Paired Comparison Models with Tie Probabilities that Depend on Competitor Strength

Mark E. Glickman

Department of Statistics, Harvard University

August 4, 2026

Outline

1. Paired comparisons with ties: what classical models do well
2. Why draw probabilities may depend on absolute strength
3. Linear strength-dependent model from the paper
4. Spline extension for the draw term
5. Bayesian fitting with informative strength priors
6. Choosing spline flexibility using LOO and sensitivity checks

The paired comparison problem

We observe outcomes between two competitors i and j:

$$Y_{ij} = \begin{cases} 0 & \text{if } i \text{ loses to } j \\ 1/2 & \text{if } i \text{ draws (ties) } j \\ 1 & \text{if } i \text{ defeats } j. \end{cases}$$

The paired comparison problem

We observe outcomes between two competitors i and j:

$$Y_{ij} = \begin{cases} 0 & \text{if } i \text{ loses to } j \\ 1/2 & \text{if } i \text{ draws (ties) } j \\ 1 & \text{if } i \text{ defeats } j. \end{cases}$$

Classical Bradley–Terry model for binary outcomes (assuming no ties):

$$\text{logit } \Pr(Y_{ij} = 1) = \theta_i - \theta_j.$$

- each competitor has latent strength θ_i,
- only strength differences matter in the binary Bradley-Terry model.

The modeling problem

For each game we observe

$$(i, j, x_{ij}, Y_{ij}),$$

where i and j are the players, x_{ij} records who has White, and Y_{ij} is the game outcome.

The modeling problem

For each game we observe

$$(i, j, x_{ij}, Y_{ij}),$$

where i and j are the players, x_{ij} records who has White, and Y_{ij} is the game outcome.

A useful paired-comparison model should:

- estimate latent player strengths θ_i;
- accommodate draws as a genuine third outcome;
- allow for a White/first-move advantage;
- produce interpretable win, draw, and loss probabilities.

The modeling problem

For each game we observe

$$(i, j, x_{ij}, Y_{ij}),$$

where i and j are the players, x_{ij} records who has White, and Y_{ij} is the game outcome.

A useful paired-comparison model should:

- estimate latent player strengths θ_i;
- accommodate draws as a genuine third outcome;
- allow for a White/first-move advantage;
- produce interpretable win, draw, and loss probabilities.

The starting point is Bradley–Terry; the next step is Davidson's extension for ties.

Davidson (1970) model for ties

A three-outcome model due to Davidson (1970) assigns probabilities to win, loss, and draw:

$$
\begin{aligned}
\Pr(Y_{ij} = 1) &\propto \exp(\theta_i), \\
\Pr(Y_{ij} = 0) &\propto \exp(\theta_j), \\
\Pr(Y_{ij} = 1/2) &\propto \nu \exp\left\{ \frac{\theta_i + \theta_j}{2} \right\}.
\end{aligned}
$$

Davidson (1970) model for ties

A three-outcome model due to Davidson (1970) assigns probabilities to win, loss, and draw:

$$\begin{aligned}
\Pr(Y_{ij} = 1) &\propto \exp(\theta_i), \\
\Pr(Y_{ij} = 0) &\propto \exp(\theta_j), \\
\Pr(Y_{ij} = 1/2) &\propto \nu \exp\left\{ \frac{\theta_i + \theta_j}{2} \right\}.
\end{aligned}$$

Key feature:

$$\frac{\Pr(i \text{ wins})}{\Pr(j \text{ wins})} = \exp(\theta_i - \theta_j),$$

so introducing a draw outcome does not change the win–loss odds.

Adding order effects

For chess, let $x_{ij} = 1$ if player i has White and -1 if player i has Black.

An extension by David (1988) combines ties and order effects:

$$\Pr(Y_{ij} = 1) \propto \exp\left(\theta_i + \frac{\alpha x_{ij}}{4}\right),$$
$$\Pr(Y_{ij} = 0) \propto \exp\left(\theta_j - \frac{\alpha x_{ij}}{4}\right),$$
$$\Pr(Y_{ij} = 1/2) \propto \nu \exp\left\{\frac{\theta_i + \theta_j}{2}\right\}.$$

Adding order effects

For chess, let $x_{ij} = 1$ if player i has White and -1 if player i has Black.

