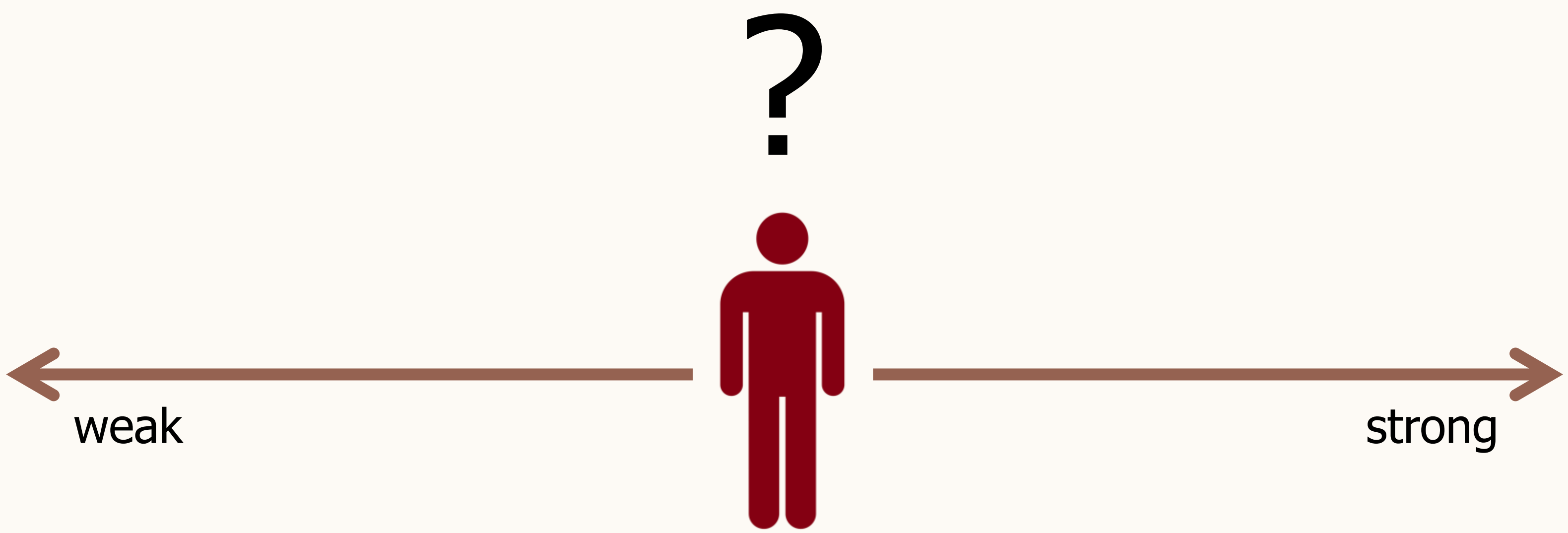
Athlete rating in score-outcome multi-competitor games

JONATHAN CHE & MARK GLICKMAN

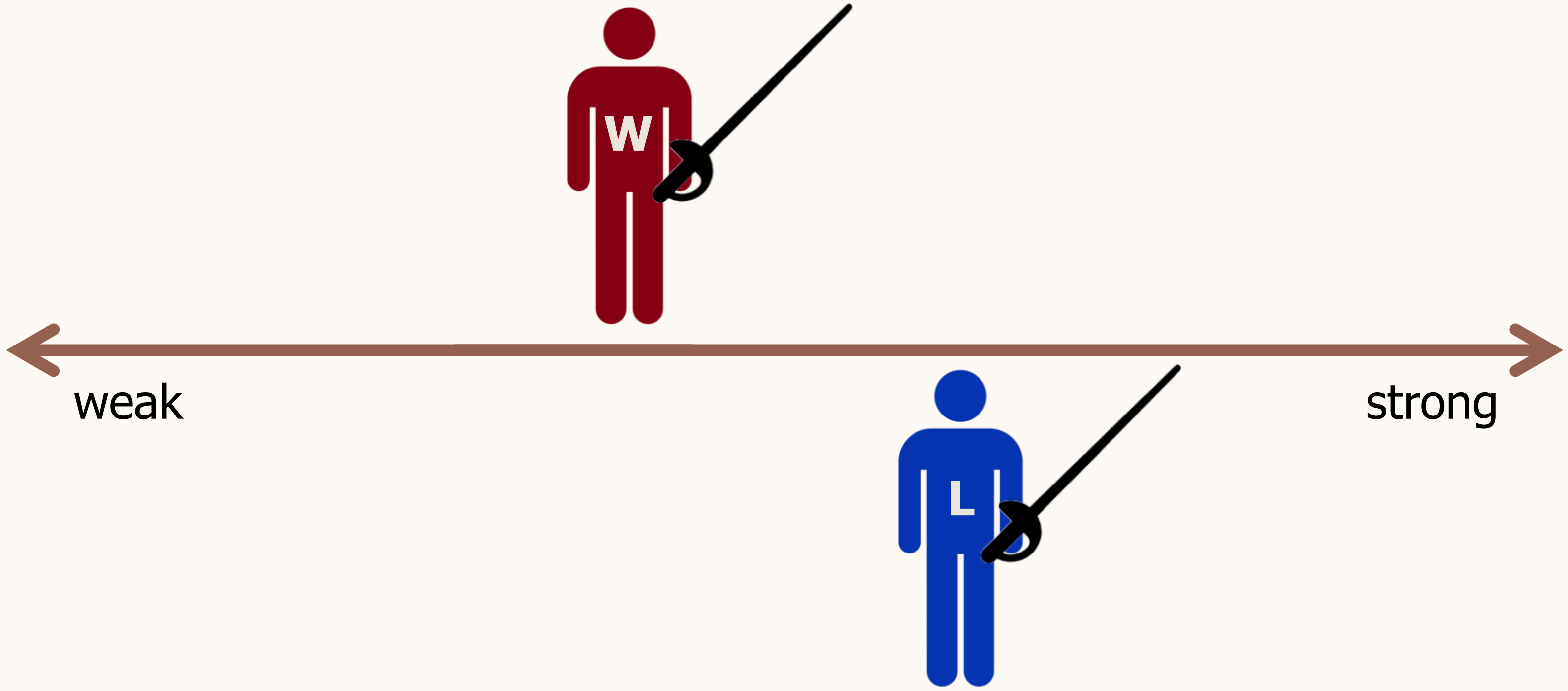
SUPPORTED BY US OLYMPIC & PARALYMPIC COMMITTEE

