FLEXIBLE EMPIRICAL BAYES ESTIMATION FOR WAVELETS

Merlise Clyde

Institute of Statistics and Decision Sciences

Duke University

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

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

WAVELET REGRESSION 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 $\mathbf{e}$: $\mathbf{Y} = \mathbf{f} + \mathbf{e}$

Apply Discrete Wavelet Transform (DWT):

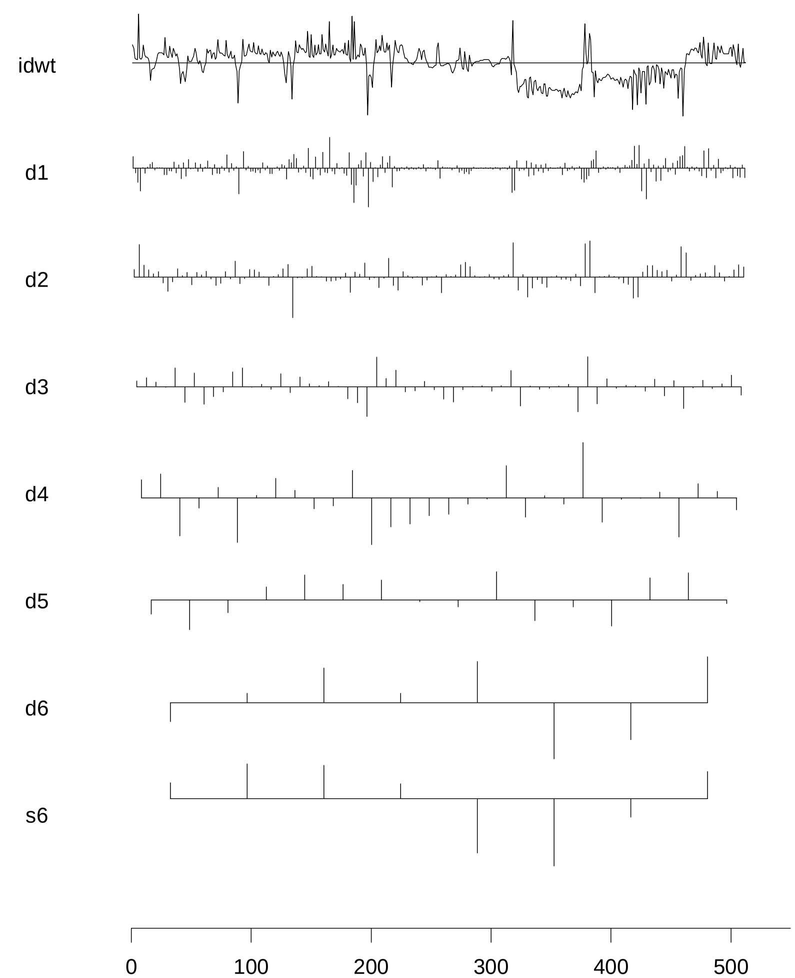
$\mathbf{W} = n \times n$ orthonormal matrix, $\mathbf{W}'\mathbf{W} = \mathbf{W}\mathbf{W}' = I$

