FLEXIBLE EMPIRICAL BAYES ESTIMATION FOR WAVELETS

Merlise Clyde

Institute of Statistics and Decision Sciences

Duke University

Clyde and George: ISDS DP 98-21 and DP 99-06

URL `http://www.stat.duke.edu/~clyde/wavelets.html`

OUTLINE

1. Bayesian Hierarchical Model for Wavelets
2. Threshold and Multiple Shrinkage Estimators
3. Empirical Bayes Prior Distributions
4. Models for Outliers
5. Performance
6. Open problems

MODEL

Non-parametric Regression Model:

Observations $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots \mathbf{Y}_n)'$

Unknown Signal $\mathbf{f} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)'$

Signal observed with error e: $\mathbf{Y} = \mathbf{f} + e$

Apply Discrete Wavelet Transform (DWT):

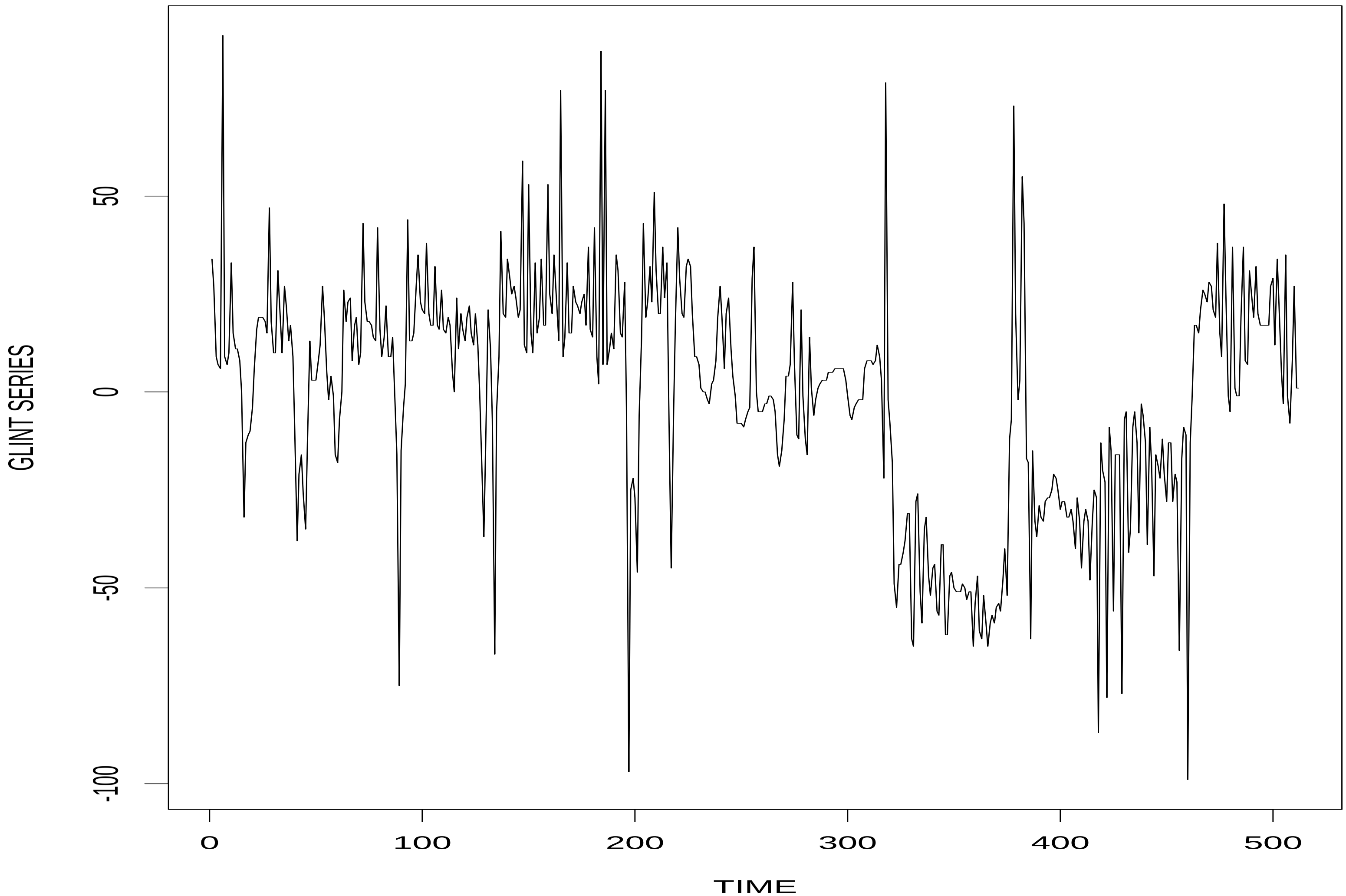
$\mathbf{W} = n \times n$ orthonormal matrix, $\mathbf{W}'\mathbf{W} = \mathbf{W}\mathbf{W}' = I$

$\mathbf{f} = \mathbf{W}\boldsymbol{\beta}$, $\boldsymbol{\beta}$ are wavelet coefficients of $\mathbf{f}$

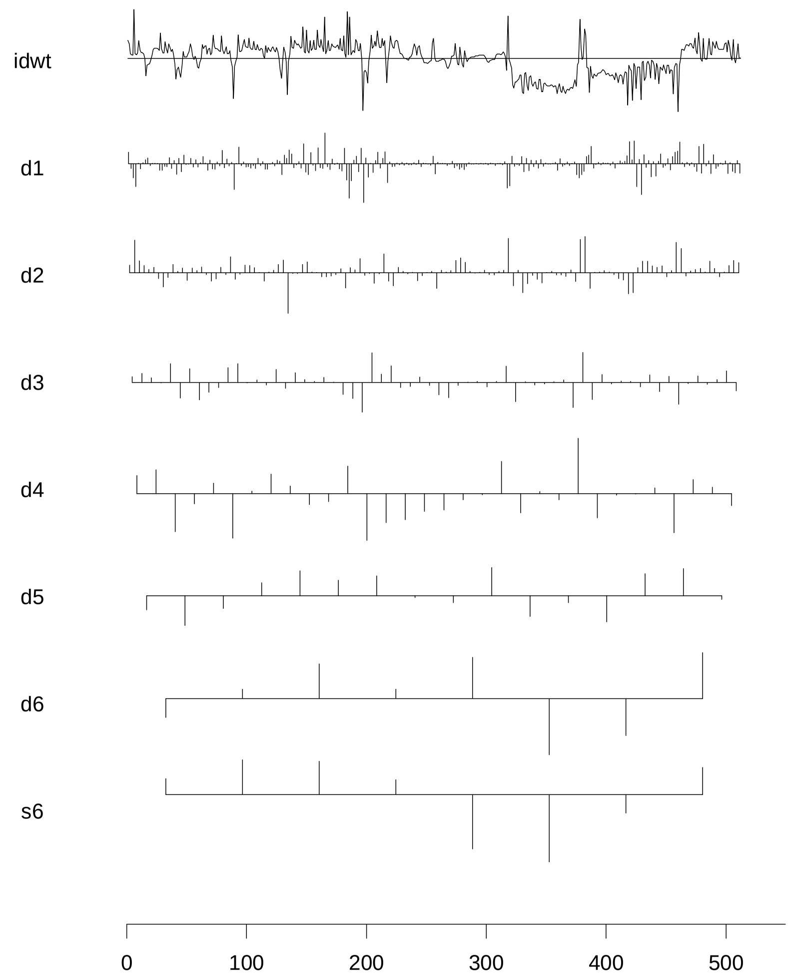
$\mathbf{W}'\mathbf{Y} \equiv \mathbf{D}$, DWT of data or empirical wavelet coefficients

