



On referee bias, crowd size, and home advantage in the English soccer Premiership

Ron Johnston

To cite this article: Ron Johnston (2008) On referee bias, crowd size, and home advantage in the English soccer Premiership, Journal of Sports Sciences, 26:6, 563-568, DOI: [10.1080/02640410701736780](https://doi.org/10.1080/02640410701736780)

To link to this article: <https://doi.org/10.1080/02640410701736780>



Published online: 20 May 2008.



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



Article views: 1423



View related articles [↗](#)



Citing articles: 27 View citing articles [↗](#)

On referee bias, crowd size, and home advantage in the English soccer Premiership

RON JOHNSTON

School of Geographical Sciences, University of Bristol, Bristol, UK

(Accepted 9 October 2007)

Abstract

In a recent paper in this journal, Boyko and colleagues (2007) identified differences in attendance and referee bias as factors influencing home advantage at soccer matches in the English Premiership. A replication of their study using more recent data found no evidence to sustain either of their claims.

Keywords: *Soccer, English Premiership, home advantage*

Introduction

There is a large literature reporting analyses of the extent of home advantage in a range of sports (for recent reviews, see Carron, Loughhead, & Bray, 2005; Pollard & Pollard, 2005), including association football (soccer). Many ways of modelling that advantage taking other variables into account have been presented. For soccer, a model proposed by Clarke and Norman (1995) – in which the core independent variables are the ability of the home and away sides – has been followed by many subsequent authors. The model has been extended in a number of ways, including, for example, the influence of crowd size and variations among referees on the extent of home advantage, as in a recent paper by Boyko and colleagues (Boyko, Boyko, & Boyko, 2007), who presented analyses of results in the English Premiership. This note is a response to that paper.

Boyko and colleagues' (2007) model is conceptually similar to that of Clarke and Norman, modified (as they indicate) to take account of the "imbalance" in individual referees' schedules. (Clarke and Norman were able to assume that each home and away side's ability was constant across an entire season because all pairs of clubs play each other both home and away. Referees' schedules are not balanced, however; the number of matches at which they officiate varies, as does the relative abilities of the clubs whose games they cover. Consequently, Boyko and colleagues introduced team ability metrics in their regression model – an approach that is

replicated here.) Their results indicated not only that the larger the attendance at a match the greater the home side's advantage (in terms of goals scored), but also that there were statistically significant differences (at the 0.05 level or better) among referees in the extent of that home advantage. They concluded that "individual referees influence the outcome of English Premier League matches... [and] are likely responsible for at least some of the observed home advantage in EPL football" (pp. 1191–1192), probably because whereas most referees are associated with substantial home advantage, "a few deviate by giving lower home advantage" (p. 1192). Together with their finding relating home advantage to crowd size, this, they suggest, reflects one or more of three possibilities: (1) referees vary in their response to a partisan crowd, but not in their marginal response to crowd size; (2) referees vary in their marginal response to crowd size, but not in their baseline response to the presence of a partisan crowd; and (3) referees differ in both their baseline and marginal response to crowd size.

Not surprisingly, these results generated considerable popular interest.¹ However, some aspects of the data used raise questions about the robustness of the findings, and so it was decided to replicate their study using data from the most recent (2006–2007)

¹For example, stories at <http://www.eufootball.biz/Competitions/011106-Study-referees-favour-home-teams.html> and <http://www.midcomb.com/MIDLAND%20REF.htm>; and in the *Daily Mail* newspaper, http://www.dailymail.co.uk/pages/live/articles/technology/technology.html?in_article_id=453087&in_page_id=1965.

season. All of the data have been taken from the website <http://soccerfactsuk.co.uk/>

Boyko and colleagues' analyses relate home advantage to crowd size and referee through a core model with the following form:

$$HA_{ij} = a + b_1HG_i - b_2AG_j + b_3A_{ij} \pm \Sigma b_k R_k$$

where:

HA_{ij} is the home advantage in the match between teams i (the home team) and j (the away team), defined as the goal difference $[G_i - G_j]$, where G_i and G_j are the goals scored by teams i and j , respectively; HG_i is the mean number of goals scored by team i in home matches, excluding the match between teams i and j ;

AG_j is the mean number of goals scored by team j in away matches, excluding the match between teams i and j ;

A_{ij} is the attendance at the match between teams i and j , where i is the home team; and

R_k is a series of dummy variables for referee k (contrasted with the referee having the highest mean value of HA_{ij} for the season).