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

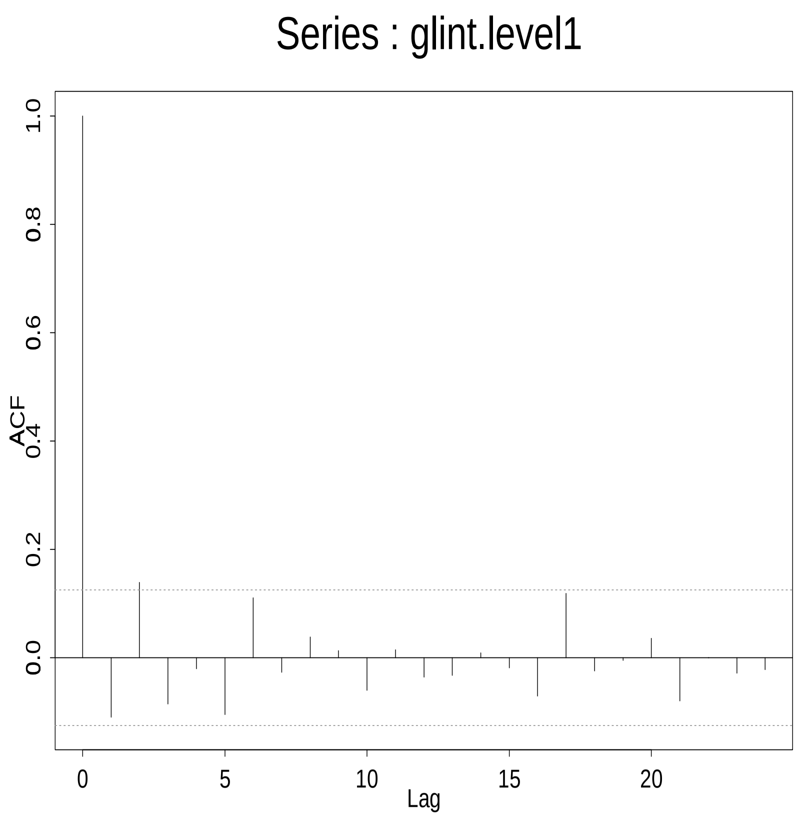
$\mathbf{W}'\mathbf{Y} \equiv \mathbf{D}$, DWT of data or empirical wavelet coefficients

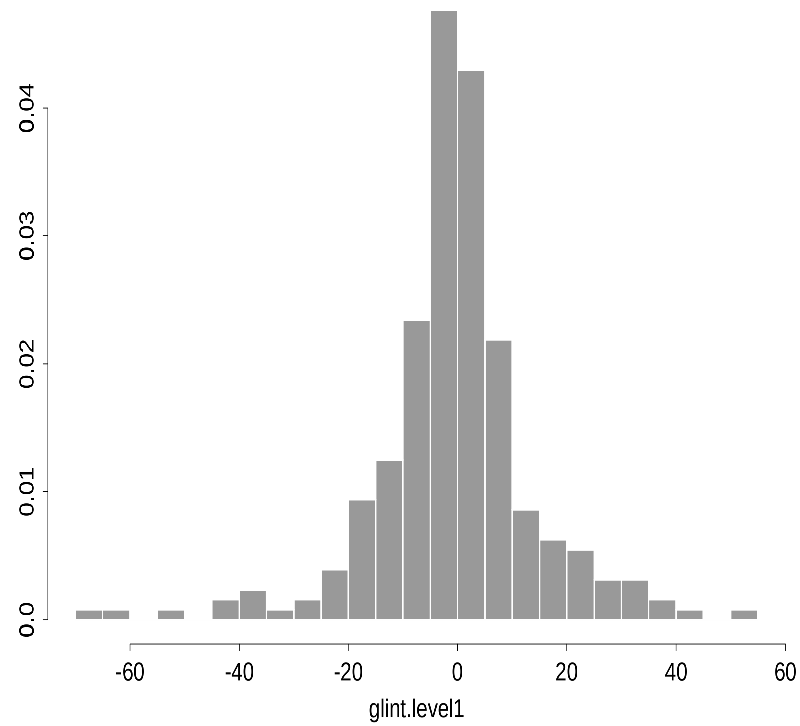
Wavelet Regression Model: $\mathbf{D}_{jk} = \boldsymbol{\beta}_{jk} + \boldsymbol{\epsilon}_{jk}$

Multi-resolution Discrete Wavelet Transformation



Distributions





Uncorrelated

Heavy Tails and Outliers

Hierarchical Model in Wavelet Domain

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

$(\boldsymbol{\lambda}_{jk}, \boldsymbol{\lambda}^*_{jk}) \sim h$ i.e.independent Gamma distributions

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

Parameters: $\boldsymbol{\sigma}^2, \mathbf{c}_j$, and ω_j

Posterior Distribution of γ

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

$$\pi(\gamma_{jk} = 1|\mathbf{Y}) = \frac{\mathcal{O}_{jk}}{1 + \mathcal{O}_{jk}} \qquad \mathcal{O}_{jk} = \left(\frac{\omega_j}{1-\omega_j}\right)\frac{m_1\left(\mathbf{D}_{jk}|\gamma_{jk} = 1\right)}{m_0\left(\mathbf{D}_{jk}|\gamma_{jk} = 0\right)}$$

m_0 is the marginal distribution of the data when $\gamma_{jk} = 0$:

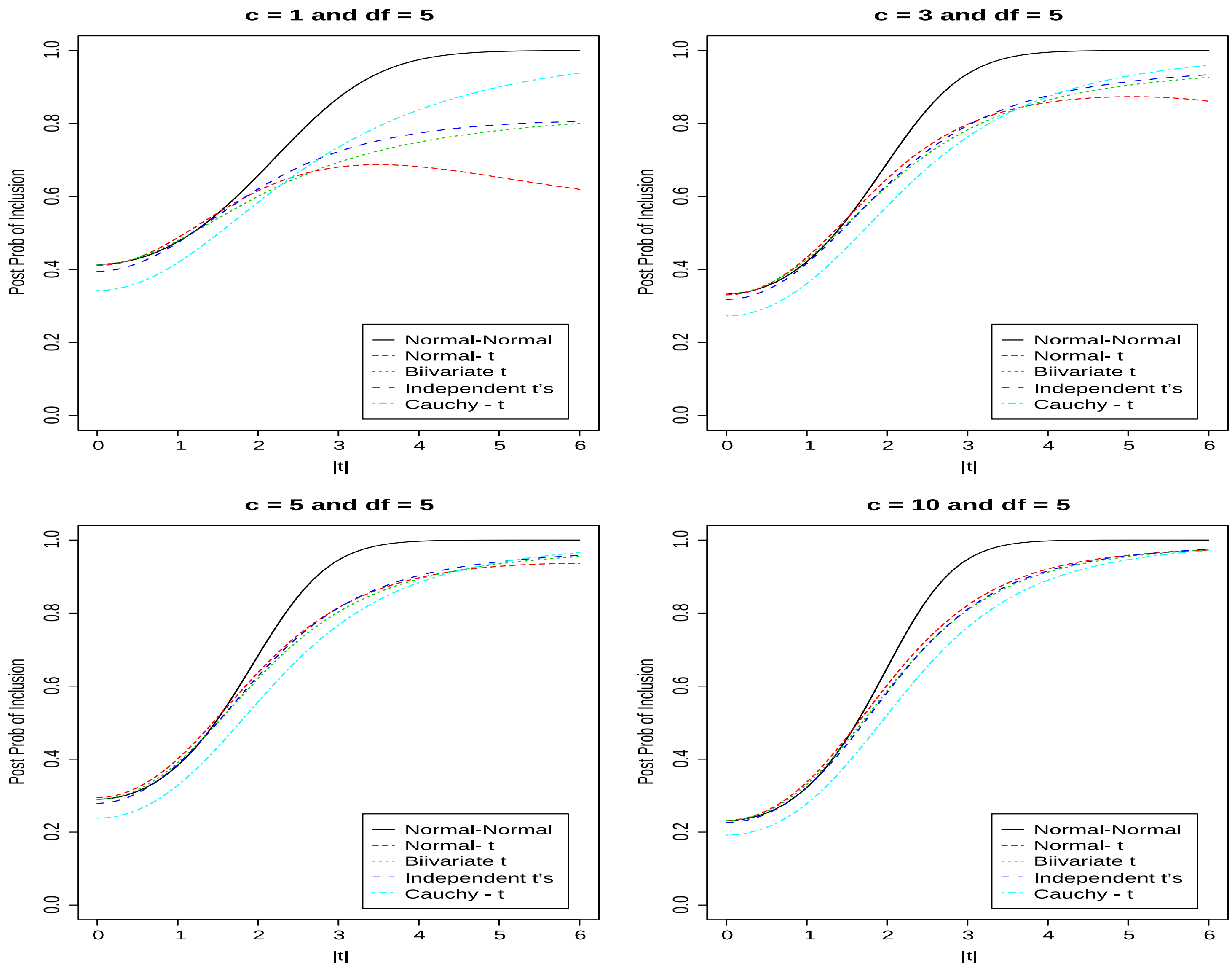
$$m_0\left(\mathbf{D}_{jk}|\gamma_{jk} = 0\right) = \int \left(\frac{\boldsymbol{\lambda}_{jk}}{2\pi\boldsymbol{\sigma}^2}\right)^{1/2} \exp\left\{-\frac{1}{2}\mathbf{D}_{jk}^2\boldsymbol{\lambda}_{jk}/\boldsymbol{\sigma}^2\right\} h(d\boldsymbol{\lambda}_{jk})$$

