The Regularized Horseshoe Prior for Covariate Selection Improves Convenience and Predictive Performance in Population PK/PD Models

Arya Pourzanjani[1], Casey Davis[1]

[1]Daiichi Sankyo Inc., USA

Motivation

Covariate selection for population PK/PD models is a crucial task in pharmacometrics. Several methods have been used [1] each with their own pros, cons, and performance [2, 3] summarized in Table 1. **The Regularized Horseshoe (RHS)** [4] is a relatively new approach with several advantages. We investigate its use in population PK/PD models and find that it offers a **convenient single-fit approach, improved estimation performance**, and **improved predictive performance on held-out data** compared to popular stepwise approaches SCM and COSSAC.

	Full (FFEM/FREM)	Stepwise (SCM/COSSAC)	LASSO	RHS
Avoids upwardly biased estimates	✓	✗	✓	✓
Avoids downwardly biased estimates	✓	✓	✗	✓
Unbiased confidence intervals	✓	✗	✓	✓
Low variance	✗	✓	✓	✓
Sparse coefficient estimates	✗	✓	✓	✓
Only single model fit needed	✓	✗	✓	✓
Averaging over posterior uncertainty	✗	✗	✗	✓

Table 1: Pros and cons of various covariate selection methods.

Methods

The Regularized Horsehoe is a prior on regression coefficients θ_j favoring sparse values but having heavy tails to allow for some larger values that are minimally shrunk (Figure 1 Left). If we define κ_j to be a shrinkage factor between zero (no shrinkage) and one (full shrinkage) that tells us how much each coefficient is shrunk from its least-squares estimate, its prior distribution resembles a horseshoe (Figure 1 Right).

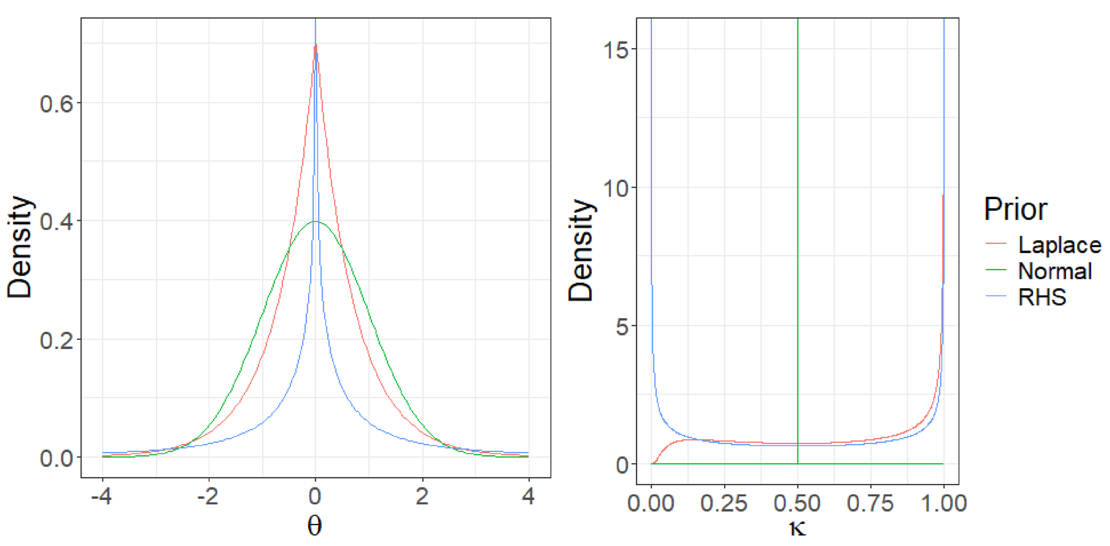

Figure 1: Prior density of θ (left) and κ (right).

Mathematically, the RHS prior is defined hierarchically as follows [4]:

$$\begin{aligned} \theta_j \mid \tau, \lambda_j, c &\sim \mathrm{N}(0, \tau\tilde{\lambda}_j) \\ \lambda_j &\sim \mathrm{Half-Cauchy}(0,1) \\ \tau &\sim \mathrm{Half-Cauchy}(0,\tau_0) \\ c^2 &\sim \mathrm{Inv-Gamma}(\nu/2, \nu s^2/2) \\ \tilde{\lambda}_j^2 &:= \frac{c^2\lambda_j^2}{c^2+\tau^2\lambda_j^2} \end{aligned}$$

Setting Hyperparameters for Population PK/PD Models

To aim for generality and wide-use within the pharmacometrics community we aimed to define a set of RHS hyperparameters applicable to general population PK/PD models encountered in practice.

A hyperparameter value of $\tau_0 = 0.032$ was tuned so that approximately 14% of coefficients will be non-zero on average with a 90% probability that the percentage of non-zero coefficients is between 2% and 54% a priori, values that are similar to the 10%, 0%, and 49%, respectively, reported across the 17 datasets in [5].