Wavelet Regression Model: $\mathbf{D}_{jk} = \boldsymbol{\beta}_{jk} + \epsilon_{jk}$

Glint Radar Signal



Discrete Wavelet Transformation



Hierarchical Model in Wavelet Domain

$$\begin{array}{lrcl} & \mathbf{D}_{jk} & = & \boldsymbol{\beta}_{jk} + \epsilon_{jk} \\ \text{Normal Errors} & \epsilon_{jk} & \sim & N\left(0, \boldsymbol{\sigma}^2\right) \\ \text{Mixture Prior} & \boldsymbol{\beta}_{jk} | \boldsymbol{\gamma}_{jk} & \sim & N\left(0, \mathbf{c}_j \boldsymbol{\gamma}_{jk} \boldsymbol{\sigma}^2\right) \\ & \boldsymbol{\gamma}_{jk} & \sim & Bernoulli(\omega_j) \end{array}$$

Sparse Representation: $\boldsymbol{\beta}_{jk} = 0$ when $\boldsymbol{\gamma}_{jk} = 0$

ω_j represents **proportion of non-zero coefficients** at level j

Posterior Distribution of γ

Because of the conditional independence structure in the prior and orthogonality of the DWT, the γ's a posteriori independent Bernoulli random variables:

$$\pi(\gamma_{jk} = 1|\mathbf{Y}) = \frac{\mathcal{O}_{jk}}{1 + \mathcal{O}_{jk}}$$

where $\mathcal{O}_{jk}$ is the posterior odds that the coefficient is non-zero:

$$\mathcal{O}_{jk} = \frac{\omega_j}{1 - \omega_j}(1 + \mathbf{c}_j)^{-1/2} \exp\left\{\frac{1}{2}\mathbf{D}^2_{jk}\boldsymbol{\sigma}^2 \frac{\mathbf{c}_j}{1 + \mathbf{c}_j}\right\}$$

The posterior probability of a model γ is

$$\pi(\gamma) = \prod_{jk} \pi(\gamma_{jk} = 1|\mathbf{D})^{\gamma_{jk}} (1 - \pi(\gamma_{jk} = 1|\mathbf{D}))^{1-\gamma_{jk}}$$

Bayes Threshold and Shrinkage Estimators

Multiple Shrinkage Estimator:

Posterior Mean (minimizes posterior expected squared error loss):

$$\hat{\beta}_{jk} = \pi(\gamma_{jk} = 1 | \mathbf{D}) \frac{\mathbf{c}_j}{1 + \mathbf{c}_j} \mathbf{D}_{jk}$$

Threshold Estimator:

Posterior Mean given highest Posterior Probability model (optimal under squared error loss with model selection)

$$\hat{\beta}_{jk} = \hat{\gamma}_{jk} \frac{\mathbf{c}_j}{1 + \mathbf{c}_j} \mathbf{D}_{jk}$$

where

$$\hat{\gamma}_{jk} = 1 \quad \text{if} \quad \pi(\gamma_{jk} = 1 | \mathbf{D}_{jk}) \geq 0.5$$
$$\hat{\gamma}_{jk} = 0 \quad \text{if} \quad \pi(\gamma_{jk} = 1 | \mathbf{D}_{jk}) < 0.5$$

Transform estimates via IDWT to estimate $\mathbf{f}$

EMPIRICAL BAYES

Estimate ω_j and $\mathbf{c}_j$ and $\boldsymbol{\sigma}^2$ from the marginal likelihood of the data:

$$\mathcal{L}(\mathbf{c}, \omega, \boldsymbol{\sigma}^2) = \sum_j \sum_k \log \left[\omega_j m_1(\mathbf{D}_{jk}; \mathbf{c}_j, \boldsymbol{\sigma}^2) + (1 - \omega_j) m_0(\mathbf{D}_{jk}; \boldsymbol{\sigma}^2) \right]$$

EM algorithm

1. Introduce "missing data" γ; complete data = $(\mathbf{D}, \gamma)$
2. **E-step:**

   Expectation of missing data given estimates of $(\mathbf{c}, \omega, \boldsymbol{\sigma}^2)$
3. **M-step:**

   Maximization of complete data likelihood

   New estimates of $(\mathbf{c}, \omega, \boldsymbol{\sigma}^2)$
4. repeat 2 and 3 until convergence

Closed Formed Iterative Solutions

E-step:

$$E(\gamma_{jk}|\mathbf{D}_{jk}) = \hat{\gamma}_{jk}^{(i)} = \frac{\mathcal{O}_{jk}^{(i)}}{1+\mathcal{O}_{jk}^{(i)}} = \hat{\pi}(\gamma_{jk}=1|\mathbf{Y})$$

