HOW OFTEN DOES THE BEST TEAM WIN?

Michael J. Lopez, Gregory J. Matthews, and Benjamin S. Baumer

Srikanth Aravamuthan and Sean Kent
Sports Analytics Reading Group
April 21, 2022

Luck and parity in sports

Using statistics to assess luck in sports

- The best team does not always win.

- How often does the best team win in each sport?

- How to untangle luck from skill?

What parity looks like

What parity does not look like

Parity checklist

- Equality at a fixed time
- Postseason tournament
- Within season equality
- Between season equality

A cross-sport model

Prior work

A State-Space Model for National Football League Scores

Mark E. GLICKMAN and Hal S. STERN

This article develops a predictive model for National Football League (NFL) game scores using data from the period 1988–1993. The parameters of primary interest—measures of team strength—are expected to vary over time. Our model accounts for this source of variability by modeling football outcomes using a state-space model that assumes team strength parameters follow a first-order autoregressive process. Two sources of variation in team strengths are addressed in our model; week-to-week changes in team strength due to injuries and other random factors, and season-to-season changes resulting from changes in personnel and other longer-term factors. Our model also incorporates a home-field advantage while allowing for the possibility that the magnitude of the advantage may vary across teams. The aim of the analysis is to obtain plausible inferences concerning team strengths and other model parameters, and to predict future game outcomes. Iterative simulation is used to obtain samples from the joint posterior distribution of all model parameters. Our model appears to outperform the Las Vegas "betting line" on a small test set consisting of the last 110 games of the 1993 NFL season.

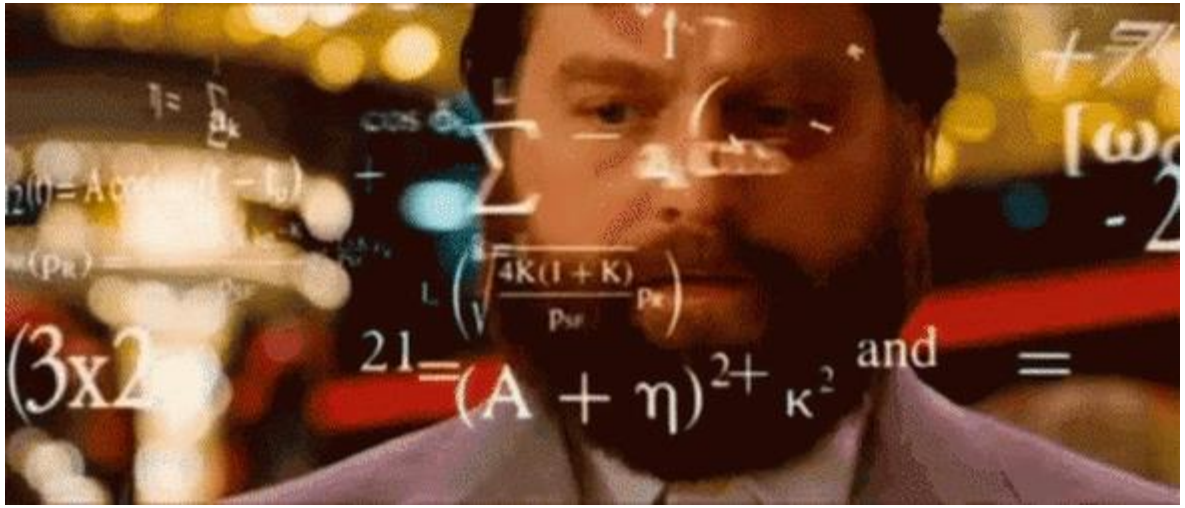
KEY WORDS: Bayesian diagnostics; Dynamic models; Kalman filter; Markov chain Monte Carlo; Predictive inference.

Challenges

- Problem 1: wins and losses alone insufficient (noisy)
- Problem 2: point differential non-generalizable

Challenges

- Problem 1: wins and losses alone insufficient (noisy)
- Problem 2: point differential non-generalizable
- Solution: gambling

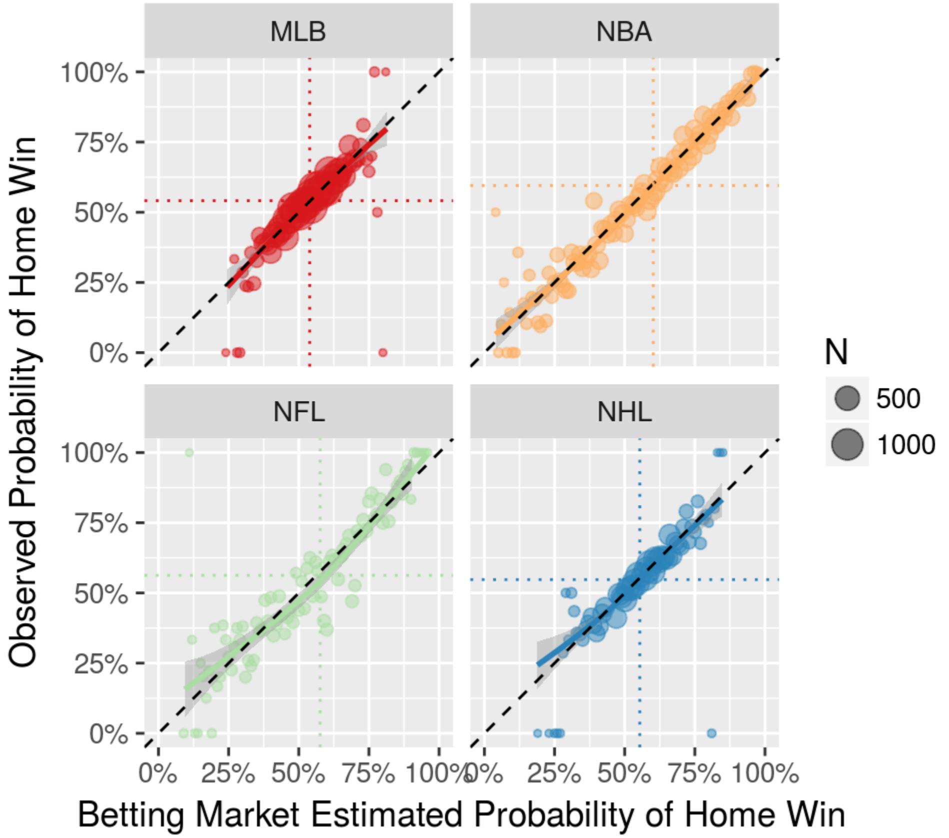


Moneylines

Team	Line (l)	Probability (p)	Normalized
THE ORIGINAL HARLEM GLOBETROTTERS	- 127	0.559	0.548
WASHINGTON GENERALS	+117	0.461	0.452

$$p_i(l_i) = \begin{cases} \dfrac{100}{100 + l_i}; if\ l_i \geq 100 \\ \dfrac{|l_i|}{100 + |l_i|}; if\ l_i \leq -100 \end{cases}$$

Checks out



MLB
NBA
NFL
NHL
100%
75%
50%
25%
0%
0%
25%
50%
75%
100%
Observed Probability of Home Win
Betting Market Estimated Probability of Home Win
N
500
1000

Model definitions

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

- $p_{(q,s,k)i,j}$ is the probability that team i will beat team j in season s during week k of sports league q, for $q \in \{MLB, NBA, NFL, NHL\}$.
- α_{q_0} be the league-wide home advantage (HA) in q.
- $\alpha_{(q)i^*}$ be the extra effect (+ or -) for team i when playing at home.
- $\theta_{(q,s,k)i}$ and $\theta_{(q,s,k)j}$ be season-week team strength parameters.

