Jordan Keener

Estimating the True Value of Home-Court Advantage in NCAA Division I Men's Basketball

Introduction

The theory that a team playing at their "home" venue has an inherent advantage is widespread in elite competitive sports, with substantial evidence of such an effect across a wide range of sports. This theory is commonly referred to as the "home-field" or "home-court" advantage, which can be defined as the expected change in scoring margin when a team plays a game at their home venue rather than a neutral location, holding all other factors constant. This effect is easily determined in settings where game location is randomly assigned, which allows for the average scoring margin of teams playing at home to be interpreted as the average treatment effect of playing at home on scoring margin, relative to a neutral site. However, in many elite competitive sports environments, game location is highly non-random in assignment, making the determination of the true causal impact of playing at home more difficult. One such example is NCAA Division 1 Men's Basketball, where budgetary constraints of smaller universities and competitive incentives for elite teams lead to an endogeneity problem where game location is correlated with factors that influence game outcomes, the most problematic of which being team quality. In this paper, I use two panel data techniques to control for relative team and opponent quality in determining the effect of game location on scoring margin in NCAA Division I Men's Basketball from 2013 to 2016. I find home-court advantage to be worth between 3.02 and 3.23 points per game in terms of scoring margin over this time[1]. I also use these estimates to determine the effect of game location on a team's win probability across various levels of relative team and opponent quality, and find teams that would be approximately 1.5 point underdogs against a given opponent on a neutral court receive the largest increase in win probability from playing the same opponent at home.

Data

To estimate the effect of playing at home relative to a neutral court, I use game result data for each regular season game in the 2013-14, 2014-15, and 2015-16 seasons.[2] The level of observation in the full sample is team-game, meaning each game appears twice in the dataset, with each school appearing as the team of interest. For each observation, there are dummy variables indicating whether the team of interest played the game at home, on the road, or if the game was played at a neutral site, as well as a variable for the scoring margin (team score minus opponent score) of the game.

[1] All tables and figures reported in this paper can be replicated using the files provided at: https://github.com/jordkeen/HCProjectJK_Nov28Submission

[2] The data is made publicly available by the website spreadsheetsports.com, and can be downloaded via Dropbox using the URL https://www.dropbox.com/sh/pt08h91kszrdypj/n9QERQa20k

The only initial data restriction for the full sample is to limit the dataset to games between two Division I teams. This restriction is executed by dropping all observations where either the team or opponent appears less than 15 times in the dataset for the season in which the game was played. Since the original dataset contained all games involving a Division I team, this serves as an adequate proxy, as each Division I team plays at least 15 games in a season, and no non-Division 1 team plays 15 or more games against Division I opponents in a season.

For most of the analysis, I use one of the three reduced samples shown in Table 1. The reason for these reductions is to bring the data to the game level, as the team-game level data suffers from a serial dependence problem where each entry has a related entry with the same scoring margin but of the opposite sign. The most commonly used reduced sample is to only use team-game observations where the team was at home, which both reduces to the game level and eliminates neutral site games. The “home and home” sample reduction also restricts the sample to observations where the team was at home, but also requires that the team played the given opponent in a road game as well. This restriction method is explained in more detail later in the within-matchup estimator analysis section. Finally, all figures presented use one randomly selected team-game observation from each game in the full sample, unless otherwise noted.

Simple Home Court Advantage Model

A naïve regression of scoring margin on game location dummies would be one simple measure of the impact of game location on scoring margin, given by,

$$1)\quad Margin_{tg} = \beta_0 + \beta_1 Home_{tg} + \beta_2 Away_{tg} + \varepsilon_{tg}$$

where *Margin* is the scoring margin (team score minus opponent score) of team *t* in game *g*, *Home* is a dummy variable indicating whether team *t* played at home in game *g*, *Away* is a dummy variable indicating whether team *t* played on the road in game *g*, and ε is a mean-zero error term. Under this model, the reference state given by the intercept is a game played at a neutral site, so the home-court advantage effect would be given by $\beta_1 - \beta_0$. Assuming the expectation of margin at a neutral site is zero, this measure is identical to simply calculating the average scoring margin of teams playing at home. I choose to find the estimates using this method for the sake of simplicity and to eliminate the serial dependence problem arising from the use of the full dataset. Table 2 reports the mean scoring margin of home teams for each season, which range from 5.27 to 5.44. Across all three seasons, teams playing at home have an average scoring margin of 5.37 points.

