

How often does the best team win? A unified approach to  
understanding randomness in North American sport

Michael J. Lopez

Skidmore College

[mlopez1@skidmore.edu](mailto:mlopez1@skidmore.edu)

Gregory J. Matthews

Loyola University Chicago

[gmatthews1@luc.edu](mailto:gmatthews1@luc.edu)

Benjamin S. Baumer

Smith College

[bbaumer@smith.edu](mailto:bbaumer@smith.edu)

January 24, 2017

## **Abstract**

Statistical applications in sports have long centered on how to best separate signal, such as team talent, from random noise. However, most of this work has concentrated on a single sport, and the development of meaningful cross-sport comparisons has been impeded by the difficulty of translating luck from one sport to another. In this manuscript, we use betting market data to develop a Bayesian state-space model that can be uniformly applied across sporting leagues to better understand the role of randomness in game outcomes. Our model can be used to extract estimates of team strength, the between-season, within-season, and game-to-game variability of team strengths, as well each team's home advantage. We implement our approach across a decade of play in each of the National Football League (NFL), National Hockey League (NHL), National Basketball Association (NBA), and Major League Baseball (MLB), finding that the NBA demonstrates both the largest dispersion in talent and the largest home advantage. Additionally, the NHL and MLB stand out for their relative randomness in game outcomes. We conclude by proposing a new metric for judging league competitiveness that works in absence of factors outside of team control. Keywords: sports analytics, Bayesian modeling, competitive balance, MCMC

# 1 Introduction

Most observers of sport can agree that game outcomes are to some extent subject to chance. The line drive that miraculously finds the fielder’s glove, the fumble that bounces harmlessly out-of-bounds, the puck that ricochets into the net off of an opponent’s skate, or the referee’s whistle on a clean block can all mean the difference between winning and losing. Yet game outcomes are not *completely* random—there are teams that consistently play better and worse. To what extent does luck influence our perceptions of team strength over time?

One way in which statistics can lead this discussion lies in the untangling of signal and noise when comparing the caliber of each league’s teams. For example, is team  $i$  better than team  $j$ ? And if so, how confident are we in making this claim? Central to such an understanding of sporting outcomes is that if we know each team’s relative strength, then, *a priori*, game outcomes—including wins and losses—can be viewed as unobserved realizations of random variables. As a simple example, a 75% probability of team  $i$  beating team  $j$  at time  $k$  implies that in a hypothetical infinite number of games between the two teams at time  $k$ ,  $i$  wins three times as often as  $j$ .

Given both national public interest and an academic curiosity that has extended across disciplines, many innovative techniques have been developed to estimate team strength. These approaches typically blend past game scores with game, team, and player characteristics in a statistical model. Corresponding estimates of talent are often checked or calibrated by comparing out-of-sample estimated probabilities of wins and losses to observed outcomes. Such exercises do more than drive water-cooler conversation as to which team may be better. Indeed, estimating team rankings has driven the development of advanced statistical models (Bradley and Terry, 1952; Glickman and Stern, 1998) and occasionally played a role in the decision of which teams are eligible for continued postseason play (CFP, 2014).

However, because randomness manifests differently in different sports, a limitation of sport-specific models is that inferences cannot generally be applied to other competitions. As a result, researchers who hope to contrast one league to another often focus on the one outcome common to all sports: won-loss ratio. Among other flaws, measuring team strength using wins and losses performs poorly in a small sample size, ignores the game’s final score (which is known to be more predictive of future performance than won-loss ratio (Boulier and Stekler, 2003)), and is unduly impacted by, among other sources, fluctuations in league scheduling, season length, injury to key players, and the general

advantage of playing at home. As a result, until now, analysts and fans have never quite been able to quantify inherent differences between sports with respect to randomness and the dispersion and evolution of team strength. We aim to fill this void.

In the sections that follow, we present a novel approach for estimating team ability that accounts for league scheduling, measures each team’s home advantage, and uncovers inherent differences in North American sport. First, we validate an assumption that game-level probabilities provided by betting markets provide unbiased and low-variance estimates of the true probabilities of wins and losses in each professional contest. Second, we implement a modified Bayesian state-space model that uses these probabilities to capture implied team strength and variability. Next, by examining posterior estimates of within and between season variability, as well as the overall dispersion in team strength estimates, we present unique league-level contrasts that to this point have been difficult to capture. Finally, we conclude by showing that our estimates of team strength improve upon both won-loss ratio and point differential with respect to correlating to future performance, and use our posterior draws propose a novel metric of assessing parity. We find that, on account of both narrower talent distributions and smaller home advantages, a typical contest in the NHL or MLB is much closer to a coin-flip than one in the NBA or NFL. Additionally, the NHL (from one season to the next) and NBA (from one week to the next) boast the least consistency in team strength estimates over time.

## 1.1 Literature review

The importance of quantifying team strength in sport extends across disciplines. This includes contrasting league-level characteristics in economics (Leeds and Von Allmen, 2004), estimating game-level probabilities in statistics (Glickman and Stern, 1998), and classifying future game winners in forecasting (Boulier and Stekler, 2003). We discuss and synthesize below.

### 1.1.1 Competitive balance

Assessing the competitive balance of sports leagues is particularly important in economics and management (Leeds and Von Allmen, 2004). While competitive balance can purportedly measure several different quantities, in general it refers to levels of equivalence between teams. This could be equivalence within one time frame (e.g. how similar was the distribution of talent within a season?), between time frames (e.g. year-to-year variations in talent), or from the beginning of a time frame

until the end (e.g. the likelihood of each team winning a championship at the start of a season).

The most widely accepted within-season competitive balance measure is the Noll-Scully (Noll, 1991; Scully, 1989), computed as the ratio of the observed standard deviation in team win totals to the idealized standard deviation, defined as that which would have been observed due to chance alone if each team were equal in talent. Larger Noll-Scully values are believed to reflect greater imbalance in team strengths.

While Noll-Scully has the positive quality of allowing for interpretable cross-sport comparisons, a reliance on won-loss outcomes entails undesirable properties as well (Owen, 2010; Owen and King, 2015). For example, Noll-Scully increases, on average, with the number of games played (Owen and King, 2015), hindering any comparisons of the NFL (16 games) to MLB (162). Additionally, each of the leagues employ some form of an unbalanced schedule. Teams in each of MLB, the NBA, NFL, and NHL play intradivisional opponents more often than interdivisional ones, and intraconference opponents more often than interconference ones, meaning that one team's won-loss record may not be comparable to another team's due to differences in the respective strengths of their opponents (Lenten, 2015). Moreover, the NFL structures each season's schedule so that teams play interdivisional games against opponents that finished in the same spot in the standings in the prior year. In expectation, this punishes teams that finish atop standings with tougher games, potentially driving winning percentages toward 0.500. Unsurprisingly, unbalanced scheduling and interconference play can lead to imprecise competitive balance metrics derived from winning percentages (Utt and Fort, 2002). As one final weakness, varying home advantages between sports leagues, as shown in Moskowitz and Wertheim (2011), could also impact comparisons of relative team quality that are predicated on wins and losses.

Although metrics for league-level comparisons have been frequently debated, the importance of competitive balance in sports is more uniformly accepted, in large part due to the uncertainty of outcome hypothesis (Rottenberg, 1956; Knowles et al., 1992; Lee and Fort, 2008). Under the uncertainty of outcome hypothesis, league success—as judged by attendance, engagement, and television revenue—correlates positively with teams having equal chances. Outcome uncertainty is generally considered on a game-level basis, but can also extend to season-level success (i.e, teams having equivalent chances at making the postseason). As a result, it is in each league's best interest to promote some level of *parity*—in short, a narrower distribution of team quality—to maximize revenue (Crooker and Fenn, 2007). Related, the Hirfindahl-Hirschman Index (Owen et al., 2007) and Competitive

Balance Ratio (Humphreys, 2002) are two metrics attempting to quantify the relative chances of success that teams have within or between certain time frames.

### 1.1.2 Approaches to estimating team strength

Competitive balance and outcome uncertainty are rough proxies for understanding the distribution of talent among teams. For example, when two teams of equal talent play a game without a home advantage, outcome uncertainty is maximized; e.g., the outcome of the game is equivalent to a coin flip. These relative comparisons of team talent began in statistics with paired comparison models, which are generally defined as those designed to calibrate the equivalence of two entities. In the case of sports, the entities are teams or individual athletes.

The Bradley-Terry model (BTM, Bradley and Terry (1952)) is considered to be the first detailed paired comparison model, and the rough equivalent of the soon thereafter developed Elo rankings (Elo, 1978; Glickman, 1995). Consider an experiment with  $t$  treatment levels, compared in pairs. BTM assumes that there is some true ordering of the probabilities of efficacy,  $\pi_1, \dots, \pi_t$ , with the constraints that  $\pi_i \geq 0$  and  $\sum \pi_i = 1$ . When comparing treatment  $i$  to treatment  $j$ , the probability that treatment  $i$  is preferable to  $j$  (i.e. a win in a sports setting) is computed as  $\frac{\pi_i}{\pi_i + \pi_j}$ .

Glickman and Stern (1998) and Glickman and Stern (2016) build on the BTM by allowing team-strength estimates to vary over time through the modeling of point differential in the NFL, which is assumed to follow an approximately normal distribution. Let  $y_{(s,k)ij}$  be the point differential of a game during week  $k$  of season  $s$  between teams  $i$  and  $j$ . In this specification,  $i$  and  $j$  take on values between 1 and  $t$ , where  $t$  is the number of teams in the league. Let  $\theta_{(s,k)i}$  and  $\theta_{(s,k)j}$  be the strengths of teams  $i$  and  $j$ , respectively, in season  $s$  during week  $k$ , and let  $\alpha_i$  be the home advantage parameter for team  $i$ , for  $i = 1, \dots, t$ . Glickman and Stern (1998) assume that for a game played at the home of team  $i$  during week  $k$  in season  $s$ ,

$$E[y_{(s,k)ij} | \theta_{(s,k)i}, \theta_{(s,k)j}, \alpha_i] = \theta_{(s,k)i} - \theta_{(s,k)j} + \alpha_i,$$

where  $E[y_{(s,k)ij} | \theta_{(s,k)i}, \theta_{(s,k)j}, \alpha_i]$  is the expected point differential given  $i$  and  $j$ 's team strengths and the home advantage of team  $i$ .

The model of Glickman and Stern (1998) allows for team strength parameters to vary stochastically

in two distinct ways: from the last week of season  $s$  to the first week of season  $s + 1$ , and from week  $k$  of season  $s$  to week  $k + 1$  of season  $s$ . As such, it is termed a ‘state-space’ model, whereby the data is a function of an underlying time-varying process plus additional noise.

Glickman and Stern (1998) propose an autoregressive process to team strengths, whereby over time, these parameters are pulled toward the league average. One attractive property of this specification is that past and future season performances are incorporated into season-specific estimates of team quality. Perhaps as a result, Koopmeiners (2012) identifies stronger fits when comparing state-space models to BTM’s fit separately within each season. Additionally, unlike BTM’s, state-space models would not typically suffer from identifiability problems were a team to win or lose all of its games in a single season (a rare, but extant possibility in the NFL).<sup>1</sup> For additional and related state-space resources, see Knorr-Held (2000), Cattelan et al. (2013), Baker and McHale (2015), and Manner (2015). Additionally, Matthews (2005), Owen (2011), Koopmeiners (2012), Tutz and Schauberger (2015), and Wolfson and Koopmeiners (2015) implement related versions of the original BTM.

Although the state-space model summarized above appears to work well in the NFL, a few issues arise when extending it to other leagues. First, with point differential as a game-level outcome, parameter estimates would be sensitive to the relative amount of scoring in each sport. Thus, comparisons of the NHL and MLB (where games, on average, are decided by a few goals or runs) to the NBA and NFL (where games, on average, are decided by about 10 points) would require further scaling. Second, it is unclear if a Normal model of goal or run differential is appropriate in low scoring sports like the NHL and MLB. Finally, NHL game outcomes would entail an extra complication, as roughly 25% of regular season games are decided in overtime or a shootout.

In place of paired comparison models, alternative measures for estimating team strength have also been developed. Massey (1997) used maximum likelihood estimation and American football outcomes to develop an eponymous rating system. A more general summary of other rating systems for forecasting use is explored by Boulier and Stekler (2003). In addition, support vector machines and simulation models have been proposed in hockey (Demers, 2015; Buttrey, 2016), neural networks and naïve Bayes implemented in basketball (Loeffelholz et al., 2009; Miljković et al., 2010), linear models and probit regressions in football (Harville, 1980; Boulier and Stekler, 2003), and two stage Bayesian models in baseball (Yang and Swartz, 2004). While this is a non-exhaustive list, it speaks to the depth and variety of coverage that sports prediction models have generated.

---

<sup>1</sup>In the NFL, the 2007 New England Patriots won all of their regular season games, while the 2008 Detroit Lions lost all of their regular season games.

## 1.2 Betting market probabilities

In many instances, researchers derive estimates of team strength in order to predict game-level probabilities. Betting market information has long been recommended to judge the accuracy of these probabilities (Harville, 1980; Stern, 1991). Before each contest, sports books—including those in Las Vegas and in overseas markets—provide a price for each team, more commonly known as the money line.

Mathematically, if team  $i$ 's money line is  $\ell_i$  against team  $j$  (with corresponding money line  $\ell_j$ ), where  $|\ell_i| \geq 100$ , then the boundary win probability for that team,  $p_i(\ell_i)$ , is given by:

$$p_i(\ell_i) = \begin{cases} \frac{100}{100 + \ell_i} & \text{if } \ell_i \geq 100 \\ \frac{|\ell_i|}{100 + |\ell_i|} & \text{if } \ell_i \leq -100 \end{cases}.$$

The boundary win probability represents the threshold at which point betting on team  $i$  would be profitable in the long run.

As an example, suppose the Chicago Cubs were favored ( $\ell_i = -127$  on the money line) to beat the Arizona Diamondbacks ( $\ell_j = 117$ ). The boundary win probability for the Cubs would be  $p_i(-127) = 0.559$ ; for the Diamondbacks,  $p_j(117) = 0.461$ . Boundary win probabilities sum to greater than one by an amount collected by the sportsbook as profit (known colloquially as the “vig” or “vigorish”). However, it is straightforward to normalize boundary probabilities to sum to unity to estimate  $p_{ij}$ , the implied probability of  $i$  defeating  $j$ :

$$p_{ij} = \frac{p_i(\ell_i)}{p_i(\ell_i) + p_j(\ell_j)}. \quad (1)$$

In our example, dividing each boundary probability by  $1.02 = (0.559 + 0.461)$  implies win probabilities of 54.8% for the Cubs and 45.2% for the Diamondbacks.