Model assumptions

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

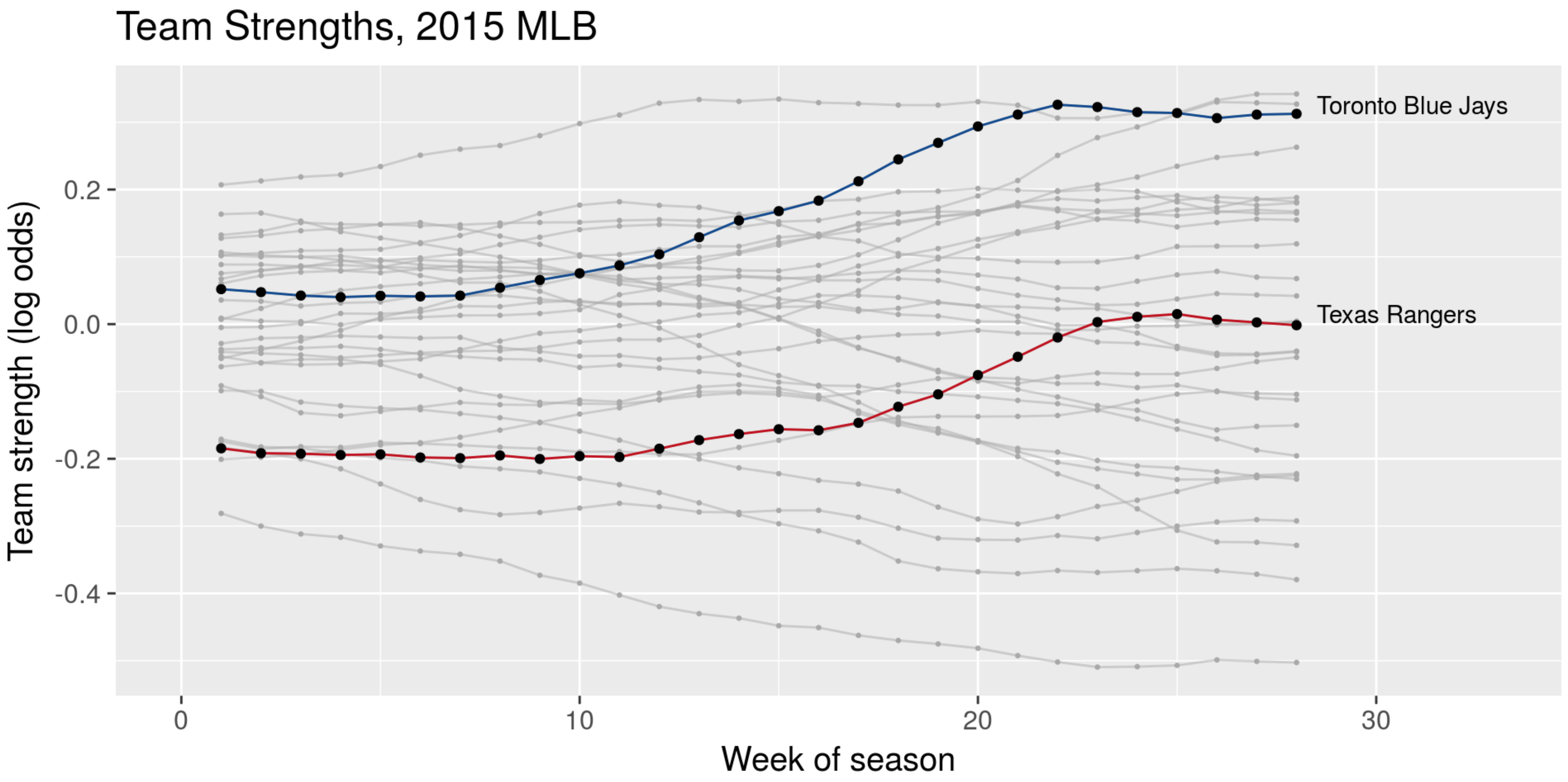
- $\sum_{i=1}^{t_q} \theta_{(q,s,k)i} = 0.$
- $E\left[\theta_{(q,s,k+1)i}\right] = \gamma_{q,week}\theta_{(q,s,k)i}.$
- $E\left[\theta_{(q,s+1,k)i}\right] = \gamma_{q,season}\theta_{(q,s,k)i}.$
- $\gamma_{q,week}\theta_{(q,s,k)i}$ and $\gamma_{q,season}\theta_{(q,s,k)i}$ *week/season level autoregressive parameters.*

