Wavelet Shrinkage Estimates For Heteroscedastic Regression Models

Hong-Ye Gao
MathSoft, Inc.
1700 Westlake Ave. N, Suite 500
Seattle, WA 98109
email: gao@statsci.com

February 13, 1997

Abstract

We study the following heteroscedastic nonparametric regression model:

$$y_i = f(t_i) + \sigma(t_i) z_i$$

where $\{z_i\}$ is independent identically distributed random noise with $z_i \sim N(0,\ 1)$ and $\sigma^2(t)$ is the variance function. We want to estimate f. We extend Donoho and Johnstone's wavelet shrinkage technique (known as WaveShrink) to this model. We address the issue of variance estimation and propose a procedure based on non-decimated wavelet transform and running MAD (Median Absolute Deviation from the median).

Key Words and Phrases: Heteroscedasticity; Nonparametric Regression; Running MAD; Wavelet Transform; Variance Estimation;

1 Introduction

Suppose we observe data $\mathbf{y} = (y_1, \ldots, y_n)'$ given by

$$y_i = f(t_i) + \varepsilon_i \qquad i = 1, 2, ..., n \tag{1}$$

with $t_i = (i-1)/n$ and $\{\varepsilon_i\}$ drawn from some noisy process. The unknown function f is potentially complex, spatially inhomogeneous. Our goal is to estimate $f = (f_1, \ldots, f_n)'$ with small mean-square-error (L_2 risk):

$$R(\hat{f},\ f) = \frac{1}{n} \sum_{i=1}^{n} E(\hat{f}_i - f_i)^2. \tag{2}$$

In the case where $\{\varepsilon_i\}$ is i.i.d. Gaussian noise, Donoho and Johnstone (1994) showed that an estimator based on shrinking wavelet coefficients has very desirable minimax properties, for example, WaveShrink achieves, within a factor of $\log n$, the optimal minimax risk over each functional class in a variety of smoothness classes and with respect to a variety of losses, including L_2 risk. Johnstone and Silverman (1994) extended Donoho and Johnstone's WaveShrink to deal with stationary $\{\varepsilon_i\}$.

In this paper, we extend Donoho and Johnstone's results to deal with non-constant variance case. Let $\varepsilon_i = \sigma(t_i)z_i$ where $\{z_i\}$ are independent Gaussian random variables with mean zero, variance one and $\sigma^2(t)$ is the variance curve.

Heteroscedasticity also arises from converting unequally spaced data to equally spaced data. Suppose we observe data $(y_1, y_2, \ldots, y_n)'$ given by (1) with $t_1 \leq \cdots \leq t_n$ unequally spaced and $\{z_i\}$ i.i.d. $N(0,\ \sigma^2)$. Procedures based on various spatial adaptive wavelets and interpolation schemes have been proposed, see Sardy et al. (1995); Sweldens (1995). The interpolated data usually have heteroscedastic noise structure.

The paper is organized as follows. In section 2, we study the case that the variances are known. We derive the covariance structure of the wavelet coefficients and show that near-optimality property still holds. In section 3, we address the issue of variance estimation. We propose a procedure based on non-decimated wavelet transform and running MAD (Median Absolute Deviation from the median). Numerical examples are used to demonstrate our procedure. Section 4 gives a brief summary. Proofs and technical details are given in section 5.

2 Wavelet Shrinkage Estimates

2.1 WaveShrink Procedure

WaveShrink is based on the principle of shrinking wavelet coefficients towards zero to remove noise. Let $w = (w_1, \ldots, w_n)'$ be the empirical wavelet coefficients. The wavelet coefficients are computed by the linear transform $w = H\mathbf{y}$, where

$$H = \begin{pmatrix} h_{11} & \cdots & h_{1n} \\ & \cdots & \\ h_{n1} & \cdots & h_{nn} \end{pmatrix}. \tag{3}$$

We refer to H as the wavelet transform matrix. The elements of the matrix h_{jk} have a very special structure corresponding to a sequence of linear filtering operations. In practice, we do not form the matrix H, but instead use the very fast (order n) pyramid algorithm, Mallat (1989), to compute the wavelet and inverse wavelet transforms.

The WaveShrink estimate is obtained by the following three steps:

[1] Compute wavelet transform $\mathbf{w} = H\mathbf{y}$.

[2] Apply a non-linear shrinkage rule $\delta_\lambda(\cdot)$ to the wavelet coefficients to obtain

$$\hat{w}_k = \delta_{\lambda\hat{\sigma}_k}(w_k). \tag{4}$$

where $\hat{\sigma}_k$ is an estimate of the standard deviation σ_k of wavelet coefficient w_k.

[3] Invert the wavelet transform $\hat{\mathbf{f}}_\lambda = H^{-1}\hat{\mathbf{w}}$.

The wavelet coefficients correspond to J different resolution levels. Typically, a double indexing scheme (j,k) is used where j is the resolution level and k is the shift parameter. The resolution levels range from the finest scale detail ($j = J$) to coarsest scale detail ($j = 1$). At the coarsest resolution level, we also have the smooth coefficients. The results in this paper do not depend on resolution level. Hence, to simplify the notation, we typically use a single index i for the wavelet coefficients, where i ranges from 1 to n.

For i.i.d. Gaussian noise, the empirical wavelet coefficients $\mathbf{w}$ are also independent Gaussian random variables with the same variance. For heteroscedastic model, $\mathbf{w}$ are not only dependent but have different variances.

The choice of shrinkage functions are the *hard*, the *soft* (Donoho and Johnstone, 1994), the *firm* (Gao and Bruce, 1996) and the garrote (Breiman, 1995; Gao, 1996) shrinkage functions:

$$\begin{aligned}
\delta_\lambda^H(x) &= \begin{cases} 0 & \text{if } |x| \le \lambda \\ x & \text{if } |x| > \lambda \end{cases} \\
\delta_\lambda^S(x) &= \begin{cases} 0 & \text{if } |x| \le \lambda \\ \operatorname{sgn}(x)(|x| - \lambda) & \text{if } |x| > \lambda \end{cases} \\
\delta_\lambda^G(x) &= \begin{cases} 0 & |x| \le \lambda \\ x - \lambda^2/x & |x| > \lambda \end{cases} \\
\delta_{\lambda_1,\lambda_2}^F(x) &= \begin{cases} 0 & \text{if } |x| \le \lambda_1 \\ \operatorname{sgn}(x)\frac{\lambda_2(|x|-\lambda_1)}{\lambda_2-\lambda_1} & \text{if } \lambda_1 < |x| \le \lambda_2 \\ x & \text{if } |x| > \lambda_2 \end{cases}
\end{aligned}$$

where $0 < \lambda, \lambda_1 < \lambda_2 \in [0, \infty)$. The pros and cons of these shrinkage functions have been studied by Donoho and Johnstone (1994); Donoho and Johnstone (1997); Bruce and Gao (1996); Marron et al. (1996); Gao (1996). Figure 1 displays the four shrinkage functions.