and m_1 is the marginal distribution of the data when $\gamma_{jk} = 1$:

$$\int\int \frac{1}{\sqrt{2\pi\boldsymbol{\sigma}^2}}\left(\frac{\mathbf{c}_j}{\boldsymbol{\lambda}_{jk}^*} + \frac{1}{\boldsymbol{\lambda}_{jk}}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{D}_{jk}^2\left(\frac{\boldsymbol{\sigma}^2\mathbf{c}_j}{\boldsymbol{\lambda}_{jk}^*} + \frac{\boldsymbol{\sigma}^2}{\boldsymbol{\lambda}_{jk}}\right)^{-1}\right\} h(d\boldsymbol{\lambda}_{jk}^*, d\boldsymbol{\lambda}_{jk})$$

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

Posterior Probabilities



Bayes Threshold and Shrinkage Estimators

Multiple Shrinkage Estimator (BMA):

Posterior Mean (minimizes posterior expected squared error loss):

$$\hat{\beta}_{jk} = \pi(\gamma_{jk} = 1 | \mathbf{D}) E[\beta_{jk} | \mathbf{D}_{jk}, \gamma_{jk} = 1]$$

Threshold Estimator (Model Selection):

Posterior Mean under highest Posterior Probability model

(optimal under squared error loss with model selection)

$$\hat{\beta}_{jk} = \hat{\gamma}_{jk} E[\beta_{jk} | \mathbf{D}_{jk}, \gamma_{jk} = 1]$$

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. Re-Introduce "latent data" $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$; In Normal-Normal and Bivariate t complete data $(\mathbf{D}, \gamma, \boldsymbol{\lambda})$ are from a regular exponential family
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 $(\mathbf{c}, \omega, \boldsymbol{\sigma}^2)$

EM Algorithm for Bivariate t

E-step:

$$E(\boldsymbol{\lambda}_{jk}\gamma_{jk}|\mathbf{D}) = \frac{\nu_j + 1}{\nu_j + \mathbf{D}^2_{jk}\hat{\boldsymbol{\tau}}_j^{(i)}}\hat{\gamma}^{(i)}_{jk} = [\hat{\boldsymbol{\lambda}}_{jk}\hat{\gamma}_{jk}]^{(i)} \qquad \boldsymbol{\tau}_j = [(1+\mathbf{c}_j)\boldsymbol{\sigma}^2]^{-1}$$

$$E(\boldsymbol{\lambda}_{jk}(1-\gamma_{jk})|D) = \frac{\nu_j + 1}{\nu_j + \mathbf{D}^2_{jk}\hat{\boldsymbol{\phi}}^{(i)}}(1-\hat{\gamma}^{(i)}_{jk}) = [\hat{\boldsymbol{\lambda}}_{jk}(1-\hat{\gamma}_{jk})]^{(i)} \qquad \boldsymbol{\phi} = 1/\boldsymbol{\sigma}^2$$

$$\hat{\gamma}^{(i)}_{jk} = \pi\left(\gamma_{jk} = 1|\mathbf{D}_{jk}, \hat{\mathbf{c}}_j^{(i)}, \hat{\omega}_j^{(i)}\hat{\boldsymbol{\sigma}}^{(i)}\right)$$