Athlete rating

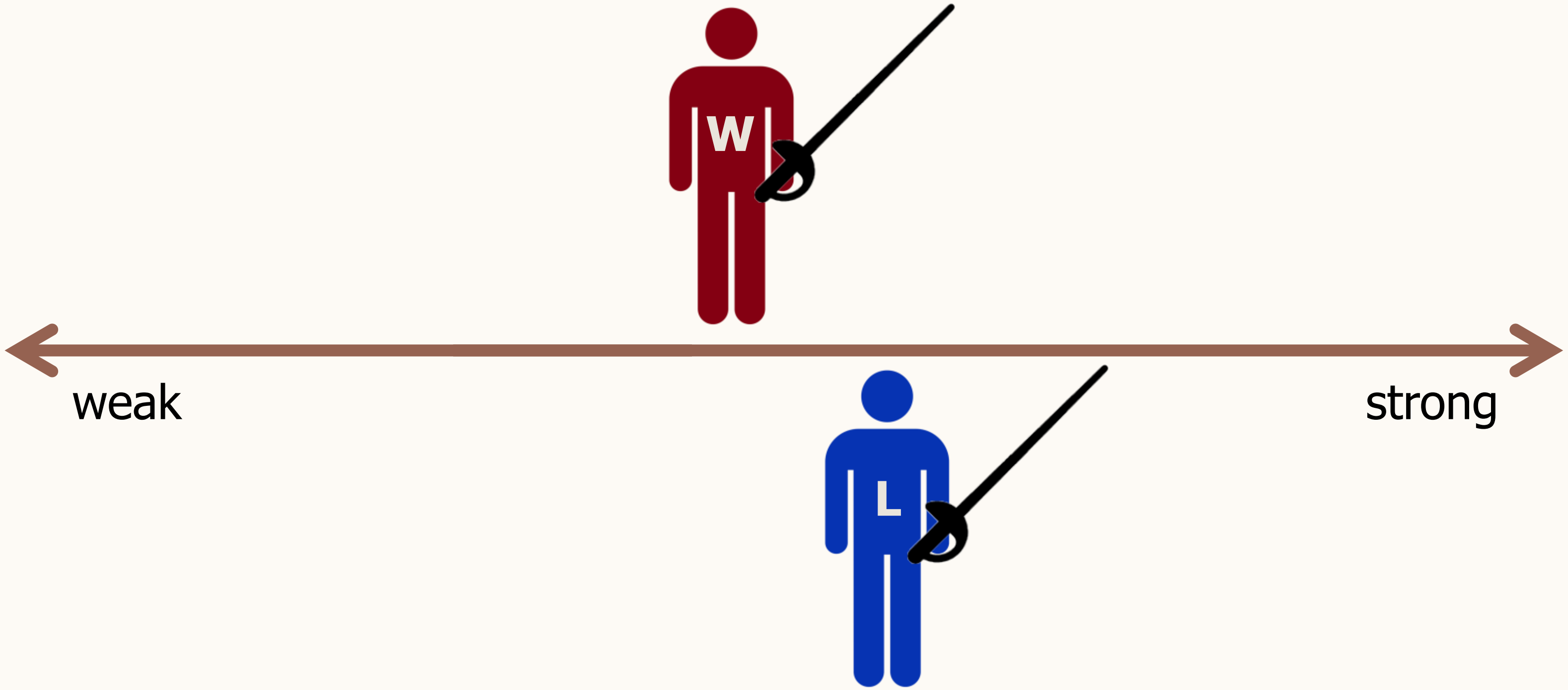
Athlete rating



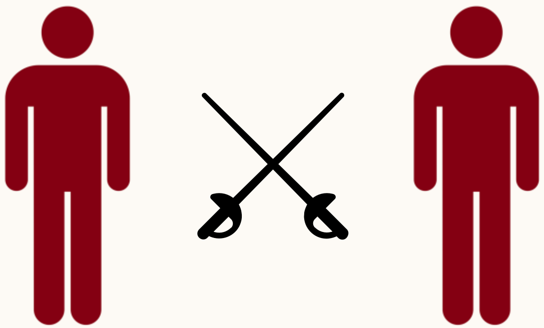
Athlete rating



Athlete rating



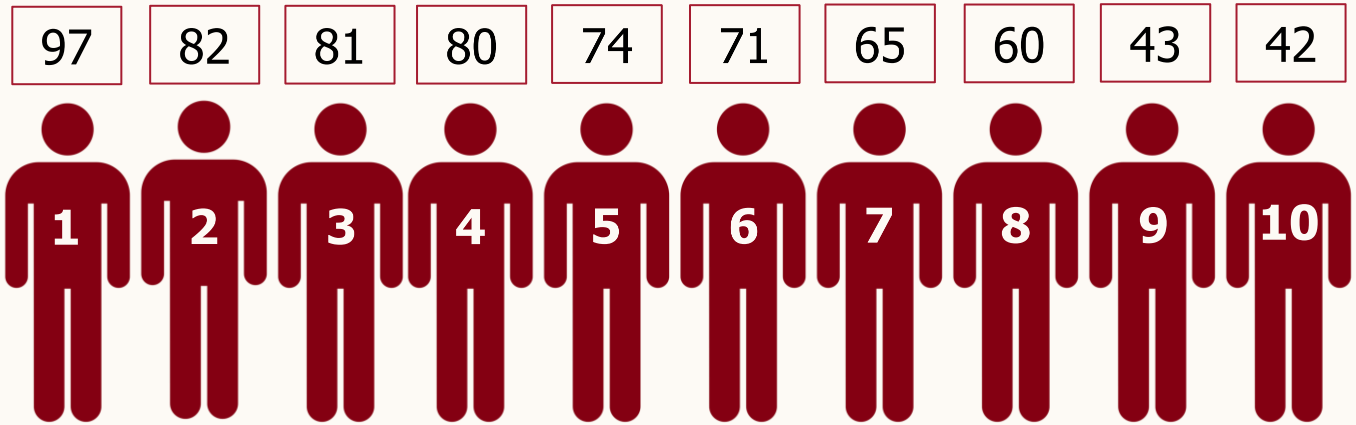
Wrinkle 1: “multi-competitor”



Wrinkle 1: “multi-competitor”

"A stochastic rank ordered logit model for rating multi-competitor games and sports“
(Glickman & Hennessy 2015)

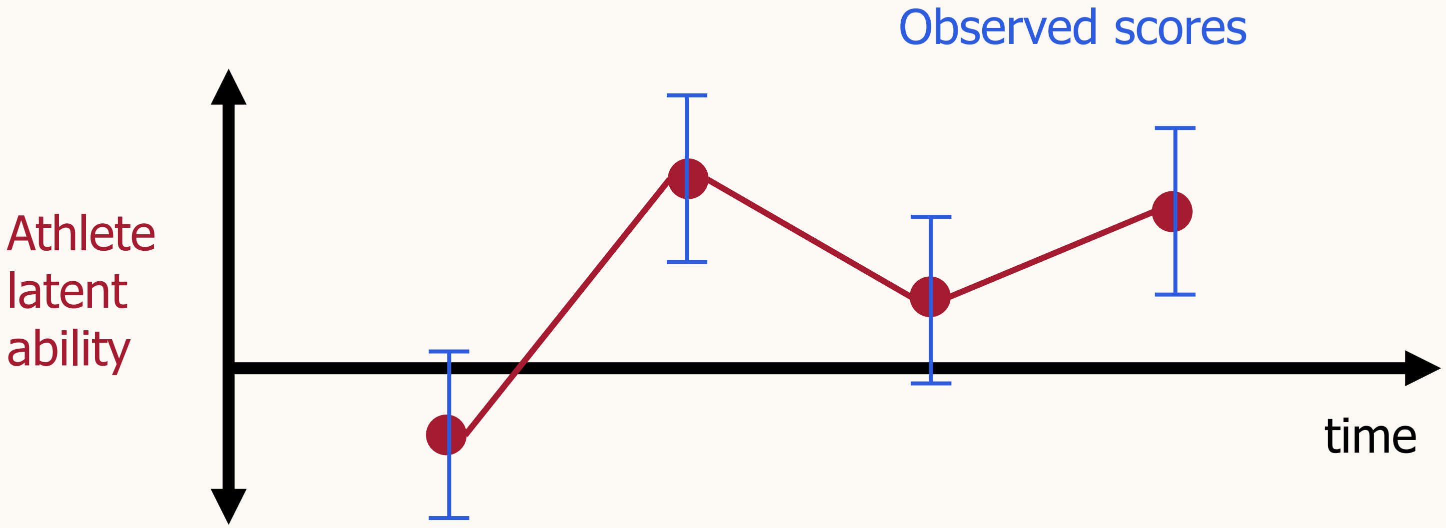
Wrinkle 2: "score-outcome"



A model for athlete rating

Dynamic linear model (DLM): visual idea

Harville (1977); Glickman & Stern (1998)



DLM: standard equations

For each athlete in game g within time period t:

observed score | latent ability | observation variance

$$p(\, y_{gt} \mid \theta_t, \sigma^2 \,) = N(\theta_t, \sigma^2)$$

$$p(\, \theta_{t+1} \mid \theta_t, \sigma^2 \,) = N(\theta_t, \sigma^2 W)$$

innovation variance ratio

$$p(\sigma^2) = IG(a_0, b_0)$$

$$p(\, \theta_1 \mid \sigma^2 \,) = N(0, \sigma^2 V_0)$$

initial ability variance ratio

We can fit this model using Kalman Filter equations

Modification 1: game-specific effects

For each athlete in game g within time period t:

$$p\left(y_{gt} \mid \theta_t, \sigma^2 \right) = N(\theta_t, \sigma^2)$$

$$p(\theta_{t+1} \mid \theta_t, \sigma^2) = N(\theta_t, \sigma^2 W)$$

Modification 1: game-specific effects

For each athlete in game g within time period t:

$$p\left(y_{gt} - \bar{y}_{gt} \mid \theta_t, \sigma^2 \right) = N(\theta_t - \bar{\theta}_{gt}, \sigma^2)$$

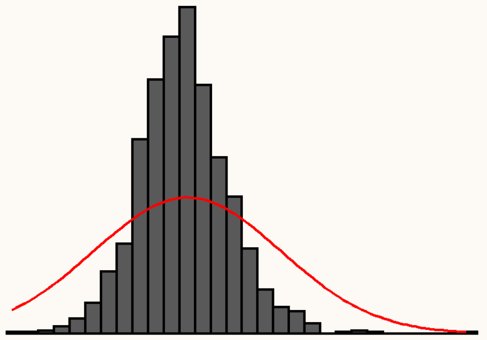
$$p(\theta_{t+1} \mid \theta_t, \sigma^2) = N(\theta_t, \sigma^2 W)$$

Modification 2: data transformations

For each athlete in game g within time period t:

$$p\left(y_{gt} - \bar{y}_{gt} \mid \theta_t, \sigma^2 \right) = N(\theta_t - \bar{\theta}_{gt}, \sigma^2)$$

$$p\left(\theta_{t+1} \mid \theta_t, \sigma^2 \right) = N(\theta_t, \sigma^2 W)$$



Modification 2: data transformations

For each athlete in game g within time period t:

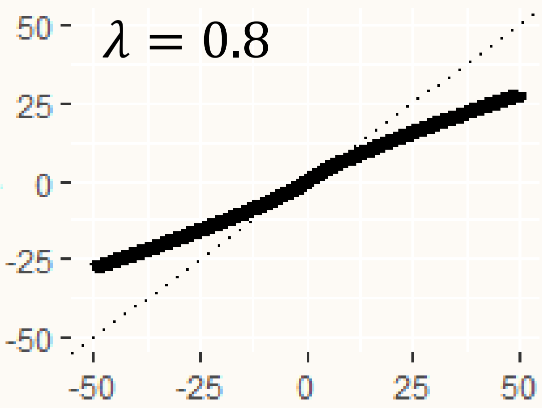
$$p\left(t_\lambda(y_{gt} - \bar{y}_{gt}) \mid \theta_t, \sigma^2 \right) = N(\theta_t - \bar{\theta}_{gt}, \sigma^2)$$

