A Time-Varying Bradley Terry Ranking Model

Heejong Bong, Wanshan Li, Shamindra Shrotriya

CMSAC: November 2nd, 2019

CMU Dept. of Statistics & Data Science

We seek principled approaches to global ranking

Global ranking of objects is fundamental problem in daily life

IDEAS Economic literature Authors Institutions Rankings Help/FAQ

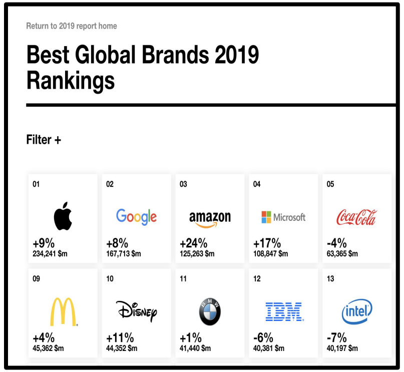
IDEAS/RePEc Aggregate Rankings for Journals

Explanations Ranking More

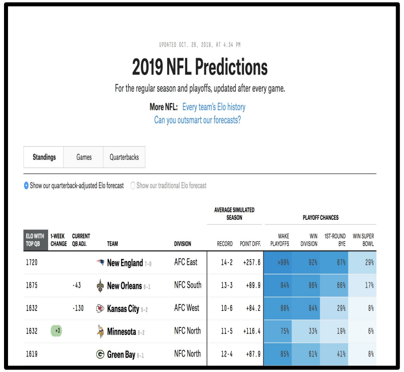
The rankings

Rank	Journals	Score	Items Listed	Simple IF	Recursive IF	Discounted IF	Recursive Discounted IF
1	The Quarterly Journal of Economics, Oxford University Press	2.25	5293	1	2	1	1
2	Journal of Political Economy, University of Chicago Press	2.78	5717	3	1	5	2
3	American Economic Review, American Economic Association	3.03	10077	9	14	13	18
4	Econometrica, Econometric Society (also covers Econometrica, Econometric Society)	3.52	3729	4	3	3	3
5	Journal of Economic Literature, American Economic Association	5.92	951	2	5	2	5
6	Journal of Financial Economics, Elsevier	6.25	2882	6	10	7	19

Journals



Brands



Sports

Ranking is a fundamentally **unsupervised** statistical problem

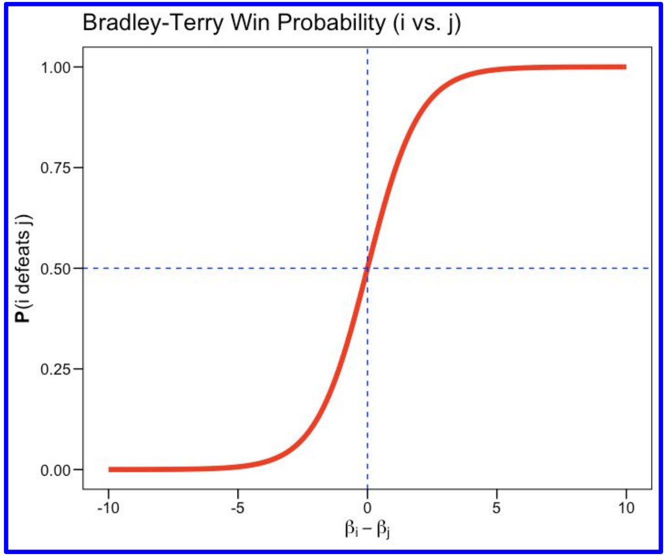
A principled statistical approach is provided by the **Bradley-Terry (BT) Model 52'**

BT-model obtains global rankings using pairwise data

Consider N distinct teams, each with a positive "strength" score, $\beta_i, \forall i \in [N]$

Assumption 1:

$$\mathbb{P}(i \text{ defeats } j) = \text{logistic}(\beta_i - \beta_j) \iff$$



Assumption 2: Matches are independent

Seek principled approach to dynamic global ranking

Typically observe paired comparisons over multiple (discrete) time periods

How to model the Bradley-Terry global rankings **over time**?

