0
20/1/73	Crystal Palace	28,660	24,250	−15·4

Mean % error = −0·6. S.D. (% points) = 23·0.

TABLE 5

Southampton predictions 1972/73

Date	*Opponents*	*Attendance*	*Prediction*	*% Error*
26/8/72	Wolverhampton	19,456	19,970	+2·7
9/9/72	Ipswich	13,919	19,140	+37·5
23/9/72	Crystal Palace	15,649	21,390	+36·7
14/10/72	Liverpool	24,100	26,510	+10·0
28/10/72	West Bromwich Albion	15,810	18,750	+18·6
18/11/72	Chelsea	24,164	25,320	+4·8
2/12/72	Tottenham	16,486	25,790	+56·5
23/12/72	West Ham	19,429	22,670	+16·7
30/12/72	Coventry	15,261	20,920	+38·5
20/1/73	Sheffield United	12,125	19,830	+63·5

Mean % error = +28·6. S.D. (% points) = 21·2.

in association negligible. There was thus little hope for adaptive expectations schemes, and this was confirmed by the (maximum-likelihood) estimates which were obtained.

Another idea tested was the interaction between X and Y which might be expected when they take similar values. The term $\log|X-Y|$ was included in equation (14); this would be expected to take a negative sign since its minimum value, zero, corresponds to the greatest competition ($X = Y \pm 1$). A similar effect occurs when "relegation battles" are played. Here terms of the form $\log X(N-X)$ and $\log Y(N-Y)$ were included, where N is the size of the Division. However, none of these variables proved to be particularly strong.

The fourth and last experiment was more successful. In Section 1 it was suggested that there might be common contemporaneous influences on most or all of the club's attendances, such as televised counter-attractions, the weather and seasonality (the "Christmas shopping" syndrome). This is confirmed by the correlation matrix of the contemporaneous residuals from the four regressions. The actual magnitudes depend on whether or not non-contemporaneous observations are included when computing the residual variances. This was in fact done and the corresponding results are presented in Table 6.

TABLE 6

Cross-correlations of OLS residuals

	Leeds	*Newcastle*	*Notts*	*Southampton*
Leeds	1			
Newcastle	0·320	1		
Notts	0·133	0·316	1	
Southampton	0·136	0·422	0·124	1

It can be seen that all the correlations are positive, as such hypotheses would require.

It is possible to use this information to improve upon the efficiency of the OLS estimates. The latter are no longer efficient since one of the Gauss–Markov conditions, that the distribution of the residuals in each equation is independent of that of any other random variables, is seen not to be upheld. The technique used here is Seemingly Unrelated Least Squares (SULS), developed by Zellner (1962). Essentially it involves estimation of M regressions simultaneously by means of the Aitken Generalized Least Squares estimator. The regressions are written in the form

$$\begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_M \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_M \end{pmatrix}$$

or simply

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{X}$ is $MT \times \sum_1^M K_i$ and $\mathbf{Y}$, $\boldsymbol{\beta}$ and $\mathbf{u}$ are "long" vectors. Denoting the covariance of $\mathbf{u}_i$ and $\mathbf{u}_j$ by σ_{ij} then

$$E(\mathbf{uu}') = (\sigma_{ij}) \otimes \mathbf{I}_T$$
$$= \boldsymbol{\Omega},$$

say, where (σ_{ij}) is the $M \times M$ matrix of variances and covariances, and the appropriate estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}.$$

Since (σ_{ij}) is unknown it may be estimated from the OLS residuals; this gives the "two-stage Aitken SULS estimator" which Kmenta and Gilbert (1968) suggest has desirable small sample properties.

The problem in appling SULS in this case is that the observations are not equal in number and are not all contemporaneous. However, it is clear that inserting "dummy matches" consisting of zero observations (actually blank computer data cards) in the series to bring the total number of observations up to the total number of weeks on which at least one team is playing and in such a way as to render the correct matches contemporaneous, does not affect the usual sums of squares or cross-products, providing the intercept is estimated (as a constant vector) in the same manner as the other regression coefficients.