An extension by David (1988) combines ties and order effects:

$$\Pr(Y_{ij} = 1) \propto \exp\left(\theta_i + \frac{\alpha x_{ij}}{4}\right),$$
$$\Pr(Y_{ij} = 0) \propto \exp\left(\theta_j - \frac{\alpha x_{ij}}{4}\right),$$
$$\Pr(Y_{ij} = 1/2) \propto \nu \exp\left\{\frac{\theta_i + \theta_j}{2}\right\}.$$

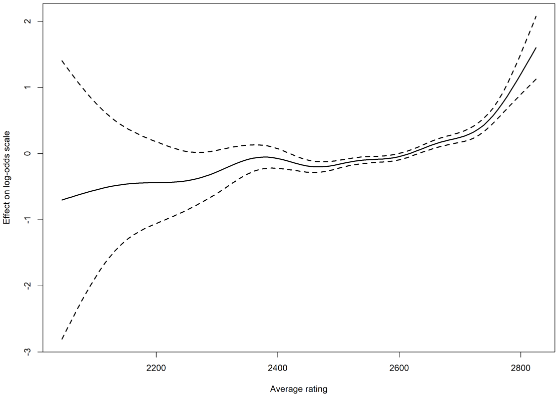
<u>Limitation</u>: if both players' strengths are shifted upward by the same amount, the model's win/draw/loss probabilities are unchanged.

Empirical motivation from chess

Glickman (2025) motivated strength-dependent ties using chess outcomes.

A preliminary analysis of FIDE (International Chess Federation) games found that the fitted log-odds of a draw increased with the average rating of the two players, after controlling for rating difference.

This suggests the draw mechanism is not only about the strength difference between two players, but also their absolute level of strength.



Published extension: linear strength dependence

Let

$$\bar{\theta}_{ij} = \frac{\theta_i + \theta_j}{2}.$$

Published extension: linear strength dependence

Let

$$\bar{\theta}_{ij} = \frac{\theta_i + \theta_j}{2}.$$

Glickman (2025) replaces scalar parameters by linear functions of average strength:

$$\begin{aligned}
\Pr(Y_{ij} = 1) &\propto \exp\left(\theta_i + \frac{x_{ij}(\alpha_0 + \alpha_1 \bar{\theta}_{ij})}{4}\right), \\
\Pr(Y_{ij} = 0) &\propto \exp\left(\theta_j - \frac{x_{ij}(\alpha_0 + \alpha_1 \bar{\theta}_{ij})}{4}\right), \\
\Pr(Y_{ij} = 1/2) &\propto \exp\left(\beta_0 + (1 + \beta_1)\bar{\theta}_{ij}\right).
\end{aligned}$$

Published extension: linear strength dependence

Let

$$\bar{\theta}_{ij} = \frac{\theta_i + \theta_j}{2}.$$

Glickman (2025) replaces scalar parameters by linear functions of average strength:

$$\Pr(Y_{ij} = 1) \propto \exp\left(\theta_i + \frac{x_{ij}(\alpha_0 + \alpha_1 \bar{\theta}_{ij})}{4}\right),$$
$$\Pr(Y_{ij} = 0) \propto \exp\left(\theta_j - \frac{x_{ij}(\alpha_0 + \alpha_1 \bar{\theta}_{ij})}{4}\right),$$
$$\Pr(Y_{ij} = 1/2) \propto \exp\left(\beta_0 + (1 + \beta_1)\bar{\theta}_{ij}\right).$$

<u>Special case</u>: $\alpha_1 = \beta_1 = 0$ is the model by David (1988).

Why move beyond linearity?

The linear draw term can be written as:

$$\eta_{ij,d} = \beta_0 + (1 + \beta_1)\bar{\theta}_{ij}.$$

Why move beyond linearity?

The linear draw term can be written as:

$$\eta_{ij,d} = \beta_0 + (1 + \beta_1)\bar{\theta}_{ij}.$$

Main issue for today: The relationship could be nonlinear.