In principle, money line prices account for all determinants of game outcomes known to the public prior to the game, including team strength, location, and injuries. Across time and sporting leagues, researchers have identified that it is difficult to estimate win probabilities that are more accurate than the market; i.e., that the betting markets are efficient. As an incomplete list, see Harville (1980); Gandar et al. (1988); Lacey (1990); Stern (1991); Carlin (1996); Colquitt et al. (2001); Spann and

Skiera (2009); Nichols (2012); Paul and Weinbach (2014); Lopez and Matthews (2015). Interestingly, Colquitt et al. (2001) suggested that the efficiency of college basketball markets was proportional to the amount of pre-game information available—with the amount known about professional sports teams, this would suggest that markets in the NFL, NBA, NHL and MLB are as efficient as they come. Manner (2015) merged predictions from a state-space model with those from betting markets, finding that the combination of both predictions only occasionally outperformed betting markets alone.

We are not aware of any published findings that have compared leagues using market probabilities. Given the varying within-sport metrics of judging team quality and the limited between-sport approaches that rely on wins and losses alone, we aim to extend paired comparison models using money line information to better capture relative team equivalence in a method that can be applied generally.

## 2 Validation of betting market data

We begin by confirming the accuracy of betting market data with respect to game outcomes. Regular season game result and betting line data in the four major North American professional sports leagues (MLB, NBA, NFL, and NHL) were obtained for a nominal fee from Sports Insights. Although these game results are not official, they are accurate and widely-used. The 2006–2016 seasons were included in our models, except for the NFL, which used only the 2006–2015 seasons.

These data were more than 99.3% complete in each league, in the sense that there existed a valid betting line for nearly all games in these four sports across this time period. Betting lines provided by Sports Insights are expressed as payouts, which we subsequently convert into implied probabilities. The average vig in our data set is 1.93%, but is always positive, resulting in revenue for the sportsbook over a long run of games. In circumstances where more than one betting line was available for a particular game, we included only the line closest to the start time of the game. A summary of our data is shown in Table 1.

We also compared the observed probabilities of a home win to the corresponding probabilities implied by our betting market data (Figure 1). In each of the four sports, the efficient market hypothesis cannot be rejected for any range of implied home win probabilities, based on visual inspection of

Sport ( $q$ )	$t_q$	$n_{games}$	$\bar{p}_{games}$	$n_{bets}$	$\bar{p}_{bets}$	Coverage
MLB	30	26728	0.541	26710	0.548	0.999
NBA	30	13290	0.595	13245	0.615	0.997
NFL	32	2560	0.563	2542	0.589	0.993
NHL	30	13020	0.548	12990	0.565	0.998

Table 1: Summary of cross-sport data.  $t_q$  is the number of unique teams in each sport  $q$ .  $n_{games}$  records the number of actual games played, while  $n_{bets}$  records the number of those games for which we have a betting line.  $\bar{p}_{games}$  is the mean observed probability of a win for the home team, while  $\bar{p}_{bets}$  is the mean implied probability of a home win based on the betting line. Note that we have near total coverage (betting odds for almost every game) across all four major sports.

a LOESS regression model. Thus, we find no evidence to suggest that the probabilities implied by our betting market data are biased or inaccurate—a conclusion that is supported by the body of academic literature referenced above. Accordingly, we interpret these probabilities as “true.”

### 3 Bayesian state-space model

Our model below expands the state-space specification provided by Glickman and Stern (1998) to provide a unified framework for contrasting the four major North American sports leagues.

Let  $p_{(q,s,k)ij}$  be the probability that team  $i$  will beat team  $j$  in season  $s$  during week  $k$  of sports league  $q$ , for  $q \in \{MLB, NBA, NFL, NHL\}$ . The  $p_{(q,s,k)ij}$ 's are assumed to be known, calculated using sportsbook odds via Equation (1). In using game probabilities, we have a cross-sport outcome that provides more information than only knowing which team won the game or what the score was.

In our notation,  $i, j = 1, \dots, t_q$ , where  $t_q$  is the number of teams such that  $t_{MLB} = t_{NBA} = t_{NHL} = 30$  and  $t_{NFL} = 32$ . Additionally,  $s = 1, \dots, S_q$  and  $k = 1, \dots, K_q$ , where  $S_q$  and  $K_q$  are the number of seasons and weeks, respectively in league  $q$ . In our data,  $K_{NFL} = 17$ ,  $K_{NBA} = 25$ ,  $K_{MLB} = K_{NHL} = 28$ , with  $S_{NFL} = 10$  and  $S_{MLB} = S_{NBA} = S_{NHL} = 11$ .

Our next step in building a model specifies the home advantage, and one immediate hurdle is that in addition to having different numbers of teams in each league, certain franchises may relocate from one city to another over time. In our data set, there were two relocations, Seattle to Oklahoma City (NBA, 2008) and Atlanta to Winnipeg (NHL, 2011). Let  $\alpha_{q_0}$  be the league-wide home advantage (HA) in  $q$ , and let  $\alpha_{(q)i^*}$  be the extra effect (positive or negative) for team  $i$  among game's played in city  $i^*$ , for  $i^* = 1, \dots, t_q^*$ . Here,  $t_q^*$  is the total number of home cities; in our data,  $t_{MLB}^* = 30$ ,  $t_{NBA}^* = t_{NHL}^* = 31$ , and  $t_{NFL}^* = 32$ .

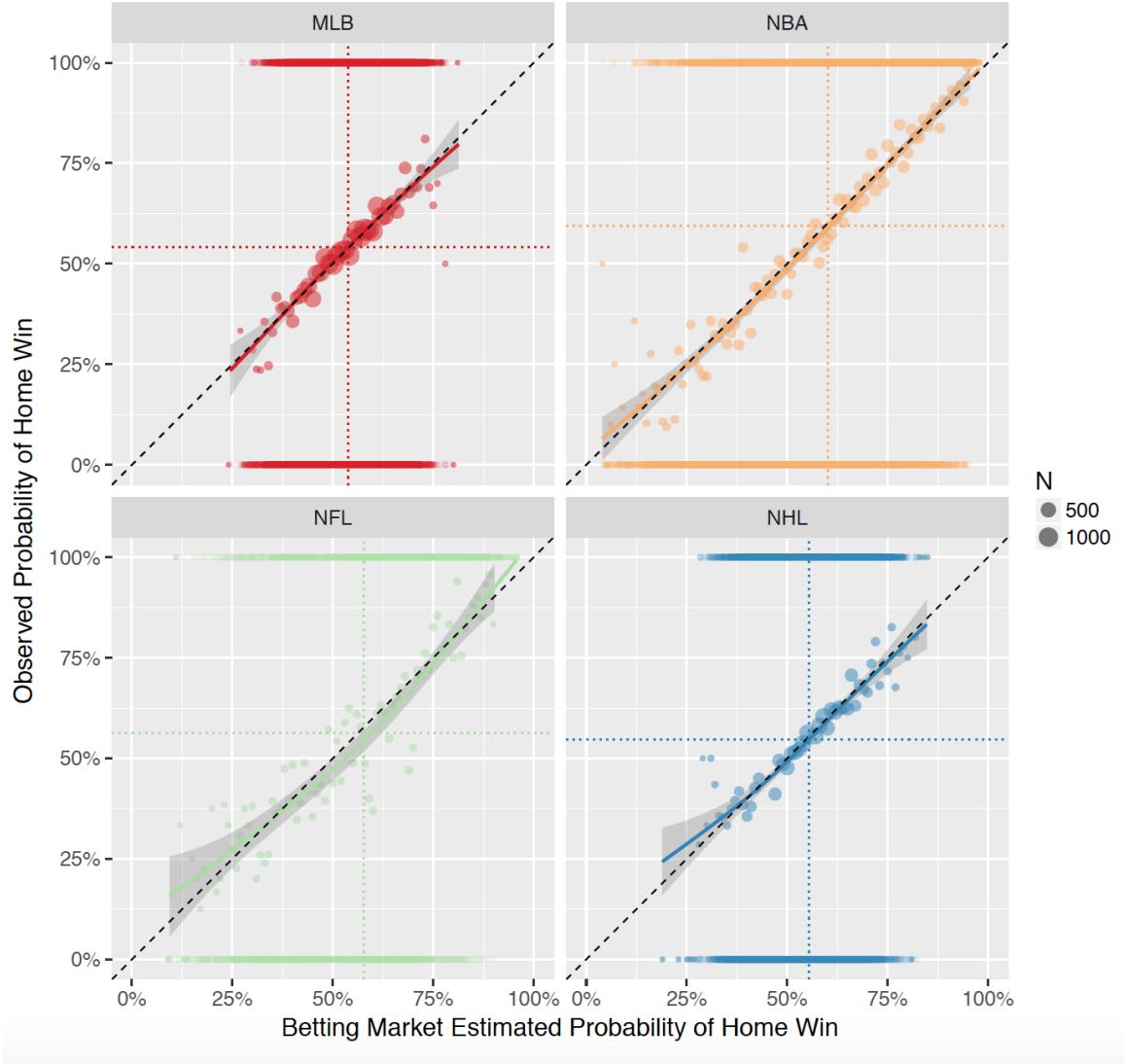


Figure 1: Accuracy of probabilities implied by betting markets. Each dot represents a bin of implied probabilities rounded to the nearest hundredth. The size of each dot ( $N$ ) is proportional to the number of games that lie in that bin. We note that across all four major sports, the observed winning percentages accord with those implied by the betting markets. The dotted diagonal line indicates a completely fair market where probabilities from the betting markets correspond exactly to observed outcomes. In each sport, this diagonal line lies entirely within the standard error surrounding a LOESS regression line, suggesting that an efficient market hypothesis cannot be rejected.

Letting  $\theta_{(q,s,k)i}$  and  $\theta_{(q,s,k)j}$  be season-week team strength parameters for teams  $i$  and  $j$ , respectively, we assume that

$$E[\text{logit}(p_{(q,s,k)ij}) | \theta_{(q,s,k)i}, \theta_{(q,s,k)j}, \alpha_{q_0}, \alpha_{(q)i^*}] = \theta_{(q,s,k)i} - \theta_{(q,s,k)j} + \alpha_{q_0} + \alpha_{(q)i^*},$$

where  $\text{logit}(.)$  is the log-odds transform.

Let  $\mathbf{p}_{(q,s,k)}$  represent the vector of length  $g_{(q,s,k)}$  containing all of league  $q$ 's probabilities in week  $k$  of season  $s$ . Our first model of game outcomes, henceforth referred to as the individual home advantage model (Model IHA), assumes that

$$\text{logit}(\mathbf{p}_{(q,s,k)}) \sim N(\theta_{(q,s,k)} \mathbf{X}_{(q,s,k)} + \alpha_{q_0} \mathbf{J}_{g_{(q,s,k)}} + \boldsymbol{\alpha}_q \mathbf{Z}_{(q,s,k)}, \sigma_{q,\text{game}}^2 \mathbf{I}_{g_{(q,s,k)}}),$$

where  $\theta_{(q,s,k)}$  is a vector of length  $t_q$  containing the team strength parameters in season  $s$  during week  $k$  and  $\boldsymbol{\alpha}_q = \{\alpha_{(q)1}, \dots, \alpha_{(q)t_q^*}\}$ . Note that  $\boldsymbol{\alpha}_q$  does not vary over time (i.e. HA is assumed to be constant for a team over weeks and seasons).  $\mathbf{X}_{(q,s,k)}$  and  $\mathbf{Z}_{(q,s,k)}$  contain  $g_{(q,s,k)}$ —the number of games in league  $q$  during week  $k$  of season  $s$ —rows and  $t_q$  and  $t_q^*$  columns, respectively. The matrix  $\mathbf{X}_{(q,s,k)}$  contains the values  $\{1, 0, -1\}$  where for a given row (i.e. one game) the value of  $i^{th}$  column in that row is a 1/-1 if the  $i^{th}$  team played at home/away in the given game and 0 otherwise.  $\mathbf{Z}_{(q,s,k)}$  is a matrix containing a 1 in column  $i^*$  if the corresponding game was played in city  $i^*$ , and 0 otherwise. Finally,  $\sigma_{q,\text{game}}^2$  is the game-level variance,  $\mathbf{J}_{g_{(q,s,k)}}$  is a column vector of length  $g_{(q,s,k)}$  containing all 1's, and  $\mathbf{I}_{g_{(q,s,k)}}$  is an identity matrix with dimension  $g_{(q,s,k)} \times g_{(q,s,k)}$ .

In addition, we propose a simplified version of Model IHA, labelled as Model CHA (constant home advantage), which assumes that the HA within each sport is identical for each franchise, such that

$$\text{logit}(\mathbf{p}_{(q,s,k)}) \sim N(\theta_{(q,s,k)} \mathbf{X}_{(q,s,k)} + \alpha_{q_0} \mathbf{J}_{g_{(q,s,k)}}, \sigma_{q,\text{game}}^2 \mathbf{I}_{g_{(q,s,k)}}).$$

In Model CHA, matrices  $\mathbf{p}_{(q,s,k)}$ ,  $\mathbf{X}_{(q,s,k)}$ ,  $\mathbf{J}_{g_{(q,s,k)}}$ , and  $\mathbf{I}_{g_{(q,s,k)}}$  are specified identically to Model IHA. As a result, for a game between home team  $i$  and away team  $j$  during week  $k$  of season  $s$ ,  $E[\text{logit}(p_{(q,s,k)ij})] = \theta_{(q,s,k)i} - \theta_{(q,s,k)j} + \alpha_{q_0}$  under Model CHA.