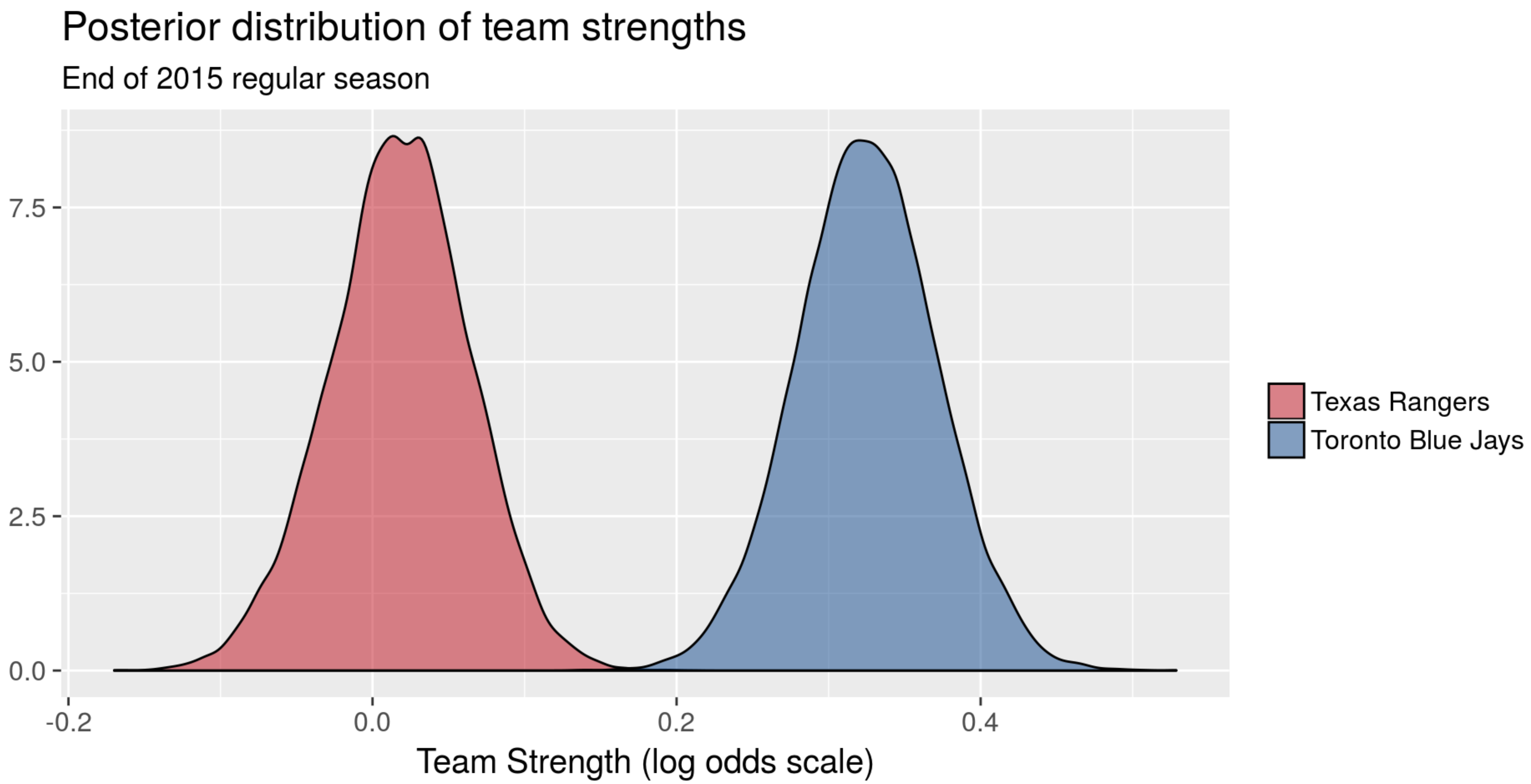
Fitting a cross-sport model

- Data
  - 2006–2016 reg. season in MLB, NBA, NFL, NHL (Sports Insights)

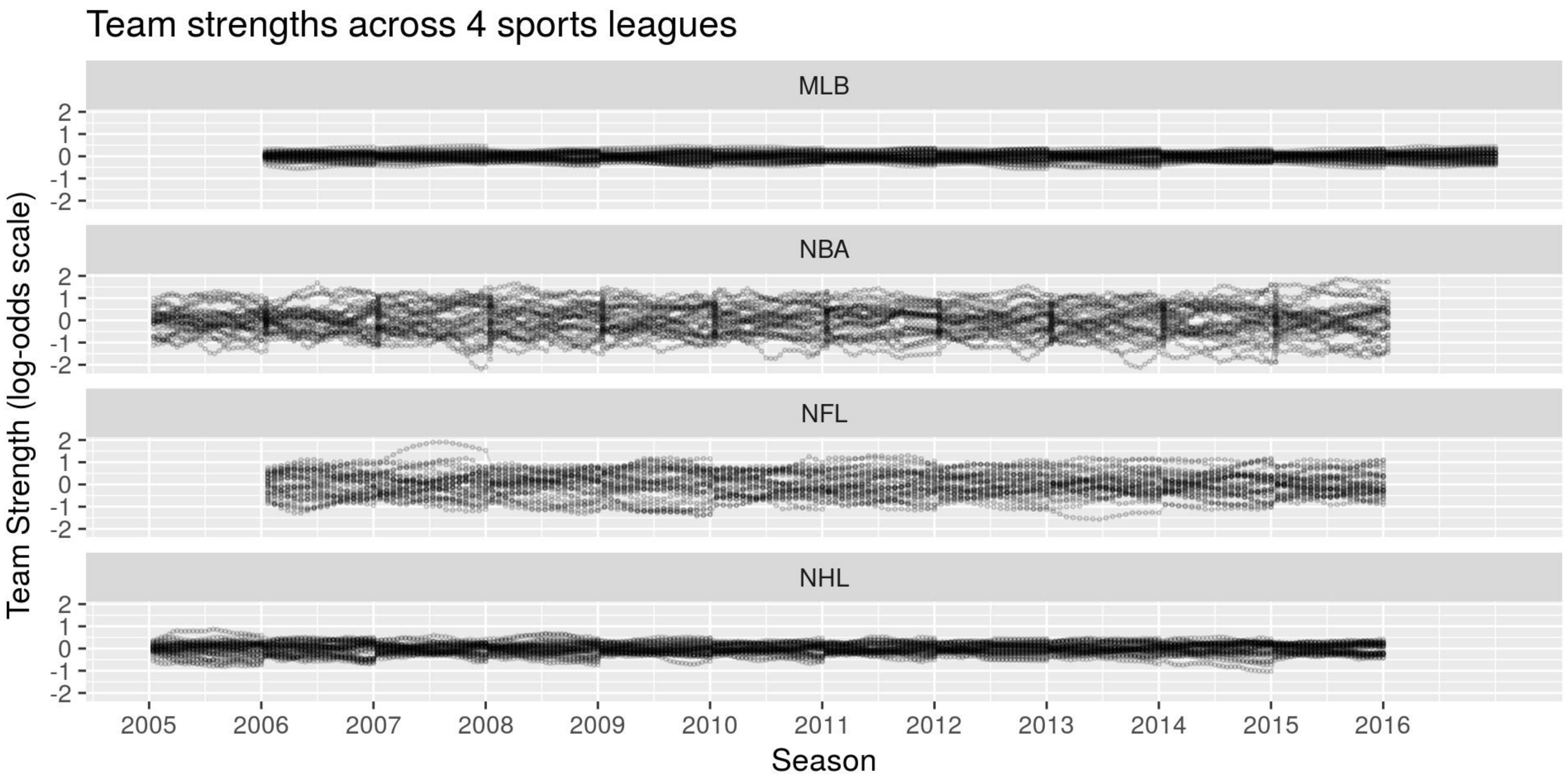
- Priors
  - Uniform (variance parameters) and Normal (team strength parameters)