- draw rates may rise sharply only among very strong players;
- very weak players may also draw frequently, for example if neither player can reliably convert a winning position;
- there may be a plateau, threshold, or U-shaped pattern.

Why move beyond linearity?

The linear draw term can be written as:

$$\eta_{ij,d} = \beta_0 + (1 + \beta_1)\bar{\theta}_{ij}.$$

Main issue for today: The relationship could be nonlinear.

- draw rates may rise sharply only among very strong players;
- very weak players may also draw frequently, for example if neither player can reliably convert a winning position;
- there may be a plateau, threshold, or U-shaped pattern.

Proposal:

$$\eta_{ij,d} = \beta_0 + \bar{\theta}_{ij} + f(\bar{\theta}_{ij}),$$

where f is a spline.

Spline model for draw propensity

Use a fixed knot sequence $\kappa_1 < \cdots < \kappa_K$.

Spline model for draw propensity

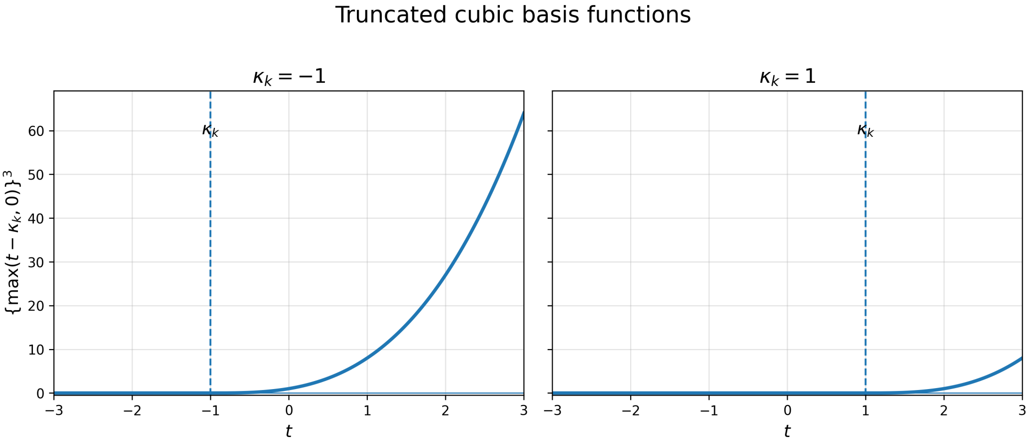
Use a fixed knot sequence $\kappa_1 < \cdots < \kappa_K$.

Can use a truncated cubic spline:

$$f(t) = b_1 t + b_2 t^2 + b_3 t^3 + \sum_{k=1}^{K} \gamma_k (t - \kappa_k)_+^3,$$

where $(t - \kappa_k)_+^3 = \{\max(t - \kappa_k, 0)\}^3$.

Truncated cubic basis components



Each component is zero until t passes its knot κ_k, after which it contributes a cubic term to the spline.

Spline model for draw propensity (continued)

Then

$$\Pr(Y_{ij} = 1/2) \propto \exp\{\beta_0 + \bar{\theta}_{ij} + f(\bar{\theta}_{ij})\}.$$

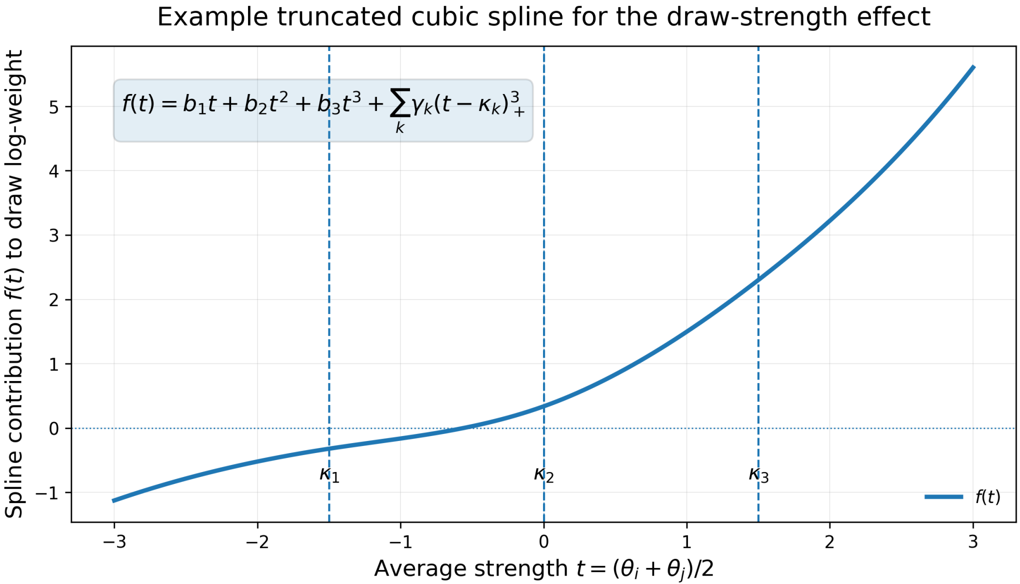
Spline model for draw propensity (continued)

