Shrinkage Estimation Toward the Data Chosen Reduced Model with Applications to Wavelet Analysis and Principal Component Regression Analysis

Huibin Zhou

Department of Mathematics

Cornell University

Feb 25, 2002

Joint with J.T. Gene Hwang

Overview

1. Motivation
2. Main Results
3. Some Interpretations
4. Simulation Studies
5. Related work
6. Discussion

Motivation

We have the following Gaussian regression model:

$$Z_i = \theta_i + \xi_i,\ i = 1, 2, ..., n$$

where ξ_i are i.i.d. $N(0, 1)$.

- Shrinkage Estimators based on asymptotia , including Hard thresholding, Soft thresholding, and many others, can not dominate the naive estimator uniformly. Sometimes those estimators are even much worse than the naive etimator .

- James -Stein Estimator

$$\hat{\theta}_i = \left(1 - \frac{a}{\sum_{i=1}^{n} Z_i^2}\right)_+ Z_i, \ \ i = 1, 2, ..., n$$

$0 < a \leq 2\,(n-2)\,,\ n \geq 3$

dominates the naive estimator uniformly. But James-Stein estimator selects only between two models: the origin and the full model.

Main Results

Theorem 1. Let

$$g_i(Z) = \left(1 - a\frac{|Z_i|^{\beta-2}}{\sum_{i=1}^{n}|Z_i|^{\beta}}\right)_+ Z_i,$$

then $E_\theta||g_i(Z) - \theta||^2 < n$, for any $\theta \in R^n$,if and only if

$$0 < a \leq 2\left(\beta - 1\right)B - 2\beta,$$

where

$$B = \inf_\theta \frac{E_\theta\left(\sum_{i=1}^{n}|Z_i|^{\beta-2} / \left(\sum_{i=1}^{n}|Z_i|^{\beta}\right)\right)}{E_\theta\left(\sum_{i=1}^{n}|Z_i|^{2\beta-2} / \left(\sum_{i=1}^{n}|Z_i|^{\beta}\right)^2\right)}$$
$$\geq n.$$

Remark: In particular, for $0 < a \leq 2(\beta - 1)n - 2\beta$, the shrinkage estimators given above dominate the naive estimator, and for $\beta = 2$, $0 < a \leq 2n - 2$, that is exactly James-Stein estimator.

Theorem 2. Assume the prior distribution that θ_i are i.i.d. $N(0, \tau^2)$. Then for each $\beta \geq \frac{1}{2}$ the Bayes risk of the shrinkage estimaor is no greater than the naive estimator if and only if

