Modelling association football scores

by M. J. MAHER*

Abstract Previous authors have rejected the Poisson model for association football scores in favour of the Negative Binomial. This paper, however, investigates the Poisson model further. Parameters representing the teams' inherent attacking and defensive strengths are incorporated and the most appropriate model is found from a hierarchy of models. Observed and expected frequencies of scores are compared and goodness-of-fit tests show that although there are some small systematic differences, an independent Poisson model gives a reasonably accurate description of football scores. Improvements can be achieved by the use of a bivariate Poisson model with a correlation between scores of 0.2.

Key Words: Poisson goals distribution, iterative maximum likelihood.

1 Introduction

MORONEY (1951) demonstrated that the number of goals scored by a team in a football match was not well fitted by a Poisson distribution but that if a "modified Poisson" (the Negative Binomial, in fact) was used, the fit was much better. REEP, POLLARD and BENJAMIN (1971) confirmed this, using data from the English Football League First Division for four seasons, and then proceeded to apply the Negative Binomial distribution to other ball games. The implication of this result is that the same Negative Binomial distribution applied to the number of goals scored by a team, regardless of the quality of that team or the quality of the opposition. In fact in an earlier paper, REEP and BENJAMIN (1968) remarked that "chance does dominate the game". HILL (1974) was unconvinced by this and showed that football experts were able, before the season started, to predict with some success the final league table positions. Therefore, certainly over a whole season, skill rather than chance dominates the game. This would probably be agreed by most people who watch the game of football; that whilst in a single match, chance plays a considerable role (missed scoring opportunities, dubious offside decisions and shots hitting the crossbar can obviously drastically affect the result), over several matches luck plays much less of a part. Teams are not identical; each one has its own inherent quality, and, surely then we should expect that when a good team is playing a weak team, the good team will have a high probability of winning and scoring several goals. By using data from the whole or just a part of the season, these inherent qualities of the teams in a league can be inferred by, for example, maximum likelihood estimation (as in THOMPSON (1975)) or by linear model methodology (as in HARVILLE (1977) and LEEFLANG and VAN PRAAG (1971)).

2 The Model

There are good reasons for thinking that the number of goals scored by a team in a match is likely to be a Poisson variable: possession is an important aspect of football,

* Department of Probability and Statistics, Sheffield University, Sheffield S3 7RH, England.

and each time a team has the ball it has the opportunity to attack and score. The probability p that an attack will result in a goal is, of course, small, but the number of times a team has possession during a match is very large. If p is constant and attacks are independent, the number of goals will be Binomial and in these circumstances the Poisson approximation will apply very well. The mean of this Poisson will vary according to the quality of the team and so if one were to consider the distribution of goals scored by *all* teams, one would have a Poisson distribution with variable mean, and hence something like the Negative Binomial observed by MORONEY (1951) and REEP, POLLARD and BENJAMIN (1971) could arise.

Therefore, in this paper, at least for the present, an independent Poisson model for scores will be adopted. In particular, if team i is playing at home against team j and the observed score is (x_{ij}, y_{ij}), we shall assume that X_{ij} is Poisson with mean $\alpha_i \beta_j$, that Y_{ij} is also Poisson with mean $\gamma_i \delta_j$, and that X_{ij} and Y_{ij} are independent. Then we can think of α_i as representing the strength of team i's attack when playing at home, β_j the weakness of team j's defence when playing away, γ_i the weakness of team i's defence at home and δ_j the strength of team j's attack away. In a league with 22 teams there are 88 such parameters (and 924 observations on the scores); however if all the α's are multiplied by a factor k and all the β's divided by k, all the $\alpha_i \beta_j$ products are unaffected and, therefore, in order to produce a unique set of parameters the constraint

$$\sum_i \alpha_i = \sum_i \beta_i$$

may be imposed. In the same way the constraint

$$\sum_i \gamma_i = \sum_i \delta_i$$

may be imposed and so only 86 independent parameters need to be specified. Since the $\underline{X}$ and $\underline{Y}$ are assumed to be independent of each other (representing separate "games" at the two ends of the pitch), the estimation of the $\underline{\alpha}$ and $\underline{\beta}$ will be entirely from the $\underline{x}$ and the estimation of the $\underline{\gamma}$ and $\underline{\delta}$ by means of the $\underline{y}$ alone.

