The Home-Court Advantage in Basketball

Dale Zimmerman
University of Iowa

January 27, 2017
Home advantage in sports

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

- The magnitude of the home advantage varies among sports:

<table>
<thead>
<tr>
<th>Sport</th>
<th># of Studies</th>
<th># of Games</th>
<th>Home winning %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseball</td>
<td>7</td>
<td>133,560</td>
<td>54.3</td>
</tr>
<tr>
<td>Football</td>
<td>5</td>
<td>2,592</td>
<td>57.3</td>
</tr>
<tr>
<td>Hockey</td>
<td>5</td>
<td>5,312</td>
<td>61.2</td>
</tr>
<tr>
<td>Basketball</td>
<td>9</td>
<td>13,686</td>
<td>64.4</td>
</tr>
<tr>
<td>Soccer</td>
<td>3</td>
<td>40,493</td>
<td>68.3</td>
</tr>
</tbody>
</table>
Home advantage in sports, continued

- The degree of home advantage within each sport has been relatively consistent over time until recently (will say more about this later)

- The degree of home advantage between the college and professional levels for a given sport is relatively similar.
Why is there a home advantage?

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

  1. crowd factors (direct effect on team and indirect effect via referees)
  2. learning factors (familiarity with distinctive features of the playing facilities)
  3. travel factors (fatigue, disruption of routine)
  4. rule factors (rule differences based on game location)
Evidence for/against crowd factors

- Baseball teams playing in domed stadiums won 10.5% more games at home than on the road ($N > 35,000$ games); the comparable value for teams playing in open-air stadiums was 7.2% (Zeller and Jurkovac, 1988).

- Home advantage did not vary across the 4 divisions of the English Football League, even though average crowd size varied from 1,500 in Division 4 to 25,000 in Division 1 (Dowie, 1982).

- Home advantage in baseball increased from 48% when attendance was 20% of capacity to 55% and 57% when attendance was 20-40% and >40% of capacity (Schwartz and Barsky, 1977).
Evidence for/against crowd factors, continued

- In basketball, the home team’s superiority in points scored, turnovers, and fouls increased significantly during periods of booing greater than 15 seconds (Greer, 1983).

- Significantly fewer fouls were called on star players for the LA Lakers in the 1984-85 season (Lehman and Reifman, 1987).

- All-Star pitchers have a significantly larger called strike zone than other pitchers (Kim and King, 2014).
Evidence for/against crowd factors, continued

- In a controlled experiment, trained soccer observers viewed videotape of 52 tackles/challenges from a televised match, of which half were made by the home player (and half by the visitor). Half of the observers observed the video with sound, the other half without sound. The response was whether the player was carded. A significant sound*team location effect was found (Nevill et al., 1999).
Evidence for/against learning factors

- Home advantage for 37 baseball, basketball, and hockey teams that moved to new stadiums within the same metropolitan area from 1987 to 2001 was significantly lower in their first season in the new stadium than in their final season in the old stadium (Pollard, 2002).
Evidence for/against travel factors

- Several studies have investigated whether the home advantage increases as the season progresses (as effects of travel might begin to accumulate). None have found a significant time-of-season effect.

- Distance traveled has been found not to have a significant effect in soccer (Pollard, 1986), minor league baseball (Courneya and Carron, 1991), and hockey (Pace and Carron, 1993).
Evidence for/against rule factors

- In baseball, the home team gets to bat last. Does this accrue to the home field advantage? In a study of recreational slow-pitch softball games, each meeting between teams was a doubleheader with alternating home-visitor status; batting last did not provide an advantage (Courneya and Carron, 1990).
Other factors?

- Biological: A study of English soccer players showed that salivary testosterone levels were significantly higher before a home game than an away game (Neave and Wolfson, 2003).
References


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 factors noted previously (crowd support, referee bias, 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 through 2015-16 seasons played by the 344 teams that were NCAA members for all 6 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 6 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

\[
y = \begin{pmatrix} -1 \\ 9 \\ 15 \\ 20 \\ 6 \\ -7 \end{pmatrix} = x\beta = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \lambda \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \varepsilon
\]
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 $HCA_i = \text{expected difference in score in a game played by team } i \text{ on its home court minus the expected difference in score in a game played by team } i \text{ on a neutral court against the same opponent.}$

- 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

\[
\hat{se}(c^T \hat{\beta}) = \left[ c^T (X^T X) c - cs^2 \right]^{1/2},
\]

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:

$$F = \frac{(RSS_{red} - RSS_{full})}{RSS_{full}/df_{full}} \times \frac{df_{red} - df_{full}}{RSS_{red}/df_{red}}$$
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 are all less than $1.0 \times 10^{-8}$ for all years, for both NCAA and NBA

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

<table>
<thead>
<tr>
<th>Year</th>
<th>NCAA</th>
<th>NBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAA</td>
<td>3.50 (.16)</td>
<td>3.66 (.15)</td>
</tr>
<tr>
<td>NBA</td>
<td>3.18 (.30)</td>
<td>2.95 (.36)</td>
</tr>
</tbody>
</table>
More results

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

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

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>
<tr>
<td></td>
<td></td>
<td>5356</td>
<td>1196351</td>
<td></td>
</tr>
</tbody>
</table>

\[ P\text{-value: } Pr(F_{343, 4669} > 1.15) = 0.030 \]
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 evidence for it in the NBA is lacking.

