A Comprehensive Survey of the Home Advantage in American Football

Thompson Bliss [1], Luke Benz [2], and Michael Lopez [1]

[1]National Football League

[2]Harvard T.H. Chan School of Public Health, Department of Biostatistics

Introduction

The existence and justification to the home advantage (HA)—the benefit a sports team receives when playing at home—has been studied across sport. Most research on this topic is restricted to individual leagues in short time frames, which limits a thorough understanding of the drivers of home advantage and prevents extrapolation. Using nearly two decades of data from the National Football League (NFL), the National Collegiate Athletic Association (NCAA), and high schools from across the United States, we provide a framework for estimating home advantage. **In particular, we are interested in the question of whether/how much home advantage has changed over time.**

Statistical Framework

Our statistical framework leverages STAN to fit a suite of Bayesian linear regression based paired-comparison models, with various plausible temporal trends for home advantage: a constant home advantage over time, a linear home advantage trend over time, and a trend in which home advantage is allowed to vary freely over time.

Priors:

$\theta_i \sim \text{Normal}(0, \sigma_\theta^2)$ Team Strength

$\sigma_\theta \sim \text{InverseGamma}(1,1)$ Team Strength Variance

$\alpha \sim \text{Normal}(0, \sigma_\alpha^2)$ Home Advantage

$\sigma_\alpha \sim \text{InverseGamma}(1,1)$ Home Advantage Variance

$\sigma \sim \text{InverseGamma}(1,1)$ Score Differential Variance

Models:

Let Y_{ijt} denote the score differential of a game where team i plays at team j in year t

$Y_{ijt} \sim N(\theta_i - \theta_j + \alpha, \sigma^2)$ **Constant HA** (1)

$Y_{ijt} \sim N(\theta_i - \theta_j + \alpha_0 + \alpha_1 t, , \sigma^2)$ **Linear HA** (2)

$Y_{ijt} \sim N(\theta_i - \theta_j + \alpha_t, \sigma^2)$ **Time-Varying HA** (3)

