

A Trail of "Buts"...

GRwHS: The Evolution of Group-Aware Sparse Regression

Real-World Challenge: Data with Group Structure

Natural grouping (pathways, families)
Mixed signal strength (strong/medium/weak)

High within-group correlation
Need adaptive group-wise shrinkage

Penalized Regression Path

Lasso

Individual Sparsity

Flaw: No group structure

Ridge

No Sparsity

Flaw: Dense solutions

SGL

Sparse Group Lasso

Solution: Group + Individual sparsity
But: Rigid shrinkage patterns

Bayesian Path

Horseshoe

Global-Local Framework

Flaw: Group blind + Sampling instability

RHS

Regularized Horseshoe

Solution: Numerical stability
But: Still group blind

GIGG

Group Structure

Solution: Group hierarchy
But: Computationally complex

GRwHS

Group Ridge within Horseshoe

Solution: Group-aware + Computationally stable + Adaptive shrinkage





Model Comparison

Table: Structural comparison of sparse regression models

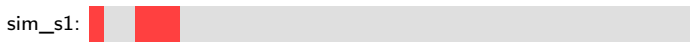
Model	Global	Group	Local Shrinkage
Lasso	α	N/A	$\alpha \sum_j \beta_j $
Ridge	α	N/A	$\alpha \sum_j \beta_j^2$
SGL	(α_1, α_2)	$\alpha_2 \sum_g w_g \ \beta_g\ _2$	$\alpha_1 \sum_j \beta_j $
Horseshoe	$\tau \sim C^+(0, \sigma)$	N/A	$\lambda_j \sim C^+(0, 1)$
RHS	$\tau \sim C^+(0, \tau_0)$	N/A	$\lambda_j \sim C^+(0, 1), \tilde{\lambda}_j^2 = \frac{c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}$
GIGG	$\tau \sim C^+(0, \sigma)$	$\gamma_g^2 \sim \text{Gamma}(a_g, 1)$	$\lambda_{gj}^2 \sim \text{InvGamma}(b_g, 1), \gamma_g^2 \lambda_{gj}^2 \sim \beta'(a_g, b_g)$
GRwHS	$\tau \sim C^+(0, \tau_0)$	$\phi_g \sim \mathcal{N}^+(0, \frac{\eta^2}{p_g})$	$\lambda_j \sim C^+(0, 1), \tilde{\lambda}_j^2 = \frac{c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}$

Synthetic Scenarios

- Design: $n=1300$, Train/test: 300/1000, 5-fold CV, $p=200$, $G=8$ groups [5,10,15,20,30,40,40,40]
- Correlation: group block ($\rho_{in}=0.7$, $\rho_{out}=0.1$); Signal-to-noise ratio(SNR) $\in \{0.1, 0.5, 1, 3\}$

Legend:  strong dense  medium  weak / sparse  null

Each bar represents $p=200$ features grouped as [5,10,15,20,30,40,40,40].



Group-sparse strong signals: two active groups; others ideally shrunk.



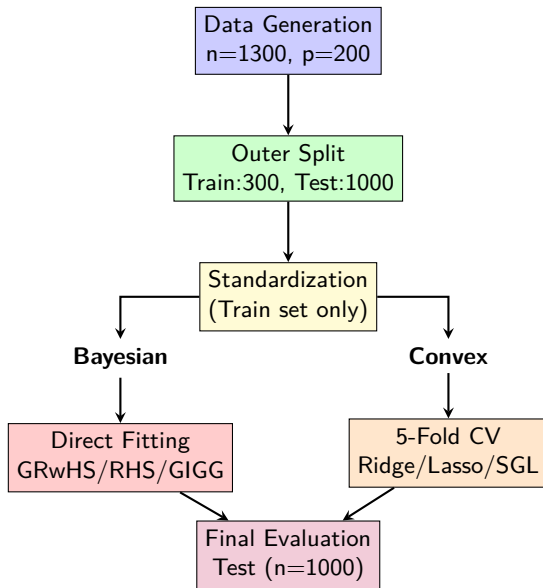
Dense weak signals: all groups slightly active; tests robustness to over-shrinkage.



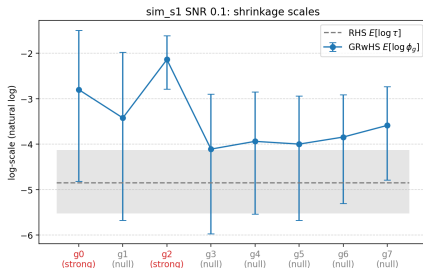
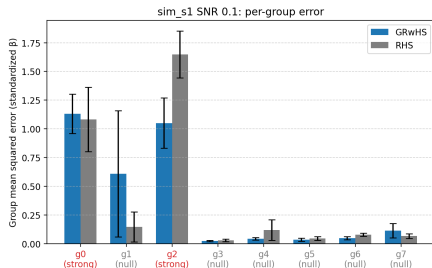
Mixed hierarchy: strong G_0 , medium G_3 , sparse weak tails in G_{5-7} .