For the home teams' scores, therefore, the log likelihood function is:

$$\log L(\underline{\alpha}, \underline{\beta}) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log (\alpha_i \beta_j) - \log (x_{ij}!))$$

Therefore,

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left(-\beta_j + \frac{x_{ij}}{\alpha_i} \right)$$

and so the maximum likelihood estimates $\hat{\underline{\alpha}}, \hat{\underline{\beta}}$ satisfy:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j} \quad \text{and} \quad \hat{\beta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}$$

An iterative technique, such as NEWTON-RAPHSON, enables these MLEs to be determined. One simpler scheme which works well is to be use the $\hat{\alpha}$'s to estimate the $\hat{\beta}$'s and then to use the $\hat{\beta}$'s to estimate the $\hat{\alpha}$'s, and so on alternately. Good initial estimates can be gained by regarding the denominator terms in the expressions above as summations over all teams: that is,

$$\hat{\alpha}_i = \sum_{j \neq i} x_{ij}/\sqrt{S_X} \quad \text{and} \quad \hat{\beta}_j = \sum_{i \neq j} x_{ij}/\sqrt{S_X}, \quad \text{where} \quad S_X = \sum_i \sum_{j \neq i} x_{ij}.$$

In a similar way, using the y_{ij}, $\hat{\underline{\gamma}}$ and $\hat{\underline{\delta}}$ may be found.

3 Results

Data were obtained, in a convenient matrix form, from the Rothmans Football Yearbook (1973, 1974, 1975). This gave 12 separate leagues (the four English Football League Divisions for each of three seasons) for analysis. The MLEs of the four types of parameter $\underline{\alpha}, \underline{\beta}, \underline{\gamma}$ and $\underline{\delta}$ are shown in Table 1 for just one data set: Division 1 in the season 1971-1972.

Table 1. Maximum likelihood estimates of the parameters for Division 1 1971-1972.

	home attack α	away defence β	home defence γ	away attack δ
Arsenal	1.36	1.03	0.64	1.06
Chelsea	1.55	1.18	0.97	0.83
Coventry City	1.05	1.66	1.12	0.84
Crystal Palace	0.99	1.28	1.49	0.65
Derby County	1.62	0.89	0.50	1.24
Everton	1.06	1.17	0.81	0.44
Huddersfield Town	0.46	1.37	1.06	0.74
Ipswich Town	0.72	1.27	0.93	0.98
Leeds United	2.02	0.82	0.49	0.91
Leicester City	0.69	1.31	0.54	1.10
Liverpool	1.78	0.54	0.78	0.78
Manchester City	1.82	1.17	0.75	1.40
Manchester United	1.49	1.35	1.31	1.49
Newcastle United	1.14	1.29	0.88	0.93
Nottingham Forest	0.98	1.96	1.43	1.10
Sheffield United	1.49	1.31	1.28	1.09
Southampton	1.21	1.98	1.38	1.05
Stoke City	0.99	1.17	1.20	0.64
Tottenham Hotspur	1.71	1.12	0.63	0.87
West Bromwich Albion	0.84	1.16	1.13	0.99
West Ham United	1.18	1.22	0.92	0.78
Wolverhampton Wanderers	1.34	1.30	1.15	1.48

The question arises of whether all these parameters are necessary for an adequate description of the scores. Intuitively it seems that there must be real differences between teams, but are these differences more apparent in the attacks or defences, and is it really necessary to have separate parameters for the quality of a team's attack at home and away? Consideration of such questions leads to a possible hierarchy of models which could be tested. At the bottom is model 0 in which $\alpha_i = \alpha, \beta_i = \beta, \gamma_i = \gamma$ and $\delta_i = \delta \, \forall i$; that is, all teams are identical in all respects. At the top is model 4, previously described, in which all four types of parameter are allowed to take different values for the different teams. The hierarchy is shown in Table 2. In this the notation is designed to show whether a set of parameters (such at the β) are free to take different values for the different teams (shown as β_i) or whether the same value applies to all teams (shown as β).