However, these measures do not provide a reliable estimate of the true treatment effect of playing at home. Game location is extremely non-random in assignment, specifically for non-conference games where teams are free to schedule opponents at their own discretion. Due to budgetary constraints of athletic programs at smaller colleges, teams from those schools are more likely to seek road games that include cash payouts in their non-conference schedule, typically against superior opponents. Additionally, superior teams generally have greater scheduling power, and seek as many home games as possible in their non-conference schedule to minimize the chances of losses to inferior opponents, which would decrease a team’s probability of being selected for postseason tournaments. For example, the team at my college, Indiana University, finished in the top 15 of most major polls for the 2015-16 season, and played the likes of Eastern Illinois, Austin Peay, Alcorn State, Morehead State, Indiana-Purdue Fort Wayne, and McNeese State during its non-conference schedule that season, none of which are known as basketball powers. Predictably, IU won those games by an average of 35 points.

Because of these factors, simple measures likely provide an upper bound on the treatment effect of playing at home, as a team playing at home appears to be positively correlated with team quality relative to opponent, which would increase expected scoring margin. Comparing the average scoring margin of home teams before January 1st in each season, which coincides with the non-conference season where game location is highly non-random and relative team quality is more variable, and after January 1st, where these problems are reduced by in-conference scheduling, supports this claim. Since it is unlikely that home teams receive a significantly different boost in scoring margin form playing at home before and after New Year's Day, it is obvious there are confounding factors in at least one of the time periods biasing the estimate. Assuming these factors are not present at all after January 1st, the regressions in Panel B of Table 2 shows that the value of home-court advantage is closer to 3.2 points per game, although we can still not be sure that game location is completely random in assignment during this time period. I propose two alternative models that directly control for relative team and opponent quality to determine the true average treatment effect of playing at home relative to a neutral site.

Alternative Models

<u>Within Matchup Estimator</u>

The first method estimates the effect of changing game location within a matchup, which is herein defined as a team-opponent combination.[3] To estimate this, I restrict the data to observations where the team and opponent play each other in multiple non-neutral site games where each team plays at least one game at home. The data is also restricted to the first home game for each team within a matchup, so that each matchup has two games where each team is the home team once. Additionally, only games where a team was at home are used to reduce from the team-game level to the game level.

By restricting to within matchup variation and looking only at home teams, the effect of playing at home relative to a neutral site can be given by the within-matchup scoring margin difference, which is defined as,

$$2)\ Within\ Matchup\ \ Margin\ Difference = \ (Margin_{tg} - \overline{Margin_{to}})$$

where *Margin* is the scoring margin of team *t* in game *g*, $\overline{Margin}$ is the average margin between team *t* and opponent *o* over the two games included in the sample. Panel A of Table 3 shows the mean within-matchup margin differences across all three seasons, with the average across all years being 3.15 points per game. This result is consistent with the hypothesis that the naïve

[3] Teams and opponents are defined as a combination of school-year, meaning for analysis covering multiple years, teams from the same school are treated as independent entities from year-to-year.

regression of scoring margin on just game location dummy variables overestimates the value of home-court advantage by not controlling for team and opponent quality. Panel B shows the results of a regression of within-matchup margin difference on a dummy variable indicating if the team played the first non-neutral site game of the matchup at home, to determine if there was a differential effect of home court advantage based on game location sequence. However, there is no evidence of such a differential effect, as none of the coefficients on playing at home first are statistically significant at a 90% level of confidence.

Because this model only evaluates the impact of game location on within matchup variation in scoring margin, this model effectively controls for relative team and opponent quality, under the key assumption that variation in team-to-opponent relative quality, holding fixed game location, within a matchup is uncorrelated with whether the team is playing at home. While a team's quality can certainly vary over the course of a season due to a variety of factors, there is no reason to believe that relative team to opponent quality would be correlated with the time of season when a team plays the given opponent at home. Since the model only looks within matchups, time trends, such as the fact that teams would be theorized to improve as the season progresses, would not bias the estimate. Even differential time trends among teams with respect to team quality would not be an issue, assuming this difference is not correlated with the sequence in which teams play their home game within the matchup. Since the sequence of game location within these home-and-home matchups is essentially random, as most are in-conference matchups where schools have little to no scheduling power, the assumption that game location does not systematically vary with relative team to opponent quality appears to hold. The lack of evidence for a differential effect of playing at home first on within-matchup margin difference provides empirical support for this assumption.

Team and Opponent Fixed Effects

A related approach is to use team and opponent fixed effects to directly control for team and opponent quality. This framework is, in many ways, superior to the within estimator, as it 1) allows for the full sample of games to be used, opposed to the home-and-home requirement for the within estimator, 2) still controls for relative team and opponent quality when estimating the effect of game location on scoring margin, and 3) the team fixed effects can be interpreted as team quality rankings denominated in relative scoring margin to a reference team.