Fitting this model to all 380 matches played during the 2006–2007 Premiership season produced the following regression equation (standard errors and probabilities in parentheses):

$$HA_{ij} = 0.616 + 1.202HG_i - 1.646AG_j + 0.007A_{ij}$$

(0.437)	(0.210)	(0.229)	(0.006)
(0.150)	(0.000)	(0.000)	(0.270)

$$R^2 = 0.235$$

Only the coefficients for the mean of home and away goals scored were significantly different from zero at the 0.05 level or better; in addition, there were only two statistically significant (both negative) regression coefficients at the 0.05 level or better for individual referees (these are not shown here). This indicates, in line with Boyko and colleagues' findings, that the greater the mean number of goals that the home team scores per home match, the greater its home advantage, and the larger the mean number of goals that the away team scores per away match, the lower its opponent's home advantage. There is, however, no significant relationship with attendance at the match and only two of the 19 referees had a significantly different mean HA_{ij} , not only from the contrasted referee but also, by implication, from all of the others. (The two referees with significant coefficients officiated at only a few matches: see below.)

These findings cast some doubt on the wider applicability of Boyko and colleagues' findings,

therefore. The results indicate that the only substantial determinants of the results of English Premiership matches in the 2006–2007 season were the relative strengths of the home and away teams. Variations in crowd size did not systematically influence outcomes and there was little substantial evidence of variations in the extent of home advantage according to which referee was officiating. Before accepting that conclusion, however, some aspects of the data – especially with regard to crowd size – are explored further.

Crowd size and home advantage

Boyko and colleagues' argument is that the larger the attendance at a match – which they assume, implicitly, means a larger crowd supporting the home team – the better the home team's performance. There are two problems with this in the English Premiership, however. First, many of the clubs perform before a (virtual) capacity crowd at every home match, which means that there is no variation in the independent variable (A_{ij}). Second, among the teams which do not attract a capacity crowd for every home match, there is a tendency for the largest attendances to be for matches against the strongest teams in the league (which usually perform well away from home). This suggests a negative and not a positive relationship between crowd size and home advantage.

Table I illustrates the first of these points with data on the minimum, maximum, mean, and standard deviation attendance at home matches for each of the 20 teams involved. Not only did they vary substantially in their mean (from 18,750 to 75,826) and maximum attendance (19,830 to 76,098), they also differed substantially in the variation around the means. Eight of the teams – Arsenal, Chelsea, Liverpool, Manchester United, Portsmouth, Reading, Tottenham, and West Ham – played, in effect, before a capacity crowd at all of their home matches, and a further four – Bolton, Charlton, Newcastle, and Watford – had standard deviations around their mean attendance of less than 2000. Such teams should probably be excluded from the analyses, since they have no variation in one of the variables. (Of the eight in the first group, only one – West Ham – did not occupy one of the top ten places at the end of the season, so mean crowd size tended to be greatest for the more successful teams. Charlton and Watford were both relegated; Watford had only been promoted to the Premiership – for the first time – at the start of the season and, as is usual for such clubs, attracted large attendances to the club's relatively small capacity ground.

The second point concerns other variables possibly related to crowd size for teams which do not attract a full attendance for every home match. One of those is

Table I. Summary statistics for matches in the English Premiership during the 2006–2007 season.

Team	LP	Home advantage		Attendance			
		H	A	Min.	Max.	Mean	s
Arsenal	4	1.37	−0.06	59,912	60,132	60,045	70
Aston Villa	11	0.32	0.21	27,450	42,551	36,214	5015
Blackburn	10	0.37	0.21	16,035	29,342	21,275	3828
Bolton	6	0.32	0.58	21,140	27,229	23,605	1877
Charlton	19	−0.05	1.31	23,423	27,111	26,193	1088
Chelsea	2	1.32	−0.74	38,000	41,953	41,541	916
Everton	5	0.84	−0.05	32,968	40,094	36,743	2295
Fulham	16	0.00	1.16	17,000	24,554	22,279	2440
Liverpool	3	1.68	0.11	41,370	44,403	43,561	928
Manchester City	14	−0.32	0.42	35,776	47,224	39,997	3012
Manchester United	1	1.79	−1.16	75,115	76,098	75,826	255
Middlesbrough	12	0.37	0.63	23,638	32,013	27,729	2601
Newcastle	13	0.16	0.63	48,145	52,305	50,686	1576
Portsmouth	9	0.68	0.37	19,105	20,223	19,862	361
Reading	8	0.47	0.21	21,954	24,122	23,829	543
Sheffield United	18	0.00	1.37	25,011	32,752	30,521	2298
Tottenham	4	0.63	0.47	34,154	36,170	35,739	509
Watford	20	−0.32	1.26	13,766	19,830	18,750	1356
West Ham	15	−0.11	1.26	33,805	35,000	34,722	374
Wigan	17	−0.58	0.53	14,636	24,726	18,158	2978