Prior Work: *Cattelan et. al. 13', Lopez et. al. 18', Glickman et. al. 98',*
Grossglauser et. al. 19'

Typically require strong domain knowledge and parametric assumptions

Goal: Extend BT-model dynamically with **minimal additional assumptions**

We propose a convex time-varying BT-model

		Negative log-likelihood	Smoothness penalty (convex)	Additive constraint
(Static) BT-model	$\min\limits_{\boldsymbol{\beta}}$	$-\ell(\boldsymbol{\beta})$		$, \text{s.t.} \sum_{i=1}^{N} \boldsymbol{\beta}_i = 0$
Time-varying BT-model	$\min\limits_{\{\boldsymbol{\beta}^{(t)}\}_{t\in[T]}}$	$-\sum_{t=1}^{T} \ell_t(\boldsymbol{\beta}^{(t)})$	$+\lambda \sum_{t=1}^{T-1} \lVert \boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)} \rVert$	$, \text{s.t.} \sum_{i=1}^{N} \boldsymbol{\beta}_i^{(1)} = 0$

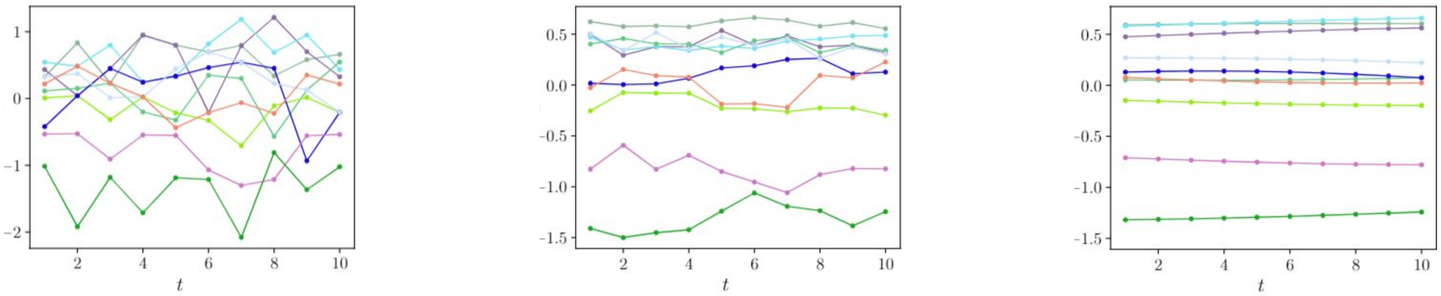
No specific distribution on parameters, use of convex opt. methods

Hyperparameter λ controls how smooth $\boldsymbol{\beta}^{(t)}$ change over time

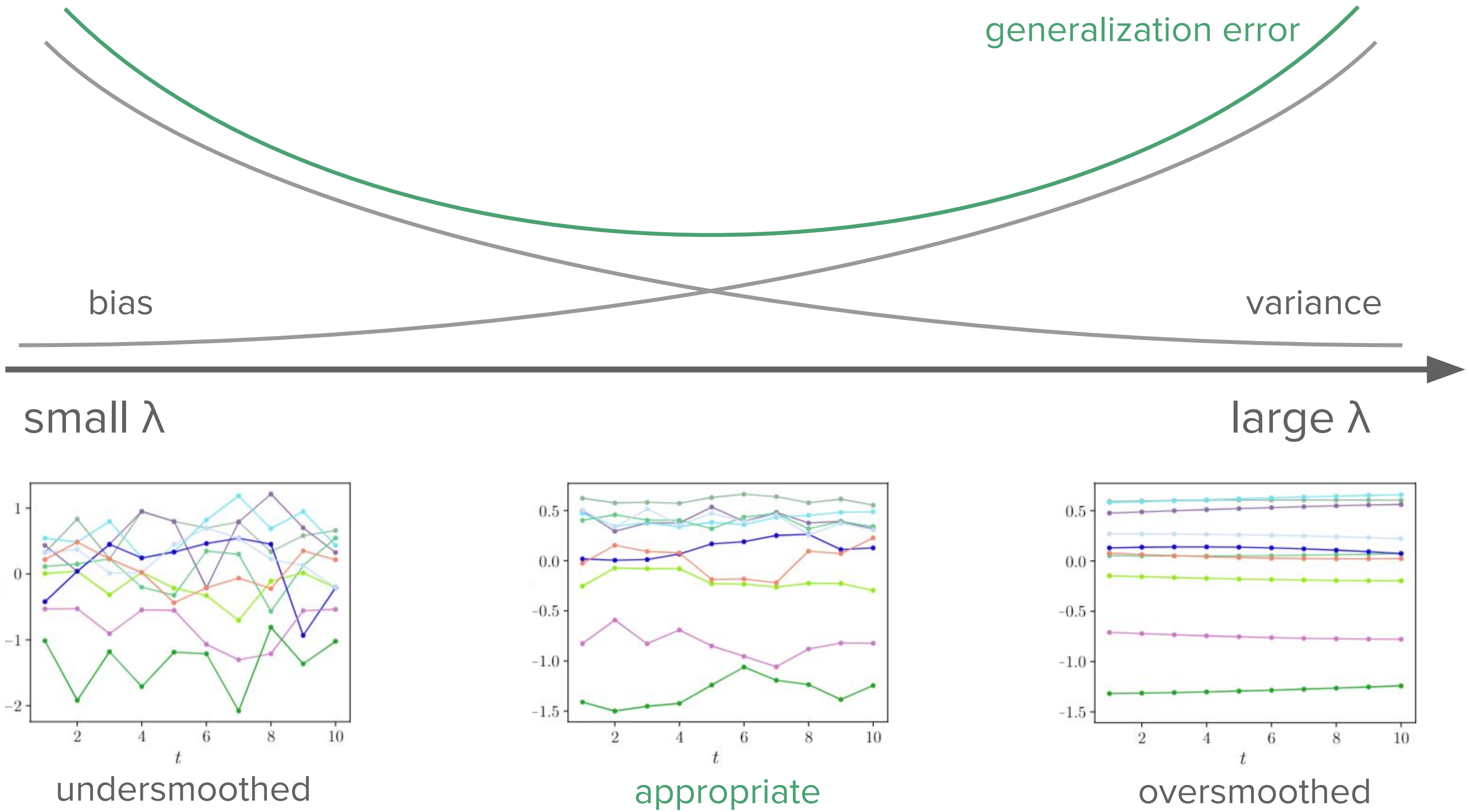
Negative log-likelihood | **Smoothness penalty (convex)**

