Jordan Keener

No Place Like Home: Estimating the True Value of Home-Court Advantage in College 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 the 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 playing "on the road". 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 and competitive incentives lead to an endogeneity problem where game location is correlated with factors that influence game outcomes, most problematically 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 1 Men's Basketball from 2013 to 2016. 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. I find that home-court advantage to be worth between 3.02 and 3.32 points per game in terms of scoring margin over this time, and that teams who would be approximately 1.5 point underdogs against a given opponent on neutral court receive the largest increase in win probability from instead playing the opponent at home.

Data

To estimate the effect of playing at home relative to a neutral court, I use game result data for the 2013-14, 2014-15, and 2015-16 seasons. The level of observation is team-game, meaning each game appears twice in the dataset, with each team appearing as the team of interest. The game number variable represents the sequential count of games played by each team. For each observation, whether the team of interest played the game at home or away is recorded using the dummy variables *teamHome* and *teamAway*, as well as the scoring margin (team score minus opponent score) of the game. Additionally, the dummy variable *neutral* indicates whether the game was played at a neutral site or not, with *teamHome*, *teamAway*, and *neutral* being mutually exclusive and collectively exhaustive. An example snapshot of six observations is provided in Table 1.

The only initial data restriction is to limit the dataset to games between two Division 1 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 1 team, this serves as an adequate proxy, as each Division 1 team plays at least 15 games in a season, and no non-Division 1 team plays 15 or more games against Division 1 opponents in a season.

Simple Home Court Advantage Models

A naïve regression of *margin* on *teamHome* and *teamAway*[1] would be one simple measure of the impact of game location on scoring margin. Across all three seasons in the data, the estimate for the coefficient on the *teamHome* from running such a regression is 5.36. [2]

However, the coefficient on *teamHome* provides simply a measure of the average scoring margin of teams playing at home over the given time period.[3] 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. Because of these factors, the coefficient on *teamHome* likely has an upward bias, as a team playing at home appears to be positively correlated with team quality relative to opponent, which would increase expected scoring margin. I propose two alternative models that 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

Within Matchup Estimator

The first method estimates the effect of changing game location from away to home within a matchup, which is herein defined as a team-opponent combination.[4] To estimate this, we 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 contains exactly two games where each team is the home team once.

---

[1] teamHome and teamAway are the game location dummy variables used, omitting the third possible value of neutral. This is done to simplify the interpretation of the coefficient of teamHome by making the reference provided by the intercept a game played on a neutral court.
[2] See Table 2 for a full table of regression results.
[3] For example, it is well known that Indiana is a superior basketball team to Purdue, so attributing the entire difference in margin for the game Table 1 to the fact that Indiana played at home would lead to a biased estimate of the true treatment effect of playing at home. Go Hoosiers!
[4] 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.

The basic framework for the within estimator model is given by,

$$1)\quad (Margin_{tg} - \overline{Margin_{to}}) = \beta_0 + \beta_1 teamHome_{tg} + \varepsilon_{tg}$$

where *teamHome* is a dummy variable indicating whether team *t* played game *g* at home, and ε_{tg} is a mean-zero error term.

The transformed outcome in this model is now the difference between scoring margin for team *t* in game *g* and the average margin of team *t* when facing opponent *o*. Using this model, β_1 can be interpreted as the expected change in the outcome when team *t* plays game *g* at home relative to playing on the road.

The restriction eliminating neutral site games alters the interpretation of the teamHome coefficient slightly compared to its interpretation in the naïve regression of *margin* on *teamHome* and *teamAway*, as it is now the expected change in the deviation of margin from the matchup average when a team is home relative to playing on the road, rather than comparing to a neutral site game. However, the coefficient on *teamHome* in this model can simply be divided by two to obtain the estimate for home-court advantage as previously defined.[5]

Because this model only evaluates the impact of game location on within matchup variation in scoring margin[6], this model effectively controls for relative team and opponent quality, under the key assumption that variation in team-to-oppenent relative quality at neutral site 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, relative team to opponent quality is the only factor that must be controlled for, so 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.