Table 2. Hierarchy of models, with changes in the value of twice the maximised log likelihood shown for Division 1 1971-1972

Model 4		$(\alpha_i, \beta_i, \gamma_i, \delta_i)$		
	35.4			16.8
Models 3C, 3D	$(\alpha_i, \beta_i, \gamma_i, \alpha_i)$		$(\alpha_i, \beta_i, \beta_i, \delta_i)$	
	17.2			35.8
Model 2		$(\alpha_i, \beta_i, k\beta_i, k\alpha_i)$		
	71.4			72.8
Models 1A, 1B	$(\alpha_i, \beta, \gamma, \alpha_i)$		$(\alpha, \beta_i, \beta_i, \delta)$	
	75.4			74.0
Model 0		$(\alpha, \beta, \gamma, \delta)$		

In model 0 there are four parameters but in order to have a unique set of parameter estimates, the constraints $\alpha = \beta$ and $\gamma = \delta$ are imposed (or, equivalently, $\alpha = \beta$, $\gamma = k\beta$ and $\delta = k\alpha$), giving just two independent parameters. Details of the constraints imposed in the other models are as follows:

Model 1A $\delta_i = \alpha_i, \beta_i = \beta, \gamma_i = \gamma \, \forall i$; $\Sigma\alpha_i = \Sigma\beta_i$. Therefore, there are $n+1$ independent parameters (where n is the number of teams in the league).

Model 1B $\gamma_i = \beta_i$, $\alpha_i = \alpha$, $\beta_i = \beta \, \forall i$; $\Sigma\alpha_i = \Sigma\beta_i$. Again, there are $(n+1)$ independent parameters.

Model 2 $\delta_i = k\alpha_i$, $\gamma_i = k\beta_i \, \forall i$; $\Sigma\alpha_i = \Sigma\beta_i$. There are $2n$ independent parameters.

Model 3C $\delta_i = \alpha_i \, \forall i$; $\Sigma\alpha_i = \Sigma\beta_i$. $3n-1$ independent parameters.

Model 3D $\gamma_i = \beta_i \, \forall i$; $\Sigma\alpha_i = \Sigma\beta_i$. Again, $3n-1$ independent parameters.

Model 4 $\Sigma\alpha_i = \Sigma\beta_i$ and $\Sigma\gamma_i = \Sigma\delta_i$. Therefore, there are $4n-2$ independent parameters.

It can be seen, therefore, that moving up one level in the hierarchy of models leads to the introduction of $(n-1)$ further parameters. Under the null hypothesis that these extra parameters are unnecessary, $2 \log_e \lambda$ will be asymptotically χ^2_{n-1} distribution by the usual likelihood ratio test, where $\log_e \lambda$ is the increase in the log likelihood in moving from one model to the other.

For Division 1 in the season 1971-1972 the changes in the value of the maximised log likelihood when moving from one model to another are shown in Table 2 ($n=22$; $\chi^2_{.95}(21)=32.7$ and $\chi^2_{.99}(21)=38.9$).

This table shows that when inequality of the α_i is allowed (moving from model 0 to model 1A or moving from model 1B to model 2), a highly significant increase in the log likelihood results. Similarly, when inequality of the β_i is allowed (model 0 to 1B, or 1A to 2), again the log likelihood increases very significantly. When the δ_i are freed from being proportional to the α_i (model 2 to 3D or 3C to 4), a marginally significant increase in the log likelihood is obtained. However, when the γ_i are freed from their linking with the β_i, no significant increase results. It should be noticed that the order in which the freeing of these parameters occurred had virtually no effect on the increase in the log likelihood due to each one; this was true for all the twelve data sets. Therefore, it is possible to associate an increase in the log likelihood with each of the four types of parameters, and, in parallel with the ideas of linear models in which factors are introduced into the model one at a time, the "inclusion of $\underline{\gamma}$", for example, means the freeing of the γ_i from their linking with the β_i. Table 3 shows the increases in log likelihood due to the inclusion of each of the four types of parameter, for each of the twelve data sets. There are 22 teams in divisions 1 and 2 and 24 teams in divisions 3 and 4. The numbers of degrees of freedom, therefore, in the asymptotic χ^2 distribution for $2 \log_e \lambda$ are 21 and 23 respectively.

Overall, then, it can be seen that the parameters $\underline{\alpha}$ and $\underline{\beta}$ should certainly be included in the model but the parameters $\underline{\gamma}$ and $\underline{\delta}$ need not be included. (Not only can the null hypotheses not be rejected in these latter cases, but they appear perfectly consistent with the data.) This means that a single parameter α_i can be used to describe the quality of team i's attack, and the parameter β_i to describe the weakness of the team's defence, whether the team is playing at home or away. So although home ground advantage is a highly significant factor, it applies with equal effect to all teams, and each team's inherent scoring power is diminished by a constant factor when playing away.