Risk objective: $$-\sum_{t=1}^{T} \ell_t(\boldsymbol{\beta}^{(t)}) \quad + \lambda \sum_{t=1}^{T-1} \|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\|$$

small λ → large λ



Bias-variance trade-off by λ improves prediction



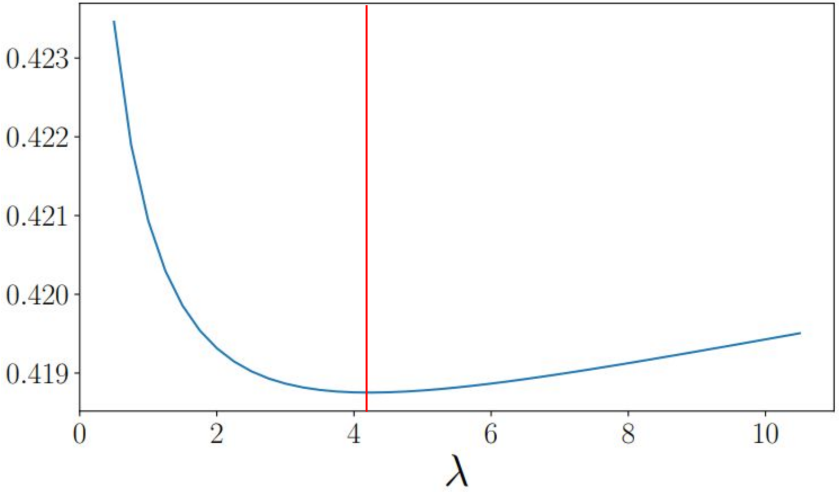
We suggest to tune λ via CV

Cross-validation

Estimate the generalization error for each λ by sample splitting (e.g., LOOCV, k-fold CV, etc.).

⇨ Choose λ with the smallest error.