Then

$$\Pr(Y_{ij} = 1/2) \propto \exp\{\beta_0 + \bar{\theta}_{ij} + f(\bar{\theta}_{ij})\}.$$

- $\bar{\theta}_{ij}$ preserves the Davidson (1970) baseline when $f = 0$;
- $f(\bar{\theta}_{ij})$ captures departures from linear strength dependence;
- K controls flexibility of the spline.

Example: truncated cubic spline



Why an unconstrained spline?

The linear model assumes a single direction and rate of change in draw propensity:

$$\beta_0 + (1 + \beta_1)\bar{\theta}_{ij}.$$

The spline model relaxes this while keeping the paired-comparison structure:

$$\beta_0 + \bar{\theta}_{ij} + f(\bar{\theta}_{ij}).$$

Why an unconstrained spline?

The linear model assumes a single direction and rate of change in draw propensity:

$$\beta_0 + (1 + \beta_1)\bar{\theta}_{ij}.$$

The spline model relaxes this while keeping the paired-comparison structure:

$$\beta_0 + \bar{\theta}_{ij} + f(\bar{\theta}_{ij}).$$

- nests the linear draw-strength model as a special case;
- permits curvature without committing to a particular parametric shape;
- keeps the win/loss odds structure unchanged, so the extension targets only the draw component;
- allows model comparison across different choices of K.

Bayesian formulation: likelihood

Let $g = 1, \ldots, G$ index games, and let

$$\mathbf{p}_g = \left(p_{g,L}, p_{g,D}, p_{g,W}\right)$$

denote the model probabilities of White loss, draw, and White win.

Bayesian formulation: likelihood

Let $g = 1, \ldots, G$ index games, and let

$$\mathbf{p}_g = \big(p_{g,L}, p_{g,D}, p_{g,W}\big)$$

denote the model probabilities of White loss, draw, and White win.

The likelihood is a product over game outcomes:

$$L(\Theta) = \prod_{g=1}^{G} p_{g,L}^{I(Y_g=L)} p_{g,D}^{I(Y_g=D)} p_{g,W}^{I(Y_g=W)}.$$

Equivalently, each game contributes one categorical-logit term.

Bayesian formulation: priors

A general Bayesian version specifies strength priors such as

$$\theta_i \mid \mu_i, \sigma_i^2 \sim N(\mu_i, \sigma_i^2), \qquad i = 1, \ldots, n.$$

The non-strength parameters receive regularizing priors:

$$\alpha_0, \alpha_1, \beta_0, \boldsymbol{\eta}, \qquad \boldsymbol{\eta} = (b_1, b_2, b_3, \gamma_1, \ldots, \gamma_K).$$

The vector $\boldsymbol{\eta}$ contains the coefficients in

$$f(\bar{\theta}_{ij}) = b_1\bar{\theta}_{ij} + b_2\bar{\theta}_{ij}^2 + b_3\bar{\theta}_{ij}^3 + \sum_{k=1}^{K} \gamma_k(\bar{\theta}_{ij} - \kappa_k)_+^3.$$

Data: US Chess Open tournaments

Application data:

- US Chess Open tournaments, 2006–2019;
- 24,888 games;
- 6,075 player-year entries (treated same player in different years as different players);
- player IDs, color assignment, game outcome, and pre-tournament rating when available.