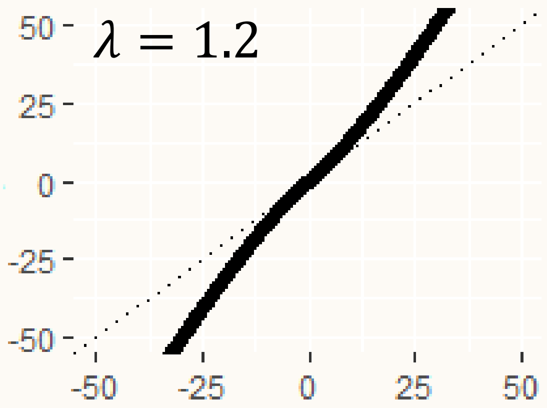
$$p(\theta_{t+1} \mid \theta_t, \sigma^2) = N(\theta_t, \sigma^2 W)$$

Transformations: Yeo-Johnson

(Yeo & Johnson 2000)

$$t_\lambda^{YJ}(y) = \begin{cases} ((y+1)^\lambda - 1)/\lambda & y \geq 0 \\ -((-y+1)^{2-\lambda} - 1)/(2-\lambda) & y < 0 \end{cases}$$

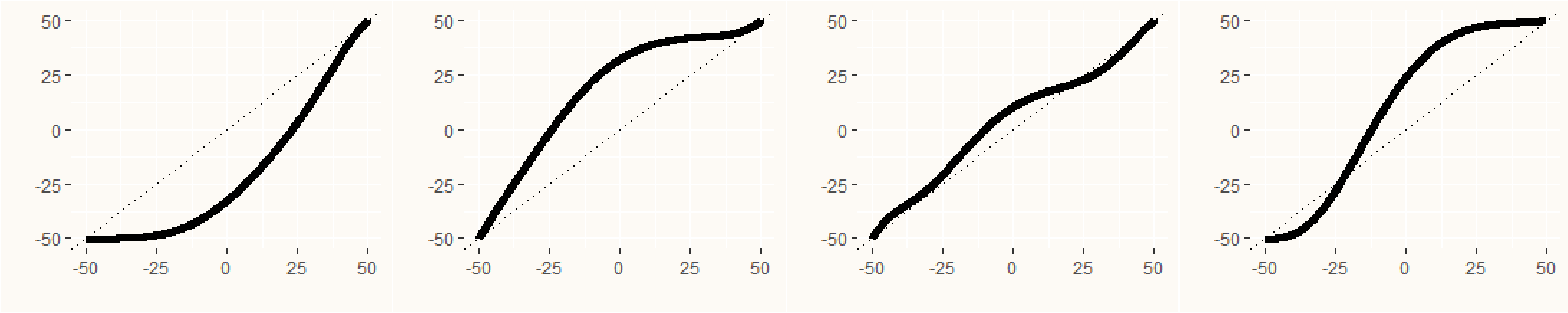




Transformations: monotone spline

(Ramsay 1988)

$$t_{\boldsymbol{\lambda}}^{MS}(y) = \sum_{b=1}^{B} \lambda_b \cdot I_b(y)$$



Fitting the model

DLM with transformation: full model

For each athlete in game g within time period t:

$$p\left(t_\lambda(y_{gt} - \bar{y}_{gt}) \mid \theta_t, \sigma^2 \right) = N(\theta_t - \bar{\theta}_{gt}, \sigma^2)$$

$$p(\theta_{t+1} \mid \theta_t, \sigma^2) = N(\theta_t, \sigma^2 W)$$

DLM with transformation: full model

For each athlete in game g within time period t:

$$p(\boldsymbol{\psi}_t^{\lambda} \mid \theta_t, \sigma^2) = N(\theta_t - \bar{\theta}_{gt}, \sigma^2)$$

$$p(\theta_{t+1} \mid \theta_t, \sigma^2) = N(\theta_t, \sigma^2 W)$$

DLM with transformation: full model

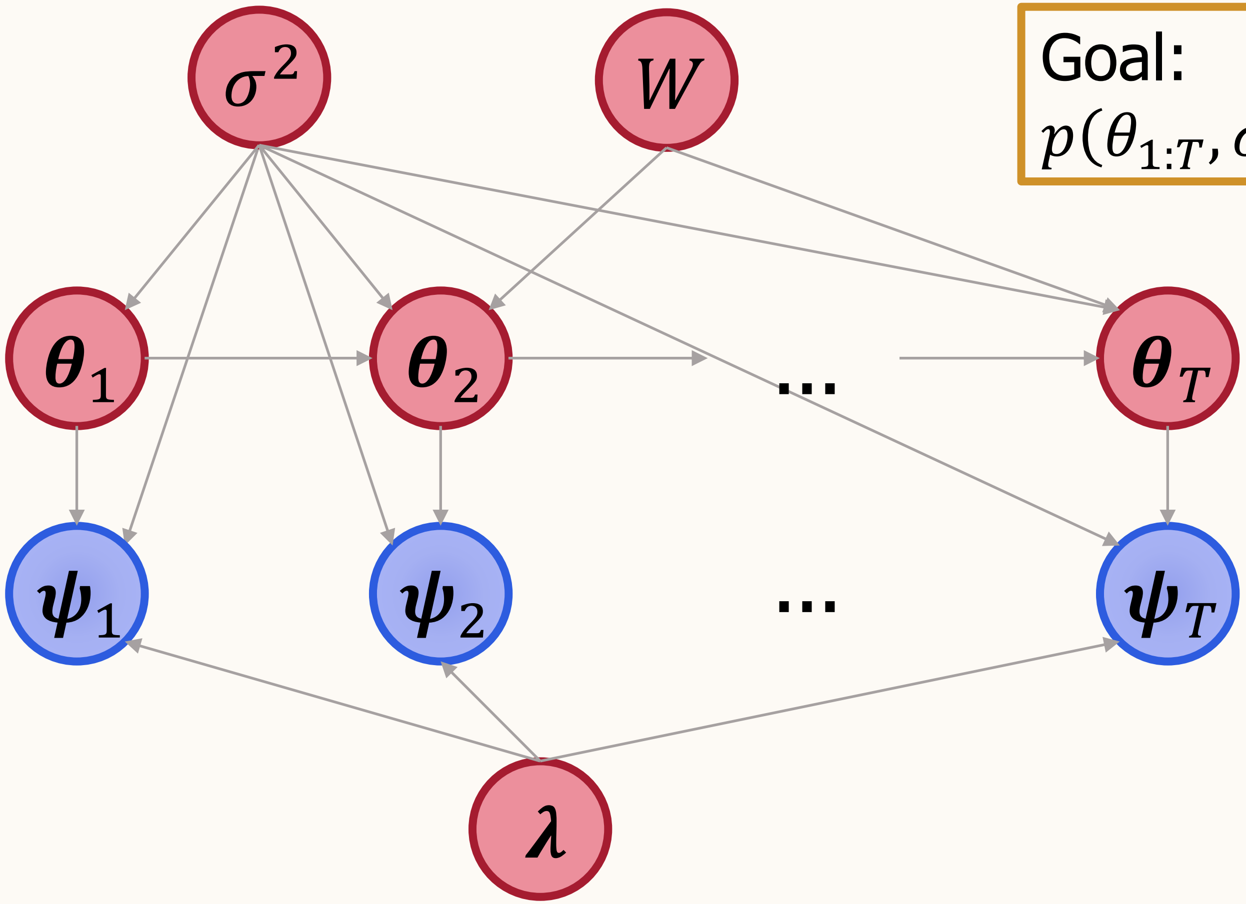
$$p(y_{1:T}, \theta_{1:T}, \sigma^2, W, \lambda) =$$

$$J\left(\psi_{1:T}^{\lambda} \rightarrow y_{1:T}\right) \times \underbrace{p(\sigma^2)p(W)p(\lambda)}_{\text{priors for } \sigma^2, W, \lambda}$$

$$\times \prod_{t=1}^{T} \underbrace{p(\psi_t^{\lambda} \mid \theta_t, \sigma^2, \lambda)}_{\text{likelihood}} \times \prod_{t=1}^{T} \underbrace{p(\theta_t \mid \theta_{t-1}, \sigma^2, W)}_{\text{innovation}}$$

