Research Article

C.J.(Chaojie) Duan*

How Much Does Home-Field Advantage Matter in Professional Soccer?

A Multilevel Bayesian Investigation

Received 5/21/2018

Abstract: Home Field Advantage (HFA) was traditionally defined in terms of winning percentage of home games at the team level. In this article, we present a hierarchical model of HFA, spanning from the top sport level to the middle league level and all the way to lowest club level. Using scoring performance data from ESPN FC, we fit a Bayesian multilevel nested model to the parameters in the hierarchical model of HFA, allowing information obtained from the team level to inform the inferences at the upper league and sport levels.

Keywords: European Professional Soccer Leagues, Home Field Advantage, Poisson generative process, Stan

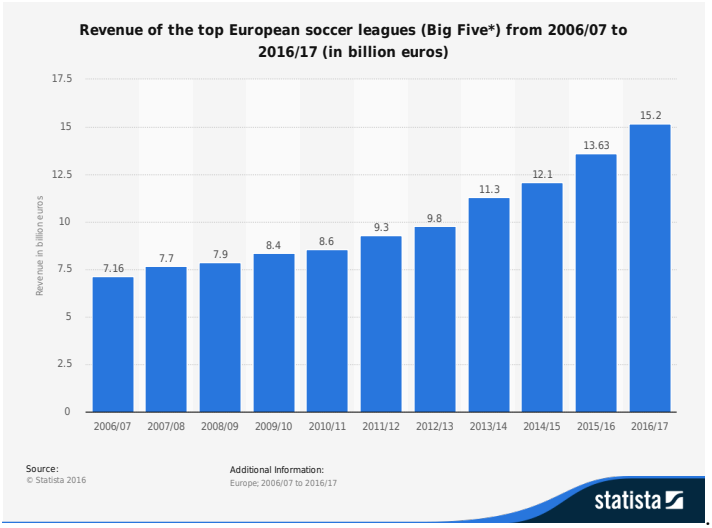
1 Introduction

In professional team sports, the term home field advantage (HFA) – also called home advantage, home ground or home court advantage, defender's advantage, home-ice advantage – describes the benefit that the home team is believed to gain over the visiting opponent. Its scientific definition is "the consistent finding that home teams in sport competition win over 50% of the games played under s balanced home and away schedule" (Courneya and Carron, 1992, p. 13). Due to the existence of HFA, many vital games, such as playoff or elimination matches, in major professional sports have special rules for determining which match is played at which place. As shown in Figure 1, the combined revenue of the Big Five European soccer leagues (English Premier League, Spanish La Liga, French Ligue 1, Bundesliga, Italian Serie A) more than doubled to 15 billion euros in 10 years from 2006/07 to 2016/17. The financial implications might partially explain UEFA's (the Union of European Football Associations) decision that

***Corresponding author: C.J.(Chaojie) Duan,** Dulun Consulting Group, research@dulun.com

a second leg of any Champions League knock-off series is favorable to playing away with the the scores still in balance after the first leg competition (Atkins, 2013).

Fig. 1: *Revenue of the top European soccer leagues (Big Five*) from 2006/07 to 2016/17 (in billion euros)*



The existence of HWP (home winning percentage) -denominated HFA measure has been well documented for a variety of sports. In their book *Scorecasting*, Moskowitz and Wertheim (2012) compiled the HWPs in all the major sports with some datasets going back as further as 1903 for MLB and 1966 for NFL. MLS figures date back to only 2002, but show the strongest evidence of HWP of 69.1%. MLB figures, on the other hand, yield the lowest HWP of only 53.9%. This disparity raises an important high-profile question: "Are all sports created equal in terms of HFA?". A subsequent but related question is "Is HFA primarily determined by the sport being played or teams and who play the sport?". Answering such questions demands a completely new way of conceptualizing HFA and signals a major departure from the currently reigning framework proposed by Courneya and Carron (1992) back in 1992.

A second motivator for this study is related to the treatment of sports data in general, and scoring in soccer matches in particular. HWP based measures tend to upstage and upgrade the originally discrete count-based outcome to continuous type, while ignoring the underlying data generating process.

The contributions we made in this paper are:

1. Propose a fresh new vertical hierarchical model of HFA, complementing the existing horizontal framework.
2. Highlight the different generative process underlying most sports performance metrics and suggest corresponding approaches for analysis.
3. Reveal sources of HFA simultaneously at sport, league, team levels.
4. Presenting a new way of measuring HFA via contrast the same performance metric at home and away venues.