The slab scale and degrees of freedom parameters, $s = 0.5$ and $\nu = 3$, were chosen to approximate the distribution of coefficient values typically seen in population PK/PD models with some loosening to err on the side of the data. More on setting hyperparameters for the RHS and using simulation is discussed in [4] and [6], respectively.

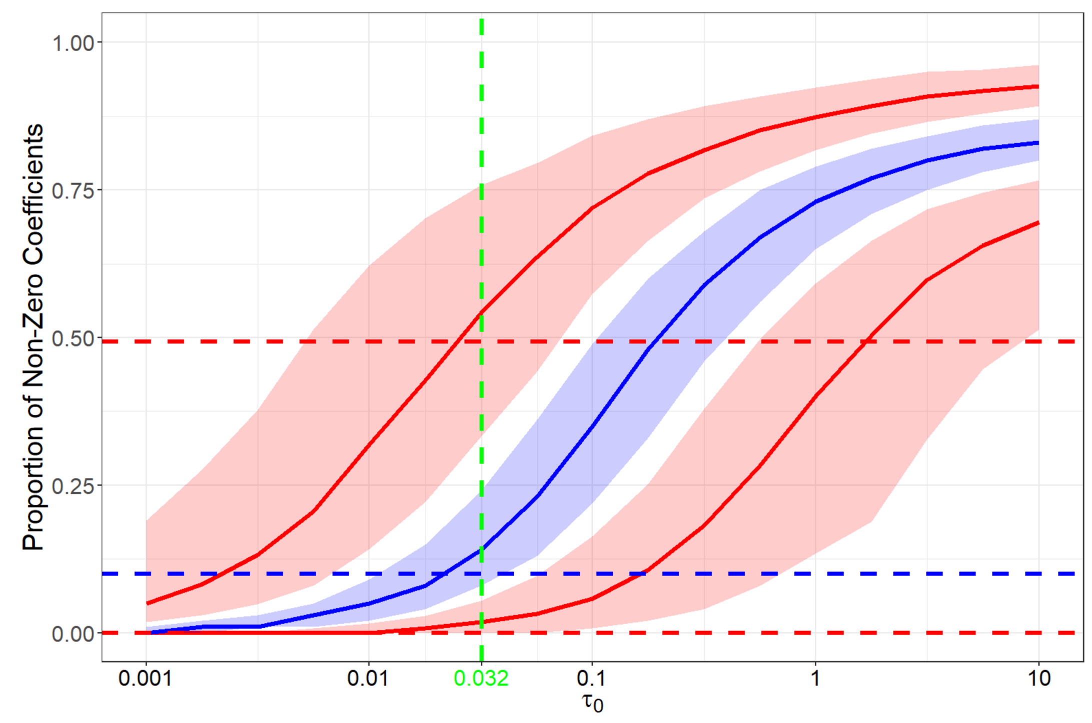

Figure 2: Median and 90% quantiles of the sparsity in simulated datasets for varying values of τ_0. Observed values are overlaid as dotted lines. Ribbons represent the 90% quantiles of these quantities.

Evaluating Estimation Performance

A simulation study using a two-compartment IV PK model with 10 possible covariates with varying levels of sparsity (0%, 10%, 50%, and 70%), three levels of dataset sizes (12, 40, and 300 subjects), and effect sizes varying across a range of realistic values was performed, assessing bias, mean-squared error, Type 1 error, Type 2 error, Type S error, and Type M error and compare to COSSAC and SCM.

Evaluating Predictive Performance

To evaluate the predictive performance of each method we used the log pointwise predictive density (LPPD)[7] using leave-one-group-out cross-validation (LOGO-CV):

$$\mathrm{lppd}_{\mathrm{logo-cv}} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\eta_{is})\right)$$

We calculate the LPPD for each method and simulated dataset from the simulation study. We also fit RHS, SCM, and COSSAC to four real-world datasets (remifentanil PK, theophylline PK, tobramycin PK, and warfarin PK/PD) found in [5] and compare predictive performance via LPPD.

Results

Sparse Posteriors

Figure 3 shows the marginal posterior median and 90% posterior intervals for all of the regression coefficients in the RHS fit of the remifentanil model. The model has six parameters and six covariates thus 36 coefficients in total. As expected, most of the coefficients have posteriors that concentrate around zero, but five coefficients have posterior 90% intervals that do not encompass zero.