Under this assumption, the coefficient on *teamHome* represents the true average treatment effect of moving from an away game to a home game, controlling for team and opponent relative quality. Using the method described above to convert this to a home-to-neutral comparison rather than home-to-away, this model estimates the home-court advantage to be about 3.2 points per game over the 2013-14, 2014-15, and 2015-16 seasons. This result is consistent with the hypothesis that the naïve regression of scoring margin on just game location dummy variables

[5] It can also be obtained by multiplying the intercept by -1.

[6] Now, using the previous example, controlling for Indiana's clear superiority in talent relative to Purdue.

overestimates the value of home-court advantage by not controlling for team and opponent quality.

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,

$$2)\quad Margin_{tg} = \beta_0 + \beta_1 teamHome_{tg} + \beta_2 teamAway_{tg} + \gamma_t team_t + \alpha_o opponent_{o_{tg}} + \varepsilon_{tg}$$

where *teamHome* is a dummy variable indicating if team *t* played at home in game *g*, *teamAway* is a dummy variable indicating if team *t* played on the road in game *g* [7], *team* represents dummy variables for each team *t* indicating if team[8] *t* played in the game, and *opponent* represents 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 team, holding fixed game location and opponent quality. More formally, the team fixed effects coefficients for each team *t* represent,

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

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

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

which can be combined into the full model given below:

$$3)\quad E(Margin_{Team\ t,\ Opponent\ o,\ g}) = \beta_0 + \beta_1 teamHome_{tg} + \beta_2 teamAway_{tg} + E\left(Margin_{Team\ t,\ reference\ opponent}\right) + E(Margin_{reference\ team,\ Opponent\ o})$$

[7] The *teamHome* and *teamAway* variables leave neutral site as the reference state, given by the intercept.
[8] 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.

This equation is particularly useful when it is necessary to make predictions on margin between a given team and opponent, which will be revisited in the next section evaluating the impact of game location on win probabilities.

Unsurprisingly, this model yields very similar results to the within matchup estimates.[9] The coefficient on *teamHome* is 3.02, 3.20, and 3.21 for the 2013-14, 2014-15, and 2015-16 seasons respectively. The necessity of common opponents to estimate team and opponent fixed effects prevents this model from being applied across seasons.

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 that are correlated with game location, which would be differential treatment effects of game location by team. This does not prevent 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*.

$$4)\quad (\widehat{Margin}_{tg} \mid neutral) = \gamma Team_t + \alpha_o Opponent_{o_{tg}}$$

[9] A full table of regression results across years and model specifications with standard errors is available in Table 2.

Assuming heteroscedasticity and normally distributed errors[10], the standard error of the forecast for predicted margin can 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,

$$5)\quad (Margin_{tg}|neutral) \sim N((\widehat{Margin}_{tg}|\ neutral), SE(\widehat{Margin}_{tg})$$

where $SE(\widehat{Margin}_{tg})$ is the standard error of the prediction of margin for team t in game g. 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.

$$6)\quad \Pr(Win_{tg}) = \Pr(Margin_{tg} > 0)$$

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,

$$7)\quad \Delta\Pr(Win_{tg}) = \Pr(Margin_{tg} > 0 |\ home) - \Pr(Margin_{tg} > 0 |\ neutral)$$

which can be re-written as using Equation 2:

$$8)\quad \Delta\Pr(Win_{tg}) = \Pr(Margin_{tg} > -\beta_1 |\ neutral) - \Pr(Margin_{tg} > 0 |\ neutral)$$

Each probability on the right side of Equation 8 can be calculated using the distribution of margin given neutral location in Equation 5.

Table 3 and Figure 4 show the change in 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 -β_1 against a given opponent at a neutral site receives the largest increase in win probability from playing the game at home instead.

[10] Tests for each shown in Figures 1-3

Table 1

Example of team-game panel data structure

gameNum	***team***	***opponent***	***margin***	***teamHome***	***teamAway***	***neutral***
1	Indiana-2016	Purdue-2016	16	1	0	0
2	Indiana-2016	Kentucky-2016	7	0	0	1
1	Purdue-2016	Indiana-2016	-16	0	1	0
1	Kentucky-2016	Louisville-2016	-4	1	0	0
2	Kentucky-2016	Indiana-2016	-7	0	0	1
1	Louisville-2016	Kentucky-2016	4	0	1	0

