Model-Based Prediction in General and as Applied to the Outcomes of College Football Games

David A. Harville

May 8, 2018

Underlying model

- Data are regarded as the elements of the realization of an $N \times 1$ observable random vector $\mathbf{y}$
- Quantities to be predicted are regarded as the elements of an $M \times 1$ unobservable random vector $\mathbf{w}$
- Model consists of the specification of the joint distribution of $\mathbf{y}$ and $\mathbf{w}$ (or of various of its characteristics such as the 1st and 2nd moments) up to the value of a $P \times 1$ vector $\boldsymbol{\theta}$ of unknown parameters
- Joint distribution of $\mathbf{y}$ and $\mathbf{w}$ is “conditional” on the information in some collection $\mathcal{X}$
- Regard a prior distribution as part of the model (a model that is hierarchical in nature)
- Special case: $P = 0$

Example: mixed-effects linear model

- $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, with $\mathrm{E}(\mathbf{u}) = \mathbf{0}$, $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{u}, \mathbf{e}) = \mathbf{0}$, $\mathrm{var}(\mathbf{u}) = \sigma_u^2\mathbf{A}$, and $\mathrm{var}(\mathbf{e}) = \sigma_e^2\mathbf{I}$

- $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{A}$ are known (i.e., are matrices whose elements are determinable from the information in $\mathcal{X}$)

- $\mathbf{w} = \boldsymbol{\Lambda}'\boldsymbol{\beta} + \boldsymbol{\Gamma}'\mathbf{u}$ or $\mathbf{w} = \boldsymbol{\Lambda}'\boldsymbol{\beta} + \boldsymbol{\Gamma}'\mathbf{u} + \mathbf{d}$, where $\mathrm{E}(\mathbf{d}) = \mathbf{0}$ and $\mathrm{var}(\mathbf{d}) = \sigma_e^2\mathbf{I}$ (with $\mathbf{d}$ being uncorrelated with $\mathbf{u}$ and $\mathbf{e}$)

- $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma_u^2 \\ \sigma_e^2 \end{pmatrix}$ or, alternatively, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \\ \sigma_e^2 \end{pmatrix}$, where $\gamma = \sigma_u^2/\sigma_e^2$

Underlying model in the special case covered by Efron and Hastie (in *Computer Age Statistical Inference*)

- Random sample: $\begin{pmatrix} y_1 \\ \mathbf{x}_1 \end{pmatrix}, \ldots, \begin{pmatrix} y_N \\ \mathbf{x}_N \end{pmatrix}, \begin{pmatrix} y_{N+1} \\ \mathbf{x}_{N+1} \end{pmatrix}, \ldots, \begin{pmatrix} y_{N+M} \\ \mathbf{x}_{N+M} \end{pmatrix}$

- The $\mathbf{x}_i$'s are the values of a $Q \times 1$ vector of "predictors"

- $y_1, y_2, \ldots, y_N$ are observable; wish to predict the values of $y_{N+1}, y_{N+2}, \ldots, y_{N+M}$

- For some functions $\mu(\cdot)$ and $v(\cdot)$, $\mathrm{E}(y_i \,|\, \mathbf{x}_i) = \mu(\mathbf{x}_i)$ and $\mathrm{var}(y_i \,|\, \mathbf{x}_i) = v(\mathbf{x}_i)$

- $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$; $\mathbf{w} = (y_{N+1}, y_{N+2}, \ldots, y_{N+M})'$;
  $\mathcal{X}$ consists of $\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_{N+1}, \ldots, \mathbf{x}_{N+M}$

- A simple form: $\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ and $v(\mathbf{x}) = \sigma^2$ for all $\mathbf{x}$ [in which case, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$]

Predictive inference

Forms of predictive inference

- “Point” prediction/classification (with or without an estimate of error)
- Prediction intervals or sets
- “Posterior” probabilities

Approaches to predictive inference

- Algorithmic
- Model-based

Underlying model in the special case covered by Efron and Hastie (in *Computer Age Statistical Inference*)

- Random sample: $\begin{pmatrix} y_1 \\ \mathbf{x}_1 \end{pmatrix}, \ldots, \begin{pmatrix} y_N \\ \mathbf{x}_N \end{pmatrix}, \begin{pmatrix} y_{N+1} \\ \mathbf{x}_{N+1} \end{pmatrix}, \ldots, \begin{pmatrix} y_{N+M} \\ \mathbf{x}_{N+M} \end{pmatrix}$