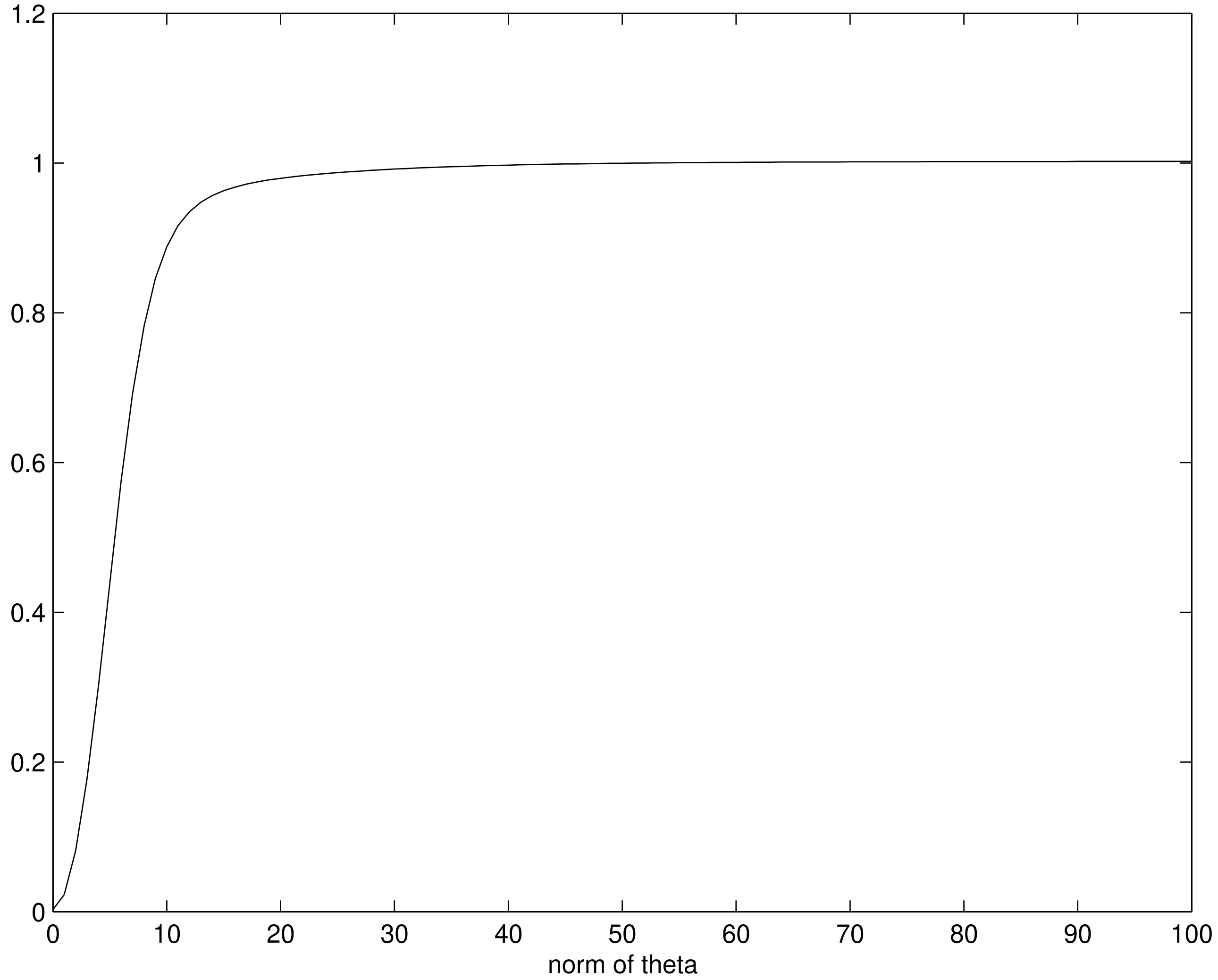
$$0 \leq a \leq \frac{2}{E\frac{\sum_{i=1}^n |\xi_i|^{2(\beta-2)}}{\left(\sum_{i=1}^n |\xi_i|^\beta\right)^2}}$$

and we know

$$\lim_{n\to\infty} \frac{2}{nE\frac{\sum_{i=1}^n |\xi_i|^{2(\beta-2)}}{\left(\sum_{i=1}^n |\xi_i|^\beta\right)^2}} = \frac{4\left(\Gamma\left(\frac{\beta}{2}+\frac{1}{2}\right)\right)^2}{\pi^{1/2}\Gamma\left(\beta-\frac{1}{2}\right)},$$

where ξ_i are i.i.d. $N(0,1)$.

Figure 1. Relative Frequentist risk of the proposed estimate with ($\beta = 4/3$ and $a = 5(n-2)/3$) to $p = 50$



Some Interpretations

- Brown (1971 Ann Stat)

$$\begin{aligned} \pi(\theta) &= c, \ \ \|\theta\|_\beta \ \leq 1 \\ &= c/\|\theta\|_\beta^a, \ \ \|\theta\|_\beta \ \geq 1 \end{aligned}$$

where C is the constant for normalization. The Bayes estimator is given by

$$Z + \nabla \log m(Z)$$

where

$$m(Z) = \int \cdots \int \frac{e^{-\frac{1}{2}\|Z-\theta\|^2}}{\left(\sqrt{2\pi}\right)^n} \pi(\theta)\, d\theta.$$

Theorem 3

$$\lim_{|Z_i|\to\infty} \frac{\nabla_i \log m(Z)}{\nabla_i \log \pi(Z)} = 1.$$

Then the ith component of the Bayes estimator equals approximately

$$Z_i + \nabla_i \log \pi(Z) = Z_i - a\frac{sign(Z_i)\,|Z_i|^{\beta-1}}{\sum_{i=1}^{n} |Z_i|^{\beta}}$$

And the positive part is

$$\left(1 - a\frac{|Z_i|^{\beta-2}}{\sum_{i=1}^{n} |Z_i|^{\beta}}\right)_+ Z_i.$$

- Mallat (1989 IEEE PAMI)

$$\pi\left(\theta\right) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|\theta|}{\alpha}\right)^{\beta}}$$

Similar argument leads to an approximate Bayes estimator

$$Z_i - \frac{\beta}{\alpha^{\beta}} sign\left(Z_i\right)\left|Z_i\right|^{\beta-1}.$$

And MLE will tell us α^{β} is approximately

$$\frac{\beta \sum_{i=1}^{n}\left|Z_i\right|^{\beta}}{n}.$$

Then the empirical Bayes estimator is approximately equal to

$$Z_i - n\frac{sign\left(Z_i\right)\left|Z_i\right|^{\beta-1}}{\sum_{i=1}^{n}\left|Z_i\right|^{\beta}}.$$

