مجله‌ی پژوهش‌های آماری ایران
پژوهشکده‌ی آمار دوره‌ی ۹، شماره‌ی ۲، پاییز و زمستان ۱۳۹۱، صص ۱۳۳–۱۴۶

J. Statist. Res. Iran **9** (2012): 133–146

Parametric Empirical Bayes Test and Its Application to Selection of Wavelet Threshold

Mona Shokripour[†], Adel Mohammadpour[*,‡] and Mina Aminghafari[‡]

[†] Shahid Beheshti University
[‡] Amirkabir University of Technology

Abstract. In this article, we propose a new method for selecting level dependent threshold in wavelet shrinkage using the empirical Bayes framework. We employ both Bayesian and frequentist testing hypothesis instead of point estimation method. The best test yields the best prior and hence the more appropriate wavelet thresholds. The standard model functions are used to illustrate the performance of the proposed method and make comparisons with other traditional methods.

Keywords. Bayes test; parametric empirical Bayes; most powerful test; heavy tailed distribution; unbiased test; wavelet thresholding.

MSC 2010: 62F03, 62C12, 65T60.

1 Introduction

There are several ways to consider prior information in a prior distribution which is the key of Bayesian inference. In *Parametric Empirical Bayes* (PEB) method, the prior distribution which belongs to a parametric family is determined by estimating prior parameters (hyperparameters) (Martiz and Lwin, 1989, e.g. see). This method employs marginal likelihood of hyperparameters given observations which yields a complete data based method of choosing the prior parameters. This approach seems so far unique in combining the

* Corresponding author

properties of fast computation, good theoretical properties and performance in computation of wavelet thresholds (Johnstone and Silverman, 2005).

Consider the non-parametric regression model:

$$y_i = f(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the t_i's are equally spaced points, ε_i are independently distributed as $\mathcal{N}(0, \sigma_\varepsilon^2)$ and f is an unknown function to be recovered through the observations. Recovering f opens a wide class of solution called spatially adaptation. A heuristic solution to spatial adaptation has been put forward by Donoho and Johnstone (1994) and some other methods have been discussed therein. In a special class of soft and hard thresholds, Donoho and Johnstone (1994) explore the admissible level thresholds. Their wavelet-based method for estimation of f proceed by taking the empirical wavelet coefficients of the data y_i and replace small magnitude wavelet coefficients with zero and keep or shrink the other coefficients. The resulting coefficients are denoised and then transform back to obtain the estimate. The quality of estimation is quite sensitive to the choice of threshold, with the best choice being dependent on the problem setting. Abramovich and Benjamini (1996) present a multiple testing to construct a proportion of the universal threshold for each coefficients based on the false recovery rate. Ogden and Parzen (1996) employ change-point approaches and consequently Kolmogorov-Smirnov test to keep or eliminate wavelet coefficients by considering the position of the coefficients as well as their magnitudes. Bayesian approach for adaptation has been studied by Chipman et al. (1997), Clyde et al. (1998) and Silverman (1999) which they all consider prior as a mixture of a Gaussian density and a Dirac mass function at zero. In both theoretically and practically points, the distribution of the wavelet coefficients has tails heavier than Gaussian, so, the heavy tailed density is more appropriate. Johnstone and Silverman (2005) present a class of empirical Bayes methods for level dependent threshold selection in wavelet shrinkage using a heavy tailed mixture prior. The remaining is computing of a posteriori statistic such as mean, median and maximum a posteriori with respect to loss function. Although the posterior mean is the most popular a posteriori characteristics, practical facts may induce some other statistics, for instance Abramovich et al. (1998) investigate the use of the posterior median. Many methods for choosing thresholds have been mentioned by Vidakovic (1999), Chapter 6.

According to the hyperparameter determination for wavelet coefficients, we use testing approach, proposed by Aminghafari and Mohammadpour

(2006), to select the best prior which considerably improves the inference about the wavelet shrinkage, for convenience sake, this method is called PEB test.

The paper is organized as follows. In Section 2, we find the best test function using both frequentist and Bayesian criteria which lead to the best choice for prior. Section 3 contains an introduction of the prior for wavelet coefficients and procedure of selecting the best prior is presented. The last section is dedicated to the comparison of proposed method with traditional PEB approach.

