The Decline in Home Advantage in the English Premier league in 2015/16: Is it Random Fluctuation or A Systematic Change?

Dr Alun Owen
Institute of Science and the Environment, University of Worcester

The presence of a home advantage in football is a phenomenon that has been well-researched and documented in a number of published papers, such as for example, Boyko et. al. (2007). The research literature suggests several possible causes of a home advantage effect. These include the impact of playing away from home on player performance, a pre-disposition for referees to favour the home team, the pressure of the larger home crowd on the referee and the distance/time spent travelling or staying away from home when playing away. Most statistical models that appear in the published research literature therefore include parameters that account for and attempt to quantify home advantage, see for example Boshnakov et. al. (2015).

However, at the time of writing (7th March 2016) the English Premier League this season (2015/16) has so far seen the lowest proportion of home wins since the inception of the Premier League in 1992. Therefore, in this article we explore whether this apparent change is likely to reflect a true fall in home advantage or is it perhaps more likely a facet of random variation that we experience in football results from time to time (as in many walks of life).

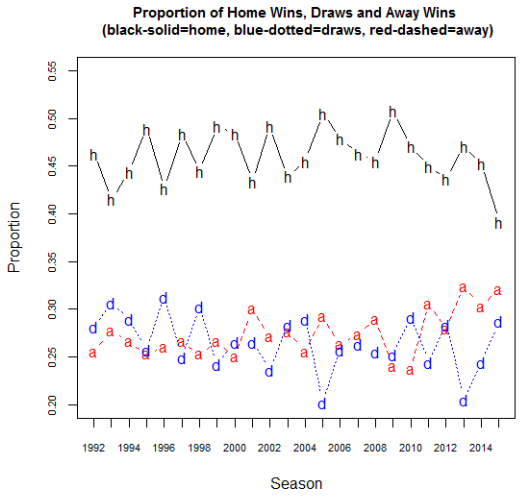
1. Data.

Data from www.football-data.co.uk provides match results and other data from the 1993/1994 season in the English Premier League, up to and including the results so far in the current season (2015/16) where so far 288 games out of a scheduled 380 have been played.

2. Exploratory Data Analysis.

Like all good analyses we start by exploring the data. Figure 1 shows the proportion of home wins, draws and away wins each season, since 1992/93 (note that 1992 refers to the 1992/93 season etc.).

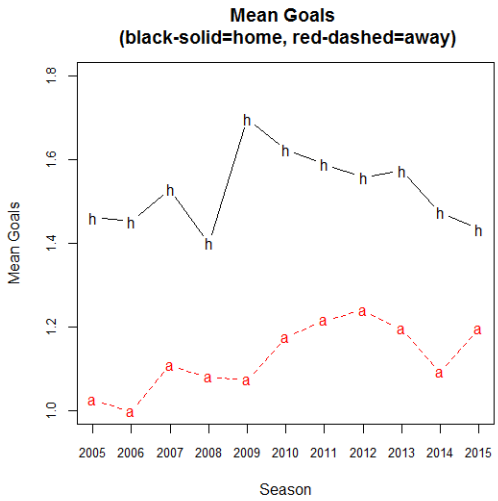
Figure 1



The results for the current season are based on data up to and including 7th March 2016. This illustrates that, if anything, since 1992/93 there was a slight increase in the proportion of home wins up until around 2005, but since then the size of this home advantage has been in reasonable decline over the last 10 seasons. The most obvious exception to this is an increase in the proportion of home wins in the 2009/10 season, which actually saw the highest proportion of home wins ever in a season over the period we have examined at around 51%. However, the current season so far has seen a marked drop back below the levels first seen in the early 1990's.

Looking deeper we examine the goals scored by the home and away teams in Figure 2, which concentrates on the period since 2005/06 in order to examine any decline in home advantage in more detail. Apart from the sudden increase in goals scored by the home team in the 2009/10 season, the changes in home advantage seem to be mostly explained by a steady increase in the mean number of goals scored by the away teams.

