RUSH: Ratings Using Score Histories

Todd Graves[1], Kary Myers[1], Earl Lawrence[1], Shane Reese[2]

[1]Statistical Sciences Group
Los Alamos National Laboratory

[2]Department of Statistics
Brigham Young University

New England Symposium on Statistics in Sports 2011

Pre-1998

- “National champion”(s) named by Associated Press poll of sportswriters and poll of coaches
- No effort to match up top teams in bowl games
  - Miami or Washington, 1991
  - Nebraska or Penn State, 1994
  - Michigan or Nebraska, 1997
- Much agitation about college football being the only major sport without a playoff system to determine the national champion

Formation of the Bowl Championship Series (BCS)

- In 1998, two teams selected to play in national championship game using a formula
- Formula elements:
  - poll of sportswriters
  - poll of coaches
  - “computer rankings”: statistical rating systems using that season’s game results as data
- Much agitation about nerds who know nothing about football affecting the postseason, and how there is still no playoff system

Discontent with BCS

- Annual complaints about the BCS (especially “computers”)
- Generally, the number of deserving teams is not 2.
  - 2003 NCG: LSU vs. Oklahoma.
    - Sportswriters’ poll had USC as #1.
- Hal Stern, “In Favor of a Quantitative Boycott of the Bowl Championship Series”, Journal of Quantitative Analysis in Sports, 2006
- People want a playoff already!!!

Goals for a new rating system

- Help select teams for a playoff system involving 2 or more teams
- Compare undefeated teams from weaker conferences with teams from stronger conferences
- Use the sparse information efficiently
- Do not reward running up the score

Methods that ignore the score

- BCS requires its official ratings to do this, for sportsmanship reasons
- Example: Bradley-Terry

$$\Pr\{\text{Team } i \text{ beats Team } j\} = \frac{a_i}{a_i + a_j}$$

- Blowouts are treated the same as nailbiters
- Can overvalue undefeated teams with weak schedules

Margin of victory (MOV)

- Example: least squares fit of (visitor score - home score) on dummy variables for visiting team and home team
- Losing by 1 is about the same as winning by 1
- Thought to have contributed to unpopular BCS choices

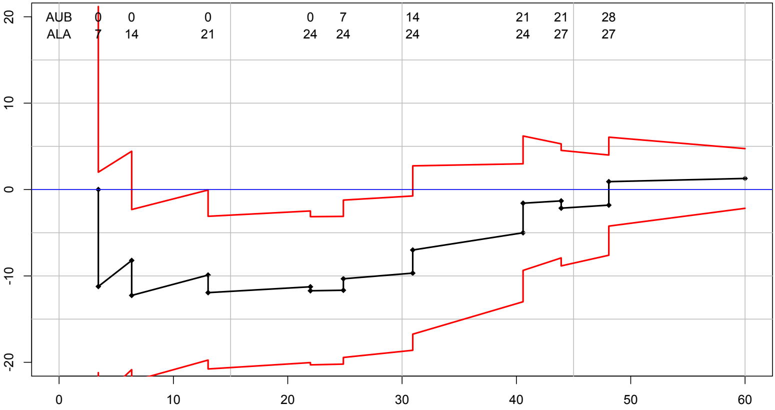
RUSH: Ratings Using Score Histories

- Use the *score process*
  - the score of a game at every point in time.
- Use this to downweight meaningless scoring plays.