is the posterior mean of γ_{jk}

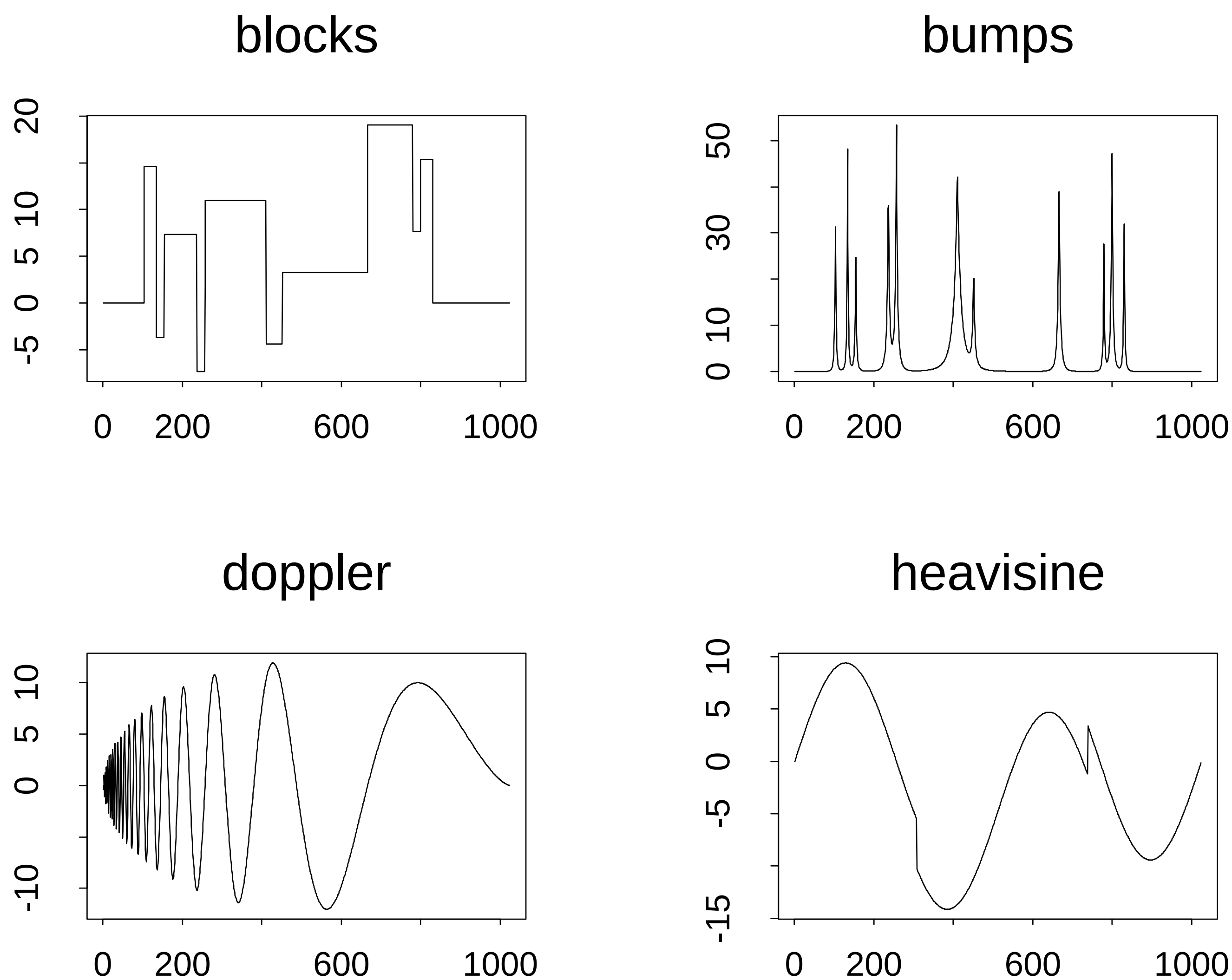
M-Step:

$$\begin{aligned}
\hat{\boldsymbol{\sigma}}^{2(i+1)} &= \frac{\sum_{jk}(1-\hat{\gamma}_{jk}^{(i)})\mathbf{D}_{jk}^2}{n-\sum_{jk}\hat{\gamma}_{jk}^{(i)}} \\
\hat{\mathbf{c}}_j^{(i+1)} &= \max\left(0, \frac{\sum_k \hat{\gamma}_{jk}^{(i)}\mathbf{D}_{jk}^2}{\hat{\boldsymbol{\sigma}}^{2(i+1)}\sum_k \hat{\gamma}_{jk}^{(i)}} - 1\right) \\
\hat{\omega}_j^{(i+1)} &= \frac{\sum_k \hat{\gamma}_{jk}^{(i)}}{n_j}
\end{aligned}$$

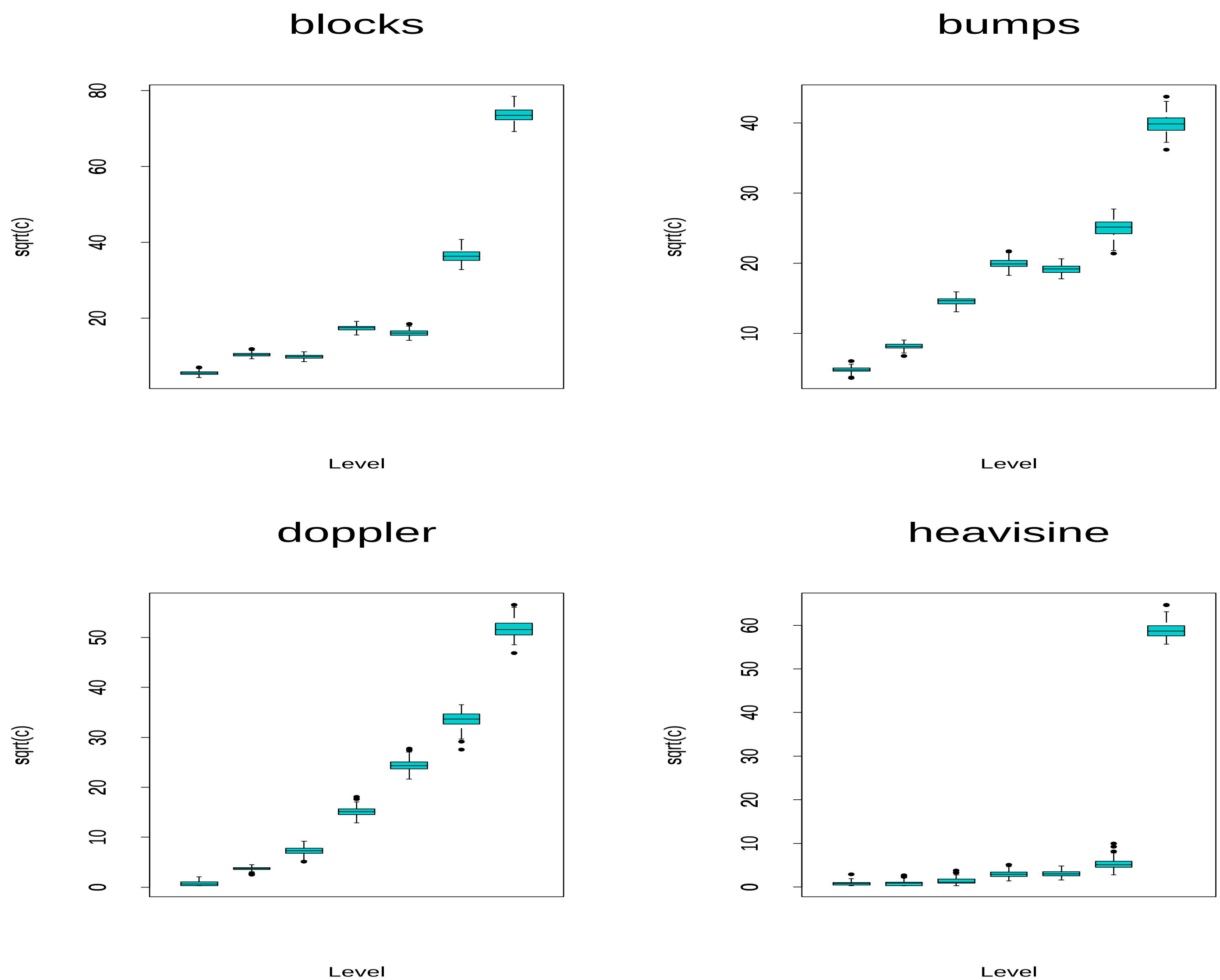
Estimate for $\boldsymbol{\sigma}^2$ uses data at all levels