Results

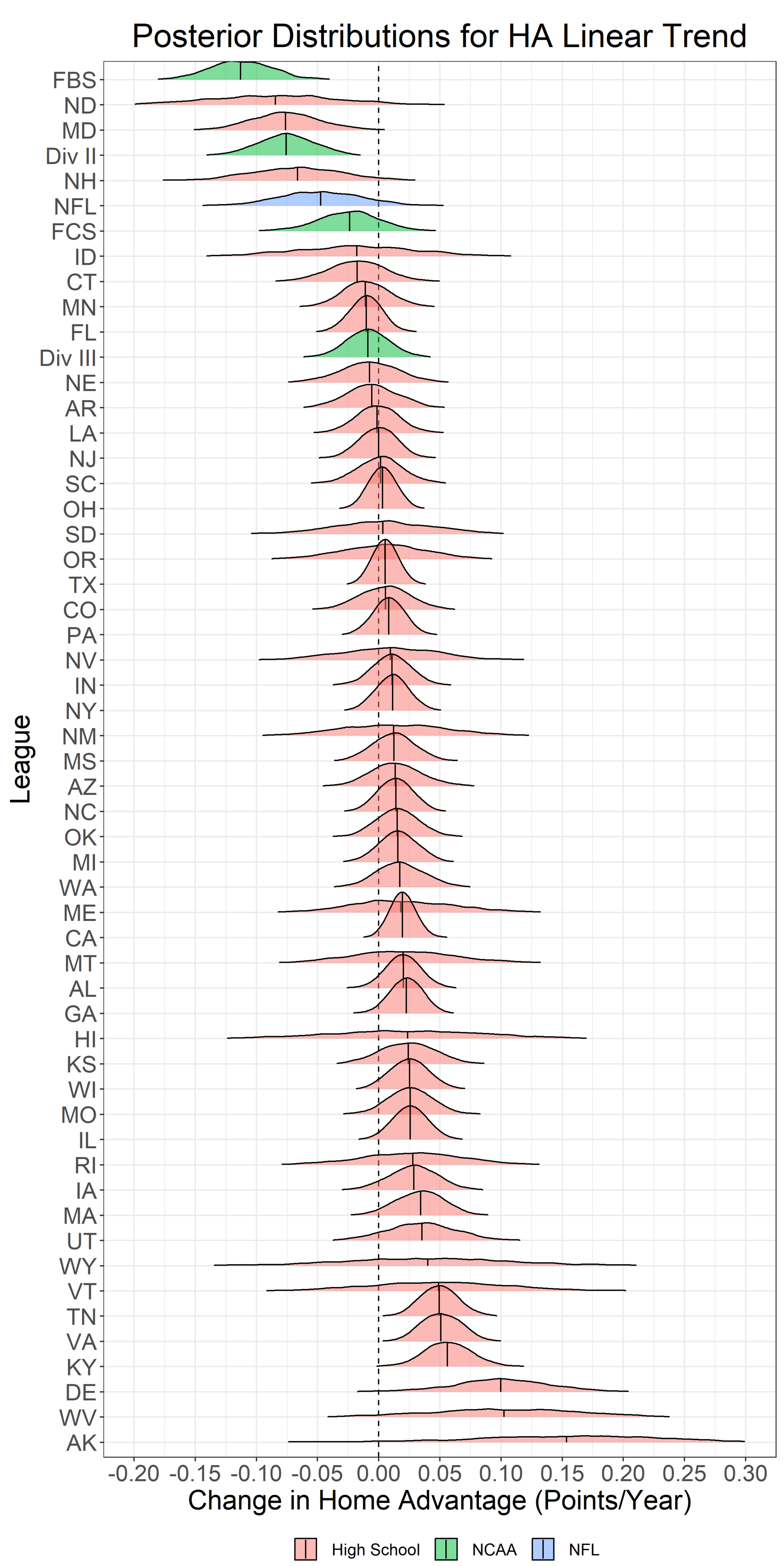

Figure 1: Posterior distributions for α_1, the slope of the linear trend for HA in Model (2), which denotes the change in home advantage in points/year. Negative values of α_1 denote a decline in HA while positive values of α_1 denote an increase in HA.

