Supplementary Material

This supplementary document provides additional model development details, diagnostic plots, and extended results that support the main manuscript. Figures and tables are numbered sequentially as S1, S2, etc., and are referenced in the order they are discussed.

1. Additional Model Structures and Equations

1.1. Model 2: Hierarchical Linear Regression (HLR)

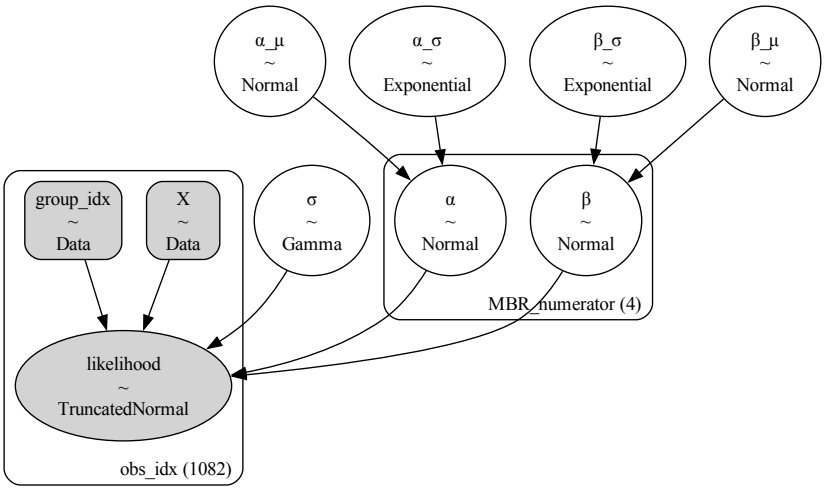

Figure S1: Model 2 Directed Acyclic Graph (DAG). Group-specific intercepts α_g and slopes β_g are modeled hierarchically by MBR numerator group g. Gaussian priors are used for all group-level parameters, with hyperpriors on their means and standard deviations to allow partial pooling. A shared dispersion parameter σ is modeled with a Gamma prior. Observations are modeled via a truncated normal distribution.

The model specification is:

$$\alpha_\mu \sim \mathcal{N}(0, 1), \qquad \alpha_\sigma \sim \text{Exponential}(1), \tag{1}$$

$$\beta_\mu \sim \mathcal{N}(0, 1), \qquad \beta_\sigma \sim \text{Exponential}(1), \tag{2}$$

$$\sigma_\gamma \sim \text{Exponential}(1), \qquad \sigma_\phi \sim \text{Exponential}(1), \tag{3}$$

$$\alpha_g \sim \mathcal{N}(\alpha_\mu, \alpha_\sigma), \qquad \beta_g \sim \mathcal{N}(\beta_\mu, \beta_\sigma), \tag{4}$$

$$\sigma \sim \text{Gamma}(\sigma_\gamma, \sigma_\phi), \tag{5}$$

$$\mu_i = \alpha_{g[i]} + \beta_{g[i]}\, x_i, \tag{6}$$

$$y_i \sim \text{TruncatedNormal}(\mu_i, \sigma, -3, 3.2). \tag{7}$$

1.2. Model 3: HLR with Group-Specific Dispersion

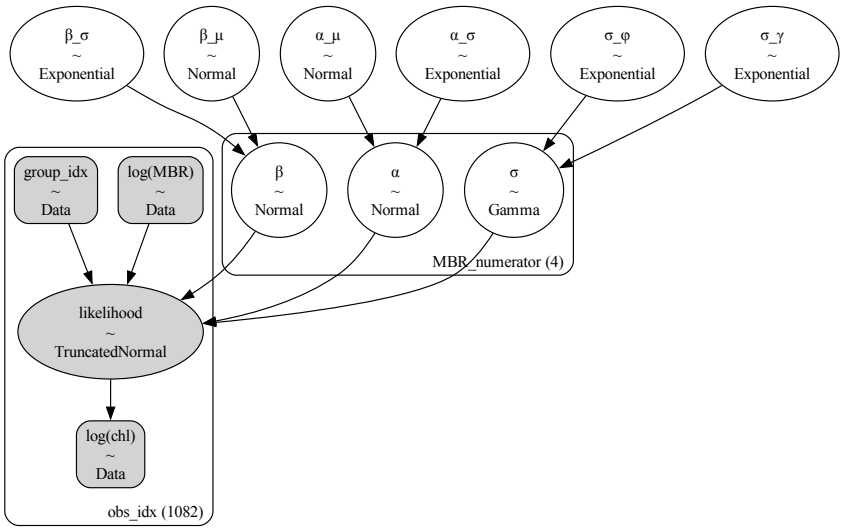

Figure S2: Model 3 DAG. Extends Model 2 by allowing the dispersion parameter σ_g to vary by MBR numerator group, modeled with a Gamma prior for each group.

The model specification is:

$$\alpha_\mu \sim \mathcal{N}(0, 1), \qquad \alpha_\sigma \sim \text{Exponential}(1), \tag{8}$$

$$\beta_\mu \sim \mathcal{N}(0, 1), \qquad \beta_\sigma \sim \text{Exponential}(1), \tag{9}$$

$$\sigma_\gamma \sim \text{Exponential}(1), \qquad \sigma_\phi \sim \text{Exponential}(1), \tag{10}$$

$$\alpha_g \sim \mathcal{N}(\alpha_\mu, \alpha_\sigma), \qquad \beta_g \sim \mathcal{N}(\beta_\mu, \beta_\sigma), \tag{11}$$

$$\sigma_g \sim \text{Gamma}(\sigma_\gamma, \sigma_\phi), \tag{12}$$

$$\mu_i = \alpha_{g[i]} + \beta_{g[i]}\, x_i, \tag{13}$$

$$y_i \sim \text{TruncatedNormal}(\mu_i, \sigma_{g[i]}, -3, 3.2). \tag{14}$$

1.3. Model 4: HLR with Input-Dependent Dispersion

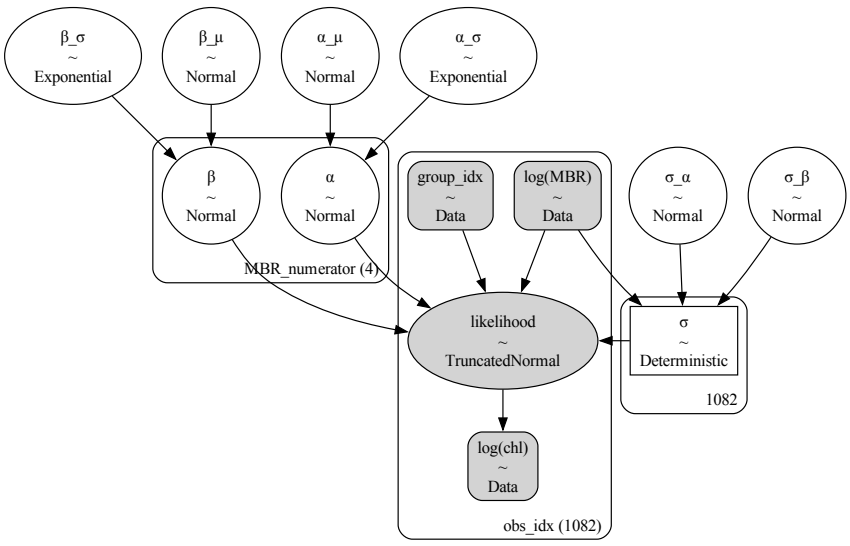

Figure S3: Model 4 DAG. Extends Model 2 by modeling the log-dispersion as a linear function of log(MBR) with shared slope and intercept, allowing variance to change with the predictor.

The model specification is:

$$\alpha_\mu \sim \mathcal{N}(0, 1), \qquad \alpha_\sigma \sim \text{Exponential}(1), \tag{15}$$

$$\beta_\mu \sim \mathcal{N}(0, 1), \qquad \beta_\sigma \sim \text{Exponential}(1), \tag{16}$$

$$\alpha_g \sim \mathcal{N}(\alpha_\mu, \alpha_\sigma), \qquad \beta_g \sim \mathcal{N}(\beta_\mu, \beta_\sigma), \tag{17}$$

$$\sigma_\alpha \sim \mathcal{N}(0, 1), \qquad \sigma_\beta \sim \mathcal{N}(0, 1), \tag{18}$$

$$\mu_i = \alpha_{g[i]} + \beta_{g[i]}\, x_i, \tag{19}$$

$$\log \sigma_i = \sigma_\alpha + \sigma_\beta\, x_i, \tag{20}$$

$$y_i \sim \text{TruncatedNormal}(\mu_i, e^{\log \sigma_i}, -3, 3.2). \tag{21}$$

1.4. Model 5: Heteroskedastic HLR with Group-Specific Dispersion Slopes

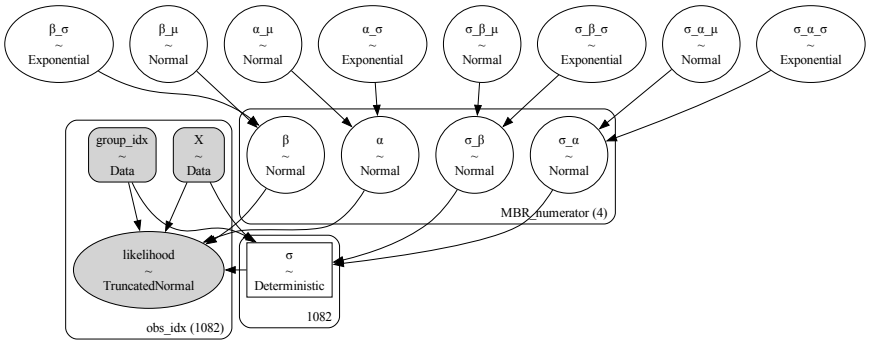

Figure S4: Model 5 DAG. Extends Model 4 by allowing both the intercept and slope of the log-dispersion relationship to vary by MBR numerator group, enabling group-specific heteroskedasticity.

The model specification is:

$$\alpha_\mu \sim \mathcal{N}(0,1), \qquad \alpha_\sigma \sim \text{Exponential}(1), \tag{22}$$
$$\beta_\mu \sim \mathcal{N}(0,1), \qquad \beta_\sigma \sim \text{Exponential}(1), \tag{23}$$
$$\sigma_\alpha^\mu \sim \mathcal{N}(0,1), \qquad \sigma_\alpha^\sigma \sim \text{Exponential}(1), \tag{24}$$
$$\sigma_\beta^\mu \sim \mathcal{N}(0,1), \qquad \sigma_\beta^\sigma \sim \text{Exponential}(1), \tag{25}$$
$$\alpha_g \sim \mathcal{N}(\alpha_\mu, \alpha_\sigma), \qquad \beta_g \sim \mathcal{N}(\beta_\mu, \beta_\sigma), \tag{26}$$
$$\sigma_{\alpha,g} \sim \mathcal{N}(\sigma_\alpha^\mu, \sigma_\alpha^\sigma), \qquad \sigma_{\beta,g} \sim \mathcal{N}(\sigma_\beta^\mu, \sigma_\beta^\sigma), \tag{27}$$
$$\mu_i = \alpha_{g[i]} + \beta_{g[i]}\, x_i, \tag{28}$$
$$\log \sigma_i = \sigma_{\alpha,g[i]} + \sigma_{\beta,g[i]}\, x_i, \tag{29}$$
$$y_i \sim \text{TruncatedNormal}(\mu_i, e^{\log \sigma_i}, -3, 3.2). \tag{30}$$

2. Posterior Diagnostics: Trace plots and Forest Plots

This section presents Markov chain Monte Carlo (MCMC) diagnostics for all five models. For each model, I show:

1. **Trace plots** for all parameters, displaying individual chain trajectories and combined posterior densities. Well-mixed, stationary chains with substantial overlap between chains indicate stable sampling.
2. **Forest plots** for most model parameters, displaying posterior distributions (central tendency and dispersion), effective sample sizes (ESS), and Gelman–Rubin convergence diagnostics ($\hat{R}$). For Models 4 and 5, dispersion parameters σ_i are excluded due to their being observation-specific.