And the positive part is

$$\left(1 - n\frac{\left|Z_i\right|^{\beta-2}}{\sum_{i=1}^{n}\left|Z_i\right|^{\beta}}\right)_{+} Z_i.$$

- Johnstone (1999)

$$\delta_i(Z) = Z_i - a sign(Z_i) |Z_i|^{\beta-1}.$$

where it is a nonlinear form if $\beta \neq 2$. The unbiased estimate of the risk $U(a, Z)$ of this estimator is

$$U(a, Z) = n - 2a\beta \sum_{i=1}^{n} Z_i^{\beta-2} + a^2 ||Z^{\beta-1}||_2^2.$$

And $U(a, Z) \leq n$ implies

$$0 \leq a \leq \frac{2\beta \sum_{i=1}^{n} |Z_i|^{\beta-2}}{\sum_{i=1}^{n} |Z_i|^{2\beta-2}}$$

where

$$\begin{aligned}
\frac{2\beta \sum_{i=1}^{n} |Z_i|^{\beta-2}}{\sum_{i=1}^{n} |Z_i|^{2\beta-2}} &= \frac{2\beta \sum_{i=1}^{n} |Z_i|^{\beta-2} \sum_{i=1}^{n} |Z_i|^{\beta}}{\sum_{i=1}^{n} |Z_i|^{2\beta-2} \sum_{i=1}^{n} |Z_i|^{\beta}} \\
&\geq \frac{2\beta n}{\sum_{i=1}^{n} |Z_i|^{\beta}}
\end{aligned}$$

Simulation Studies

- Wavelet regression. Wavelet transformation is a linear transformation like Fourier transformation. You can think of it as an action of a special orthonormal matrix $A_{n\times n}$.
  In general, wavelet regression has the folloing three steps

Wavelet transformation: $A_{n\times n}Y_{n\times 1} = Z_{n\times 1}$

$$\downarrow$$

$$\begin{pmatrix} Z_1, ..., Z_{2^k}, Z_{2^k+1}, ..., Z_{2^{k+1}}, \\ Z_{2^{k+1}+1}, ..., Z_{2^{k+2}}, ..., Z_{2^{m-1}+1}, ..., Z_{2^m} \end{pmatrix}^T$$

$$\rightarrow \begin{pmatrix} \widehat{\theta}_1, ..., \widehat{\theta}_{2^k}, \widehat{\theta}_{2^k+1}, ..., \widehat{\theta}_{2^{k+1}}, \\ \widehat{\theta}_{2^{k+1}+1}, ..., \widehat{\theta}_{2^{k+2}}, ..., \widehat{\theta}_{2^{m-1}+1}, ..., \widehat{\theta}_{2^m} \end{pmatrix}^T$$

$$\downarrow$$

Inverse transformation: $A^T_{n\times n}\widehat{\theta}_{n\times 1} = \widehat{f}_{n\times 1}$

Figure 2. The curves represent the true curves.