• The root mean squared deviation of the NCAA’s HCA’s is at most only about 25% of the overall HCA, so the variation in HCA among teams is not of very great practical significance even in the NCAA.

• 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 large enough to be of great practical importance. It is robust over years (see the subsequent discussion on trends, however), 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, the variation is small relative to the mean.

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

1. How, if at all, has the HCA changed over time?

There’s significant evidence that it has been decreasing over recent years; see handouts from kenpom and espn.com

Theories as to why:

- Increase in proportion of total score consisting of made 3-pointers (smaller proportion consisting of made free throws), thus reducing referees’ effect on score
- Travel not as inconvenient (charter vs. commercial flights, improved video scouting technology)
- Home crowds don’t have as much effect
2. 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 (≈ # 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 in the NCAA may be explained by made free throws (see next page). But I don’t yet have an answer to parts (a) and (b) of the question.
HCA estimates by type of score

<table>
<thead>
<tr>
<th>Year</th>
<th>NCAA HCA</th>
<th></th>
<th></th>
<th>NBA HCA</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FT</td>
<td>2-pt</td>
<td>3-pt</td>
<td>Total</td>
<td>FT</td>
<td>2-pt</td>
<td>3-pt</td>
</tr>
<tr>
<td>2010-11</td>
<td>1.85</td>
<td>1.11</td>
<td>0.54</td>
<td>3.50</td>
<td>0.97</td>
<td>1.51</td>
<td>0.70</td>
</tr>
<tr>
<td>2011-12</td>
<td>1.87</td>
<td>0.94</td>
<td>0.85</td>
<td>3.66</td>
<td>1.00</td>
<td>1.70</td>
<td>0.25</td>
</tr>
<tr>
<td>2012-13</td>
<td>1.78</td>
<td>1.15</td>
<td>0.61</td>
<td>3.53</td>
<td>0.74</td>
<td>1.82</td>
<td>0.70</td>
</tr>
<tr>
<td>2013-14</td>
<td>1.74</td>
<td>0.92</td>
<td>0.45</td>
<td>3.11</td>
<td>0.97</td>
<td>1.80</td>
<td>-0.18</td>
</tr>
<tr>
<td>2014-15</td>
<td>1.77</td>
<td>1.00</td>
<td>0.41</td>
<td>3.17</td>
<td>0.70</td>
<td>1.06</td>
<td>0.63</td>
</tr>
<tr>
<td>2015-16</td>
<td>1.88</td>
<td>0.74</td>
<td>0.59</td>
<td>3.20</td>
<td>0.94</td>
<td>1.68</td>
<td>0.37</td>
</tr>
</tbody>
</table>
## Percentage of HCA by type of score

<table>
<thead>
<tr>
<th>Year</th>
<th>NCAA HCA</th>
<th></th>
<th></th>
<th>NBA HCA</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FT</td>
<td>2-pt</td>
<td>3-pt</td>
<td>Total</td>
<td>FT</td>
<td>2-pt</td>
<td>3-pt</td>
</tr>
<tr>
<td>2010-11</td>
<td>53</td>
<td>32</td>
<td>15</td>
<td>100</td>
<td>30</td>
<td>47</td>
<td>22</td>
</tr>
<tr>
<td>2011-12</td>
<td>51</td>
<td>26</td>
<td>23</td>
<td>100</td>
<td>34</td>
<td>58</td>
<td>9</td>
</tr>
<tr>
<td>2012-13</td>
<td>50</td>
<td>32</td>
<td>17</td>
<td>100</td>
<td>23</td>
<td>56</td>
<td>22</td>
</tr>
<tr>
<td>2013-14</td>
<td>56</td>
<td>29</td>
<td>15</td>
<td>100</td>
<td>37</td>
<td>70</td>
<td>-7</td>
</tr>
<tr>
<td>2014-15</td>
<td>56</td>
<td>31</td>
<td>13</td>
<td>100</td>
<td>29</td>
<td>45</td>
<td>26</td>
</tr>
<tr>
<td>2015-16</td>
<td>59</td>
<td>23</td>
<td>18</td>
<td>100</td>
<td>31</td>
<td>56</td>
<td>12</td>
</tr>
</tbody>
</table>