Table 2

Regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
margin	2013-14 Naive	2013-14 Within	2013-14 FE	2014-15 Naive	2014-15 Within	2014-15 FE	2015-16 Naive	2015-16 Within	2015-16 FE	2013-2016 Naive	2013-2016 Within
teamHome	5.268 (0.450)	6.339 (0.207)	3.023 (0.346)	5.389 (0.462)	6.469 (0.200)	3.204 (0.350)	5.431 (0.465)	6.065 (0.203)	3.212 (0.352)	5.363 (0.265)	6.290 (0.117)
teamAway	-5.268 (0.450)		-3.023 (0.346)	-5.389 (0.462)		-3.204 (0.350)	-5.431 (0.465)		-3.212 (0.352)	-5.363 (0.265)	
_cons	-1.20e-14 (0.404)	-3.169 (0.147)	-4.91e-14 (3.214)	7.59e-14 (0.417)	-3.235 (0.141)	3.99e-14 (2.767)	2.75e-14 (0.418)	-3.033 (0.144)	-3.06e-14 (2.876)	2.88e-13 (0.239)	-3.145 (0.0830)
N	10722	4972	10722	10710	5172	10710	10792	5168	10792	32224	15312

- Naïve and Fixed Effects models use all games where both team and opponent are NCAA Division 1 over the given time period
- Within Matchup model restricts data to observations where the team played opponent multiple times, at least once at home and once on the road, and to only the first such home and away game

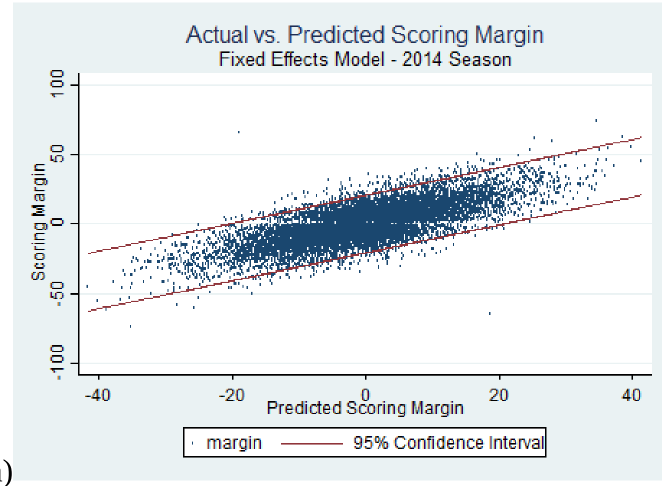
Table 3

Expected win probability changes from playing at home, 2016 fixed effects model

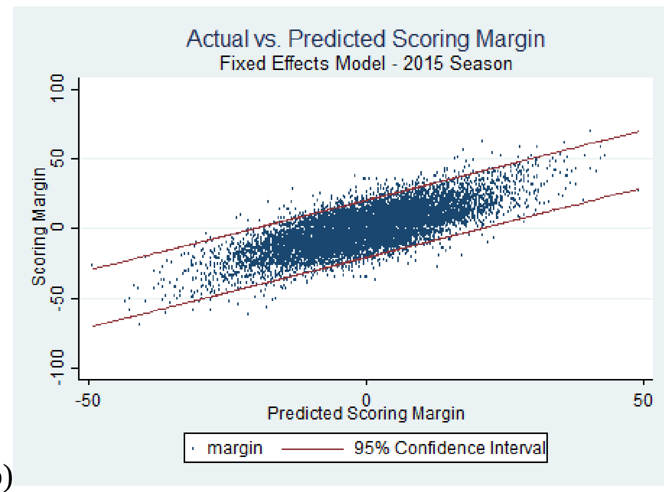
Predicted Margin (Neutral)	Pr(win) Neutral	Pr(win) Home	Δ Pr(win) Home
-20	3.0%	5.7%	2.7%
-19	3.7%	6.9%	3.2%
-18	4.5%	8.2%	3.7%
-17	5.5%	9.8%	4.2%
-16	6.6%	11.5%	4.8%
-15	7.9%	13.4%	5.5%
-14	9.4%	15.5%	6.1%
-13	11.1%	17.9%	6.8%
-12	13.0%	20.4%	7.5%
-11	15.1%	23.2%	8.1%
-10	17.4%	26.2%	8.8%
-9	19.9%	29.3%	9.4%
-8	22.6%	32.6%	10.0%
-7	25.5%	36.1%	10.6%
-6	28.6%	39.7%	11.0%
-5	31.9%	43.3%	11.4%
-4	35.3%	47.0%	11.7%
-3	38.9%	50.8%	11.9%
-2	42.5%	54.5%	12.0%
-1	46.3%	58.2%	12.0%
0	50.0%	61.9%	11.9%
1	53.7%	65.4%	11.6%
2	57.5%	68.8%	11.3%
3	61.1%	72.0%	10.9%
4	64.7%	75.1%	10.5%
5	68.1%	78.0%	9.9%
6	71.4%	80.7%	9.3%
7	74.5%	83.1%	8.7%
8	77.4%	85.4%	8.0%
9	80.1%	87.4%	7.3%
10	82.6%	89.3%	6.6%
11	84.9%	90.9%	6.0%
12	87.0%	92.4%	5.3%
13	88.9%	93.6%	4.7%
14	90.6%	94.7%	4.1%
15	92.1%	95.7%	3.6%
16	93.4%	96.5%	3.1%
17	94.5%	97.1%	2.6%
18	95.5%	97.7%	2.2%
19	96.3%	98.2%	1.9%
20	97.0%	98.5%	1.6%