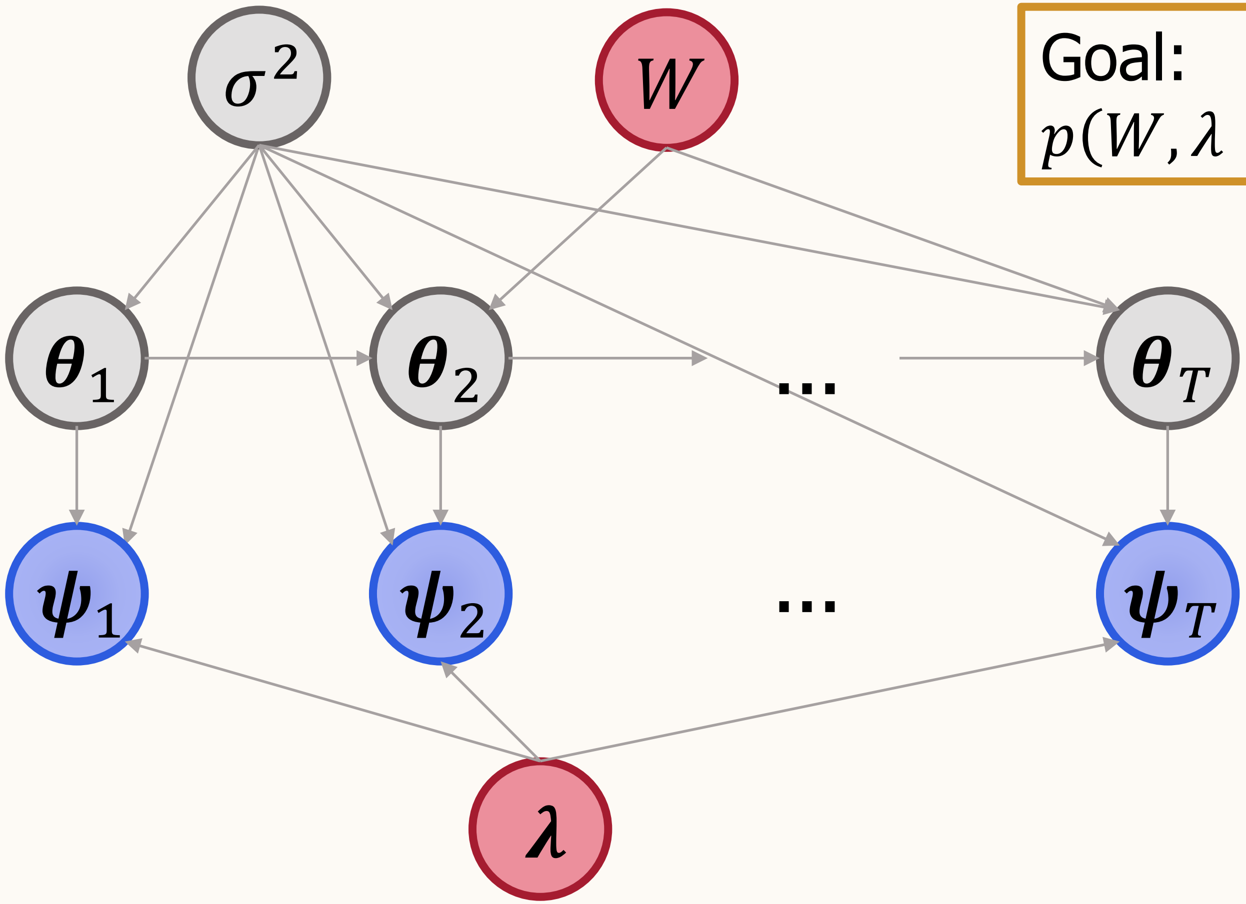
transformation Jacobian → $J\left(\psi_{1:T}^{\lambda} \rightarrow y_{1:T}\right)$

DLM with transformation: model fitting



Goal:
$p(\theta_{1:T}, \sigma^2, W, \lambda \mid y_{1:T})$

DLM with transformation: model fitting



Goal:
$p(W, \lambda \mid y_{1:T})$

DLM with transformation: model fitting

$$p(W, \lambda \mid y_{1:T})$$

$$= \int p(\theta_{1:T}, \sigma^2, W, \lambda \mid y_{1:T}) d\theta_{1:T} d\sigma^2$$

DLM with transformation: model fitting

Given W and $\boldsymbol{\lambda}$:

1. Use Kalman Filter to help evaluate $p(\psi_t^\lambda \mid \psi_{1:t-1}^\lambda, W, \lambda)$ for $t = 1, \dots, T$
2. Compute $p(W, \lambda \mid y_{1:T})$

$$p(W, \lambda \mid y_{1:T})$$

$$= \int p(\theta_{1:T}, \sigma^2, W, \lambda \mid y_{1:T}) d\theta_{1:T} d\sigma^2$$

$$= \underbrace{J(\psi_{1:T}^\lambda \to y_{1:T})}_{\text{transformation Jacobian}} \times \underbrace{p(W)p(\lambda)}_{\text{priors for } W, \lambda} \times \prod_{t=1}^T \underbrace{p(\psi_t^\lambda \mid \psi_{1:t-1}^\lambda, W, \lambda)}_{\text{posterior predictive t-distributions}}$$

DLM with transformation: model fitting

Goal: find posterior mode of $p(W, \boldsymbol{\lambda} \mid \boldsymbol{y}_{1:T})$

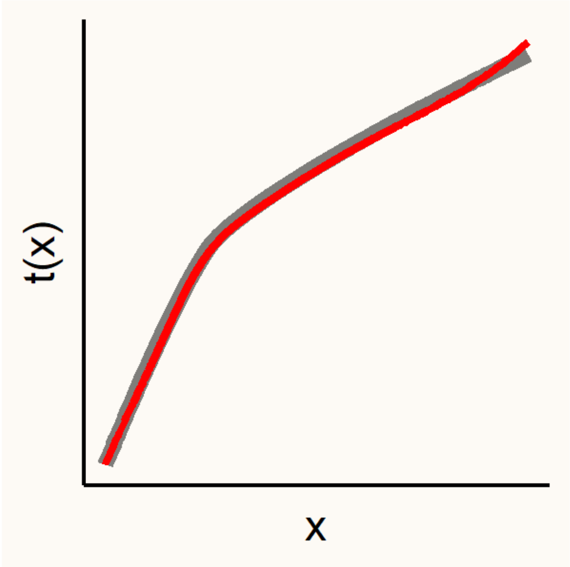
$$\max_{W,\boldsymbol{\lambda}} p(W, \boldsymbol{\lambda} \mid \boldsymbol{y}_{1:T})$$