2.1. Trace Plots

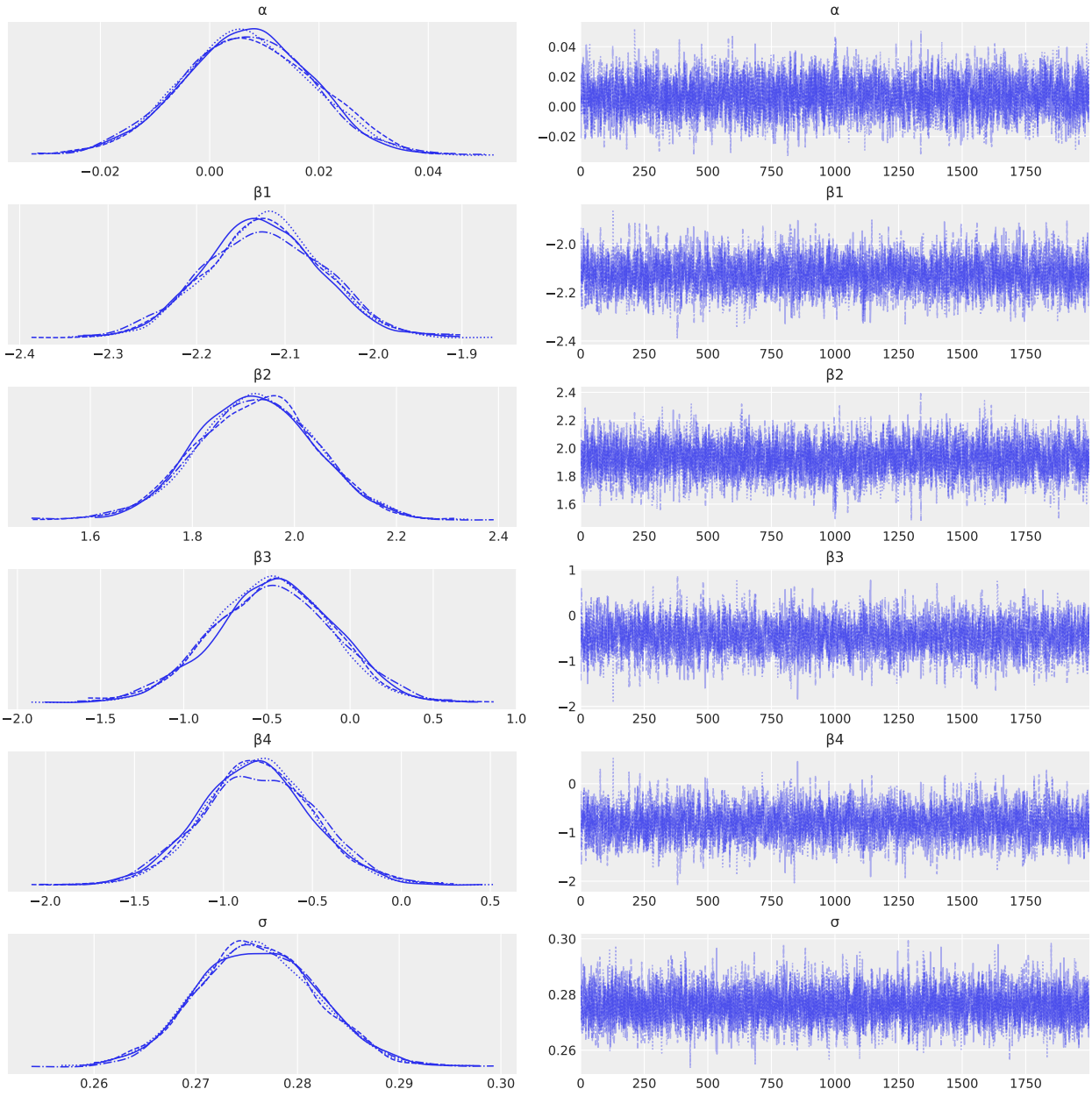

Figure S5: Traceplots for **Model 1** showing MCMC sampling behavior for all parameters. Left sub-panels show the parameter posterior densities. Right corresponding sub-panels display chain trajectories; well-mixed and stationary patterns with substantial overlap across chains indicate stable sampling and good convergence.

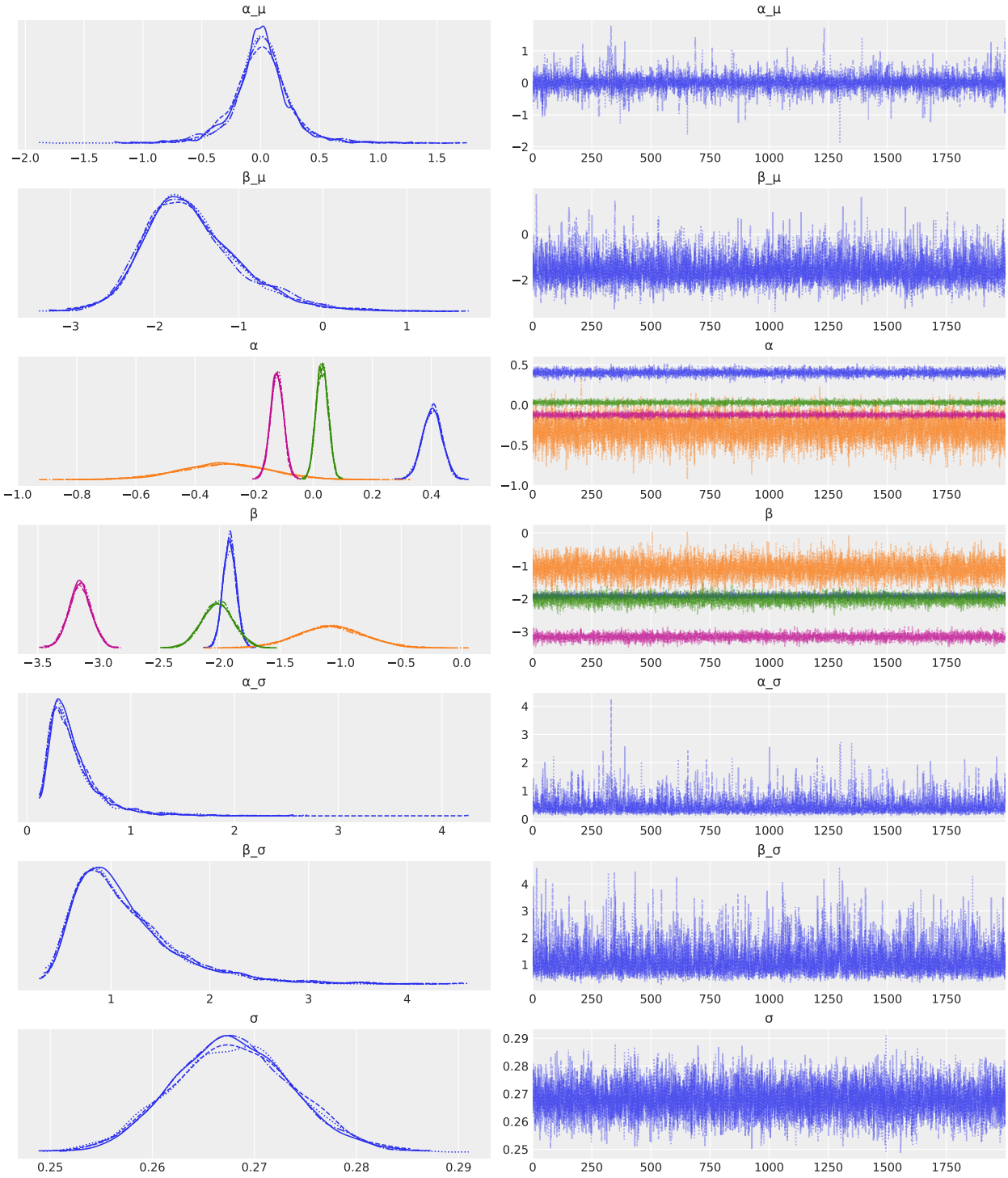

Figure S6: Traceplots for **Model 2** showing MCMC sampling behavior for all parameters, including group-indexed effects where applicable. Sub-panels show parameter posterior densities. Right sub-panels display chain trajectories; well-mixed and stationary patterns with substantial overlap across chains indicate stable sampling and good convergence. For hierarchical parameters, separate chains are shown for each group-level effect, allowing visual assessment of convergence across all levels.

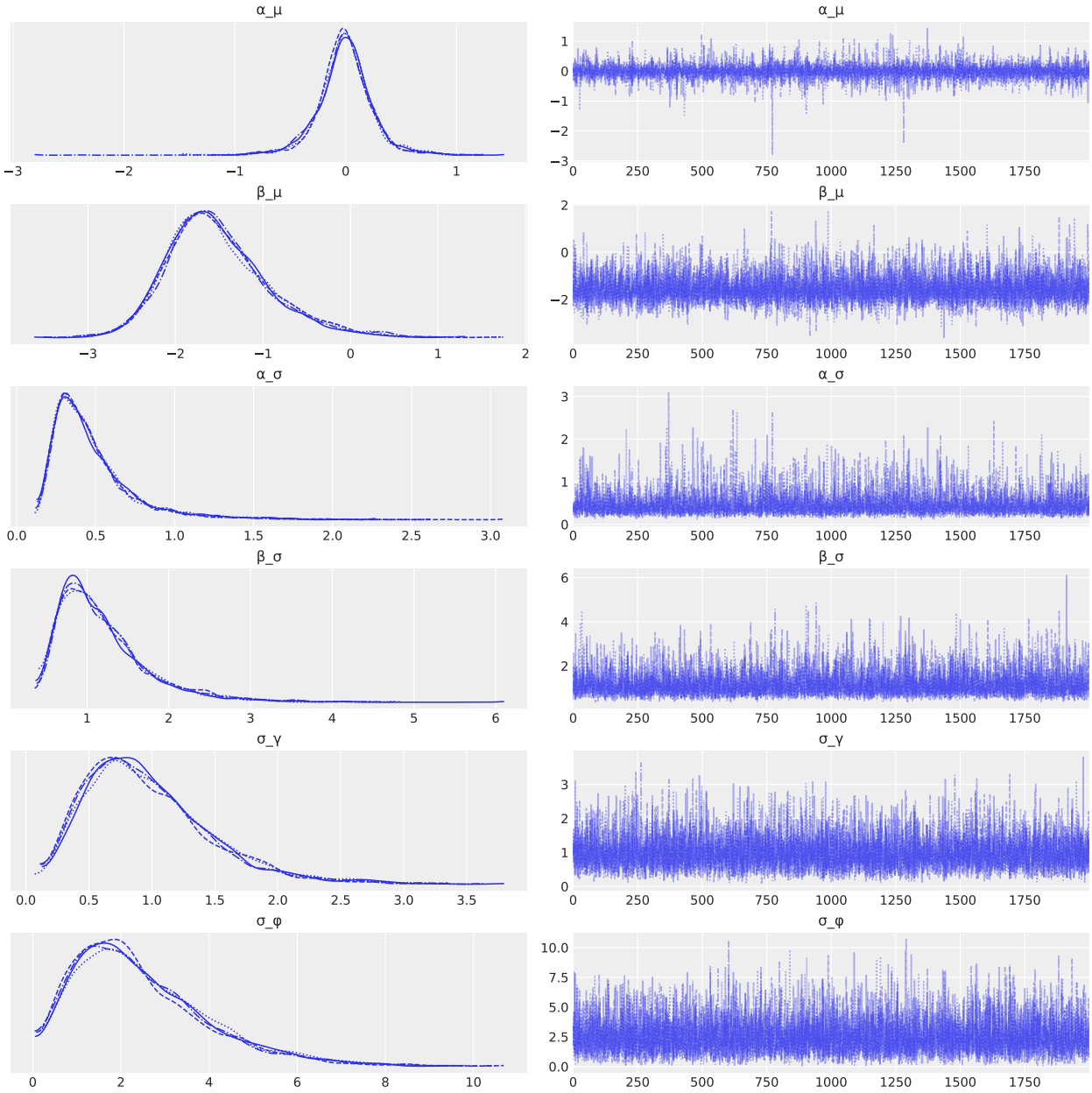

Figure S7: Traceplots for **Model 3** (Part 1 of 2): *hyperparameters*. Posterior densities (left) and MCMC chains (right) for hyperparameters governing the group-indexed mean parameters, $(\alpha_\mu, \alpha_\sigma, \beta_\mu, \beta_\sigma)$, and for the dispersion hyperparameters $(\sigma_\gamma, \sigma_\phi)$. Chains exhibit stable mixing and convergence across all hyperparameters.

