Home Ground Advantage of individual clubs in English Soccer

By STEPHEN R CLARKE[†]

Swinburne University of Technology, Melbourne, Australia
and JOHN M NORMAN
University of Sheffield, UK

SUMMARY

Least squares is used to fit a model to the individual match results in English football and produce a Home Ground Advantage (HA) effect for each team in addition to a team rating. We show for a balanced competition this is equivalent to a simple calculator method using only data from the final ladder. The existence of a spurious HA is discussed. HA's for all teams in the English Football league from 1981-82 to 1990-91 are calculated, and some reasons for their differences investigated. A paired HA is defined and shown to be linearly related to the distance between club grounds.

Keywords: Football statistics; Home ground advantage; Soccer; Performance measures; Least squares.

1. Introduction

The existence of a home advantage (HA) in most sports is now well documented. Courneya and Carron (1992) give a summary of the work done on HA. They make the point that future research should be directed to the causes of HA rather than document its existence. However this requires the calculation of the HA's of individual clubs, so differences can be related to the playing characteristics of the clubs. Pollard (1986) quantifies HA (in a competition where each team plays an equal number of matches at home and away) as the number of games won by teams playing at home expressed as a percentage of all games played, with 50% indicating no HA. While this method is acceptable when averages over a whole competition are taken, it is obviously inadequate when the performance of individual clubs is studied. Here a team may well win more (or less) than 50% at home because it is a relatively strong (or weak) team. Snyder and Purdy (1985) show the limitations of this approach, when in looking at a universities basketball competition they found division 2 teams won only 40% of their

---

home matches against division 1 teams. This implies the quality of opposition effect overshadowed the HA effect. Because the quality of teams differ, we must allow for differences in ability and measure HA by comparing a team's home and away performance.

TABLE 1
End of season ladder for division 1, 1986-87.

TEAM	HW	HD	HL	Hf	Ha	HGD	AW	AD	AL	Af	Aa	AGD	GD	Pnts	h	u
Everton	16	4	1	49	11	38	10	4	7	27	20	7	45	86	0.84	0.92
Liverpool	15	3	3	43	16	27	8	5	8	29	26	3	30	77	0.49	0.76
Tottenham Hotspur	14	3	4	40	14	26	7	5	9	28	29	-1	25	71	0.64	0.57
Arsenal	12	5	4	31	12	19	8	5	8	27	23	4	23	70	0.04	0.82
Norwich City	9	10	2	27	20	7	8	7	6	26	31	-5	2	68	-0.11	0.42
Wimbledon	11	5	5	32	22	10	8	4	9	25	28	-3	7	66	-0.06	0.51
Luton Town	14	5	2	29	13	16	4	7	10	18	32	-14	2	66	0.79	-0.03
Nottingham Forest	12	8	1	36	14	22	6	3	12	28	37	-9	13	65	0.84	0.20
Watford	12	5	4	38	20	18	6	4	11	29	34	-5	13	63	0.44	0.40
Coventry City	14	4	3	35	17	18	3	8	10	15	28	-13	5	63	0.84	0.01
Manchester United	13	3	5	38	18	20	1	11	9	14	27	-13	7	56	0.94	0.01
Southampton	11	5	5	44	24	20	3	5	13	25	44	-19	1	52	1.24	-0.28
Sheffield Wednesday	9	7	5	39	24	15	4	6	11	19	35	-16	-1	52	0.84	-0.12
Chelsea	8	6	7	30	30	0	5	7	9	23	34	-11	-11	52	-0.16	0.15
West Ham United	10	4	7	33	28	5	4	6	11	19	39	-20	-15	52	0.54	-0.29
Queens Park Rangers	9	7	5	31	27	4	4	4	13	17	37	-20	-16	50	0.49	-0.29
Newcastle United	10	4	7	33	29	4	2	7	12	14	36	-22	-18	47	0.59	-0.38
Oxford United	8	8	5	30	25	5	3	5	13	14	44	-30	-25	46	1.04	-0.77
Charlton Athletic	7	7	7	26	22	4	4	4	13	19	33	-14	-10	44	0.19	-0.00
Leicester City	9	7	5	39	24	15	2	2	17	15	52	-37	-22	42	1.89	-1.12
Manchester City	8	6	7	28	24	4	0	9	12	8	33	-25	-21	39	0.74	-0.53
Aston Villa	7	7	7	25	25	0	1	5	15	20	54	-34	-34	36	0.99	-0.95
						---						---			----	----
						297						-297			14.14	0.00

H, home; A, away; W, win; D, Draw; L, Loss; f, goals for; a, goals against; GD, goal difference h, Home advantage; u, team rating.

Table 1 shows a typical ladder (Division 1 1986) as published for English soccer, with the addition of two extra columns to be explained later. Sports followers have long recognised the importance of a HA, and soccer tables have traditionally separated

a team's home and away performance. At the bottom of the table, Aston Villa, under the percentage of games or points scored at home definition, could be construed as having no HA. Counting a draw as 0.5 of a win, they have won exactly 10.5 or 50% of games at home and scored exactly 50% of the goals in their home matches. However this is in large part due to their low team ability relative to their opponents. The away matches can be used to allow for team ability, and we see Aston Villa have won only 3.5 of their away matches and scored 34 less goals than their opponents. A supporter might say over the season they enjoyed a 10.5–3.5=7 game and a 0–(−34)=34 goal HA. Virtually all teams, irrespective of ability, have a HA when measured in this way. However, we shall show that this includes the HA of all the other teams, and so overestimates the true individual HA effect. It is not generally appreciated that each team with a 'real' home ground advantage automatically gives each other team in the competition a 'spurious' or apparent HA. This is best demonstrated by a simple example which is given in Appendix 1.