- The $\mathbf{x}_i$'s are the values of a $Q \times 1$ vector of "predictors"

- $y_1, y_2, \ldots, y_N$ are observable; wish to predict the values of $y_{N+1}, y_{N+2}, \ldots, y_{N+M}$

- For some functions $\mu(\cdot)$ and $v(\cdot)$, $\mathrm{E}(y_i \,|\, \mathbf{x}_i) = \mu(\mathbf{x}_i)$ and $\mathrm{var}(y_i \,|\, \mathbf{x}_i) = v(\mathbf{x}_i)$

- $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$; $\mathbf{w} = (y_{N+1}, y_{N+2}, \ldots, y_{N+M})'$;
  $\mathcal{X}$ consists of $\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}_{N+1}, \ldots, \mathbf{x}_{N+M}$

Model-based predictive inference: some thoughts

- "... all models are wrong, but some are useful" (George Box); an addendum: a model may be useful for some purposes , but not for others

- Performance in "repeated application" is important—minimum requirement: being well-calibrated (Dawid, JASA 1982)

- Ideally, $\mathcal{X}$ would include all of the variables and factors that are important and no others

- The relationships implicit in the model should reflect the "actual relationships"

- Success may require flexibility and improvisation

Model-based prediction: an application

Application: Predict (on a weekly basis) the outcomes (winners and MOVs) of those college football games involving a Division I-FBS team

3 groups of Division I teams: (1) power-5 FBS; (2) non-power-5 FBS; (3) FCS

Notation: $p(i)$ = group number of the ith team

$y_{kjj'\ell}$ = score of the home team minus the score of the visiting team in the ℓth game played in Year k by Teams j and j' on the home "field" of Team j

$x_{kjj'\ell}$ = 0 or 1 depending on whether or not the $kjj'\ell$th game is played at a neutral site

$\mathbf{y}$ = $y_{kjj'\ell}$'s for past games

$\mathbf{w}$ = $y_{kjj'\ell}$'s for future games

$\mathcal{X}$ includes the $x_{kjj'\ell}$'s, the identities of the home and visiting teams, and the $p(i)$'s

A model

$$y_{kjj'\ell} = x_{kjj'\ell}\lambda + \alpha_{p(j)} + t_{jk} - (\alpha_{p(j')} + t_{j'k}) + e_{kjj'\ell}$$

$e_{kjj'\ell}$'s: uncorrelated random variables with mean 0 and variance σ^2

t_{jk}'s: random variables with $\mathrm{E}(t_{jk}) = 0$, $\mathrm{var}(t_{jk}) = \gamma_{p(j)}\sigma^2$, and

$$\mathrm{corr}(t_{jk}, t_{jk'}) = \rho_{p(j)}^{|k'-k|}$$

$\boldsymbol{\theta} = (\lambda, \alpha_2, \alpha_3, \sigma^2, \gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2, \rho_3)'$

Accounting for discreteness and for OT (overtime)

To account for the inherent discreteness of the scores, redefine $y_{kjj'\ell}$ to be a "latent" variable

Let $z_{kjj'\ell}$ = the score of the "home" team minus the score of the "visiting" team

"Thresholds": $0 < \xi_0 < \xi_1 < \xi_2 < \cdots$

Correspondence:

$$\begin{aligned} z_{kjj'\ell} = i \quad &\Leftrightarrow \quad \xi_{i-1} < y_{kjj'\ell} \le \xi_i \quad (i = 1, 2, \ldots) \\ z_{kjj'\ell} = -i \quad &\Leftrightarrow \quad -\xi_i < y_{kjj'\ell} \le -\xi_{i-1} \quad (i = 1, 2, \ldots) \\ z_{kjj'\ell} = 0 \quad &\Leftrightarrow \quad -\xi_0 < y_{kjj'\ell} \le \xi_0 \end{aligned}$$

Rounding to the nearest integer: take $\xi_i = i + 0.5 \quad (i = 0, 1, 2, \ldots)$

OT adjustment:

$$\begin{aligned} z_{kjj'\ell} > 0 \quad &\Leftrightarrow \quad 0 < y_{kjj'\ell} \le \xi_0 \\ z_{kjj'\ell} < 0 \quad &\Leftrightarrow \quad -\xi_0 < y_{kjj'\ell} \le 0 \end{aligned}$$