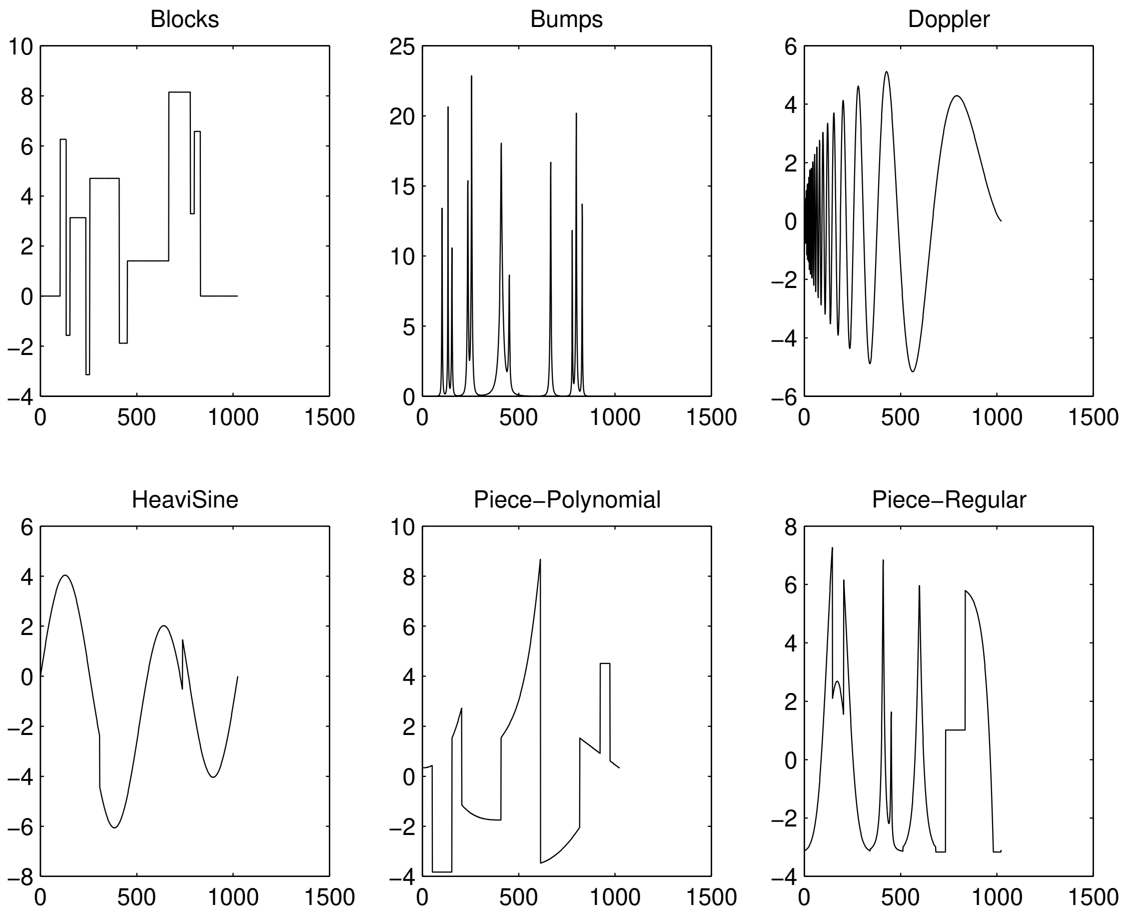


Figure 3. The data are generated from the true curves with random noise N(0,1).