Remark: All the shrinkage functions have the property:

$$\delta_{\lambda\hat{\sigma}}(X) = \hat{\sigma}\delta_\lambda\left(\frac{X}{\hat{\sigma}}\right) = \sigma\delta_{\lambda\hat{\sigma}/\sigma}\left(\frac{X}{\sigma}\right).$$

2.2 Global Risk of WaveShrink

Let $X \sim N(\theta,\ 1)$, for shrinkage function $\delta(\cdot)$, define the L_2 risk function of the shrinkage estimator of θ

$$R_\lambda(\theta) = E\left\{\delta_\lambda(X) - \theta\right\}^2.$$

Computational formulas of $R_\lambda(\theta)$ for shrinkage function given in section 2.1 can be found in Donoho and Johnstone (1994); Marron et al. (1996); Bruce and Gao (1996); Gao (1996); Gao and Bruce (1996).

We now turn our attention to the global risk of WaveShrink. Let

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{pmatrix}$$

and $\Theta = H\mathbf{f}$, where $\{\sigma_i\}$ are the standard deviations of the noise in model (1) and H is given in (3). Then

$$\mathbf{w} = H\mathbf{y} \sim N(\Theta,\ HD(HD)') = N(\Theta,\ HD^2H').$$

Remark: Matrix multiplication HD can be performed efficiently by taking advantage of the sparsity of D.

Let $(\eta_1^2, \ldots, \eta_n^2)$ be the diagonal elements of HD^2H', then $w_i \sim N(\theta_i,\ \eta_i^2)$ and

$$\eta_i^2 = \sum_{\ell=1}^{n} \sigma_\ell^2 h_{i\ell}^2. \tag{5}$$

When H is orthonormal, $c_1 \le \sigma_i \le c_2$ implies $c_1 \le \eta_i \le c_2$ and then the global risk defined in (2) is

$$\begin{aligned} R(\hat{f},\ f) &= \frac{1}{n}\sum_{i=1}^{n} E\left(\delta_{\lambda_{\eta_i}}(w_i) - \theta_i\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} \eta_i^2 R_\lambda\left(\frac{\theta_i}{\eta_i}\right). \end{aligned} \tag{6}$$

Consider Besov Body (Donoho and Johnstone, 1997):

$$\Theta_{p,q}^s(C) = \left\{(\theta_{jk}) : \sum_{j\ge 0}\left(2^{js}\left(\sum_k |\theta_{jk}|^p\right)^{1/p}\right)^q \le C^q\right\}. \tag{7}$$

From the definition of Besov body, we can easily see that

Lemma 1 *Let $\theta'_{jk} = \theta_{jk}/\eta_{jk}$. If $\eta_{jk} \geq C_0 > 0$ for all j and k, then*

$$(\theta_{jk}) \in \Theta^s_{p,q}(C) \quad \textit{implies} \quad \left(\theta'_{jk}\right) \in \Theta^s_{p,q}\left(C/C_0\right).$$

On the other hand, if $0 < \eta_{jk} \leq C_1$ for all j and k, then

$$\left(\theta'_{jk}\right) \in \Theta^s_{p,q}(C) \quad \textit{implies} \quad (\theta_{jk}) \in \Theta^s_{p,q}\left(C_1 C\right).$$

Combining Lemma 1 and (6), we have

Theorem 1 *Under model (1) and orthogonal wavelet transform, if $\{\sigma_i\}$ are finite and bounded from 0, i.e.*

$$0 < C_0 \leq A_n \equiv \min_{1 \leq i \leq n} \{\sigma_i\} \leq \max_{1 \leq i \leq n} \{\sigma_i\} \equiv B_n < \infty.$$

Then from (6),

$$\frac{A_n^2}{n} \sum_{i=1}^{n} R_\lambda\left(\theta'_i\right) \leq R(\hat{f},\ f) \leq \frac{B_n^2}{n} \sum_{i=1}^{n} R_\lambda\left(\theta'_i\right) \tag{8}$$

and hence when $B_n \leq C_1$,

$$\sup_{f \in B^s_{p,q}} R(\hat{f},\ f) = O\left(\frac{\log n}{n^{2s/(2s+1)}}\right). \tag{9}$$

The uniform convergence rate in (9) is the same as in the white noise model, see Donoho and Johnstone (1997); Donoho and Johnstone (1994).

Remark: When the noise variances (σ_i) are known, one may consider dividing both side of model (1) by σ_i

$$\frac{y_i}{\sigma_i} = \frac{f_i}{\sigma_i} + \frac{z_i}{\sigma_i}. \tag{10}$$

Then we have a standard Gaussian white noise model. This may not be a good idea for wavelet shrinkage denoising technique, because at the same time, we also change the characteristics of the underlying signal, i.e. instead of estimating $f(t)$ we are dealing with $f(t)/\sigma(t)$. The wavelet that matches the characteristics of f may not match the characteristics of rescaled f.

2.3 Numerical Examples

Let us look at Donoho and Johnstone's four signals: "blocks", "bumps", "heavisine" and "doppler". Independent Gaussian noise with the following standard deviations are added to each of the signals:

Example 1:

$$\sigma(t)=\begin{cases} v+a(t-t_1)^2 & 0<t\le t_1 \\ v & t_1<t<t_2 \\ v+b(t-t_2)^2 & t_2<t\le 1 \end{cases}.$$

Example 2:

$$\sigma(t)=v+\left\{a-b(t-c)^2\right\}^2, \qquad 0\le t\le 1.$$

Figure 2 (a) displays the Example 1 for $v=0.001, a=1, b=2$, $t_1=0.3$ and $t_2=0.7$; and (b) displays the Example 2 for $v=0.001, a=0.4, b=2.5$ and $c=0.515$. Example 2 with different parameters was also studied by Antoniadis and Lavergne (1995).

In our numerical study, we define the signal-to-noise ratio (SNR) by

$$\text{SNR}=\frac{\text{SD(signal)}}{\int\sigma(t)dt}$$

and the denominator is computed by averaging $\sigma(t)$ over discrete t. All computations were carried out using S+Wavelets version1.1 on SUNOS5.

The sample size is 512 and the minimax thresholds ($\lambda=3.312$ for *hard* and 2.441 for *garrote*) are used. Minimax thresholds for other n can be found in Bruce and Gao (1996); Gao (1996). In those examples, we use Haar wavelet for the "blocks" example, Daubechies' least asymmetric wavelet with 8 filter coefficients (*s8*) for the other three signals. We apply the shrinkage functions to 5 finest level wavelet coefficients (totally 496 coefficients). The smoothed signals are displayed in Figures 4 and 5.

3 WaveShrink with Unknown Variances

Variance estimation in regression model has been studied by many authors. Difference-based estimation for σ^2 in the homoscedastic case has fruitful results, for example, Hall and Marron (1990) studied the mth-order differences; Hall et al. (1990) studied the asymptotically optimal mth-order differences. For heteroscedastic case, Hall and Carroll (1989); Müller and Stadtmüller (1987); Antoniadis and Lavergne (1995) used difference-based estimates as initial estimates for the variance function $\sigma^2(t)$. High-pass wavelet filter is a differencing operator, therefore the squared detail wavelet coefficients provide natural initial variance estimates.

The difference between our model and the models discussed by Hall and Carroll (1989); Müller and Stadtmüller (1987); Antoniadis and Lavergne (1995) is that we assume f is from a certain Besov space (f is potentially complex, spatially inhomogeneous) rather than a Lipschitz class of smooth functions. For functions in Besov space, under appropriate wavelet transform, the detail wavelet coefficients (high-pass coefficients) are very sparse. Figure 6 displays the detail wavelet coefficients for signal/wavelet combinations: blocks/Haar, blocks/s8, bumps/s8, heavisine/s8 and doppler/s8.