Figure 1

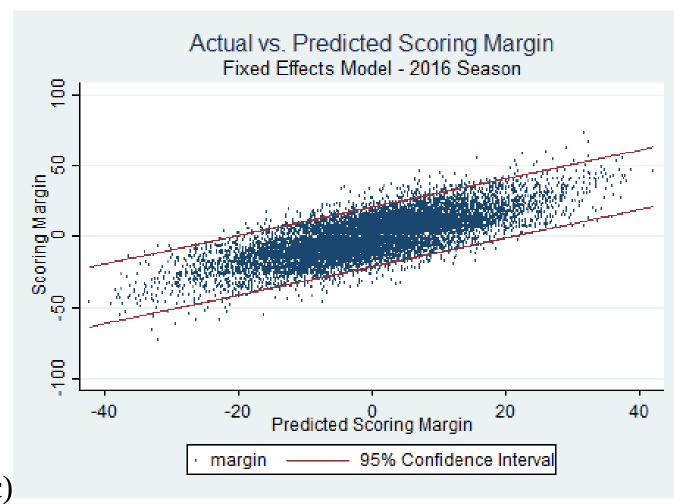
Predicted to Actual Scoring Margin – Team and Opponent Fixed Effects Model



a)



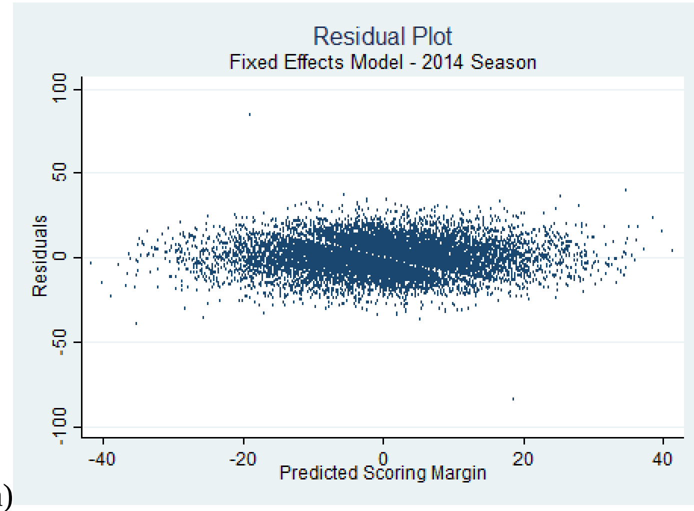
b)