2. Modelling team ability and home advantage

To correctly estimate HA we need to model for ability of team. We use a model similar to one used by Stefani (1983, 1987), Stefani and Clarke (1992) and Clarke (1993) that has proved successful in predicting match results. The winning margin w_{ij} in a match between team i and team j played at the home ground of team i is modelled as

$$w_{ij} = u_i - u_j + h_i + \varepsilon_{ij} , \qquad (1)$$

where u_i is a measure of team i's ability, h_i is a measure of team i's HA and ε_{ij} is a zero mean random error. We assume the u_i's and h_i's are constant throughout the season.

The w_{ij} can be measured either in 'win margin' (1, 0 or -1 depending on whether the home team won drew or lost) or goals margin. The second case is usually followed in prediction models, and that is the method we prefer here as it is more sensitive to HA. For example a team that wins 3-0 at home and wins 2-1 away shows no HA if win margins are used, but shows a 2 goal advantage if goal margins are used.

The above model, with the additional constraint that the u_i's (being relative) sum to zero, can be fitted to the individual match results with a standard regression package by using dummy variables for the u's (1 if a team is home, -1 if the team is away and 0 for the other teams) and h's (1 for the home team and 0 for the others). In this case the REG procedure from SAS 6.10 gave the values for u and h shown in the last two columns of Table 1. The overall model was significant at the 0.01% level, with an R^2 of 0.19. The low value of R^2 reflects the high variablility in soccer, and for the other

seasons analysed increased with the unevenness of the season. Each u had a standard error of .33 and the h's a standard error of .49. A Q-Q plot of the residuals indicated they were normally distributed, which was confirmed with a Shapiro-Wilk test of normality statistic of 0.99 (p=0.63). A plot of the residuals against predicted values showed no evidence of heteroskedasticity.

Alternatively, the Lagrange multiplier technique can be used to derive the values of u_i and h_i that minimise the sums of the squares of the errors. This is shown in Appendix 2. Surprisingly, the derived equations only use simple arithmetic on information contained in the end of year ladder. Thus instead of using complicated regression procedures on the individual match results, ability and HA effects are easily found using only a calculator and data obtained from the final ladder.

The procedure is as follows. Given a season's results in an N team competition, where each team plays the other $N-1$ teams once at home and once away, we can obtain measures u_i and h_i that describe each team's level of performance on a neutral ground and their home ground advantage.

(a) $H = \sum h_i = \sum \text{HGD}_i / (N-1)$ is the total of all the individual teams' home ground advantages; i.e. H is the total of the teams HGD column, divided by $N-1$. In the ladder of Table 1, $H = 297/21 = 14.14$.

(b) For each team, the HA $h_i = (\text{HGD}_i - \text{AGD}_i - H)/(N-2)$; i.e. for each team, their HA is the difference in their Home and Away goal difference, less the total of all the teams HA, all divided by $(N-2)$. For example for Everton $h_1 = (38-7-14.14)/20 = 16.86/20 = 0.84$, and for Aston Villa $h_{22} = \{0 - (-34) -14.14\}/20 = 0.99$.

(c) For each team, the ability measure $u_i = \{\text{HGD}_i - (N-1) h_i \}/N$. In Everton's case $u_1 = (38-21 \times 0.84)/22 = 0.92$, and for Aston Villa $u_{22} = (0-21 \times 0.99)/22 = -0.95$.

The above equations could be explained quite simply to a layman by replacing each match result in the usual home and away grid with the expected or model result and using simple addition. This derivation is given in Appendix 3.

Note the source of the spurious HA is now clearly shown in (b) above. The difference in a team's home and away performance is given by $(N-2)h_i + \sum h_i$. The difference is made up of one component due to that individual team's HA, and a second due to the total of all the teams' HA's. Thus although a team does better at home than away, this may be due to the collective advantage enjoyed by the other teams.

The final two columns of Table 1 show the results for 1986 Division 1. Although all teams do better at home than away, the sum of the HA's of all teams is 14.14 goals. Teams with 14 goals or less difference in their home and away performance will consequently have a negative HA. For Norwich, Wimbledon and Chelsea their better home than away performances are spurious and due entirely to the HA of the other teams. The u's have a range of about 2, so their difference has a range of about 4, whereas the h's have a range of about 1.4. This implies that in equation (1) ability is

about 3 times more important than HA in determining goal difference. For the above ladder, the correlation between actual ladder position and the ladder position determined by $u_i + kh_i$ is best for k of about 0.5.   As u affects a team's performance every match, and h only for half the matches this is perhaps not unexpected.

3. Data and results

Data was collected for all English soccer matches from season 1981-82 to season 1990-91, comprising 920 teams and 20,306 matches. The Official Football Association Yearbook published by Penguin contains a summary of the previous year's match results and final ladders for each division. Individual results were entered and a computer program used to produce the end of year ladders, which were checked with those published. This often showed up about a 1% error rate in the actual match results (ie about three results per year per division were incorrectly reported). Results were checked with the newspapers if necessary until agreement with the ladder was obtained. All computing work was performed with SAS.

The home teams won 9894 (48.7%) and drew another 5415 (26.7%) of their matches. Of the total 54,378 goals the home team scored 32,556 or 59.9%, which is very close to but just under the percentage of wins 48.7 + 0.5x26.7 = 62.1%. This may suggest that HA factors are slightly better at producing wins than larger margins. The proportion of wins, draws and losses was remarkably consistent across divisions, with a chi square test for independence of results and division producing a p value of 0.949.