Since the detail wavelet coefficients of the noisy data contain mostly noise, we can estimate the variances from the detail wavelet coefficients after removing the (sparse) signal components. Our procedure consists of the following steps:

[1] Apply a non-decimated wavelet transform to $\mathbf{y}$ to get the finest level detail coefficients $\mathbf{d}_J$.

[2] Apply a running MAD (Median Absolute Deviation from the median, defined by $1.4826 \cdot \mathrm{med}(|x - \mathrm{med}(x)|)$) to get the estimated standard deviation, see Hampel et al. (1986).

We apply above procedure to the previous examples. Figure 7 contains the estimated standard deviations and Figures 8 and 9 display the WaveShrink estimates using the estimated standard deviations. In these examples, we use Haar wavelet for the "blocks" example, Daubechies' least asymmetric wavelet with 8 taps (*s8*) for the other signals; the estimated standard deviations are obtained by running MAD of span 51 (roughly 10% of the sample size). The choices of wavelet and span are all empirical.

Remark 1: The non-decimated discrete wavelet transform, also known as the stationary wavelet transform or translation invariant wavelet transform, see, Mallat and Zhong (1992); Coifman and Donoho (1995), is a non-orthogonal variant of the classical DWT. In the non-decimated DWT, at each resolution level, starting with n coefficients one ends up with n coefficients, while the classical DWT has fewer coefficients at coarser scales. Unlike the orthogonal DWT, the non-decimated DWT is *translation invariant*. In other words, shifts in the signal do not change the pattern of coefficients for the non-decimated DWT.

Remark 2: For the non-decimated wavelet transform, it is easy to see that for large j, $\mathrm{Var}(z_{jk}) \approx \sigma_k^2$.

Remark 3: With estimated variances, the standized wavelet coefficients are no longer Gaussian distributed (roughly t-distributed, with heavier tails than Gaussian). Therefore, larger thresholds are expected to suppress the noise. In our examples, we use $\lambda = 2\sqrt{\log n} = 4.995$.

Theorem 2 *Under the conditions of Theorem 1, let $\hat{\sigma}(t)$ be an L_2 consistent estimate of $\sigma(t)$ and $\tilde{f}$ be the wavelet shrinkage estimator using $\hat{\sigma}(t)$, then the same near-optimal convergence rate holds for $R(\tilde{f},\ f)$.*

4 Conclusions and Discussions

In this paper, we extend Donoho and Johnstone's wavelet shrinkage smoothing technique (known as WaveShrink) to handle data with heteroscedastic noise. We derive formulas for computing the variances of the wavelet coefficients and show that the same near-optimal convergence rate as in the case of white noise can be achieved if the noise variances are known. We proposed a noise variance estimation procedure, without assuming any smoothness of the underlying function. Our procedure is based on applying running MAD (Median Absolute Deviation from the median) to the non-decimated finest level wavelet coefficients. We apply our technique to various numerical examples.

5 Technical Details

5.1 Proof of Theorem 1

We first prove the Lemma 1. If $\eta_{jk} \geq C_0$ for all j and k, then

$$\left|\theta'_{jk}\right| \leq |\theta_{jk}| / C_0$$

for all j and k. The first part of Lemma 1 follows immediately from the definition of Besov space. On the other hand, if $\eta_{jk} \leq C_1$ for all j and k, then

$$|\theta_{jk}| \leq C_1 \left|\theta'_{jk}\right|$$

for all j and k and this completes the proof of Lemma 1.

From (5), $\{\eta_i^2\}$'s are also bounded by A_n and B_n, therefore conditions in Lemma 1 hold. The inequalities (8) in Theorem 1 follows immediately from (5) and (6) and (9) follows from Lemma 1 and Donoho and Johnstone's minimax convergence rate over Besov space for white noise models, see Donoho and Johnstone (1997); Donoho and Johnstone (1994).

5.2 Proof of Theorem 2

Using the conditional expectation argument,

$$\begin{aligned} R(\tilde{f}_\lambda,\ f) &= \frac{1}{n}\sum_{i=1}^{n} E\left\{\delta_{\lambda\hat{\eta}_i}(w_i) - \theta_i\right\}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} \eta_i^2\ E\left\{\delta_{\lambda\hat{\eta}_i/\eta_i}\left(\frac{w_i}{\eta_i}\right) - \frac{\theta_i}{\eta_i}\right\}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} \eta_i^2\ E\left\{E\left(\delta_{\lambda\hat{\eta}_i/\eta_i}\left(\frac{w_i}{\eta_i}\right) - \frac{\theta_i}{\eta_i}\right)^2 \middle|\, \hat{\eta}_i\right\} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\eta_i^2 E\left\{R_{\lambda\hat{\eta}_i/\eta_i}\left(\frac{\theta_i}{\eta_i}\right)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\eta_i^2 R_\lambda\left(\frac{\theta_i}{\eta_i}\right) + \frac{1}{n}\sum_{i=1}^{n}\eta_i^2\left\{ER_{\lambda\hat{\eta}_i/\eta_i}\left(\frac{\theta_i}{\eta_i}\right) - R_\lambda\left(\frac{\theta_i}{\eta_i}\right)\right\}.
\end{aligned}
$$

Theorem 2 follows from the facts that $\hat{\eta} \to \eta$ in L_2 as $n \to \infty$, and the L_2 risk function $R_\lambda(\theta)$ is continuous in λ.

6 Acknowledgements

This work was inspired by a conversation with Dr. Daryl Pregibon of AT&T Laboratories. The author would like to thank Professor Jianqing Fan and Dr. Yihui Zhan for helpful discussions.

References

Antoniadis, A. and Lavergne, C. (1995). Variance Function Estimation in Regression by Wavelet Methods. In Antoniadis, A. and Oppenheim, G., editors, *Wavelets and Statistics*, pages 31–42. Springer-Verlag, New York. Lecture Notes in Statistics 103.

Breiman, L. (1995). Better Subset Regression Using the Nonnegative Garrote. *Technometrics*, 37(4):373–384.

Bruce, A. G. and Gao, H.-Y. (1996). Understanding WaveShrink: Variance and Bias Estimation. *Biometrika*, 83(4).

Coifman, R. R. and Donoho, D. L. (1995). Translation-Invariant De-Noising. In Antoniadis, A. and Oppenheim, G., editors, *Wavelets and Statistics*, pages 125–150. Springer-Verlag, New York. Lecture Notes in Statistics 103.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika*, 81:425–455.

Donoho, D. L. and Johnstone, I. M. (1997). Minimax Estimation via Wavelet Shrinkage. *The Annals of Statistics*. To appear.

Gao, H.-Y. (1996). Wavelet Shrinkage Denoising Using the Non-Negative Garrote. preprint.

Gao, H.-Y. and Bruce, A. G. (1996). WaveShrink with Firm Shrinkage. *Statistica Sinica*. To appear.

Hall, P. and Carroll, R. J. (1989). Variance Function Estimation in Regression: the Effect of Estimating the Mean. *Journal of the Royal Statistical Society, Series B*, 51:3–14.