- Data-driven
- Moderate computational cost
  (We suggest ways to reduce the cost)

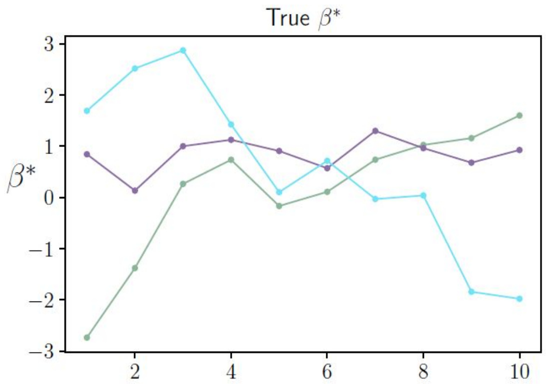


Simulation: a simple case

3 teams, 10 rounds/seasons

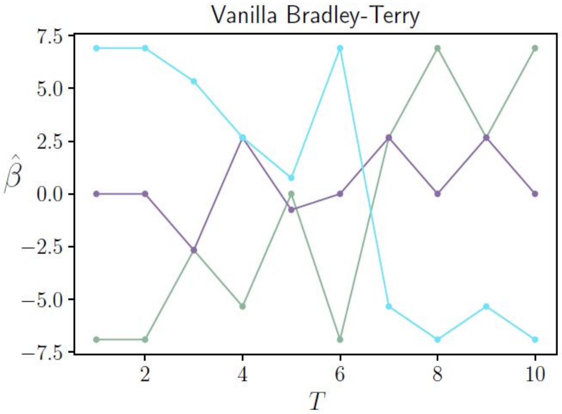
Team ability changes

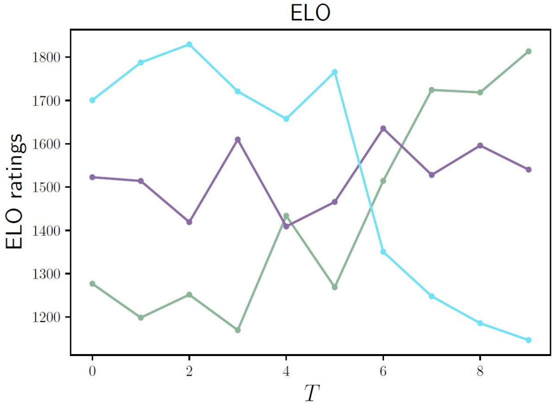
Ranking changes

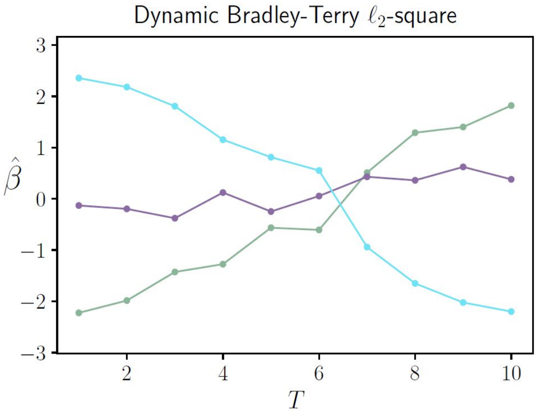


Smoother estimates are better!

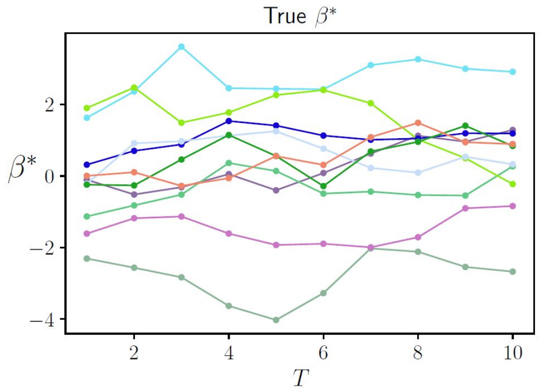
- Interpretability
- Handle small/moderate sample size