The basic framework for this model is given by,

$$3)\ Margin_{tg} = \beta_0 + \beta_1 Home_{tg} + \beta_2 Away_{tg} + \gamma_t team_t + \alpha_o opponent_{o_{tg}} + \varepsilon_{tg}$$

where *Home* is a dummy variable indicating if team t played at home in game g, *Away* is a dummy variable indicating if team t played on the road in game g [4], *team* represents dummy variables for each team t indicating if team[5] t played in the game, and *opponent* represents

[4] The *Home* and *Away* variables leave neutral site as the reference state, given by the intercept.

[5] One team is omitted from the team and opponent fixed effects dummy variables to be used as a reference team. The same team is omitted from each, leading to an expectation of zero for the intercept.

dummy variables for each opponent indicating if opponent *o* played in the game *g* against team *t*, and ε_{tg} is a mean-zero error term.

One nice feature of this model specification is that it eliminates all team-opponent level factors that were the basis of the within estimator model, eliminating the need to restrict the sample to matchups with multiple games. Another added benefit is the coefficients on the team fixed effects dummy variables indicate the effect of a team playing on scoring margin relative to a reference opponent, holding fixed game location. More formally, the team fixed effects coefficients for each team *t* represent,

$$E\left(Margin_{Team\ t,\ \ reference\ opponent} \mid neutral\right)$$

and the opponent fixed effects coefficients for each opponent *o* represent,

$$E\left(Margin_{reference\ team,\ \ Opponent\ o} \mid neutral\right)$$

which can be combined into the full model given below:

$$4)\quad E(Margin_{Team\ t,\ \ Opponent\ o,\ \ g}) = \beta_0 + \beta_1 teamHome_{tg} + \beta_2 teamAway_{tg} + E\left(Margin_{Team\ t,\ \ reference\ opponent} \mid neutral\right) + E\left(Margin_{reference\ team,\ \ Opponent\ o} \mid neutral\right)$$

This equation is particularly useful when it is necessary to make predictions of margin between a given team and opponent, which will be revisited in the next section evaluating the impact of game location on win probabilities. For this analysis, I choose to use the full sample of data at the team-game level. This is so each game that a team plays appears in the data with them as both the team of interest and as the opponent, providing richer team and opponent fixed effects. Since the reference for each fixed effect is the same team, each team's team fixed effect and opponent fixed effect are identical, but of the opposite sign.

Unsurprisingly, this model yields very similar results to the within matchup estimates. The coefficient on the home dummy variable is 3.02, 3.20, and 3.21 for the 2013-14, 2014-15, and 2015-16 seasons respectively. Table 4 shows full regression results using this model. Because the full sample of data at the team-game level is used to provide the best estimates for the team and opponent fixed effects, standard errors are clustered on game.

As the team and opponents fixed effects model is essentially the same as the within estimator, just with a slightly altered specification to handle more data and provide additional meaningful coefficients, it also requires the same key assumption to interpret the estimate for β_1 as the true average treatment effect of playing at home relative to a neutral court, holding team and opponent quality fixed. That assumption is still that game location is uncorrelated with game-to-game variation in team and opponent quality around its average, controlling for the average effect of playing at home across all games in the sample. This is still in the error term because the team and opponent fixed effects represent just average team and opponent quality. Violating this assumption is slightly more plausible under this model because it involves non-conference games scheduled strategically, eliminating the randomness of game location assignment that was present in the restricted home-and-home sample. However, by definition, each game is played at the same point in time in the season by each team in the game, so time trends relating to injuries or team development that are not team-specific would not bias the estimate. Any other potential endogeneity that would bias the estimate would arise from team-specific trends in relative quality to opponent around the full season average (team/opponent fixed effect control) that are correlated with game location, which would be differential treatment effects of game location by team. While it is certainly possible, and probably very likely, there are differential home court advantages by school, this does not prevent the estimates for β_1 from being interpreted as the unbiased average treatment effect of playing at home.

Effect of Game Location on Win Probability

Each model introduced so far has focused on estimating the point value of home-court advantage by looking at its effect on scoring margin. However, under the assumption that teams are win-maximizing rather than scoring margin-maximizing, the most relevant effect of home-court advantage would be the change in win probability when game location is changed from a neutral site to a home game, holding all else constant.