Results

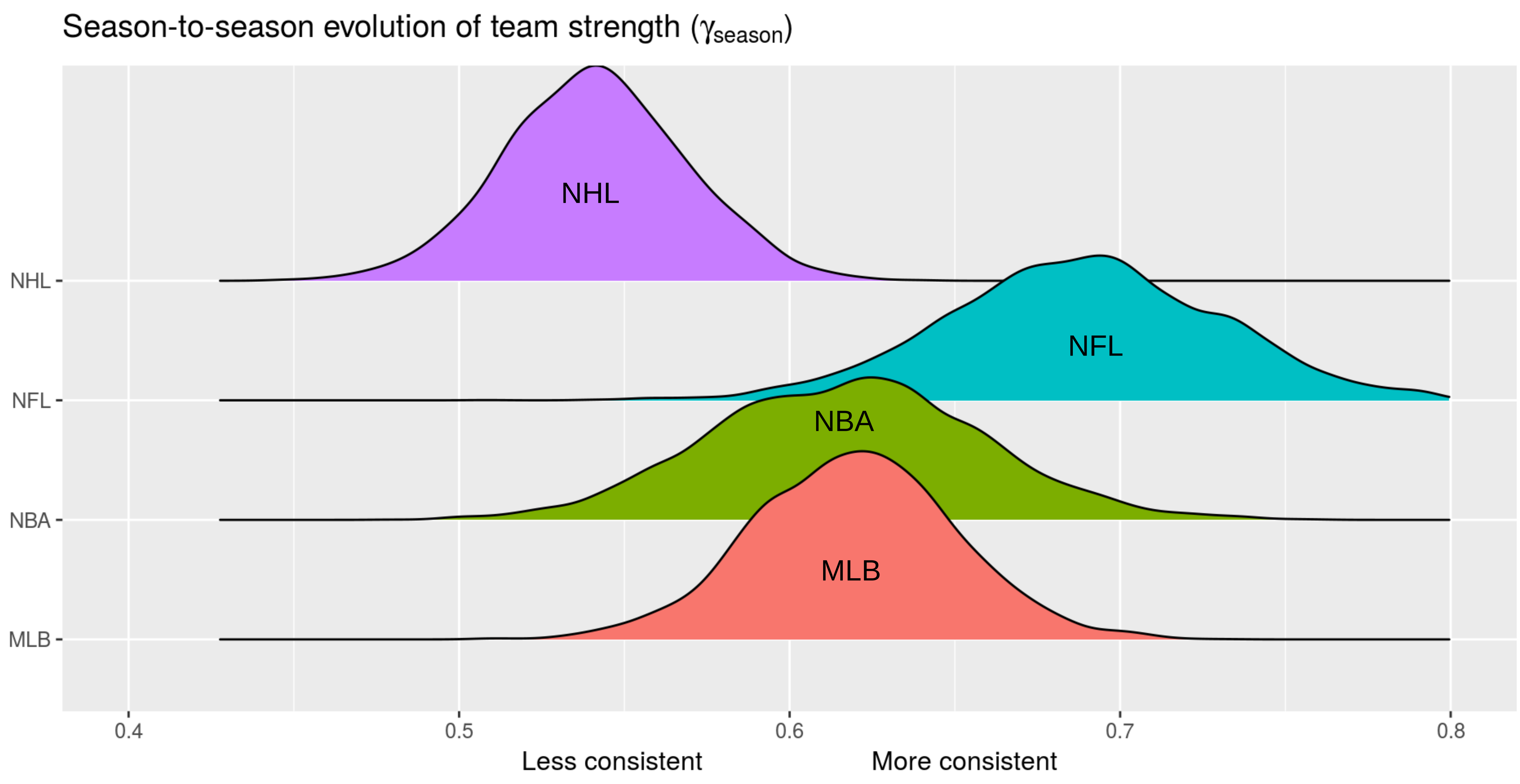

Team Strengths, 2015 MLB
Team strength (log odds)
0.2
0.0
-0.2
-0.4
0
10
20
30
Week of season
Toronto Blue Jays
Texas Rangers