All share identical design, standardization, and evaluation pipeline (RMSE, group edf, posterior shrinkage).

Experimental Design: Nested Data Splitting



Results: sim_s1



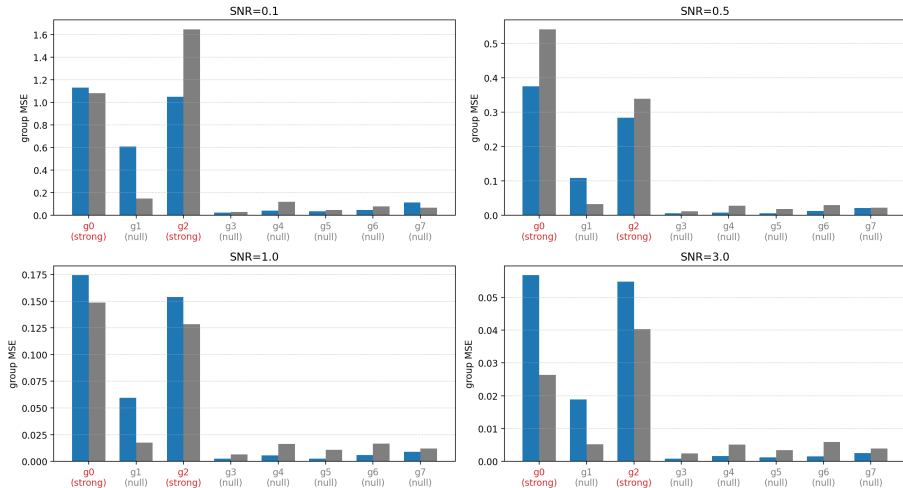
Left: GRwHS lowers error in the strong (g2) groups while matching RHS on weak/null groups.

Right: GRwHS learns larger $E[\log \phi_g]$ for signal groups, whereas RHS uses one global τ (dashed).

This group-aware shrinkage releases true signal and preserves shrinkage on noise, explaining the error gains.

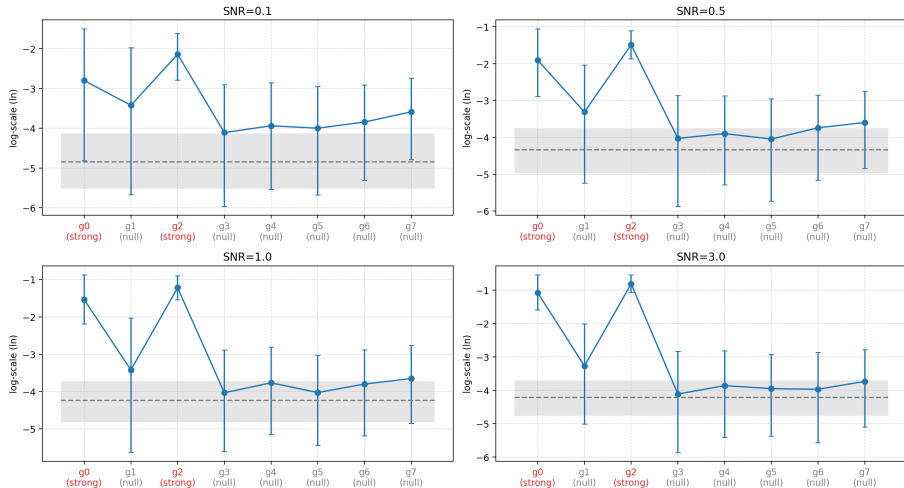
Results: `sim_s1`(Group-sparse Strong)