The remainder of the paper is organized as

2 Review of Literature

3 The Hierarchical Model of HFA

2

4 Data

Tab. 1: Descriptive Statistics

	Mean	Median	Std. Dev.	Min.	Max.	Skewness	Kurtosis
MHG	3.634	4	1.676	0	9	0.246	0.034
MAG	2.884	3	1.676	0	10	0.627	0.786

5 Computation and Results

We run 4 chains using the default sampler in Stan, the HMC variant of No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) and set

The model estimates are shown in figure 3 as shift from the 0. The outer contour lines show the 99.5% credible intervals, while the shaded area underneath

Fig. 2: The Hierarchical Model of Home Field Advantage

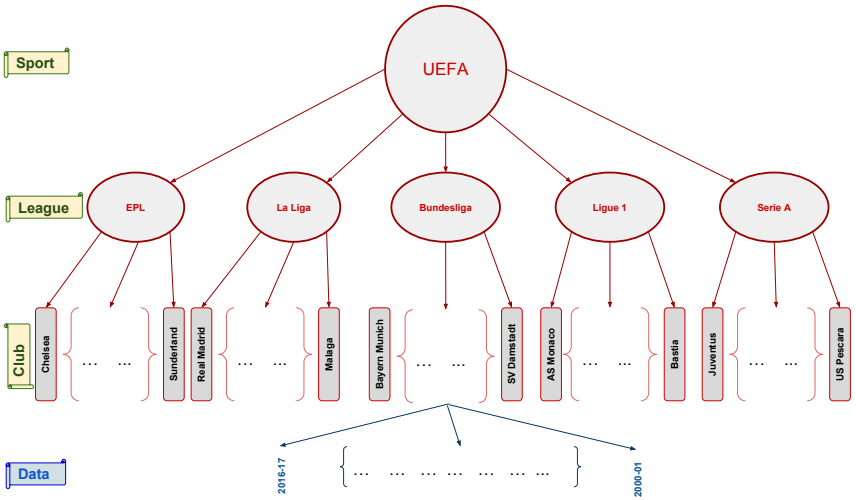
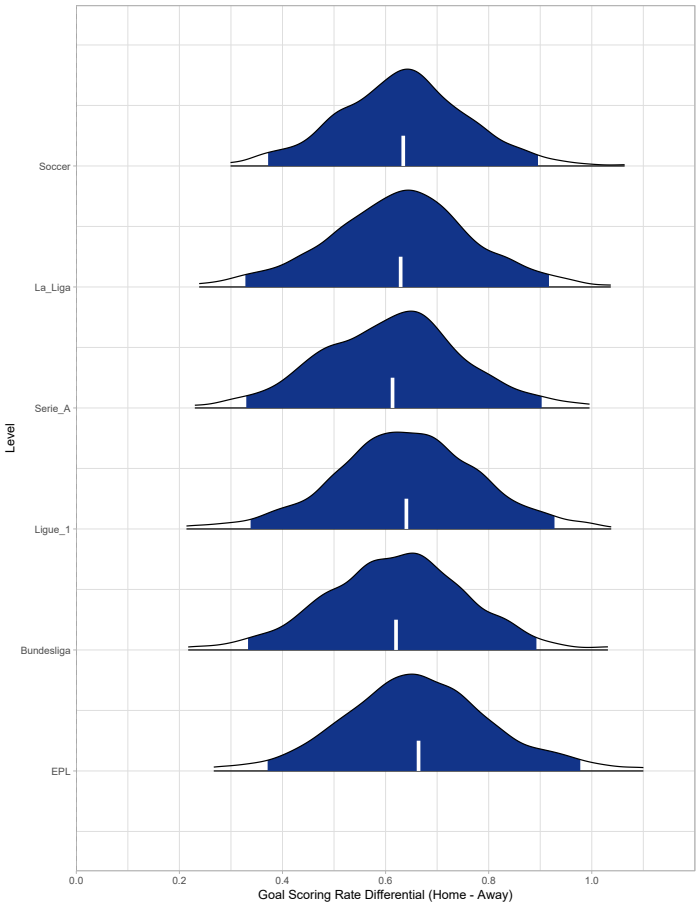


Fig. 3: Home Field Advantage Posterior Plot at Sport and League Levels



covers the corresponding 95% credible interval. The light bar in the middle represents the mean.

Acknowledgment: We would like to thank ESPN FC for compiling the season-level club performance data and allow public access.