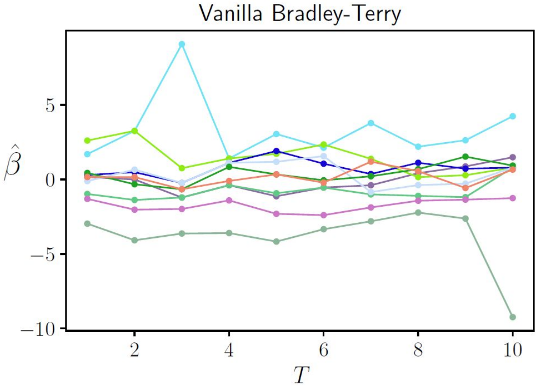


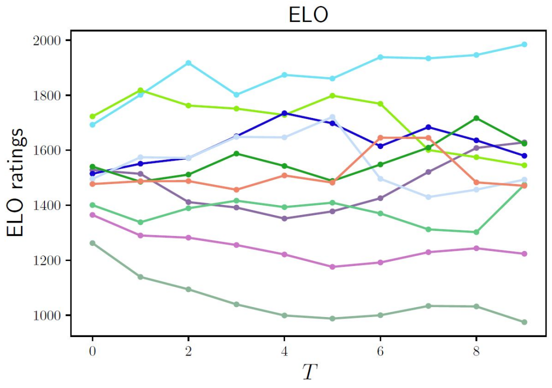


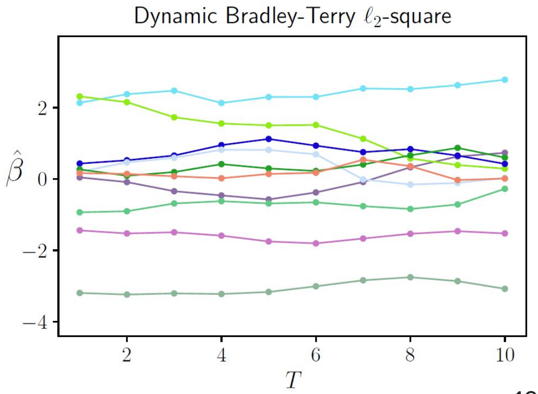


Simulation: comparison of different methods









Prediction risk:0.54

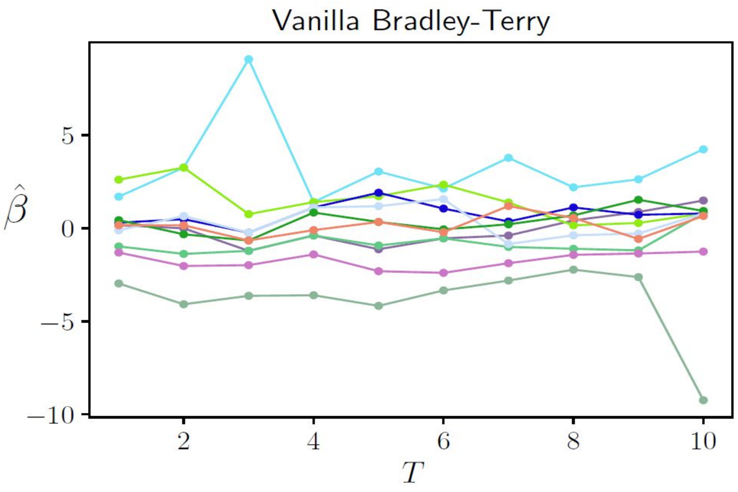
Prediction risk:0.56

Prediction risk:0.51

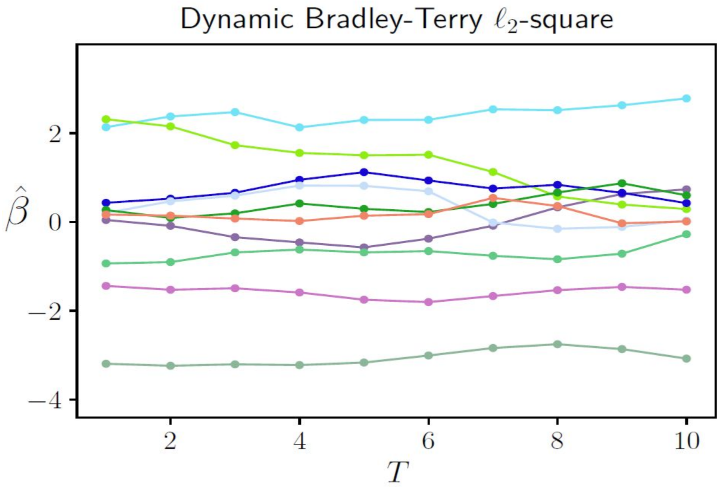
Our model ensures stable AND accurate rankings

Our model performs well both

- Qualitatively: smooth parameter paths, stable rankings, easy to interpret
- Quantitatively: recover true rankings, predict win/loss