`sim_s1`: Per-group error across SNR



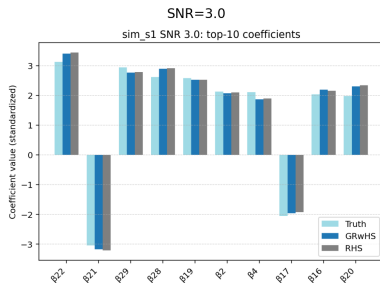
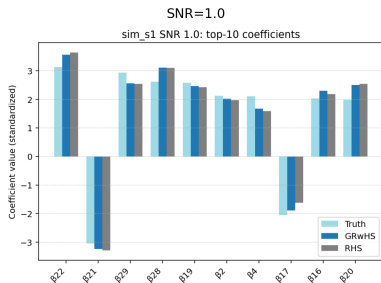
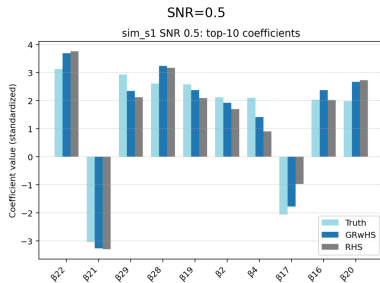
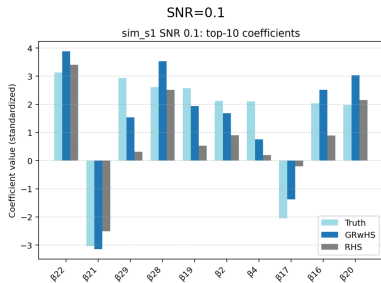
Results: sim_s1(Group-sparse Strong)

sim_s1: Group shrinkage scales across SNR



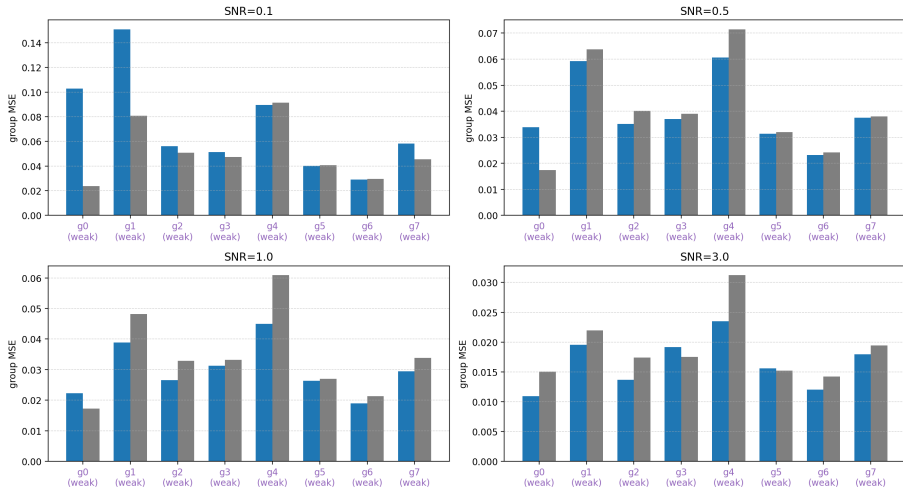
Results: sim_s1(Group-sparse Strong)

sim_s1: Top-10 coefficients (Truth vs GRwHS vs RHS)



Results: $\text{sim_s2}(\text{Dense Weak})$

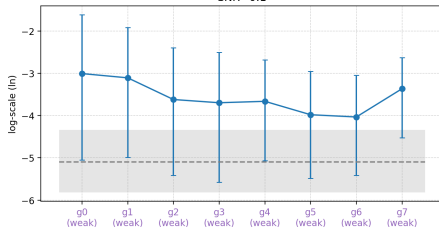
sim_s2 : Per-group error across SNR



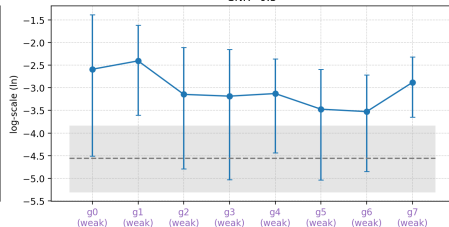
Results: $\text{sim_s2}(\text{Dense Weak})$