The team and opponent fixed effects model proves useful here, as the fixed effect coefficients for teams and opponents can be used to make a prediction for the expected scoring margin between any two teams from the same season if the game were played a neutral site, equal to the sum of the team fixed effect of team *t* and the opponent fixed effect of opponent *o* for team *t* in game g.

$$5)\quad (\widehat{Margin}_{tg} \mid neutral) = \gamma_t Team_t + \alpha_o Opponent_{o_{tg}}$$

Assuming homoscedasticity and normally distributed errors, the standard error of the predicted values for margin can be used to give probabilistic estimates for the scoring margin between two teams on a neutral site. Specifically, margin is normally distributed with a mean of predicted margin and standard deviation of the standard error of the forecast,

$$6)\quad (Margin_{tg} \mid neutral) \sim N((\widehat{Margin}_{tg} \mid neutral), SE(\widehat{Margin}_{tg}))$$

where $SE(\widehat{Margin}_{tg})$ is the standard error of the prediction of margin for team *t* in game g. Figures 1-3 show tests for the assumptions stated above. Based on Figures 1 and 2, it does not appear that heteroscedasticity is a concern. Figure 3 shows that the distribution of the residuals is a little fatter in the tails than the normal distribution. Although this does violate a necessary assumption, the distribution is close enough to normal that it should not influence the analysis within the non-extreme range of predicted neutral court margins that is analyzed.

Naturally, we can define the probability of team *t* winning game g as the probability that margin for team *t* in game g is greater than zero.

$$7)\quad \Pr(Win_{tg}) = \Pr(Margin_{tg} > 0)$$

Using this framework, we can define the change in a team's win probability from moving from a neutral location as the change in the probability margin is greater than zero when predicted margin increases by the average treatment effect of home relative to neutral site given by β_1 in the fixed effects model. We can then define the change in win probability from moving to home from a neutral site as,

$$8)\quad \Delta \Pr(Win_{tg}) = \Pr(Margin_{tg} > 0 \,|\, home) - \Pr(Margin_{tg} > 0 \,|\, neutral)$$

which can be re-written using Equation 4 as:

$$9)\quad \Delta \Pr(Win_{tg}) = \Pr(Margin_{tg} > -\beta_1 \,|\, neutral) - \Pr(Margin_{tg} > 0 \,|\, neutral)$$

Each probability on the right side of Equation 9 can be calculated using the distribution of margin given neutral location in Equation 6, and the estimates for β_1 from the fixed effects model for the 2015-16 season.

Table 5 and Figure 4 show the change in a team's win probability when moving from neutral to home across a range of predicted scoring margins at a neutral site. The key takeaway is that home-court advantage has a much bigger impact on the outcome of games between more evenly matched teams. Precisely, a team with an expected margin of $-\beta_1$ against a given opponent at a neutral site receives the largest increase in win probability from playing the game at home instead, with a change in win probability of 0.116. By comparison, a team that would be an 18-point underdog or 15-point favorite on a neutral court against a given opponent have a change in win probability from playing at home of less than 0.04.

Conclusion

After controlling for team quality using a variety of methods, I find the best estimate for the average treatment effect of playing at home relative to a neutral site in NCAA Division 1 Men's Basketball to be between 3.02 and 3.23 points per game over the 2013-14, 2014-15, and 2015-16 seasons. Using these estimates, I find that teams with an expected scoring margin between -1.5 and -1.6 points against a given opponent at a neutral site have the largest change in win probability from playing the opponent on their home floor of about 12 percentage points, whereas teams that are heavy favorites or underdogs at a neutral site receive a much smaller change in win probability from a change in game location.

Table 1

Sample Summary Statistics

	Full Sample	Random Reduced	Home Reduced	Home and Home
Number of Observations	32,168	16,084	14,384	7,656
2013-14	10,722	5,361	4,784	2,486
2014-15	10,708	5,354	4,804	2,586
2015-16	10,738	5,369	4,796	2,584
Number of Games	16,084	16,084	14,384	7,656
2013-14	5,361	5,361	4,784	2,486
2014-15	5,354	5,354	4,804	2,586
2015-16	5,369	5,369	4,796	2,584
% Neutral Site	11%	11%	0%	0%
2013-14	11%	11%	0%	0%
2014-15	10%	10%	0%	0%
2015-16	11%	11%	0%	0%

Notes: The full sample contains all regular season games for the 2013-14, 2014-15, and 2015-16 seasons at the team-game level of observation. The random reduced sample is identical to the full sample but reduces the data to the game level by randomly selecting one team-game observation from the full sample. The home reduced reduces to the game level by including only the team-game observations for teams playing at home, therefore eliminating neutral site games. The "home and home" sample reduces to the game level also by restricting to home teams, and only contains observations where the matchup occurred more than once and each team played at home at least once. Among observations that met these criteria, each team's first such home game was used in the event there were multiple, to ensure that each matchup in the data has exactly two games, with each team playing at home one time.