Prediction risk:0.54

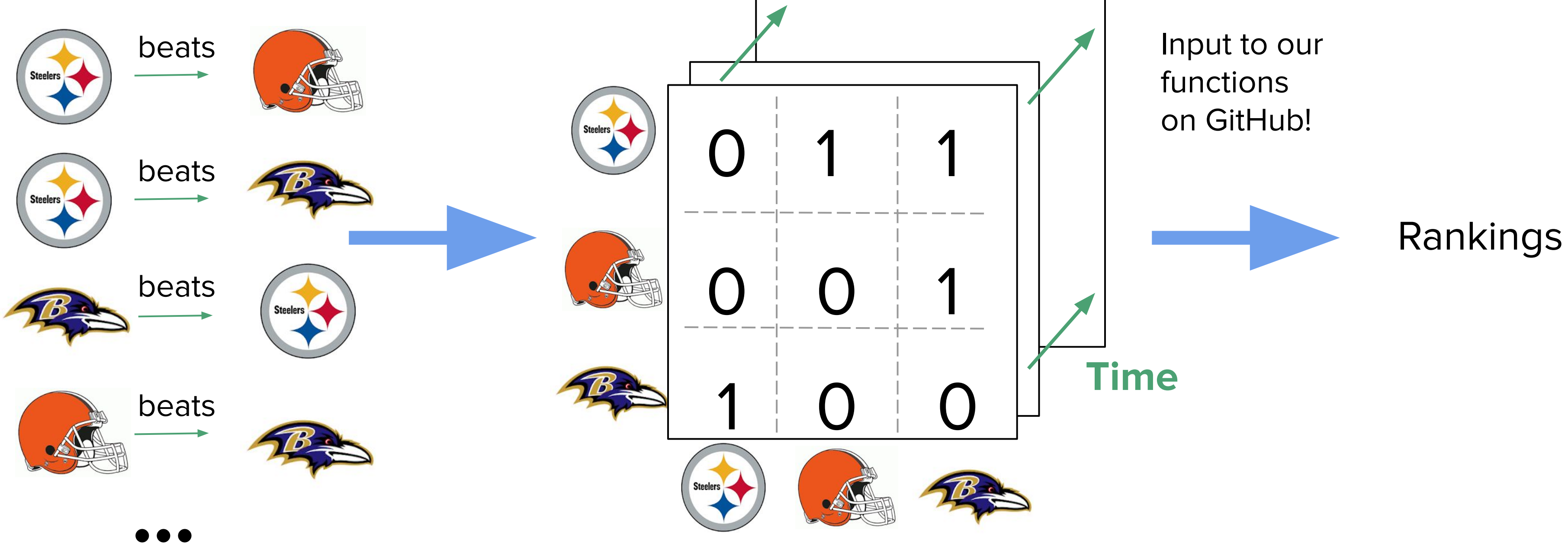


Prediction risk:0.51

Well... How does it work on real data?

Pairwise matches

Temporal array



nflscrapR

We also test our model against NFL-ELO rankings

rank	2011		2012		2013		2014		2015	
	ELO	BT	ELO	BT	ELO	BT	ELO	BT	ELO	BT
1	GB	GB	NE	DEN	SEA	SF	SEA	SEA	SEA	CAR
2	NE	NO	DEN	NE	SF	CAR	NE	DEN	CAR	ARI
3	NO	NE	GB	SEA	NE	SEA	DEN	GB	ARI	KC
4	PIT	SF	SF	MIN	DEN	ARI	GB	NE	KC	SEA
5	BAL	PIT	ATL	SF	CAR	NE	DAL	DAL	DEN	MIN
6	SF	BAL	SEA	GB	CIN	DEN	PIT	PIT	NE	DEN
7	ATL	DET	NYG	IND	NO	NO	BAL	IND	PIT	CIN
8	PHI	ATL	CIN	HOU	ARI	CIN	IND	ARI	CIN	PIT
9	SD	PHI	BAL	WAS	IND	IND	ARI	BUF	GB	GB
10	HOU	SD	HOU	CHI	SD	SD	CIN	DET	MIN	DET
Av. Diff.	2.6		3.2		2.6		1.9		2.8	

Table 1: Bradley-Terry vs. ELO NFL top 10 rankings. Blue: perfect match, yellow: top 10 match

Summary

We propose a time-varying extension of the BT model with **minimal assumptions**

Bias-variance trade-off with smoothness penalty achieves performance gain

Performance gain is confirmed in **simulated settings**

Our upcoming recent work builds on this approach to obtain theoretical results

Use it as a **minimalist dynamic ranking benchmark** for other (BT) ranking models!

Reproducibility: https://bit.ly/337r5qh

Questions?

Reproducibility: https://bit.ly/337r5qh

Bradley, Ralph Allan, and Milton E. Terry. "Rank analysis of incomplete block designs: I. The method of paired comparisons." *Biometrika* 39.3/4 (1952): 324-345.