However our main interest here is in calculating individual HAs. Using the methods shown above the HA's were calculated for all teams playing from 1980-81 to 1990-91, and are given in Table 2. The table is sorted in order of decreasing average HA, which is shown in the last column of the table. By looking at the HA's of individual clubs we may discover the mechanism behind HA.

TABLE 2
Home Ground Advantages (goals per match) for all teams in English soccer, 1981-2 to 1990-91, in order of decreasing average home advantage.

Rank	TEAM	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	Aver
1	Plymo	0.36	0.72	1.38	0.50	1.03	0.84	1.22	1.08	0.38	1.59	0.911
2	Alder	0.43	0.76	0.79	0.93	1.52	0.86	1.14	0.75	0.58	0.73	0.849
3	Maids	.	.	.	.	.	.	.	.	1.31	0.32	0.816
4	Luton	0.31	0.20	-0.22	1.33	0.75	0.79	1.60	1.44	0.94	0.89	0.803
5	Exete	1.31	0.81	-0.35	0.70	-0.03	1.18	0.54	1.63	1.18	1.00	0.798
6	Leeds	0.99	-0.26	1.06	0.94	0.96	1.39	0.75	0.26	0.83	0.94	0.788

7	Leice	0.21	0.09	0.58	1.13	0.50	1.89	0.79	0.72	0.51	1.28	0.770
8	Evert	0.49	0.65	0.23	1.18	0.70	0.84	0.60	1.06	1.55	0.39	0.768
9	Carli	0.95	1.09	0.56	0.74	0.76	0.47	1.27	-0.55	1.08	1.27	0.765
10	Colch	0.52	1.35	1.38	0.25	0.84	1.09	-0.23	0.68	0.81	.	0.743
11	Millw	0.27	1.36	1.11	0.96	0.91	0.79	0.56	0.61	0.71	0.00	0.728
12	Bolto	0.91	0.74	1.11	1.19	0.43	0.33	1.13	0.93	0.20	0.27	0.724
13	Soton	1.24	0.75	0.73	-0.02	1.35	1.24	-0.13	0.33	0.49	1.22	0.720
14	Orien	0.56	0.86	0.83	0.05	0.29	0.77	1.04	1.59	0.06	1.14	0.719
15	Stock	0.89	1.12	0.29	1.65	0.34	0.68	0.13	0.32	0.86	0.87	0.714
16	Queen	1.41	0.84	0.43	1.38	1.25	0.49	0.71	0.11	-0.01	0.44	0.705
17	Hartl	0.43	1.26	1.29	0.47	1.20	-0.05	-0.19	0.68	1.58	0.32	0.700
18	Bourn	0.02	1.00	0.65	1.32	0.57	0.60	0.56	0.90	0.29	1.09	0.700
19	Stoke	0.44	1.00	1.08	0.38	0.76	0.79	0.89	0.81	0.42	0.37	0.694
20	York	1.21	1.62	0.52	0.69	1.39	1.06	-0.18	0.91	-0.14	-0.27	0.679
21	BrisR	0.49	1.13	0.92	0.82	0.98	0.42	1.14	-0.16	0.52	0.46	0.673
22	Oldha	-0.04	-0.06	1.21	0.79	0.51	0.34	0.56	0.81	1.38	1.23	0.673
23	ShefU	0.52	1.50	1.15	0.39	0.51	0.54	0.70	0.84	0.06	0.44	0.666
24	Scunt	0.30	0.22	0.88	1.02	0.47	1.09	0.36	0.13	0.36	1.77	0.659
25	Gilli	0.81	-0.00	0.88	0.96	0.93	1.28	0.95	0.30	0.22	0.23	0.657
26	Rochd	0.39	1.35	0.93	0.34	1.34	0.50	0.27	1.13	0.08	0.14	0.646
27	BrisC	0.22	0.49	1.20	0.69	1.12	0.74	0.73	0.11	0.06	1.09	0.645
28	Brigh	0.64	0.95	1.36	0.49	0.56	0.44	0.41	0.90	0.33	0.32	0.641
29	Newca	0.91	0.49	1.01	0.78	1.15	0.59	0.39	0.39	0.65	0.00	0.636
30	Shrew	1.11	-0.11	1.11	1.14	0.86	0.99	0.08	0.08	0.56	0.50	0.633
31	Charl	1.26	1.44	1.26	0.49	0.61	0.19	0.50	0.56	0.16	-0.13	0.633
32	Bpool	0.48	1.03	1.02	0.34	0.57	0.19	0.73	0.21	0.25	1.41	0.622
33	Prest	0.08	1.27	0.56	0.96	0.11	0.04	0.64	0.80	1.25	0.50	0.620
34	Burnl	0.13	1.29	0.97	0.59	0.02	0.13	0.54	1.18	0.04	1.23	0.612
35	Rothe	1.11	-0.16	0.61	0.55	1.62	0.47	0.50	-0.09	0.93	0.59	0.611
36	MancU	-0.11	1.15	0.43	1.23	-0.00	0.94	0.34	0.67	0.88	0.56	0.608
37	Halif	0.25	0.40	1.84	0.11	0.56	0.22	0.81	1.04	0.04	0.73	0.600
38	Oxfor	-0.19	0.04	0.61	1.79	1.00	1.04	0.39	0.31	0.29	0.46	0.574
39	Walsa	0.54	0.91	0.88	0.14	1.43	1.01	0.23	-0.19	0.34	0.23	0.551
40	Soend	0.08	0.36	1.06	0.20	0.70	0.41	1.14	0.89	0.81	-0.18	0.547
41	WestH	0.44	0.65	0.68	0.58	0.90	0.54	0.39	-0.17	1.01	0.41	0.544
42	Ipswi	0.69	0.05	0.63	0.58	0.15	0.49	0.98	0.67	0.92	0.23	0.540
43	Bburn	0.31	1.09	0.36	0.54	0.81	0.69	0.18	1.26	0.01	0.09	0.535
44	Chfld	0.13	0.36	0.29	0.56	0.48	0.38	0.55	0.98	0.90	0.68	0.531
45	Coven	0.59	0.90	0.43	0.88	0.15	0.84	-0.29	-0.06	0.49	1.33	0.527