Table 3. Increase in log likelihood due to inclusion of each of the four types of parameter in the model

season	division	$\underline{\alpha}$	$\underline{\beta}$	$\underline{\gamma}$	$\underline{\delta}$
1971-1972	1	37.7**	35.7**	8.6	17.7*
	2	23.4**	32.4**	8.7	6.1
	3	40.6**	28.1**	11.5	19.7*
	4	29.4**	34.2**	12.5	11.0
1972-1973	1	24.8**	7.1	8.8	8.2
	2	23.2**	17.9*	3.9	13.2
	3	27.8**	18.4*	12.3	12.1
	4	26.3**	30.2**	8.9	15.0
1973-1974	1	12.5	19.5**	14.8	10.6
	2	19.8**	20.1**	15.1	10.2
	3	23.4**	39.9**	13.5	14.9
	4	31.0**	28.4**	8.5	13.0

* indicates a significant increase at the 5% level
** indicates a significant increase at the 1% level

In the light of the results above, then, Model 2 was adopted as being the most appropriate, and further analyses were made of its adequacy as a description of the mechanism underlying football scores.

4 Goodness-of-fit-tests

For a match between team i and team j the MLEs from model 2 may be used to estimate μ_{ij} and λ_{ij}, the means of X_{ij} and Y_{ij}. Since X_{ij} and Y_{ij} are assumed to be Poisson and independent, the probabilities that $X_{ij}=x$ and $Y_{ij}=y$ may be easily calculated. By repeating this for all pairs of i and j, the expected score distributions may be found and compared with the observed score distributions. For Division 1 in the season 1971-1972, for example, these observed and expected frequencies are shown in Table 4.

Table 4. Observed and expected frequencies of home and away scores for Division 1 1971-1972

	home		away	
no. of goals	obs.	exp.	obs.	exp.
0	117	111.2	184	189.3
1	127	144.6	157	159.5
2	115	106.1	88	75.9
3	66	58.0	30	26.9
$\geqslant 4$	37	42.1	3	10.5
	$\chi^2=4.90$		$\chi^2=7.79$	

For the fitted model 2 the MLEs of the parameters are:

$$\hat{\alpha}_i=\frac{\sum\limits_{j\neq i}(x_{ij}+y_{ji})}{(1+\hat{k}^2)\sum\limits_{j\neq i}\hat{\beta}_j}\qquad \hat{\beta}_j=\frac{\sum\limits_{i\neq j}(x_{ij}+y_{ji})}{(1+\hat{k}^2)\sum\limits_{i\neq j}\hat{\alpha}_i}\ \forall i,j$$

and

$$\hat{k}^2=\frac{\sum\limits_{i}\sum\limits_{j\neq i}y_{ij}}{\sum\limits_{i}\sum\limits_{j\neq i}x_{ij}}$$

It follows from this that

$$\sum_{i}\sum_{j\neq i}\hat{\alpha}_i\hat{\beta}_j=\sum_{i}\sum_{j\neq i}x_{ij}$$

and that

$$\sum_{i}\sum_{j\neq i}\hat{k}^2\alpha_j\beta_i=\sum_{i}\sum_{j\neq i}y_{ij},$$

which means that the sum of the means of the fitted Poisson distributions is equal to the observed numbers of goals scored. The estimation of the parameters gives rise, therefore, to one linear constraint on the expected frequencies in each of the two χ^2 goodness-of-fit tests in Table 4. The resulting statistics will be approximately χ_3^2 distributed under the hypotheses that home and away teams' scores are Poisson distributed. This was repeated for each of the other eleven data sets, and the resulting χ^2 statistics are listed in Table 5.

Table 5. Values of the χ^2 statistic for home and away teams' scores for the independent Poisson model

		χ^2 values	
season	division	homes	aways
1971-1972	1	4.90	7.79
	2	5.71	1.08
	3	10.05*	8.96*
	4	4.62	1.07
1972-1973	1	6.08	13.41**
	2	3.44	9.77*
	3	4.94	4.31
	4	0.78	3.22
1973-1974	1	7.91*	1.33
	2	1.97	1.12
	3	0.89	5.28
	4	3.61	1.92
critical values	5% level 7.81 1% level 11.3		

* indicates a significant value at the 5% level
** indicates a significant value at the 1% level