Hall, P., Kay, J. W., and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77:521–528.

Hall, P. and Marron, J. S. (1990). On Variance Estimation in Nonparametric Regression. *Biometrika*, 77:415–419.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M., and Stahel, W. A. (1986). *Robust Statistics: the approach based on influence functions*. John Wiley, New York.

Johnstone, I. and Silverman, B. (1994). Wavelet Threshold Estimators for Data with Correlated Noise. Technical report, Stanford University.

Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Mallat, S. and Zhong, S. (1992). Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):710–732.

Marron, S. J., Adak, S., Johnstone, I. M., Neumann, M. H., and Patil, P. (1996). Exact Risk Analysis of Wavelet Regression. *Journal of Computational and Graphical Statistics*. To appear.

Müller, H. G. and Stadtmüller, U. (1987). Estimation of Heteroscedasticity in Regression Analysis. *The Annals of Statistics*, 15:610–625.

Sardy, S., Percial, D. B., Bruce, A. G., Gao, H.-Y., and Stuetzle, W. (1995). Wavelet De-Noising for Unequally Spaced Data. Technical report, StatSci Division of MathSoft, Inc., 1700 Westlake Ave. N, Seattle, WA 98109-9891.

Sweldens, W. (1995). The Lifting Scheme: A new philosophy in biorthogonal wavelet constructions. In *SPIE Proceedings, Signal and Image Processing III*, volume 2569, pages 68–79, San Diego, CA.

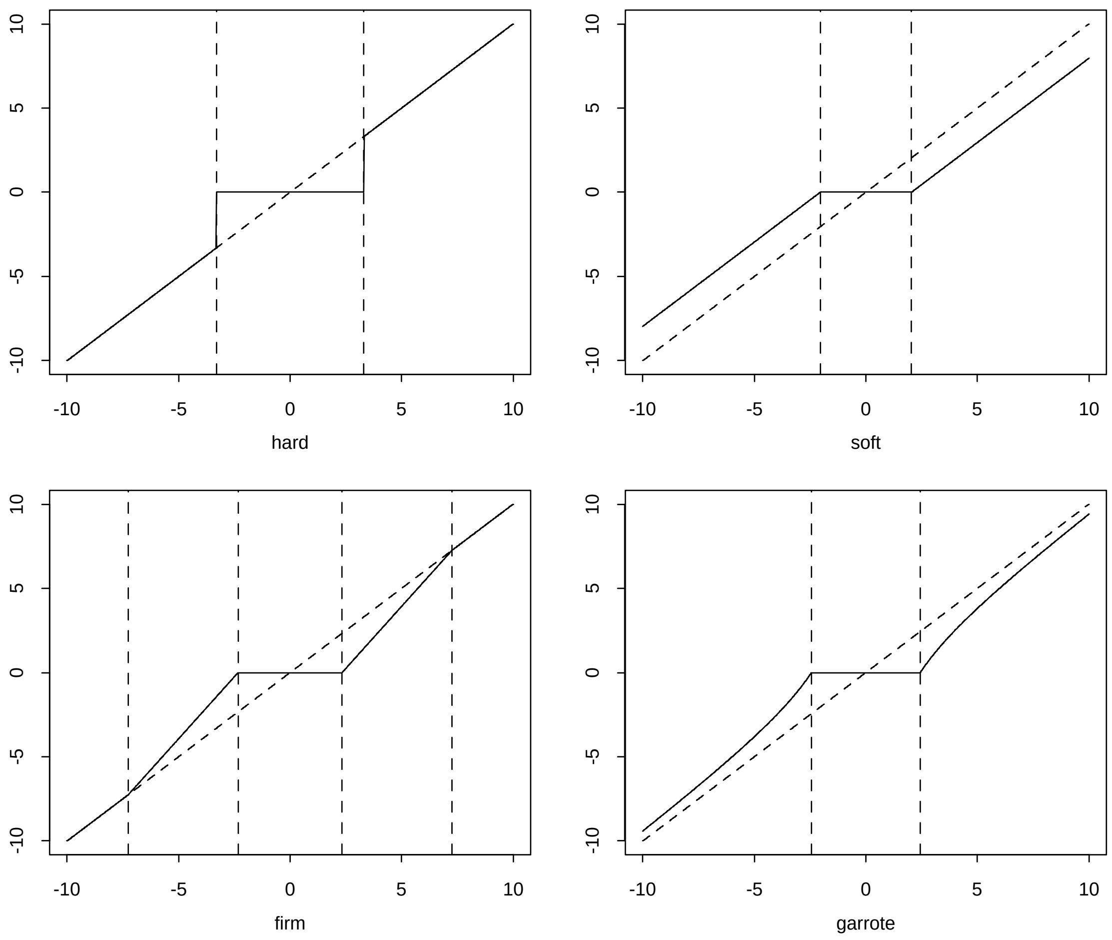

Figure 1: The hard (top left), soft (top right), semisoft (bottom left) and garrote (bottom right) shrinkage functions (solid lines). Hard and semisoft shrinkage functions agree with the identity function (dashed lines) for large enough $|x|$. Garrote shrinkage approaches the identity function for large $|x|$. All but the hard shrinkage functions are continuous.