46	Swans	1.04	0.35	0.96	-0.04	1.34	0.68	-0.14	0.48	0.15	0.41	0.524
47	Torqu	0.48	0.49	0.79	0.02	1.11	0.18	-0.01	0.91	0.45	0.77	0.519
48	Arsen	0.24	0.75	-0.07	1.03	0.70	0.04	0.55	-0.61	1.60	0.83	0.507
49	Wigan	0.57	-0.05	0.38	0.91	1.16	0.38	0.27	-0.16	0.88	0.59	0.494
50	Hull	0.25	0.58	0.52	0.46	0.86	0.39	0.65	0.67	-0.53	1.05	0.490
51	Cambr	0.86	1.19	0.46	-0.72	0.65	0.91	0.31	0.82	0.40	0.00	0.488
52	PVale	-0.02	-0.24	1.02	0.70	0.79	-0.12	0.95	0.30	0.83	0.59	0.480
53	Wrexh	0.21	1.41	0.06	0.65	0.79	-0.05	0.81	0.23	0.22	0.46	0.479
54	Psmth	1.04	-0.09	0.26	-0.06	0.56	0.74	0.50	0.90	0.20	0.69	0.473
55	Heref	0.21	0.67	0.15	0.15	1.56	0.50	0.13	0.86	-0.32	0.77	0.469
56	Cryst	-0.14	1.24	-0.19	0.24	-0.19	0.99	0.89	0.67	1.27	-0.11	0.467
57	Wolve	1.04	0.79	0.78	0.19	-0.43	-0.23	0.27	1.07	0.74	0.41	0.463
58	Linco	0.54	1.04	0.52	0.59	0.25	0.72	.	0.41	-0.23	0.32	0.462
59	North	1.16	1.08	0.43	0.38	-0.25	0.54	0.27	0.48	-0.07	0.59	0.460
60	Hudde	0.49	1.09	-0.34	-0.01	0.86	0.54	0.51	0.71	0.20	0.55	0.460
61	Bradf	0.16	0.54	0.29	-0.09	1.16	0.29	0.65	0.36	0.79	0.46	0.460
62	Aston	-0.11	1.75	0.73	0.73	0.10	0.99	-0.92	0.61	0.21	0.50	0.460
63	Readi	0.63	0.72	1.38	-0.63	0.16	0.74	0.18	0.30	0.43	0.68	0.459
64	Watfo	0.71	1.20	0.38	0.63	0.85	0.44	-0.13	0.31	0.79	-0.59	0.458
65	Mansf	0.30	0.31	1.02	0.20	0.47	0.28	0.36	0.48	1.02	0.14	0.458
66	Scarb	.	.	.	.	.	.	0.86	-0.37	0.72	0.59	0.452
67	Grims	-0.09	1.24	0.76	0.59	0.66	0.09	-0.41	0.50	0.36	0.73	0.443
68	MancC	0.54	0.60	0.61	0.79	-0.00	0.74	0.56	0.13	0.55	-0.11	0.441
69	NottC	-0.11	1.15	0.03	0.19	0.30	1.10	0.64	0.30	0.61	0.09	0.430
70	Peter	1.02	0.44	1.47	0.11	0.97	-0.05	-0.28	0.13	-0.01	0.46	0.428
71	Derby	1.06	0.14	1.11	0.78	-0.16	0.49	0.18	0.11	0.21	0.28	0.421
72	Norwi	0.91	0.65	0.88	0.63	0.61	-0.11	0.18	-0.17	0.49	0.11	0.418
73	Swind	0.36	0.67	0.20	1.15	0.56	-0.22	0.51	0.54	0.42	-0.04	0.416
74	NottF	-0.21	0.15	0.88	0.63	-0.10	0.84	0.50	0.17	0.16	0.89	0.390
75	Tranm	-0.11	0.44	-0.07	1.29	-0.25	-0.09	1.22	0.45	0.75	0.27	0.389
76	Totte	0.14	1.90	0.18	-0.52	0.50	0.64	0.39	-0.33	0.05	0.78	0.372
77	WBA	0.04	0.55	0.63	0.83	0.60	-0.01	0.60	0.58	-0.49	0.37	0.371
78	Newpo	0.18	0.04	0.70	-0.04	0.12	0.47	1.13	.	.	.	0.370
79	Donca	0.77	0.50	0.20	0.37	-0.43	0.97	0.50	0.73	-0.19	0.18	0.359
80	Barns	0.41	0.19	0.11	0.99	-0.19	-0.16	0.41	0.49	0.42	0.91	0.359
81	Bury	0.98	0.17	-0.39	0.61	1.34	0.15	0.09	0.39	0.02	0.23	0.358
82	Liver	-0.31	0.55	1.18	-0.62	1.20	0.49	0.44	0.06	-0.23	0.56	0.332
83	Sunde	-0.11	0.60	0.78	-0.17	0.46	0.09	0.14	0.95	0.06	0.50	0.329
84	Chels	0.11	0.79	0.61	0.38	-0.15	-0.16	1.23	-0.42	-0.18	0.89	0.311