Table 2

Panel A - Average Scoring Margin of Home Team

Mean Scoring Margin of Home Team	**(1) 2013-14**	**(2) 2014-15**	**(3) 2015-16**	**(4) All Years**
After January 1	3.18 (0.23)	3.31 (0.24)	3.19 (0.24)	3.23 (0.13)
Before January 1	9.06 (0.35)	9.15 (0.37)	9.43 (0.37)	9.22 (0.21)
All Games	5.27 (0.20)	5.39 (0.20)	5.44 (0.21)	5.37 (0.12)

Panel B – Regression of Home Team Scoring Margin on Time of Season Indicator

Outcome = Home Team Scoring Margin	**(1) 2013-14**	**(2) 2014-15**	**(3) 2015-16**	**(4) All Years**
Game Before Jan. 1st	5.884 (0.408)	5.841 (0.411)	6.238 (0.420)	5.989 (0.238)
Constant	3.181 (0.243)	3.305 (0.245)	3.194 (0.252)	3.227 (0.142)
N	4784	4804	4796	14384

Notes: Game being played before January 1st is meant to be a proxy variable for non-conference vs. in-conference play. Standard errors of estimates in parentheses. Sample size measures relates to the "All Games" mean in Panel A and regressions in Panel B.

Table 3

Panel A – Average Within-Matchup Margin Difference of Home Team

Mean Within-Matchup Scoring Margin Difference	**(1) 2013-14**	**(2) 2014-15**	**(3) 2015-16**	**(4) All Years**
All Games	3.17 (0.15)	3.23 (0.14)	3.03 (0.14)	3.15 (0.08)

Panel B – Regression of Within-Matchup Margin Difference of Home Team on Home First Indicator

Outcome = Within-Matchup Margin Diff	**(1) 2013-14**	**(2) 2014-15**	**(3) 2015-16**	**(4) All Years**
Played at Home First	0.101 (0.508)	0.268 (0.502)	-0.0186 (0.517)	0.117 (0.294)
Constant	3.119 (0.359)	3.101 (0.355)	3.042 (0.365)	3.087 (0.208)
N	2486	2586	2584	7656

Notes: Within-matchup margin difference is the difference between the scoring margin of the home team in any given game and the average of the two games they played against the given opponent. Each matchup included in the data contains exactly two games, with each team as the home team one time. Played at Home First is a dummy variable indicating if the team played the first non-neutral site game of the matchup at home. Standard errors of estimates in parentheses. Sample size measures relate to the "All Games" mean in Panel A and regressions in Panel B.

Table 4

Team and Opponent Fixed Effects Regressions on Scoring Margin

Outcome = Scoring Margin	**(1)** **2013-14**	**(2)** **2014-15**	**(3)** **2015-16**
Team Home	3.023 (0.156)	3.204 (0.155)	3.205 (0.158)
Team Away	-3.023 (0.156)	-3.204 (0.155)	-3.205 (0.158)
Constant	-5.77e-14 (3.11e-7)	-1.06e-13 (2.00e-7)	-2.31e-13 (2.99e-7)
N	10722	10708	10738

Notes: Sample includes observations for each team for every game, so number of games for each season is N/2. Standard errors in parentheses and clustered on game. Team and opponent fixed effects used to control for relative team and opponent quality. Necessity of common opponents for useful estimates of team and opponent fixed effects prevent this model from being used across years.

Table 5

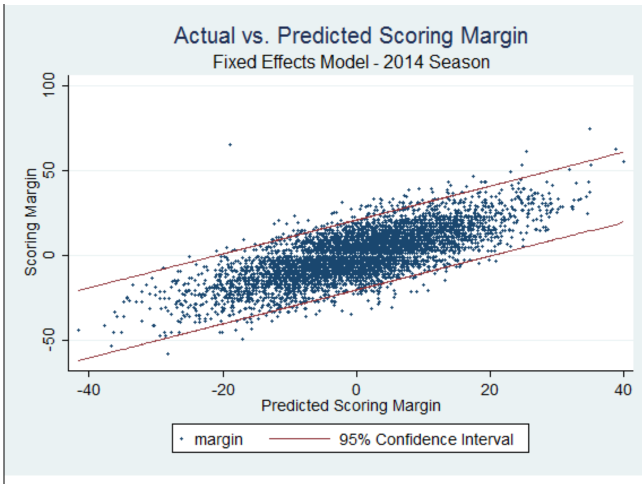
Expected win probability changes from playing at home, 2015-16 season, fixed effects model