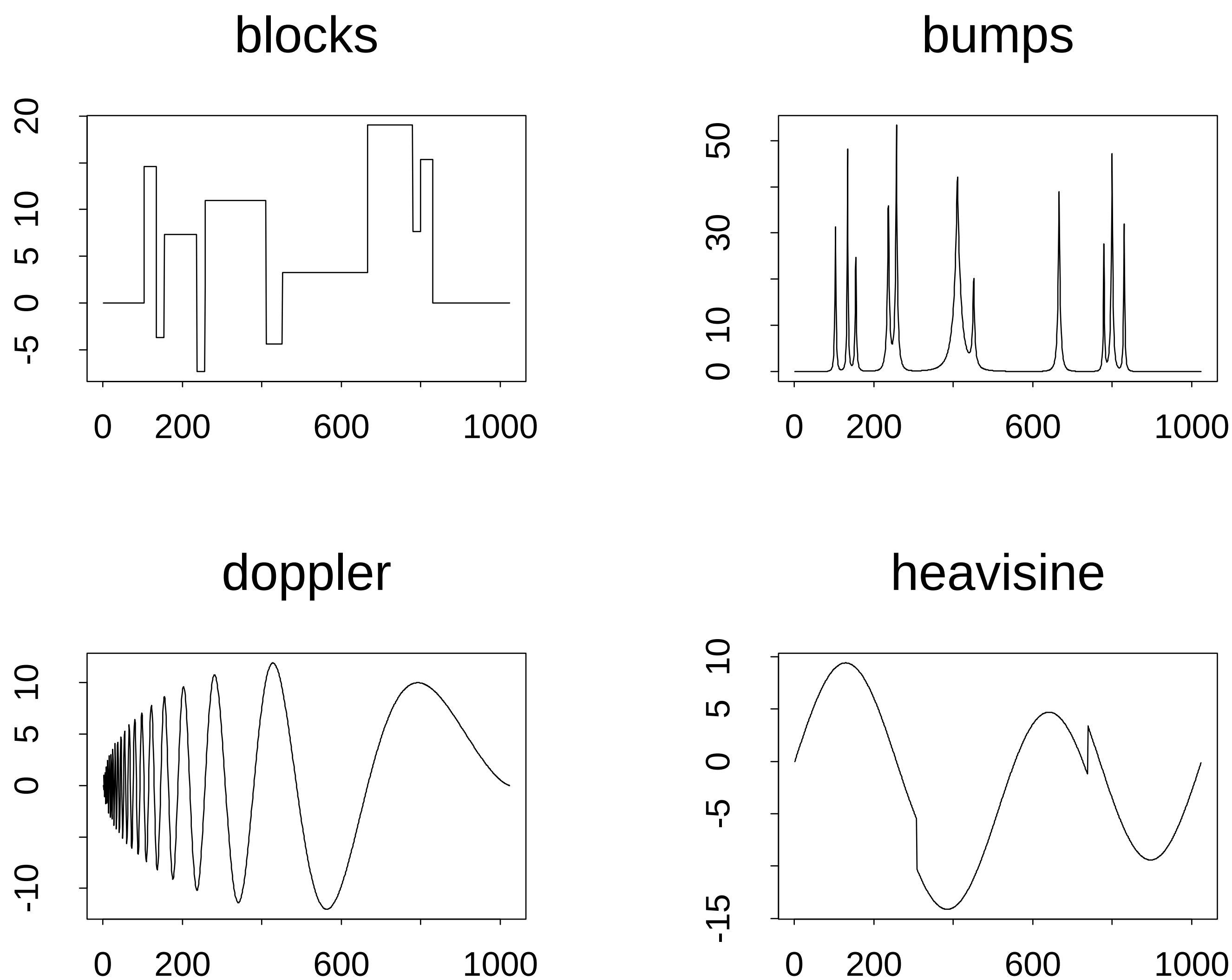
M-Step:

$$\hat{\boldsymbol{\sigma}}^{2(i+1)} = \frac{\sum_{jk}\mathbf{D}^2_{jk}[\hat{\boldsymbol{\lambda}}_{jk}(1-\hat{\gamma}_{jk})]^{(i)}}{n - \sum_{jk}\hat{\gamma}^{(i)}_{jk}}$$