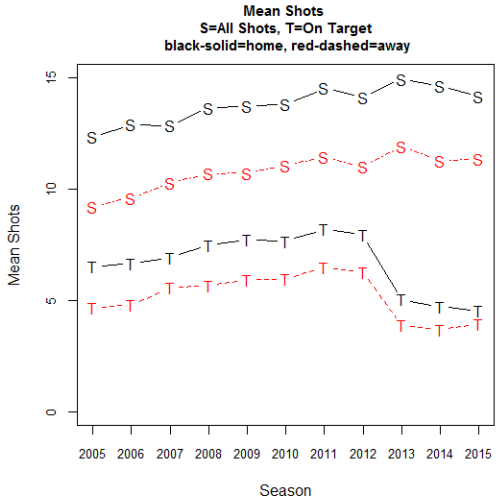
Figure 2



So it seems the decline in home advantage is perhaps more about the away teams scoring more goals. But are away teams attacking more or are they creating similar numbers of scoring opportunities in the past but creating better quality ones? To explore this, in Figure 3 we examine the mean shots per match for the home and away teams and also the mean shots on target.

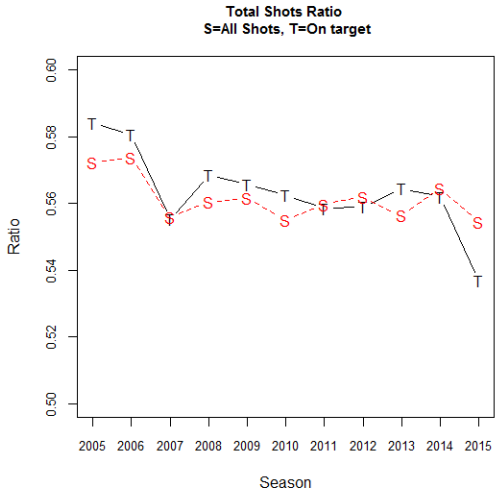
Figure 3 suggests that since 2005/06 there has been a continual increase in the total number of shots for both the home and away teams. Although we cannot verify whether this is as a result of changes to what is considered a "shot" in the data collected, this would seem unlikely to manifest itself with the pattern of a continual increasing trend. Therefore the evidence would indeed seem to point to an increase in shots for both home and away teams. Figure 3 also suggests an initial increase in shots on target by both the home and away teams since 2005/06, but interestingly both exhibit sudden decreases in 2013/14 onwards. This sudden fall does seem more likely to be due to changes in the manner in which shots on target are recorded, but it could still reflect, at least to some degree, real decreases in shots on target. This evidence therefore might indicate that teams are taking more shots at goal which are of a lower quality and less likely to be on target.

Figure 3



The key highlight however in Figure 3, is how the mean shots and shots on target for the away teams have increased over the last few seasons to the point where they are much closer to the numbers achieved by the home team, and in the current season they are almost the same. This last point is illustrated more clearly in Figure 4, which shows the mean total shots ratio (TSR) for the home teams. The TSR for the home teams is calculated as $\frac{home\ shots}{(home\ shots + away\ shots)}$, such that lower values for TSR indicate lower levels of home advantage. The TSR based on both all shots and also shots on target are shown in Figure 4, which suggests a general decline in home advantage over time, but a much greater decline in the current season.

Figure 4



So, there does seem to be a much greater decline in home advantage this season, whether this is framed in terms of results, goals or shots. Based on all these metrics, the current season appears to reflect a much lower degree of home advantage. However, an important consideration for the current season is that we do not have a balanced data set in that not all teams have played each other twice. Therefore we need to account for the abilities of the

teams that have played more at home and those that have played more away and also the abilities of the teams they have played. In the next section we account for this by fitting a simple Poisson goals model, equivalent to that described by Maher (1982) and described by many other authors since then including Dixon and Coles (1997) and Owen (2011). The simple Poisson model was also considered by Boshnakov et. al. (2015) who proposed a more complex model based in a Weibull count distribution. For simplicity, here we use the simple Poisson model, which does not include any of the adjustments or additional complexities added by those authors. One reason for this is (as we will show in a subsequent paper) that the adjustments proposed by Dixon and Coles no longer apply in the same way that existed at the time their paper was published almost 20 years ago now. However, it is possible that the analyses which follow here could be improved using more sophisticated models.