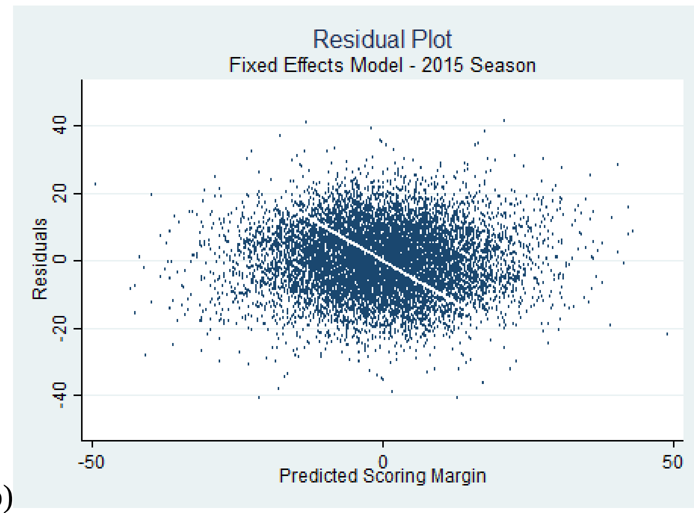
c)

Figure 2

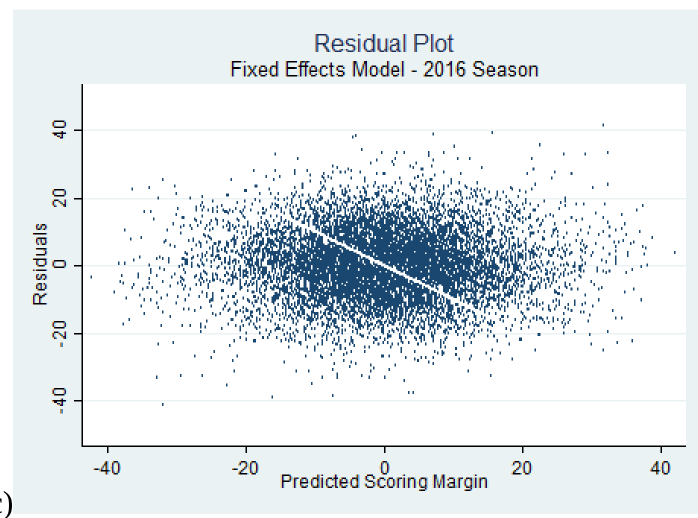
Predicted Scoring Margin Residual Plots – Team and Opponent Fixed Effects Model



a)



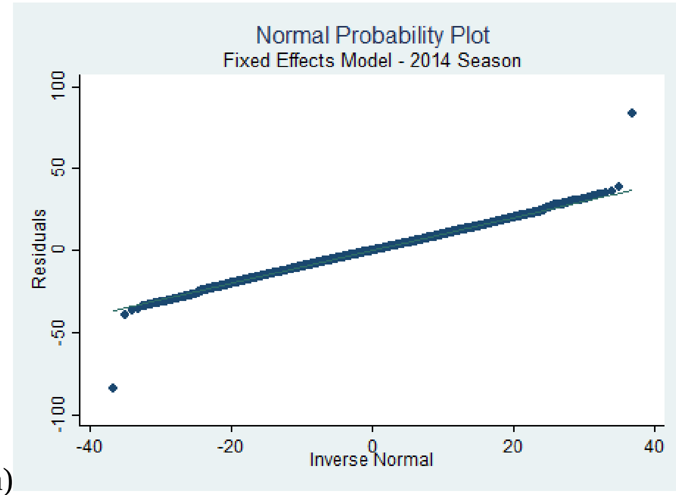
b)



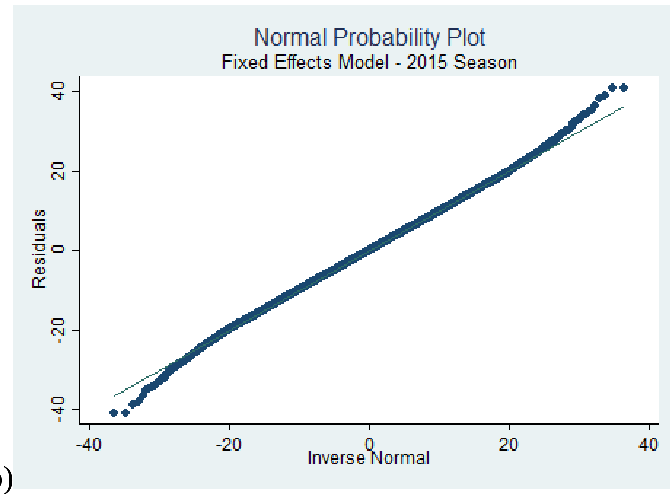
c)