Similar to Glickman and Stern (1998), we allow the strength parameters of the teams to vary autoregressively from season-to-season and from week-to-week. In general, this entails that team strength parameters are shrunk towards the league average over time in expectation. Formally,

$$\theta_{(q,s+1,1)} | \theta_{q,s,K_q}, \gamma_{q,season}, \sigma_{q,season}^2 \sim N(\gamma_{q,season} \theta_{(q,s,K_q)}, \sigma_{q,season}^2 \mathbf{I}_{t_q}) \text{ for all } s \in 2, \dots, S_q,$$

and

$$\theta_{(q,s,k+1)} | \theta_{(q,s,k)}, \gamma_{q,week}, \sigma_{q,week}^2 \sim N(\gamma_{q,week} \theta_{(q,s,k)}, \sigma_{q,week}^2 \mathbf{I}_{t_q}) \text{ for all } s \in 1, \dots, S_q, k \in 2, \dots, K_q.$$

In this specification,  $\gamma_{q,week}$  is the autoregressive parameter from week-to-week,  $\gamma_{q,season}$  is the autoregressive parameter from season-to-season, and  $\mathbf{I}_{t_q}$  is the identity matrix of dimension  $t_q \times t_q$ .

Given the time-varying nature of our specification, we use a Bayesian approach to obtain model estimates. For sport  $q$ , the team strength parameters for week  $k = 1$  and season  $s = 1$  have a prior distribution of

$$\theta_{(q,1,1)i} \sim N(0, \sigma_{q,season}^2), \quad \text{for all } i \in 1, \dots, t_q.$$

Team specific home advantage parameters have a similar prior, namely,

$$\alpha_{(q)i^\star} \sim N(0, \sigma_{q,\alpha}^2), \quad \text{for } i \in 1, \dots, t_q^\star.$$

Finally, letting  $\tau_{q,game}^2 = 1/\sigma_{q,game}^2$ ,  $\tau_{q,season}^2 = 1/\sigma_{q,season}^2$ ,  $\tau_{q,week}^2 = 1/\sigma_{q,week}^2$ , and  $\tau_{q,\alpha}^2 = 1/\sigma_{q,\alpha}^2$ , we assume the following prior distributions:

$$\begin{aligned} \tau_{q,game}^2 &\sim \Gamma(0.0001, 0.0001) & \alpha_{q_0} &\sim N(0, 10000) \\ \tau_{q,season}^2 &\sim \Gamma(0.0001, 0.0001) & \gamma_{q,season} &\sim Uniform(0, 2) \\ \tau_{q,week}^2 &\sim \Gamma(0.0001, 0.0001) & \gamma_{q,week} &\sim Uniform(0, 2) \\ \tau_{q,\alpha}^2 &\sim \Gamma(0.0001, 0.0001) \end{aligned}$$

Our primary interest lies in three levels of variability with respect to the  $\theta_{(q,s,k)}$ 's. First, there is

variability at any fixed time  $s$  and  $k$  across  $i$ . This reflects the between-team variability in team strength; in other words, how equivalent are the teams to one another at a given snapshot in time? Second, there is variability across  $k$ , reflected in the week-to-week autoregressive parameter,  $\gamma_{q,week}$ . This generalizes to how teams can improve or worsen over the course of a season. Third, there is variability across  $s$ , corresponding to the season-to-season autoregressive parameter,  $\gamma_{q,season}$ . This accounts for larger changes to team ability that can occur between seasons.

Posterior distributions of each parameter are estimated using Markov Chain Monte Carlo (MCMC) methods. We used Gibbs sampling via the `rjags` package (Plummer, 2016) in the `R` statistical computing environment to obtain posterior distributions, done separately for each  $q$ . Three chains—using 20,000 iterations after a burn-in of 2,000 draws, fit with a thin of 5 to reduce the autocorrelation within chains—yielded 4,000 posterior samples in each  $q$ . Visual inspection of trace plots with parallel chains are used to confirm convergence. Comparisons of Models IHA and CHA are made using the Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)).

While we are unable to share the exact betting market data due to licensing restrictions, a simplified version of our game-level data, the data wrangling code, Gibbs sampling code, posterior draws, and the code used to obtain posterior estimates and figures are all posted to a GitHub repository, available at <https://github.com/bigfour/competitiveness>.

## 4 Results

In this section we present our results. We begin by validating and comparing the fits of Models IHA and CHA. We discuss the implications of our estimates of team strength and home advantage, as well as the interpretation of our variance and autoregressive parameters. We conclude by evaluating our team strength parameters and illustrating how they could be used to build a league parity metric.

### 4.1 Model fit

We identified no concerns with the fit of Models IHA and CHA. Trace plots of  $\alpha_{q_0}$ ,  $\gamma_{q,season}$ ,  $\gamma_{q,week}$ ,  $\sigma_{q,game}$ ,  $\sigma_{q,season}$ , and  $\sigma_{q,week}$  are shown for each  $q$  in Figures 8–11 in the Appendix. Visual inspection of these plots does not provide evidence of a lack of convergence or of autocorrelation between draws. These trace plots stem from Model IHA; conclusions are similar when plotting draws

from Model CHA.

Table 2 shows the deviance information criterion (DIC) for each fit in each league, along with the difference in DIC values and the associated standard error (SE). In each of the leagues, fits with a team-specific HA (Model IHA) yielded lower DIC’s (lower is better) by a statistically significant margin, with the most noticeable difference in fit improvement in the NBA.

	Model IHA	Model CHA	Difference (SE)
MLB	-8548	-8522	-25.7 (10.8)
NBA	6886	7224	-337.4 (26.6)
NFL	-1230	-1216	-13.6 (3.3)
NHL	-18335	-18148	-187.6 (20.7)

Table 2: Deviance information criterion (DIC) by sport and model, along with the difference in DIC and the associated standard errors (SE, in parentheses). IHA: individual home advantage, CHA: constant home advantage

These results suggest that chance alone likely does not account for observed differences in the home advantage among teams in each league, with the NBA showing the largest team-to-team differences in home advantage. As such, results that follow use model estimates from Model IHA.

## 4.2 Team strength

Table 3 shows summary statistics of the team strength estimates, approximated using posterior mean draws for all weeks  $k$  and seasons  $s$  across all four sports leagues. Overall, there tends to be a larger variability in team strength at any given point in time in both the NFL and NBA, with posterior coefficient estimates tending to vary between -1.3 and 1.2 in the NBA and -1.1 and 1.0 in the NFL (on the logit scale). For reference, a team-strength of 1.0 on the log-odds scale implies a  $\frac{e^{1.0}}{1+e^{1.0}} = 73.1\%$  chance of beating a league average team in a game played at a neutral site. The standard deviation of team strength is smallest in MLB, suggesting that—relative to the other leagues—team talent is more tightly packed. Relative to MLB, spread of team strengths are about 1.3, 3.0, and 3.6 times wider in the NHL, NFL, and NBA, respectively.

League ( $q$ )	N*	min	2.5 <sup>th</sup>	Q1	mean	Q3	97.5 <sup>th</sup>	max	sd
MLB	9240	-0.561	-0.384	-0.145	-0.012	0.116	0.324	0.470	0.183
NBA	8250	-2.177	-1.267	-0.485	0.001	0.478	1.209	1.864	0.660
NFL	5440	-1.570	-1.084	-0.390	0.010	0.427	1.034	1.909	0.558
NHL	9240	-1.035	-0.532	-0.169	-0.007	0.173	0.429	0.869	0.246

Table 3: Summary of average week-level team strength parameters, taken on the log-odds scale. N\*: number of unique team strength draws (teams  $\times$  seasons  $\times$  weeks)

Figure 2 shows estimated team strength coefficients over time. Figures 12–15 (shown in the Appendix) provide an individual plot for each sport, which include divisional facets to allow easier identification of individual teams. Teams in Figures 2 and 12–15 are depicted using their two primary colors, scraped from <http://jim-nielsen.com/teamcolors/> via the `teamcolors` package (<https://github.com/beanumber/teamcolors>) in R.

As in Table 3, these figures suggest that the NBA and NFL boast larger between-team gaps in talent than the NHL and MLB, implying more competitive balance in the latter pair of leagues. On one level, this stands somewhat in contrast to competitive balance as measured using Noll-Scully, which alternatively argues that the NFL is more competitively balanced than MLB (Berri, 2014). One likely explanation for this difference is Null-Scully’s link to number of games played, which artificially makes MLB (162 games) appear less balanced than it actually is and the NFL (16) appear more balanced. Like Noll-Scully, we conclude that the NBA does not show competitive balance relative to other leagues.

Our figures also illustrate several other observations. For example, the New England Patriots of the NFL stand out as having the top performance in the last decade, with an average team strength of 1.91 on the log-odds scale, observed during Week 11 of 2007. In that season, New England finished the regular season 16-0 before eventually losing in the Super Bowl. The worst performance belongs to the NBA’s Miami Heat, who during week 23 of the 2007–08 season had a posterior mean team strength of -2.18. That Heat team finished with an overall record of 15-67, at one point losing 15 consecutive games. Related, it is interesting that the team strength estimates of bad teams in the NBA (e.g. the Heat in 2007–08) lie further from 0 than the estimates for good teams. This possibly reveals the tendency for teams in this league to “tank”—a strategy of fielding a weak team intentionally to improve the chances of having better selection preference in the upcoming player draft (Soebbing and Humphreys, 2013).

Another observation is that in the NHL, top teams appear less dominant than a decade ago. For example, there are seven NHL team-seasons in which at least one team reached an average posterior strength estimate of 0.55 or greater; each of these came during or prior to the 2008–09 season. In addition to increased parity, the league’s point system change in 2005–06—which unintentionally encouraged teams to play more overtime games (Lopez, 2013)—could be responsible. More overtime contests could lead to different perceptions in how betting markets view team strengths, as overtime sessions and the resulting shootouts are roughly equivalent to coin flips (Lopez and Schuckers, 2016).

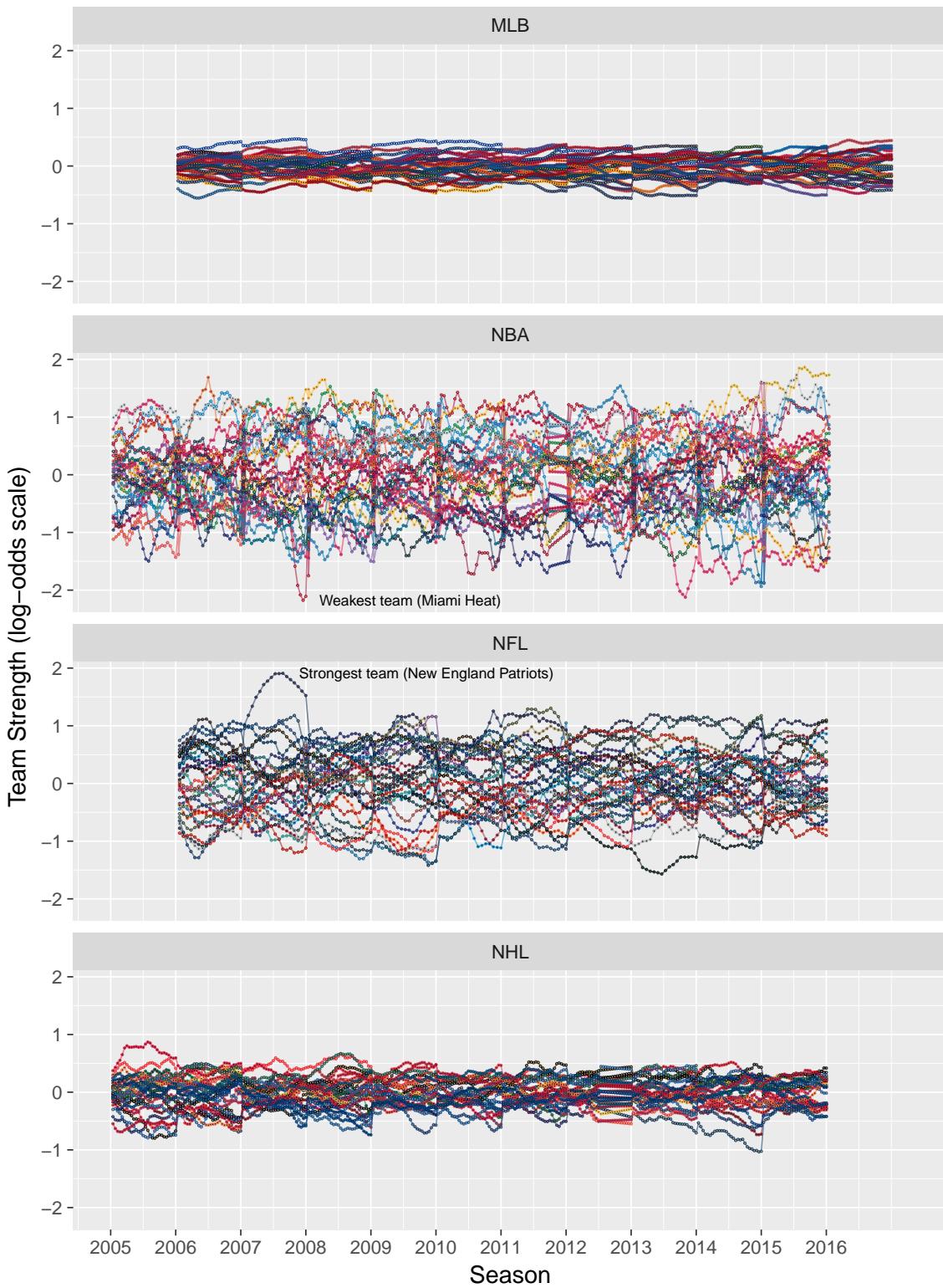


Figure 2: Mean team strength parameters over time for all four sports leagues. MLB and NFL seasons follow each yearly tick mark on the  $x$ -axis, while NBA and NHL seasons begin during years labeled by the preceding tick marks.

As a final point of clarification in Figures 2, 13, and 15, the periods of time with straight lines of team strength estimates during the 2012–13 season (NHL) and 2011–12 season (NBA) reflect time lost due to lockouts.

### 4.3 Variance and autoregressive parameters

Table 4 shows the mean and standard deviation of posterior draws for  $\gamma_{q,season}$ ,  $\gamma_{q,week}$ ,  $\sigma_{q,game}$ ,  $\sigma_{q,season}$ , and  $\sigma_{q,week}$  for each  $q$ .

League ( $q$ )	$\gamma_{q,season}$	$\gamma_{q,week}$	$\sigma_{q,game}$	$\sigma_{q,season}$	$\sigma_{q,week}$
MLB	0.621 (0.031)	1.002 (0.002)	0.201 (0.001)	0.094 (0.005)	0.027 (0.001)
NBA	0.618 (0.041)	0.977 (0.003)	0.274 (0.002)	0.443 (0.02)	0.166 (0.003)
NFL	0.688 (0.041)	0.978 (0.005)	0.232 (0.008)	0.333 (0.02)	0.148 (0.006)
NHL	0.542 (0.028)	0.993 (0.003)	0.105 (0.001)	0.122 (0.006)	0.053 (0.001)

Table 4: Mean posterior draw (standard deviation) by league.

