The Home-Court Advantage in Basketball

Dale Zimmerman
University of Iowa

February 3, 2015
Home advantage in sports

- The so-called home advantage has been documented in every major team sport and many individual-competitor sports.

- A large volume of literature exists on the topic, in diverse journals. Some examples:


The home advantage

• The conclusion of all this literature is that the home advantage is largest in basketball and hockey, not as large in American football or soccer, and smallest in baseball and individual sports (though Dr. Z doesn’t find the evidence for this compelling).

• The literature hypothesizes that home advantage may be the net effect of several factors, including:

  1. crowd support (direct effect on team and indirect effect via referees)
  2. distinctive features of the playing facilities
  3. travel
The home court advantage (HCA) in basketball

- In basketball, the game venue is enclosed, with fans very close to the players and referees; travel can be wearisome.

- In any given season, 65–70% of NCAA Division I men’s basketball games, and roughly 60% of NBA games, are won by the home team.

- In the NCAA, teams from the major (stronger) conferences play more games at home than away; in the NBA, all teams play the same number of home and away games, but the schedule is still slightly “unbalanced.”

- So team strength must be accounted for in a proper analysis of HCA (especially for NCAA games).
Harville and Smith’s paper

Via a relatively standard regression analysis of a slightly unusual linear model for basketball game score differences, Harville and Smith address the following questions:

- Does a HCA exist in basketball? If so, how large is it?
- Are there differences in the HCA from team to team? If so, do the better teams tend to have the greater home-court advantages?
Other questions of interest

Many other questions could be of interest, e.g.

- Are there any notable differences between the HCA in the NCAA and the HCA in the NBA?

- Can we determine how much of the HCA is attributable to some of the three factors noted previously (crowd support, venue familiarity, travel)?

These questions might also be addressed with some variations on Harville and Smith’s methodology.
Data

I will apply Harville and Smith’s methodology to two sets of data:

1. All NCAA Division I men’s basketball games from the 2010-11, 2011-12, 2012-13, and 2013-14 seasons played by the 343 teams that were NCAA members for all 4 seasons (each team plays roughly 30-35 games each season). Data source:

   http://www.sports-reference.com/cbb/play-index/tgl_finder.cgi

2. All NBA games from the same 4 seasons (30 teams each season, each team plays 82 games). Data source:

   http://www.basketball-reference.com/play-index/tgl_finder.cgi
Some notation

- $r_{ij} =$ the number of games in which the $i$th team is the home team and the $j$th team is its opponent (if a game is played on a neutral court we arbitrarily label one of the teams the home team)

- $y_{ijk} =$ the difference in score between teams $i$ and $j$ in the $k$th of these $r_{ij}$ games

- $x_{ijk} = \begin{cases} 0 & \text{if } ijk\text{th game is played on a neutral court} \\ 1 & \text{otherwise.} \end{cases}$
Three models

- Model 1

\[ y_{ijk} = \beta_i - \beta_j + e_{ijk} \]

- Model 2

\[ y_{ijk} = \begin{cases} 
\lambda + \beta_i - \beta_j + e_{ijk} & \text{if } x_{ijk} = 1 \\
\beta_i - \beta_j + e_{ijk} & \text{if } x_{ijk} = 0 
\end{cases} \]

- Model 3

\[ y_{ijk} = \begin{cases} 
\alpha_i - \beta_j + e_{ijk} & \text{if } x_{ijk} = 1 \\
\beta_i - \beta_j + e_{ijk} & \text{if } x_{ijk} = 0 
\end{cases} \]

The \( \beta_i \)'s, \( \lambda \), and the \( \alpha_i \)'s are regarded as unknown parameters; and the \( e_{ijk} \)'s are taken to be iid \( N(0, \sigma^2) \) random variables.
A toy example for Model 2

Suppose 3 teams played 6 games amongst each other, in which:

- Team 1 plays twice at home, losing to Team 2 by 1 point and defeating Team 3 by 9 points
- Team 2 plays twice at home, defeating Team 1 by 15 points and Team 3 by 20 points; and once on a neutral court, defeating Team 3 by 6 points
- Team 3 plays once at home, losing to Team 2 by 7 points

\[
y = \begin{pmatrix} -1 \\ 9 \\ 15 \\ 20 \\ 6 \\ -7 \end{pmatrix} = \mathbf{x}\beta = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \lambda \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \epsilon\]
Explanation of models

- Model 1 makes no allowance for a HCA.

- Model 2 allows for the possibility of a HCA but takes it to be identical for every team; it also implicitly assumes that the expected difference in score between any two teams in a game played on a neutral court is halfway between the expected differences in scores of games played by the same two teams on each other’s home courts.

- Model 3 allows for the possibility of a HCA and the possibility that it may vary from team to team; it also implicitly assumes that the expected difference in score between any two teams in a game played on a neutral court equals the difference in expected differences in scores of games played by them against a common opponent on the opponent’s home court.
Definition of HCA in terms of model parameters

Define $HCA_i =$ expected difference in score in a game played by team $i$ on its home court minus the expected difference in score in a game played by team $i$ on a neutral court against the same opponent.