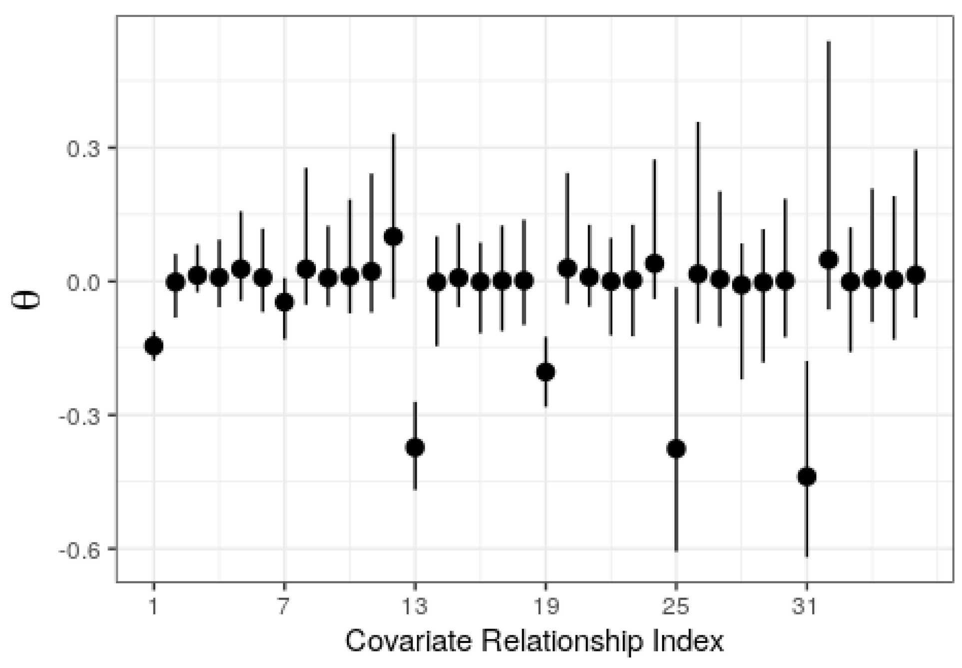

Figure 3: Posterior medians and 90% quantiles for each coefficient in the remifentanil model.

Simulated Data

Figure 4 shows model estimates for the effect of age on central compartment volume across five simulated datasets and four sparsity settings for the 12 subject dataset, and Figure 5 shows estimation metrics for the effect of age on clearance.

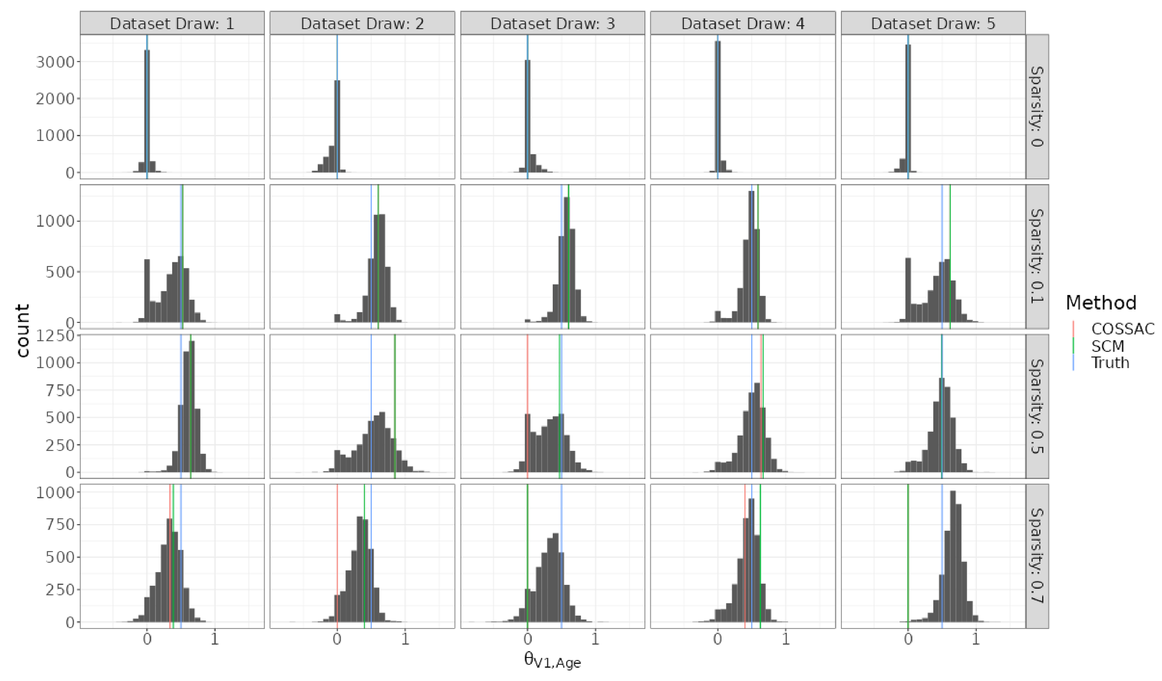

Figure 4: Posterior distributions of the $\theta_{V1,Age}$ parameter. Ground truth values and point estimates are overlaid as vertical lines.