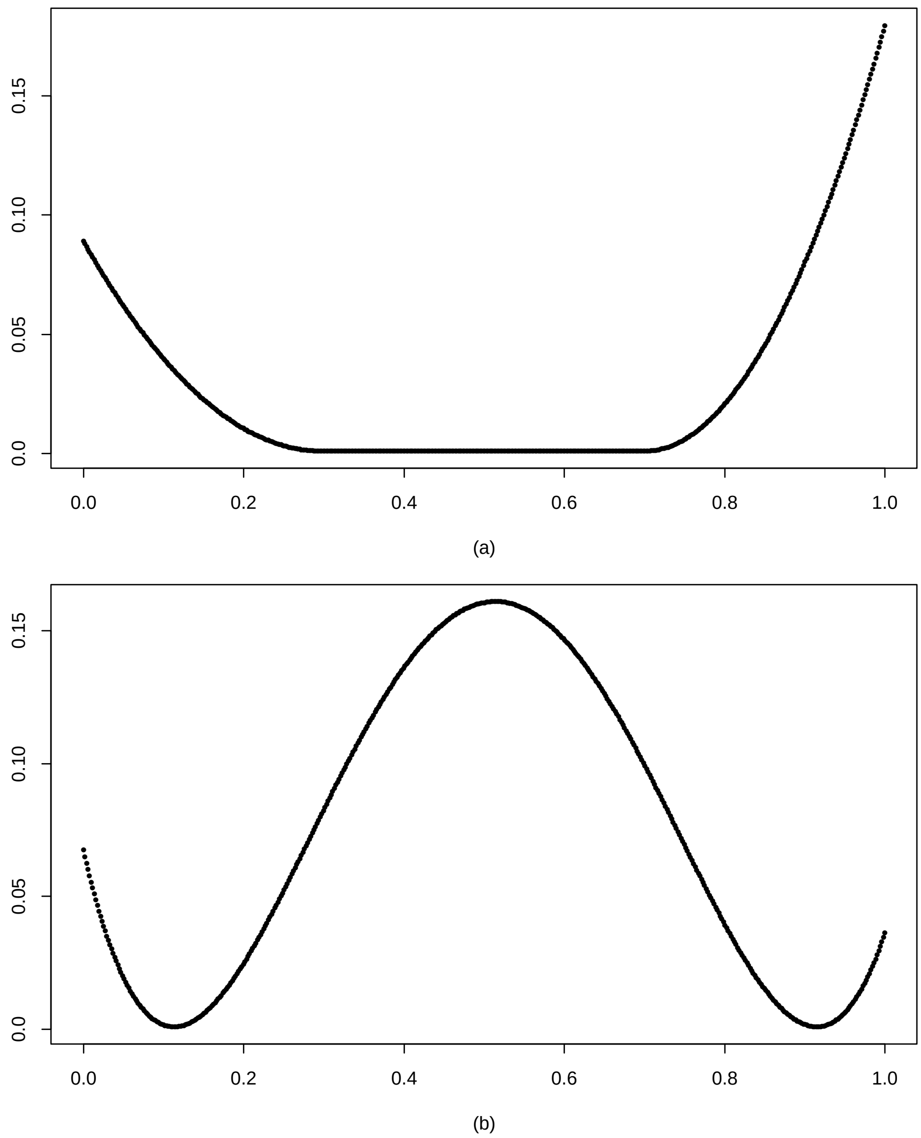

Figure 2: Standard deviations used for generating random noise. For example 1, $v = 0.001$, $t_1 = 0.3$, $t_2 = 0.7$, $a = 1$ and $b = 2$; for example 2, $v = 0.001$, $a = 0.4$, $b = 2.5$ and $c = 0.515$;

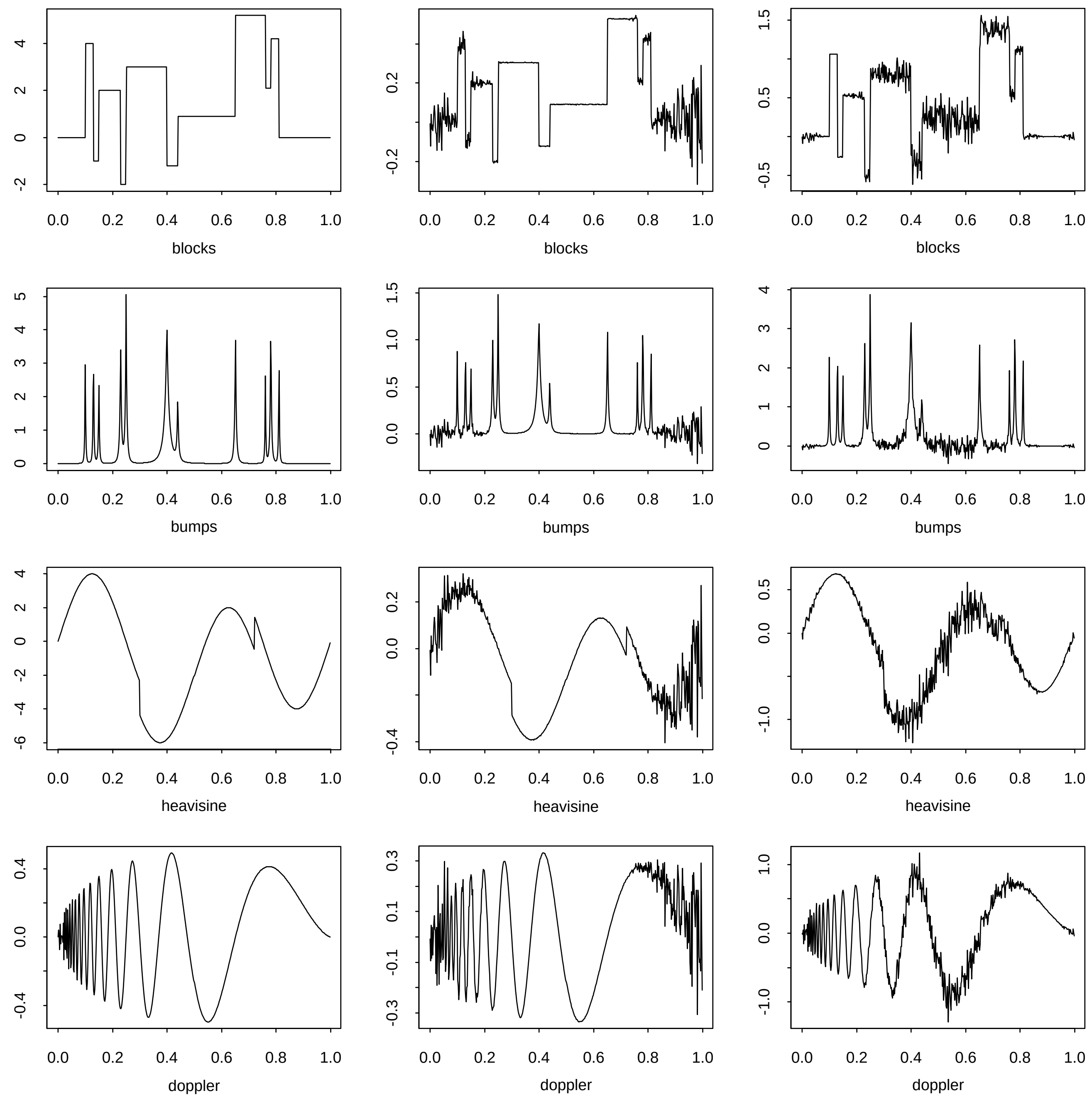

Figure 3: Clean and noisy signals with SNR = 7.