$$\text{subject to } W, \lambda_i \geq 0 \ \forall i$$

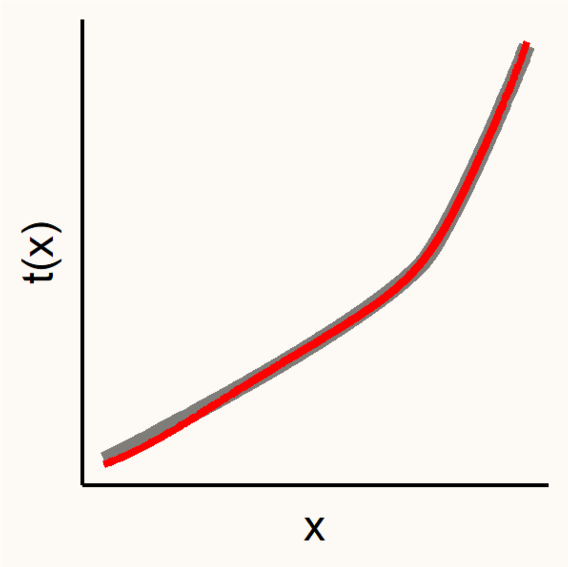
Results

Recovering the true transformation

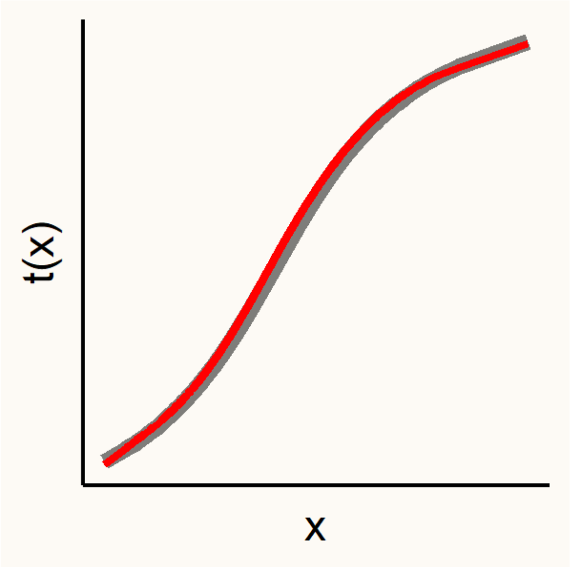
$t^{YJ}_{0.8}(x)$



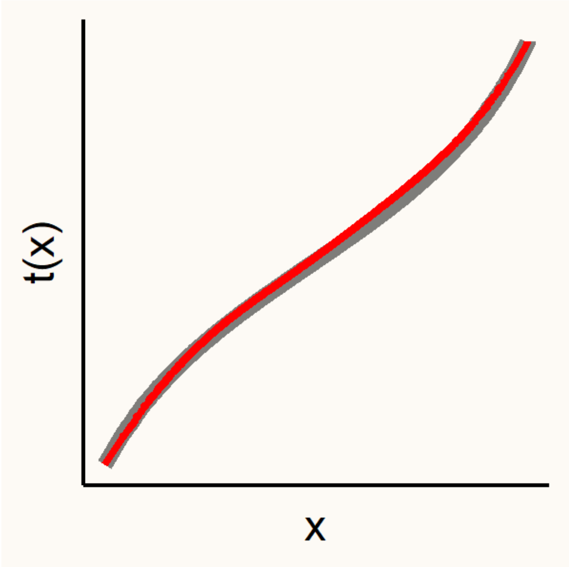
$t^{YJ}_{1.2}(x)$



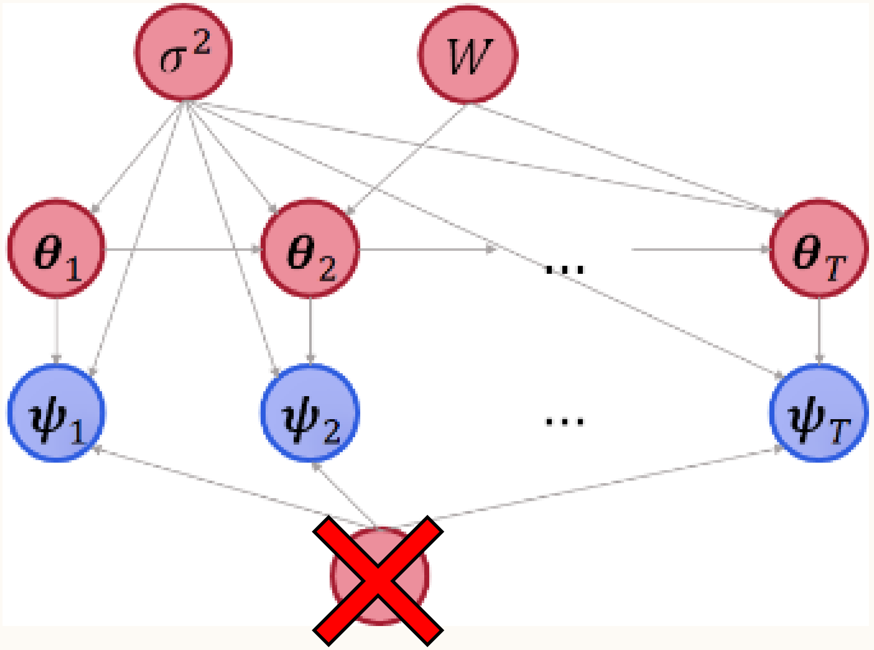
$50 \cdot \mathrm{atan}(\frac{x}{50})$



$50 \cdot \tan(\frac{x}{50})$



Model comparison: other models



Vanilla DLM (VLM)

Rank-order logit model (ROL)

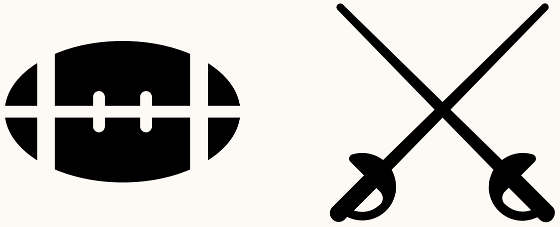
Model comparison: train/test split

Training set	Test set
Find $\widehat{W}$ & $\hat{\lambda}$	Make predictions

Data

Multicompetitor:

Head-to-head:



Model comparison: multi-competitor data

Games g, time periods t, correlations ρ_{gt}, competition sizes n_{gt}:

$$\rho = \frac{\sum_{t=s+1}^{T} \sum_{g=1}^{N_{gt}} (n_{gt}-1)\rho_{gt}}{\sum_{t=s+1}^{T} \sum_{g=1}^{N_{gt}} (n_{gt}-1)}$$

	Diving			Biathlon			Biathlon relay		
Metric	VLM	LMT	ROL	VLM	LMT	ROL	VLM	LMT	ROL
Spearman	**0.62**	0.58	0.61	0.61	0.60	**0.61**	0.75	**0.77**	0.75
Kendall	**0.48**	0.45	0.47	0.44	0.44	**0.44**	0.59	**0.60**	0.58

Model comparison: head-to-head data

	Rugby			Fencing (15 pts)		
Metric	VLM	LMT	GLO	VLM	LMT	GLO
Accuracy	.72	**.73**	.70	.66	**.70**	0.68

Initial conclusions

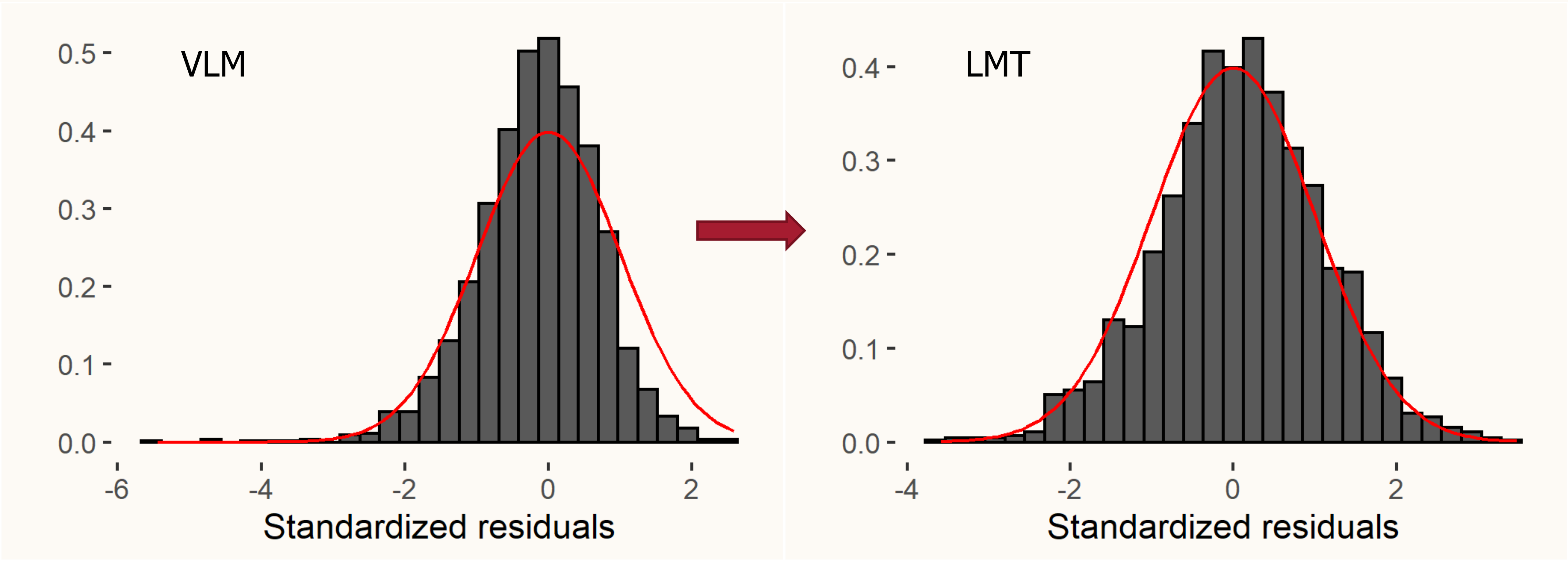
1. Difficult to improve on ROL predictions for given datasets
2. Scores may help more in low-data settings

Does the LMT have any other nice features?