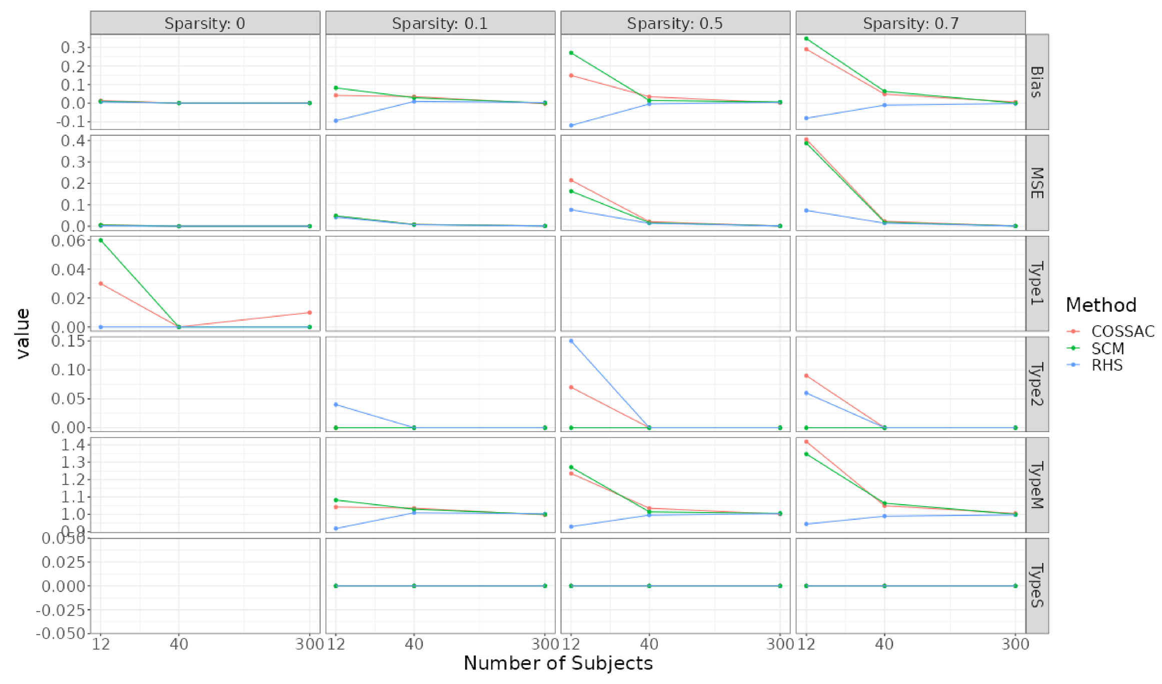

Figure 5: Evaluation metrics of estimators of $\theta_{CL,Age}$ as a function of ground-truth sparsity value and number of subjects fit.

These figures show that SCM and COSSAC provide point estimates that make a definite decision on whether a coefficient is zero or non-zero while often overestimating the true effect size, while RHS sometimes exhibits bimodality to reflect the true uncertainty in non-zero coefficients given the data while not having the same problems wih overestimation of the true effect size.

Predictive Performance on Held-out Data

The RHS outperforms both stepwise methods in predictive performance on held-out data on all four examples (see Table 2). Some of this outperformance can be attributed to the upward bias in coefficient estimates described above that occurs with stepwise methods, and some can be attributed to averaging over posterior uncertainty to reduce overfitting.

Dataset	Subj.	Pars.	Covs.	LPPD SCM	LPPD COSSAC	LPPD RHS
Remifentanil PK	65	6	6	-3763	-3804	-3600
Theophylline PK	12	7	3	-197	-197	-182
Tobramycin PK	97	2	4	-324	-324	-290
Warfarin PK/PD	32	8	3	-1289	-1289	-1112

Table 2: Description of datasets/models and LPPD for each method.

References

[1] Mélanie Karlsen et al. "Covariate Model Selection Approaches for Population Pharmacokinetics: A Systematic Review of Existing Methods, From SCM to AI". In: *CPT: Pharmacometrics & Systems Pharmacology* (2025).

[2] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.

[3] Frank E Harrell et al. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Vol. 608. Springer, 2001.

[4] Juho Piironen and Aki Vehtari. "Sparsity information and regularization in the horseshoe and other shrinkage priors". In: (2017).

[5] Géraldine Ayral et al. "A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models: The COSSAC approach". In: *CPT: pharmacometrics & systems pharmacology* 10.4 (2021), pp. 318–329.

[6] Andrew Gelman, Daniel Simpson, and Michael Betancourt. "The prior can often only be understood in the context of the likelihood". In: *Entropy* 19.10 (2017), p. 555.

[7] A. Gelman et al. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN: 9781439840955. URL: https://books.google.com/books?id=ZXL6AQAAQBAJ.