More on accounting for discreteness and for OT

Reference: "Collegiate football scores, U.S.A." (Mosteller, JASA 1970)

Assumption: The values of $i \geq 24$ were partitioned into intervals, within each of which the ξ_i's were taken to be equally spaced

Definition: $o_{kjj'\ell} = 1$ or 0 depending on whether or not the $kjj'\ell$th game is decided in OT

$\mathbf{y}$ = $z_{kjj'\ell}$'s and $o_{kjj'\ell}$'s for past games;

$\mathbf{w}$ = $z_{kjj'\ell}$'s for future games;

$\mathcal{X}$ as before

Regard $\lambda, \alpha_2, \alpha_3$, and the ξ_i's as random, with a marginal (prior) distribution that is noninformative

$\boldsymbol{\theta} = (\gamma_1, \gamma_2, \gamma_3, \rho_1, \rho_2, \rho_3)'$

Some background

Weekly predictions were made during the 2017 season

Data were those for games between Division I teams and included the data from the 2014, 2015, and 2016 seasons

Computations for models involving latent variables: Gibbs sampler can be used to advantage (Albert and Chib, JASA 1993; Sorensen et al., Genet. Sel. Evol. 1995)

<u>Modification</u>: $e_{kjj'\ell}$'s were taken to have a t distribution with 3 df

Methodology for making weekly predictions

The predictions were made via a 4-step process: ,

(1) Using the data from 2014, 2015, and 2016, draws were made from what was regarded as the marginal posterior distribution of λ and the ξ_i's

(2) Using the data from 2014, 2015, and 2016 and the draws from Step (1), estimates were obtained for γ_1, γ_2, γ_3, ρ_1, ρ_2, and ρ_3

(3) Using all of the data (including those from the preceding weeks of 2017), using the draws from Step (1), and acting as though the parameter estimates from Step (2) were true values, draws were obtained from the posterior distribution of α_2, α_3, and the t_{jk}'s

(4) Using the draws from Steps (1) and (3) and continuing to act as though the parameter estimates from Step (2) were true values, predictions were made for the $z_{kjj'\ell}$'s (those representing that week's games)