85	Middl	0.09	0.24	0.41	-0.11	0.11	0.28	0.79	0.39	0.38	0.37	0.296
86	ShefW	0.01	-0.26	0.51	0.48	0.25	0.84	-0.03	0.17	0.82	0.00	0.279
87	Fulha	0.36	0.14	0.06	0.19	0.26	-0.22	0.18	0.57	0.56	0.68	0.279
88	Brent	-0.32	0.72	0.65	0.91	-0.52	0.24	0.14	0.66	0.43	-0.27	0.264
89	Crewe	0.80	0.44	0.52	-0.03	0.47	-0.09	-0.14	0.18	0.11	0.37	0.262
90	Cardi	-0.19	0.86	0.46	-0.71	0.07	0.18	0.63	1.21	-0.07	0.14	0.257
91	Birmi	0.39	0.40	0.08	-0.31	0.35	0.59	0.22	0.36	0.75	-0.27	0.256
92	Darli	0.25	-0.47	0.43	0.38	0.66	0.69	0.68	-0.46	.	0.00	0.241
93	Chstr	0.18	-0.06	-0.12	0.43	0.38	-0.22	-0.09	1.07	0.70	0.09	0.236
94	Wimbl	0.72	0.12	0.33	0.79	0.51	-0.06	0.23	0.67	-1.01	-0.33	0.198
	Average	0.45	0.66	0.64	0.52	0.59	0.51	0.47	0.50	0.45	0.50	0.528

4. Discussion

The results show that a team's HA is quite variable from year to year and that in some years some teams have a negative HA. In fact 126 of 920 or about 14% are negative - in any one division in any year about 3 teams actually have a negative HA. Because of the inherent variation in soccer matches, an average over a number of years is necessary to obtain a reasonable measure of HA.

On average the home ground advantage is worth just over 0.5 of a goal, and that is amazingly constant over the divisions (0.521, 0.529, 0.529, 0.533 for division 1 to 4). The general linear models framework was used to perform various ANOVA tests which indicated that the year is significant, division is not, and differences between the clubs were only borderline significant. For example a test on all the data, 920 values gave for year effect $p = 0.014$, division effect $p = 0.990$ and club effect $p = 0.085$. The residuals passed the usual tests for normality, and the R^2 for this model was 14% so there is lots of variation in HA. It is clear from these results that

1. There is no division effect. This is in contrast to the results of Pollard (1986) and seems to negate the crowd factor as a cause of HA.

2. There is a highly significant year effect. The HA is about 10% higher than average in 82, 83 and 85, and 10% lower in 81, 87 and 89.

3. There is some evidence for a significant club effect, but this is not conclusive. Certainly the club effect is weaker than the year effect.

4.1 Special Clubs

Pollard (1986) singled out 5 clubs for special attention when looking at the effect of local conditions: Bristol Rovers and Halifax (small pitch), Manchester City and Carlisle (large pitch), Queens Park Rangers (artificial turf). Pollard found the points gained at

home were not significantly different for these clubs. However as we argued above this would be affected greatly by the relative strengths of those clubs. Table 2 shows that most of them are in the top third - with ranks 9, 16, 21, 37 and 68 in a table with 94 values. A rank sum test gives $R = 151$ which has a p value of 0.076 which is some evidence for these teams having a higher than average HA.

What about the 13 London clubs - Millwall (ranked 11), Orient (14), QPR (16), Charlton (30), West Ham (41), Arsenal (48), Crystal Palace (56), Watford (64), Tottenham (76), Chelsea (84), Fulham (86), Brentford (88) and Wimbledon (94)? Because of their close proximity we might expect them to have low HA's, and there are 4 in the bottom 11 rankings. Again, a rank sum test gives $R = 708$ with a p value of 0.161. Since QPR has already been singled out as having special properties that may give it a large HA, it might be argued that we should exclude it in this analysis. Doing this gives $R = 692$ which has a p value of 0.072. So again there is some evidence that the London clubs have a lower than average HA.

Barnett and Hilditch (1993) looked specifically at the effect of artificial pitch on HA. The above table confirms their finding of an artificial pitch effect. Queens Park Rangers 1981/82-87/88, Luton 1985/86-89/90, Oldham 1986/87-90/91 and Preston 1986/87-90/91 all had artificial pitches. The 22 seasons played on an artificial pitch had a mean HA of 0.889, compared with 0.519 for the other 898 seasons - significant at the 0.1% level. As this may be due to a year or team effect, including type of pitch along with year, division and club in an ANOVA test showed type of pitch to be significant at the 1% level ($p = 0.0013$).

4.2 HA versus time in division

A possible reason often advanced for HA is the home teams familiarity with the 'quirks' of their home ground. Alternatively we could argue the visiting teams are unfamiliar with the home team's facility. If this was so one would expect the HA to be greatest when a team is new to the division. To test this the current length of time continuously in the division for all teams was calculated. The results were the reverse of that expected, with a small nonsignificant positive correlation between HA and years in division. Perhaps when teams are new to the competition the opposition put effort into counteracting their peculiarities, but relax this effort after they (mistakenly) believe they are familiar with the opposition. To make sure this was not due to a year effect, averages by year and continuous time in divisions were looked at and the results confirmed the finding. For 0 or 1 years in the division, most of the averages were below the yearly average, whereas for 2 to 3 years in the division, most of the averages were above the yearly averages. Contrary to expectation, the teams that are new or have been in the division only one year do not appear to have higher than average HA's.