Can adapt to allow level dependent variances $\boldsymbol{\sigma}_j^2$

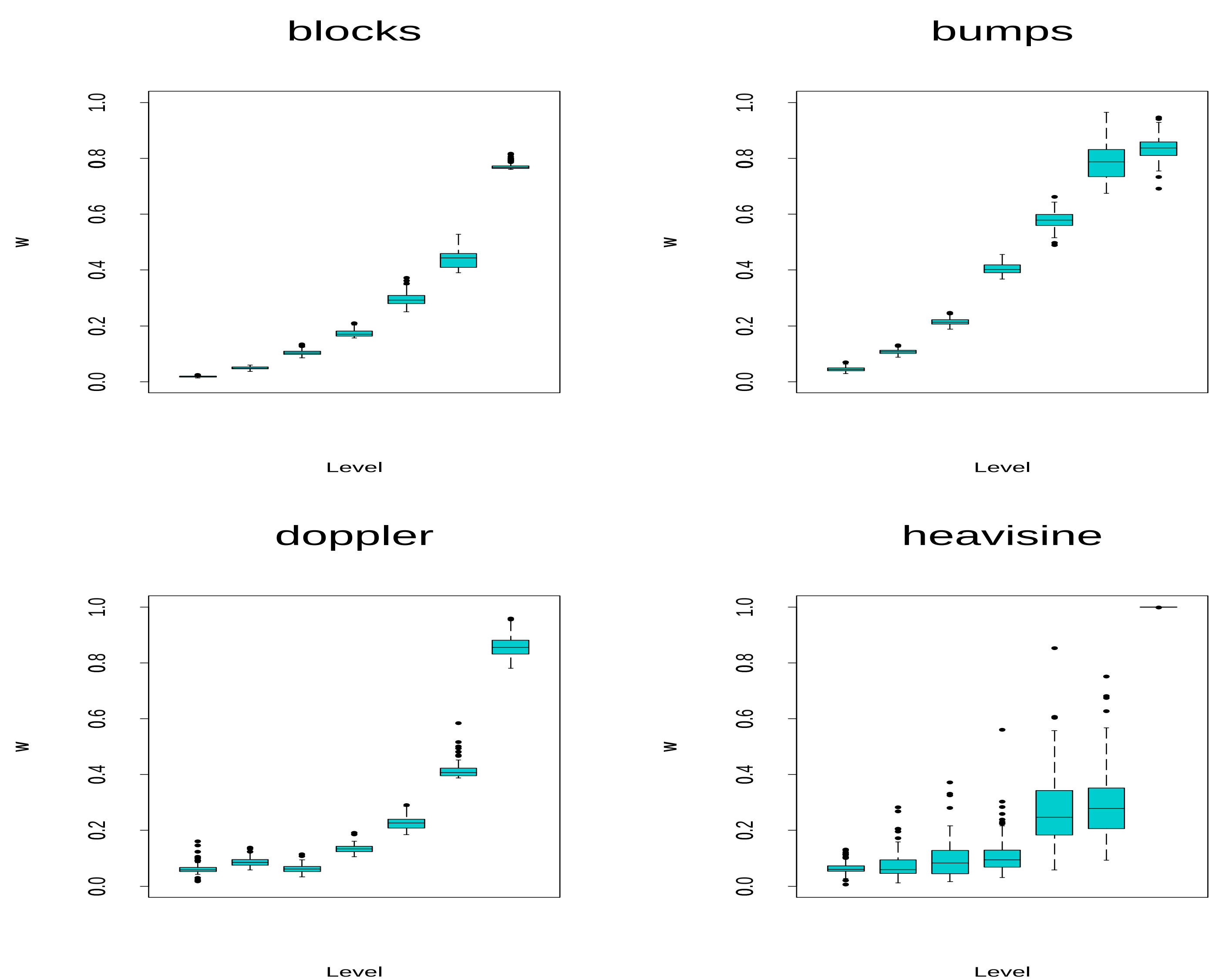
Test Functions



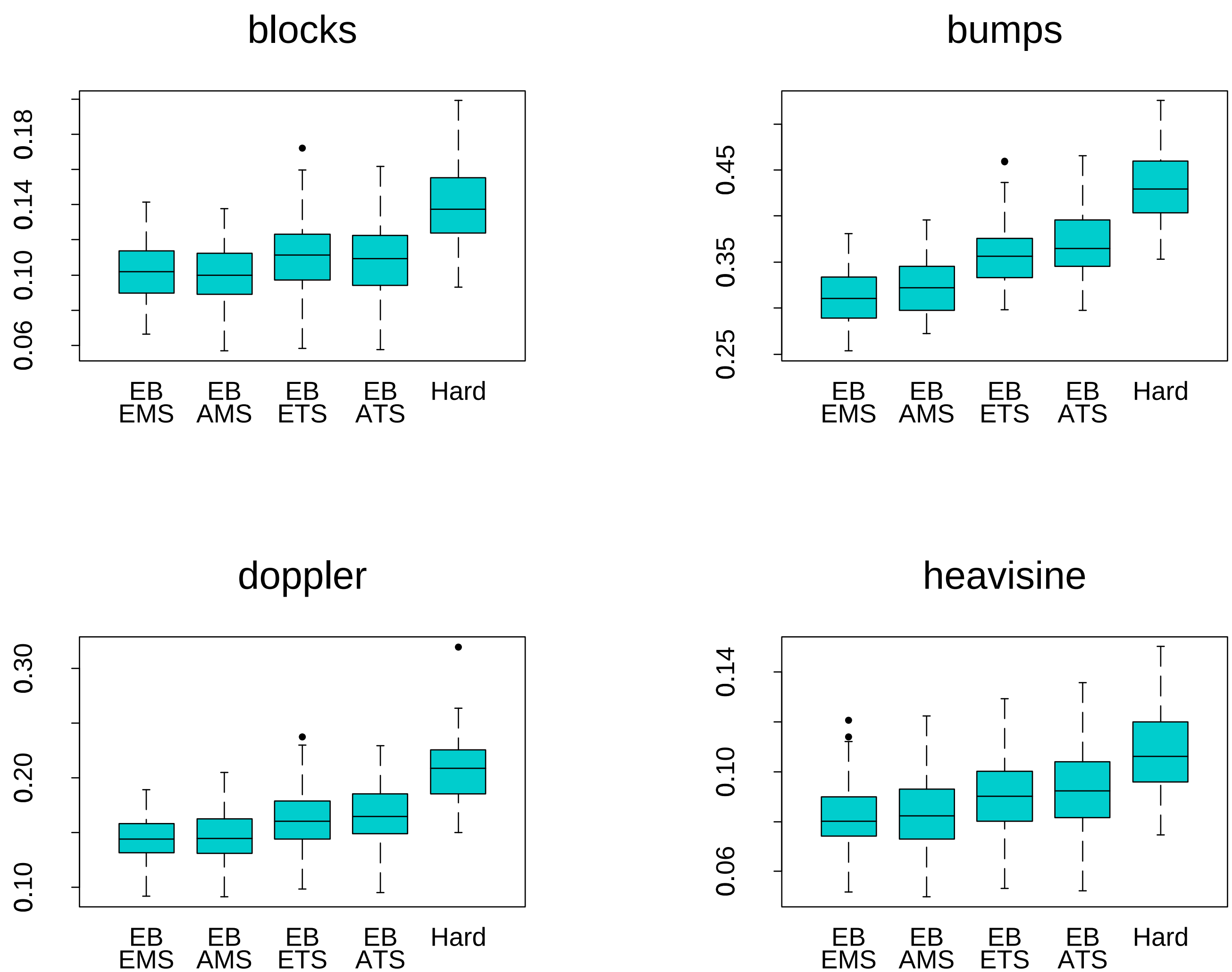
EB Estimates of $c_j^{1/2}$



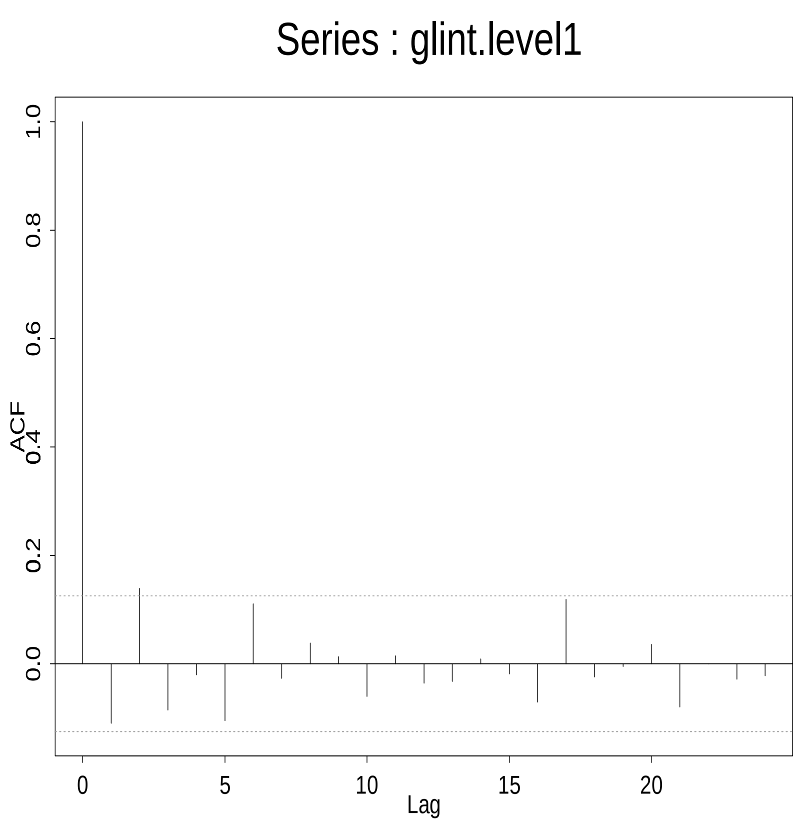
EB Estimates of ω_j

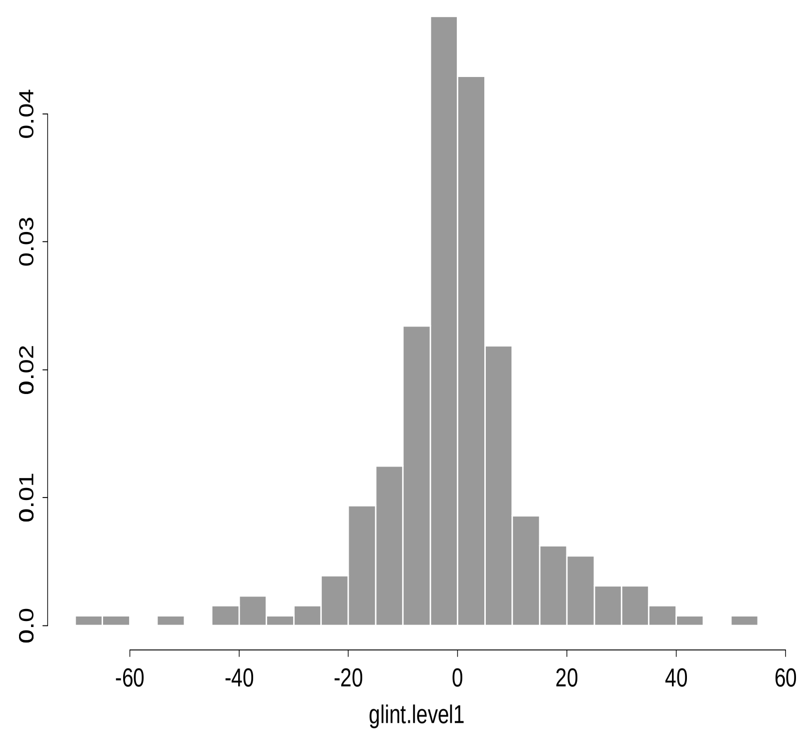


Normal Error MSE Comparisons



Distributions