Note: LP = position in the league at the end of the season; H = mean home advantage at home matches; A = mean home advantage at away matches.

the strength of the opposition. (Others could include the day of the week on which the game is played, each team's position in the league, and when during the season the match was held. Boyko and colleagues did not include these, and as the purpose of this note is only to evaluate their findings on a further data set, they are not included here either.) In general, strong away teams attract a larger attendance than weak ones. Four teams dominated the Premiership in the six seasons leading up to 2006–2007: Arsenal, Chelsea, Liverpool, and Manchester United were never out of the top six places at the end of the season and three of them (excluding Liverpool) won the Premiership on at least one occasion. If we analyse the mean attendance at the home matches for the 12 teams that did not attract a full attendance for every match – as a percentage of their maximum attendance during the season to hold constant the substantial differences in ground capacity – we find that for home matches against one of the “top four” teams the mean attendance was 94% of the maximum, whereas for matches against other teams it was only 86%. Teams attracted larger attendances when they were playing stronger opponents, which, as Table I shows, averaged the largest values of HA_{ij} when they were playing away from home. This suggests that the relationship between attendance and HA_{ij} for those teams that experienced variations in crowd size should be negative.

Table II explores these differences by fitting the basic model – without including referees – to three separate populations: all matches; only those home

Table II. Regression analysis of home advantage (standard errors and significance levels in parentheses).

Model	I	II	III
Constant	0.142 (0.343: 0.678)	0.469 (0.525: 0.373)	0.588 (0.644: 0.363)
Home goals	1.140 (0.198: 0.000)	0.953 (0.286: 0.001)	0.894 (0.288: 0.002)
Away goals	− 1.618 (0.225: 0.000)	− 1.793 (0.278: 0.000)	− 1.987 (0.322: 0.000)
Attendance (in 000s)	0.008 (0.006: 0.169)	0.010 (0.009: 0.273)	0.017 (0.013: 0.210)
R^2	0.234	0.182	0.228
N	380	228	152

Key to models: I = all matches; II = matches excluding teams that played to full-capacity crowds at all games; III = matches excluding teams that played to near-full-capacity crowds at all games.

Note: Coefficients significantly different from zero at the 0.05 level or better are shown in **bold**.

matches for the 12 teams whose grounds were not virtually full on all occasions; and those matches for the eight teams whose attendances varied most over the season. They show a consistent pattern. First, no matter which teams are included in the analysis, there is no evidence that variations in attendance have any impact on the outcome. Second, the stronger the home team the greater the home advantage, and the stronger the away team the smaller that advantage. But the relative importance of those statistically significant coefficients varies. When all matches are included, the coefficient for the home team's strength is 1.140; for every extra goal

that it scores on average in home matches, its advantage over its opponent increases by 1.14 goals. But if we exclude the eight teams that normally play before capacity crowds, this advantage is reduced to 0.953 goals, and if the additional four teams that usually have full grounds are also excluded, it falls further to 0.894. Countering this, the stronger the away team, the smaller the home team's advantage, and the size of this impact increases as the number of teams included declines (from -1.618 through -1.793 to -1.987). Since the clubs with the greatest variations in attendance tend to be among the weaker teams (only Charlton and West Ham are exceptions to this), they are the ones that suffer most when the stronger teams visit them.

Referees and home advantage

Twenty referees were used during the 2006–2007 Premiership season. Most of them officiated at 20 or more matches but four (Marriner, Probert, Stroud, and Tanner; see Table III) were involved in only a few. Those four officiated at matches with relatively small mean attendances (as a percentage of the maximum); only one other referee (Gallagher, who officiated at the next smallest number of matches, i.e. 12) had a mean attendance below 90%.

Those 20 referees differed substantially in the mean home advantage goal difference for the matches they officiated at, ranging from 0.96 for Webb to -2.0 for Stroud (who refereed one match

only). Apart from the three referees who officiated at only one or two matches, every referee had a positive coefficient – that is, the home team outperformed the away team – but the differences shown in Table III provide circumstantial evidence to sustain Boyko and colleagues' argument that some referees apparently were more favourably inclined towards the home team than others. This, however, does not take account of variations in the relative strength of the home and away teams at the matches they refereed; some might have officiated at more matches involving the stronger home teams than others.