5. Paired home advantage

The arguments advanced earlier for going from a competition level to a club level can be extended one stage further. Just as the 'competition' level HA is an average of the HA's of all the clubs, so an individual club's HA is an average of its paired HA with all the other clubs it plays. For example, suppose HA is due entirely to distance travelled. A particular club would travel a short distance to some clubs (with no HA) and a long distance to others (with a consequent large HA). As its HA is an average of these it would have an average HA and the effect of distance would be lost. Thus the HA of one club is really the average of all its paired HA's with the HA's of the other clubs removed. It includes matches with nearby clubs, far clubs etc and so to some extent averages out the effects (of distance, crowd etc). Can we get a more refined measure by looking at the paired HA?

Stefani and Clarke (1992) state that the home ground advantage can be thought of as h_{ij}. For each pair of teams the difference in home and away matches - in our previous notation $w_{ij} + w_{ji}$ which is equivalent to $h_i + h_j$ - gives a measure of this for each year. Note that we need the actual match results for this - it cannot be calculated from the ladder. For our data, this gave 10153 match pairs, played between 2865 club pairs, with an average of 1.057 paired HA. This agrees with the previous estimate that a HA is worth about 1/2 a goal, as the paired HA incorporates two individual HA's. However the values are highly variable, ranging from -7 to +11.

To investigate if distance had an effect, the grid coordinates of the home ground of each club on a map were estimated and the straight line distance between each pair calculated. The correlation between distance and HA was 0.07 - very small but because of the great number of observations is highly significant with $p = 0.0001$. The low correlation is due to the high variation in the individual data, making it difficult to explain high proportions of the variation. Several averaging methods were tried to reduce the variation, and all showed a clear relationship between paired HA and distance. For example, by averaging the paired HA's for each of the 2865 club pairs the correlation between average paired HA and distance became 0.11. By restricting the analysis to the 1303 pairs of teams which played each other for 4 seasons or more, the correlation increased to 0.14.   The pairs of clubs were separated into groups in multiples of 50 miles apart and the average paired HA was calculated for each group. The results are shown in Fig. 1 and clearly show increasing paired HA for increasing distance. To some extent this effect is reflected in Table 2. One referee pointed out that 3 clubs in the top ten, Plymouth, Exeter and Carlisle, are geographically isolated.
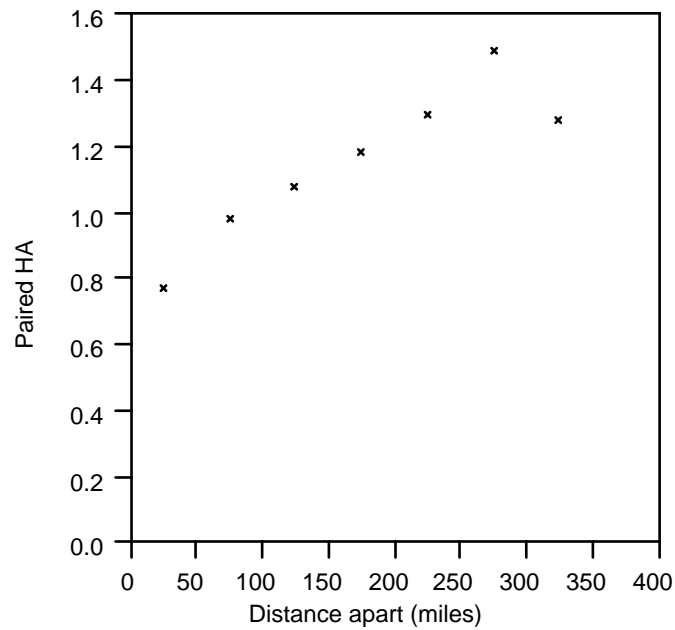
Fig. 1. Average paired HA (goals) versus distance apart of clubs (miles).

6. Win/loss home advantage

In the above analysis we have used goal difference to measure a team's performance. However the analysis can be repeated using win (or point) margins as the measure of performance. Although goal difference should be a more sensitive measure than wins, it may be that HA works to produce wins rather than large margins. Replacing a team's score by 1 for a win, 0.5 for a draw and 0 for a loss produces win margins of 1, 0 and -1. The analysis can be repeated exactly, but the measures obtained would now be in terms of win margins rather than goal margins. Alternatively, using 3,1 and 0 points produces point margins of 3, 0 and -3, but this is only an exact multiple of the win/lose case.

Using win margins produced similar results to the above, with a tendency to produce slightly more significant results. For example the overall average HA is 0.472, or nearly 1/2 of a win. The ranks of the 5 clubs singled out for special characteristics by Pollard (1986) now go to 4, 9, 11, 41 and 80 with a rank sum 145 now significant with $p = 0.06$. The ranks of the 12 London clubs (without QPR) are 28, 31, 34, 40, 47, 70, 72, 75, 81, 83, 91, 93. This gives a rank total of 745 significant with $p = 0.02$. The fact that both these have moved in the direction indicating enhanced HA suggests that HA may have a greater effect on winning than goal difference. Thus whatever it is that produces HA tends to operate more effectively in determining winners rather than just larger winning margins.