2 Parametric Empirical Bayes Test

In this section, we find the best test function in the framework of PEB assumptions. Let X be a random variable, $f(x|\theta)$ the probability density function (pdf) of X and θ has a prior pdf $\pi(\theta|\tau)$, where τ is the only unknown hyperparameter. First, we test simple versus simple hypothesis according to the following equivalent hypotheses:

$$\begin{cases} H_0 : \tau = \tau_0 \\ H_1 : \tau = \tau_1 \end{cases} \quad \text{or} \quad \begin{cases} H_0 : \pi(\theta|\tau) = \pi(\theta|\tau_0) \\ H_1 : \pi(\theta|\tau) = \pi(\theta|\tau_1) \end{cases}. \tag{1}$$

Using marginal pdf of X, $m(x|\tau)$, we can recover $\pi(\theta|\tau)$ and vice versa, i.e.,

$$m(x|\tau) = \int_{\mathbb{R}} f(x|\theta)\ \pi(\theta|\tau)d\theta \quad \text{and} \quad \pi(\theta|\tau) = \int_{\mathbb{R}} \pi(\theta;\tau|x)m(x|\tau)dx,$$

where $\pi(\theta;\tau|x) \propto \pi(\theta|x)\pi(\theta|\tau)$. So, (1) may replace by:

$$\begin{cases} H_0 : \tau = \tau_0 \\ H_1 : \tau = \tau_1 \end{cases} \quad \text{or} \quad \begin{cases} H_0 : m(x|\tau_0) = \int_{\mathbb{R}} f(x|\theta)\pi(\theta|\tau_0)d\theta = m_0(x) \\ H_1 : m(x|\tau_1) = \int_{\mathbb{R}} f(x|\theta)\pi(\theta|\tau_1)d\theta = m_1(x) \end{cases}, \tag{2}$$

Mentioned hypotheses testing in (2) is equivalent to prior selection, hence the best prior is given by the best test.

Now, let $(X_1, \theta_1), \ldots, (X_n, \theta_n)$ be n independent copies of (X, θ), where only X_i's are observable. Note that, recovering $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ by $\boldsymbol{x} = (x_1, \ldots, x_n)$ is not possible, but we can recover $\pi(\boldsymbol{\theta}|\tau)$ using $\boldsymbol{x}$. Thus, the problem of hyperparameter determination reduces to a classical hypothesis testing.

In the case of simple null versus simple alternative hypothesis, the best test, which is the most powerful (MP) test, is derived via the Neyman-Pearson lemma. To determine the best test function, $\phi(\boldsymbol{x})$, it suffices to minimize the probability of type II error, β, for the fix level of probability of type I error, α. Formally,

$$\begin{aligned}\alpha &= \int_{\mathbb{R}^n} \phi(\boldsymbol{x})\ m_0(\boldsymbol{x})\ d\boldsymbol{x} = E(\phi(\boldsymbol{X}); H_0),\\ \beta &= \int_{\mathbb{R}^n} (1-\phi(\boldsymbol{x}))\ m_1(\boldsymbol{x}) d\boldsymbol{x} = E(1-\phi(\boldsymbol{X}); H_1),\end{aligned} \tag{3}$$

where $m_0(\boldsymbol{x})$ and $m_1(\boldsymbol{x})$ are marginal pdfs under hypotheses H_0 and H_1, respectively and $\mathbf{X} = (X_1, \ldots, X_n)$.

In the general case where the hypothesis are composite, there is no MP test. However, we can find the best test in the class of unbiased test functions and the best test is the uniformly most powerful unbiased (UMPU) (e.g. see Lehmann and Romano, 2005).

For two mentioned cases, the hyperparameter determination returns to test hypotheses. When there is no preferable hypothesis since there is no idea about the true prior, we could assume that the probabilities of two types of errors are equal. In this situation, if there is an MP test, it is also minimax, see also (Fergusen, 1967). The following example is devoted to the first case.