Posterior draws of  $\sigma_{q,game}$  suggest that the highest game-level errors in our log-odds probability estimates occur in the NBA (median posterior draw of  $\sigma_{NBA,game} = 0.273$ ), followed in order by the NFL, MLB, and the NHL. Interestingly, although Figure 2 identifies that the talent gap between teams is smallest in MLB,  $\sigma_{MLB,game} \approx 2 \times \sigma_{NHL,game}$  in our posterior draws. We posit that this additional game-level error in MLB is a function of the league’s pitching match-ups, in which teams rotate through a handful of starting pitchers of varying calibers.

We also examine the joint distribution of the variability in team strength on a season-to-season ( $\sigma_{q,season}$ ) and week-to-week ( $\sigma_{q,week}$ ) basis via the contour plot in Figure 3, using separate colors for each  $q$ . Figure 3 reveals that the highest uncertainty with respect to team strength occurs in the NBA, followed in order by the NFL, NHL, and MLB.

There are a couple of plausible explanations regarding the increased uncertainty in NBA team strength on a weekly basis. Injuries, the resting of starters, and in-season trades would seemingly have a larger impact in a sport like basketball where fewer players are participating at a single point in time. In particular, our model cannot precisely gauge team strength when star players, who could play, are rested in favor of inferior players. Relative to the other professional leagues, star players take on a more important role in the NBA (Berri and Schmidt, 2006), an observation undoubtedly known in betting markets. That said, while there is increased variability in our estimate of NBA team strengths, when considering differences in team talent to begin with, these absolute differences

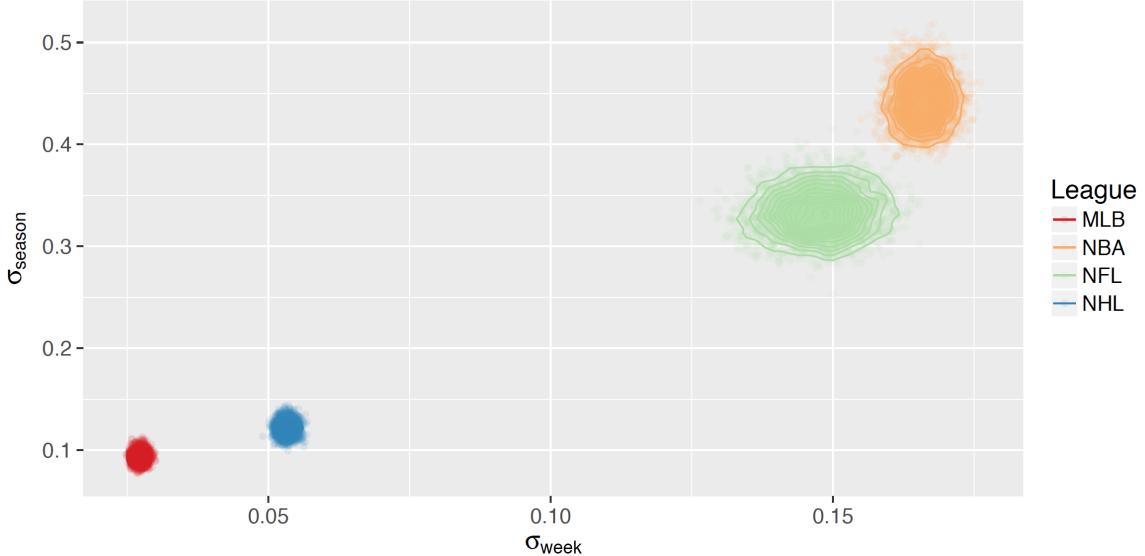


Figure 3: Contour plot of the estimated season-to-season and week-to-week variability across all four major sports leagues. By both measures, uncertainty is lowest in MLB and highest in the NBA.

are not as extreme (e.g., a difference in team strength of 0.05 means less in the NBA than in the NHL).

Similarly, Figure 4 displays the joint posterior distribution of  $\gamma_{q,\text{season}}$  and  $\gamma_{q,\text{week}}$  via contour plots for each  $q$ . On a season-to-season basis, team strengths in each of the leagues tend to revert towards the league average of zero as all draws of  $\gamma_{q,\text{season}} < 1$  for all  $q$ . Reversion towards the mean is largest in the NHL (estimated  $\gamma_{NHL,\text{season}} = 0.54$ , implying 46% reversion), followed by the NBA (38%), MLB (38% reversion), and the NFL (31%). However, the only pair of leagues with non-overlapping credible intervals are the NFL and NHL.

For each of the NHL, NBA, and NFL, posterior estimates of  $\gamma_{q,\text{week}}$  (as well as 95% credible intervals) imply an autoregressive nature to team strength within each season. Interestingly, the NBA and NFL are the least consistent leagues on a week-to-week basis. In MLB, however, team strength estimates quite possibly follow a random walk (i.e.,  $\gamma_{MLB,\text{week}} = 1$ ), in which the succession of team strength is unpredictable.

Finally, it is worth noting that our estimates for  $\gamma_{NFL,\text{week}}$  and  $\gamma_{NFL,\text{season}}$ —0.98 and 0.69, respectively—do not substantially diverge from the estimates observed by Glickman and Stern (1998) (0.99 and 0.82). Further, our estimates are more precise. For example, our 95% credible interval for  $\gamma_{NFL,\text{season}}$  of (0.61, 0.77) is entirely contained within the interval of (0.52, 1.28) reported by Glick-

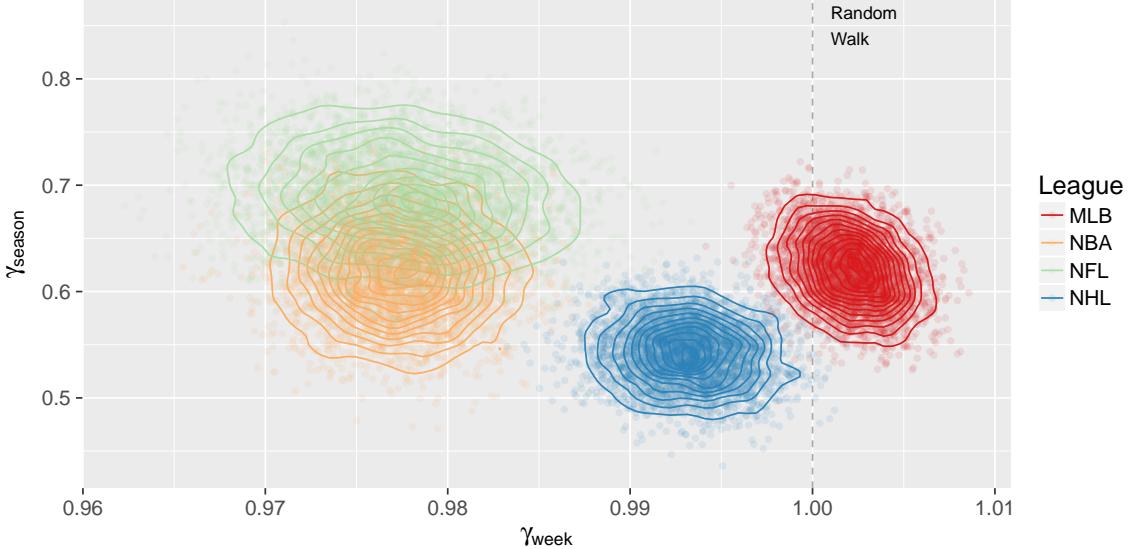


Figure 4: Contour plot of the estimated season-to-season and week-to-week autoregressive parameters across all four major sports leagues.

man and Stern (1998). In fairness, it is unclear if this increased precision is a function of our model specification (using log-odds of the probability of a win as the outcome, as opposed to point differential) or because we used a larger sample(10 seasons, compared to 5).

Like Glickman and Stern (1998), we also observe an inverse link in posterior draws of  $\gamma_{NFL,week}$  and  $\gamma_{NFL,season}$ . Given that total shrinkage across time is the composite of within- and between-season shrinkage, such an association is not surprising (Glickman and Stern, 1998). If one source of reversion towards the average were to increase, the other would likely compensate by decreasing.

#### 4.4 The home advantage

Figure 5 shows the 2.5th percentile, median, and 97.5th percentile draws of each team's estimated home advantage parameter, presented on the probability scale. These are calculated by summing draws of  $\alpha_{q_0}$  and  $\alpha_{(q)i^*}$  for all  $i^*$ . HAs are shown in descending order to provide a sense of the magnitude of differences between the home advantage provided in MLB (league-wide, a 54.0% probability of beating a team of equal strength at home), NHL (55.5%), NFL (58.8%), and NBA (62.0%). The two franchises that have relocated in the last decade, the Atlanta Thrashers (NHL) and Seattle Supersonics (NBA), are also included for the games played in those respective cities.

Figure 5 depicts substantial between-franchise differences within both the NBA and NHL. Con-

### Estimated Home Advantage by Franchise

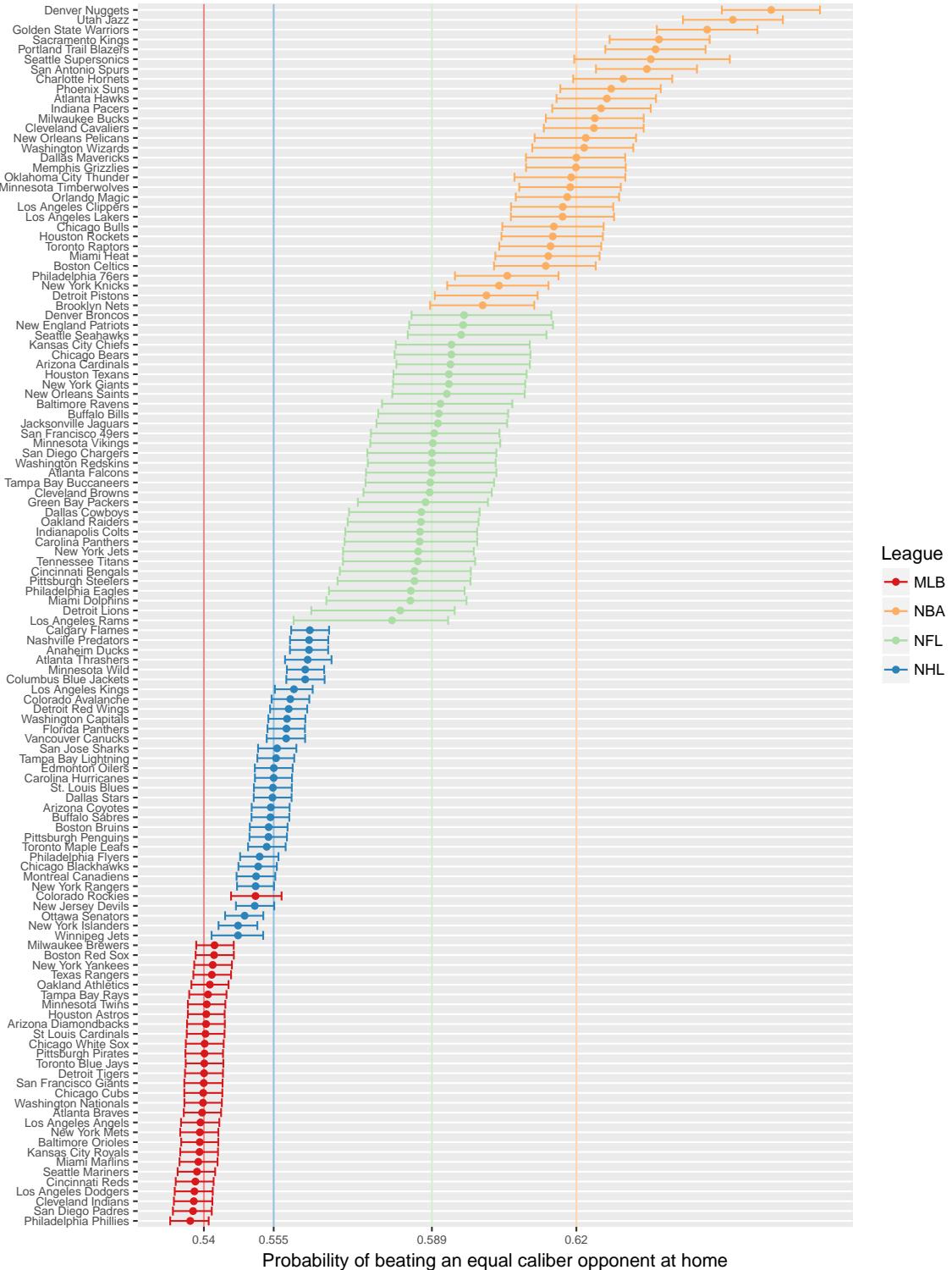


Figure 5: Median posterior draw (with 2.5th, 97.5th quantiles) of each franchise's home advantage intercept, on the probability scale. We note that the magnitude of home advantages are strongly segregated by sport, with only one exception (the Colorado Rockies). We also note that no NFL team, nor any MLB team other than the Rockies, has a home advantage whose 95% credible interval does not contain the league median.

versely, HA estimates within the NFL and MLB are, with the exception of the Colorado Rockies, indistinguishable across franchises. Interestingly, the draws of the home advantage parameters for a few NFL franchises are skewed (see Denver and Seattle, relative to Detroit), potentially the result of a shorter regular season. Alternatively, the NFL’s HA may vary by season, game time, or the day of the game. Anecdotally, night games (Thursday, Sunday, or Monday) conceivably offer a larger HA than those played during the day (Crabtree, 2014). Informally, NFL team-level HA estimates are similar in effect size to those depicted by Koopmeiners (2012).

In the NBA, Denver (first) and Utah (second) post the best home advantages, with Brooklyn showing the worst. This matches the results of Paine (2013), who found significantly better performances when comparing Denver and Utah to the rest of the league with respect to home and road point differential. In MLB, the Colorado Rockies stand out for having the highest home advantage, while the remaining 29 teams boast overlapping credible intervals. We note that teams playing at home in Denver have the largest home advantages in MLB, the NBA, and the NFL, and the 8th-highest in the NHL. We speculate that this consistent advantage across sports is related to the home team’s acclimation to the city’s notably high altitude.