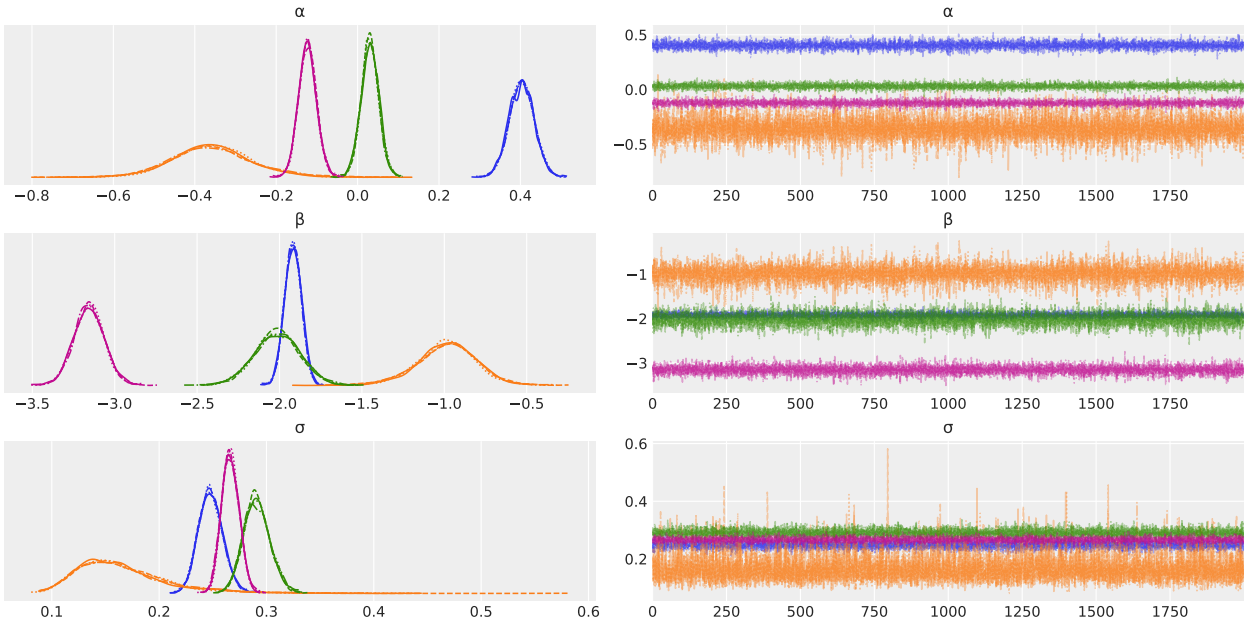

Figure S8: Traceplots for **Model 3** (Part 2 of 2): *group-level parameters*. Posterior densities (left) and MCMC chains (right) for α (vector of group-indexed intercepts), β (vector of group-indexed slopes), and σ (vector of group-indexed dispersions). Colors distinguish MBR numerator groups for vector-valued parameters; the mapping is omitted as the traceplots are intended for diagnosing mixing and convergence rather than comparing magnitudes (see forest plots for inter-group comparisons).

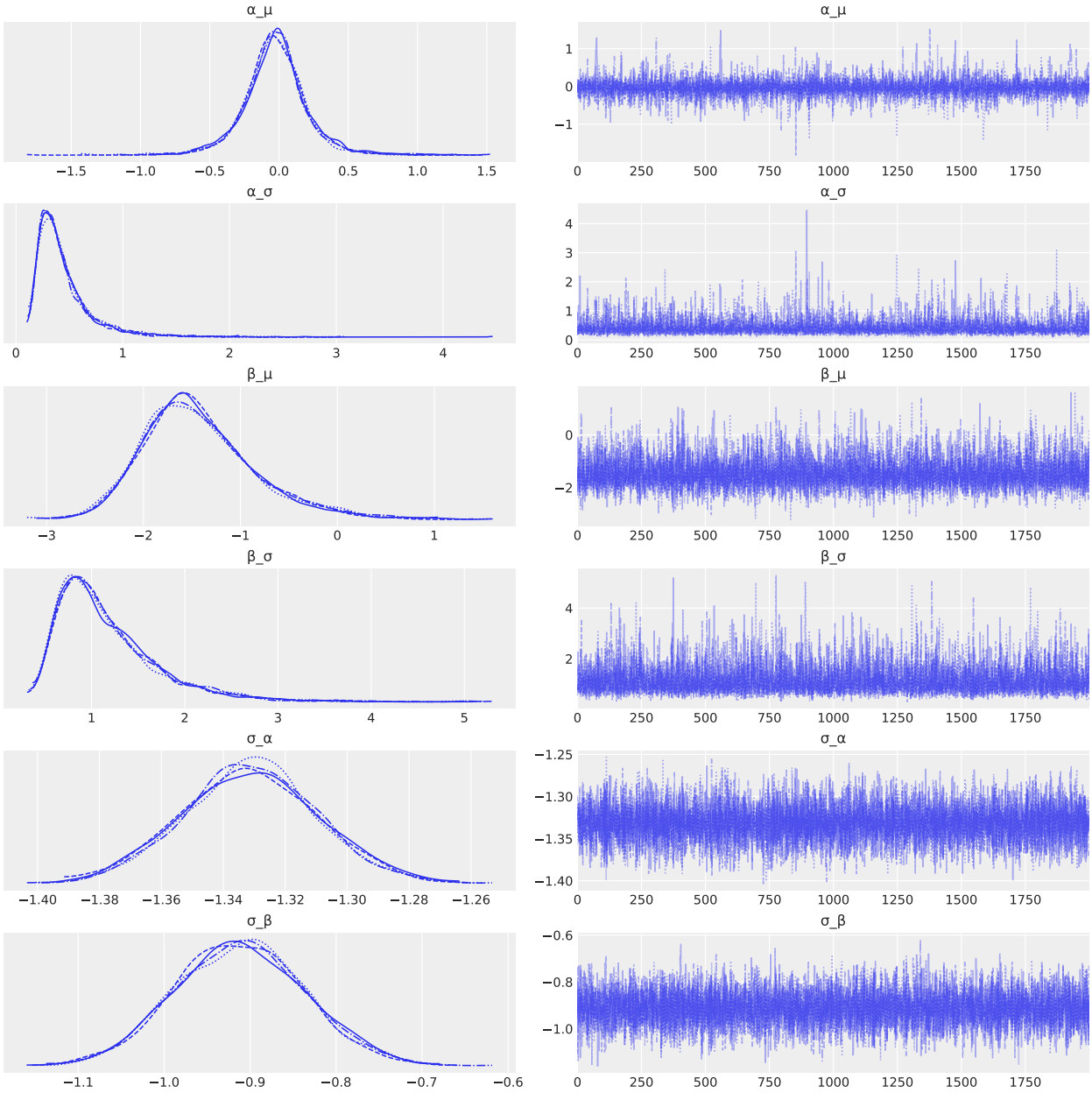

Figure S9: Traceplots for **Model 4** (Part 1 of 2): *hyperparameters and global variance-function coefficients*. Panels show posterior densities (left column) and MCMC chain trajectories (right column) for hyperparameters governing the group-level intercepts and slopes $(\alpha_\mu, \alpha_\sigma, \beta_\mu, \beta_\sigma)$, together with the global variance-function coefficients $(\sigma_\alpha, \sigma_\beta)$ from $\log \sigma_i = \sigma_\alpha + \sigma_\beta \log(\mathrm{MBR}_i)$.

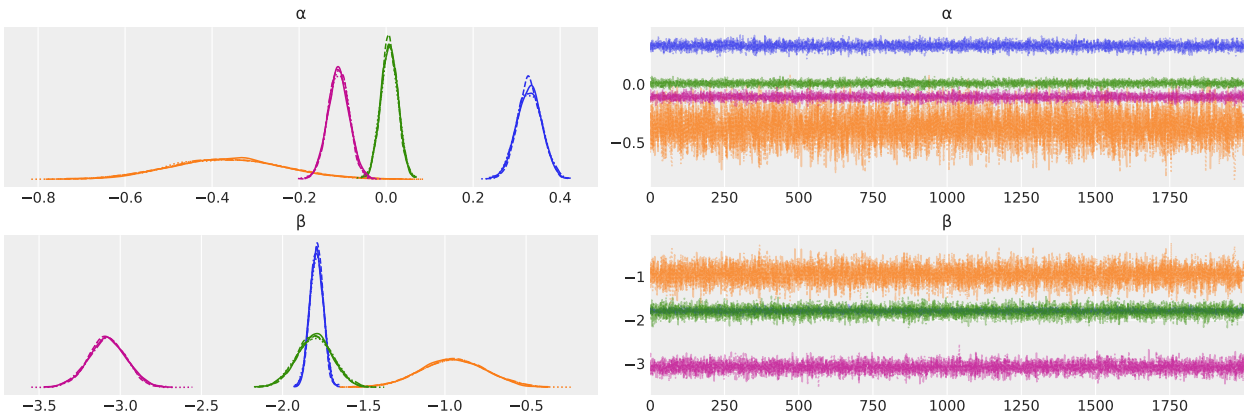

Figure S9: Traceplots for **Model 4** (Part 2 of 2): *group-level mean parameters*. Posterior densities (left) and MCMC chains (right) are shown for group-indexed intercepts and slopes (α, β) by MBR numerator group. Colors denote groups; per-panel color–group mapping is not shown since the purpose is to diagnose mixing and convergence rather than to compare group magnitudes.