Figure 3

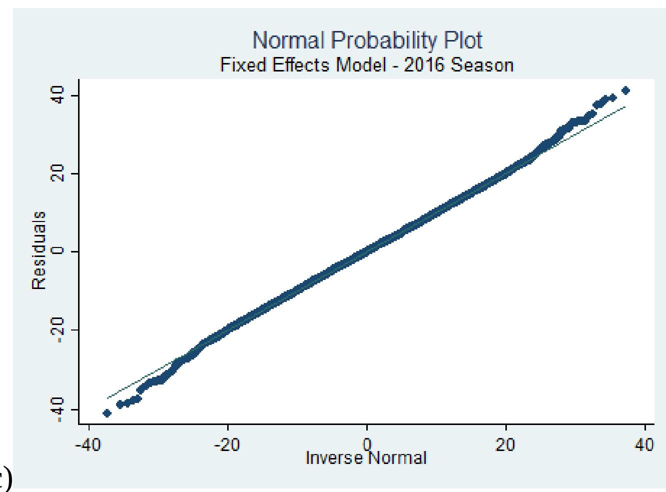
Normal Probability Plots of Residual – Team and Opponent Fixed Effects Model



a)



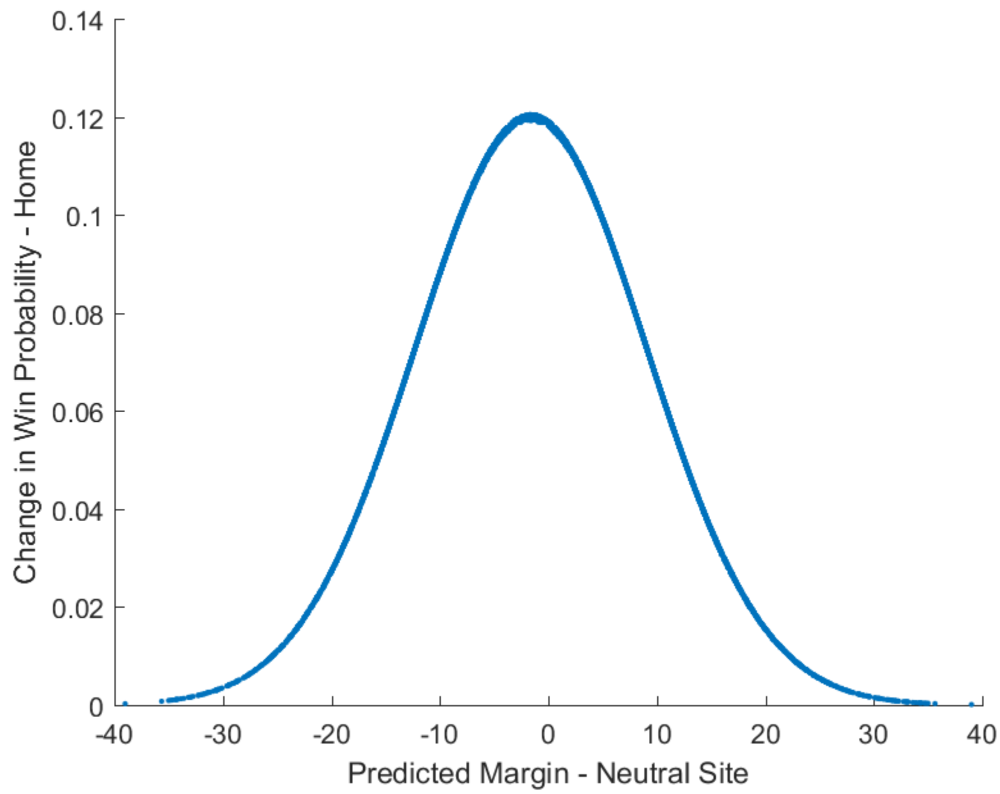
b)



c)

Figure 4

Change in Win Probability from playing at home – estimates from 2015-16 season team and opponent fixed effects model



Replication Files

All results reported in this paper can be replicated using the files provided at:

https://github.com/jordkeen/HomeCourt-October31_16Submission