Exploring results: post. predictive intervals

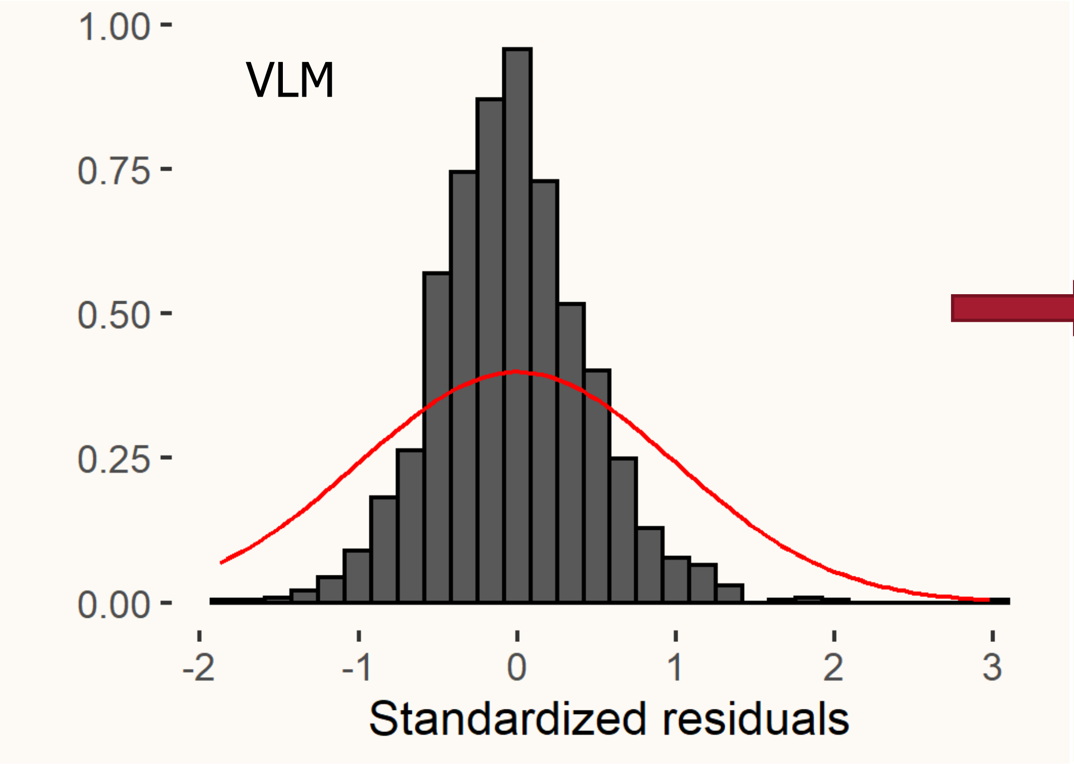
Exploring results: residuals

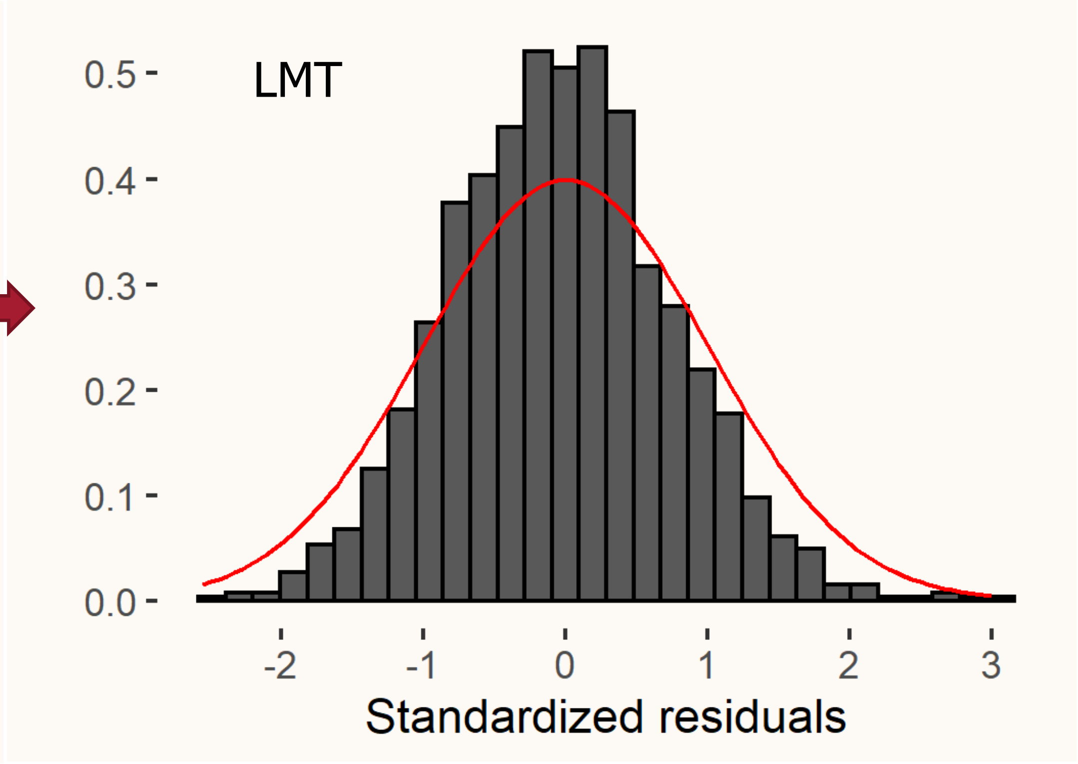
Biathlon



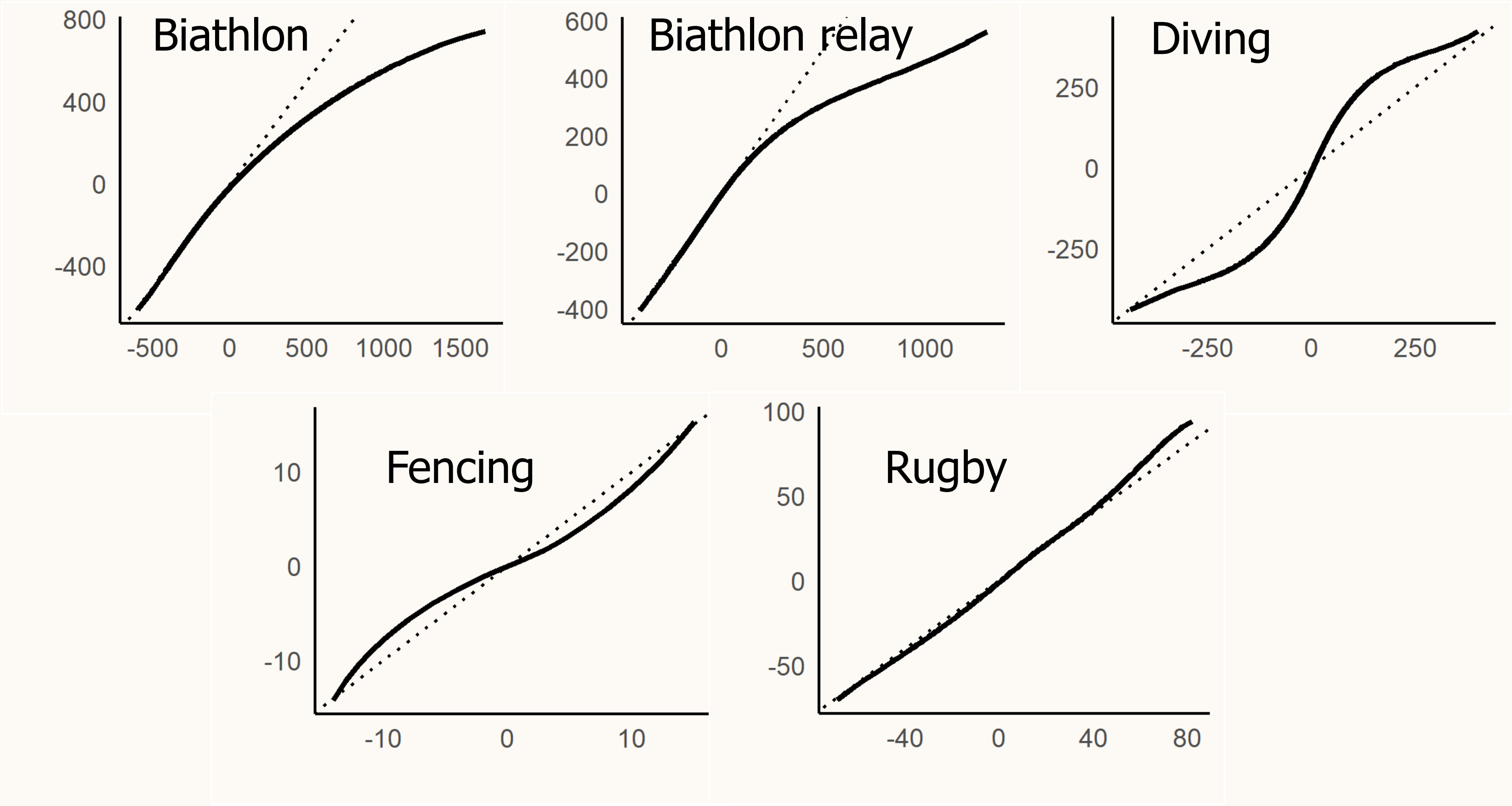
Exploring results: residuals

Diving

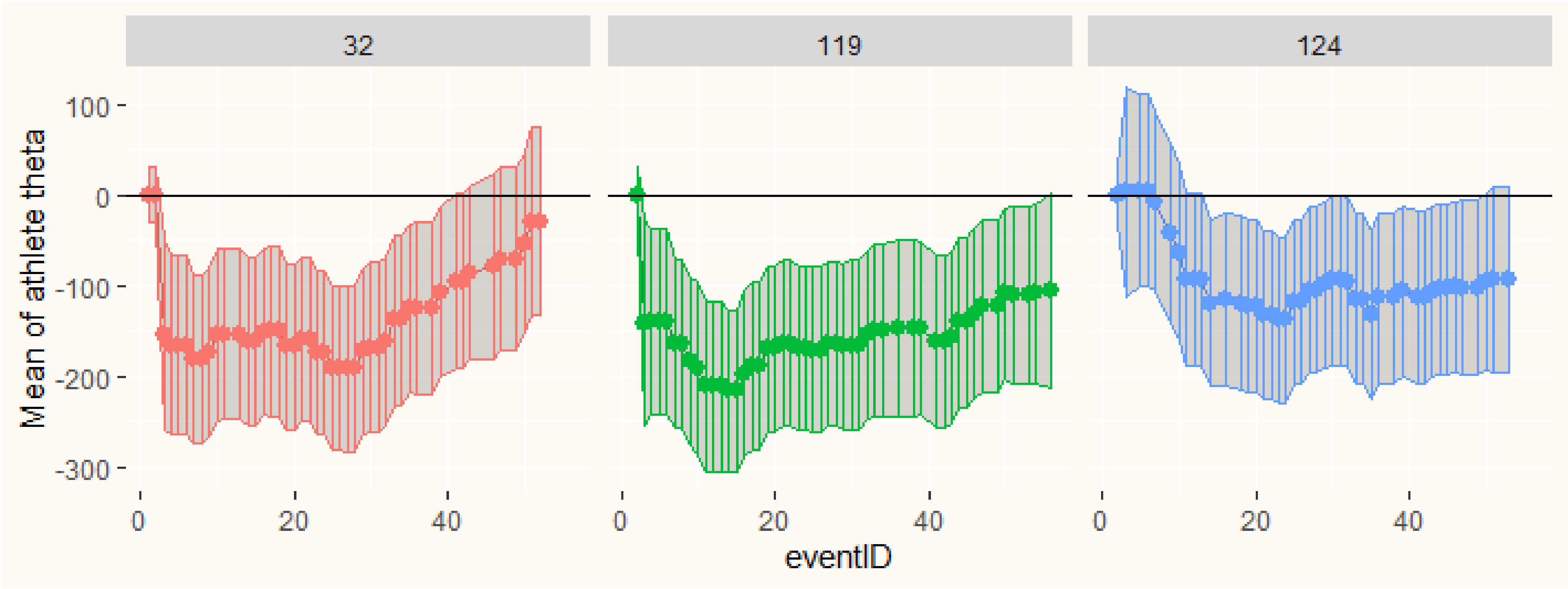




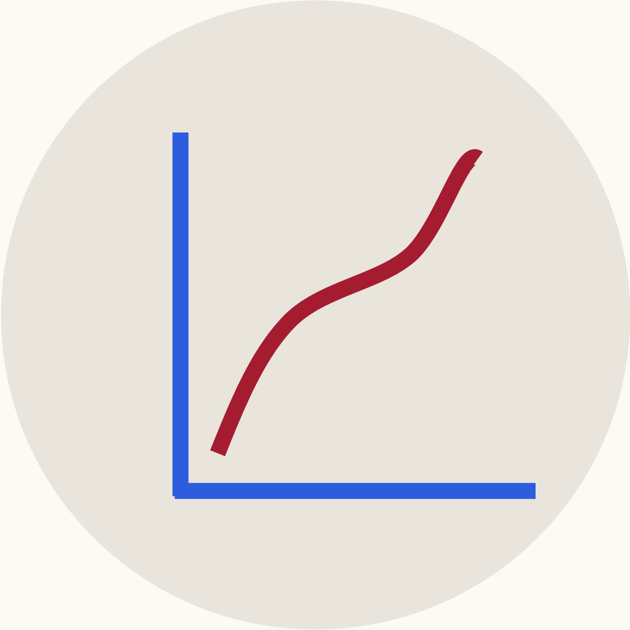
Exploring results: transformations



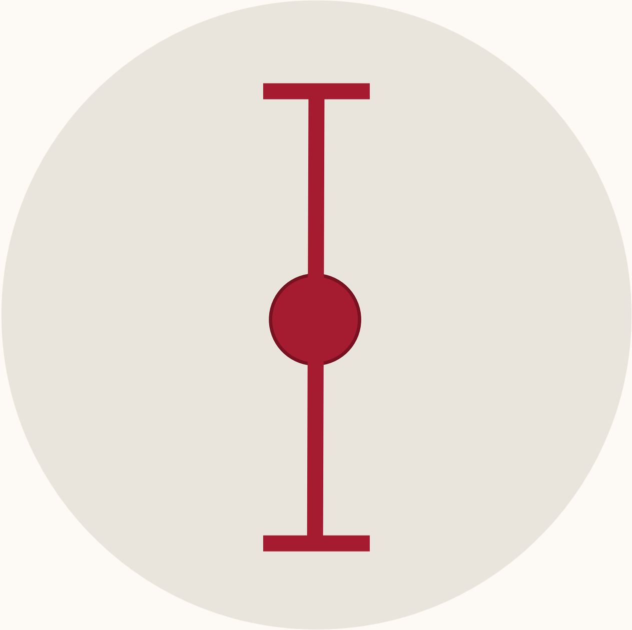
Exploring results: athlete strengths



Conclusion



ρ



References

Glickman, Mark E., and Jonathan Hennessy. "A stochastic rank ordered logit model for rating multi-competitor games and sports." Journal of Quantitative Analysis in Sports 11.3 (2015): 131-144.

Glickman, Mark E., and Hal S. Stern. "A state-space model for National Football League scores." Journal of the American Statistical Association 93.441 (1998): 25-35.

Harville, David. "The use of linear-model methodology to rate high school or college football teams." Journal of the American Statistical Association 72.358 (1977): 278-289.

Ramsay, James O. "Monotone regression splines in action." Statistical science 3.4 (1988): 425-441.

Yeo, In-Kwon, and Richard A. Johnson. "A new family of power transformations to improve normality or symmetry." Biometrika 87.4 (2000): 954-959.