Hierarchical Model for Outliers

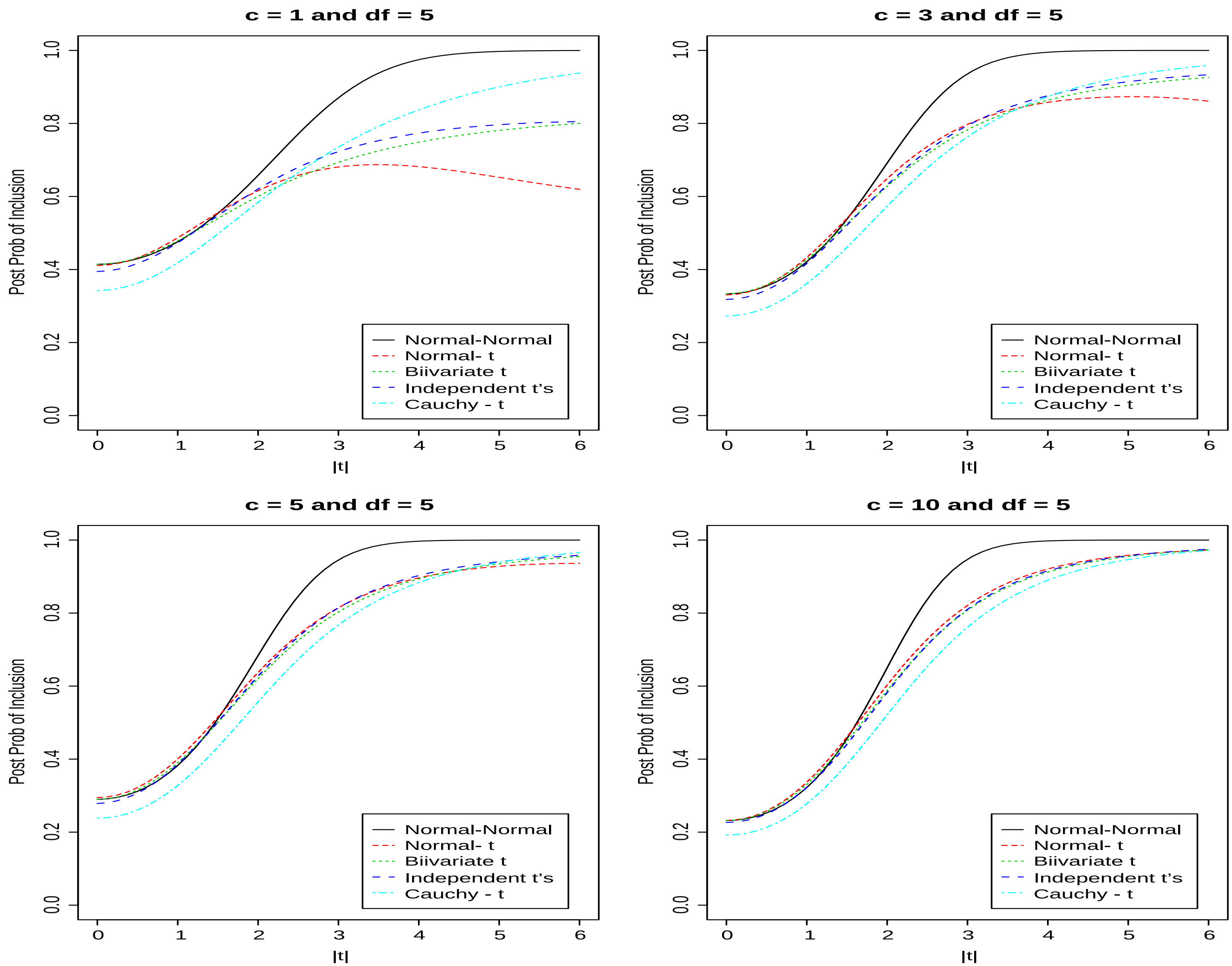
$$\begin{aligned}
\mathbf{D}_{jk} &= \boldsymbol{\beta}_{jk} + \epsilon_{jk} \\
\epsilon_{jk} | \boldsymbol{\lambda}_{jk} &\sim N\left(0, \frac{\boldsymbol{\sigma}^2}{\boldsymbol{\lambda}_{jk}}\right) \\
\boldsymbol{\beta}_{jk} | \gamma_{jk} &\sim N\left(0, \mathbf{c}_j \frac{\gamma_{jk}}{\boldsymbol{\lambda}^*_{jk}} \boldsymbol{\sigma}^2\right) \\