7. Conclusion

At a competition level, variations in percentage of home matches won may arise because of differences in team ability as well as variations in HA. To calculate HA at a club level, we need to take account of team ability, by looking at the difference in home and away performances. Least squares can be used on the match results to estimate team and HA effects. However for a balanced competition such as English soccer, this is equivalent to simple calculation methods on the final ladder results.

Using 10 years data we have calculated HA in terms of goal and win difference for all 94 clubs in English soccer. These showed no division effects but significant year effects. There was some evidence that clubs with special facilities have significantly higher HA, and that London clubs have less than average HA. There was no evidence that clubs new to a division have higher HA. It also appeared that HA effects have more leverage on winning than goal margins.

Paired HA is a more sensitive measure of HA, but individual match results are needed for its calculation. A definite linear relationship exists between a pair of clubs paired HA and their distance apart.

Appendix 1. Spurious and real home ground advantage

Consider 3 teams, A, B and C. Suppose A is better than B which is better than C, and there are no home ground advantages. Suppose both home and away A beats B 2-1 and C 3-1, while B beats C 2-1. Final results would be

		Away Team		
		A	B	C
Home	A	-	2-1	3-1
Team	B	1-2	-	2-1
	C	1-3	1-2	-

with final ladder

Team	Home				Away				Home-Away	
	wins	draws	losses	goals	wins	draws	losses	goals	wins	goals
A	2	0	0	5-2	2	0	0	5-2	0	0
B	1	0	1	3-3	1	0	1	3-3	0	0
C	0	0	2	2-5	0	0	2	2-5	0	0

Obviously each team has the same home performance as away both in terms of wins and goals.

However we now give C a 2 goal home ground advantage so that C will perform 2 goals better at home than anywhere else. Thus at home it will draw against A and beat B 3-2. The results and end of year ladder now look like

		Away Team		
		A	B	C
Home	A	-	2-1	3-1
Team	B	1-2	-	2-1
	C	3-3	3-2	-

	Home				Away				Home-Away	
Team	wins	draws	losses	goals	wins	draws	losses	goals	wins	goals
A	2	0	0	5-2	1	1	0	5-4	0.5	2
B	1	0	1	3-3	0	0	2	3-5	1	2
C	1	1	0	6-5	0	0	2	2-5	1.5	4

The final ladder shows that even though only C has a HA, all teams had better results at home than away, both in terms of wins and goal difference.

A 'naive' analysis of goal difference would incorrectly conclude that each team had a home ground advantage - A and B performing better at home than away over the season by a total of 2 goals, while C performed better by a total of 4 goals.

Appendix 2. Derivation of formula for calculation of home advantage and team performance using least squares

Let w_{ij} be the winning margin for home team i against away team j, (negative if loss). For N teams, this gives an $N \times N$ matrix with no diagonals. Adding across a row gives the home goal difference (HGD), while adding down a column gives the negative of away goal difference (AGD).

ie. for team I,
$$HGD_I = \sum_{j=1(j\neq I)}^{j=N} w_{Ij} \, , \qquad AGD_I = - \sum_{i=1(i\neq I)}^{i=N} w_{iI}$$

Thus since we are merely summing all the w_{ij} in a different order

$$\sum_{i=1}^{i=N} HGD_i = - \sum_{i=1}^{i=N} AGD_i$$

If u_i is a measure of team ability, rating, or skill level etc of team i and h_i is the home ground advantage of team i, and ε_{ij} is random error, then as before in equatiion (1) we model the winning margin by

$$w_{ij} = u_i - u_j + h_i + \varepsilon_{ij} \tag{1}$$

Since only differences of the u_i are used, they are relative, and we make the arbitrary restriction that $\sum_{i=1}^{i=N} u_i = 0$. So minimising the sums of squares of the errors subject to this condition, we have using the usual Lagrange multiplier expression.

$$\text{Minimise} \quad S = \sum_{i=1}^{i=N} \sum_{j=1(j\neq i)}^{j=N} (w_{ij} - u_i + u_j - h_i)^2 + \lambda \sum_{i=1}^{i=N} u_i$$

In the normal manner, partial differentiating with respect to u_I, $I = 1$ to N; h_I, $I = 1$ to N; and λ we get $2N + 1$ equations.

$$\sum_{j=1(j\neq I)}^{j=N} 2(w_{Ij} - u_I + u_j - h_I)(-1) + \sum_{i=1(i\neq I)}^{i=N} 2(w_{iI} - u_i + u_I - h_i) + \lambda = 0, I=1 \text{ to } N \tag{2}$$

$$\sum_{j=1(j\neq I)}^{j=N} 2(w_{Ij} - u_I + u_j - h_I)(-1) = 0, \quad I = 1 \text{ to } N \tag{3}$$

$$\sum_{i=1}^{i=N} u_i = 0 \tag{4}$$

Expanding (3) gives $\quad \displaystyle\sum_{j=1(j\neq I)}^{j=N} w_{Ij} = (N-1)\, u_I + (N-1)\, h_I - \sum_{j=1(j\neq I)}^{j=N} u_j$

$$\text{i.e. HGD}_I = N\, u_I + (N-1)\, h_I + \sum_{j=1}^{j=N} u_j$$

$$\text{HGD}_I = N\, u_I + (N-1)\, h_I \tag{5}$$

So adding for $I = 1$ to N,

$$\text{i.e.} \quad \sum_{I=1}^{I=N} \text{HGD}_I \quad = N \sum_{I=1}^{I=N} u_I \; + (N{-}1) \sum_{I=1}^{I=N} h_I$$

$$\text{HGD} \quad = (N{-}1) \; H \tag{6}$$

where $H = \sum_{i=1}^{i=N} h_i$ is the total of all the individual team's home ground advantages.