This result is immediately confirmed by running a SULS program in these circumstances with the off-diagonal terms in (σ_{ij}) suppressed; the OLS estimates and t-statistics are indeed obtained.

The SULS estimates and asymptotic t-values are presented in Table 7, and the corresponding model parameter values in Table 8.

TABLE 7

SULS *estimates*

	b_0	$b_1(X)$	$b_2(Y)$	$b_3(P^A)$	$b_4(D)$	$b_5(T)$
Leeds	9·273	−0·015	−0·098	0·137	0·002	0·045
	(23·31)	(−0·39)	(−3·11)	(3·75)	(0·07)	(2·25)
Newcastle	8·922	0·076	−0·084	0·305	−0·243	−0·070
	(13·84)	(0·86)	(−2·49)	(5·35)	(−5·31)	(−2·73)
Notts	10·220	−0·132	−0·108	0·187	−0·289	−0·036
	(13·48)	(−0·47)	(−2·08)	(3·11)	(−5·37)	(−0·65)
Southampton	9·605	−0·086	−0·107	0·131	−0·072	−0·020
	(21·15)	(−2·07)	(−5·08)	(3·51)	(−1·91)	(−1·31)

The individual F-statistics are no longer available and the $R^{2\prime}$s do not differ from the OLS values to two decimal places, though they are of course slightly lower.

The asymptotic gain in efficiency is quite evident, and is aided by the near-orthogonality of the independent variables (only X and P exhibit any significant correlation). The values of a_1, β and γ are little changed, but those of θ and ϕ are rather better than before.

TABLE 8

SULS *values for model parameters*

	a_1	β	γ	θ	ϕ
Leeds	0·863	0·015	0·045	0·001	−0·113
Newcastle	0·695	0·797	−0·070	0·202	−0·209
Notts	0·813	1·545	−0·036	−0·140	−0·100
Southampton	0·869	0·550	−0·020	−0·082	−0·111

7. CONCLUSIONS

We have tried in this paper to present an objective analysis of the behaviour of football supporters. While our model applies to four specific clubs within the First Division of the English Football League, tentative conclusions may be drawn concerning professional football in general. In particular, distance is seen to provide an impediment to away supporters, but only at certain grounds. Further, supporters seem to regard the calibre of the opposing team with some importance. We suspect that supporters' response to the performance of their own team is more sensitive than could be gauged from our model.

The issues at stake are obviously of some importance to the Football League since, especially in recent times, they have shown considerable interest in the factors which motivate football supporters. While there is clearly a large irrational element in the psychology of football support, we have shown that a quantitative approach to the problem can account for a significant proportion of the short-term variation in attendances.

ACKNOWLEDGEMENTS

An earlier version of this paper was presented at a meeting of the Royal Statistical Society, Highlands Group. We are grateful to the participants at this meeting for a number of suggestions which have since been incorporated. We would also like to express our gratitude to Mr G. A. Mackay, Dr P. J. Sloane and a referee. Finally, the assistance of Leeds United, Newcastle United, Nottingham Forest and Southampton football clubs in providing information on match attendances is gratefully acknowledged.

REFERENCES

DEPARTMENT OF EDUCATION AND SCIENCE (1968). *Report of the Committee on Football.* ("The Chester Report".) London: H.M.S.O.

KMENTA, J. and GILBERT, R. F. (1968). Small sample properties of alternative estimators of seemingly unrelated regressions. *J. Amer. Statist. Ass.*, **63**, 1180–1200.

NEALE, W. C. (1964). The peculiar economics of professional sport. *Qu. J. of Econ.*, **78**, 1–14.

SLOANE, P. J. (1971). The economics of professional football; the football club as a utility maximiser. *Scott. J. of Pol. Econ.*, **18**, 121–146.

ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Ass.*, **57**, 348–368.