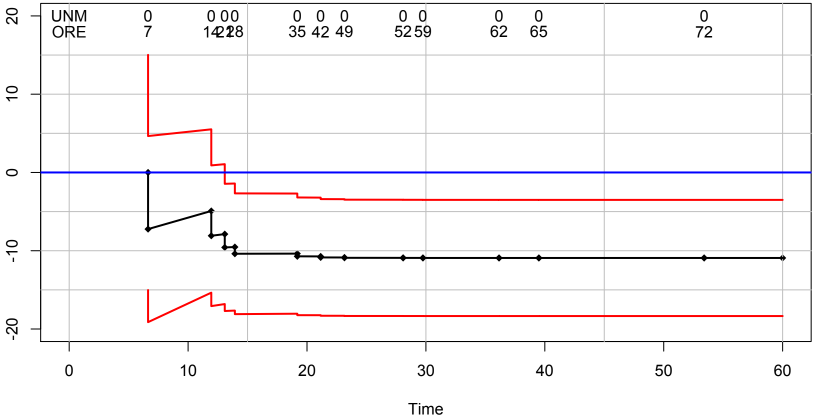
Example data from one game

Week	Visitor	Home	Time	Score Margin
13	AUB	ALA	3.43	-7
13	AUB	ALA	6.35	-14
13	AUB	ALA	13.03	-21
13	AUB	ALA	21.98	-24
13	AUB	ALA	24.87	-17
13	AUB	ALA	30.93	-10
13	AUB	ALA	40.58	-3
13	AUB	ALA	43.92	-6
13	AUB	ALA	48.08	1

Auburn 28, Alabama 27



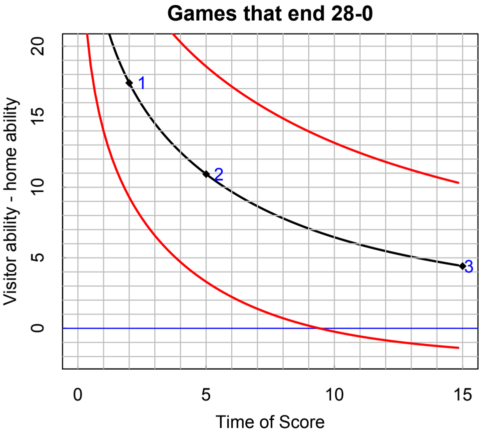
Oregon 72, New Mexico 0



Family of hypothetical 28-0 games

Consider games that the visitor wins 28-0, with 7-point touchdowns at evenly spaced intervals, e.g.

1. 2, 4, 6, 8 minutes
2. 5, 10, 15, 20 minutes
3. 15, 30, 45, 60 minutes



Notation and elements of the model

- θ_j = ability of team j
- η = home field advantage
- λ = expected number of scores per unit time
  - different for each game, equals $\tau_{Visitor} + \tau_{Home}$
  - τ_j = one team's contribution to λ
- $S = \{-8, -7, -6, -4, -3, -2, 2, 3, 4, 6, 7, 8\}$ set of possible scores
- $\pi_s = \pi_{-s}$ fraction of scores of each type

$\pi_{-2} + \pi_2$	$\pi_{-3} + \pi_3$	$\pi_{-4} + \pi_4$	$\pi_{-6} + \pi_6$	$\pi_{-7} + \pi_7$	$\pi_{-8} + \pi_8$
1 in 133	1 in 4	1 in 700	1 in 15	2 in 3	1 in 60

Markov point process

- Let $\alpha = \theta_v - \theta_h - \eta$ be the difference between the abilities of the two teams.
- Let $M(t) = V(t) - H(t)$ be the score margin at time t.

Define

$$f(t, m, s) = \Pr_0\{M(T) > 0 | M(t) = m+s\} - \Pr_0\{M(T) > 0 | M(t) = m\}.$$

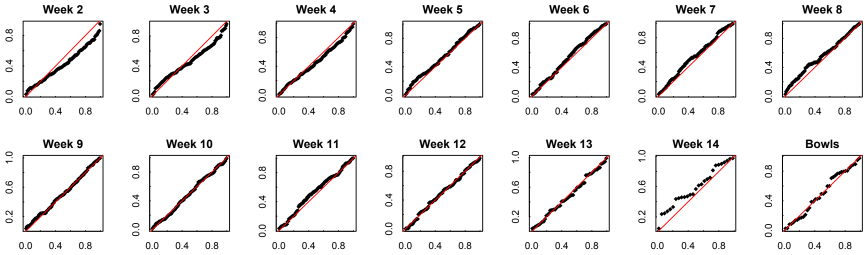
Assume

$$\begin{aligned} \Pr_\alpha\{M(t+\Delta t) = m + s | M(t) = m\} &\propto \lambda \Delta t \, \pi_s \exp\{\alpha f(t, m, s)\} \\ \Pr_\alpha\{M(t+\Delta t) = m | M(t) = m\} &\propto (1 - \lambda \Delta t) \end{aligned}$$