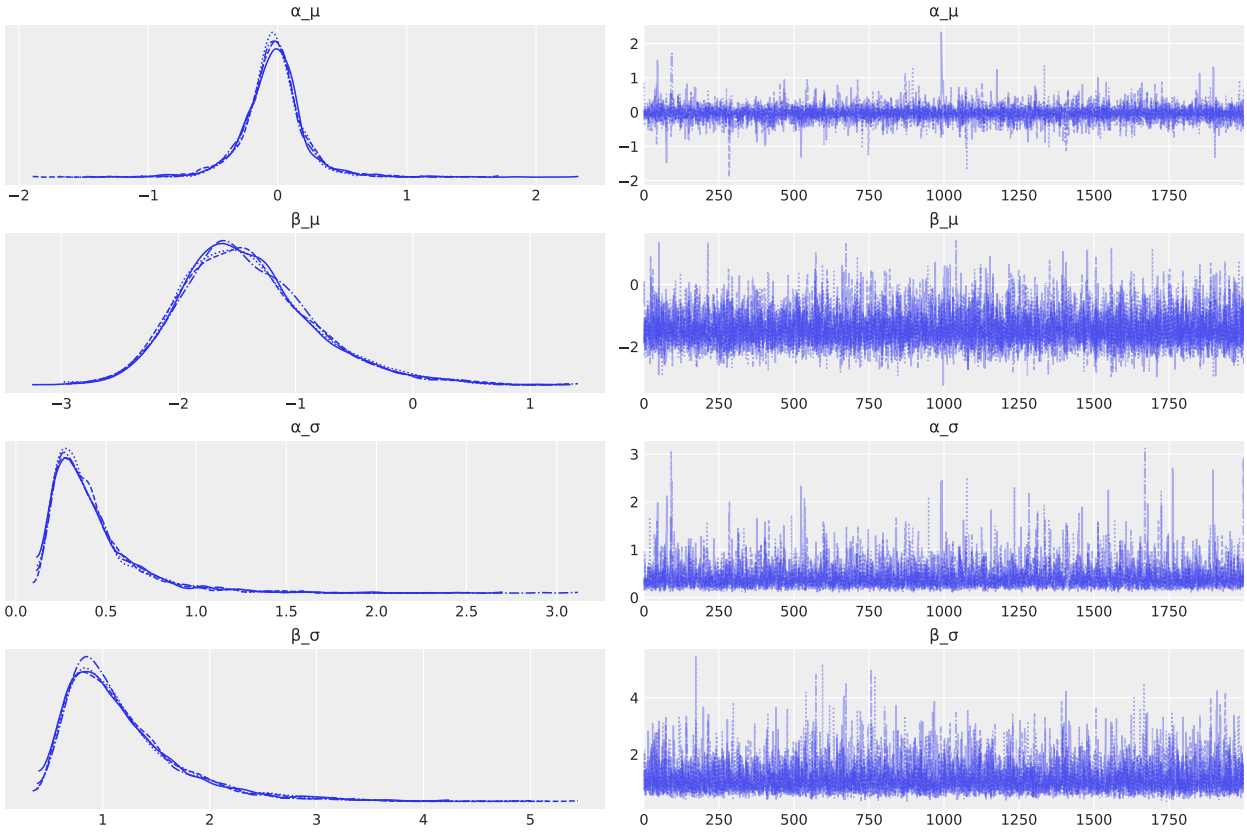

Figure S10: Traceplots for **Model 5** (Part 1 of 3): *hyperparameters for group-indexed mean parameters.* Posterior densities (left) and MCMC chains (right) for $(\alpha_\mu, \alpha_\sigma, \beta_\mu, \beta_\sigma)$, which govern the vectors α (group-indexed intercepts) and β (group-indexed slopes). Colors within vector-valued panels correspond to MBR numerator groups; the mapping is omitted because these plots are for mixing/convergence diagnostics (see forest plots for inter-group comparisons).

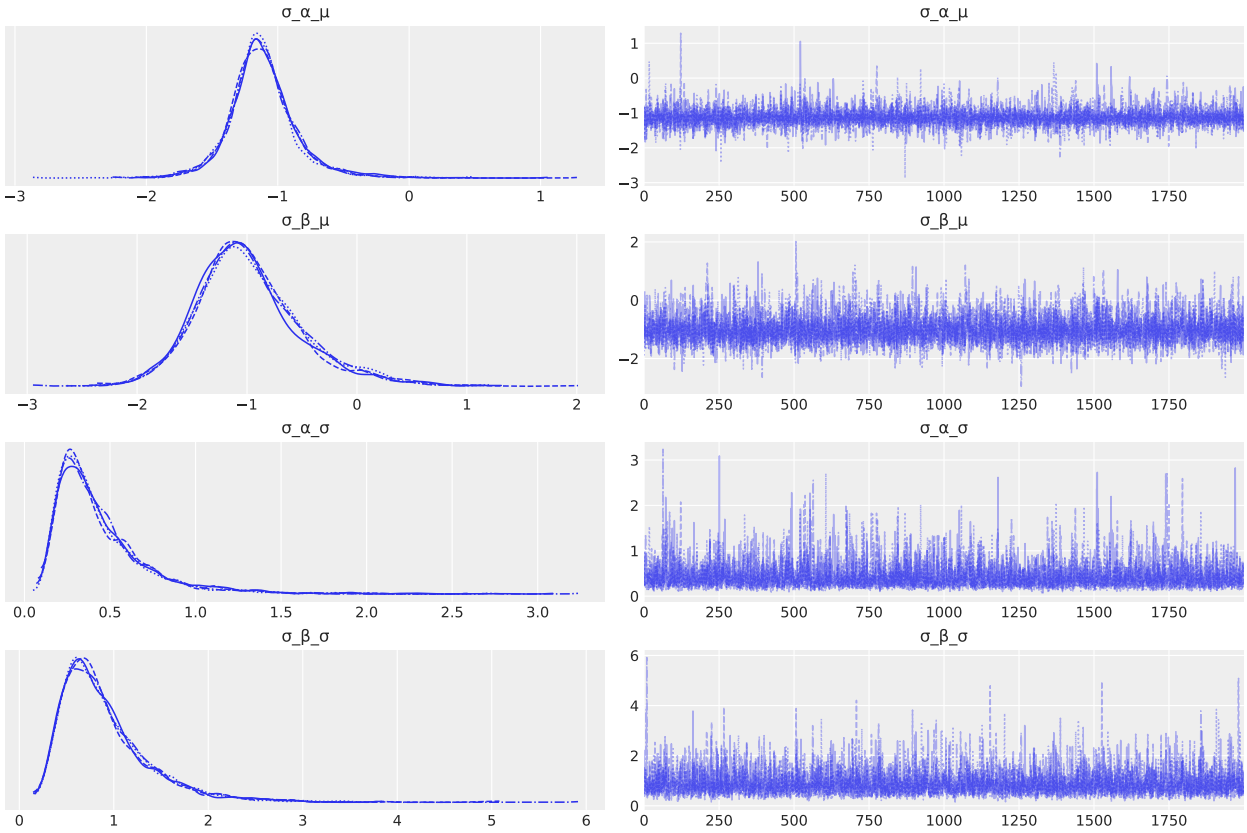

Figure S10: Traceplots for **Model 5** (Part 2 of 3): *hyperparameters for group-indexed variance-function parameters*. Posterior densities (left) and chains (right) for $(\sigma_\alpha^\mu, \sigma_\alpha^\sigma, \sigma_\beta^\mu, \sigma_\beta^\sigma)$, which govern the vectors σ_α and σ_β; *cf.* S10 part 3.

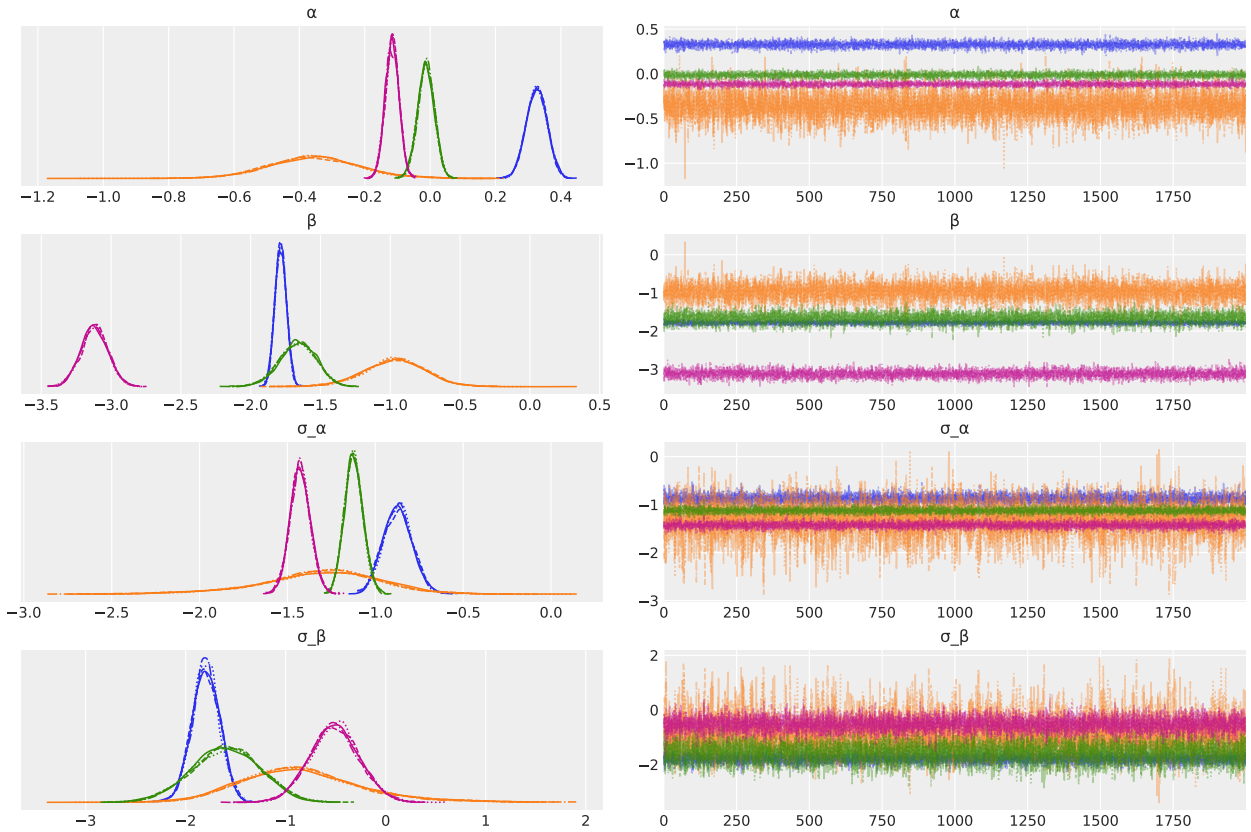

Figure S10: Traceplots for **Model 5** (Part 3 of 3): *group-indexed parameters*. Colors denote MBR numerator groups for vector-valued parameters. Posterior densities (left) and chains (right) for α (vector of group-indexed intercepts), β (vector of group-indexed slopes), and the variance-function vectors σ_α and σ_β (one intercept and slope per MBR numerator group) in $\log \sigma[\text{group}_i] = \sigma_\alpha[\text{group}_i] + \sigma_\beta[\text{group}_i] \log(\text{MBR}[\text{group}_i])$. This yields observations specific σ_j; not shown to maintain clarity, as including hundreds of traces would obscure the primary diagnostics without improving convergence assessment.

2.2. Forest Plots

In the forest plots, the left panel shows each parameter's posterior distribution with median (point) and highest density interval (whiskers). Note that for clarity only first-level parameters are shown, not their hyperparameters. Moreover, Model 4 and 5 have σ posteriors for each observation, and these are omitted from all plots, again for clarity. The middle panel displays the bulk ESS for each depicted parameter; values well above 1000 indicate stable estimation despite potential autocorrelation in chains. The right panel shows $\hat{R}$ values, with values close to 1.00 indicating convergence across chains.

Some posterior distributions have intervals overlapping zero. In a Bayesian context, this means that the posterior includes both positive and negative values with non-negligible probability, and thus the direction of the effect is not strongly supported by the data under the current model and priors.