Cattelan, Manuela, Cristiano Varin, and David Firth. "Dynamic Bradley–Terry modelling of sports tournaments." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62.1 (2013): 135-150.

Horowitz, M., R. Yurko, and S. L. Ventura. "nflscrapR: Compiling the NFL play-by-play API for easy use in R." *URL https://github. com/maksimhorowitz/nflscrapR, r package version* 1.0 (2017).

Glickman, Mark E. "Dynamic paired comparison models with stochastic variances." *Journal of Applied Statistics* 28.6 (2001): 673-689.

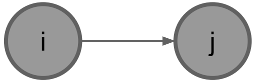
Lopez, Michael J., Gregory J. Matthews, and Benjamin S. Baumer. "How often does the best team win? A unified approach to understanding randomness in North American sport." *The Annals of Applied Statistics* 12.4 (2018): 2483-2516.

Maystre, Lucas, Victor Kristof, and Matthias Grossglauser. "Pairwise Comparisons with Flexible Time-Dynamics." *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019.

Uniqueness and existence of the solution requires a weak condition for data

Ford, Jr (1957): BT-model has a unique maximum likelihood parameter *iff*

strongly connected

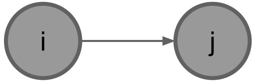
where 

implies "i defeated j at least once".

Uniqueness and existence of the solution requires a weak condition for data

We extend this condition to the time-varying case:

strongly connected

where 

implies “i defeated j at least once **throughout entire time**”.

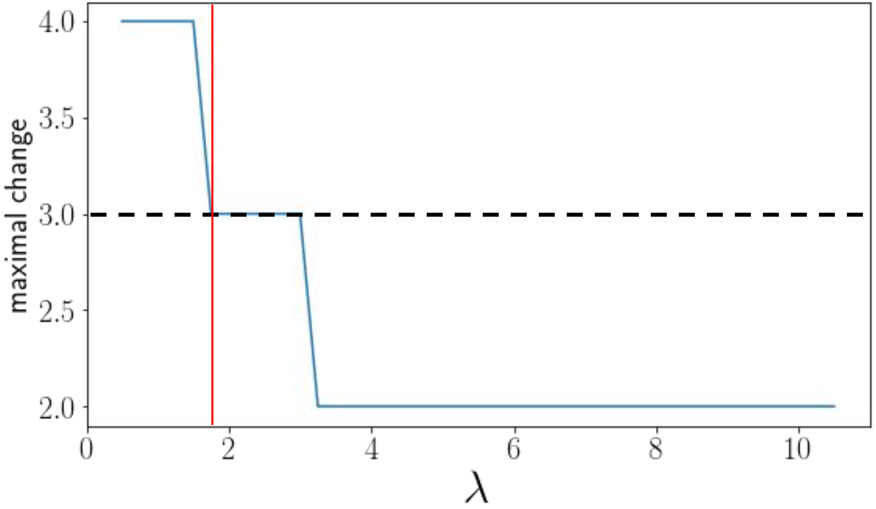
Supp: Known limitations of the BT-model?

Batch models - need to re-fit after each new time point

Unweighted strength parameters

Assumes independence in matches played (can be relaxed)

Supp: We suggest to tune λ via CV/heuristics



Heuristic

Use domain knowledge in smoothness of ranking changes to tune λ.

➩ Choose λ to control maximum global ranking movements over all time periods

- Human-judgement
- Low computational cost

Additional Questions:

Multiple team competing at the same time?
Handling Ties?
Why choose this model over ELO?
What are the limitations of your model?
What about changing history?

There is a need to extend BT-model dynamically

Typically observe paired comparisons over multiple (discrete) time periods

How to model the Bradley-Terry global rankings **over time**?

Moving Average
[Cattelan et. al. 13’]

State Space/Bayesian
[Lopez et. al. 18’]
[Glickman et. al. 98’]
[Grossglauser et. al. 19’]

Goal: Extend BT-model dynamically with **minimal additional assumptions**

Supp:

Reflect the reviews - serious comparison of methods (ELO for example) (main)

Cattelan paper comparison

NASCAR simulation (main)

WL: Put one or 2 examples up front + then BT method

Stress the use of LOOCV as a predictive benchmarking comparison tool

SS: Add reproducibility links to github

SS: How do we “borrow” information over time exactly?

SS: Can we detail the fitting process visually?