Data: US Chess Open tournaments

Application data:

- US Chess Open tournaments, 2006–2019;
- 24,888 games;
- 6,075 player-year entries (treated same player in different years as different players);
- player IDs, color assignment, game outcome, and pre-tournament rating when available.

Why useful:

- broad range of strengths;
- many games per player relative to weekend events;
- Swiss pairings produce both uneven and close matchups.

Informative strength priors for the US Open application

Most players enter the US Open with a pre-tournament US Chess rating.

For rated players, use the rating as prior information:

$$\theta_i \sim N(\mu_i, \sigma^2), \qquad \mu_i = \frac{R_i - \bar{R}}{400/\log 10} \quad \text{(centered rating scale)}$$

For unrated players:

$$\theta_i \sim N(\mu_{\mathrm{miss}}, \sigma^2_{\mathrm{miss}}).$$

Informative strength priors for the US Open application

Most players enter the US Open with a pre-tournament US Chess rating.

For rated players, use the rating as prior information:

$$\theta_i \sim N(\mu_i, \sigma^2), \qquad \mu_i = \frac{R_i - \bar{R}}{400/\log 10} \quad \text{(centered rating scale)}$$

For unrated players:

$$\theta_i \sim N(\mu_{\text{miss}}, \sigma^2_{\text{miss}}).$$

This separates two sources of uncertainty:

- uncertainty around rated players' pre-tournament strengths (likely small);
- uncertainty about the population of unrated players (likely large).

Implementation in R and Stan

All analyses were run in R and Stan.

Implementation in R and Stan

All analyses were run in R and Stan.

The spline basis cannot be precomputed in R because

$$\bar{\theta}_{ij} = \frac{\theta_i + \theta_j}{2}$$

changes at every MCMC iteration.

Implementation in R and Stan

All analyses were run in R and Stan.

The spline basis cannot be precomputed in R because

$$\bar{\theta}_{ij} = \frac{\theta_i + \theta_j}{2}$$

changes at every MCMC iteration.

Implementation strategy:

- choose knot locations in R from average pre-tournament ratings;
- pass the knot vector to Stan as data;
- evaluate the truncated cubic spline inside Stan at each sampled $\bar{\theta}_{ij}$;
- compute pointwise log-likelihood values for LOO comparison.

Implementation in R and Stan

All analyses were run in R and Stan.

The spline basis cannot be precomputed in R because

$$\bar{\theta}_{ij} = \frac{\theta_i + \theta_j}{2}$$

changes at every MCMC iteration.

Implementation strategy:

- choose knot locations in R from average pre-tournament ratings;
- pass the knot vector to Stan as data;
- evaluate the truncated cubic spline inside Stan at each sampled $\bar{\theta}_{ij}$;
- compute pointwise log-likelihood values for LOO comparison.

Models were fit in Stan using three parallel chains, each run for 4,000 iterations with the first 2,000 discarded as warm-up.

MCMC diagnostics and convergence checks

Convergence checks focused on the non-strength parameters:

$$\alpha_0, \alpha_1, \beta_0, \boldsymbol{\eta}, \sigma, \mu_{\text{miss}}, \sigma_{\text{miss}}.$$

MCMC diagnostics and convergence checks

Convergence checks focused on the non-strength parameters:

$$\alpha_0, \alpha_1, \beta_0, \boldsymbol{\eta}, \sigma, \mu_{\text{miss}}, \sigma_{\text{miss}}.$$

For each fitted model:

- convergence was assessed using $\widehat{R}$;
- trace plots were inspected after the 2,000-iteration warm-up;
- pointwise log-likelihood values were saved for LOO comparison;
- posterior summaries of θ_i were saved for plotting and interpretation.

Choosing the number of knots

Candidate values:

$$K \in \{2, 3, 4, 5, 6\}.$$

For each K:

1. choose knots from quantiles of average pre-tournament strength;
2. fit the Bayesian model in Stan;
3. compute pointwise log-likelihood values;
4. use approximate LOO-CV to estimate out-of-sample predictive accuracy.