To determine whether there was circumstantial evidence that any referees did have a significant impact on the outcome of the matches they officiated at, the regressions in Table II were rerun with a series of dummy variables incorporated to represent each referee – with Webb (the referee with the highest mean HA_{ij} – Table III) as the comparator. The dummy variables thus assess whether each referee significantly differed in his mean HA_{ij} – that is, whether the matches he refereed had either a significantly smaller or larger home advantage outcome than those refereed by Webb, holding constant each team's home and away strength and match attendance.

The first regression in Table IV includes all matches. The strongest relationship is with away goals; the better the away team's mean performance when away from home, the smaller the home advantage. There is also a significant positive link with attendance – larger attendances produce larger home advantage outcomes – and a weak positive link with the home team's mean performance. (Recall that there is some collinearity between these two variables; the teams with the largest mean attendances tended to be those with the largest ground capacities and the best overall performers.) Among referees, there are only two significant coefficients: Probert (who refereed only two matches – a 1–1 draw between Bolton and Charlton and a 0–3 win for Portsmouth at Sheffield United) and Atkinson both officiated over matches with significantly smaller mean home advantages than did Webb.

The second regression in Table IV replicates the first, except that it excludes the four referees who officiated at only a few matches (11 between them). This has the same outcomes for the first three variables and no significant differences across the 16 referees; indeed, with the exception of Atkinson, none comes anywhere close to statistical significance at the 0.05 level or better.

The next regression excludes the eight teams that played all their 19 home matches before (virtual) capacity crowds. In this, there are only two significant relationships, by far the most important of which is the relative strength of the away team. The

Table III. Referees who officiated at matches in the English Premiership, season 2006–2007.

	N	A	HA
Atkinson	24	92.5	0.17
Bennett	28	92.6	0.36
Clattenburg	27	90.1	0.04
Dean	25	93.5	0.48
Dowd	17	90.4	0.71
Foy	22	90.2	0.64
Gallagher	14	87.4	0.50
Halsey	26	94.3	0.31
Marriner	6	87.3	0.67
Mason	12	95.0	0.75
Poll	32	94.1	0.63
Probert	2	87.3	-1.50
Rennie	16	92.1	0.25
Riley	22	93.5	0.59
Stroud	1	82.6	-2.00
Styles	30	94.1	0.33
Tanner	2	82.2	-1.00
Walton	16	92.8	0.63
Webb	28	90.1	0.96
Wiley	30	93.6	0.40

Note: N = number of matches officiated; A = mean attendance at those matches, as a percentage of the season's maximum attendance; HA = mean home advantage at those matches.

Table IV. Regression analyses of home advantage, including referee effects (standard errors and significance levels in parentheses).