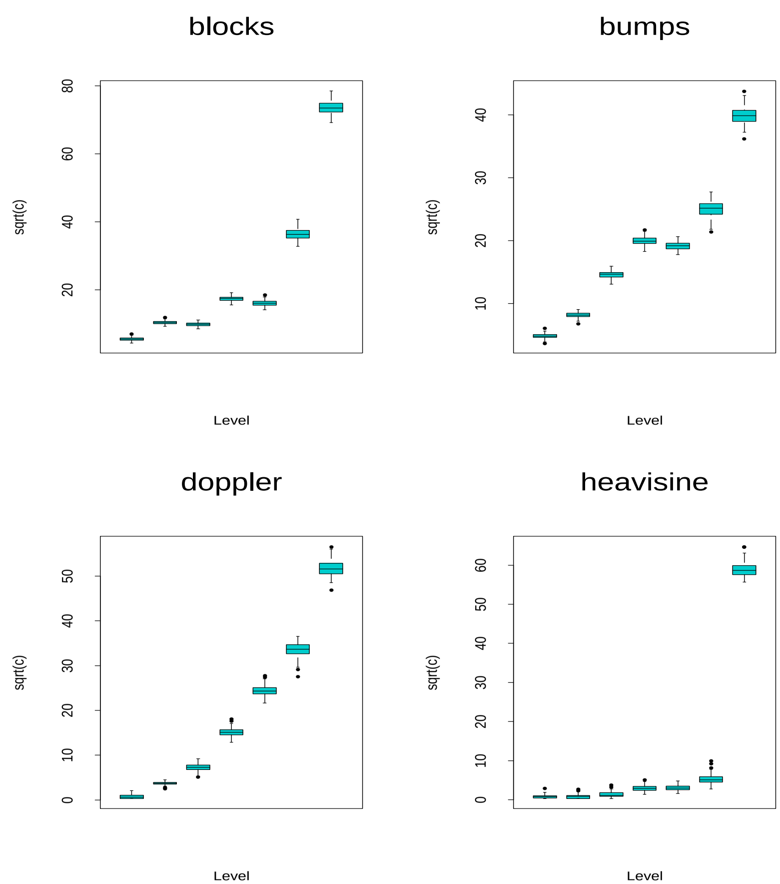
$$\hat{\mathbf{c}}_j^{(i+1)} = \max\left(0, \frac{\sum_k \mathbf{D}^2_{jk}[\hat{\boldsymbol{\lambda}}_{jk}\hat{\gamma}_{jk}]^{(i)}}{\hat{\boldsymbol{\sigma}}^{2(i+1)}\sum_k \hat{\gamma}^{(i)}_{jk}} - 1\right)$$