Choosing the number of knots

Candidate values:

$$K \in \{2, 3, 4, 5, 6\}.$$

For each K:

1. choose knots from quantiles of average pre-tournament strength;
2. fit the Bayesian model in Stan;
3. compute pointwise log-likelihood values;
4. use approximate LOO-CV to estimate out-of-sample predictive accuracy.

Guiding principle: prefer the simplest spline that captures the main draw-rate pattern without unnecessary flexibility.

LOO comparison across values of K

$$\text{elpd}_{\text{loo}} = \sum_{g=1}^{G} \log p(y_g \mid y_{-g})$$

K	elpd diff	SE diff	elpd_{loo}	p_{loo}
5	0.0	0.0	-19878.6	1648.3
6	-0.9	1.1	-19879.5	1650.3
4	-1.8	2.1	-19880.4	1639.6
2	-2.0	2.8	-19880.6	1626.9
3	-2.6	2.2	-19881.2	1633.9

The differences are small relative to their standard errors, but $K = 5$ has the highest estimated predictive accuracy and provides reasonable flexibility.

LOO comparison across values of K

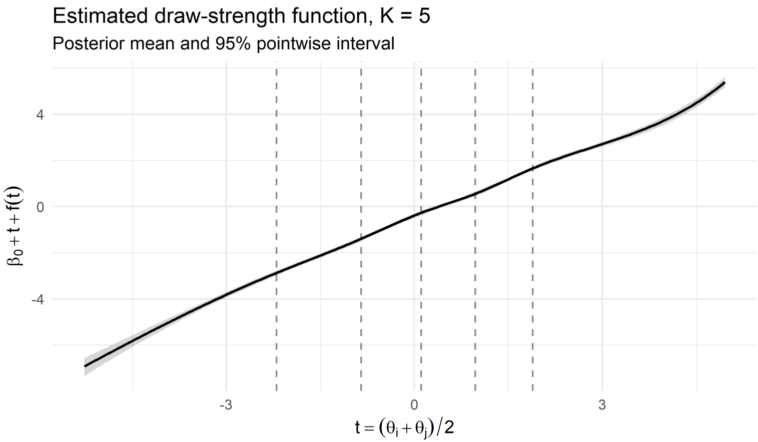
$$\text{elpd}_{\text{loo}} = \sum_{g=1}^{G} \log p(y_g \mid y_{-g})$$

K	elpd diff	SE diff	elpd_{loo}	p_{loo}
5	0.0	0.0	-19878.6	1648.3
6	-0.9	1.1	-19879.5	1650.3
4	-1.8	2.1	-19880.4	1639.6
2	-2.0	2.8	-19880.6	1626.9
3	-2.6	2.2	-19881.2	1633.9

The differences are small relative to their standard errors, but $K = 5$ has the highest estimated predictive accuracy and provides reasonable flexibility.

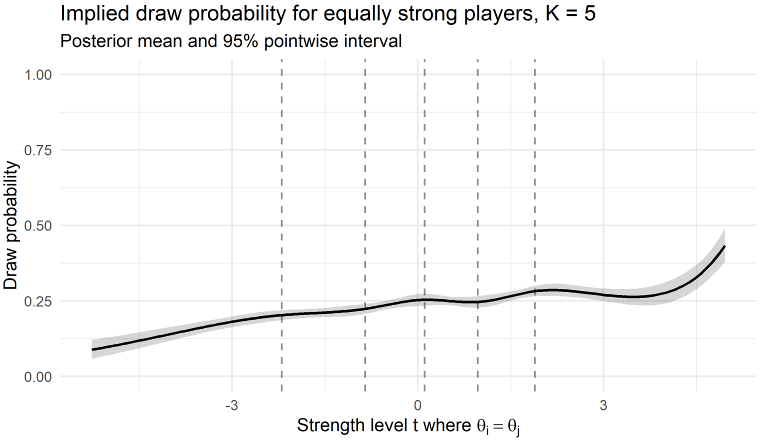
Use $K = 5$ as the main specification, while treating the nearby values as sensitivity checks.