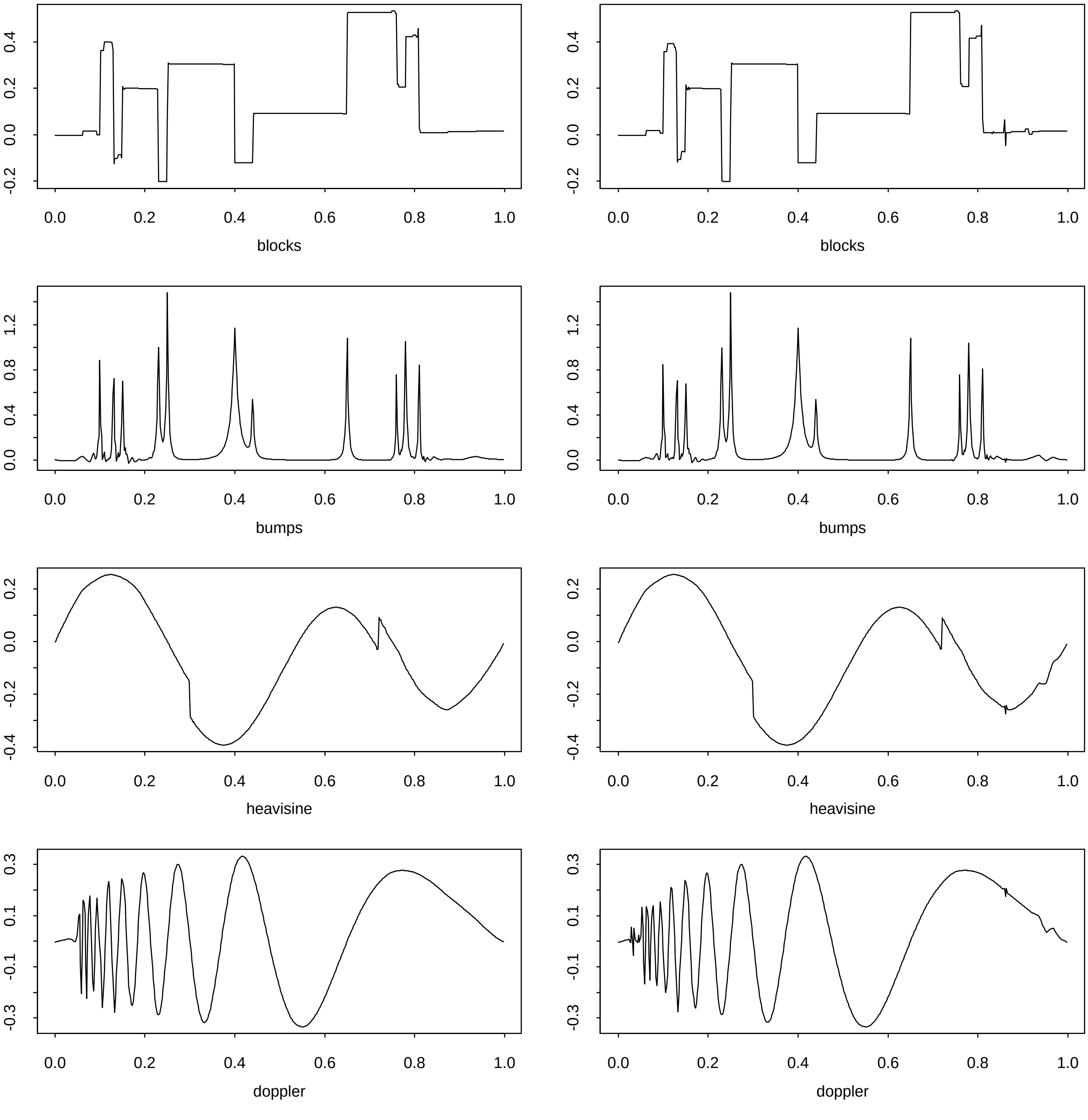

Figure 4: WaveShrink estimates with known variances (example 1): left column are by hard shrinkage with minimax threshold $\lambda = 3.312$; right column are by garrote with minimax threshold $\lambda = 2.441$.

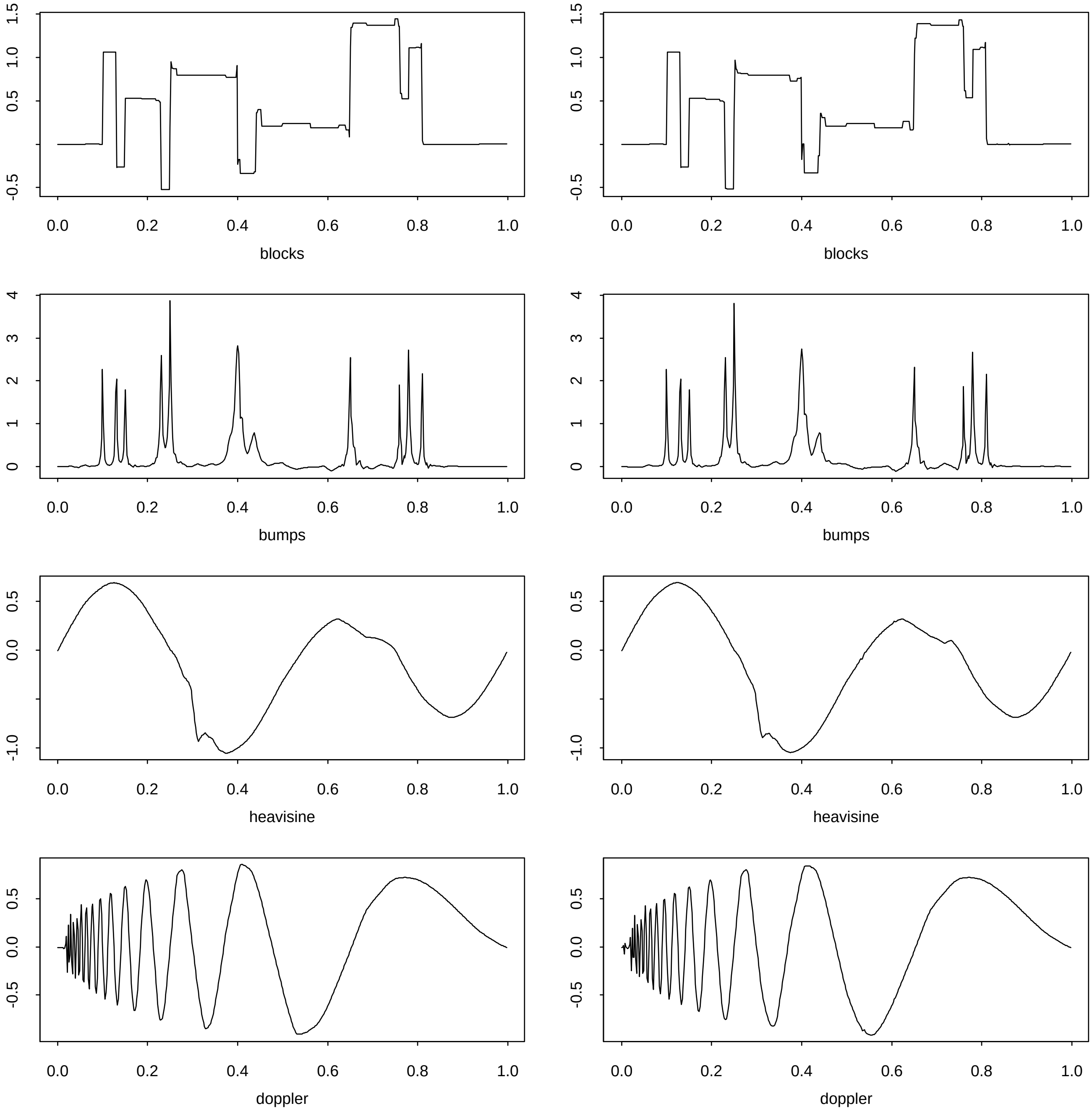

Figure 5: WaveShrink estimates with known variances (example 2): left column are by hard shrinkage with minimax threshold $\lambda = 3.312$; right column are by garrote with minimax threshold $\lambda = 2.441$.