Fig. 4: Home Field Advantage Posterior Plot for La Liga Teams

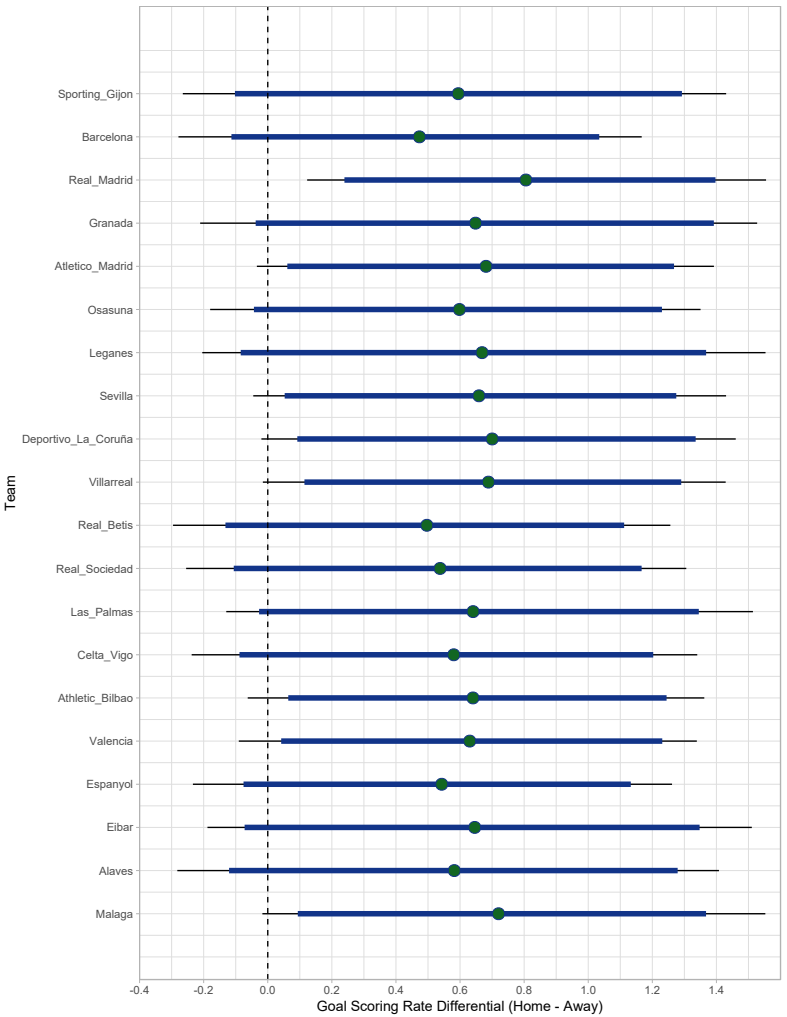


Fig. 5: Home Field Advantage Posterior Plot for Serie A Teams

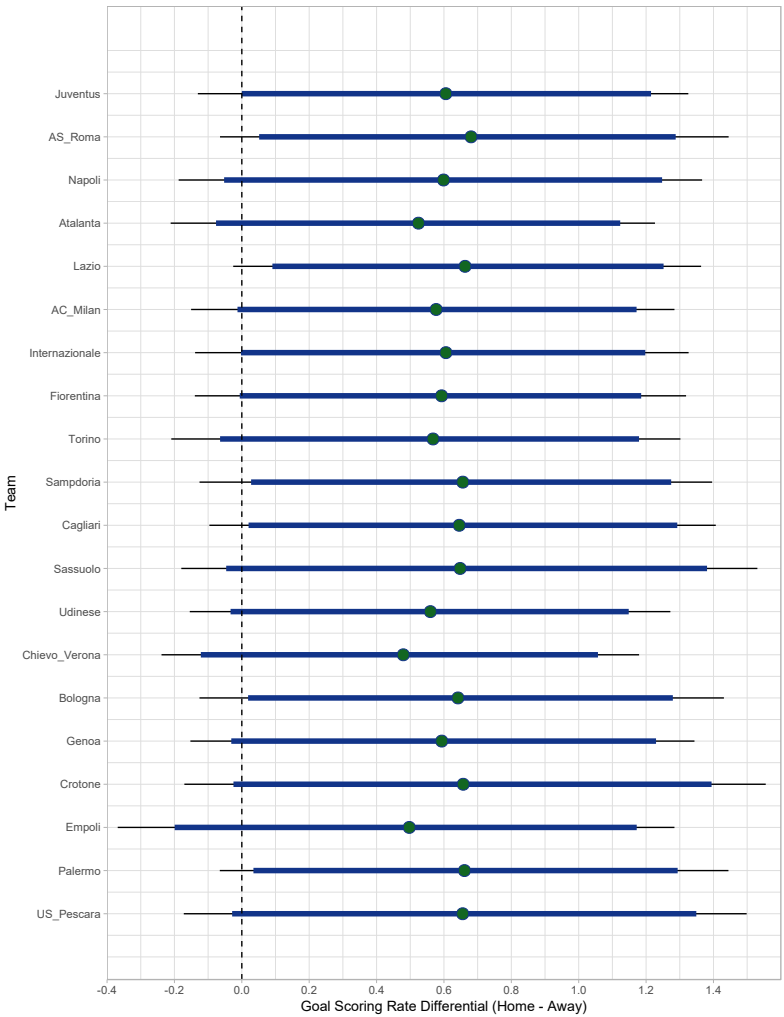


Fig. 6: Home Field Advantage Posterior Plot for Ligue 1 Teams

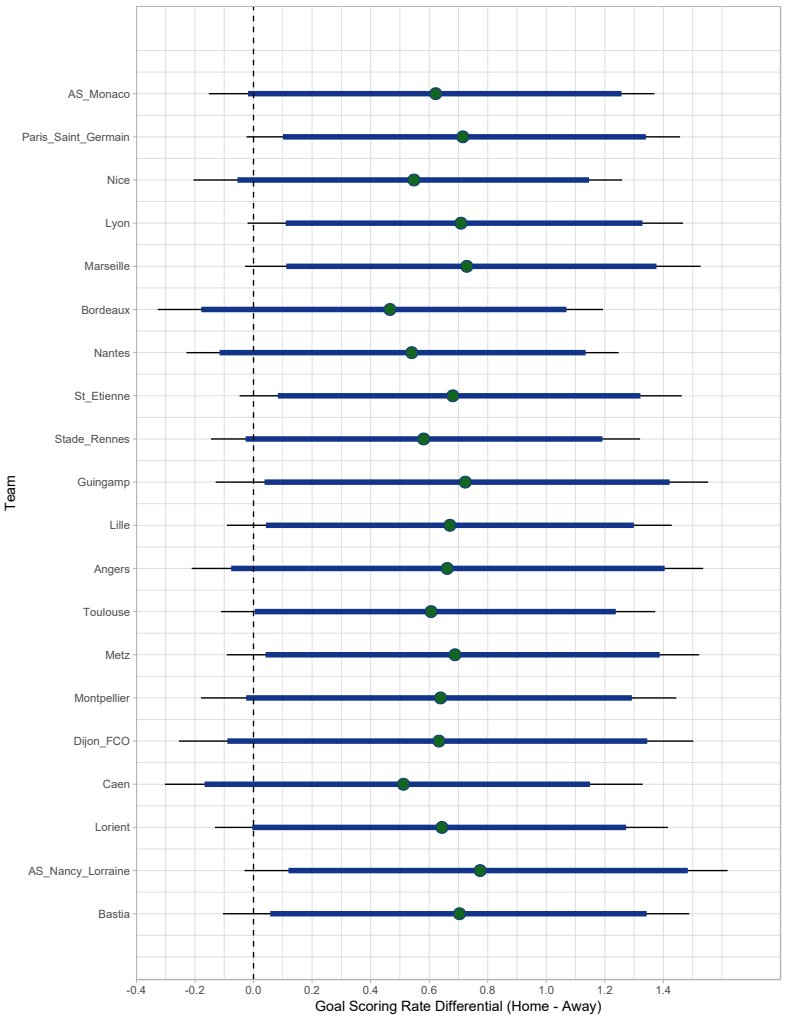