The case where the model would be rejected are shown by an asterisk. For the home teams' scores there are two such cases and for the away teams' scores there are three. Overall, then, the Poisson model may be regarded as acceptable, although with some slight doubt. If the observed and expected frequencies are compared for each of the twelve data sets, some small but systematic differences can be seen. The overall observed and expected proportions are:

HOME SCORES

no. of goals	0	1	2	3	$\geqslant 4$
observed	0.217	0.321	0.254	0.130	0.078
expected	0.230	0.318	0.238	0.128	0.086

AWAY SCORES

no. of goals	0	1	2	3	$\geqslant 4$
observed	0.388	0.371	0.177	0.051	0.014
expected	0.406	0.352	0.166	0.056	0.020

The model underestimates the number of occasions on which one and two goals are scored, and overestimates the number of times that 0 or $\geqslant 4$ goals are scored. This effect can be seen in each of the twelve data sets. The differences are small and over just one season do not seriously inflate the χ^2 value, but if the observed and expected frequencies for all twelve seasons are added the values of the χ^2 statistics (16.2 and 28.8 for home and away scores respectively) would lead to clear rejection of the model. The distribution of the number of goals scored by a team in a match is very close to a Poisson distribution, then, but is slightly "narrower". This might seem to conflict with MORONEY's (1951) and REEP and BENJAMIN's (1968) conclusion which was that a distribution which was wider (in terms of the variance to mean ratio) than the Poisson was required; the Negative Binomial was their fitted distribution. However, in both these other works a single distribution was fitted to scores from all matches, whereas here each match has a different fitted Poisson distribution.

5 A bivariate Poisson model

There is no shortage of possible explanations, of course, for the small discrepancy between the independent Poisson model and the data; in fact it is perhaps fairer to say that it is surprising that such a simple model comes so close to explaining the data so fully! A match does not consist of two independent games at opposite ends of the pitch; to the teams concerned, the result is all important, and so, for example, if a team is losing with ten minutes left to play, it must take more defensive risks in order to try to score. Therefore, an examination of the distribution of the difference between the teams' scores, $Z_{ij} = X_{ij} - Y_{ij}$ might be revealing. Table 6 shows the observed and estimated frequencies for Z under model 2 for Division 1 in 1971-1972.

Table 6. Observed and estimated frequencies for Z, the difference in the teams' scores, for Division 1 in 1971-1972, for (i) the independent Poisson model and (ii) the bivariate Poisson with $\varrho = 0.2$

Z	$\leqslant -3$	-2	-1	0	$+1$	$+2$	$+3$	$+4$	$\geqslant 5$
observed	8	26	72	129	105	69	31	16	6
estimated ($\varrho = 0$)	14.4	30.3	69.8	113.0	104.9	68.7	35.8	15.8	9.3
estimated ($\varrho = 0.2$)	9.9	25.3	68.0	126.2	111.7	67.7	32.6	13.4	7.1

In this it can be seen that the number of drawn matches ($Z = 0$) is a little underestimated. This is a systematic feature noted in all twelve data sets. The χ^2 goodness-of-fit statistics are shown in Table 7; four of the twelve are significant at the 5% level, whilst several others approach this. Only one of the twelve has a value of the χ^2 statistic which is less than the expected value of 7. (The number of degrees of freedom is reduced to 7 because of the linear constraint on the expected frequencies resulting from the estimation of the α's and β's.) This suggests that there may be some correlation between the X_{ij} and Y_{ij}. A bivariate Poisson model was tried; in this the marginal distributions are still Poisson with means $\mu_{ij} (= \alpha_i \beta_j)$ and $\lambda_{ij} (= k^2 \alpha_j \beta_i)$ but there is a correlation of ϱ between the scores. One way of thinking of such a bivariate Poisson distribution is that $X_{ij} = U_{ij} + W_{ij}$ and $Y_{ij} = V_{ij} + W_{ij}$ where U_{ij}, V_{ij} and W_{ij} are independent Poisson with means of $(\mu_{ij} - \eta_{ij})$, $(\lambda_{ij} - \eta_{ij})$ and η_{ij} respectively, with $\eta_{ij} (= \varrho\sqrt{\mu_{ij}\lambda_{ij}})$ being the covariance between X_{ij} and Y_{ij}.