- Under Model 1, $HCA_i = 0$
- Under Model 2, $HCA_i = \lambda$
- Under Model 3, $HCA_i = \alpha_i - \beta_i$

Each of the latter two $HCA_i$’s is estimable under the model in which it is defined, provided that the data are “connected” (which our data are).
Model fitting and parameter estimation

We may fit the model, and estimate regression coefficients, by the method of least squares.

- Obtain a solution, $\hat{\beta}$, to the “normal equations” $X^T X \beta = X^T y$.
- $c^T \hat{\beta}$ is the least squares estimate of $c^T \beta$ for any $c$ whose elements sum to 0.
- We may estimate the standard error of $c^T \hat{\beta}$ by

$$
\hat{se}(c^T \hat{\beta}) = c^T (X^T X)^{-1} cs^2,
$$

where $s^2$ is the residual MSE from the ANOVA of the fitted model.
Model comparisons

- Note that Model 1 $\subset$ Model 2 $\subset$ Model 3.
- Thus, we can answer questions about HCA by comparing the models using a classical full-vs.-reduced model F-testing approach:

$$F = \frac{RSS_{red} - RSS_{full}}{(df_{red} - df_{full})} \times \frac{RSS_{full} / df_{full}}{1}$$
Results

Test for existence of a HCA (Model 2 vs. Model 1), i.e., test $H_0 : \lambda = 0$ vs. $H_A : \lambda \neq 0$.

E.g., for NCAA in 2010-2011:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>343</td>
<td>627412</td>
<td>1829</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>1</td>
<td>1</td>
<td>51392</td>
<td>51392</td>
</tr>
<tr>
<td>Residual</td>
<td>5012</td>
<td>517547</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5356</td>
<td>1196351</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P$-value: $Pr(F_{1,5012} > 498) < 1.0 \times 10^{-8}$
Existence and magnitude of HCA

$P$-values:

<table>
<thead>
<tr>
<th></th>
<th>2010-11</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAA</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
</tr>
<tr>
<td>NBA</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
<td>$&lt; 1.0 \times 10^{-8}$</td>
</tr>
</tbody>
</table>

Estimates of $\lambda$ (estimated standard errors in parentheses):

<table>
<thead>
<tr>
<th></th>
<th>2010-11</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAA</td>
<td>3.50 (.16)</td>
<td>3.66 (.15)</td>
<td>3.53 (.16)</td>
<td>3.11 (.16)</td>
</tr>
<tr>
<td>NBA</td>
<td>3.18 (.30)</td>
<td>2.95 (.36)</td>
<td>3.26 (.32)</td>
<td>2.59 (.32)</td>
</tr>
</tbody>
</table>
## More results

Test for equality of HCA across teams (Model 3 vs. Model 2), i.e., test

$$H_0 : \alpha_1 - \beta_1 = \alpha_2 - \beta_2 = \cdots = \alpha_t - \beta_t \text{ vs. } H_A : \text{not } H_0$$

E.g., for NCAA in 2010-2011:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>343</td>
<td>627412</td>
<td>1829</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>1</td>
<td>1</td>
<td>51392</td>
<td>51392</td>
</tr>
<tr>
<td>Model 3</td>
<td>2</td>
<td>343</td>
<td>40479</td>
<td>118</td>
</tr>
<tr>
<td>Residual</td>
<td>4669</td>
<td>477068</td>
<td>102</td>
<td></td>
</tr>
</tbody>
</table>

|                 | 5356| 1196351|

\[ P\text{-value: } Pr(F_{343,4669} > 1.15) = 0.030 \]
Does the HCA vary from team to team?

$P$-values:

<table>
<thead>
<tr>
<th></th>
<th>2010-11</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAA</td>
<td>.030</td>
<td>.060</td>
<td>.003</td>
<td>.001</td>
</tr>
<tr>
<td>NBA</td>
<td>.217</td>
<td>.015</td>
<td>.144</td>
<td>.354</td>
</tr>
</tbody>
</table>

- Thus, HCA varies significantly from team to team in NCAA basketball, but possibly not in the NBA.
- The root mean squared deviation of the NCAA’s HCA’s is at most only about 25% of the overall HCA, so it’s not of great practical significance.
- Scatterplots of team-specific HCA versus team performance level indicate that there is no relationship between them.
Conclusions

1. The HCA undeniably exists in basketball and is of great practical importance. It is robust over years, but it appears that it is slightly (10–15%) larger in the NCAA than in the NBA.

2. The HCA varies from team to team in the NCAA, but it is not consistently higher for some teams than others over time, nor does it appear to have any relationship with a team’s overall performance level; moreover, it is not of great practical importance.

3. The HCA does not consistently vary from team to team in the NBA.
Additional questions

1. How much, if any, of the HCA is attributable to free throws? And how much of this can be attributed to
   
   (a) A difference in the number of free throws taken ($\approx$ # of fouls called against the other team)?
   
   (b) A difference in the free throw shooting percentage?

   Preliminary results suggest that as much as 50% of the HCA may be explained by free throws.

2. Is there a relationship between HCA and distance traveled by the visiting team?

Answers to these (and possibly other) questions will be reported later in the semester.