sim_s2 : Group shrinkage scales across SNR

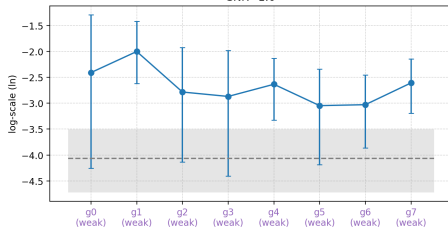
SNR=0.1



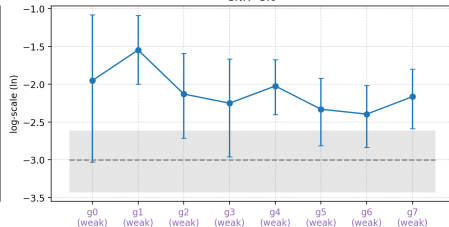
SNR=0.5



SNR=1.0

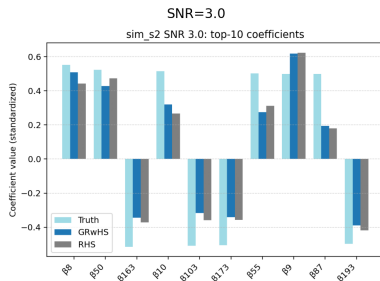
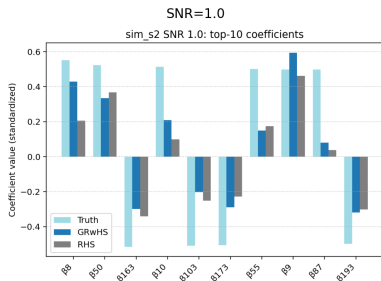
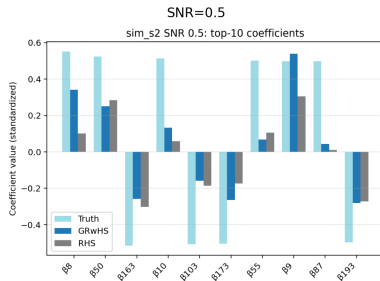
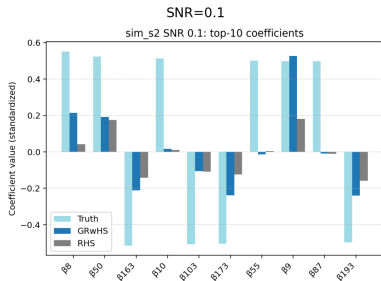


SNR=3.0



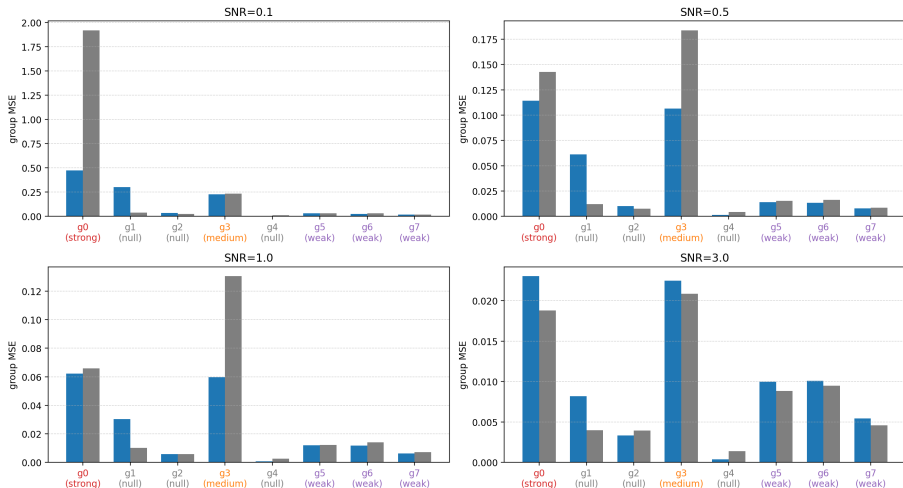
Results: sim_s2(Dense Weak)

sim_s2: Top-10 coefficients (Truth vs GRwHS vs RHS)



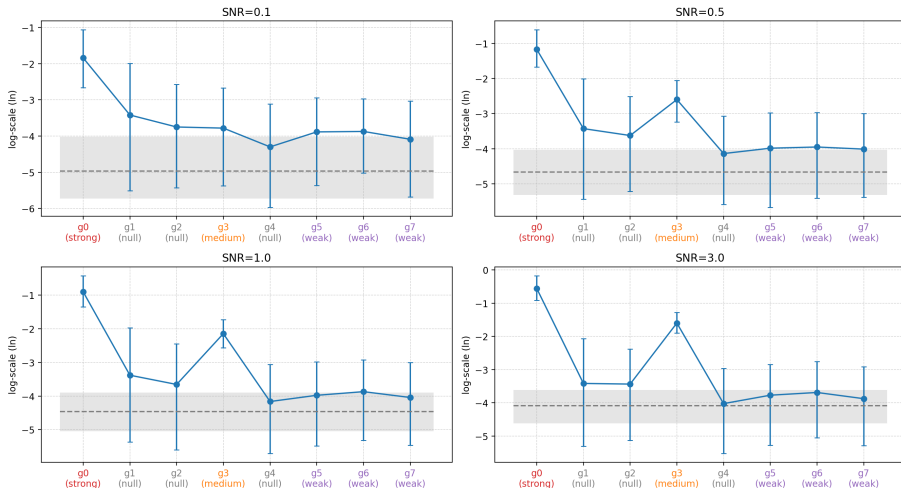
Results: sim_s3(Mixed Hierarchy)

sim_s3: Per-group error across SNR



Results: sim_s3(Mixed Hierarchy)

sim_s3: Group shrinkage scales across SNR



Results: sim_s3(Mixed Hierarchy)

sim_s3: Top-10 coefficients (Truth vs GRwHS vs RHS)