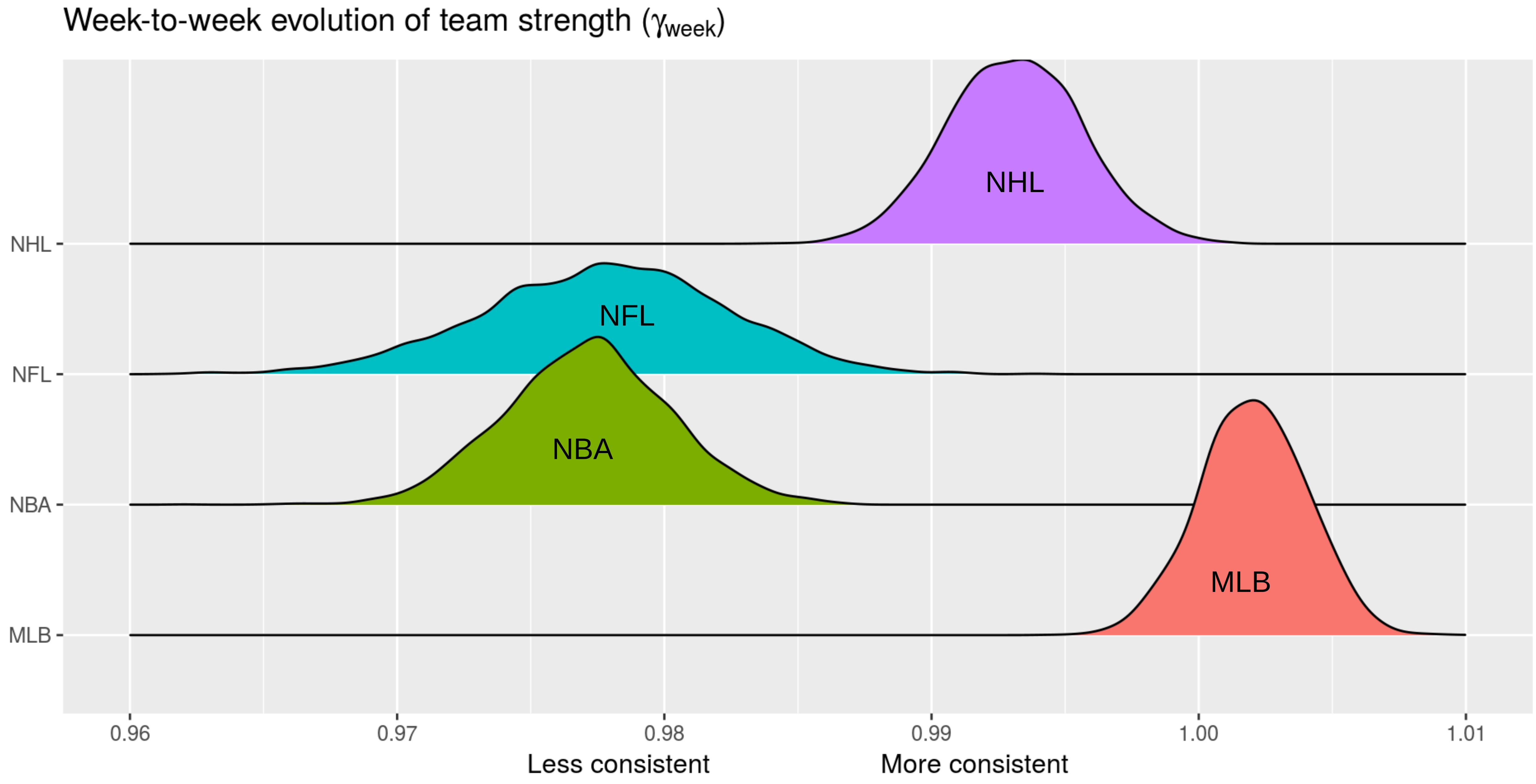


Posterior distribution of team strengths
End of 2015 regular season
7.5
5.0
2.5
0.0
-0.2
0.0
0.2
0.4
Team Strength (log odds scale)
Texas Rangers
Toronto Blue Jays

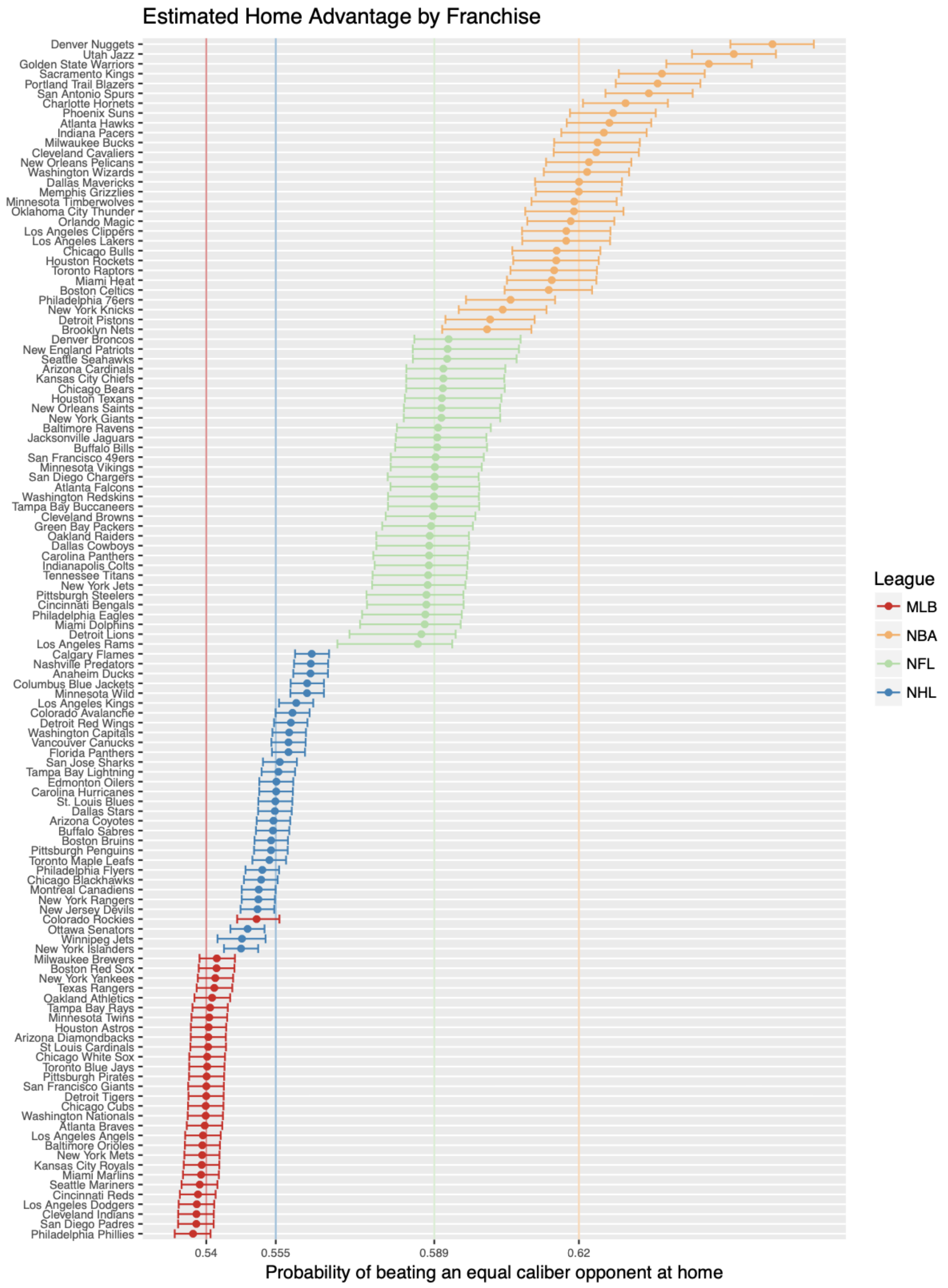


Team strengths across 4 sports leagues
MLB
NBA
NFL
NHL
Team Strength (log-odds scale)
2
1
0
-1
-2
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Season