These distinctions have plausible impacts on league standings. An NBA team with a typical home advantage can expect to win 62.0% of home games against a like-caliber opponent. Yet for Brooklyn, the corresponding figure is 59.9%, while for Denver, it is 66.1%. Across 41 games (the number each team plays at home), this implies that Denver’s home advantage is worth an extra 1.68 wins in a single season, relative to a league average team. Compared to Brooklyn, Denver’s home advantage is worth an estimated 2.54 wins per year. As one important caveat, our model estimates do not account for varying line-up and injury information. If opposing teams were to rest their star players at Denver, for example, our model would artificially inflate Denver’s home advantage.

## 4.5 Evaluation of team strength estimates

Ultimately, estimates from Model IHA are designed to estimate team quality at any given point in a season in the absence of factors such as the home advantage and opponent caliber. If these estimates more properly assess team quality than other metrics of team success (e.g., won-loss percentage or point differential), they should more accurately link to future performance, such as how well teams will perform over the remainder of the season.

### Coefficient of determination with future in-season win %

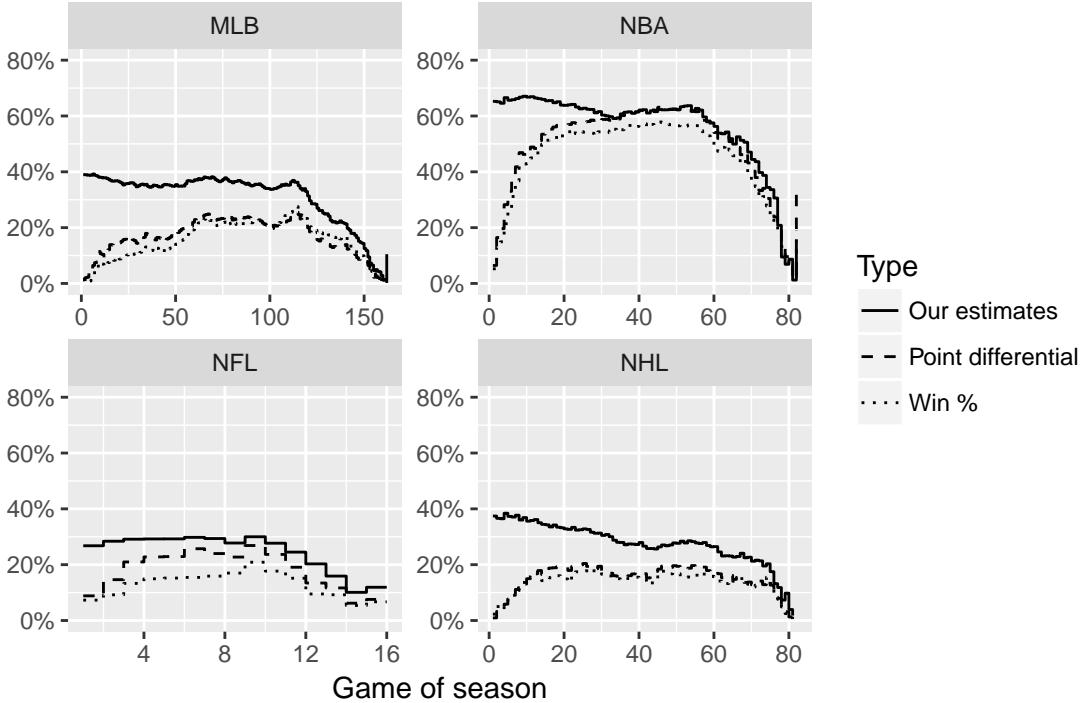


Figure 6: Coefficient of determination with future in-season win percentage. We note the improvement our team strength estimates offer over season-to-date win percentage and season-to-date point differential in all sports, especially early in the season.

Figure 6 shows the coefficient of determination ( $R^2$ ) between each team's future won-loss percentage in a season and each team's (i) average team strength estimate from Model IHA, (ii) season-to-date cumulative point differential, and (iii) season-to-date won-loss percentage. Within each sport, this is computed by game number, which helps to account for league-level differences in season length. For purposes of using our team strength estimates, we took the mean posterior draw for each team in each week a particular contest was played. The lockout-shortened seasons in the NBA (2012) and NHL (2013) were dropped.

Across each sport, our estimates of team strength consistently outperform past team win percentage and point differential in predicting future win percentage. This gap is most pronounced earlier in each season, which is not surprising given the instability of won-loss percentage and point differential in a small number of games. Differences in predictive accuracy remain throughout most of the regular season in MLB, the NHL, and the NFL. However, by the NBA's mid-season, won-loss ratio and point differential are similar to our estimates of team strength in assessing future performance. By and large, this confirms the findings of Wolfson and Koopmeiners (2015), who identified that most of the

information needed to predict the remainder of the NBA season is contained within the first third of the year.

Altogether, results suggest that across seasons and sports, team strength estimates from our state-space model more accurately assess team caliber than won-loss percentage and point differential.

#### 4.6 How often does the best team win? A new measure of league parity

We conclude by addressing our initial question about the inherent randomness of game outcomes.

One simple way to compare league randomness would be to contrast the observed distribution of  $p_{(q,s,k)ij}$ 's between each  $q$ . However, while sportsbook odds can be used to infer the probability of each team winning, these odds are only provided for scheduled games. As a result, any between-league comparisons using sportsbook odds alone would be contingent upon each league's actual schedule, and they may not accurately reflect differences that would be observed if all teams were to play one another.

A second option would be to contrast our posterior draws of  $\theta_{(q,s,k)i}$  for all  $i$ , either across time periods or at a fixed point in time, as these estimates account for league particulars such as strength of schedule. While possible with our team strength estimates, which are presented on identical scales, such a procedure would not generalize to other sports or leagues where betting market data may not be unavailable.

Instead, to assess the equivalence of all teams in each league, we consider the likelihood that—given any pair of teams chosen at random—the better team wins, by simulating estimates of  $p_{(q,s,k)ij}$  using posterior draws of team strength, home advantage, and game level error. For our purposes, we define the *better* team to be the one, *a priori*, with a higher probability of winning that game. If a contest has no inherent randomness (consider the Harlem Globetrotters), then the better team *always* wins.<sup>2</sup> Conversely, if game-level variability is large relative to the difference in team strength, then even the inferior team might win nearly half the time.

Using our posterior draws, we approximate the distribution of game-level probabilities between two randomly chosen teams using the following steps.

Given sport  $q$  with season length  $K_q$ , number of seasons  $S_q$ , and number of teams  $t_q$ ,

---

<sup>2</sup>The Harlem Globetrotters are an exhibition basketball team that plays hundreds of games in a year, rarely losing

1. Draw season  $\tilde{s}$  from  $\{1, \dots, S_q\}$ , and week  $\tilde{k}$  from  $\{1, \dots, K_q\}$ .
2. Draw teams  $\tilde{i}$  and  $\tilde{j}$  from  $\{1, \dots, t_q\}$  without replacement.
3. Sample one posterior draw of team strength for  $\tilde{i}$  and  $\tilde{j}$ ,  $\tilde{\theta}_{(q,\tilde{s},\tilde{k})\tilde{i}}$  and  $\tilde{\theta}_{(q,\tilde{s},\tilde{k})\tilde{j}}$ , respectively, from the posterior distributions of  $\tilde{i}$  and  $\tilde{j}$ 's team strength estimates during season  $\tilde{s}$  at week  $\tilde{k}$ .
4. Sample one posterior draw of the HA,  $\tilde{\alpha}_{q_0}$ , from the posterior distribution of  $\alpha_{q_0}$ .
5. Sample one posterior draw of the game-level variance parameter,  $\tilde{\sigma}_{q,game}^2$ , and draw a game-level error,  $\tilde{\epsilon}_{q,game}$ , from  $\tilde{\epsilon}_{q,game} \sim N(0, \tilde{\sigma}_{q,game})$
6. Impute the simulated log-odds of the better team winning between  $\tilde{i}$  and  $\tilde{j}$ ,  $\text{logit}(\tilde{p}_{(q,\tilde{s},\tilde{k})\tilde{i}\tilde{j}}) = \tilde{\alpha}_{q_0} + |\tilde{\theta}_{(q,\tilde{s},\tilde{k})\tilde{i}} - \tilde{\theta}_{(q,\tilde{s},\tilde{k})\tilde{j}} + \tilde{\epsilon}_{q,game}|$ , where the *better* team's log-odds are based on  $\tilde{\theta}_{(q,\tilde{s},\tilde{k})\tilde{i}}$ ,  $\tilde{\theta}_{(q,\tilde{s},\tilde{k})\tilde{j}}$ , and  $\tilde{\epsilon}_{q,game}$ .
7. Transform  $\text{logit}(\tilde{p}_{(q,\tilde{s},\tilde{k})\tilde{i}\tilde{j}})$  into a probability to obtain a simulated estimate,  $\tilde{p}_{q,sim}$ , where  $\tilde{p}_{q,sim} = \tilde{p}_{(q,\tilde{s},\tilde{k})\tilde{i}\tilde{j}}$
8. Repeat the above steps  $n_{sim}$  times to obtain  $\tilde{\mathbf{p}}_q = \{\tilde{p}_{q,1}, \dots, \tilde{p}_{q,n_{sim}}\}$ .

For each  $q$ , we simulated with  $n_{sim} = 1000$ . Additionally, to remove the effect of each league's HA on simulated probabilities, we repeated the process fixing  $\tilde{\alpha}_{q_0} = 0$  for each league to reflect game probabilities played at neutral sites.

Figure 7 shows the cumulative distribution functions (CDFs) for each set of probabilities in each league. The median probability of the best team winning a neutral site game is highest in the NBA (67%), followed in order by the NFL (65%), NHL (57%), and MLB (56%). The spread of these probabilities are of great interest. Nearly every simulated MLB and NHL game played at neutral site is less than a 3:1 proposition with respect to the best team winning (75%). Meanwhile, roughly 28% of NBA and 20% of NFL neutral site match-ups are greater than this 3:1 threshold.

Factoring in each league's home advantage works to exaggerate league-level differences. When the best team plays at home in the NBA, it is always favored to win at least 60% of the time, with the middle 50% of games ranging from a 68% probability to an 84% probability. Meanwhile, even with a home advantage, it is rare that the best MLB team is *ever* given a 70% probability of winning, with the middle 50% of games ranging from 57% to 63%.

## How often does the best team win?

Solid: neutral site, Dashed: home game for better team

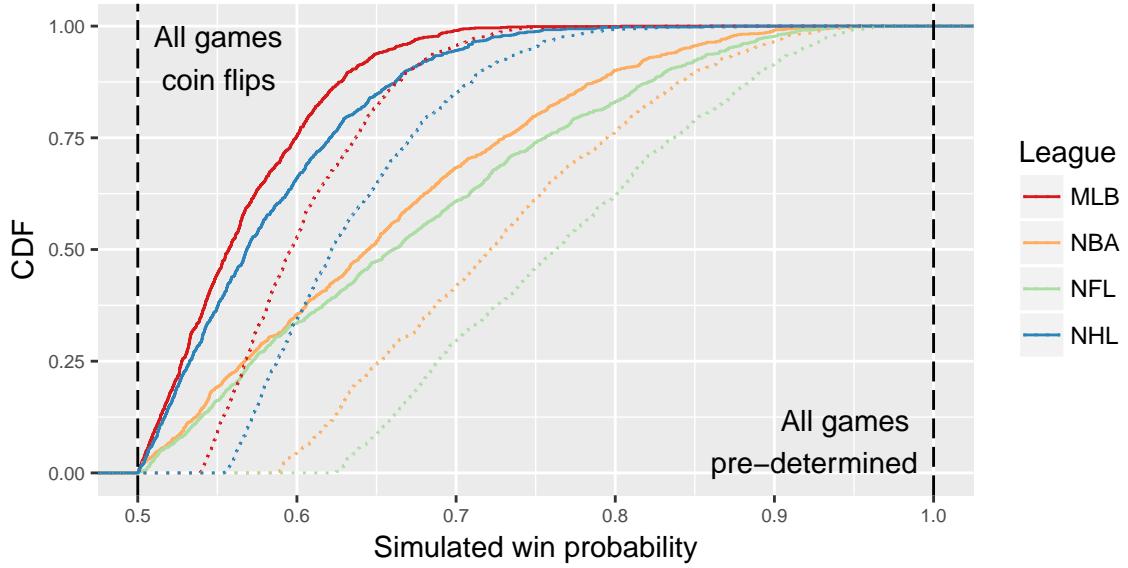


Figure 7: Cumulative distribution function (CDF) of 1000 simulated game-level probabilities in each league, for both neutral site and home games, with the better team (on average) used as the reference and given the home advantage.

Finally, we use the CDFs displayed in Figure 7 to quantify the cumulative difference between each league’s game-level probabilities and a league of coin flips by estimating the approximate area under each curve. Let  $Parity_q$  be our parity measure, such that

$$Parity_q = 2 \int_{0.5}^1 P(\tilde{\mathbf{p}}_q \leq x) dx,$$

where we multiply by 2 in order to scale so that  $0 \leq Parity_q \leq 1$ , where 1 represents complete parity (every game a coin flip) and 0 represents no parity (every game outcome pre-determined).

For games with no home advantage,  $Parity_{MLB} = 0.86$ , followed by the NHL (0.83), NFL (0.69), and NBA (0.66). When the best team has a home advantage, parity is again the greatest in the MLB (0.79), followed by the NHL (0.73), NFL (0.54), and NBA (0.46). These results suggest that when the best team is playing at home, the NBA is closer to a world where every game outcome is predetermined than to one where every game outcome is a coin flip. Meanwhile, even when giving the best team a HA, MLB game outcomes remain lightly-weighted coin flips.

## 5 Conclusion

### 5.1 Summary

Using a modified Bayesian state-space model, we estimate both time-varying team strength and league-level variance parameters in order to better understand the underlying randomness in the four major North American professional sporting leagues, the NBA, NFL, NHL, and MLB.

