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

Gregory Matthews [1], Ben Baumer [2], and Mike Lopez [3]

[1]Loyola University Chicago
[2]Smith College
[3]Skidmore College

October, 2016

Outline

- Basically everyone who is interested in sports is interested in the question "Is team i better than team j?
- How do we measure team strength within a league?
- If we knew the relative team strengths we could view the outcome as a draw from a random variable with some win probability for team i.

- When researchers try to compare sports across leagues, they often use winning percentages.
- However, these are problematic.
- More sophisiticated techniques have focused on a specific sports.
- Hard to compare across sports because of sport specific nuances.
- What are the inherent differences in the dispersion and evolution of team strength across sports?

- Instead of estimating win probabilities we work backwards.
- We assume that betting markets offer unbiased and low-variance estimates of true probabilities.

sport	games	num_teams	home_wp	N_bets	mean_home_p	N_results	coverage
mlb	21854	30	0.542	21854	0.549	24299.000	0.899
nba	10781	30	0.598	10781	0.619	12059.000	0.894
nfl	2294	32	0.567	2294	0.590	2560.000	0.896
nhl	10535	30	0.551	10535	0.566	10564.000	0.997

Table: Summary of cross-sport data. Note that we have near total coverage (betting odds for every game) across all four major sports during the 2005–2014 regular seasons.

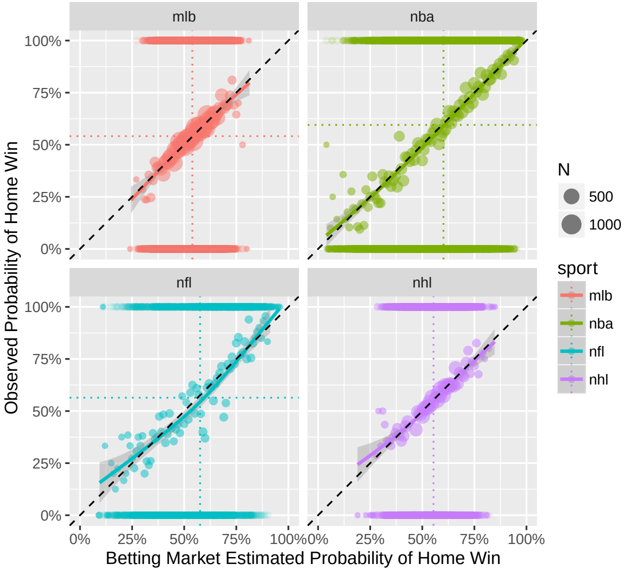

Figure: Accuracy of probabilities implied by betting markets.

For sport q, in season s, in week k:

$$logit(p_{(q,s,k)}) \sim N(\theta_{(\mathbf{q},\mathbf{s},\mathbf{k})}\mathbf{X}_{q,s,k} + \alpha_{q_0}\mathbf{J}_{g_{q,s,k}} + \alpha_q \mathbf{Z}_{q,s,k}, \tau^2_{q,game}\mathbf{I}_{g_{(q,s,k)}}),$$

$$\theta_{(q,s+1,1)} | \gamma_{q,seas}, \theta_{\mathbf{q},\mathbf{s},\mathbf{g}_{\mathbf{q},s,.}}, \tau^2_{q,seas}, \sim N(\gamma_{q,seas}\theta_{(q,s,g_{q,s,.})}, (\tau^2_{q,seas})I_{t_q})$$

and

$$\theta_{(q,s,k+1)} | \gamma_{q,week}, \theta_{\mathbf{q},\mathbf{s},\mathbf{k}}, \tau^2_{q,week}, \sim N(\gamma_{q,week}\theta_{(q,s,k)}, (\tau^2_{q,week})I_{t_q})$$