\gamma_{jk} &\sim Bernoulli(\omega_j)
\end{aligned}$$

$\boldsymbol{\lambda}_{jk}$ and $\boldsymbol{\lambda}^*_{jk}$ have heavy tailed distributions h and h^*, i.e.Gamma distributions

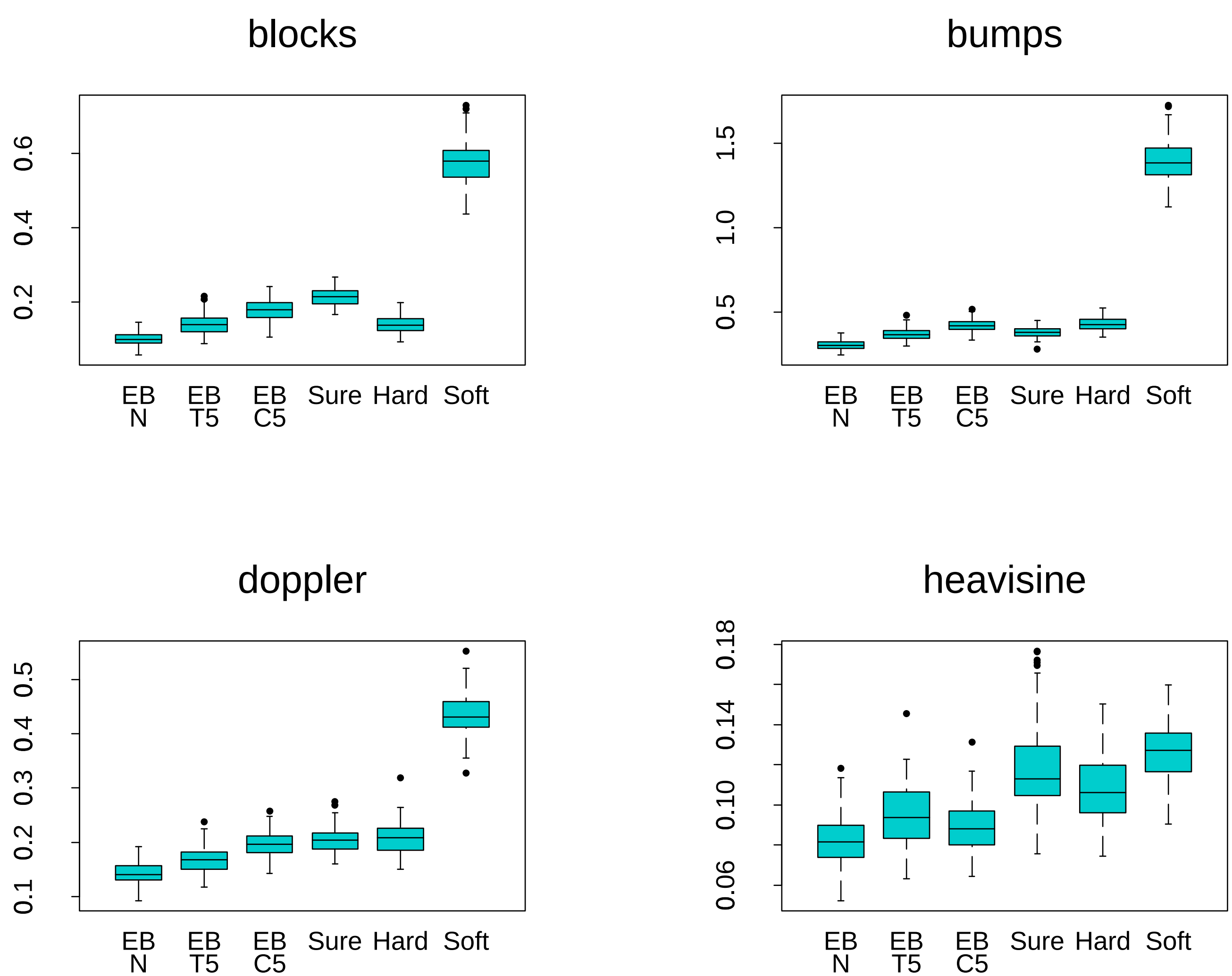
Standard Normal Model $\boldsymbol{\lambda}_{jk} = \boldsymbol{\lambda}^*_{jk} = 1$