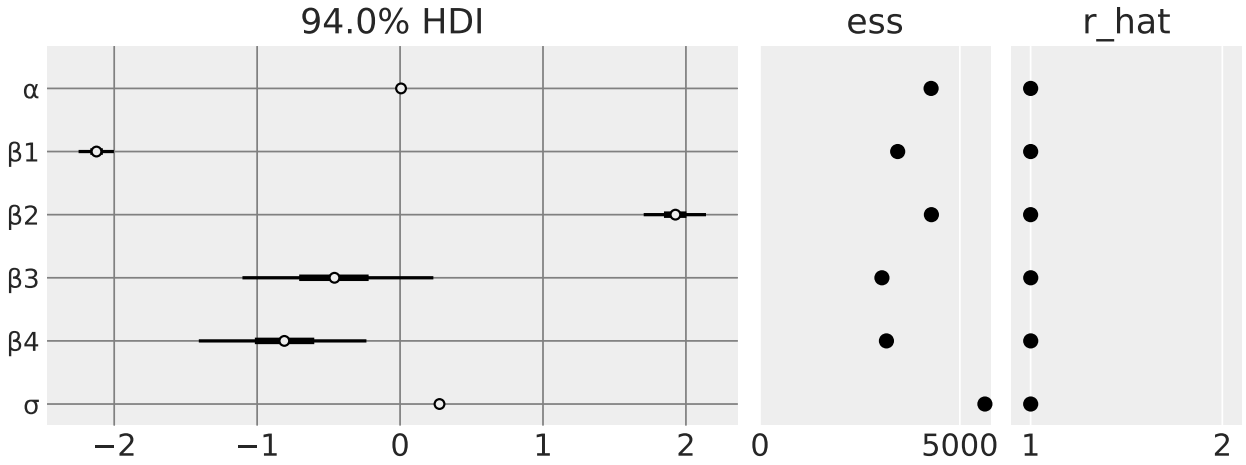

Figure S11: Forest plots of posterior distributions for Model 1. Central dots indicate posterior medians, thick bars show the inter-quartile range (25th–75th percentiles), and thin bars denote the 94% highest density interval (HDI). Effective sample sizes (ESS) and $\hat{R}$ convergence diagnostics are shown in adjacent panels, confirming robust sampling performance.

Interpretation Notes for Forest Plots

- **Relative CI length (all parameters)**: Shorter CIs imply greater certainty in the parameter estimate; longer CIs signal less information in the data about that parameter or stronger regularization from the prior.
- **Intercepts (α_g)**: If a group's credible interval (CI) overlaps zero, this indicates that the baseline log(Chl) for that group could plausibly be near $10^0 = 1$ mg m^{-3}. .
- **Slopes (β_g)**: If a slope's CI overlaps zero, the model cannot rule out the possibility of no linear relationship between log(MBR) and log(Chl) for that group. Wide overlap suggests high uncertainty in how strongly MBR explains chlorophyll variability in that group.
- **Shared σ**: For models with a common dispersion parameter, the CI length reflects uncertainty in the overall noise level. Narrow intervals imply more confidence in the model's estimate of residual variance.

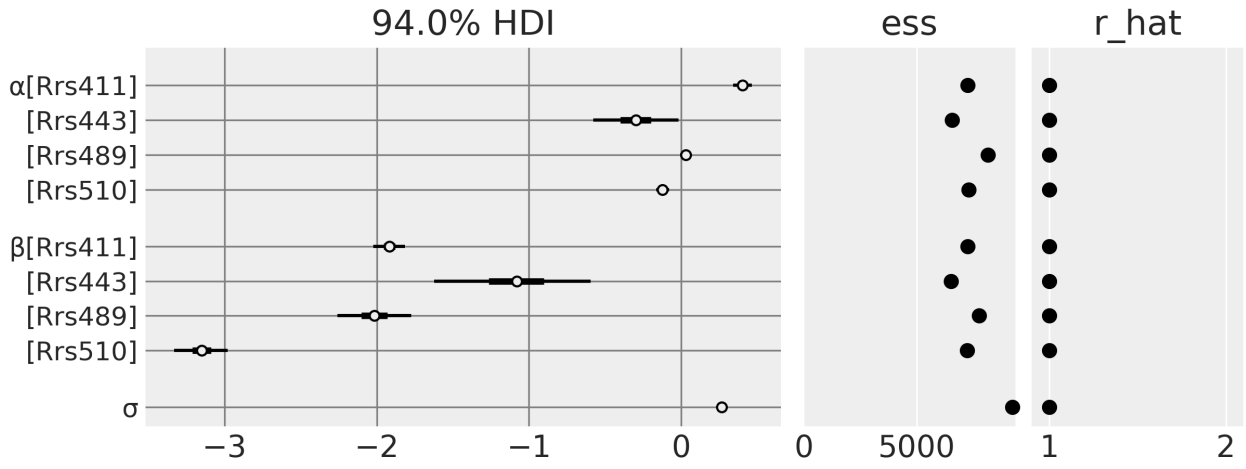

Figure S12: Forest plot for **Model 2**. Group-indexed parameters α (intercepts) and β (slopes) have one value per MBR numerator group. The shared dispersion parameter σ is shown on a single row. Posterior intervals overlapping zero indicate weakly informed parameters. ESS and $\hat{R}$ values confirm adequate sampling and convergence.

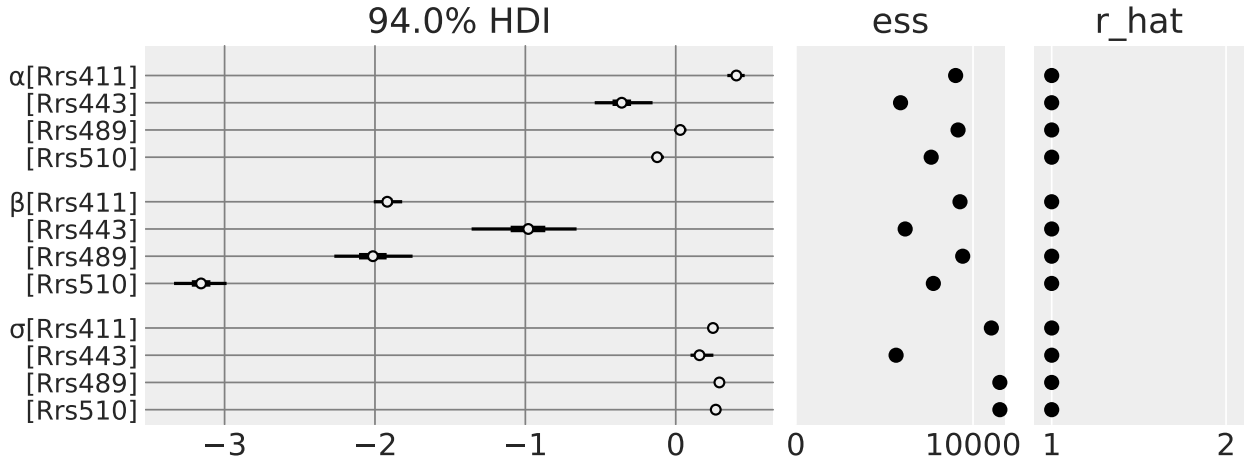

Figure S13: Forest plot for **Model 3**. Three-panel layout (posterior intervals, ESS, $\hat{R}$) for hyperparameters $(\alpha_\mu, \beta_\mu, \alpha_\sigma, \beta_\sigma, \sigma_\gamma, \sigma_\phi)$ and group-indexed parameters (α, β, σ). Group-indexed vectors show one value per MBR numerator group, with colors distinguishing groups. Posterior intervals overlapping zero again indicate limited information for some coefficients, while ESS and $\hat{R}$ panels show good effective sample size and convergence.

- **Group-specific** σ_g: When dispersion varies by group, comparing posterior medians shows which groups exhibit greater residual variability. CI length indicates uncertainty about each group's noise level, and differences in overlap reveal whether groups can be meaningfully distinguished in terms of variance.

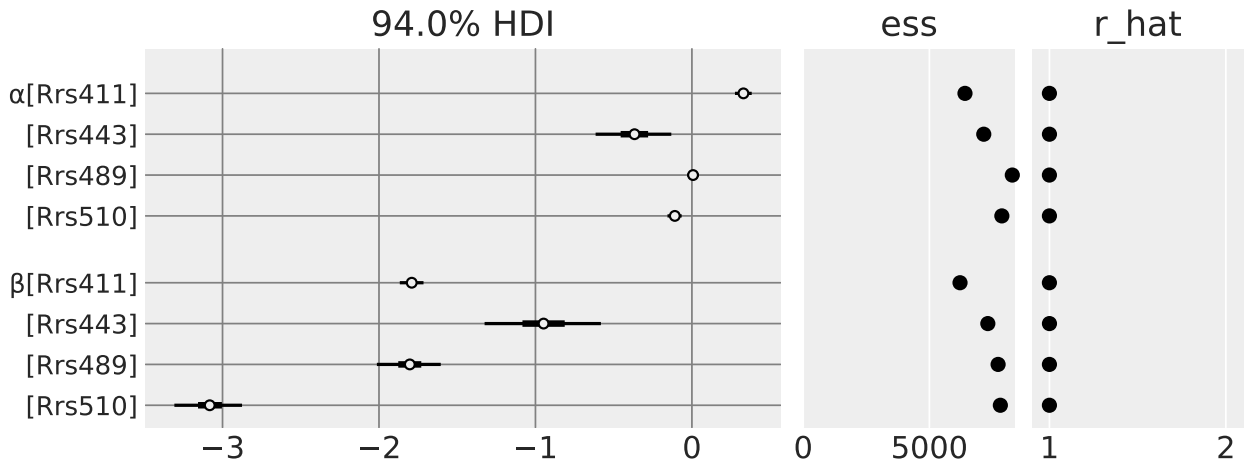

Figure S14: Forest plot for **Model 4**. Panels show posterior intervals (left), ESS (middle), and $\hat{R}$ (right) for group-indexed parameters (α, β), their hyperparameters $(\alpha_\mu, \beta_\mu, \alpha_\sigma, \beta_\sigma)$, and the global variance-function coefficients $(\sigma_\alpha, \sigma_\beta)$. Observation-specific dispersions σ_i are not included, since plotting hundreds of rows would obscure interpretation. Colors indicate MBR numerator groups for vector-valued parameters. ESS and $\hat{R}$ values indicate good mixing and convergence.

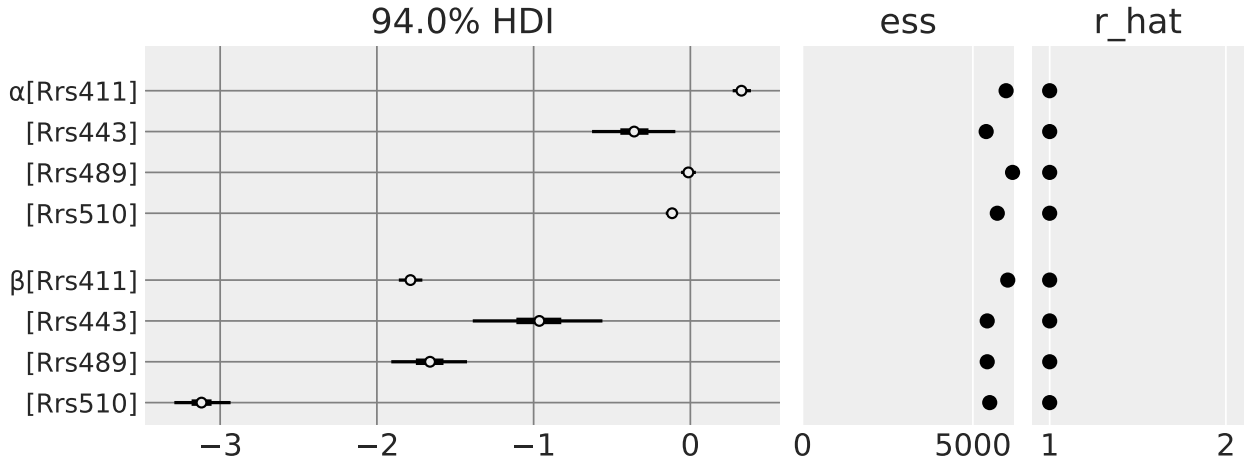

Figure S15: Forest plot for **Model 5**. Three-panel layout (posterior intervals, ESS, $\hat{R}$) for group-indexed mean parameters (α, β), variance-function parameters $(\sigma_\alpha, \sigma_\beta)$, and their hyperparameters. Each vector has one entry per MBR numerator group; colors distinguish groups but the mapping is omitted since the forest plot focuses on magnitude and uncertainty. Observation-specific dispersions σ_i are excluded for clarity. Posterior intervals, ESS, and $\hat{R}$ values confirm stable estimation and convergence across all group- and hyper-level parameters.