Numerical results: thresholds and home-field advantage

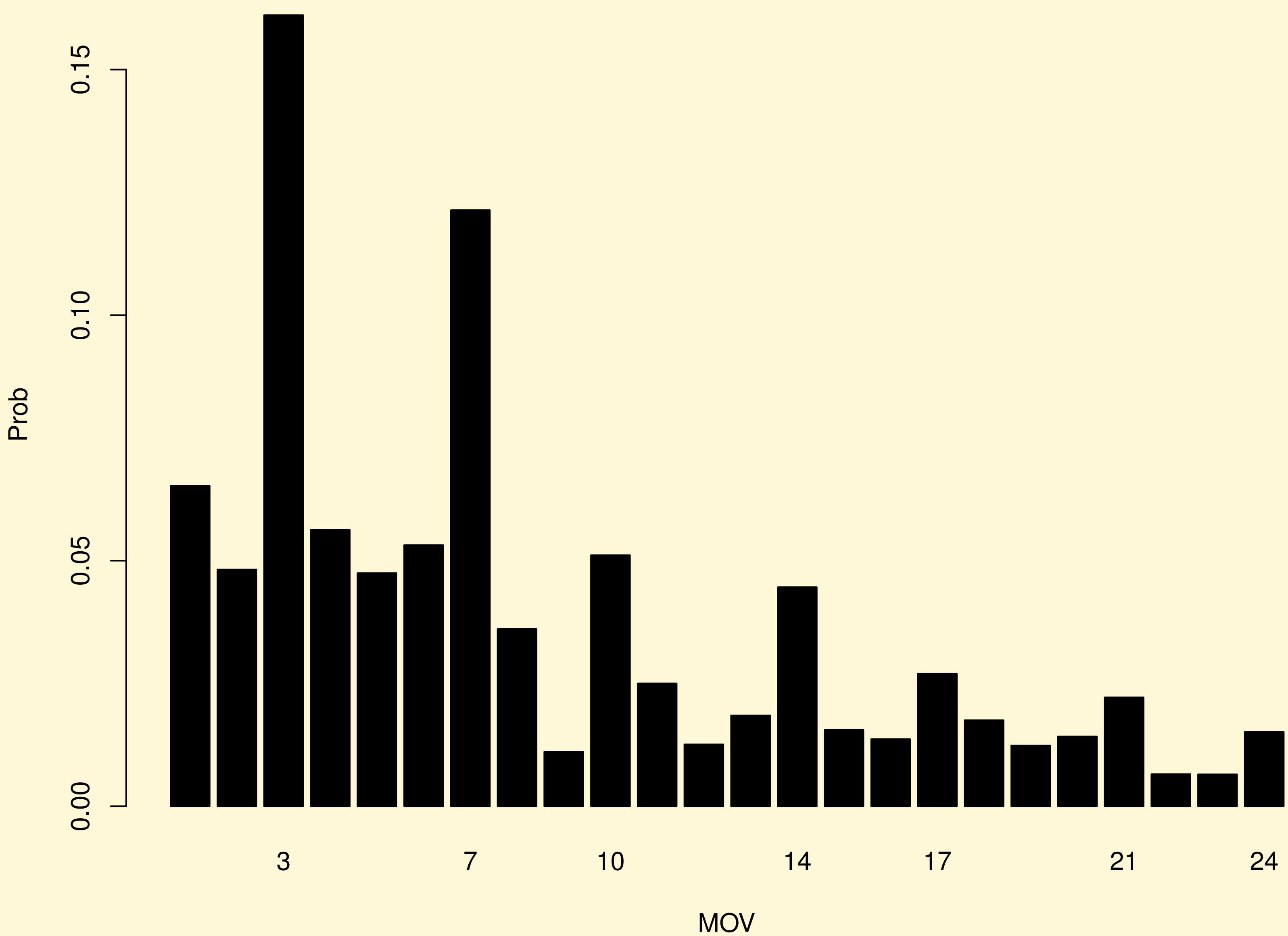
i	Estimate* of $\xi_i - \xi_{i-1}$	i	Estimate of $\xi_i - \xi_{i-1}$	i	Estimate of $\xi_i - \xi_{i-1}$	i	Estimate of $\xi_i - \xi_{i-1}$
0	0.21**	17	0.13	8	0.08	23	0.06
7	0.19	18	0.09	19	0.07	16	0.06
3	0.19	20	0.09	11	0.07	13	0.06
21	0.17	4	0.09	6	0.07	22	0.06
24	0.16	1	0.08	2	0.07	12	0.04
14	0.16	5	0.08	15	0.06	9	0.03
10	0.13						

*Posterior mean ** Estimate of $2\xi_0$

Average value of the 25 estimates in the table = 0.10

Estimate (posterior mean) of λ = 0.23

Posterior probabilities for MOVs between 1 and 24 (in a game between evenly matched teams on a neutral field)