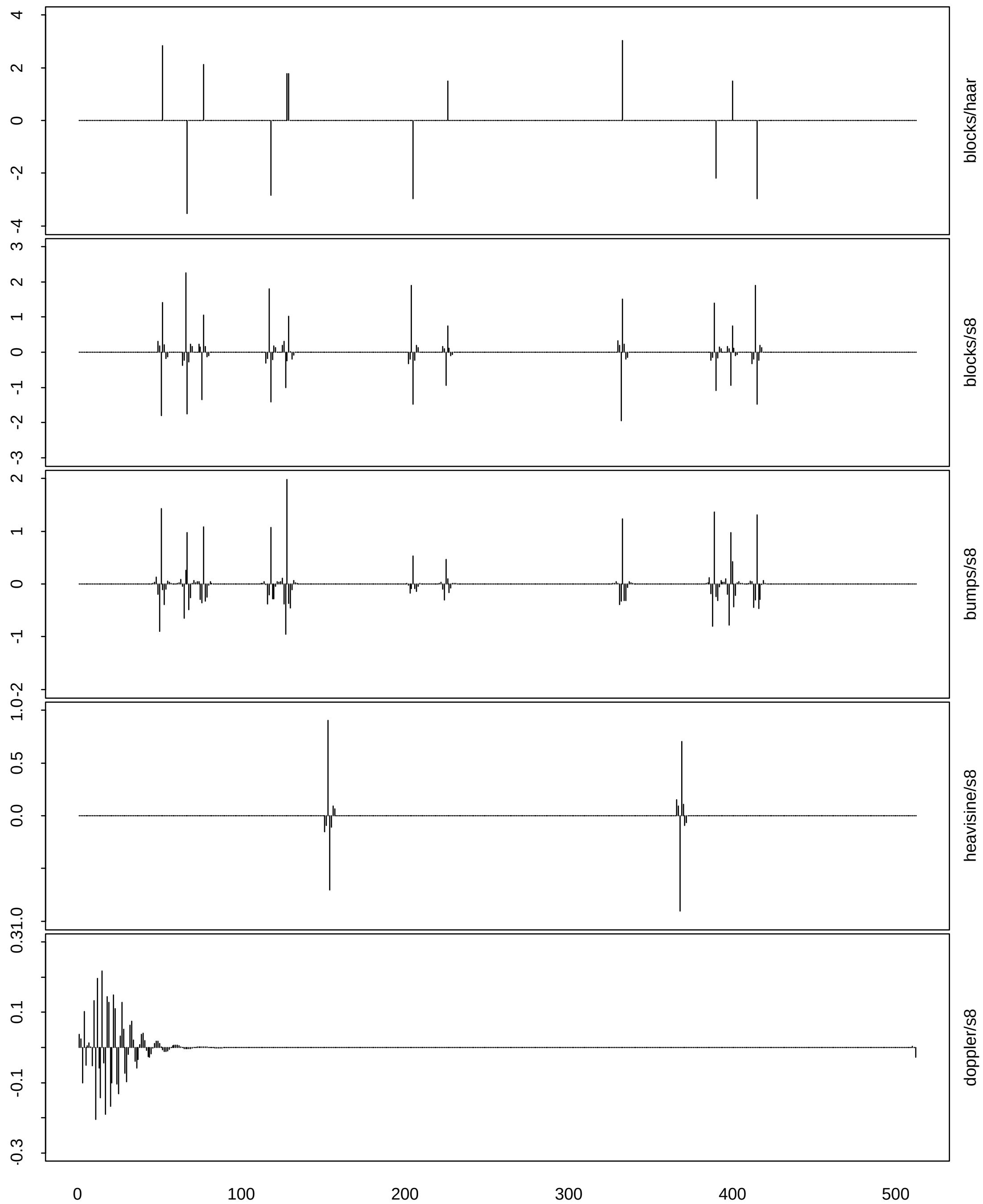

Figure 6: Finest level non-decimated wavelet coefficients for signal/wavelet combinations: (from top to bottom) blocks/Haar, blocks/s8, bumps/s8, heavisine/s8 and doppler/s8.

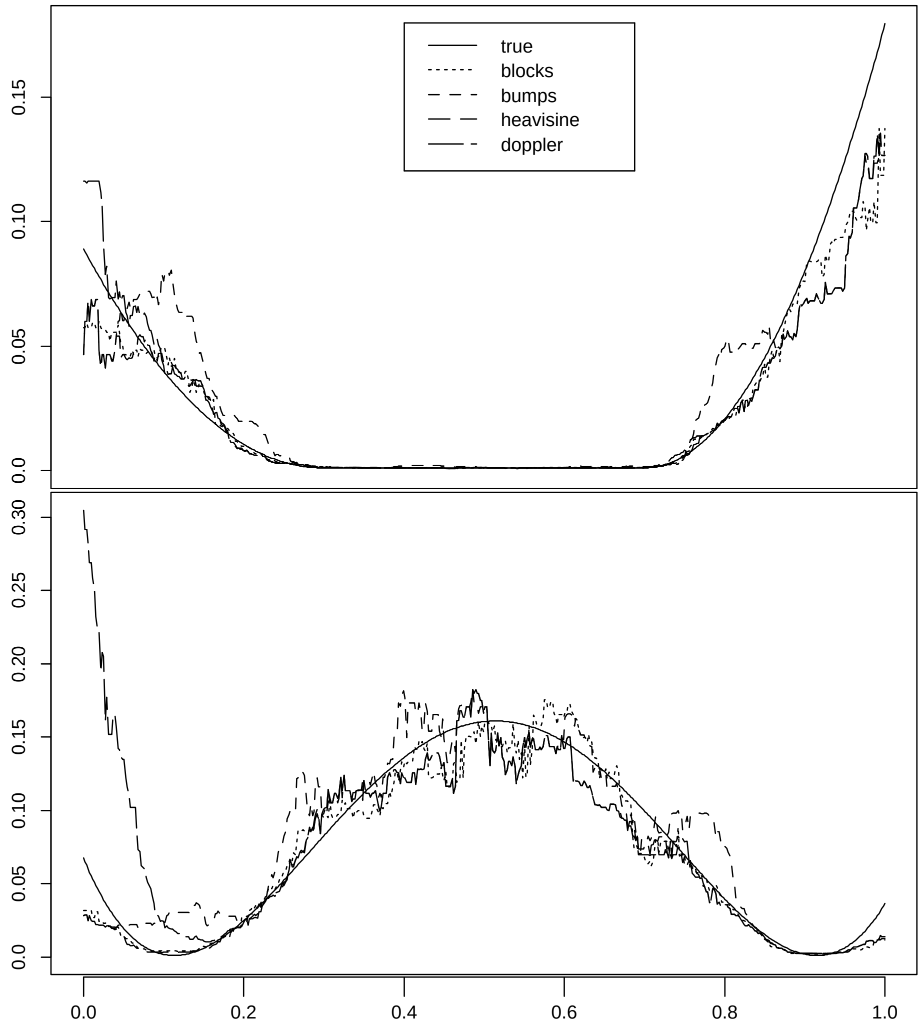

Figure 7: Estimated standard deviations by using the finest level non-decimated wavelet coefficients and running MAD smoother with a span of 51.