Computing the ratings

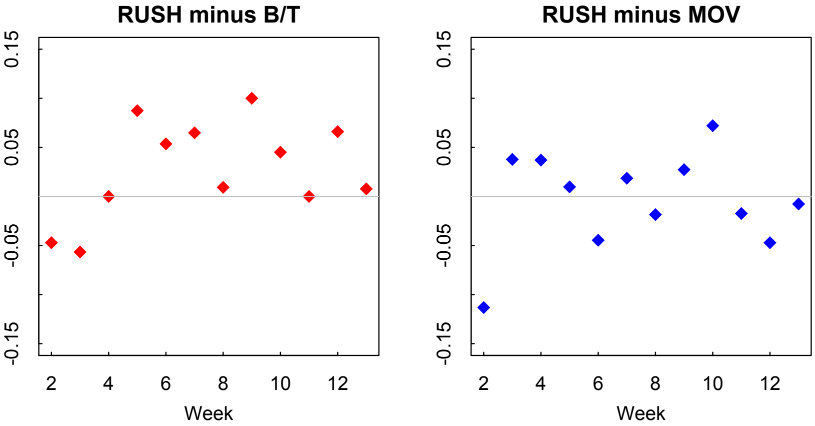
- Not trivial.
- However, with evenly matched teams, scores occur according to a homogeneous Poisson process, with the probabilities of the various types of scores fixed and known
- Many important quantities can be precomputed
- MCMC computations done in YADAS (yadas.lanl.gov)

Calibrated predictions for Visitor - Home score



Compute the quantile in the predictive distribution of each actual result. If the predictions are well-calibrated, each week should be a sample from the Uniform(0,1) distribution.

Comparing predictions to BT and MOV



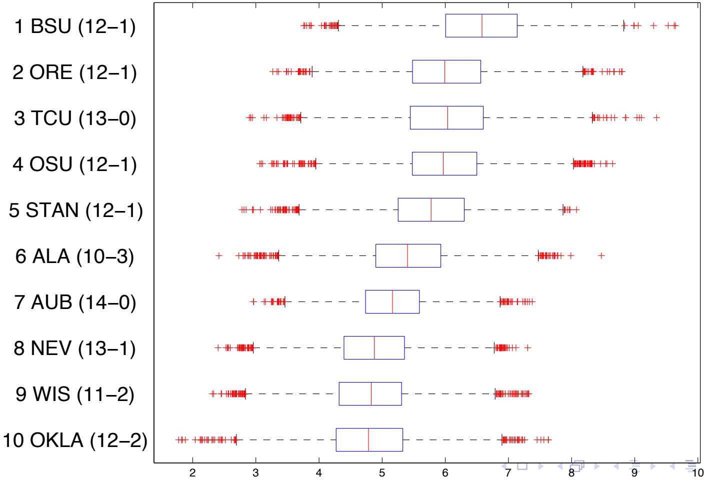
2010 Cast of Characters

- NATIONAL CHAMPIONSHIP GAME:
  - Auburn (14-0): champions of SEC, the strongest league. Defeated Oregon 22-19 in NCG.
  - Oregon (12-1): champions of Pac 10. Many impressive wins.
- DOMINANT IN MEDIUM-STRENGTH CONFERENCES:
  - TCU (13-0): champions of Mountain West.
  - Boise State (12-1), lost only to Nevada (13-1). Many lopsided wins.

2010 RUSH Ratings

	E(rank)	team	record	$E(\theta)$	sd	τ
1	2.62	Boise St	12-1	6.563	0.79	4.21
2	4.31	Oregon	12-1	6.014	0.79	4.87
3	4.37	TCU	13-0	6.031	0.84	4.01
4	4.40	Ohio St	12-1	5.973	0.75	4.21
5	5.29	Stanford	12-1	5.768	0.77	4.38
6	7.03	Alabama	10-3	5.403	0.77	4.11
7	8.15	Auburn	14-0	5.154	0.62	4.76
8	10.36	Nevada	13-1	4.874	0.70	4.60
9	10.95	Wisconsin	11-2	4.807	0.74	4.62
10	11.06	Oklahoma	12-2	4.791	0.76	4.67

Posterior distribution for RUSH ratings



Why is Auburn #7??

- Auburn won games by 1, 3, 3, 3, 3, 7, 8.
- This includes a home overtime win over #35 Clemson and a 3-point win on the game's last play against #63 Kentucky.
- And they won a game by 22 after trailing in the 4th quarter.
- They were a great story and deserving national champions, but they were very lucky: in 100 simulated seasons, they went undefeated in 3 and lost 3 or more in 48.