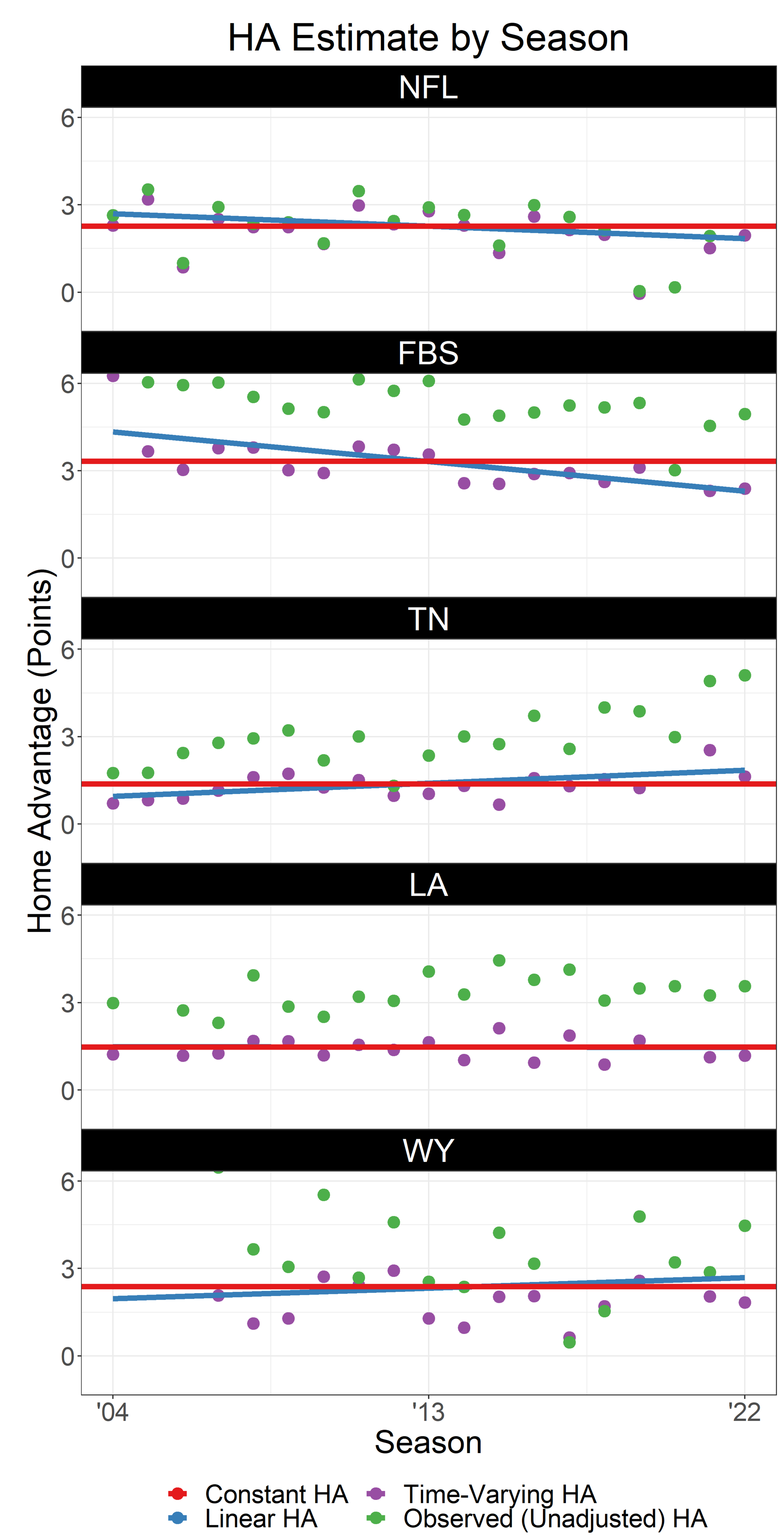

Figure 2: Comparison of HA posterior mean estimates for select state/league between 2004-2022 from each of the 3 models. Also included on the plot is the mean (HOME - AWAY) score differential for each league-season. In the vast majority of leagues, this "observed" home advantage is much larger than than any model based estimate, suggesting failure to account for team strength may lead to incorrect conclusions.

Conclusions

Across 50 high school states, 4 NCAA divisions, and the NFL, a constant home advantage trend was preferred in the majority of states. 95% posterior credible intervals for α_1, the linear home advantage trend in Model (2), do not contain 0 in only 7/55. Similar results are obtained when formally comparing models on the basis of expected log pointwise predictive density (ELPD) estimated via the leave-one-out cross-validation (LOO) approach of Gelman and colleagues [1].

Interestingly, FBS exhibits the largest estimated decline in HA, with a drop of roughly 0.11 points/year between 2004-2022. If anything, HA has declined more in NCAA/NFL over the past 2 decades than in high school, where changes in HA are far more heterogeneous.

An intuitive but perhaps underappreciated point is the importance of properly adjusting for team strength when estimating home advantage. In professional leagues such as the NFL, better teams host playoff games, and so even with more balanced schedules, one would expect model based estimates of HA to be attenuated compared to empirical observations. In lower levels of football, including college and high school, better teams are both more likely to host more home games (perhaps due to better facilities and larger athletic budgets) and win by larger amounts. As evidenced in Figure 2, failure to properly account for confounding by team strength would lead one to not only produce estimates of HA that are severely biased, but also overstate the extent to which strong temporal HA trends exist (e.g. Tennessee).

References

[1] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistical Computing*, 27(6):1413–1432, 2017.