Model	I	II	III	IV
Constant	1.972 (0.396: 0.004)	1.934 (0.393: 0.000)	2.332 (0.558: 0.000)	2.237 (0.558: 0.000)
Home goals	0.367 (0.186: 0.049)	0.381 (0.187: 0.043)	0.147 (0.264: 0.578)	0.136 (0.507: 0.613)
Away goals	-2.240 (0.182: 0.000)	-2.231 (0.184: 0.000)	-2.267 (0.226: 0.000)	-2.247 (0.229: 0.000)
Attendance (in 000s)	0.018 (0.006: 0.002)	0.017 (0.006: 0.005)	0.007 (0.009: 0.432)	0.008 (0.009: 0.400)
Referee (comparator: Webb)				
Atkinson	-0.776 (0.375: 0.039)	-0.723 (0.373: 0.053)	-0.376 (0.423: 0.375)	-0.304 (0.421: 0.471)
Bennett	-0.414 (0.362: 0.254)	-0.365 (0.360: 0.310)	-0.443 (0.452: 0.328)	-0.374 (0.451: 0.408)
Clattenburg	-0.489 (0.368: 0.185)	-0.446 (0.366: 0.224)	-0.777 (0.447: 0.084)	-0.703 (0.446: 0.116)
Dean	-0.454 (0.371: 0.222)	-0.400 (0.369: 0.278)	-0.365 (0.444: 0.412)	-0.293 (0.443: 0.508)
Dowd	0.181 (0.416: 0.665)	0.227 (0.415: 0.584)	0.307 (0.475: 0.518)	0.377 (0.474: 0.428)
Foy	-0.366 (0.385: 0.343)	-0.316 (0.383: 0.409)	-0.342 (0.444: 0.442)	-0.269 (0.443: 0.544)
Gallagher	-0.539 (0.445: 0.227)	-0.495 (0.444: 0.265)	0.142 (0.488: 0.771)	0.218 (0.488: 0.655)
Halsey	-0.600 (0.368: 0.104)	-0.548 (0.366: 0.135)	-0.197 (0.455: 0.666)	-0.131 (0.455: 0.774)
Marriner	-0.459 (0.571: 0.422)	- (0.571: 0.422)	-0.744 (0.649: 0.253)	- (0.649: 0.253)
Mason	0.212 (0.482: 0.661)	0.257 (0.480: 0.593)	0.050 (0.520: 0.923)	0.120 (0.520: 0.818)
Poll	-0.293 (0.350: 0.403)	-0.241 (0.348: 0.488)	-0.502 (0.463: 0.280)	-0.431 (0.462: 0.352)
Probert	-2.151 (0.988: 0.030)	- (0.988: 0.030)	-1.893 (0.955: 0.049)	- (0.955: 0.049)
Rennie	-0.229 (0.424: 0.590)	-0.180 (0.422: 0.669)	-0.256 (0.502: 0.610)	-0.190 (0.502: 0.706)
Riley	-0.270 (0.385: 0.483)	-0.216 (0.383: 0.573)	-0.176 (0.488: 0.719)	-0.107 (0.488: 0.826)
Stroud	-2.504 (1.377: 0.070)	- (1.377: 0.070)	-2.145 (1.326: 0.107)	- (1.326: 0.107)
Styles	-0.287 (0.355: 0.419)	-0.237 (0.353: 0.503)	0.057 (0.418: 0.892)	0.125 (0.417: 0.764)
Tanner	-1.224 (0.991: 0.218)	- (0.991: 0.218)	-0.081 (1.324: 0.951)	- (1.324: 0.951)
Walton	-0.413 (0.424: 0.331)	-0.365 (0.422: 0.388)	-0.180 (0.504: 0.721)	-0.105 (0.503: 0.834)
Wiley	-0.412 (0.355: 0.247)	-0.359 (0.353: 0.309)	0.190 (0.453: 0.675)	0.257 (0.452: 0.571)
R^2	0.330	0.321	0.305	0.209
N	380	369	228	219

Key to models: I = all matches; II = all matches excluding those officiated at by four referees; III = all matches excluding teams that played to capacity crowds at all games; IV = all matches excluding teams that played to capacity crowds at all games and those officiated at by four referees.

Note: Coefficients significantly different from zero at the 0.05 level or better are shown in **bold**.

other shows again that Probert officiated at matches where on average the away team performed significantly better than was the case at Webb's 28 matches. Finally, the fourth regression in Table IV again excludes the four referees who handled only a few matches each. In this, the only significant relationship is with the away team's mean performance

and there are no significant differences among the 16 referees.

Conclusions

There is a considerable literature on the extent of home advantage in a range of spectator sports. Boyko

and colleagues' (2007) analysis of results in the English soccer Premiership extends it by suggesting that part of the variation in this home advantage can be accounted for by both variations in attendance at matches and some referees apparently being more favourably inclined towards the home team than others.

This replication of their analysis using data for the latest English Premiership season (2006–2007) does not sustain their findings. It is argued that because a substantial proportion of the teams play before (near-) capacity crowds at all of their home games, biases may be introduced into the data set by including those games. By focusing only on those matches played at grounds where there was considerable variation in attendances over the season, the analyses reported here have shown that changes in the attendance had no significant relationship to the outcome of those matches. It has also been shown that – excluding four who officiated over only a few, probably unrepresentative, matches – there were no significant variations among referees in the home team's advantage. Indeed, the only significant variable linked to differences in home advantage at

those 219 matches was the away team's strength, as indicated by the mean number of goals they scored in other away games during the season.

One replication does not necessarily entirely nullify a study's findings, but that reported here does suggest caution with regard to Boyko and colleagues' finding regarding "referee bias". There is no evidence that it was present in a structured way during the 2006–2007 English Premiership season.

References

- Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences*, 25, 1185–1194.
- Carron, A. V., Loughhead, T. M., & Bray, S. R. (2005). The home advantage in sport competitions: Courneya and Carron's conceptual framework a decade later. *Journal of Sports Sciences*, 23, 395–407.
- Clarke, S. R., & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician*, 44, 509–521.
- Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876–2003). *Journal of Sports Sciences*, 23, 337–350.