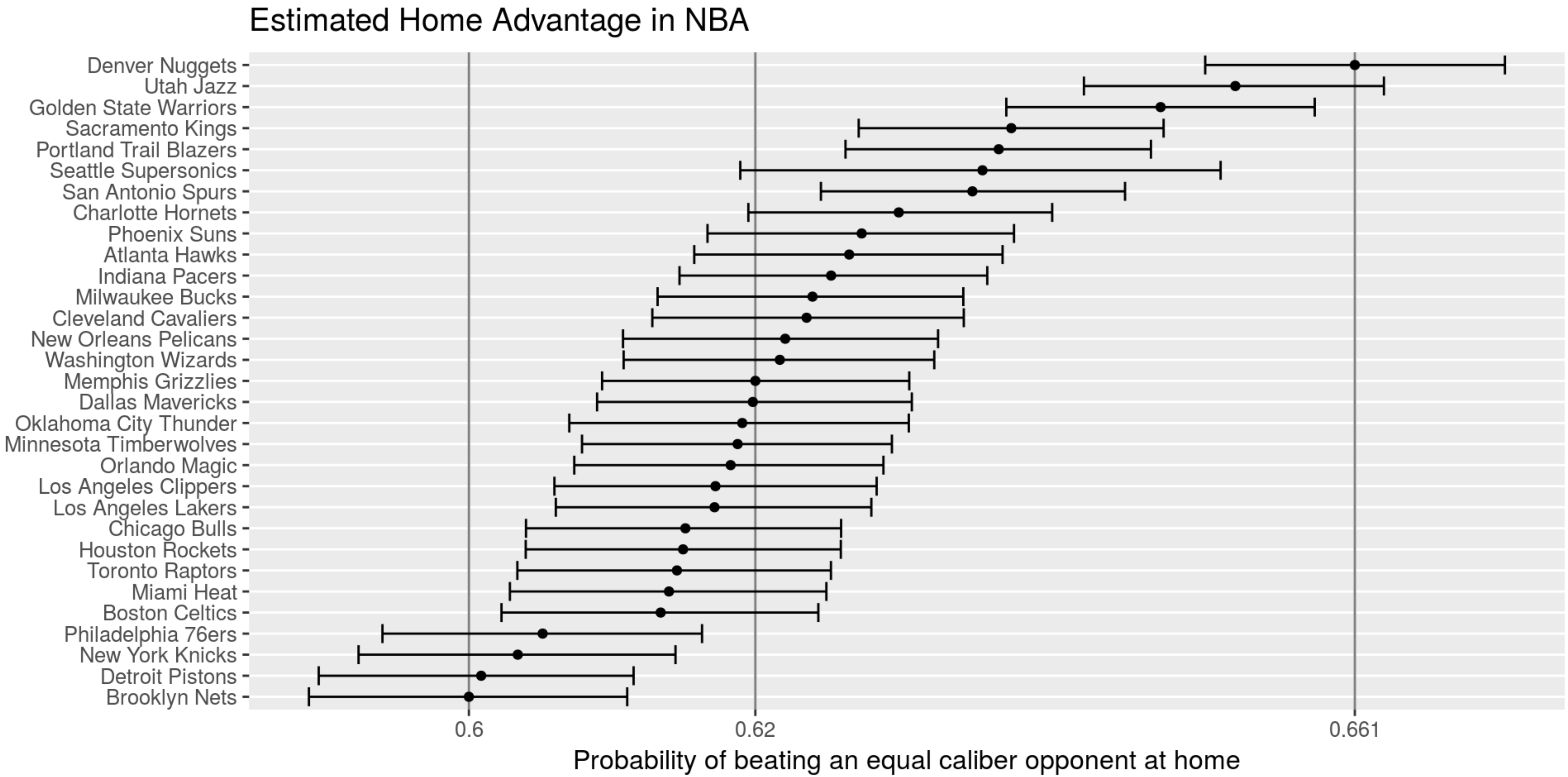

Season-to-season evolution of team strength (γ_{season})
NHL
NFL
NBA
MLB
0.4
0.5
0.6
0.7
0.8
Less consistent
More consistent



Week-to-week evolution of team strength (γ_{week})
NHL
NFL
NBA
MLB
NHL
NFL
NBA
MLB
0.96
0.97
0.98
0.99
1.00
1.01
Less consistent
More consistent



Estimated Home Advantage by Franchise
Denver Nuggets
Utah Jazz
Golden State Warriors
Sacramento Kings
Portland Trail Blazers
San Antonio Spurs
Charlotte Hornets
Phoenix Suns
Atlanta Hawks
Indiana Pacers
Milwaukee Bucks
Cleveland Cavaliers
New Orleans Pelicans
Washington Wizards
Dallas Mavericks
Memphis Grizzlies
Minnesota Timberwolves
Oklahoma City Thunder
Orlando Magic
Los Angeles Clippers
Los Angeles Lakers
Chicago Bulls
Houston Rockets
Toronto Raptors
Miami Heat
Boston Celtics
Philadelphia 76ers
New York Knicks
Detroit Pistons
Brooklyn Nets
Denver Broncos
New England Patriots
Seattle Seahawks
Arizona Cardinals
Kansas City Chiefs
Chicago Bears
Houston Texans
New Orleans Saints
New York Giants
Baltimore Ravens
Jacksonville Jaguars
Buffalo Bills
San Francisco 49ers
Minnesota Vikings
San Diego Chargers
Atlanta Falcons
Washington Redskins
Tampa Bay Buccaneers
Cleveland Browns
Green Bay Packers
Oakland Raiders
Dallas Cowboys
Carolina Panthers
Indianapolis Colts
Tennessee Titans
New York Jets
Pittsburgh Steelers
Cincinnati Bengals
Philadelphia Eagles
Miami Dolphins
Detroit Lions
Los Angeles Rams
Calgary Flames
Nashville Predators
Anaheim Ducks
Columbus Blue Jackets
Minnesota Wild
Los Angeles Kings
Colorado Avalanche
Detroit Red Wings
Washington Capitals
Vancouver Canucks
Florida Panthers
San Jose Sharks
Tampa Bay Lightning
Edmonton Oilers
Carolina Hurricanes
St. Louis Blues
Dallas Stars
Arizona Coyotes
Buffalo Sabres
Boston Bruins
Pittsburgh Penguins
Toronto Maple Leafs
Philadelphia Flyers
Chicago Blackhawks
Montreal Canadiens
New York Rangers
New Jersey Devils
Colorado Rockies
Ottawa Senators
Winnipeg Jets
New York Islanders
Milwaukee Brewers
Boston Red Sox
New York Yankees
Texas Rangers
Oakland Athletics
Tampa Bay Rays
Minnesota Twins
Houston Astros
Arizona Diamondbacks
St Louis Cardinals
Chicago White Sox
Toronto Blue Jays
Pittsburgh Pirates
San Francisco Giants
Detroit Tigers
Chicago Cubs
Washington Nationals
Atlanta Braves
Los Angeles Angels
Baltimore Orioles
New York Mets
Kansas City Royals
Miami Marlins
Seattle Mariners
Cincinnati Reds
Los Angeles Dodgers
Cleveland Indians
San Diego Padres
Philadelphia Phillies
0.54
0.555
0.589
0.62
Probability of beating an equal caliber opponent at home
League
MLB
NBA
NFL
NHL


Estimated Home Advantage in NBA
Denver Nuggets
Utah Jazz
Golden State Warriors
Sacramento Kings
Portland Trail Blazers
Seattle Supersonics
San Antonio Spurs
Charlotte Hornets
Phoenix Suns
Atlanta Hawks
Indiana Pacers
Milwaukee Bucks
Cleveland Cavaliers
New Orleans Pelicans
Washington Wizards
Memphis Grizzlies
Dallas Mavericks
Oklahoma City Thunder
Minnesota Timberwolves
Orlando Magic
Los Angeles Clippers
Los Angeles Lakers
Chicago Bulls
Houston Rockets
Toronto Raptors
Miami Heat
Boston Celtics
Philadelphia 76ers
New York Knicks
Detroit Pistons
Brooklyn Nets
0.6
0.62
0.661
Probability of beating an equal caliber opponent at home