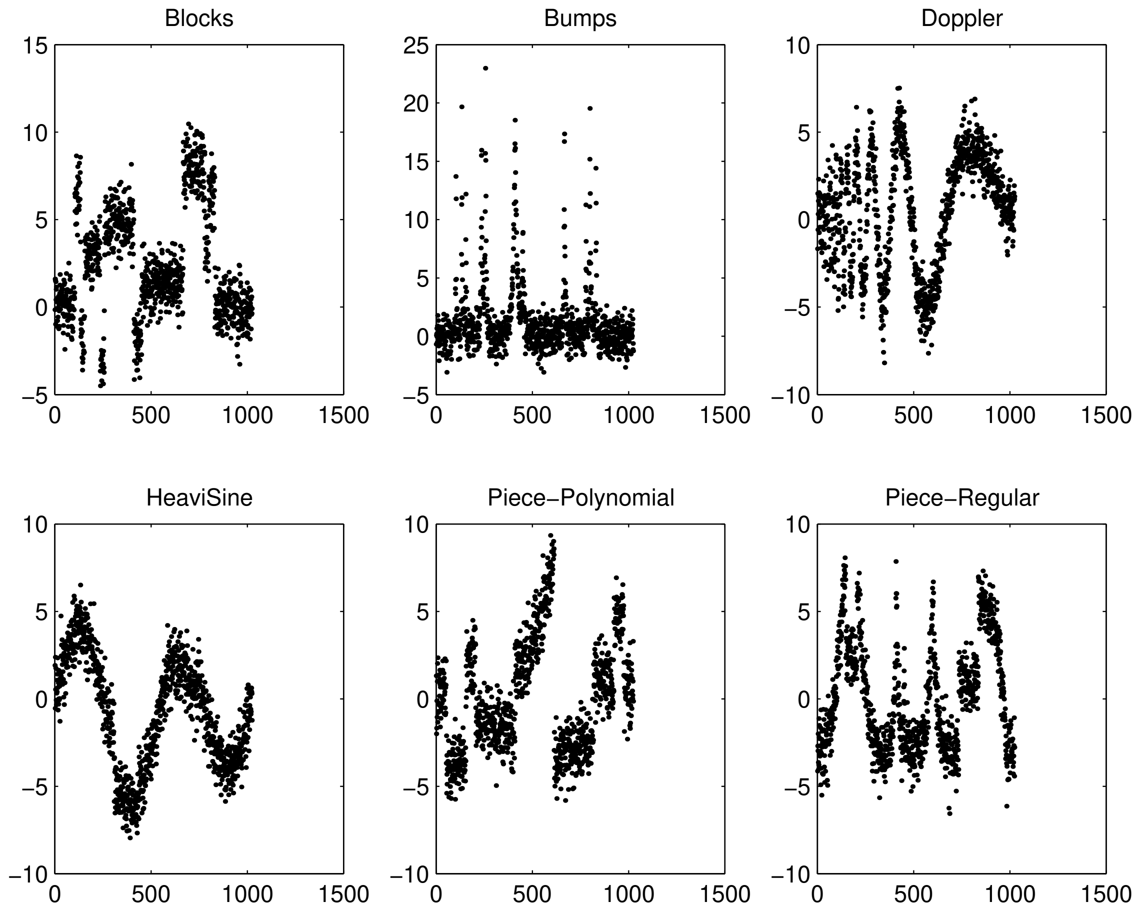


Figure 4. Proposed Estimator Applied to Reconstruct Figure 2.