Estimated spline function



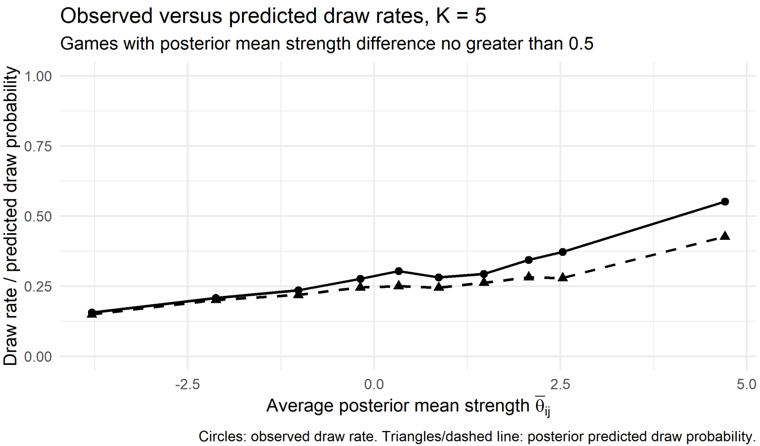
The estimated draw log-weight increases with average player strength, with only modest visible curvature beyond the linear trend.

Draw probability for evenly matched players



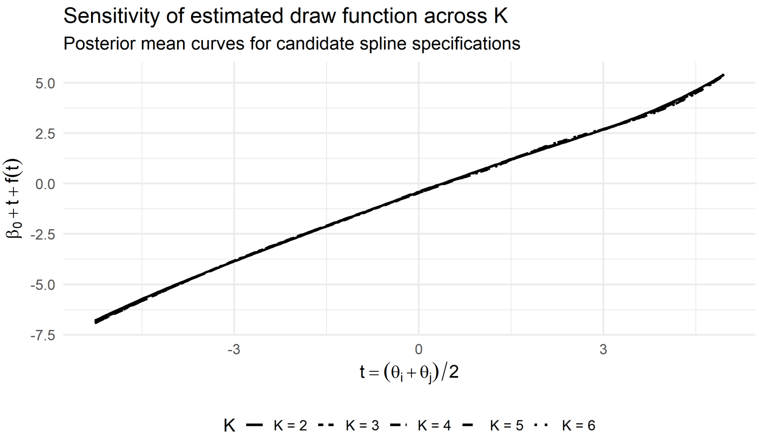
For two equally strong players, the implied draw probability rises gradually across most of the strength range and increases more sharply among the strongest players.

Observed versus predicted draw rates



The model captures the broad upward trend in draw rates, but still underpredicts draws among the strongest nearly evenly matched players.

Sensitivity across K



The estimated draw-strength function is highly stable across the candidate values of K, suggesting that the main pattern is not driven by knot choice.

What did the spline extension add?

Model selection: $K = 5$ had the largest estimated elpd, but nearby values of K gave nearly indistinguishable fitted curves.

Shape of the effect: The fitted draw-strength relationship is mostly increasing, with modest curvature over most of the observed strength range.

Substantive scale: For equally strong players, draw probability rises from roughly 0.10 among weak players to over 0.40 among the strongest players.

Model adequacy: The model captures the broad upward trend, but still underpredicts draws in the highest-strength bins.

Interpretation: the spline mainly confirms the strength-dependent draw effect, while suggesting only modest departures from linearity.

Conclusions

1. Absolute strength matters for draw propensity.
2. A spline draw term relaxes linearity while preserving the paired-comparison structure.
3. In the US Open data, the dominant pattern is increasing draw probability with player strength.
4. The spline provides robustness: the evidence for strength-dependent draws is not sensitive to the exact number of knots.

References

- Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324–345.
- Davidson, R. R. (1970). On extending the Bradley–Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329), 317–328.
- David, H. A. (1988). *The Method of Paired Comparisons*, 2nd revised ed. London: Charles Griffin.
- Glickman, M. E. (2025). Paired comparison models with strength-dependent ties and order effects. *Statistical Modeling*. OnlineFirst. doi:10.1177/1471082X251400474.
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432.

Thanks for listening