3. Model Fitting,

The model used is a Poisson goals model, similar to that originally presented in Maher (1982). The number of goals scored by team i playing at home and team j away are denoted by $X_{i,j}$ and $Y_{i,j}$ respectively, and are modelled as independent Poisson variables as follows:

$$X_{i,j} \sim Po(\mu_{i,j}),\ Y_{i,j} \sim Po(\lambda_{i,j}),$$

$$\mu_{i,j} = \alpha_i \times \beta_j \times \gamma, \quad \lambda_{i,j} = \alpha_j \times \beta_i. \tag{1}$$

The α_i and β_i measure the attack and defence abilities respectively, of team i, and are the same irrespective of whether a team is playing at home or away, whilst the multiplying parameter γ in (1) reflects the home advantage. This model was fitted to each season separately, using only the data for that season in order to obtain separate estimates of the home effect for each season. This was fitted using goals scored as the response (dependent) variable, and repeated using all shots as the response and also shots on target as the response. In this way we obtain three different potential measures of home advantage based on these three metrics. The resulting estimates of the home effect parameter γ obtained for each metric in each season are plotted in Figure 5.

Figure 5

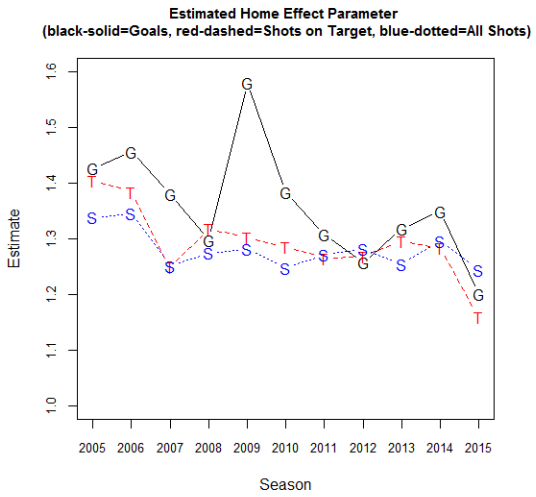


Figure 5 illustrates the sudden decline in home advantage in the current season reported earlier, and so even after accounting for the abilities of the teams that have played more at home and those that have played more away, this decline is still evident.

The question that remains however, is whether this decline reflects a true fall in home advantage or is it simply a facet of random variation? One way to explore this is to consider whether there have been any similar periods (windows) of 288 games in any of the previous seasons the English Premier League which have exhibited a similar change in home advantage in the shorter term. The same simple Poisson model, described in (1), was fitted with goals as the response, using a moving window of 288 games from the start of the 2005/06 season to the end of the 2014/15 season. Each season is again considered separately, so that we do not fit the models across seasons to avoid estimation problems when only a small number of match results data are available for the promoted teams. Therefore, in each season we fit the model to a moving window such that each window includes games i to $288+i-1$ ($i=1, \ldots,93$) from that season.