Predicted Margin (Neutral)	Pr(win) Neutral	Pr(win) Home	∆ Pr(win) Home
-20	3.5%	6.4%	2.9%
-19	4.2%	7.6%	3.4%
-18	5.1%	8.9%	3.9%
-17	6.1%	10.5%	4.4%
-16	7.3%	12.3%	5.0%
-15	8.6%	14.2%	5.6%
-14	10.2%	16.3%	6.2%
-13	11.9%	18.7%	6.8%
-12	13.8%	21.2%	7.5%
-11	15.9%	24.0%	8.1%
-10	18.2%	26.9%	8.7%
-9	20.7%	29.9%	9.3%
-8	23.4%	33.2%	9.8%
-7	26.2%	36.5%	10.3%
-6	29.3%	40.0%	10.7%
-5	32.5%	43.5%	11.1%
-4	35.8%	47.1%	11.3%
-3	39.3%	50.8%	11.5%
-2	42.8%	54.4%	11.6%
-1	46.4%	58.0%	11.6%
0	50.0%	61.5%	11.5%
1	53.6%	64.9%	11.3%
2	57.2%	68.2%	11.0%
3	60.7%	71.4%	10.6%
4	64.2%	74.4%	10.2%
5	67.5%	77.2%	9.7%
6	70.7%	79.9%	9.2%
7	73.8%	82.3%	8.6%
8	76.6%	84.6%	7.9%
9	79.3%	86.6%	7.3%
10	81.8%	88.5%	6.7%
11	84.1%	90.2%	6.0%
12	86.2%	91.7%	5.4%
13	88.1%	93.0%	4.8%
14	89.8%	94.1%	4.3%
15	91.4%	95.1%	3.7%
16	92.7%	96.0%	3.3%
17	93.9%	96.7%	2.8%
18	94.9%	97.3%	2.4%
19	95.8%	97.8%	2.0%
20	96.5%	98.3%	1.7%

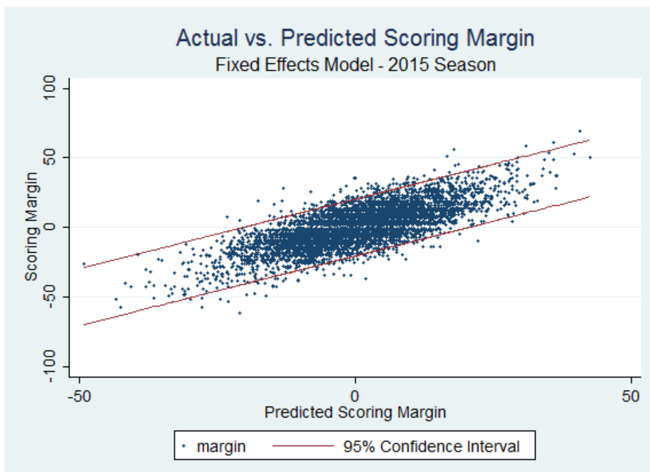
Notes: Generating forecast standard errors is not possible after clustering standard errors, so the reduced data featuring one randomly selected team-game observation for each game is used.

Figure 1

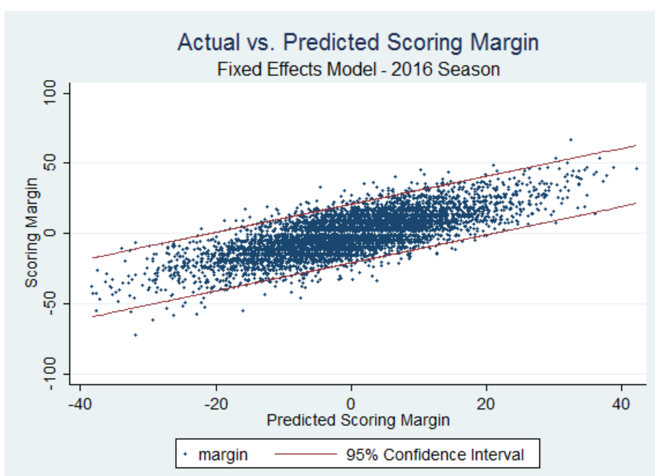
Predicted to Actual Scoring Margin – Team and Opponent Fixed Effects Model



a)



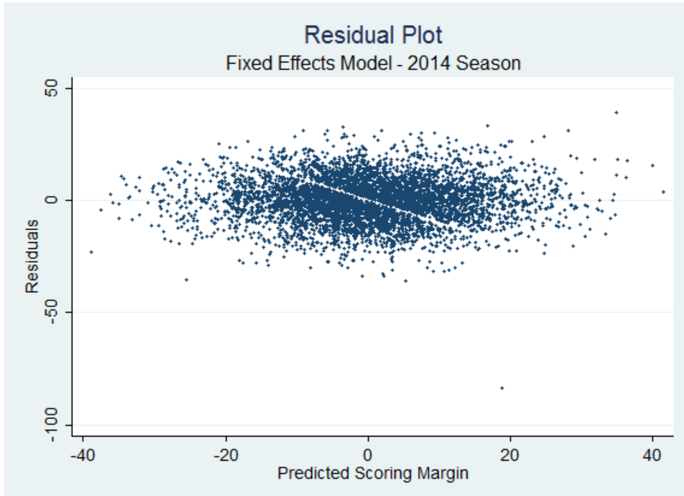
b)



