Shrinkage Estimation Toward the Data Chosen Reduced Model with Application to Wavelet Regression

Huibin Zhou

Cornell University

Aug 22, 2002

Jointly with J.T. Gene Hwang

Overview

1. **Motivation**

2. **Main Results**

3. **Application in wavelet regression**

4. **Discussion**

Motivation

We have the following Gaussian regressin model:

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

where ξ_i are i.i.d. $N(0, 1)$.

Goal: Estimation of $\theta_1, ..., \theta_n$.

A Naive Approach (doing nothing):

$$\widehat{\theta}_1 = Z_1, ..., \widehat{\theta}_n = Z_n$$

Question: Can we find an estimator which adaptively selects a reduced model and guarantee always doing better than doing nothing?

The answer is Yes.

James -Stein Estimator

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

$$0 < a \le 2\,(n-2)\,,\; n \ge 3$$

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

Main Results

Now we provide a new class of estimators such that the chosen reduced model can be a subspace with an arbitrary dimension between zero and that of a full model, and the new estimators dominate the naive estimator.

Theorem 1. For $1 < \beta \leq 2$,Let

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

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

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

where

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

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

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

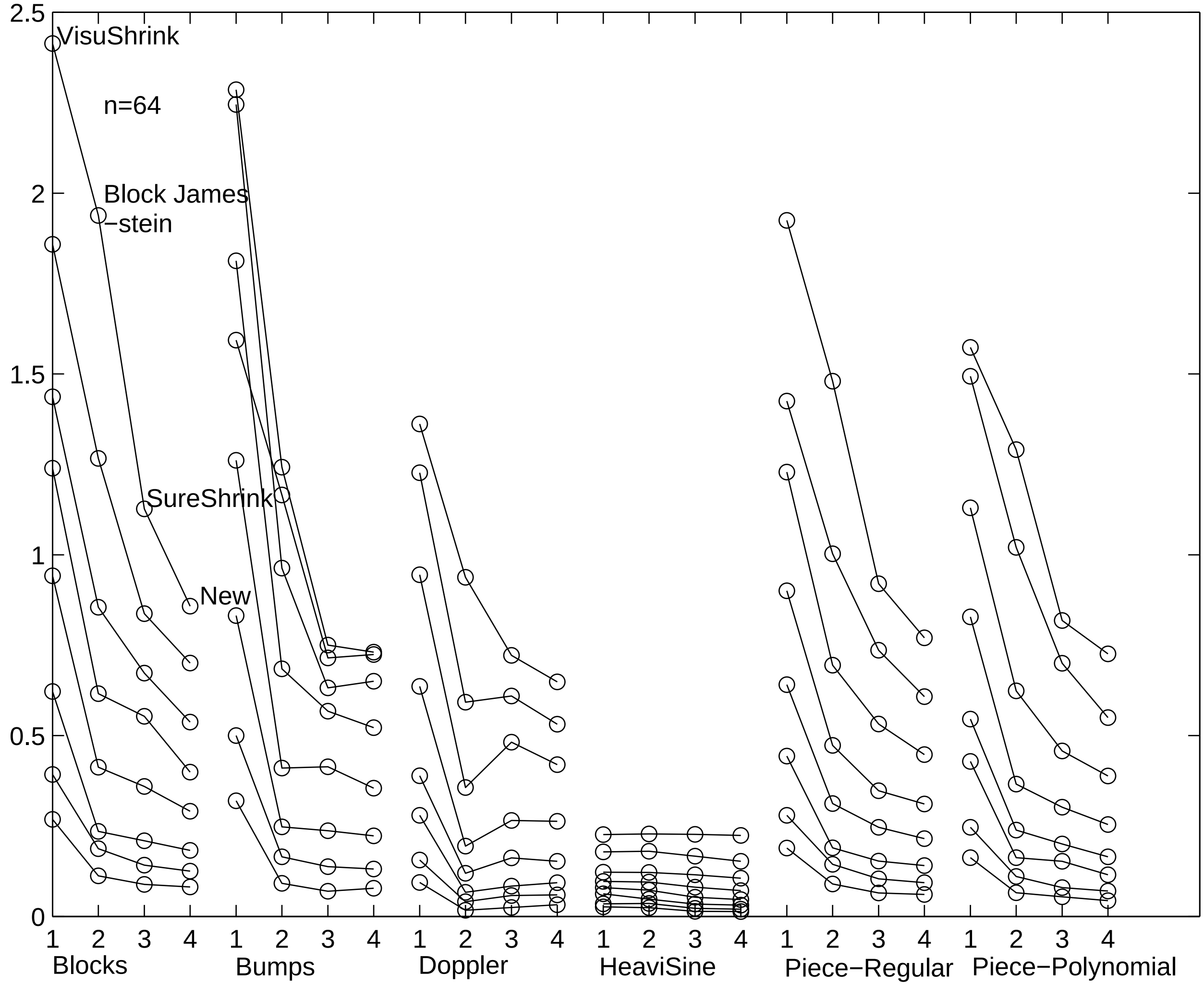
and we know

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

where ξ_i are i.i.d. $N(0,1)$.

Application in wavelet regression

Figure: 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 with $\beta = 4/3$.



VisuShrink
n=64
Block James
−stein
SureShrink
New
2.5
2
1.5
1
0.5
0
1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
Blocks
Bumps
Doppler
HeaviSine
Piece−Regular
Piece−Polynomial

Discussion

- Adaptive choice of β using SURE.

- This result can be applied to PCR (Principal Component Regression).

- we have generalized results for the Gaussian regression model: $Z\tilde{\ }N\left(\theta, V_{n\times n}\right).$ This leads to a minimax variable selection approach.