Label	$\boldsymbol{\lambda}^*_{jk}$ and $\boldsymbol{\lambda}_{jk}$ Distributions	$\boldsymbol{\beta}_{jk}$ and ϵ_{jk} Distributions	$\boldsymbol{\beta}_{jk}$ and ϵ_{jk} Dependence
Normal-Normal	$\boldsymbol{\lambda}^*_{jk} = \boldsymbol{\lambda}_{jk} \equiv 1$	Normal prior Normal errors	Independent
Normal-t	$\boldsymbol{\lambda}^*_{jk} \equiv 1$ $\boldsymbol{\lambda}_{jk} \sim \text{Gamma}(\nu/2, 2/\nu)$	Normal prior t_ν errors	Independent
Independent t	$\boldsymbol{\lambda}^*_{jk} \sim \text{Gamma}(\nu/2, 2/\nu)$ $\boldsymbol{\lambda}_{jk} \sim \text{Gamma}(\nu/2, 2/\nu)$	t_ν prior t_ν errors	Independent
Cauchy-t	$\boldsymbol{\lambda}^*_{jk} \sim \text{Gamma}(1/2, 2)$ $\boldsymbol{\lambda}_{jk} \sim \text{Gamma}(\nu/2, 2/\nu)$	Cauchy prior t_ν errors	Independent
Bivariate t	$\boldsymbol{\lambda}^*_{jk} = \boldsymbol{\lambda}_{jk}$ $\boldsymbol{\lambda}_{jk} \sim \text{Gamma}(\nu/2, 2/\nu)$	t_ν prior t_ν errors	Uncorrelated But Dependent