Our first finding relates to the relative equivalence of the four leagues. At a single point in time, team strength estimates diverge substantially more in the NBA and NFL than in the NHL and MLB. In the latter two leagues, contests between two randomly chosen teams are closer to a coin-flip, in which each team has a reasonable shot at winning. Understanding this underlying randomness would appear to be crucial for decision makers in these leagues. At critical moments in a team’s evolution, such as the a trade deadline, free agency period, or the decision to fire a coach, we recommend that team officials look past wins and losses to better understand team strength in the context of their league. As one easy example, it is insufficient to evaluate a baseball or hockey team based on their performance in the postseason alone, given that so many of those contests are nearly 50-50 outcomes.

Our second set of findings relates to the autoregressive nature of team strengths. Within a season, posterior estimates suggest that teams in each of the NBA (largest reversion), NFL, and NHL tend to revert towards the league average in the long term on a week-to-week basis, while trends of team strength in MLB are indistinguishable from a random walk. On a season-to-season basis, NHL teams exhibit the largest reversion (nearly 50%) towards the league average, with the other three leagues falling somewhere between roughly 25% and 40%.

Our next finding relates to the relative equivalence of the home advantage in each league, with the NBA well ahead of the pack, with teams averaging a 62.0% chance of winning versus a like-caliber opponent. We also show that the home advantage varies most significantly between venues within each of the NBA and the NHL. In the NBA, for example, the league’s best team home advantage is worth a few wins per year, in expectation, over the league’s worst home advantage. Moreover, with the exception of the Colorado Rockies, it is not clear that any MLB or NFL team has a statistically significant home effect.

Finally, we identify that incorporating information from betting markets can help to more accurately

gauge the caliber of each league’s teams, as shown by an improved ability to predict future team performance. Unlike wins and losses or point differential, our estimates of team strength account for league characteristics such as unbalanced schedules and season length. We conclude by using these team strength draws to propose a parity metric that can compare team equivalence without being affected by league-level characteristics like unbalanced schedules.

## 5.2 Future work

Opportunities to extend our model are plentiful. One approach would use our team strength estimates to examine how each league’s scheduling quirks impact resulting won-loss standings. For example, what is the impact of the unbalanced schedule used in the NFL? A second question concerns the relationship of our estimates of team strength to performance in the postseason. How likely is it for the best team to win each league’s title? Conversely, how likely is it that the team that won the postseason tournament was actually the strongest team at the end of the regular season? Finally, one could use time-varying estimates of team strength to consider the existence of tanking, in which teams—in order to secure a better draft position—are better off losing games later in the season. While this has been demonstrated in basketball using betting market data (Soebbing and Humphreys, 2013), it would also be worth looking at tanking in other leagues, or if team interest in tanking corresponds to the perceived talent available in the upcoming draft.

To maintain consistency with the NFL’s calendar, we considered time on a weekly basis. More refined approaches may be appropriate in other sports. As an example, investigation into starting pitchers in baseball—who change daily—could lead to novel findings. Additionally, another model specification could consider the possibility that time-varying estimates of team strength follow something other than an autoregressive structure. One alternative specification, for example, is a stochastic volatility process (Glickman, 2001). In this respect, our model can be considered a starting point for those looking to dig deeper in any sport without losing an ability to make cross-league comparisons.

## References

- Baker, R. D. and McHale, I. G. (2015), “Time varying ratings in association football: the all-time greatest team is..” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178, 481–492.

- Berri, D. (2014), “Noll-Scully,” <http://wagesofwins.com/noll-scully/>, accessed May 19, 2016.
- Berri, D. J. and Schmidt, M. B. (2006), “On the road with the National Basketball Association’s superstar externality,” *Journal of Sports Economics*, 7, 347–358.
- Boulier, B. L. and Stekler, H. O. (2003), “Predicting the outcomes of National Football League games,” *International Journal of Forecasting*, 19, 257–270.
- Bradley, R. A. and Terry, M. E. (1952), “Rank analysis of incomplete block designs: I. The method of paired comparisons,” *Biometrika*, 39, 324–345.
- Buttrey, S. E. (2016), “Beating the market betting on NHL hockey games,” *Journal of Quantitative Analysis in Sports*, 12, 87–98.
- Carlin, B. P. (1996), “Improved NCAA basketball tournament modeling via point spread and team strength information,” *The American Statistician*, 50, 39–43.
- Cattelan, M., Varin, C., and Firth, D. (2013), “Dynamic Bradley–Terry modelling of sports tournaments,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62, 135–150.
- CFP (2014), “Bowl Championship Series explained,” [http://www.collegefootballpoll.com/bcs\\_explained.html](http://www.collegefootballpoll.com/bcs_explained.html), accessed May 19, 2016.
- Colquitt, L. L., Godwin, N. H., and Caudill, S. B. (2001), “Testing efficiency across markets: Evidence from the NCAA basketball betting market,” *Journal of Business Finance & Accounting*, 28, 231–248.
- Crabtree, C. (2014), “NFL wary of putting Seahawks home games in prime time,” <http://profootballtalk.nbcsports.com/2014/04/24/nfl-wary-of-putting-seahawks-home-games-in-prime-time-due-to-recent-blowouts/>, accessed October 19, 2016.
- Crooker, J. R. and Fenn, A. J. (2007), “Sports leagues and parity when league parity generates fan enthusiasm,” *Journal of Sports Economics*, 8, 139–164.
- Demers, S. (2015), “Riding a probabilistic support vector machine to the Stanley Cup,” *Journal of Quantitative Analysis in Sports*, 11, 205–218.
- Elo, A. E. (1978), *The rating of chessplayers, past and present*, Arco Pub.

- Gandar, J., Zuber, R., O'brien, T., and Russo, B. (1988), "Testing rationality in the point spread betting market," *The Journal of Finance*, 43, 995–1008.
- Glickman, M. E. (1995), "A comprehensive guide to chess ratings," *American Chess Journal*, 3, 59–102.
- (2001), "Dynamic paired comparison models with stochastic variances," *Journal of Applied Statistics*, 28, 673–689.
- Glickman, M. E. and Stern, H. S. (1998), "A state-space model for National Football League scores," *Journal of the American Statistical Association*, 93, 25–35.
- (2016), "Estimating team strength in the NFL," in *Handbook of Statistical Methods and Analyses in Sports*, eds. Albert, J., Glickman, M. E., Swartz, T. B., and Koning, R. H., Chapman and Hall/CRC Press: Boca Raton, FL, chap. 5, pp. 113–135.
- Harville, D. (1980), "Predictions for National Football League games via linear-model methodology," *Journal of the American Statistical Association*, 75, 516–524.
- Humphreys, B. R. (2002), "Alternative measures of competitive balance in sports leagues," *Journal of Sports Economics*, 3, 133–148.
- Knorr-Held, L. (2000), "Dynamic rating of sports teams," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49, 261–276.
- Knowles, G., Sherony, K., and Haupert, M. (1992), "The demand for Major League Baseball: A test of the uncertainty of outcome hypothesis," *The American Economist*, 36, 72–80.
- Koopmeiners, J. S. (2012), "A Comparison of the Autocorrelation and Variance of NFL Team Strengths Over Time using a Bayesian State-Space Model," *Journal of Quantitative Analysis in Sports*, 8, 1–19.
- Lacey, N. J. (1990), "An estimation of market efficiency in the NFL point spread betting market," *Applied Economics*, 22, 117–129.
- Lee, Y. H. and Fort, R. (2008), "Attendance and the uncertainty-of-outcome hypothesis in baseball," *Review of Industrial Organization*, 33, 281–295.
- Leeds, M. and Von Allmen, P. (2004), "The economics of sports," *The Business of Sports*, 361–366.

- Lenten, L. J. (2015), "Measurement of competitive balance in conference and divisional tournament design," *Journal of Sports Economics*, 16, 3–25.
- Loeffelholz, B., Bednar, E., Bauer, K. W., et al. (2009), "Predicting NBA games using neural networks," *Journal of Quantitative Analysis in Sports*, 5, 1–15.
- Lopez, M. J. (2013), "Inefficiencies in the national hockey league points system and the teams that take advantage," *Journal of Sports Economics*, 16, 410–424.
- Lopez, M. J. and Matthews, G. J. (2015), "Building an NCAA mens basketball predictive model and quantifying its success," *Journal of Quantitative Analysis in Sports*, 11, 5–12.
- Lopez, M. J. and Schuckers, M. (2016), "Predicting coin flips: using resampling and hierarchical models to help untangle the NHLs shoot-out," *Journal of Sports Sciences*, 1–10.
- Manner, H. (2015), "Modeling and forecasting the outcomes of NBA basketball games," *Journal of Quantitative Analysis in Sports*.
- Massey, K. (1997), "Statistical models applied to the rating of sports teams," Tech. rep., Bluefield College, honor's thesis.
- Matthews, G. J. (2005), "Improving paired comparison models for NFL point spreads by data transformation," Ph.D. thesis, Worcester Polytechnic Institute.
- Miljković, D., Gajić, L., Kovačević, A., and Konjović, Z. (2010), "The use of data mining for basketball matches outcomes prediction," in *IEEE 8th International Symposium on Intelligent Systems and Informatics*, IEEE, pp. 309–312.
- Moskowitz, T. and Wertheim, L. J. (2011), *Scorecasting: The hidden influences behind how sports are played and games are won*, Crown Archetype: New York, NY.
- Nichols, M. W. (2012), "The impact of visiting team travel on game outcome and biases in NFL betting markets," *Journal of Sports Economics*, 15, 78–96.
- Noll, R. G. (1991), "Professional Basketball: Economic and Business Perspectives," in *The Business of Professional Sports*, eds. Mangan, J. A. and Staudohar, P. D., University of Illinois Press: Urbana, IL, pp. 18–47.
- Owen, A. (2011), "Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter," *IMA Journal of Management Mathematics*, 22, 99–113.

- Owen, P. D. (2010), “Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues,” *Economics Letters*, 109, 38–41.
- Owen, P. D. and King, N. (2015), “Competitive balance measures in sports leagues: the effects of variation in season length,” *Economic Inquiry*, 53, 731–744.
- Owen, P. D., Ryan, M., and Weatherston, C. R. (2007), “Measuring competitive balance in professional team sports using the Herfindahl-Hirschman index,” *Review of Industrial Organization*, 31, 289–302.
- Paine, N. (2013), “Analyzing real home court advantage,” [http://insider.espn.com/nba/insider/story/\\_/id/9014283/nba-analyzing-real-home-court-advantage-utah-jazz-denver-nuggets](http://insider.espn.com/nba/insider/story/_/id/9014283/nba-analyzing-real-home-court-advantage-utah-jazz-denver-nuggets), accessed October 19, 2016.
- Paul, R. J. and Weinbach, A. P. (2014), “Market efficiency and behavioral biases in the wnba betting market,” *International Journal of Financial Studies*, 2, 193–202.
- Plummer, M. (2016), *rjags: Bayesian Graphical Models using MCMC*, R package version 4-6.
- Rottenberg, S. (1956), “The baseball players’ labor market,” *The Journal of Political Economy*, 242–258.
- Scully, G. W. (1989), *The Business of Major League Baseball*, University of Chicago Press: Chicago, IL.
- Soebbing, B. P. and Humphreys, B. R. (2013), “Do gamblers think that teams tank? Evidence from the NBA,” *Contemporary Economic Policy*, 31, 301–313.
- Spann, M. and Skiera, B. (2009), “Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters,” *Journal of Forecasting*, 28, 55–72.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002), “Bayesian measures of model complexity and fit,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.
- Stern, H. (1991), “On the probability of winning a football game,” *The American Statistician*, 45, 179–183.
- Tutz, G. and Schauberger, G. (2015), “Extended ordered paired comparison models with application to football data from German Bundesliga,” *AStA Advances in Statistical Analysis*, 99, 209–227.

Utt, J. and Fort, R. (2002), “Pitfalls to measuring competitive balance with Gini coefficients,” *Journal of Sports Economics*, 3, 367–373.

Wolfson, J. and Koopmeiners, J. S. (2015), “Who’s good this year? Comparing the Information Content of Games in the Four Major US Sports,” *arXiv preprint arXiv:1501.07179*.

Yang, T. Y. and Swartz, T. (2004), “A two-stage Bayesian model for predicting winners in major league baseball,” *Journal of Data Science*, 2, 61–73.