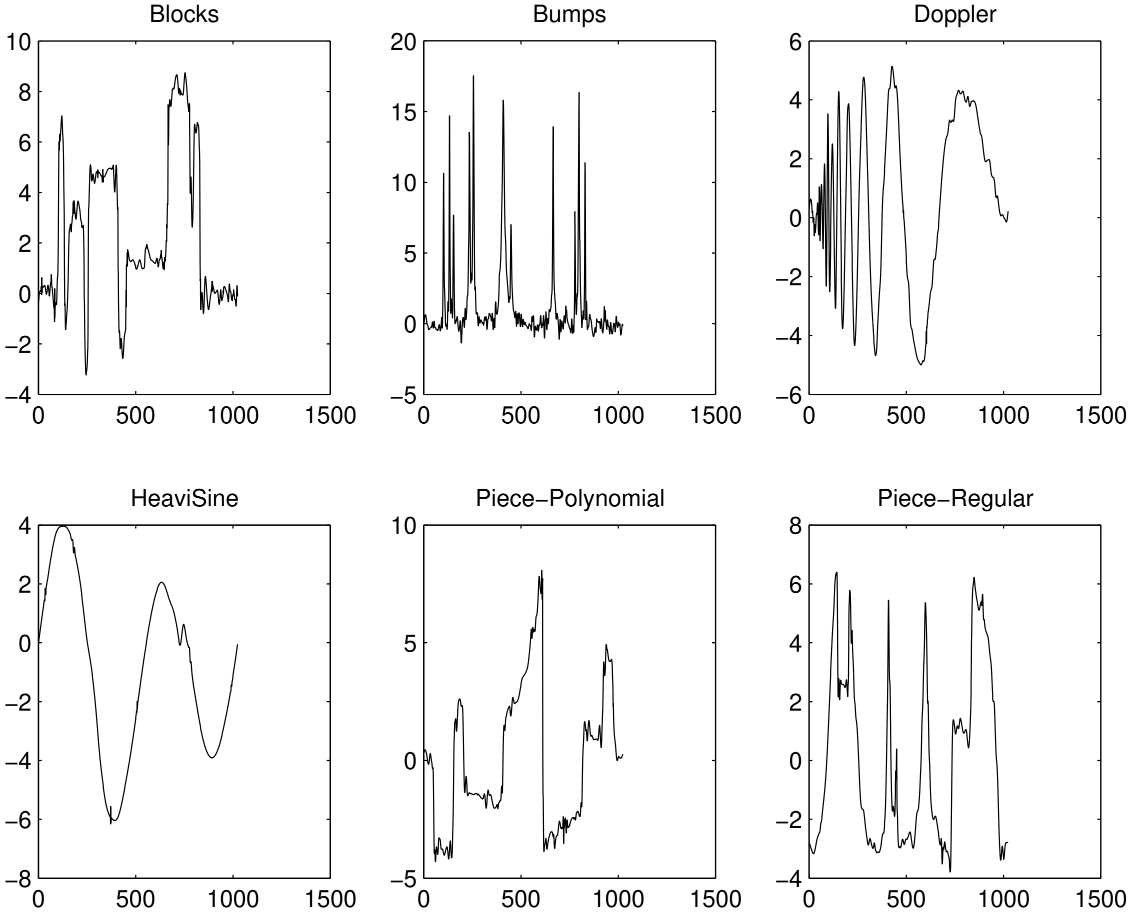
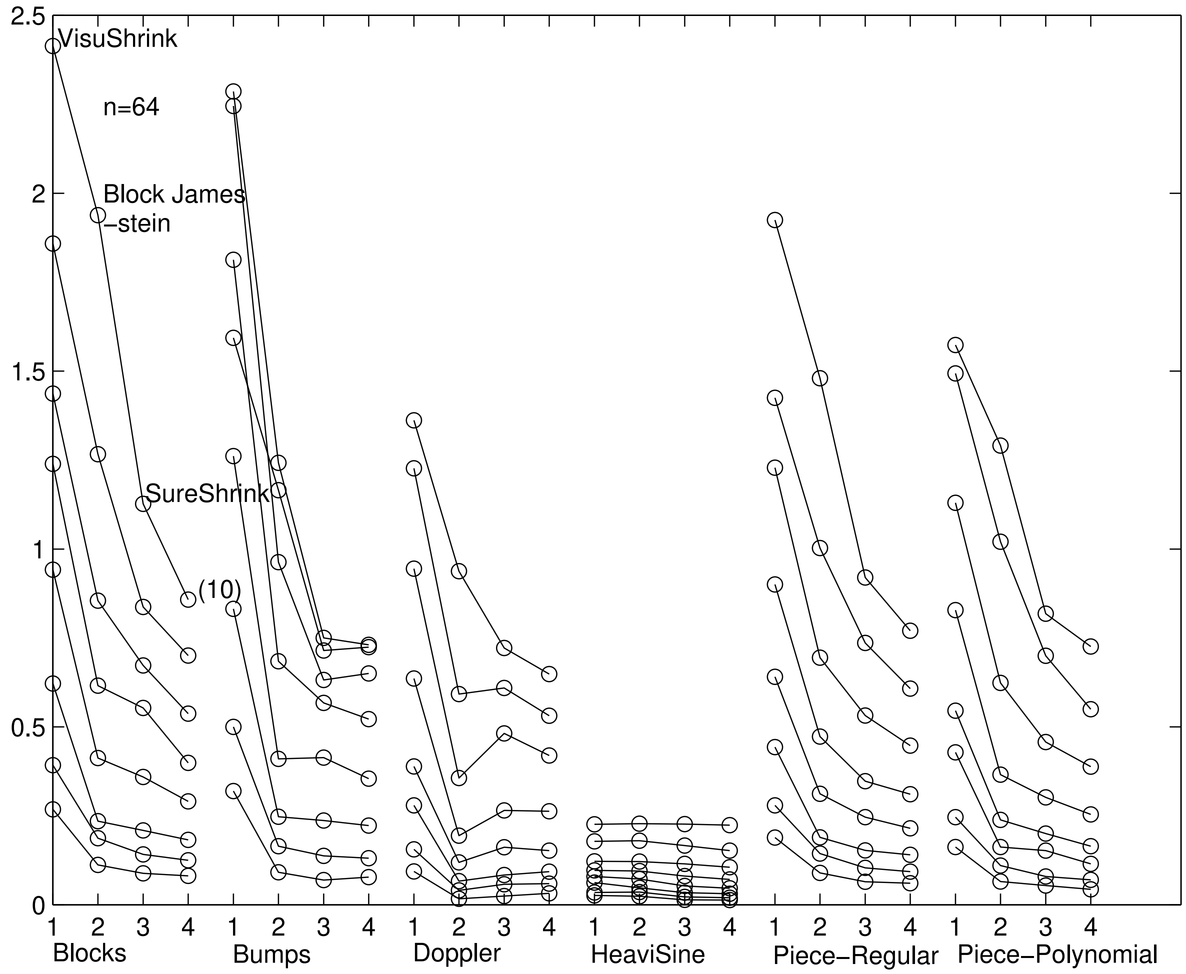