Overall, all models exhibit good convergence and adequate sampling efficiency: ESS values are consistently high, and $\hat{R}$ values are effectively 1.00-1.01 for all monitored parameters.

3. In- and Out-of-Sample Performance Assessment

3.1. Predictive Checks and Calibration Plots

This section presents four-panel diagnostic figures for each model. Top row shows prior and posterior predictive checks, providing insight into how well the model has learned from training data. Bottom row shows Leave-One-Out-Probability Integral Transform (LOO-PIT) diagnostics to assess model calibration and potential underdispersion, overdispersion or bias problems on future data. Comparing these plots between models helps understand areas of improvements, as well as identifying potential problems left to resolve.

How to Read LOO-PIT Plots

Calibration target: If the model is well-calibrated, PIT values follow a Uniform$(0, 1)$. Density plots should be flat at 1, and ECDF–uniform difference plots should lie around 0.

Dispersion cues:

- *Hump in the center ($\sim$0.5):* predictive intervals too wide (**over-dispersed**).
- *U-shape (peaks near 0 and 1):* predictive intervals too narrow (**under-dispersed**).

Bias cues:

- *More mass near 1 / ECDF curve below 0:* model **underpredicts** (observed values larger than predicted).
- *More mass near 0 / ECDF curve above 0:* model **overpredicts** (observed values smaller than predicted).

Relation between panels:

- Where PIT density > 1, the ECDF curve slopes upward.
- Where PIT density < 1, the ECDF curve slopes downward.

Together, the two views show both **dispersion** (interval width) and **bias direction** (systematic under- vs. overprediction).

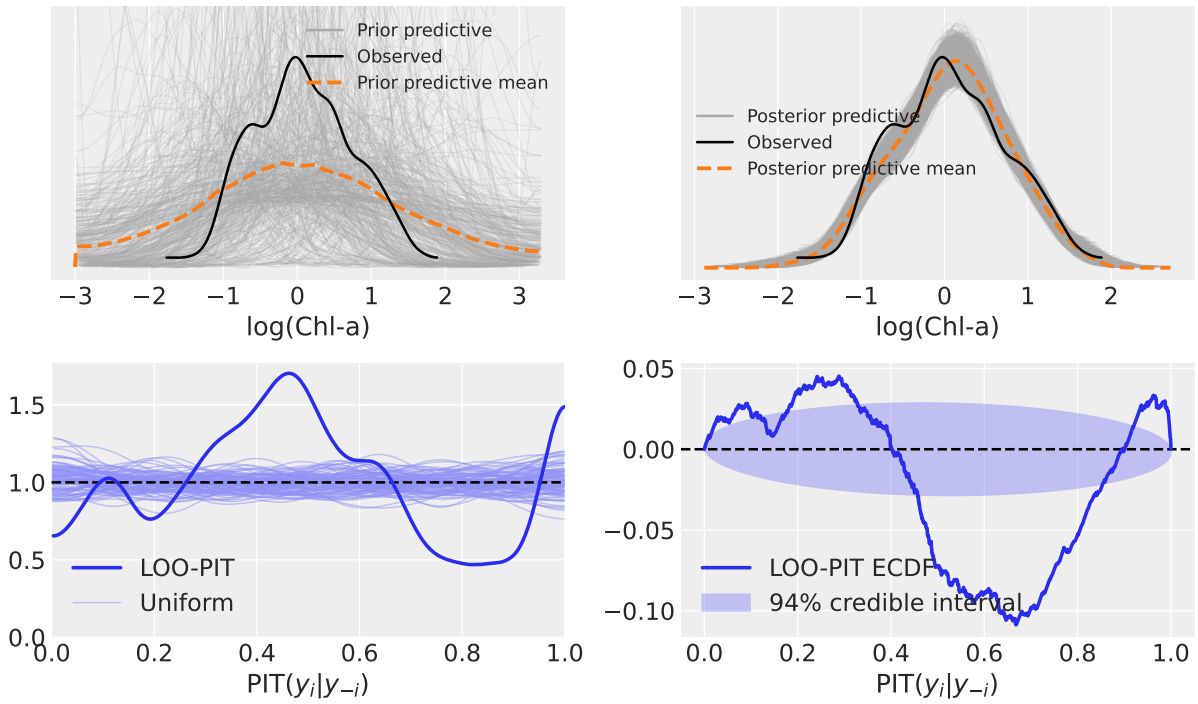

Figure S16: Model 1 4-panel diagnostics. **Top-left:** prior predictive check; observed density (black) shown for visual comparison against prior predictive draws (gray) and their mean (orange dashed). **Top-right:** posterior predictive check; observed log(Chl) density (black) overlaid with posterior predictive draws (gray) and their mean (orange dashed). **Bottom-left:** LOO-PIT kernel density estimate (KDE) in black, with reference KDEs from $\mathcal{U}(0,1)$ draws in light blue; a dashed horizontal line at $y = 1$ indicates ideal calibration. The LOO-PIT density plot departs from uniformity, with noticeable deviations around mid-quantiles, suggesting systematic under- or over-prediction in those regions.**Bottom-right:** LOO-PIT deviance from uniform empirical cumulative distribution function (ECDF) plot; the curve shows the ECDF difference relative to the identity line with a shaded 94% reference envelope under uniformity and a dashed $y = 0$ reference.

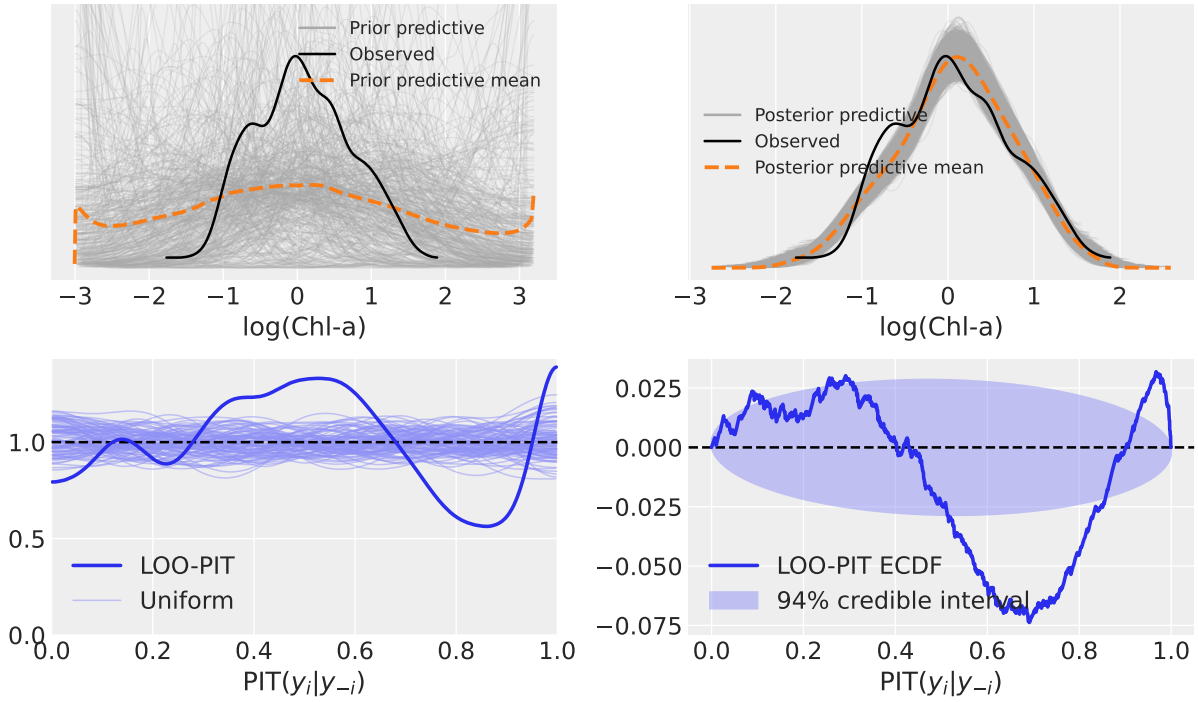

Figure S17: Model 2 - Posterior predictive checks show a closer alignment between predictive mean and observed distribution then with Model 1, particularly in the central region. LOO-PIT diagnostics also improve: the density more closely approximates uniformity, and the ECDF difference plot exhibits smaller deviations within the 94% credible interval. These improvements reflect the benefits of partial pooling across MBR groups, which reduces systematic miscalibration observed in Model 1.

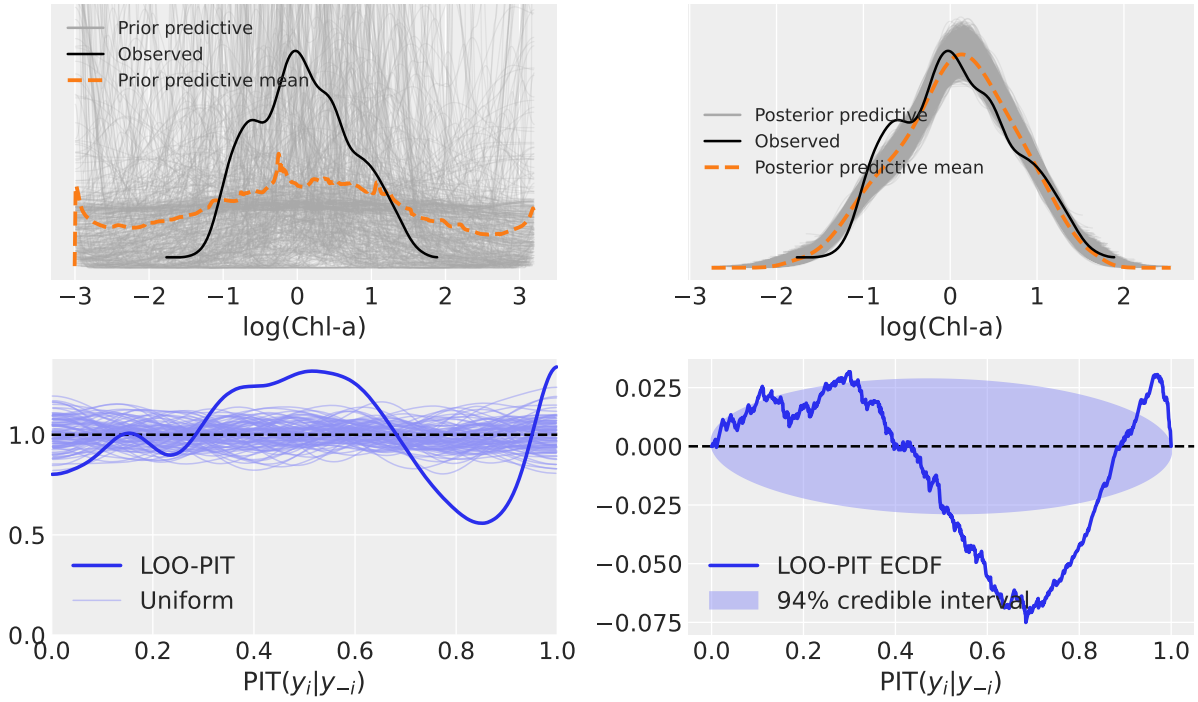

Figure S18: Model 3 - Posterior predictive checks remain well aligned with the observed distribution, with slightly improved representation of the tails. LOO-PIT diagnostics show further gains: the density curve is closer to uniform, and deviations in the ECDF difference plot are reduced in magnitude relative to Model 2. These improvements reflect the introduction of group-specific dispersion, which allows the model to better capture heterogeneity in variance across MBR numerator groups.