**Supplementary Materials for**

**“A unified approach to understanding randomness in sport”**

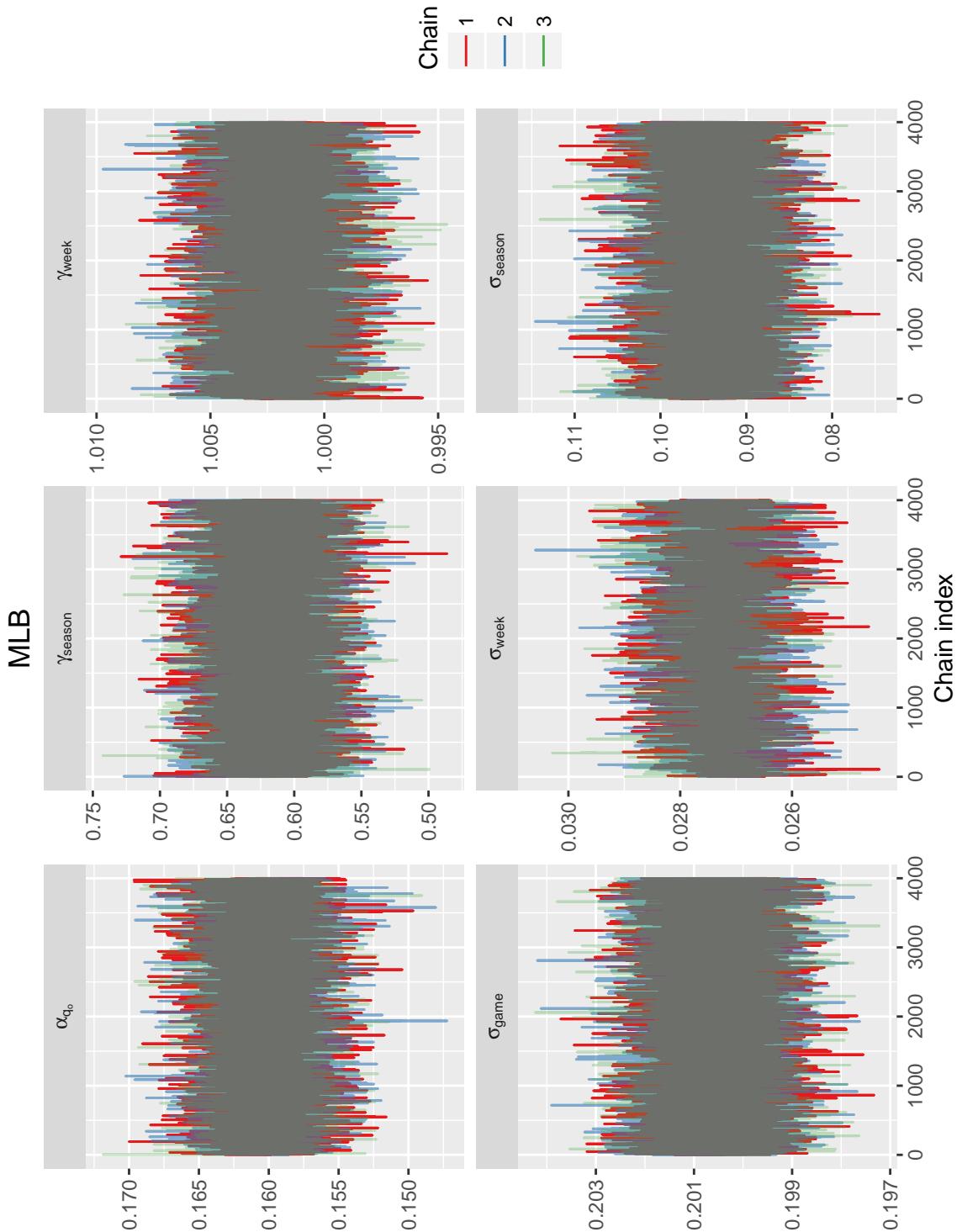


Figure 8: Trace plots of MLB parameters

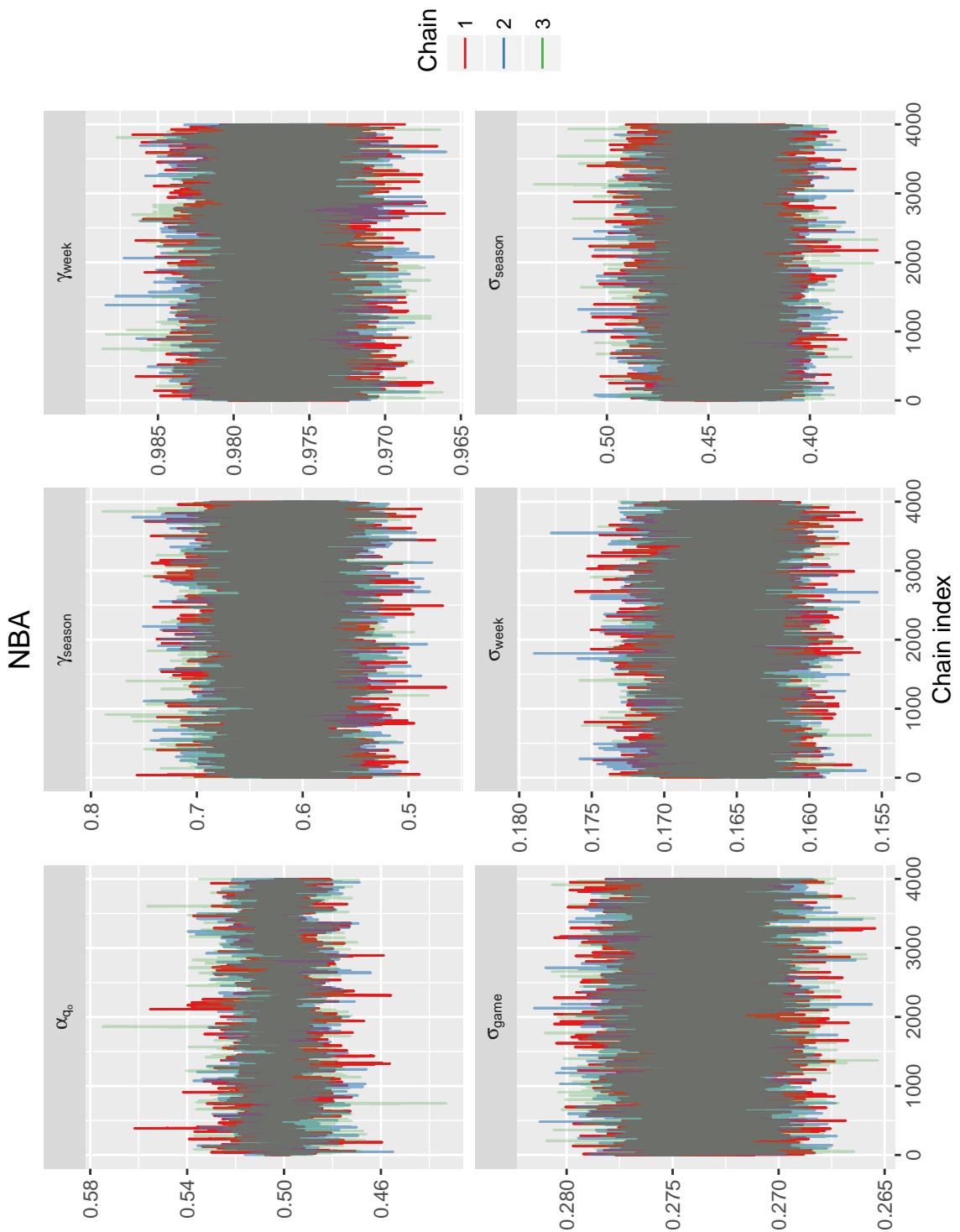


Figure 9: Trace plots of NBA parameters

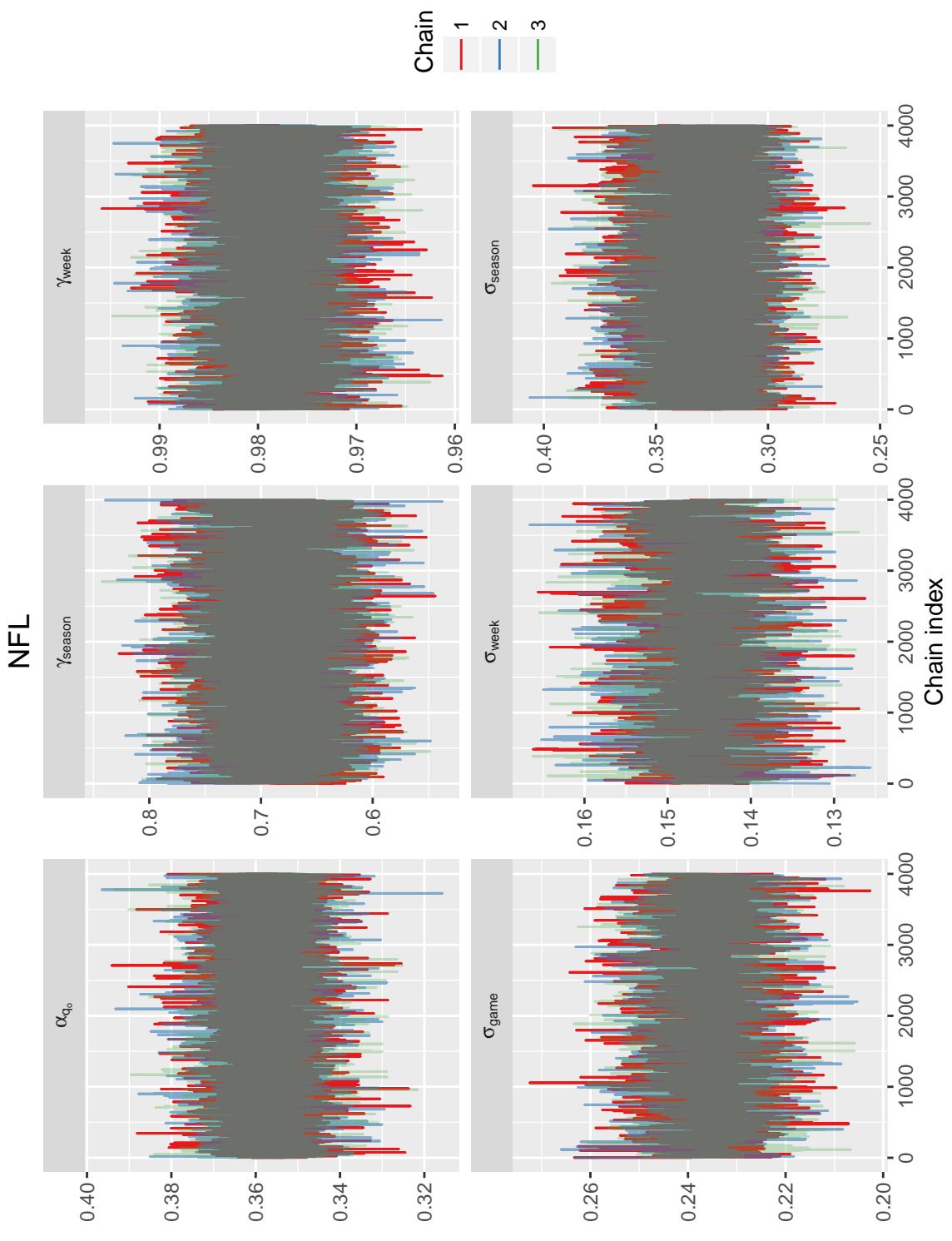


Figure 10: Trace plots of NFL parameters

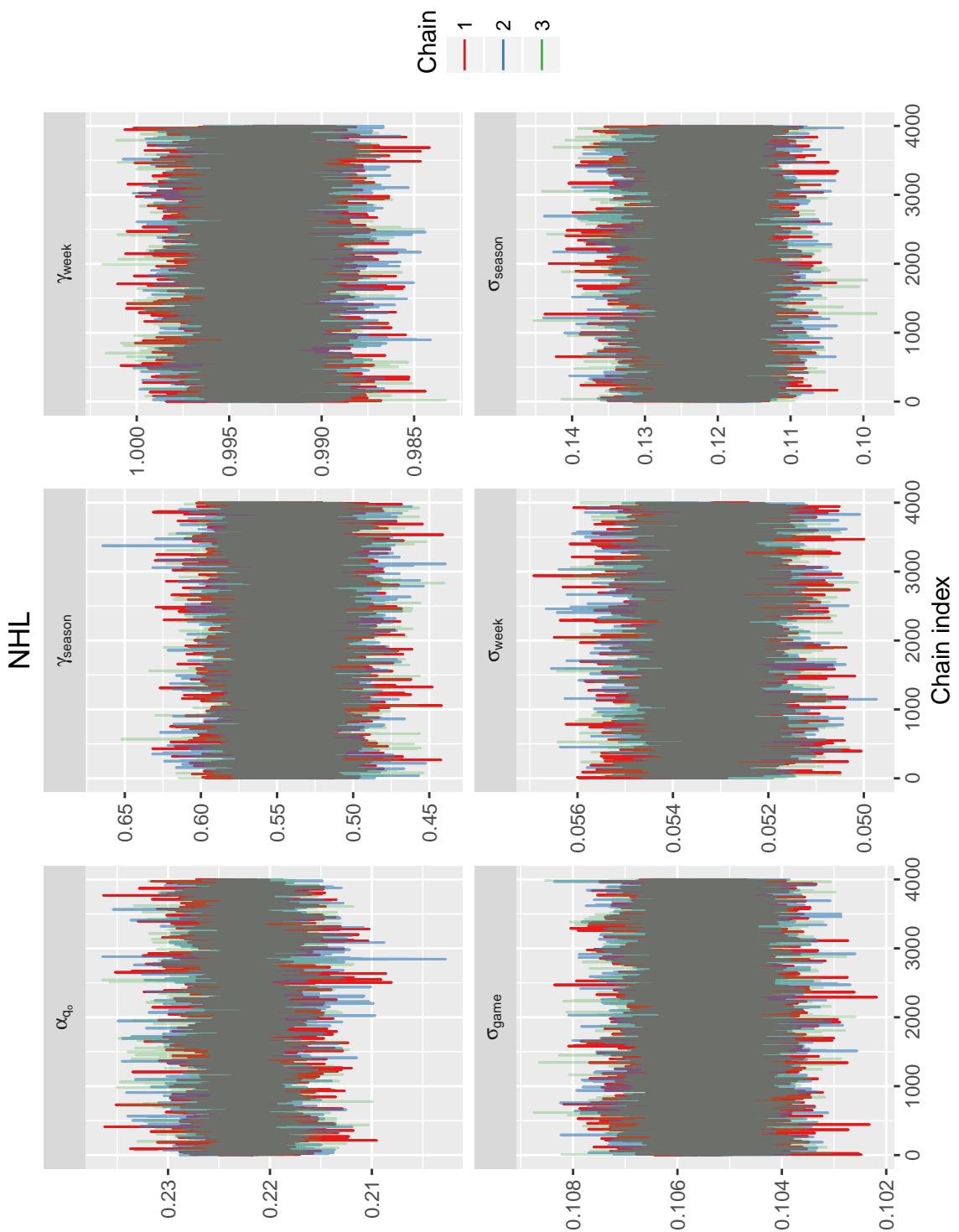


Figure 11: Trace plots of NHL parameters

## Team strength parameters over time, MLB

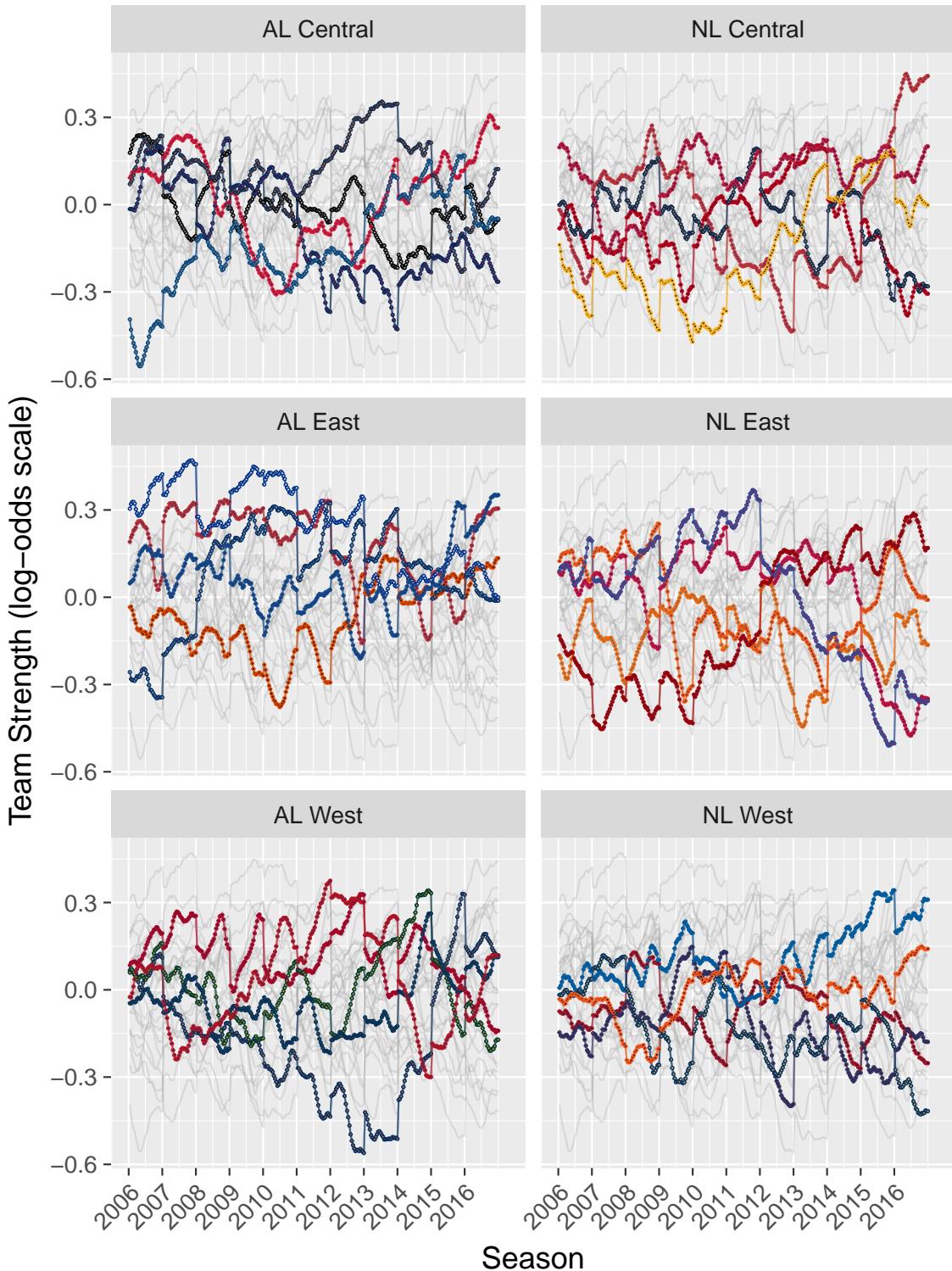