$$\hat{\omega}_j^{(i+1)} = \frac{\sum_k \hat{\gamma}^{(i)}_{jk}}{n_j}$$

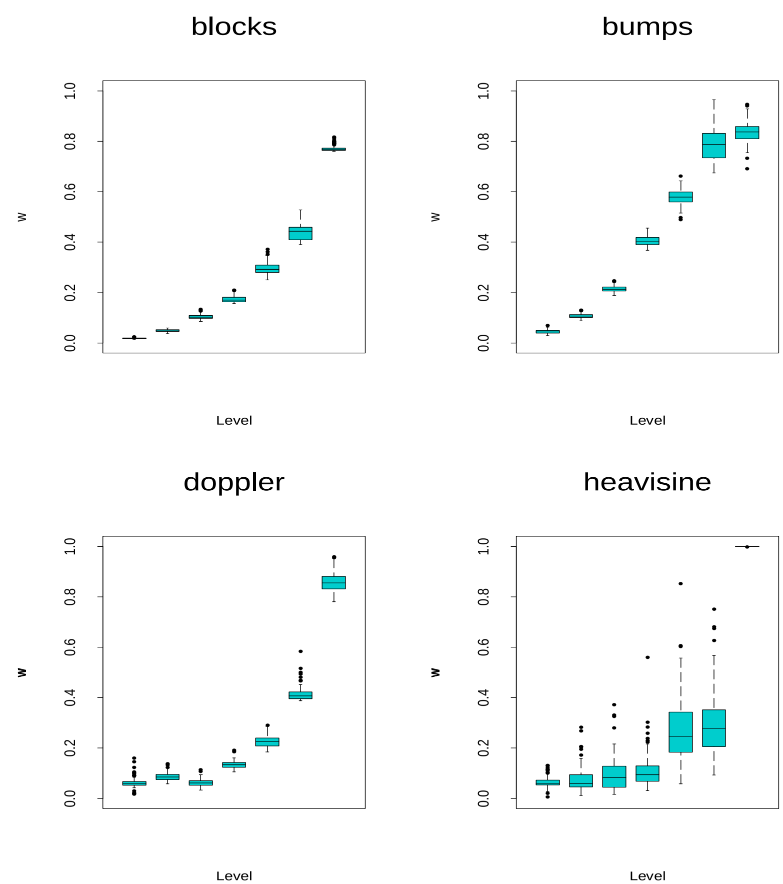
Test Functions



EB Estimates by Level

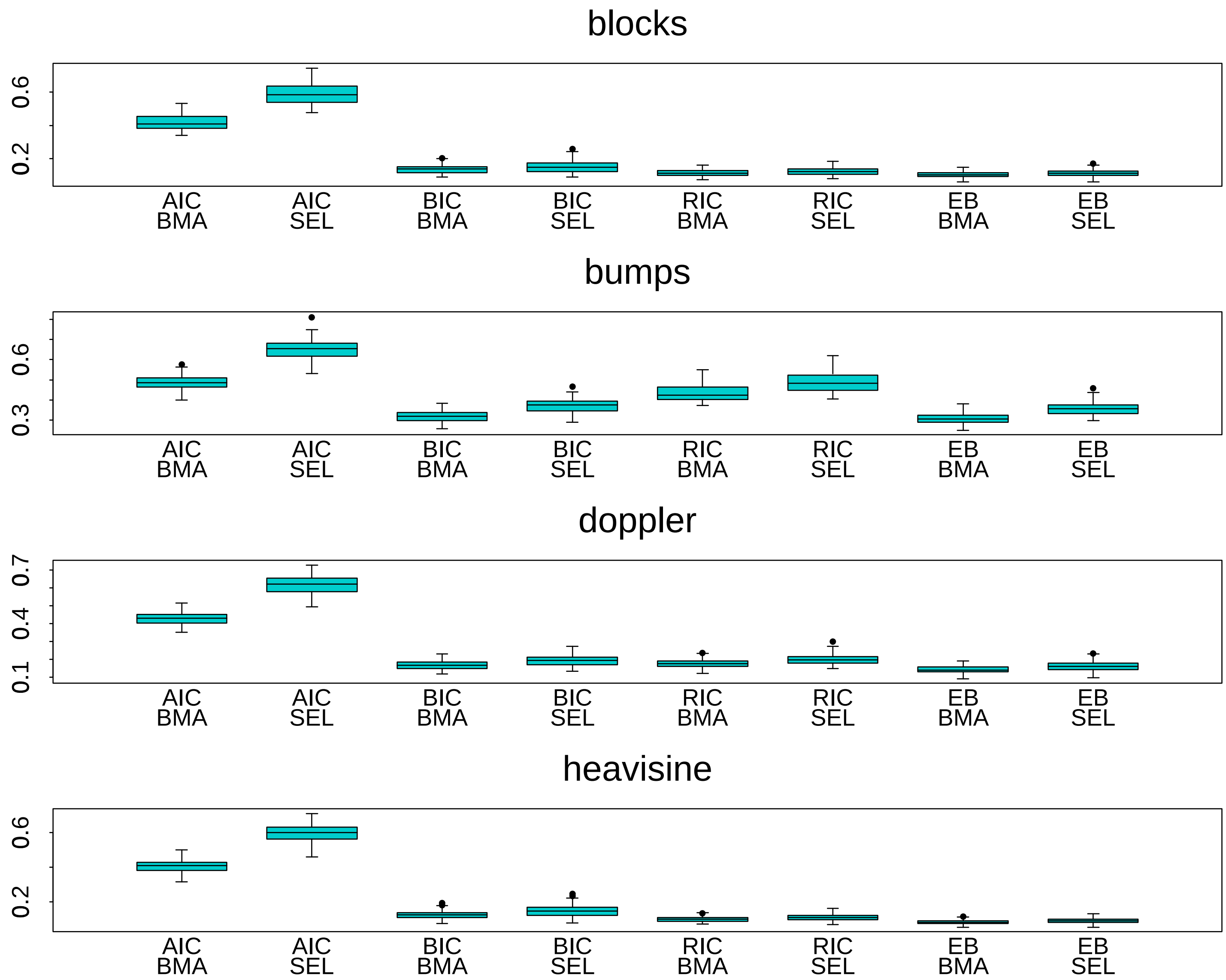