Example 1. Let x be an observation from $X|\theta$ with normal pdf, $\mathcal{N}(\theta, 1)$, where θ is a real unknown mean parameter. Consider testing the following hypotheses:

$$\begin{cases} H_0 : \pi(\theta) = \frac{1}{2}e^{-|\theta|} \\ \\ H_1 : \pi(\theta) = \frac{1}{\sqrt{2\pi}}\left[\frac{1-|\theta|\{1-\Phi(|\theta|)\}}{\phi(\theta)}\right] \end{cases} \text{or} \begin{cases} H_0 : \theta \sim \mathcal{L}(0,1) \\ \\ H_1 : \theta \sim \text{quasi} - \mathcal{C}(0,1) \end{cases},$$

where Φ and ϕ are the standard normal cumulative distribution and normal density, respectively and $\mathcal{L}(\mu, \nu)$ and quasi-$\mathcal{C}(\mu, \nu)$, which described in Johnstone and Silverman (2005) denote Laplace and quasi-Cauchy distributions, respectively with location parameter μ and scale parameter ν. For testing hypotheses like (2), we must calculate the marginal distribution for each

hypothesis, which yields

$$\begin{cases} H_0 : m_0(x) = \frac{1}{2}e^{\frac{1}{2}}[e^x\{1-\Phi(x+1)\} + e^{-x}\Phi(x-1)] \\ H_1 : m_1(x) = \frac{1}{\sqrt{2\pi}}x^{-2}\left[1-e^{\frac{-x^2}{2}}\right] \end{cases}.$$

The best test function based on n observations $\boldsymbol{x} = (x_1, \ldots, x_n)$ is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1 \quad \text{if} \quad \prod_{i=1}^{n} \frac{\frac{1}{2}e^{\frac{1}{2}}[e^{x_i}\{1-\Phi(x_i+1)\}+e^{-x_i}\Phi(x_i-1)]}{\frac{1}{\sqrt{2\pi}}x_i^{-2}\left[1-e^{\frac{-x_i^2}{2}}\right]} \leqslant k \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}.$$

The test statistic has the MLR properties in $\sum_{i=1}^{n}|X_i|$, So we have:

$$\phi(\boldsymbol{x}) = \begin{cases} 1 \quad \text{if} \quad \sum_{i=1}^{n}|x_i| > k^* \\ 0 \quad \text{if} \quad \sum_{i=1}^{n}|x_i| \leqslant k^* \end{cases},$$

where k^* is calculated by (3). Figure 1 indicates the probabilities of two types of errors for this hypothesis.

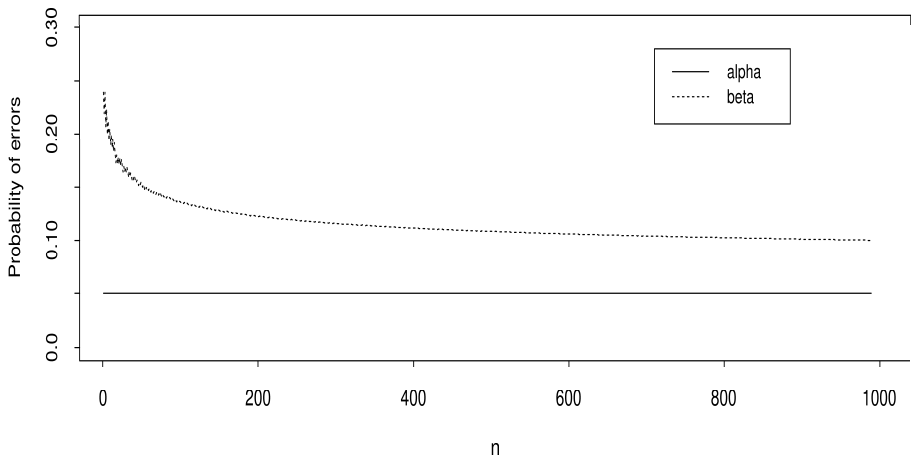

Figure 1. Comparing the probabilities of errors in the proposed method of hyperparameter determination for different sample sizes.