Unpredictability at a fixed point in time

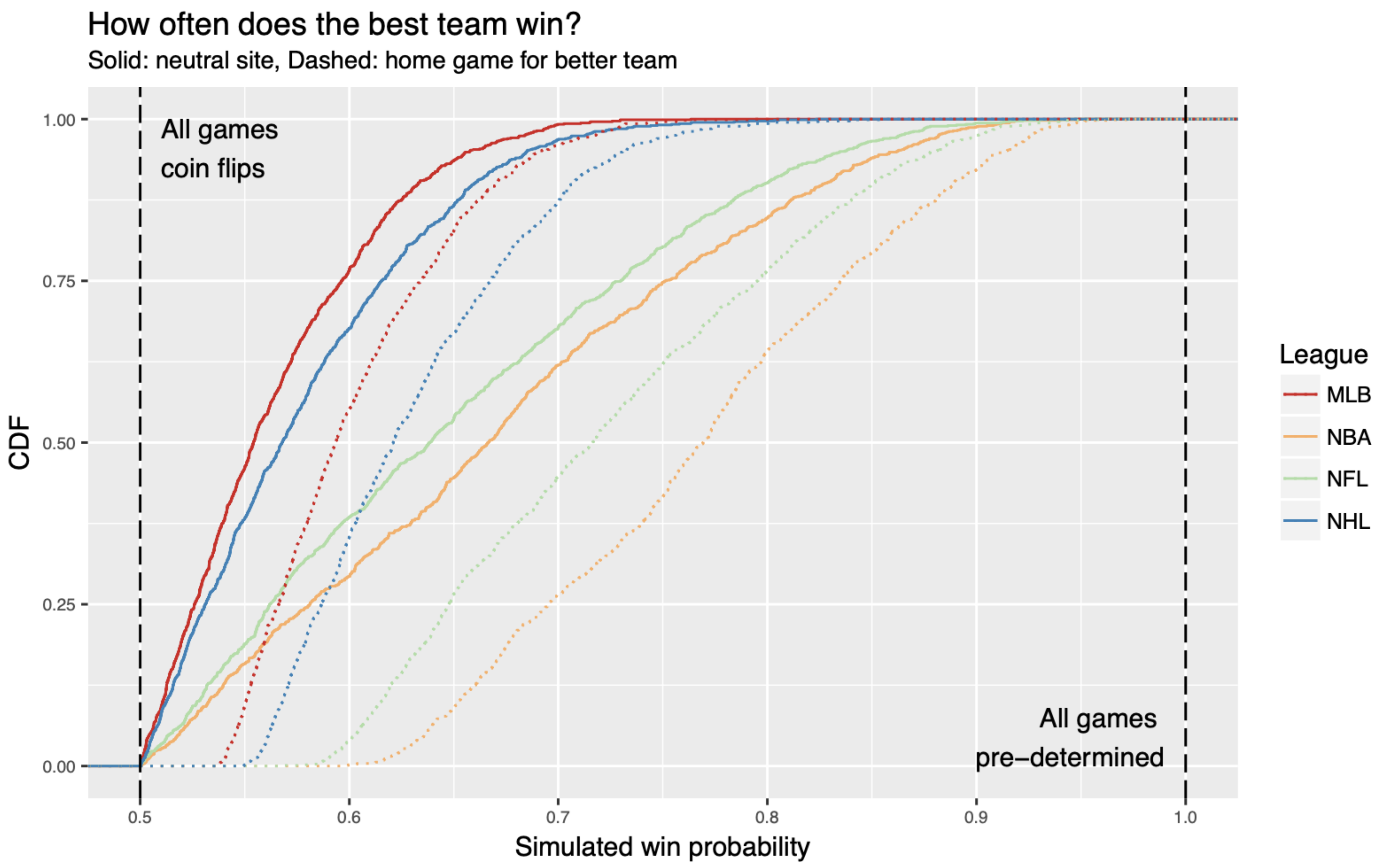
How often does the best team win?

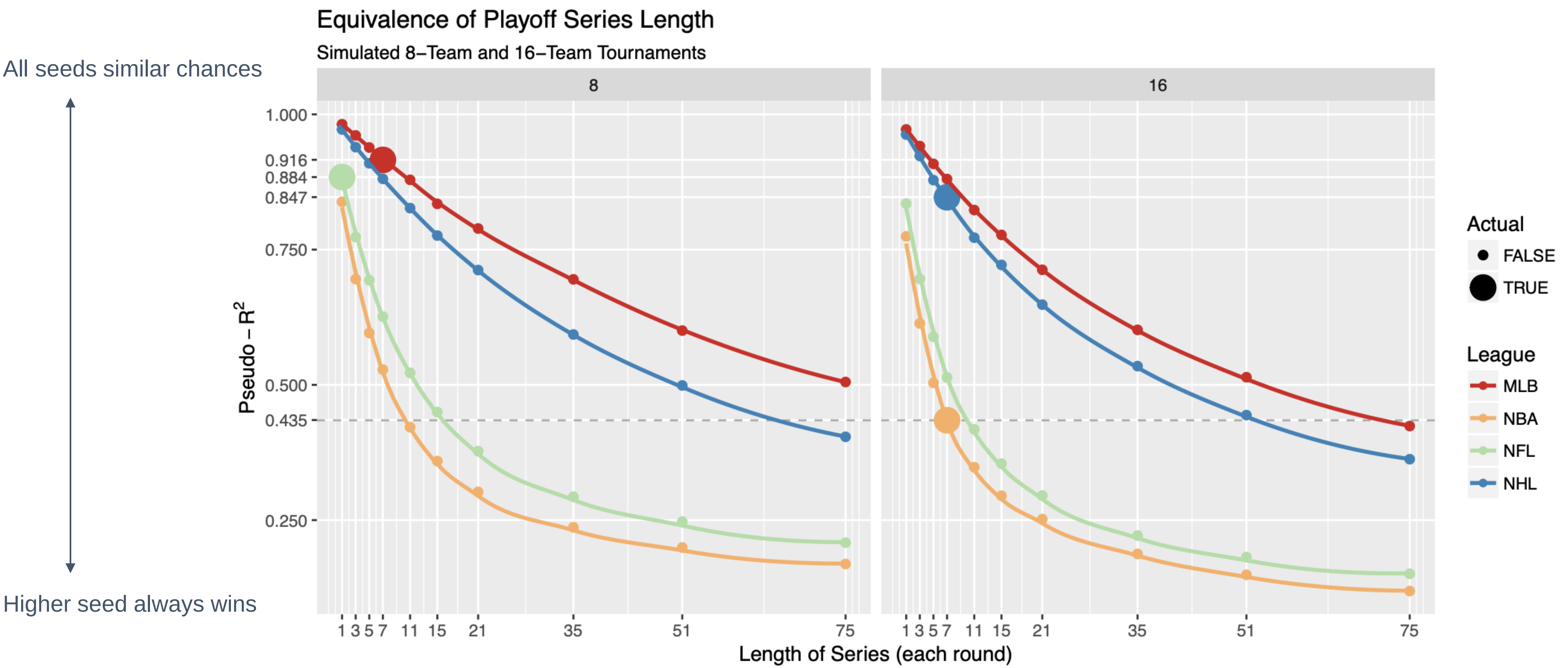
$$\text{RegParity}_q = 2\int_{0.5}^{1} \underbrace{P\left(\tilde{\mathbf{p}}_{\mathbf{q}} \leq x\right)}_{\text{CDF of game-level probability between two randomly chosen teams}} dx$$

$$\text{PostParity}_z = 1 - \frac{\left(E[\mathbf{F}] - f_z 1_z\right)'\left(E[\mathbf{F}] - f_z 1_z\right)}{\sum_{d=1}^{z} \left(\lceil \log_2 d + 1 \rceil - f_z\right)^2}$$