From (2), Substituting (3) eliminates the first summation term, so

$$-\lambda/2 \quad = \sum_{i=1(i\neq I)}^{i=N} (w_{iI} - u_i + u_I - h_i)$$

$$= \sum_{i=1(i\neq I)}^{i=N} w_{iI} \quad - \sum_{i=1(i\neq I)}^{i=N} u_i \quad - \sum_{i=1(i\neq I)}^{i=N} h_i \quad + (N{-}1) \, u_I$$

$$= \; -\text{AGD}_I + u_I - H + h_I \; + (N{-}1) \, u_I$$

$$-\lambda/2 \quad = \; -\text{AGD}_I \; - H + h_I \; + N \, u_I \tag{7}$$

$$\text{So} \quad \sum_{I=1}^{I=N} -\lambda/2 \; = \sum_{I=1}^{I=N} -\text{AGD}_I - NH + \sum_{I=1}^{I=N} h_I \; + N \sum_{I=1}^{I=N} u_I$$

$$-N\lambda/2 \quad = \sum_{I=1}^{I=N} \text{HGD}_I \; - (N{-}1) \, H + 0$$

$$= \; 0 \text{ from (6)}$$

So $\lambda = 0$ and (7) becomes

$$\text{AGD}_I \quad = - H + h_I \; + N \, u_I \tag{8}$$

So subtracting (8) from (5)

$$\text{HGD}_I - \text{AGD}_I \quad = N \, u_I + (N{-}1) \, h_I + H - h_I - N \, u_I$$

$$\text{HGD}_I - \text{AGD}_I \quad = H + (N{-}2) \, h_I \tag{9}$$

Thus H is calculated from (6), h_i from (9) and u_i from (5).

Appendix 3. Derivation of formula for calculation of home advantage and team performance from final ladder using simple explanation

As before we model the winning margin by

$$w_{ij} = u_i - u_j + h_i$$

The error term is neglected for simplicity. It could be included, and discarded later under the assumption that it sums to zero.

Adding across row i gives the home ground performance of team i as

$$\text{HGD}_i = \sum_{j=1(j\neq i)}^{j=N} w_{ij} = \sum_{j=1(j\neq i)}^{j=N} (u_i - u_j + h_i) ,$$

$$= (N-1)u_i - \sum_{j=1(j\neq i)}^{j=N} u_j + (N-1)h_i$$

$$= Nu_i - \sum_{j=1}^{j=N} u_j + (N-1) h_i$$

Now as the u_i are relative, and it is only the difference that matters, we can require that they sum to zero.

$$\text{HGD}_i = Nu_i + (N-1) h_i \quad \text{as in (5) from Appendix 2.}$$

If we sum all the home performances for the whole competition we obtain

$$\sum_{i=1}^{i=N} \text{HGD}_i = \sum_{i=1}^{i=N} \{Nu_i + (N-1)h_i\} = (N-1) \sum_{i=1}^{i=N} h_i$$

$$= (N-1) \ H \ \text{as in (6) from Appendix 2}$$

In a similar manner, a team's away performance is obtained by adding up the negatives of a column. For column j we have

$$\text{AGD}_j = \sum_{i=1(i\neq j)}^{i=N} -w_{ij} = \sum_{i=1(i\neq j)}^{i=N} (-u_i + u_j - h_i)$$

$$= - \sum_{i=1(i\neq j)}^{i=N} u_i + (N-1)u_j - \sum_{i=1(i\neq j)}^{i=N} h_i$$

$$= - \sum_{i=1}^{i=N} u_i \ + Nu_j \ - \ H \ + \ h_j$$

$$= \ Nu_j \ - \ H \ + \ h_j \quad \text{as in (8) from Appendix 2.}$$

The difference between home and away performance for any team now becomes

$$\text{HGD}_i - \text{AGD}_i \ = \ Nu_i \ + (N{-}1)\, h_i \ - Nu_i \ + \ H \ - \ h_i$$

$$= \ H \ + (N{-}2)\, h_i \quad \text{as in (9) from Appendix 2.}$$

References

Barnett, V. and Hilditch, S. (1993). The Effect of an Artificial Pitch surface on Home Team Performance in Football (Soccer). *Journal Royal Statistical Society A*, **156** (1), 39-50.

Clarke, S. R. (1993). Computer Forecasting of Australian Rules football for a daily newspaper. *Journal Operational Research Society*, **44** (8), 753-759.

Courneya, K. S. and Carron, A. V. (1992). The Home Advantage in Sport Competitions: A Literature Review. *Journal of Sport & Exercise Psychology*, **14**, 13-27.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sport Sciences*, **4**, 237-248.

Snyder, E. E. and Purdy, D. A. (1985). The home advantage in collegiate basketball. *Sociology of Sport Journal*, **2**, 352-356.

Stefani, R. (1983). Observed betting tendencies and suggested betting strategies for European football pools. *The Statistician,* **32**, 319-329.

Stefani, R. (1987). Applications of statistical methods to American football. *Journal of Applied Statistics*, **14**, 61-73.

Stefani, R. and Clarke, S. (1992). Predictions and home advantage for Australian rules football. *Journal of Applied Statistics*, **19** (2), 251-261.