Thank you!

DLM: model matrix notation

N total athletes, N_{gt} athletes playing in game g within time period t:

$$p\left(\boldsymbol{y}_{gt} \mid \theta_t, \sigma^2\right) = N(X_{gt}\boldsymbol{\theta}_t, \sigma^2 \cdot I_{N_{gt}})$$

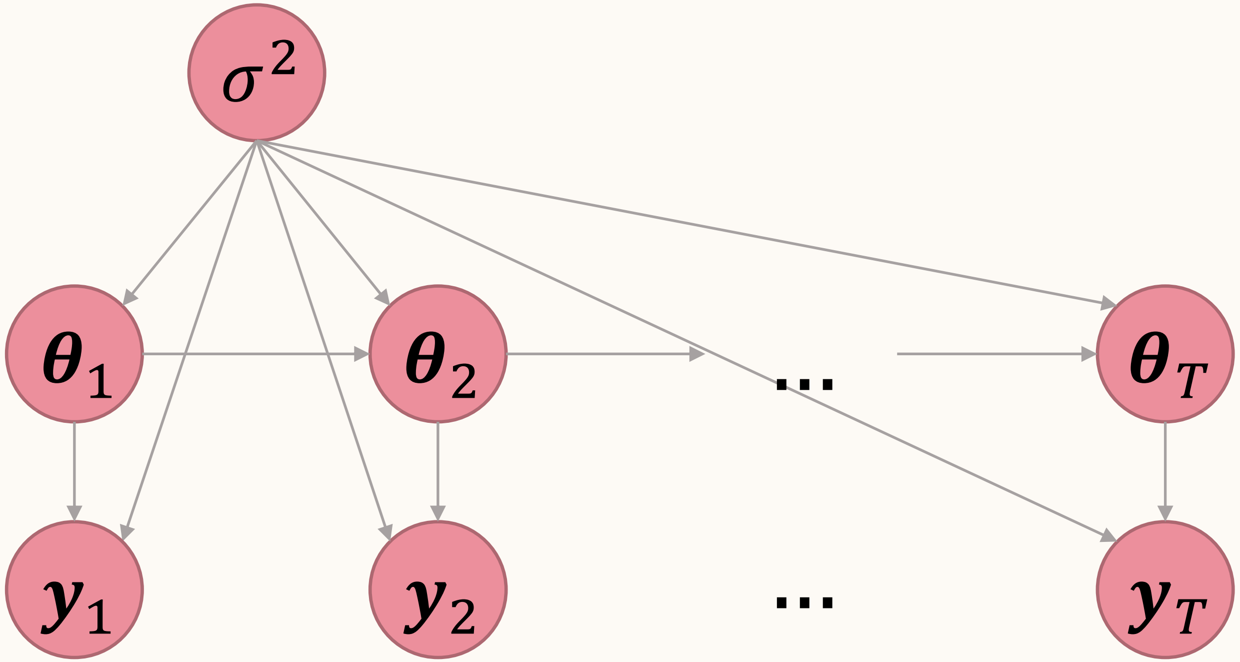
$$p(\boldsymbol{\theta}_{t+1} \mid \boldsymbol{\theta}_t, \sigma^2) = N(\boldsymbol{\theta}_t, \sigma^2 W \cdot I_N)$$

$$X_{gt} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{N_{gt} \times N}$$ (multicompetitor game)

$$X_{gt} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \end{bmatrix}$$ (head-to-head game)

DLM: full model

$$p(y_{1:T}, \theta_{1:T}, \sigma^2) = \prod_{t=1}^{T} p(y_t \mid \theta_t, \sigma^2) \times \prod_{t=1}^{T} p(\theta_t \mid \theta_{t-1}, \sigma^2) \times p(\sigma^2)$$



DLM: posterior updates

Target: $p(\boldsymbol{\theta}_{1:T}, \sigma^2 \mid \boldsymbol{y}_{1:T})$

Prior:

$$\sigma^2 \sim IG(a_{t-1}, b_{t-1})$$
$$\boldsymbol{\theta}_t | \sigma^2 \sim N(\boldsymbol{m}_{t-1}, \sigma^2 \cdot V_{t-1})$$

&

Likelihood:

$$\boldsymbol{y}_t \sim N(X_t \cdot \boldsymbol{\theta}_t, \sigma^2 \cdot I_{N_t})$$