Figure 5: In each of the six cases corresponding to Blocks ,Bumps, etc., the eight curves plot the risk function , from top to bottom, when n=64, 128, ..., 8192. For each curves, the vertices from left to the right give the risks of VisuShrink, SureShrink, Block James-Stein, and the proposed method



VisuShrink
n=64
Block James
-stein
SureShrink
(10)
0
0.5
1
1.5
2
2.5
1 2 3 4
Blocks
Bumps
Doppler
HeaviSine
Piece-Regular
Piece-Polynomial

Figure 6. James-Stein Estimator Applied to Reconstruct Figure 2.

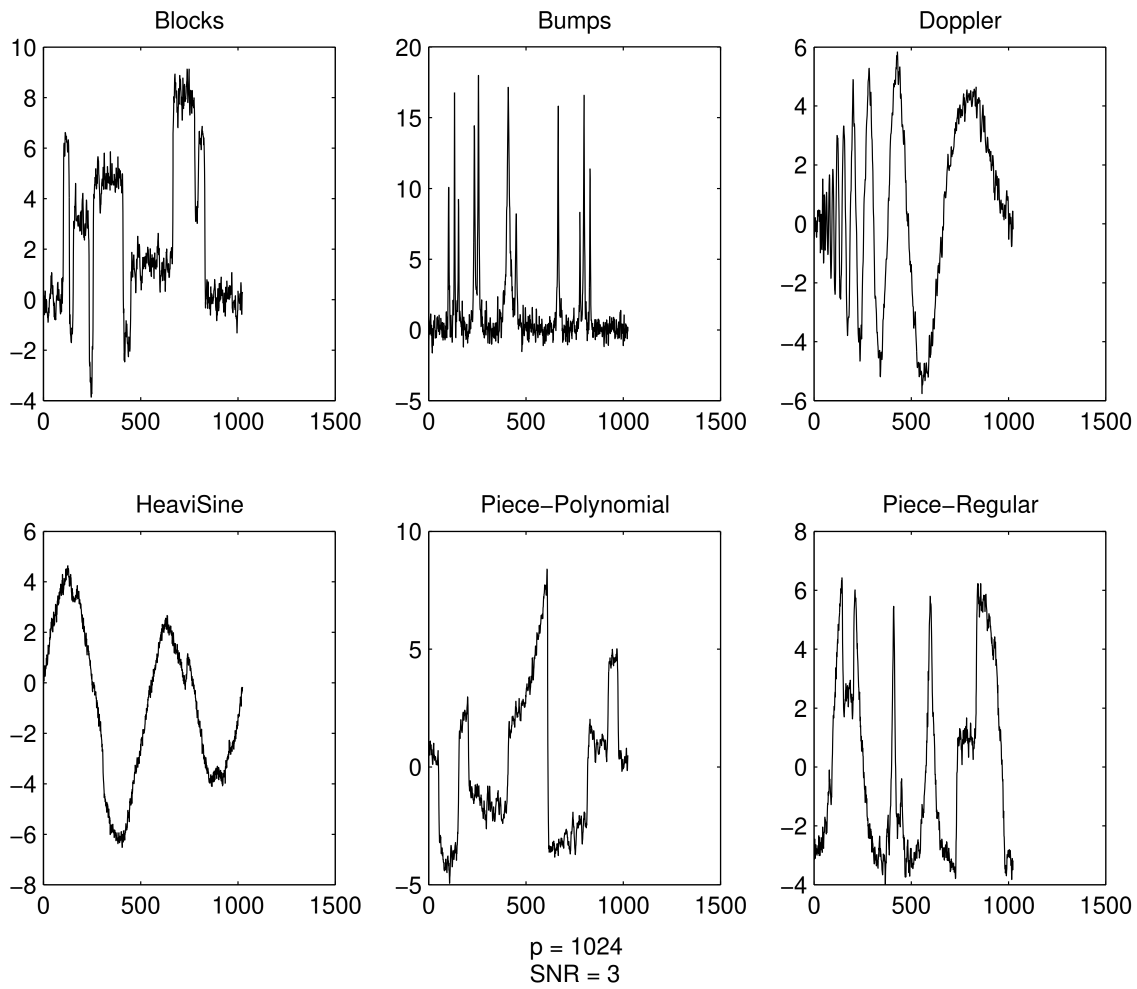
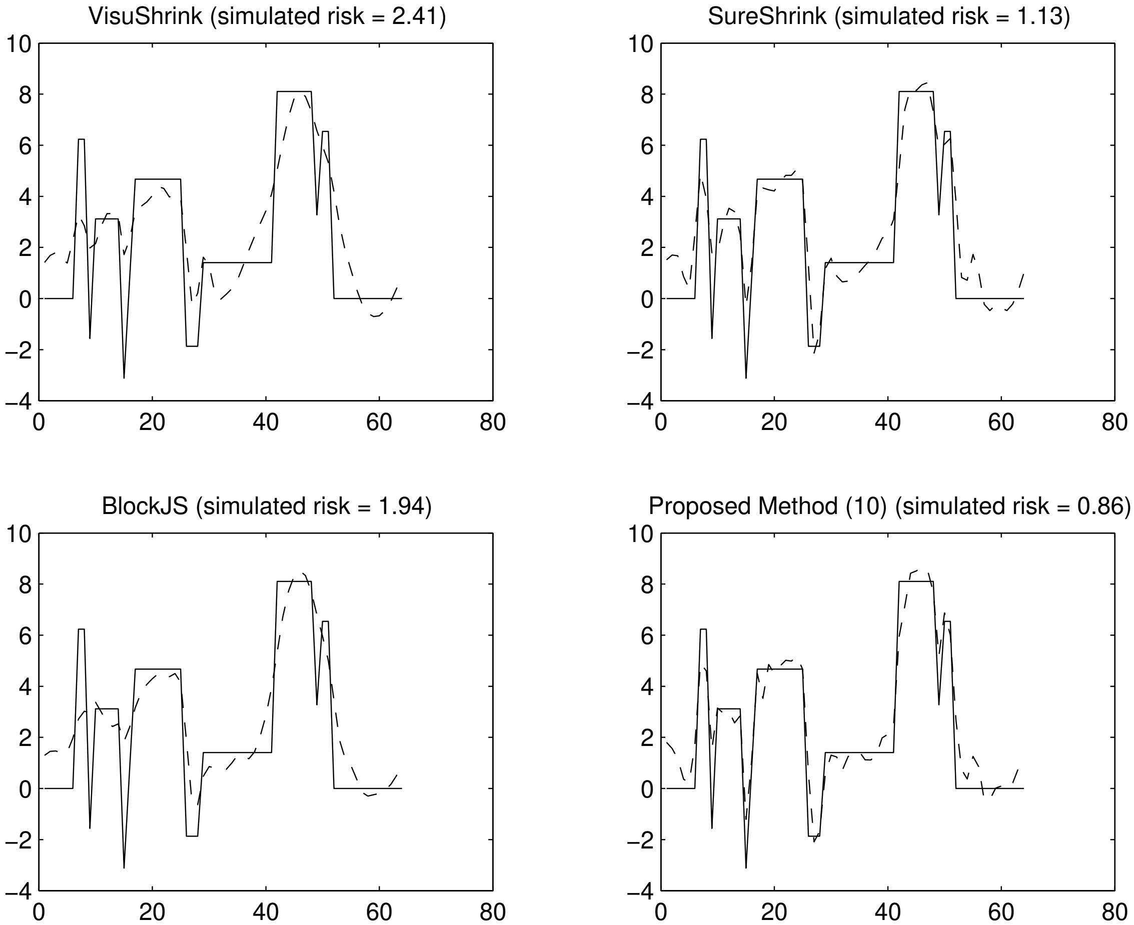