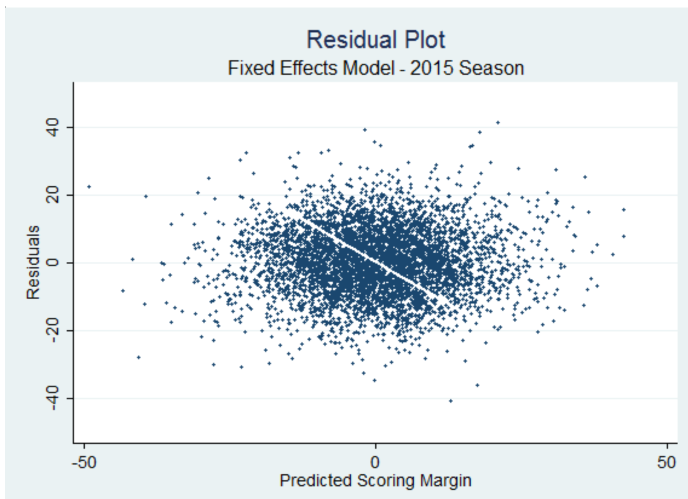
c)

Figure 2

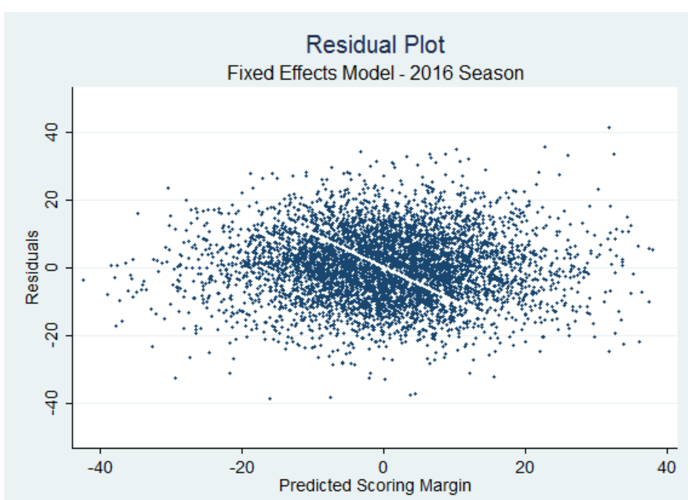
Predicted Scoring Margin Residual Plots – Team and Opponent Fixed Effects Model



a)



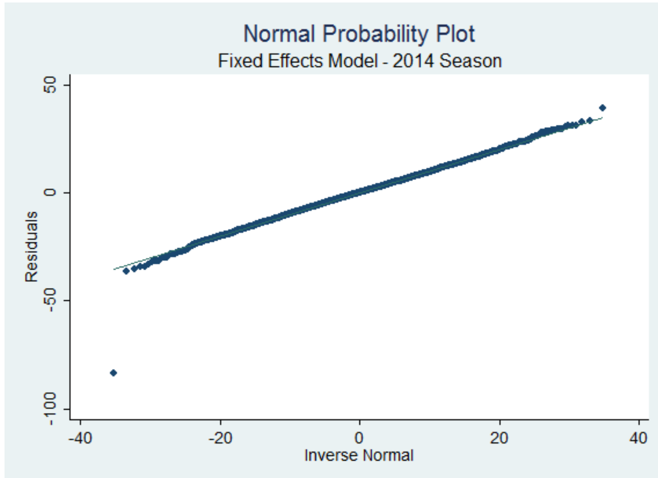
b)



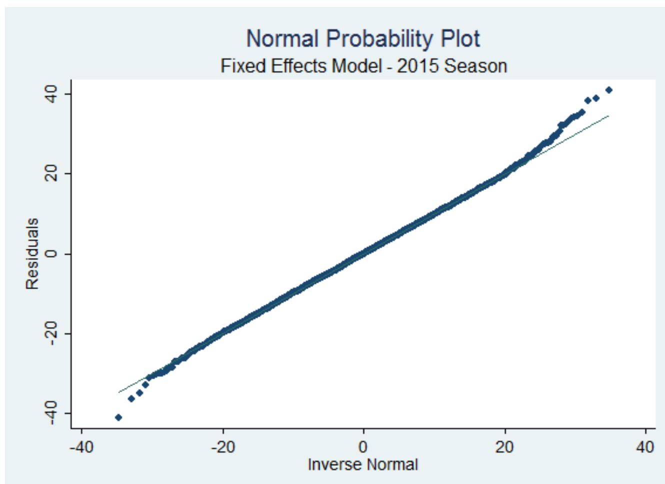
c)