=

Posterior:

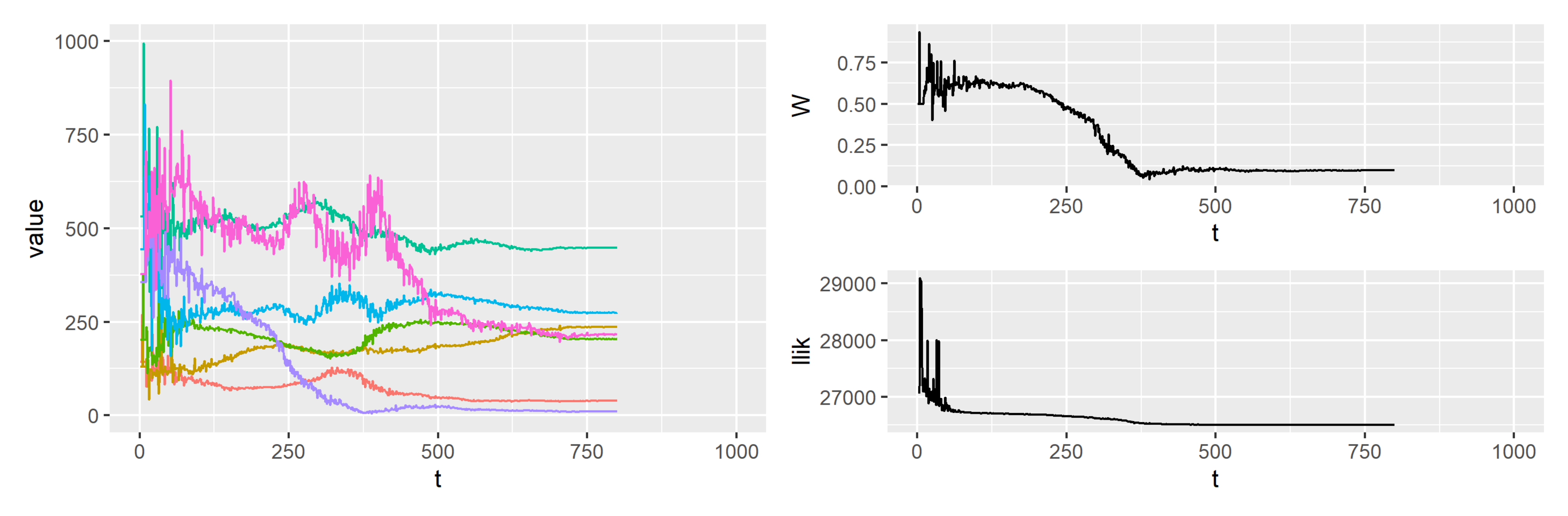
$$\sigma^2 | \boldsymbol{y}_t \sim IG(a_t, b_t)$$
$$\boldsymbol{\theta}_t | \boldsymbol{y}_t, \sigma^2 \sim N(\boldsymbol{m}_t, \sigma^2 \cdot V_t)$$

Innovation period:

$$\boldsymbol{\theta}_{t+1} \mid \boldsymbol{\theta}_t \sim N(\boldsymbol{\theta}_t, \sigma^2 W)$$

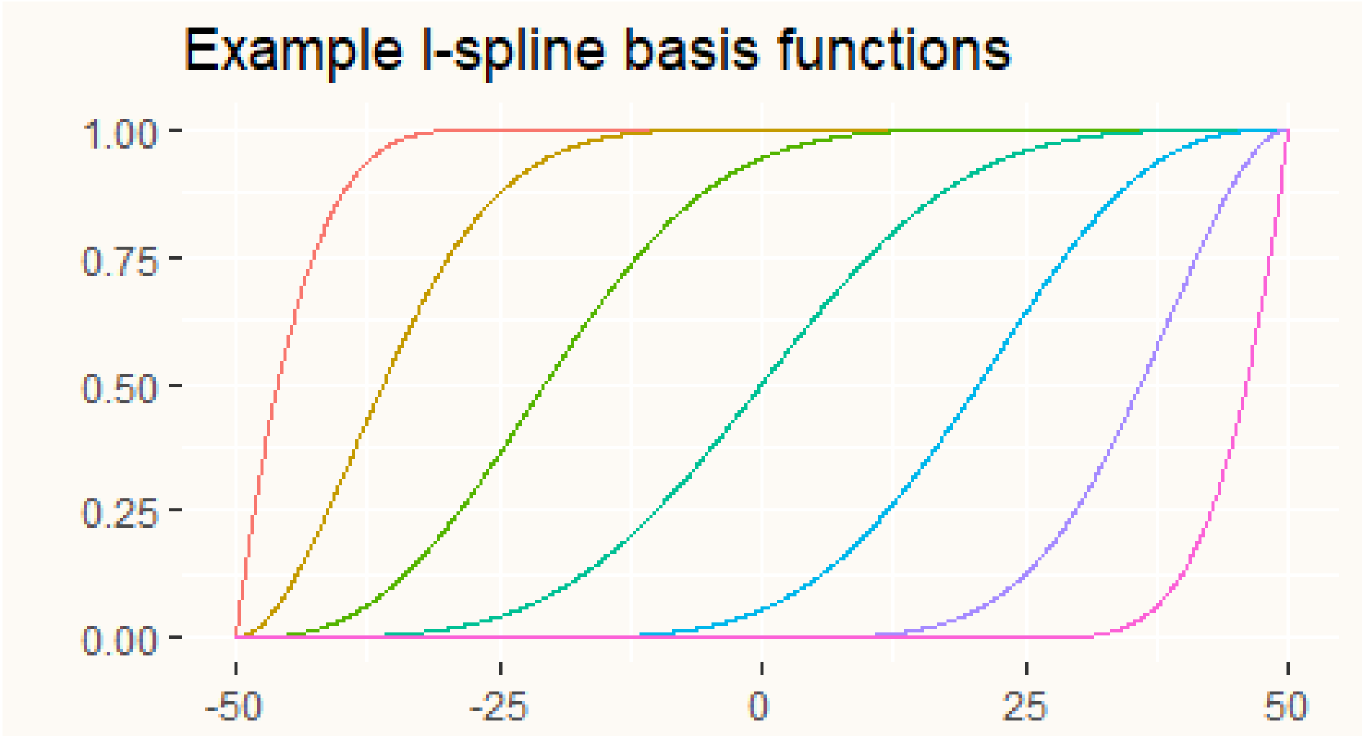
Exploring results: optim() convergence

Monotone spline for biathlon dataset



Transformations: monotone spline (Ramsay 1988)

$$t_{\boldsymbol{\lambda}}^{MS}(y) = \sum_{i=1}^{B} \lambda_i \cdot I_i(y)$$



DLM with transformation: model fitting

$$p(W, \lambda \mid y_{1:T})$$

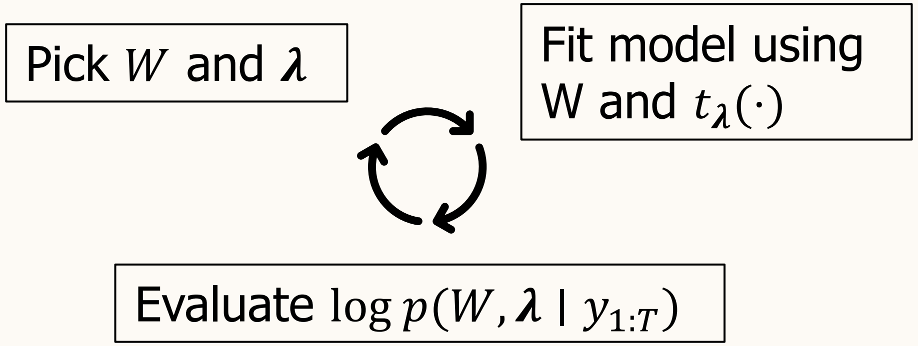
$$= \int p(\theta_{1:T}, \sigma^2, W, \lambda \mid y_{1:T}) d\theta_{1:T} d\sigma^2$$

$$= J\left(\psi_{1:T}^{\lambda} \to y_{1:T}\right) \times p(W)p(\lambda) \times \prod_{t=1}^{T} \underbrace{p(\psi_t^{\lambda} \mid \psi_{1:t-1}^{\lambda}, W, \lambda)}_{t_{2a_t-1}\left(\bar{X}_t m_{t-1}, \frac{b_{t-1}}{a_{t-1}}[I + \bar{X}_t(V_{t-1} + WI)\bar{X}_t^T]\right)}$$

DLM with transformation: model fitting

Goal: find the values of W and λ that maximize $p(W, \lambda \mid y_{1:T})$

1. Initialize: set initial $(W, \boldsymbol{\lambda})$
2. Optimize: use numerical optimization to find optimal value of $(W, \boldsymbol{\lambda})$



3. Return: optimal (MAP) value of $(W, \boldsymbol{\lambda})$

Exploring results: 95% coverage

	Diving	Biathlon	Biathlon Relay	Rugby	Fencing
LMNT	0.998	0.979	0.998	0.958	0.949
LM	0.987	0.949	0.985	0.954	0.945

Exploring results: coverage