Estimates of the γ_i's and ρ_i's

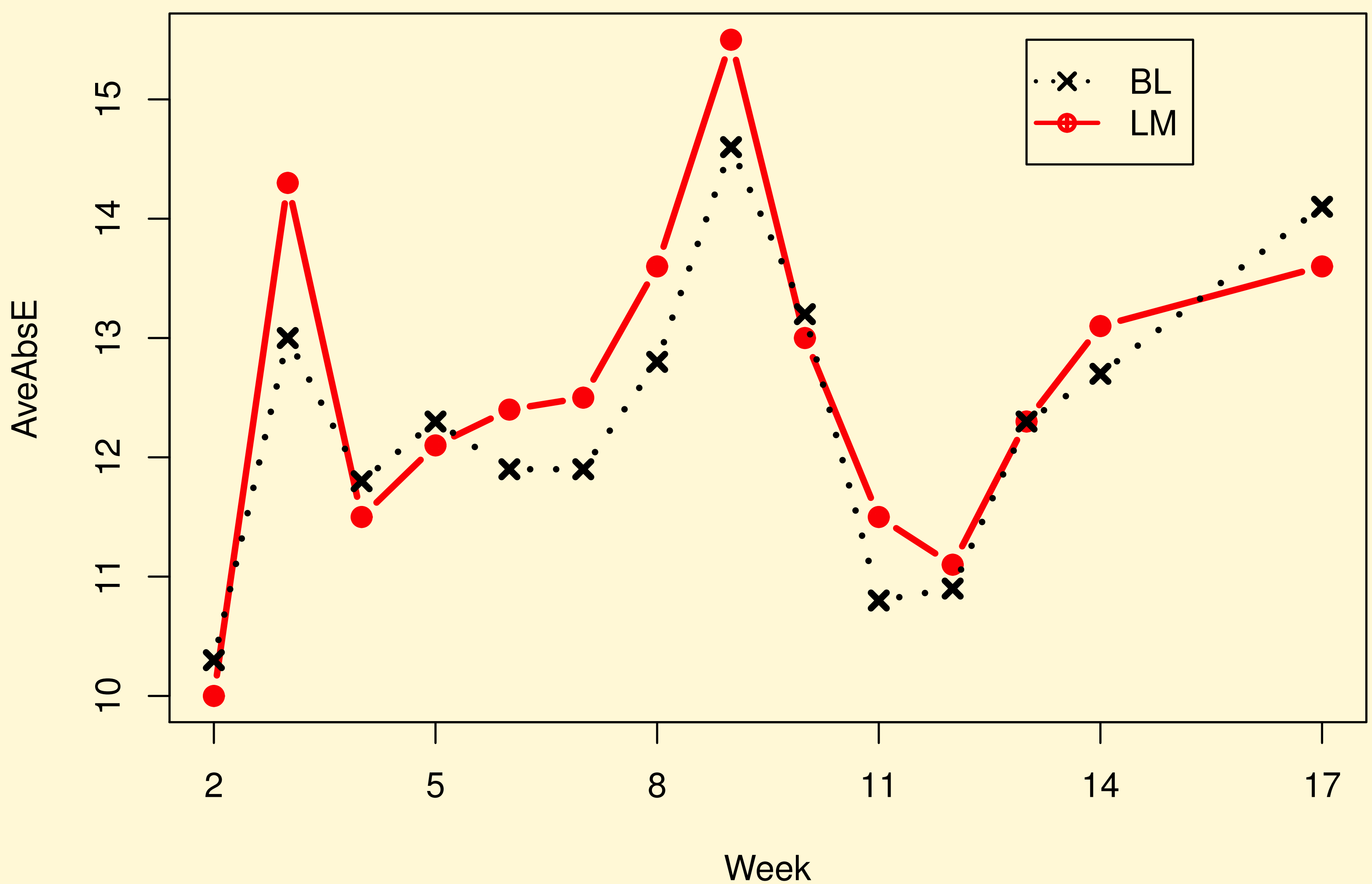
i	Group i	γ_i	ρ_i
1	Power-5 FBS teams	0.61	0.75
2	Non-power-5 FBS teams	0.67	0.55
3	FCS teams	1.18	0.86

Accuracy of the (point) predictions (for the $z_{kjj'\ell}$'s) obtained weekly over the course of the 2017 season*: Linear Model vs. Betting Line

	% winners	Ave. Absolute Error	Root MSE
Linear Model	73	12.6	15.9
Betting Line	74	12.3	15.6

*The number of games for which predictions were obtained totaled 752

Average absolute error (AveAbsE) week-by-week (for "weeks" with at least 39 predictions)*: Linear Model (LM) vs. Betting Line (BL)



*Week 17: 39 bowl games

Posterior probability of a game being won by the "favored" team* vs. the actual frequency of winning

Number of games	Average posterior probability	Actual frequency	Standard error
107	0.54	0.49	0.05
107	0.61	0.60	0.05
107	0.68	0.58	0.05
107	0.74	0.71	0.04
107	0.81	0.84	0.04
107	0.88	0.91	0.03
110	0.94	0.97	0.02

*Whichever team has the higher posterior probability of winning

(Posterior) expected value of the MOV (for a future game) vs. the actual MOV

Number of games	Average expected MOV	Average actual MOV	Standard error
107	11.4	11.2	1.0
107	12.2	9.9	0.9
107	13.4	16.2	1.2
107	15.1	13.9	1.0
107	17.6	17.3	1.1
107	21.3	22.9	1.2
110	29.7	29.1	1.4

Posterior probability of a team* winning by more (or losing by less) than indicated by the betting line vs. the actual frequency

Number of games	Average posterior probability	Actual frequency	Standard error
104	0.51	0.52	0.05
104	0.53	0.49	0.05
107	0.55	0.59	0.05
105	0.57	0.47	0.05
106	0.60	0.40	0.05
105	0.63	0.59	0.05
108	0.69	0.55	0.04

*Whichever team has the higher posterior probability of "beating the spread" (i.e., the team for which this probability exceeds 0.5)