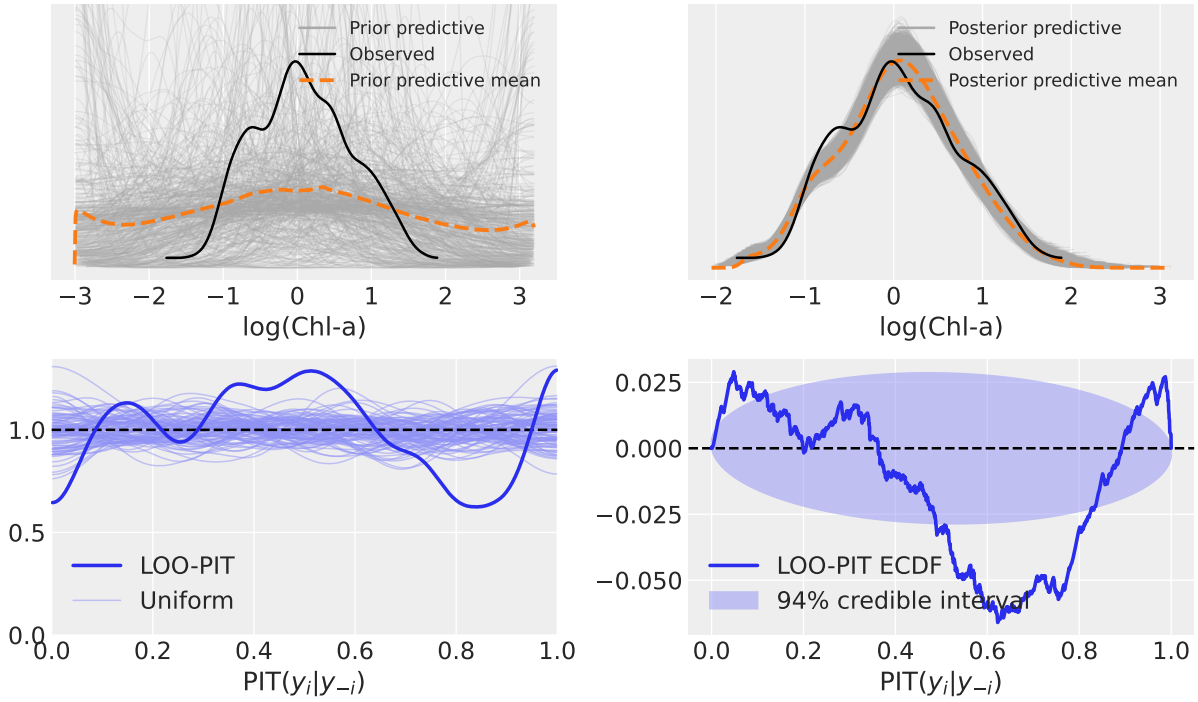

Figure S19: Model 4 - Posterior predictive checks continue to align well with the observed distribution, with modest further improvement in the fit of the tails. LOO-PIT diagnostics show that calibration remains strong: the density curve is smoother and closer to uniform than in Model 3, and the ECDF difference plot shows reduced systematic deviations across mid-quantiles. These gains arise from modeling dispersion as a function of the predictor, which allows the model to better capture input-dependent heteroskedasticity.

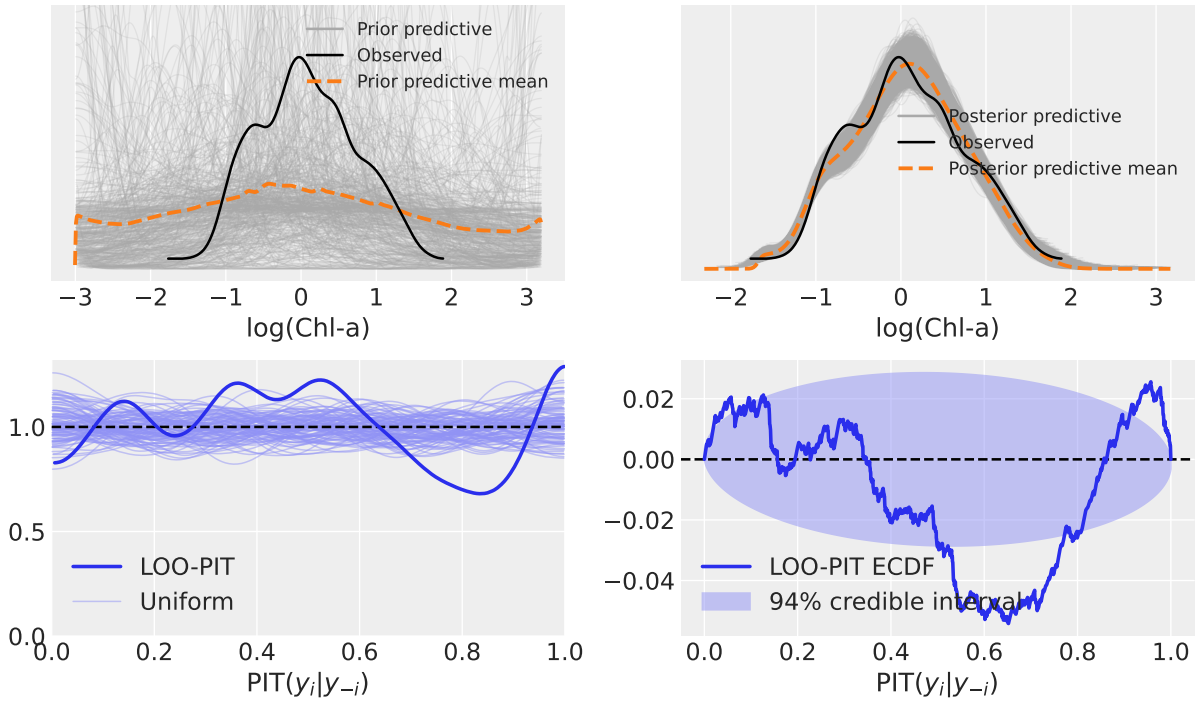

Figure S20: Model 5 - Posterior predictive distributions remain well aligned with observations, with slightly better fit in the distribution tails. The LOO-PIT diagnostics show further calibration gains: the density curve stays closer to uniform across quantiles, and the ECDF difference plot shows narrower deviations well within the 94% credible interval. These improvements stem from allowing both the intercept and slope of the log-dispersion to vary by MBR numerator group, enabling the model to flexibly capture group-specific heteroskedasticity.

3.2. Predictive Coverage Plots

Predictive coverage plots illustrate the proportion of observed chlorophyll-a concentrations that fall within the 94% highest density intervals (HDI) of the posterior predictive distribution. These plots provide a visual diagnostic of calibration: ideally, the fraction of covered observations should be close to the nominal level (94%), and deviations highlight under- or over-coverage.

In the predictive coverage plots, the shaded ribbon represents the 94% HDI of the posterior predictive distribution. At any given value of log(MBR) on the x-axis, the ribbon spans vertically along the y-axis to indicate the range within which there is a 94% probability that the predicted log(Chl) lies, conditioned on the data and the model. This direct probabilistic interpretation distinguishes Bayesian predictive intervals from classical confidence intervals, which do not provide probability statements about parameters or predictions.

Figures S21–S25 show predictive coverage for Models 1 through 5, stratified by maximum band ratio (MBR) numerator group. Each panel compares observed values against predictive intervals; black points denote observed log-chlorophyll, shaded bands indicate the posterior predictive HDI, and coverage percentages are summarized in the figure.

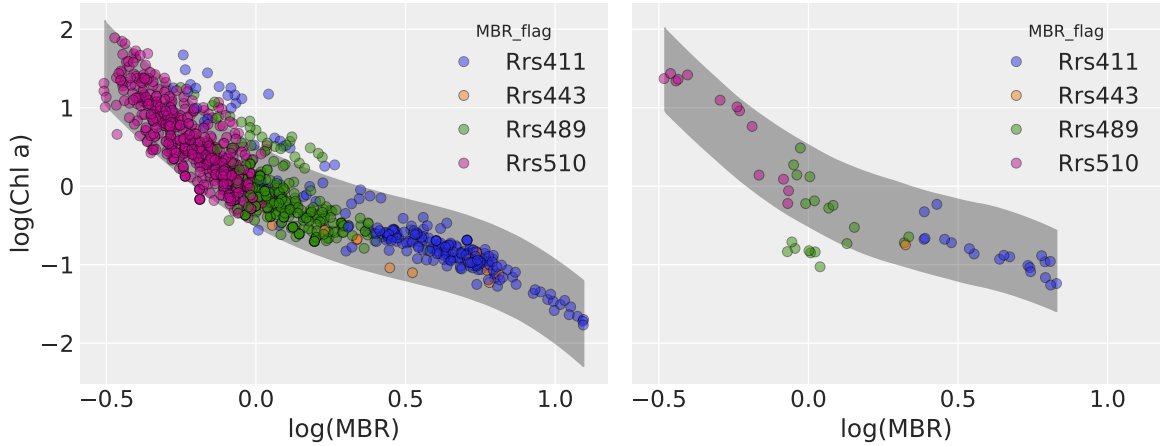

Figure S21: Model 1 predictive coverage. Left: in-sample (NOMAD). Right: out-of-sample (SeaBASS). Gray ribbons show the 94% posterior predictive HDI. In-sample, most MBR numerator groups include observations lying outside the HDI, with the exception of the Rrs443 group (which also has relatively few points). Out-of-sample, a cluster of Rrs489 points falls below the HDI, indicating underestimation of chlorophyll; this miscoverage is consistent with the absence of comparable data in the training range, potentially reflecting a structural limitation of the global polynomial fit.

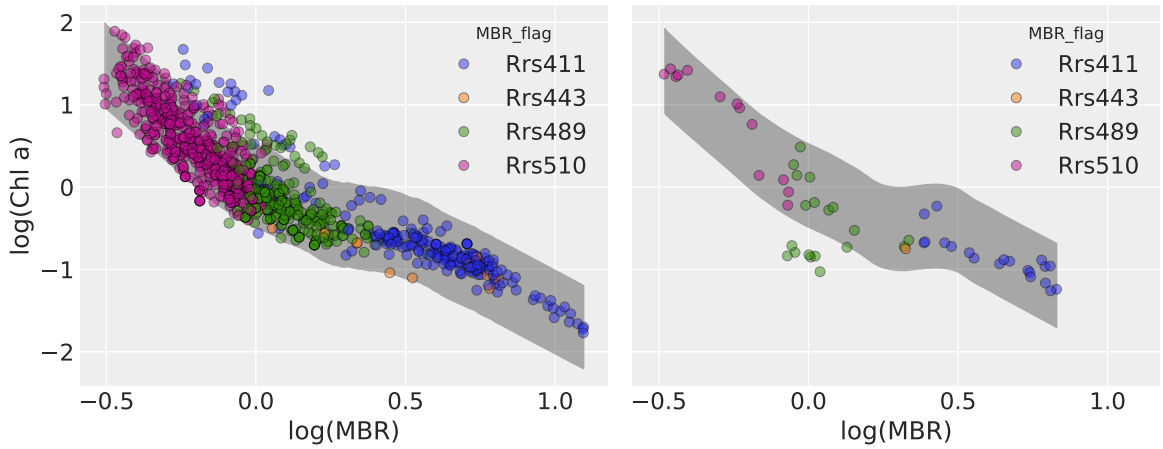

Figure S22: Model 2 predictive coverage. Left: in-sample (NOMAD). Right: out-of-sample (SeaBASS). Gray ribbons show the 94% posterior predictive HDI. The intervals are stratified by MBR numerator group, producing segmented bands that are narrower due to hierarchical shrinkage. In-sample, coverage better follows the tendencies of each group, while out-of-sample the Rrs489 cluster remains outside the HDI but coverage for other groups improves.