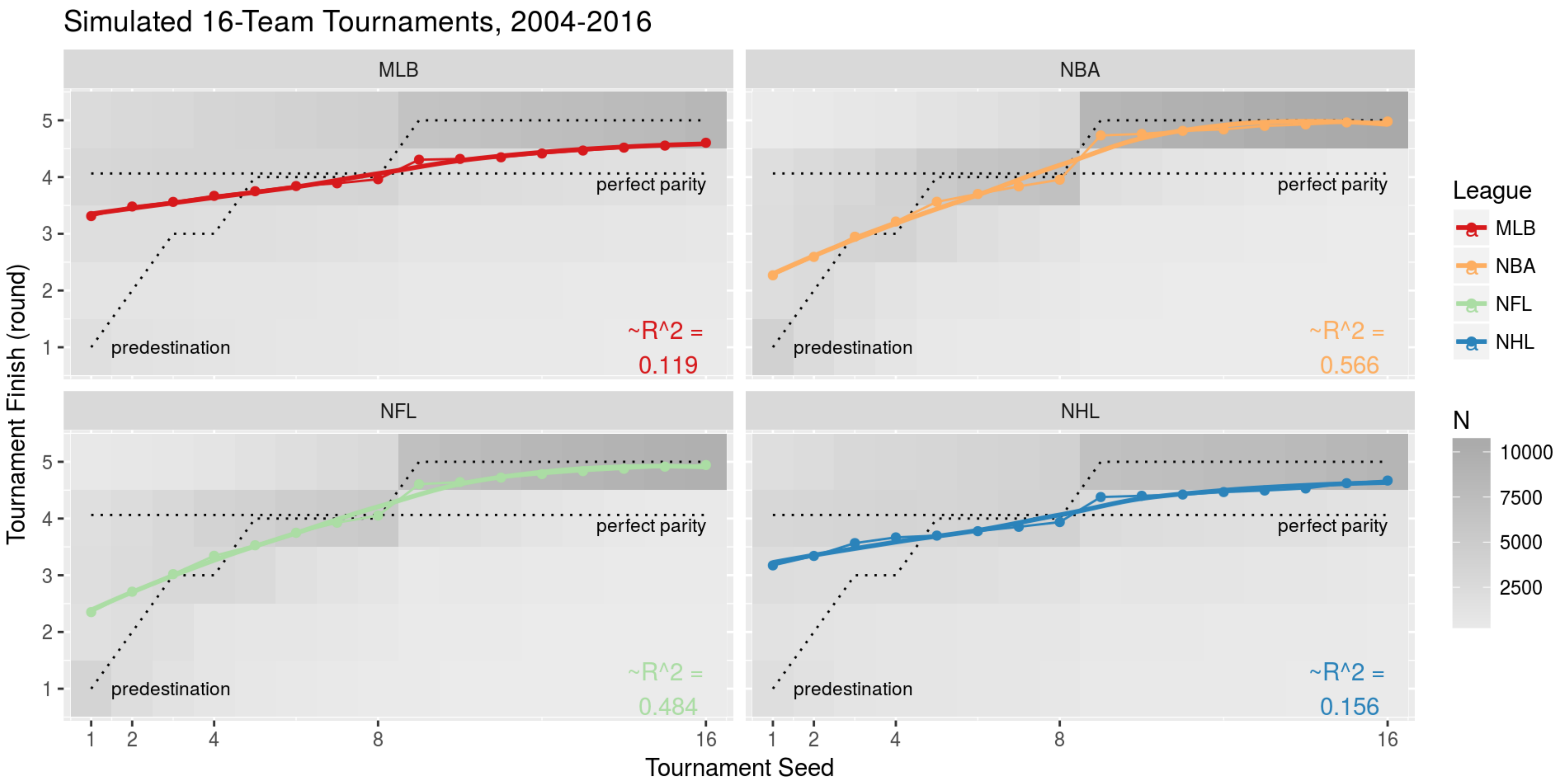
$\mathbf{F}$ z-vector of rounds finished, ordered by seed

f_z Seed-weighted expected finish, assuming teams equal


How often does the best team win?
Solid: neutral site, Dashed: home game for better team
All games
coin flips
All games
pre–determined
CDF
1.00
0.75
0.50
0.25
0.00
0.5
0.6
0.7
0.8
0.9
1.0
Simulated win probability
League
MLB
NBA
NFL
NHL



All seeds similar chances
Higher seed always wins
Equivalence of Playoff Series Length
Simulated 8–Team and 16–Team Tournaments
8
16
Pseudo – R^2
1.000
0.916
0.884
0.847
0.750
0.500
0.435
0.250
1 3 5 7 11 15 21 35 51 75
Length of Series (each round)
Actual
FALSE
TRUE
League
MLB
NBA
NFL
NHL


Simulated 16-Team Tournaments, 2004-2016
MLB
NBA
NFL
NHL
perfect parity
predestination
~R^2 = 0.119
~R^2 = 0.566
~R^2 = 0.484
~R^2 = 0.156
Tournament Finish (round)
Tournament Seed
League
MLB
NBA
NFL
NHL
N
10000
7500
5000
2500
1
2
3
4
5
1
2
4
8
16

Who cares?

GMs need to predict the future

- Buyers or sellers at trade deadline?
- Our team strengths are better at predicting future W-L

GMs need to strategize

- Patriots rest starters in 2015 against Dolphins
- Lost HFA, lost in Denver
- Home advantage matters in 🏈 and 🏀

GMs need a long-term plan

“They have to rethink their whole philosophy”

Mike Milbury on the 2016 Washington Capitals losing to Pittsburgh

- There’s an immense amount of luck involved in hockey. Rethinking your philosophy on a postseason series is ludicrous

Thank you!

- R code: https://github.com/bigfour/competitiveness
- References: https://arxiv.org/abs/1701.05976