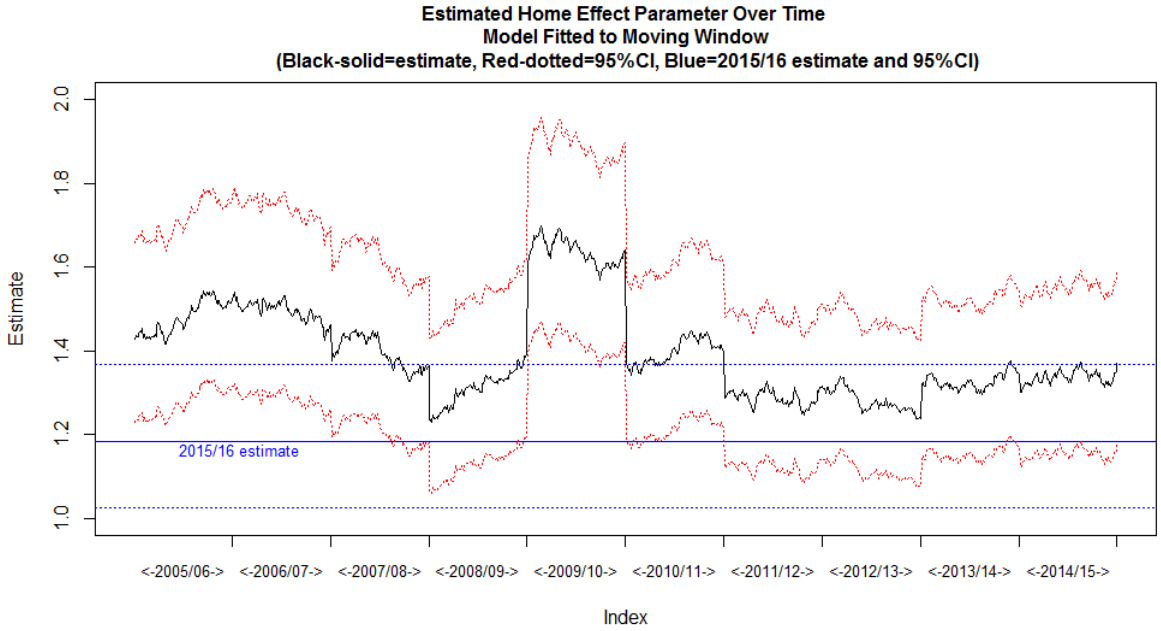
Figure 6 shows the resulting estimates of the home effect parameter (γ) over time (black-solid line). For example, the data for 2005/06 provides estimates of the home effect parameter from March 2006 through to the completion of the season in May 2006 (recall these estimates are based just on the match results from that season only using a moving window of 288 games over that period). This indicates that the home effect has generally declined steadily since 2005/06 from a figure above 1.4 to a level below that. The sharp rise in the 2009/10 season is indicative of the large proportion of home wins experienced in that season, discussed earlier in relation to Figure 1.

Figure 6 also includes a reference line (blue solid line) which indicates the estimated home effect parameter based on the data from the current 2015/16 season. This suggests that since 2005/06, the estimated home effect (black-solid line) has never been as low in any window of 288 games within any one season, as that which has so far been observed in the current season (blue-sold line).

However, the plot also includes 95% confidence intervals for the estimates (derived from the Wald standard errors when fitting the model) for each past season (red-dotted lines). This interval illustrates the uncertainty in the estimated size of the home effect at any point in time due to the element of randomness that is present in football match results data (surprising results do occur regularly in football). Since the 2011/12 season onwards, the current season's estimated home effect falls almost always within this interval, which provides evidence that the true underlying level of home advantage in the current season may well be no different to what it has been since 2011/12. Furthermore, we also need to consider the fact that the current season's estimated home effect is also subject to similar uncertainty. Therefore Figure 6 also shows the 95% confidence interval for the current season (blue-dotted lines). Taking the uncertainty in the estimated home effect for both past seasons and also the current season, we see that these intervals overlap almost continuously, apart from in 2009/10. This would seem to provide very strong evidence

indeed to suggest that any observed decline in home advantage in the English Premier League in 2015/16 is most likely due to random fluctuation rather than a systematic change.

Figure 6



References

Boshnakov, G., Kharrat, T., and McHale, I. G. (2015) Using the Weibull count distribution for predicting the results of football matches, http://www.sportstradingnetwork.com/article/journal/using-the-weibull-count-distribution-for-predicting-the-results-of-football-matches/

Boyko, R. H., Boyko, A. R., and Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. Journal of Sports Sciences, 25:11, 1185-1194, DOI: 10.1080/02640410601038576, http://dx.doi.org/10.1080/02640410601038576

Dixon, M.J. and Coles, S.G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46, 265–280.

Maher, M.J. (1982). Modelling association football scores. *Statistica Neerlandica*,36, 109–118.

Owen, A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, 22(2), 99-113. DOI:*10.1093/imaman/dpq018.*