Figure 7. All Estimators Applied to Reconstruct Blocks for n=64.



- Principal component regression

$$Y_{n\times 1} = X_{n\times p}\theta_{p\times 1} + \sigma^2 I_{n\times 1}.$$

We are going to compare the predictive error. Let

$$\begin{aligned} X_{n\times p} &= U_{n\times n}S_{n\times p}V^T_{p\times p} \\ &= \left(\lambda_1 u_1, ..., \lambda_m u_m, 0, ..., 0\right) V^T_{p\times p}. \end{aligned}$$

Here $U_{n\times n} = \left(u_1, ..., u_n\right)$, and λ_i are nonzero eigenvalues, $m \leq n$. The OLS estimaor is

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T_{n\times p} Y.$$

We see

$$E\left\|X\hat{\beta}-X\beta\right\|^2 = E\left\|U_m^T Y - S_m V^T \beta\right\|^2.$$

where

$$U^T = \begin{pmatrix} u_1^T \\ \vdots \\ u_m^T \end{pmatrix}$$

and $Z = U^T Y \sim N\left(\theta, \sigma^2 I_m\right), \theta = S V^T \beta.$

Per	OLS	VisuShrink	SureShrink	New
0.1	0.6626	0.4675	0.4389	0.2995
0.2	0.6644	0.9644	0.5453	0.4389
0.3	0.6684	1.4431	0.6089	0.5124
0.4	0.6676	1.9269	0.6326	0.5613
0.5	0.6668	2.4000	0.6470	0.5890
0.6	0.6628	2.8727	0.6531	0.6068
0.7	0.6631	3.3349	0.6583	0.6212
0.8	0.6669	3.8449	0.6658	0.6390
0.9	0.6609	4.3315	0.6602	0.6434
1.0	0.6691	4.8217	0.6694	0.6600

Related work

- Beran and Dumbgen (1997 Ann Stat)

$$\widehat{\theta}_i = h_i Z_i$$

  where h_i decreases as i getting bigger.

- Antoniadis, Leporini and Pesquet (2000)

- Fan and Antoniadi (2001 JASA). Fan and Li (2001 Ann stat)

- Brieman (1993)

- Ridge regression. $\beta = 2$

- Lasso. $\beta = 1$

Discussion

- Adaptive choice of β. Kou (2001)
- Adaptive choice of a
- Asymtotic property
- Confidence set
- Microarray data analysis
- More applications