Figure 3

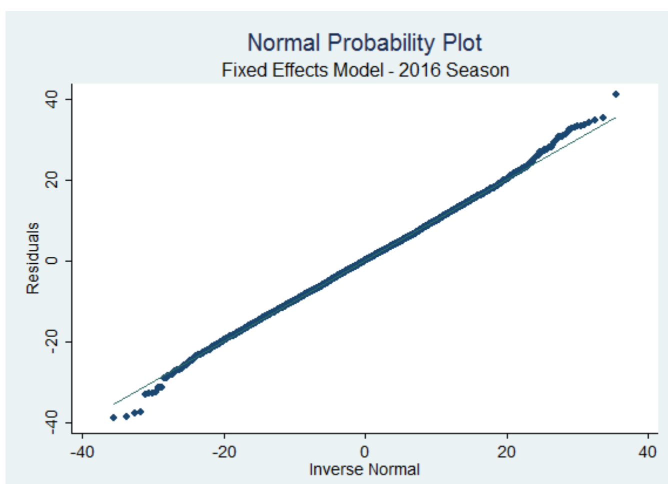
Normal Probability Plots of Residual – Team and Opponent Fixed Effects Model



a)



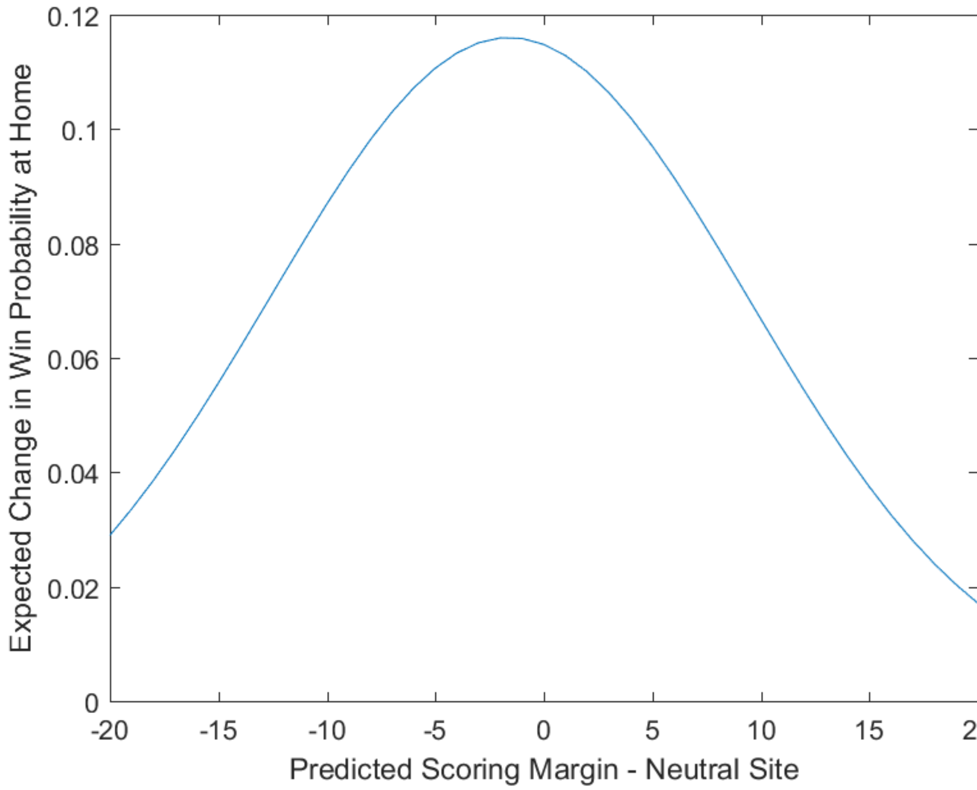
b)



c)

Figure 4

Change in Win Probability from Playing at Home Relative to Neutral Site by Predicted Margin



Notes: Shows the expected change in win probability when moving from neutral site to home game, given the predicted scoring margin of game if at neutral site. Team and opponent fixed effects model for 2015-16 season used to determine point value of home court advantage.

Replication Files

All tables and figures reported in this paper can be replicated using the files provided at:

https://github.com/jordkeen/HCProjectJK_Nov28Submission