Why is Boise State #1??

This is the complete set of their *halftime* leads:
6, 34, 14, 38, 29, 41, 21, 21, 38, 20, 17, 22, 13.

- In 100 simulated seasons they went undefeated in 48, lost twice or more in only 10.
- In the 2007-2009 seasons, the top rankings for mid-major teams were #14 BYU, #6 TCU, #3 TCU

Rankings of teams in other rating systems

RUSH	Team	Record	BT	MOV
1	Boise St	12-1	6	1
2	Oregon	12-1	3	2
3	TCU	13-0	2	5
4	Ohio St	12-1	7	6
5	Stanford	12-1	4	3
6	Alabama	10-3	11	4
7	Auburn	14-0	1	7
16	Michigan St	11-2	13	44
17	LSU	11-2	5	14
51	Top FCS Team	*	19	33

*Top FCS Team is Eastern Washington (12-2) in BT, Villanova (9-5) in MOV and RUSH.

Questions?

The resulting likelihood function

$$\lambda^n \left(\prod_{i=1}^{n} \pi_{s_i} \right) \exp \left\{ \alpha \sum_{i=1}^{n} f(t_i, m_{i-1}, s_i) - \lambda \int_0^T \sum_{\sigma \in S} \pi_\sigma \exp\{\alpha f(u, m(u), \sigma)\} du \right\}$$

$$f(t, m, s) = \Pr_0\{M(T) > 0 | M(t) = m + s\} - \Pr_0\{M(T) > 0 | M(t) = m\}.$$

where

- n is the number of scores;
- s_i is the number of points scored on the ith scoring play;
- t_i is the time of the ith score;
- $m_i = \sum_{k=1}^{i} s_i$ is the ith value of the score margin, and $m(t)$ is the margin written as a function of time

Other choices of f

$$\lambda^n \left(\prod_{i=1}^{n} \pi_{s_i} \right) \exp \left\{ \alpha \sum_{i=1}^{n} f(t_i, m_{i-1}, s_i) - \lambda \int_0^T \sum_{\sigma \in S} \pi_\sigma \exp\{\alpha f(u, m(u), \sigma)\} du \right\}$$

- $f(t, m, s) = s$ yields $\sum_{i=1}^{n} f(t_i, m_{i-1}, s_i) = m(T)$, a model for the margin of victory!
- $f(t, m, s) = \text{sign}(m + s) - \text{sign}(m)$ yields $\sum_{i=1}^{n} f(t_i, m_{i-1}, s_i) = \text{sign}(m(T))$, a Bradley-Terry-type model. (Which behaves a bit strangely.)

Point process math

- Write the probability of no scores in $(t, t + \Delta t)$ as $(1 - \lambda \Delta t)$ divided by the normalizing constant
- Write the probability of no scores in the interval (t, v) as a product of probabilities for the intervals of length Δt that make up (t, v)
- This is a product of terms like (1-small number). Approximate using $1 - x \simeq \exp(-x)$, then rewrite the product as $\exp\{-\text{sum of small numbers}\}$.
- The sum of small numbers in the exponential is a Riemann integral as $\Delta t \to 0$.

Computational remarks

The RUSH likelihood is not a delight to deal with, but:

- When $\alpha = 0$, which we need when calculating f, the number of scores remaining after time t is Poisson$(\lambda\{T - t\})$, so:
- $\Pr_0\{M(T) > 0 | M(t) = m\} = \sum_{k=0}^{\infty} \frac{e^{-\lambda(T-t)}[\lambda(T-t)]^k}{k!} \xi_k(m)$, where $\xi_k(m) = \Pr_0\{M(T) > 0 | M(t) = m, k \text{ scores remain}\}$ does not depend on λ or t so can be precomputed, if the πs are assumed fixed and known.
- $\xi_k(m)$ can be computed exactly for small k, and a normal approximation can be used for large k. (Actually we compute $\zeta_k(m, s) = \xi_k(m + s) - \xi_k(m)$.)

Computation, continued

- The integrand in the likelihood is smooth. We compute the integrals using Simpson's rule.
- Final MCMC computations done in YADAS. Previous (Newton-Raphson) version in R, which we still use to precompute the ζs and some additional quantities to help with the numerical integrations.