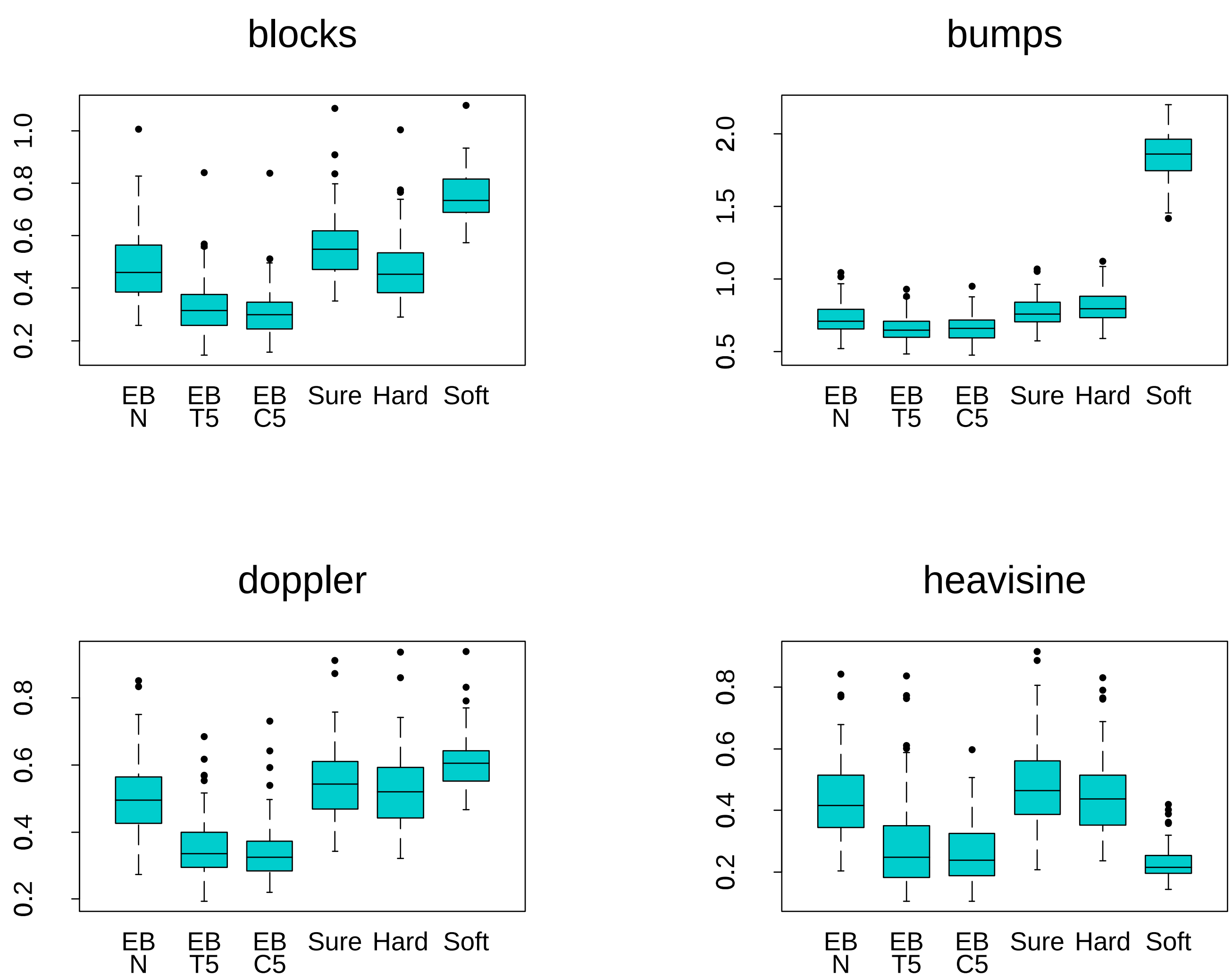
Posterior Probabilities



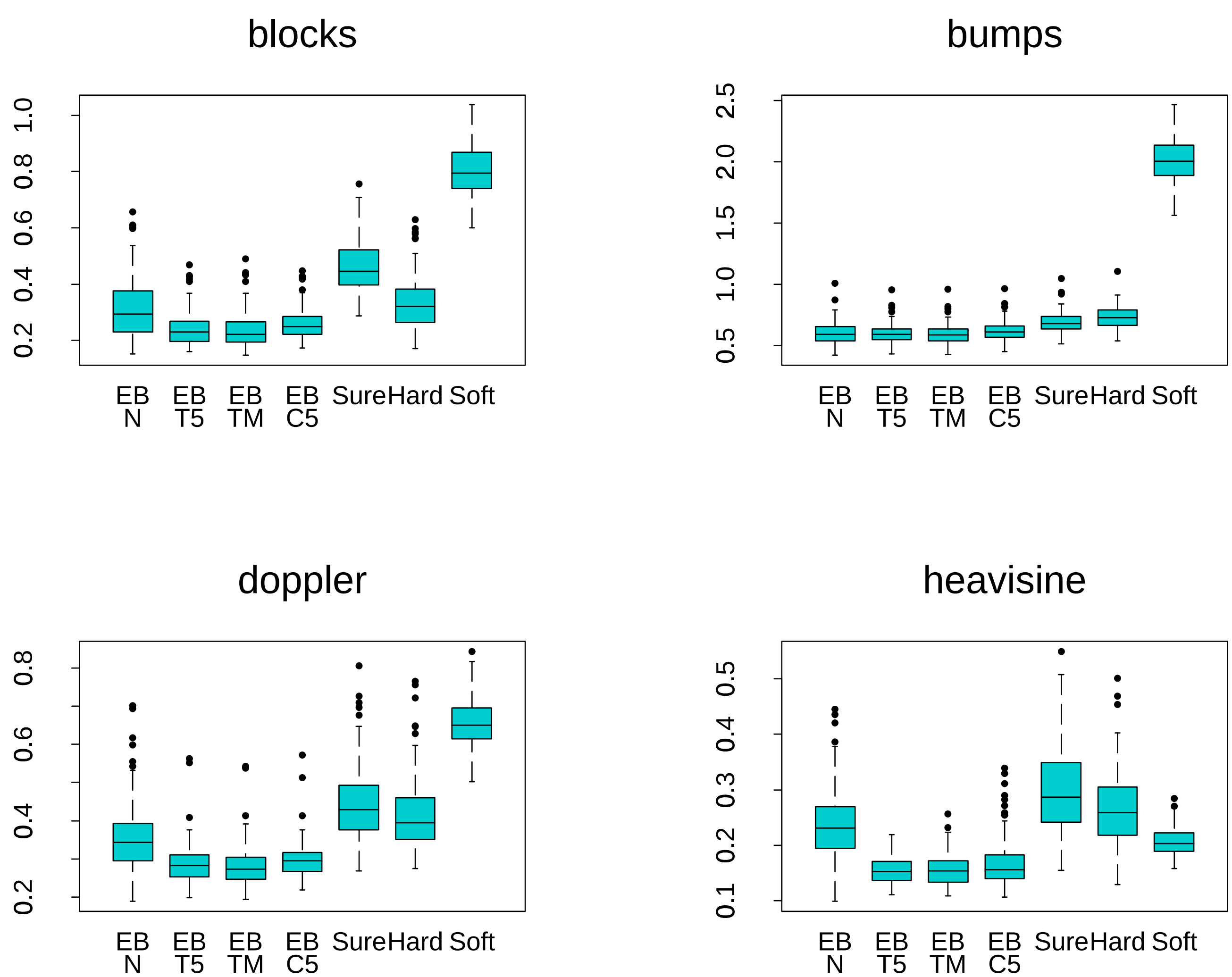
Normal Error MSE Comparisons



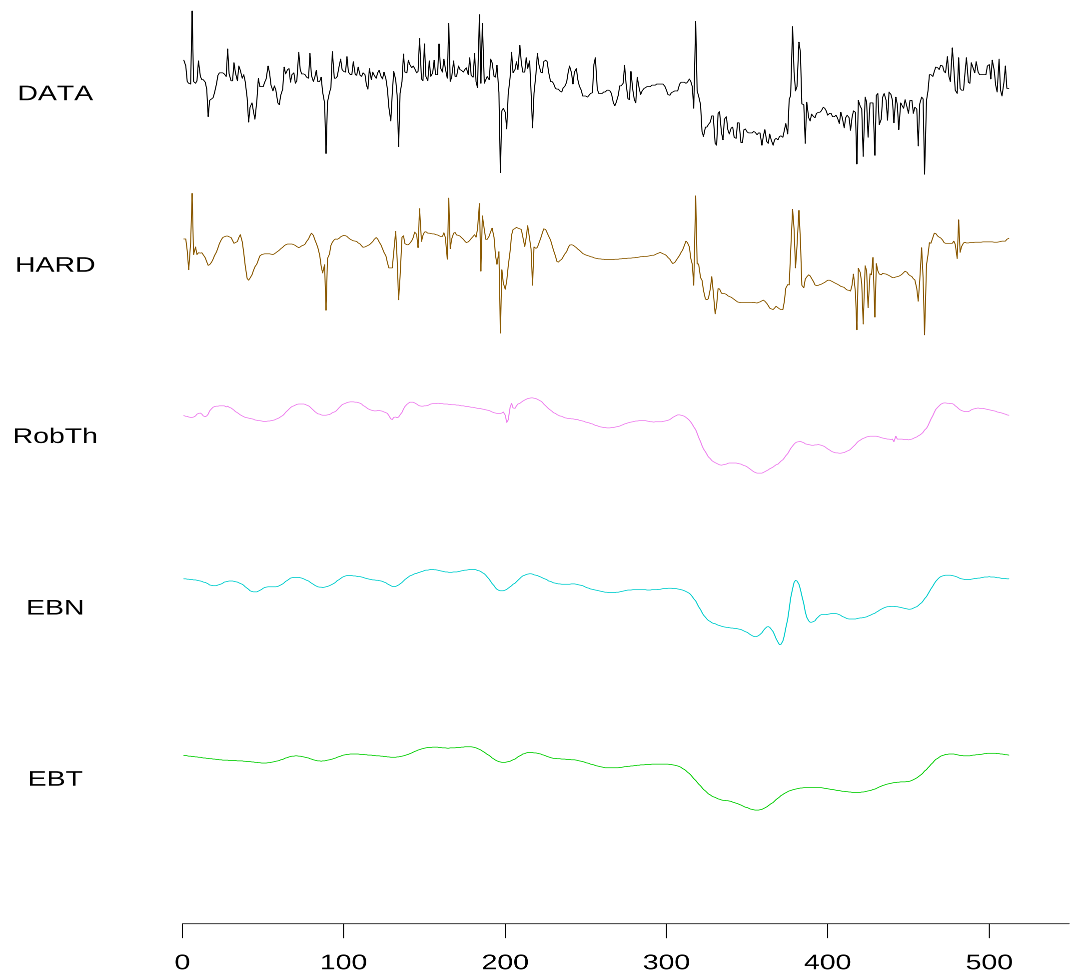
Student t Errors in Wavelet Domain – MSE Comparisons



Student t Errors in Data Domain – MSE Comparisons



Glint Estimates



Open Problems

1. Non-Decimated Wavelet Transform

    Empirical Wavelet coefficients are Dependent

2. Wavelet Packets and Basis Selection

3. Conditional Independent Priors – Tree Structures

    Clustering of Coefficients

    Persistence