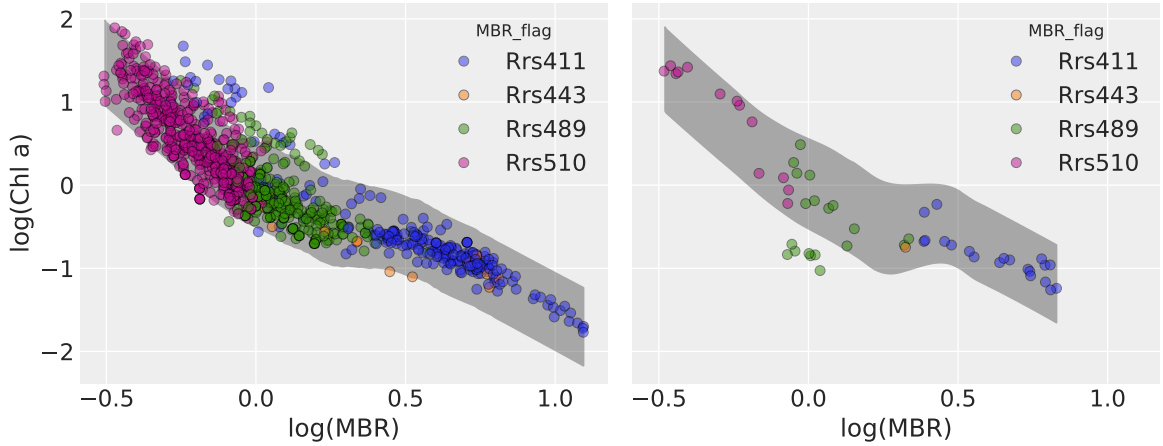

Figure S23: Model 3 predictive coverage. Left: in-sample (NOMAD). Right: out-of-sample (SeaBASS). Gray ribbons show the 94% posterior predictive HDI. Allowing group-specific dispersion produces visibly different interval widths across MBR numerator groups, aligning the HDI more closely with observed spread. In-sample coverage improves in the tails where Model 2 intervals were too uniform, while out-of-sample the persistent Rrs489 cluster remains below the HDI.

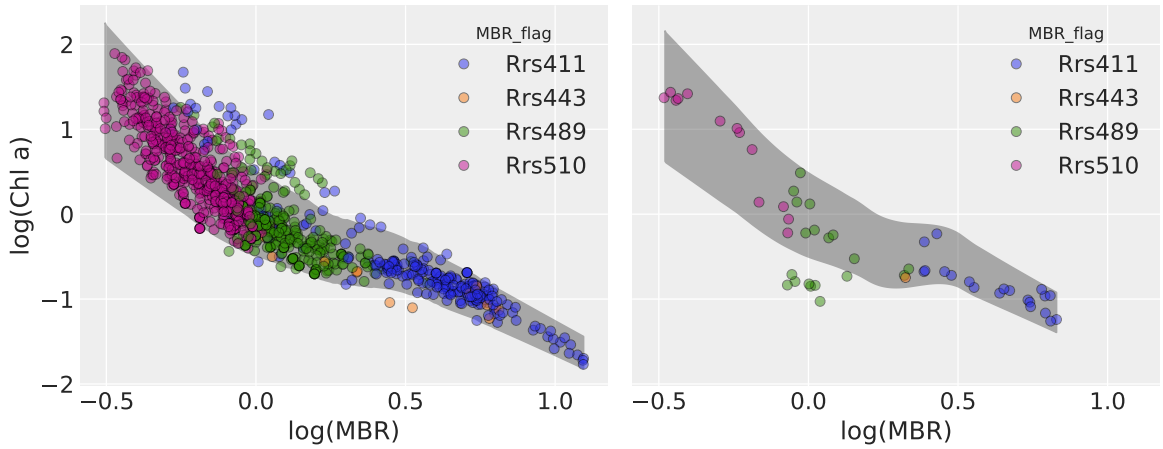

Figure S24: Model 4 predictive coverage. Left: in-sample (NOMAD). Right: out-of-sample (SeaBASS). Gray ribbons show the 94% posterior predictive HDI. Modeling dispersion as a function of log(MBR) produces a wedge-shaped envelope: intervals are much wider at low log(MBR) and narrow progressively with increasing values. This input-dependent variance structure yields closer alignment with observed spread across the predictor range. The out-of-sample panel (left) similar tightening of the HDI band.

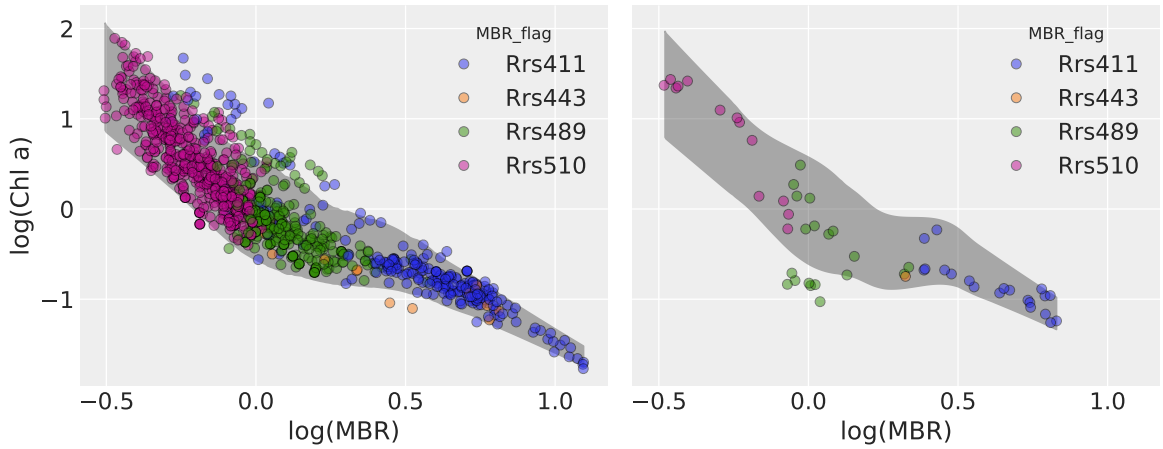

Figure S25: Model 5 predictive coverage. Left: in-sample (NOMAD). Right: out-of-sample (SeaBASS). Gray ribbons show the 94% posterior predictive HDI. Group-specific variance functions yield narrower and more flexibly shaped intervals than in Model 4, adapting to both slope and intercept differences across MBR groups. In-sample, the envelopes conform closely to each group's observed spread. Out-of-sample, coverage remains stable across groups, with predictive intervals transferring more effectively than in earlier models.

4. Five-Model Comparison

To evaluate predictive performance across all models, I used Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO). This method approximates the expected log predictive density (ELPD) that each model would achieve when predicting unseen data, while adjusting for model complexity. Unlike point-based metrics such as RMSE or R^2, PSIS-LOO evaluates the entire posterior predictive distribution, providing a more comprehensive basis for comparison.

Table S1 summarizes the PSIS-LOO results for Models 1 through 5, with smaller rank values indicating better expected predictive accuracy. These numerical results are more clearly conveyed in Figure S26, which provides a visual representation of model ranking along with the associated uncertainty.

Table S1: PSIS-LOO comparison of Models 1 through 5. **rank**: model ranking based on expected out-of-sample predictive accuracy, with 0 indicating the best-performing model. **elpd_loo**: expected log predictive density; higher values indicate better predictive performance. **p_loo**: effective number of parameters, estimated from the variance of the pointwise log-likelihood; reflects model flexibility and complexity. **elpd_diff**: difference in ELPD relative to the top-ranked model; large positive values indicate substantially worse predictive performance. **weight**: approximate model weight under Bayesian stacking, representing relative support for each model given the data; values near 1 indicate strong preference. **se**: standard error of the ELPD estimate for each model. **dse**: standard error of the ELPD difference compared to the top-ranked model. **warning**: flag indicating whether reliability issues were detected in the importance-sampling diagnostics; `False` indicates stable estimates. **scale**: the unit in which predictive densities are expressed (here, log scale).

Model	Rank	ELPD_{LOO}	p_{loo}	ΔELPD	Weight	SE	dSE	Warning
Model 5	0	-0.25	15.27	0.00	0.96	28.11	0.00	False
Model 4	1	-51.73	10.87	51.48	0.00	29.98	10.43	False
Model 3	2	-111.06	14.63	110.81	0.00	27.95	13.41	False
Model 2	3	-114.00	9.88	113.75	0.00	27.59	12.54	False
Model 1	4	-144.54	5.40	144.28	0.04	31.67	21.66	False

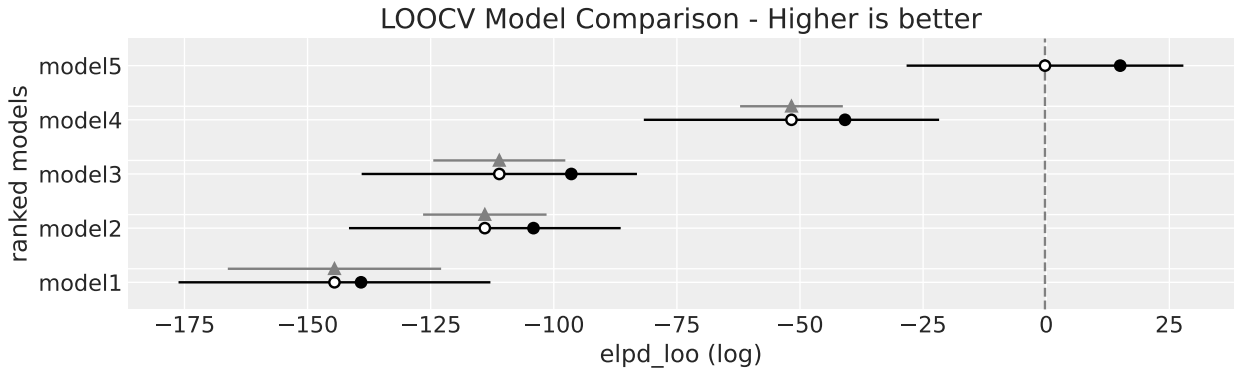

Figure S26: LOOCV model comparison (higher is better). The horizontal axis shows ELPD_{LOO} on the log scale; models are ordered top to bottom by rank (best at top). Open circles mark out-of-sample expected log predictive density estimated via PSIS-LOO; filled circles mark the corresponding in-sample predictive performance computed on the full dataset (same log-probability units), allowing direct visual comparison of apparent fit versus expected generalization. Horizontal error bars on the open circles are the SE of ELPD_{LOO} for each model. Light-gray horizontal segments (enabled by `ic_diff`) depict $\pm$dSE intervals for the ELPD differences relative to the top model (triangle at $\Delta = 0$), indicating the uncertainty in the pairwise gaps that drive ranking. Larger (less negative) ELPD values indicate better expected out-of-sample performance. When ΔELPD for a model is large relative to its dSE interval, the separation is practically decisive; when ΔELPD is on the order of its dSE, ranking uncertainty should be assumed.

Taken together, Table S1 and Figure S26 present a consistent view of model performance, with the table providing detailed numerical values and the figure highlighting both relative ranking and the uncertainty in those comparisons. The comparative rankings reinforce the

patterns seen in predictive coverage and LOO-PIT diagnostics, where progressively richer model structures yielded improved calibration and generalization.