A range of values of ϱ was tried and the most appropriate seemed to be around 0.2. In computing the expected frequencies for Z, the values of the $\hat{\alpha}$'s, $\hat{\beta}$'s and $\hat{k}$ used were those found from the fitting of the independent Poisson model. The terms in the Poisson bivariate probability function can be calculated by the following recursive relationship:

$$p_{00} = \exp(-\mu - \lambda - \eta)$$
$$x\, p_{xy} = (\mu - \eta) p_{x-1\, y} + \eta p_{x-1\, y-1}$$
$$y\, p_{xy} = (\lambda - \eta) p_{x\, y-1} + \eta p_{x-1\, y-1}$$

Table 7. Values of the χ^2 goodness-of-fit statistic for Z, the difference between the teams' scores, for (i) the independent model and (ii) the bivariate Poisson model with $\varrho = 0.2$

		χ^2 values	
season	division	independent model	bivariate model
1971-1972	1	9.67	1.86
	2	16.42*	6.50
	3	10.87	3.94
	4	12.99	5.75
1972-1973	1	15.51*	4.77
	2	13.70	11.98
	3	4.79	2.50
	4	15.30*	8.27
1973-1974	1	16.47*	9.08
	2	9.76	12.29
	3	13.53	8.00
	4	10.36	5.40
critical values	5% level	14.1	12.6
	1% level	18.5	16.8

The results of fitting this bivariate Poisson model are shown in Tables 6 and 7, where it can be seen that the introduction of the extra parameter ϱ has led to a considerable improvement in the fit. The χ^2 statistics in Table 7 are not only non-significant but are fairly representative values from a χ^2_6 distribution. (It has been assumed that the fitting of the extra parameter ϱ will be roughly equivalent to the imposition of another linear constraint on the expected frequencies, although in fact the same value of ϱ has been applied to all the twelve data sets). A bivariate Poisson model with correlation of about 0.2, therefore, would seem to give a very adequate fit to the differences in scores.

6 Summary

Previous work on the distribution of scores in football matches has rejected the Poisson model in favour of the Negative Binomial. This work, however, has not allowed for the different qualities of the teams in a league. The first model investigated here assumes that the home team's and aways team's scores in any one match are independent Poisson variables with means $\alpha_i \beta_j$ and $\gamma_i \delta_j$, where the parameters $\underline{\alpha}, \underline{\beta}, \underline{\gamma}$ and $\underline{\delta}$ represent the qualities of the teams attacks and defences, in home and away matches. Maximum likelihood estimation of these parameter shows that only the $\underline{\alpha}$ and $\underline{\beta}$ are needed, showing that the relative strength of teams' attacks is the same whether playing at home or away; the same applies to the defences.

When this model is aplied to each of the twelve data sets and observed and expected score distributions are compared by means of a χ^2 test, nineteen out of the twenty-four cases give a non-significant result at the 5% level. Overall, then, the independent Poisson model gives a reasonably good fit to the data. The deviations from this model are small but consistent in each of the data sets, there being slightly fewer occasions observed than expected on which no goals or a large number of goals are scored. When the differences in scores are investigated however, the lack of fit of the model is rather more serious and suggests that the independence assumption is not totally valid. A bivariate Poisson model was then used to model this dependence between scores and this improved the fit considerably for the differences in scores. The correlation coefficient between home teams' and away teams' scores is estimated to be approximately 0.2.

Acknowledgement

The author would like to thank an anonymous referee for his very helpful comments on an earlier version of this paper.

References

HARVILLE, D. (1977), The use of linear-model methodology to rate high school or college football teams, *J. Amer. Statist. Ass. 72, No. 358,* pp. 278–289.

HILL, I. D. (1974), Association football and statistical inference, *Appl. Statist. 23, No. 2,* pp. 203–208.

LEEFLANG, P. S. H. and B. M. S. VAN PRAAG (1971), A procedure to estimate relative powers in binary contacts and an application to Dutch football league results, *Statist. Neerlandica 25, No. 1,* pp. 63–84.

MORONEY, M. (1951), *Facts from figures,* London, Pelican.

REEP, C. and B. BENJAMIN (1968), Skill and chance in association football, *J. R. Statist. Soc. A, 131,* pp. 581–585.

REEP, C., R. POLLARD and B. BENJAMIN (1971), Skill and chance in ball games, *J. R. Statist. Soc. A, 134,* pp. 623–629.

THOMPSON, M. (1975), On any given Sunday: fair competitor orderings with maximum likelihood methods, *J. Amer. Statist. Ass. 70, No. 351,* pp. 536–541.

Received October 1981, Revised December 1981.