$\mathbf{c}_j^{1/2}$

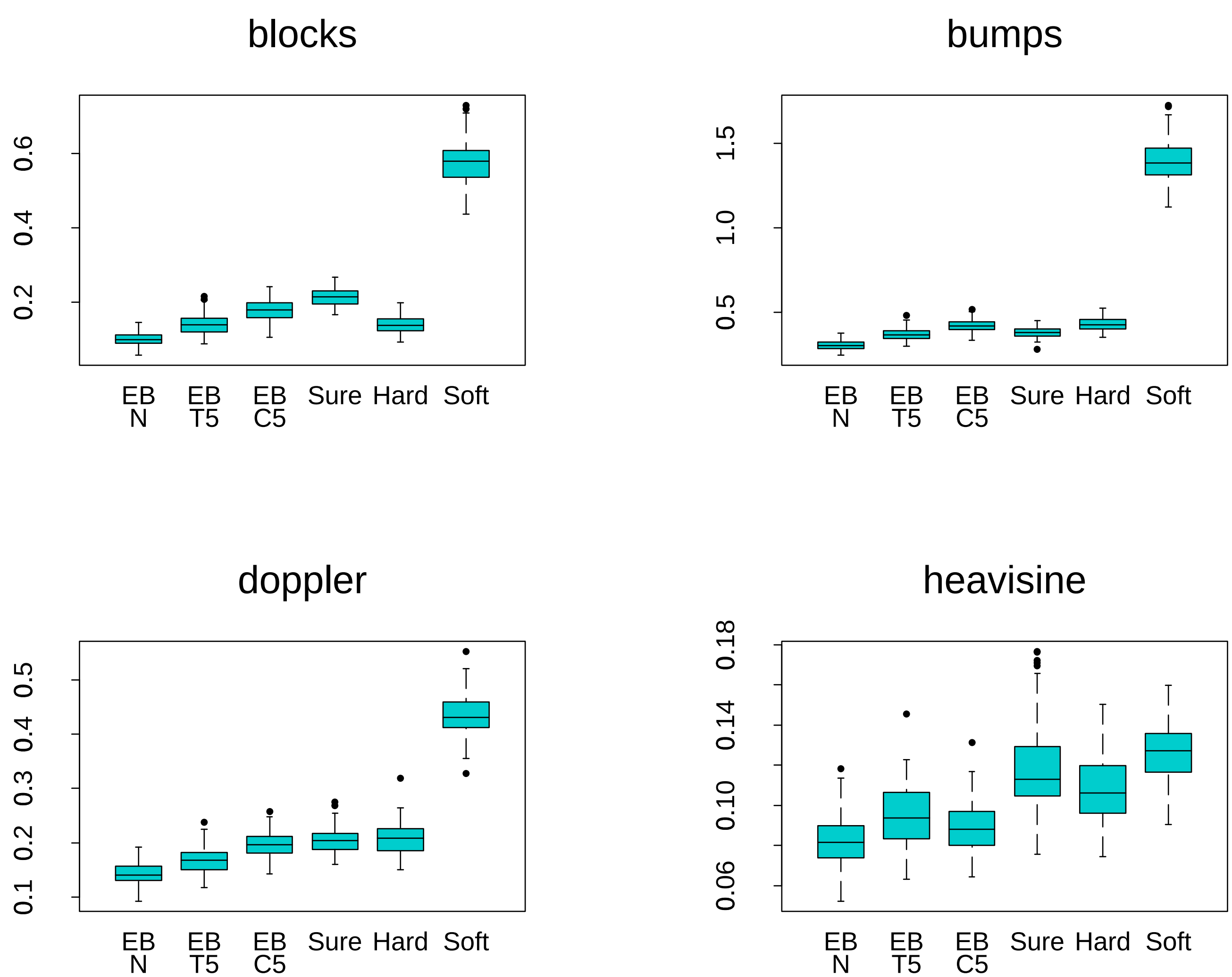


ω_j

Normal Error MSE Comparisons



Normal Error MSE Comparisons



Student t Errors in Data Domain – MSE Comparisons