Fig. 7: Home Field Advantage Posterior Plot for Bundesliga Teams

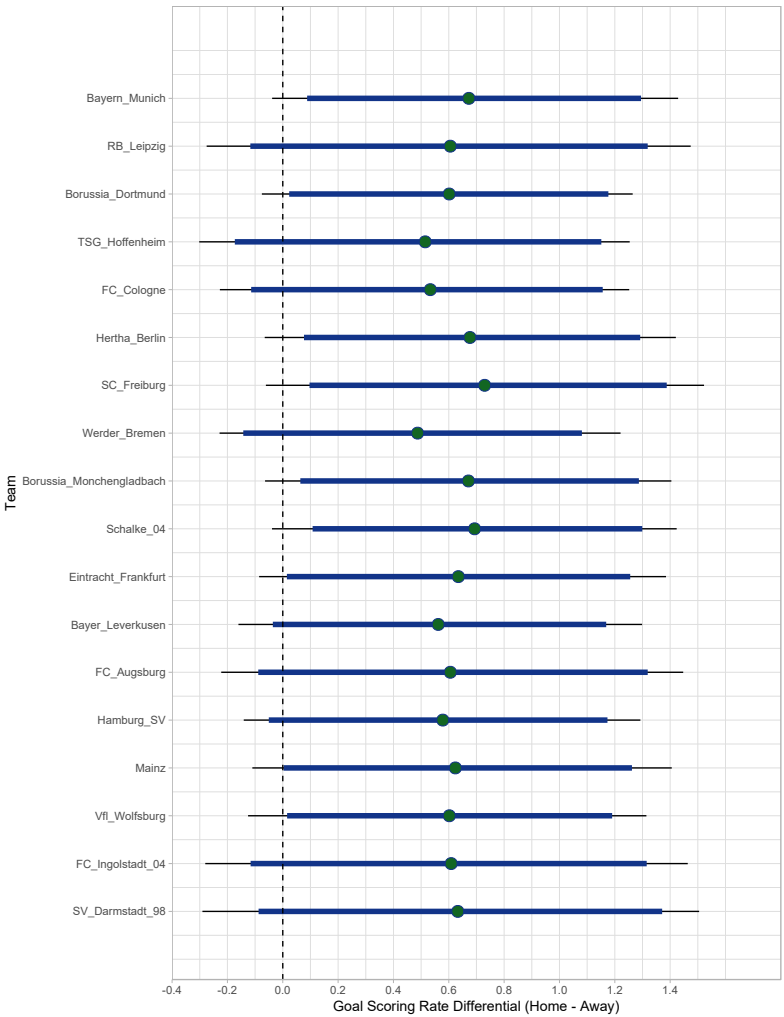
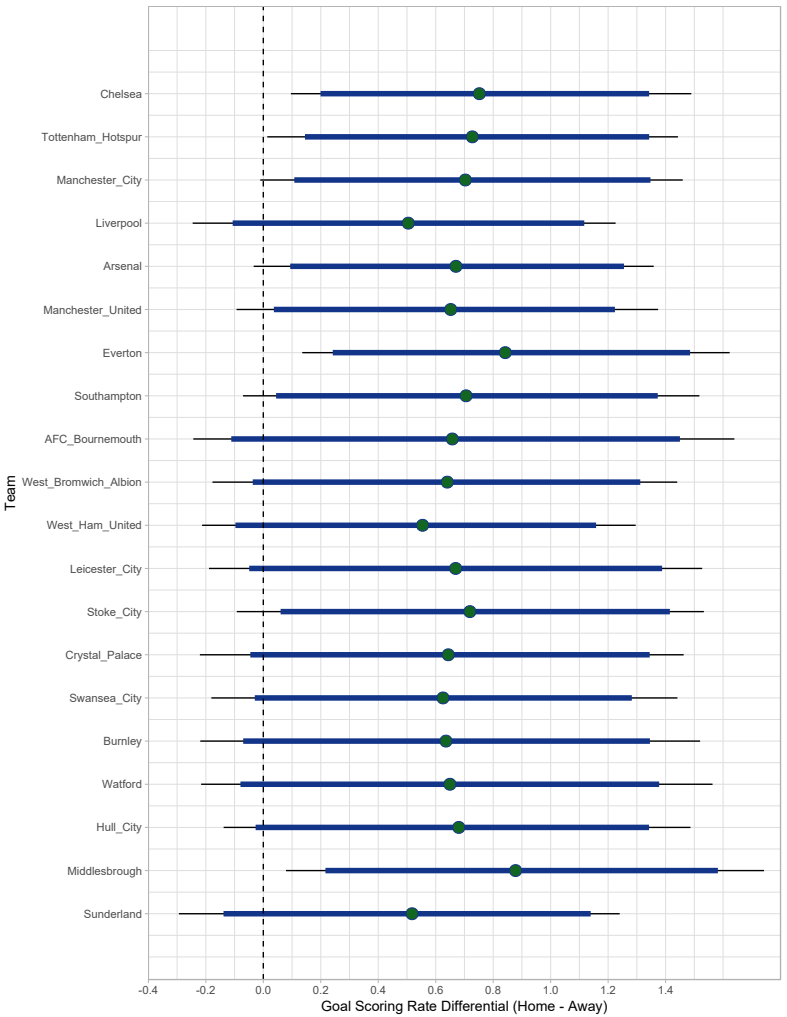


Fig. 8: Home Field Advantage Posterior Plot for English Premier League Teams



References

Atkins, C. (2013). How much does home-field advantage matter in soccer? *B/R*.

Courneya, K. S. and Carron, A. V. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Psychology*, 14(1):13–27.

Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Moskowitz, T. and Wertheim, L. J. (2012). *Scorecasting: The hidden influences behind how sports are played and games are won*. Three Rivers Press (CA).