Figure 12: Team strength coefficients over time for Major League Baseball.

## Team strength parameters over time, NBA

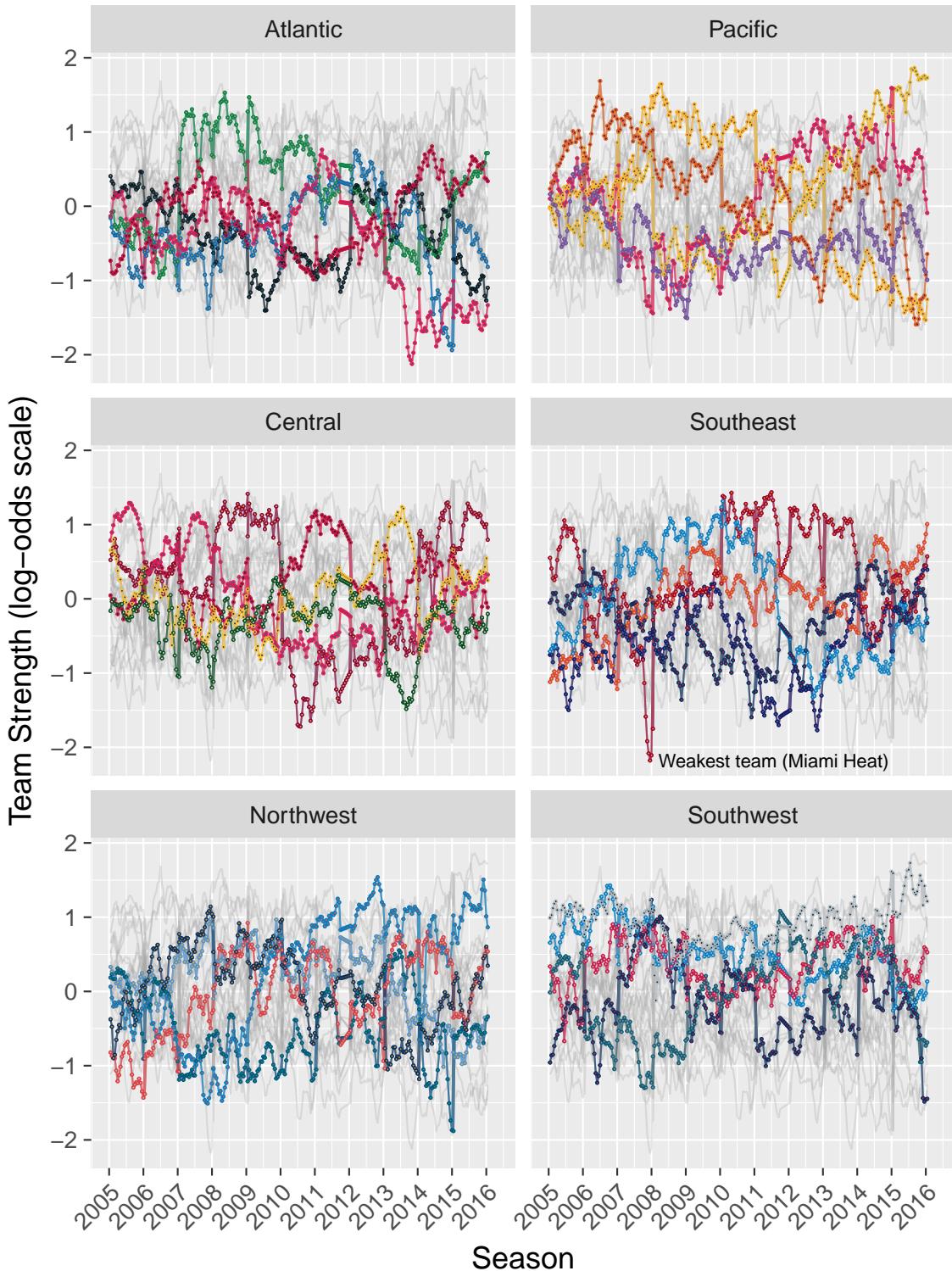


Figure 13: Team strength coefficients over time for the National Basketball Association.

## Team strength parameters over time, NFL

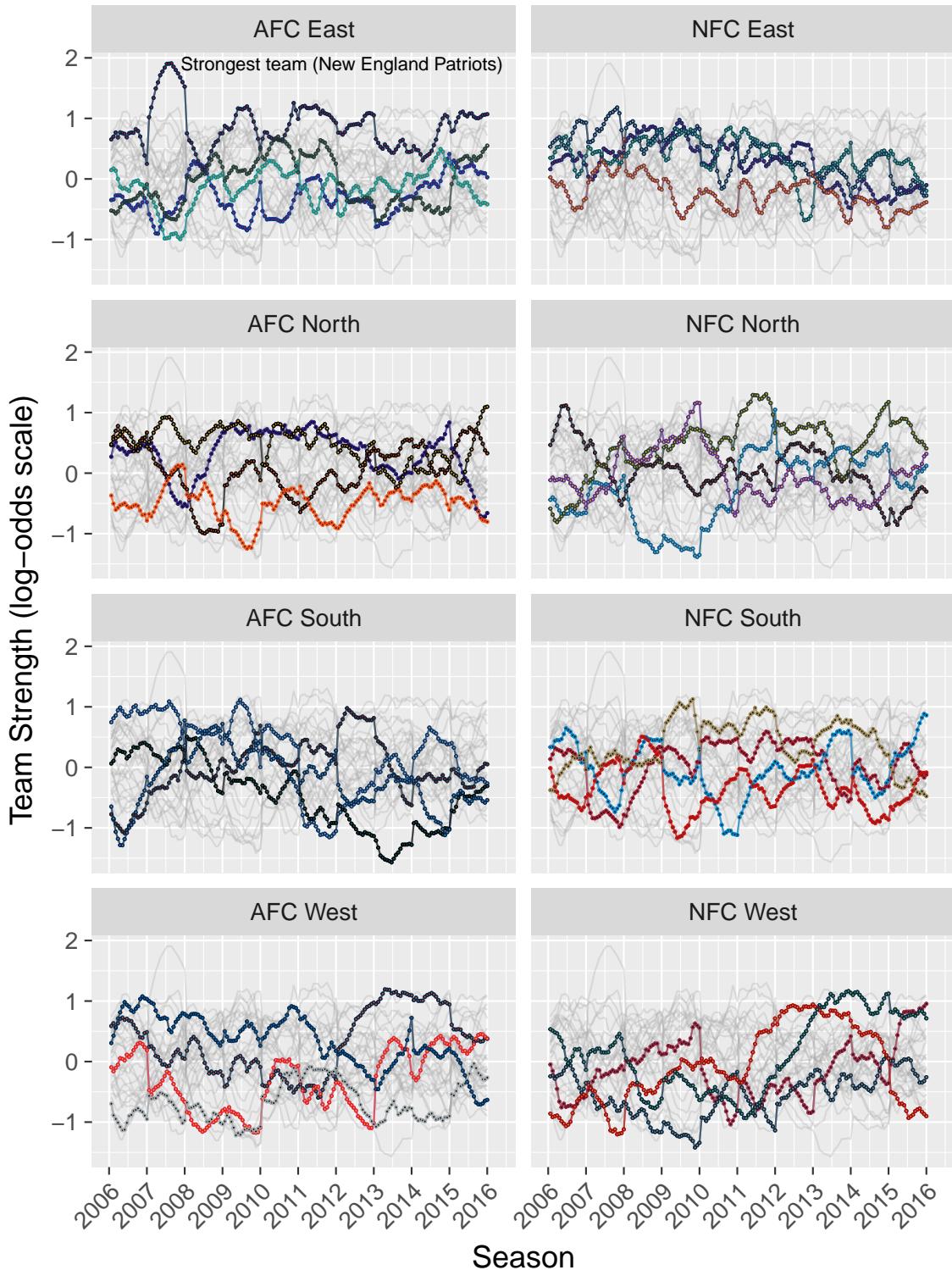


Figure 14: Team strength coefficients over time for the National Football League.

## Team strength parameters over time, NHL

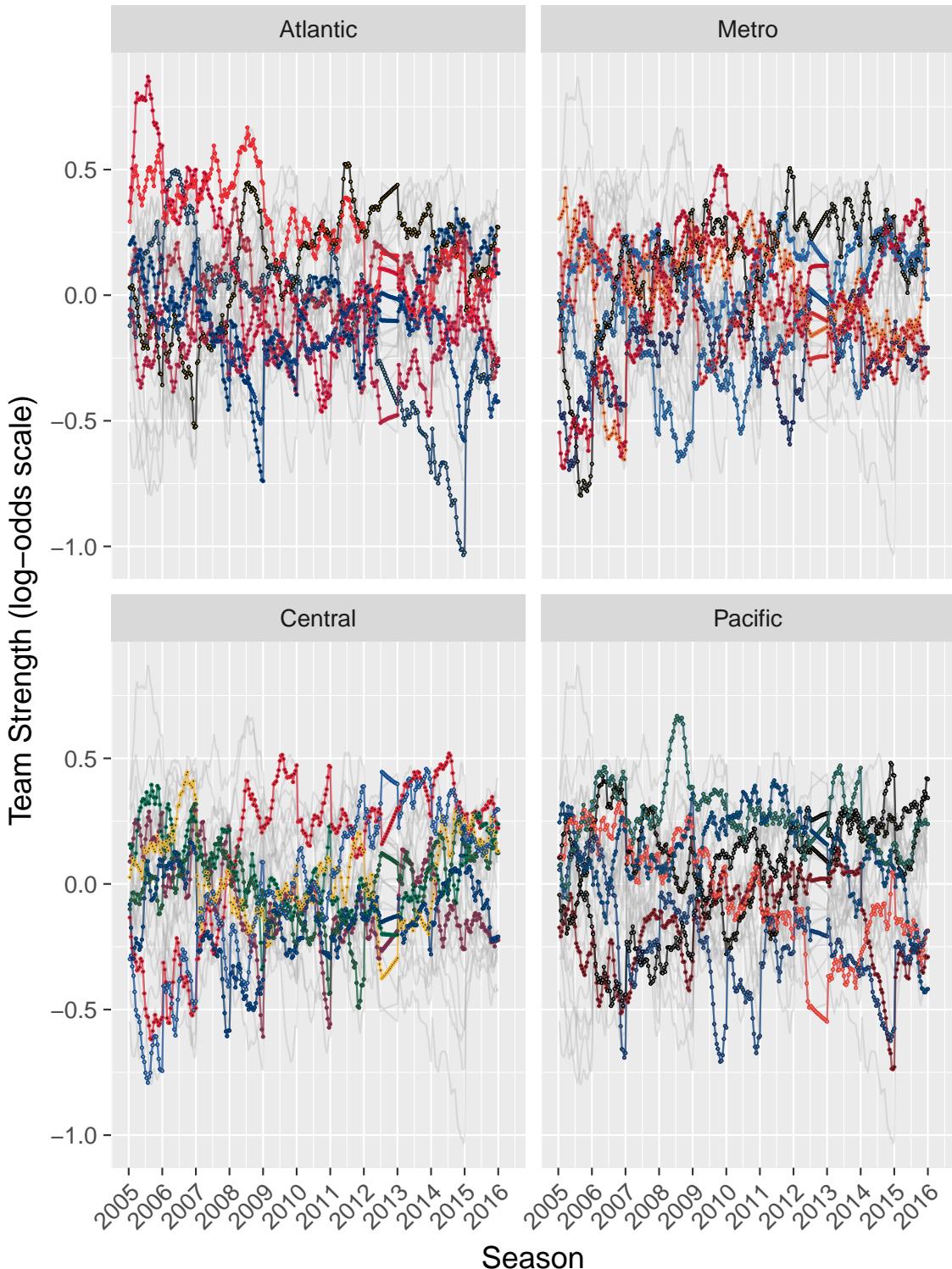


Figure 15: Team strength coefficients over time for the National Hockey League.