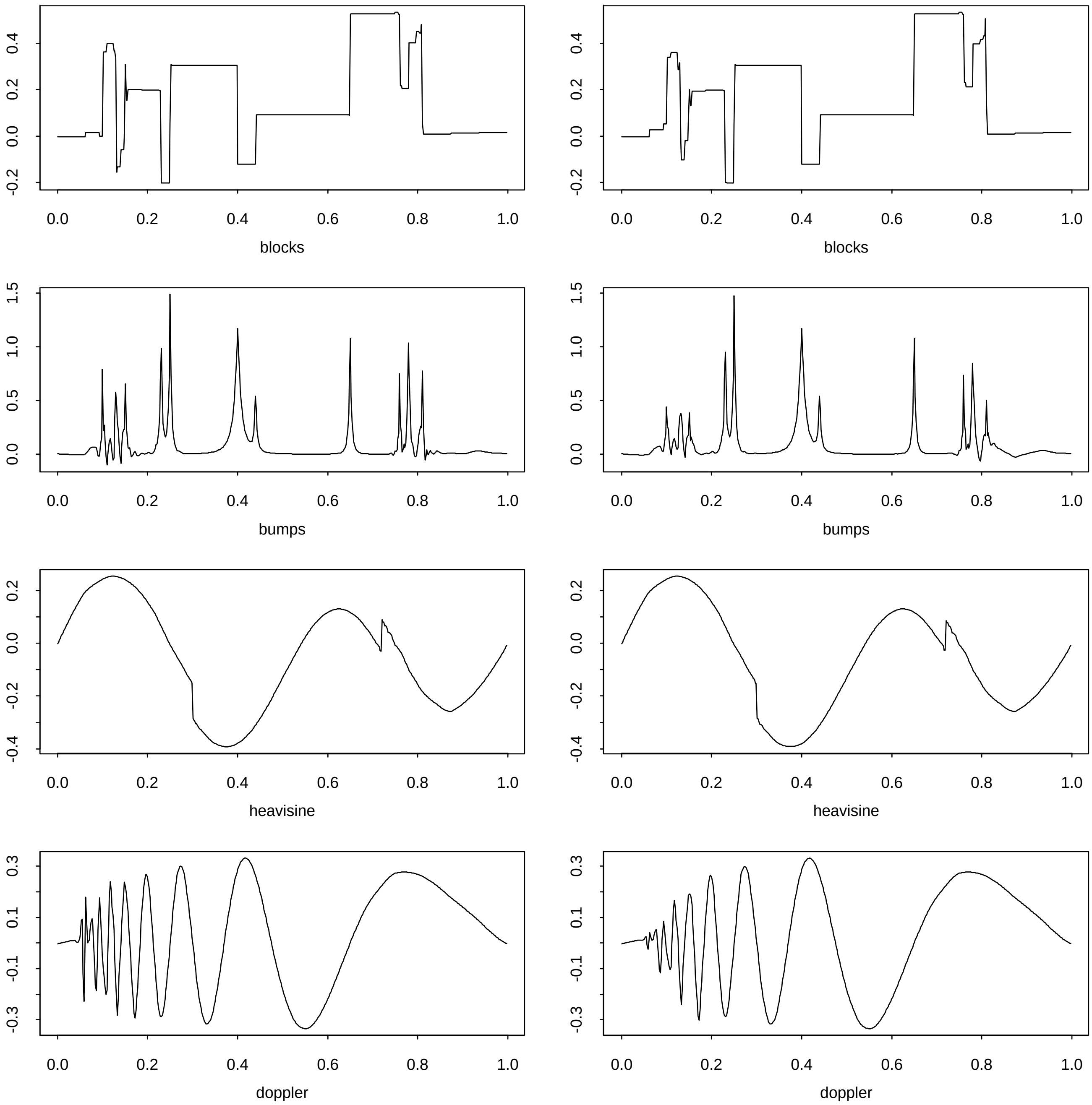

Figure 8: WaveShrink estimates with estimated variances (example 1): left column are by hard shrinkage; right column are by garrote. Threshold $\lambda = 2(\log n)^{1/2} = 4.995$ is used for both cases;

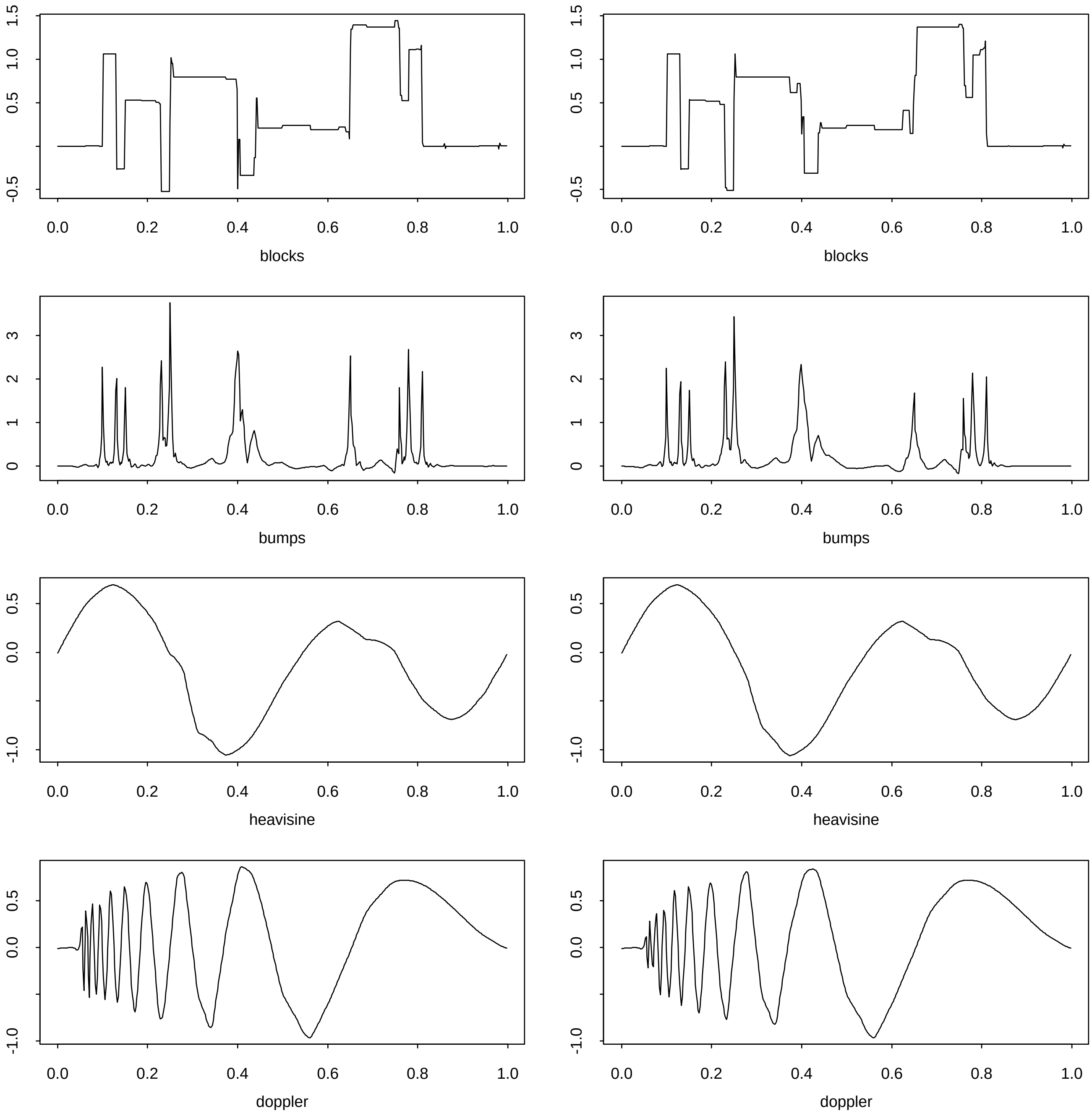

Figure 9: WaveShrink estimates with estimated variances (example 2): left column are by hard shrinkage; right column are by garrote. Threshold $\lambda = 2(\log n)^{1/2} = 4.995$ is used for both cases;