Bayes Method

In the Bayesian framework, testing hypothesis is based on Bayes factor, discussed in Robert (2001), p. 227, which is given by

$$BF_{01} = \frac{k_0(\boldsymbol{x})}{k_1(\boldsymbol{x})},$$

where $k_i(\boldsymbol{x}) = \int_{\mathbb{R}} m(\boldsymbol{x}|\tau)\ g_i(\tau)\mathrm{d}\tau$, $i = 0, 1$ and $g_1(\tau)$, $g_2(\tau)$ are priors for hyperparameter under both hypotheses. The Bayes test function for (1) is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad BF_{01} < \frac{\pi_1}{\pi_0} \\ 0 & \text{if} \quad \text{otherwise} \end{cases}, \tag{4}$$

where π_i, $i = 0, 1$ are priors probabilities related to the hypotheses H_0 and H_1, respectively.

We compare the proposed method in the classical and the Bayesian framework for hyperparameter determination. Let $X_i|\theta_i \sim \mathcal{N}(\theta_i, 1)$, $i = 1, \ldots, n$ and $\theta_i \sim \mathcal{N}(\tau, \sigma^2)$. Consider following hypotheses:

$$\begin{cases} H_0 : \theta \sim \mathcal{N}(\tau_0, \sigma^2) \\ H_1 : \theta \sim \mathcal{N}(\tau_1, \sigma^2) \end{cases} \qquad \text{or} \qquad \begin{cases} H_0 : \tau = \tau_0 \\ H_1 : \tau = \tau_1 \end{cases},$$

where τ_0 and τ_1 are known constants.

First, σ^2 considered to be known. The best test function in the proposed method is given by

$$\phi_1(\boldsymbol{x}) = \begin{cases} 1 \text{ if } \frac{\sqrt{n}(\bar{x}-\tau_0)}{\sqrt{2}} \geqslant z_{1-\alpha} \\ 0 \text{ if } \frac{\sqrt{n}(\bar{x}-\tau_0)}{\sqrt{2}} < z_{1-\alpha} \end{cases},$$

where $\bar{x} = \sum_{i=1}^{n} x_i$ and $z_{1-\alpha}$ is the $100(1-\alpha)$th percentile of normal distribution. The Bayes test function based on (4) is given by

$$\psi_1(\boldsymbol{x}) = \begin{cases} 1 \text{ if } \exp\{\frac{n}{2}(\tau_0 - \tau_1)\bar{x} - \frac{n}{4}(\tau_0^2 - \tau_1^2)\} < \frac{1-\pi_0}{\pi_0} \\ 0 \text{ if } \exp\{\frac{n}{2}(\tau_0 - \tau_1)\bar{x} - \frac{n}{4}(\tau_0^2 - \tau_1^2)\} \geqslant \frac{1-\pi_0}{\pi_0} \end{cases}.$$

Second, for unknown σ^2, the best test function in the proposed method is given by

$$\phi_2(\boldsymbol{x}) = \begin{cases} 1 \text{ if } \frac{\sqrt{n}(\bar{x}-\tau_0)}{s} \geqslant t_{n-1,1-\alpha} \\ 0 \text{ if } \frac{\sqrt{n}(\bar{x}-\tau_0)}{s} < t_{n-1,1-\alpha} \end{cases},$$

where $\bar{x}$ and s are the sample mean and standard deviation, respectively, and $t_{n-1,1-\alpha}$ is the $100(1-\alpha)$th percentile of Student's t-distribution. In the Bayesian approach, we must choose π_0 and prior for σ^2. Consider $\sigma^2 \sim I\chi^2(n-1,\sigma_0^2)$ which is inverse chi-square conjugate prior, the Bayes test function is given by

$$\psi_2(\boldsymbol{x}) = \begin{cases} 1 \text{ if } \sum_{i=1}^{n} \ln \frac{(n-1)\sigma_0^2+(x_i-\tau_1)^2}{(n-1)\sigma_0^2+(x_i-\tau_0)^2} < \frac{2}{n}\ln(\frac{1-\pi_0}{\pi_0}) \\ 0 \text{ if } \sum_{i=1}^{n} \ln \frac{(n-1)\sigma_0^2+(x_i-\tau_1)^2}{(n-1)\sigma_0^2+(x_i-\tau_0)^2} \geqslant \frac{2}{n}\ln(\frac{1-\pi_0}{\pi_0}) \end{cases}.$$

Given same sample sizes, the probabilities of two types of errors for two approaches are shown in Figure 2. We assume that there are some evidences about true prior, i.e., it can be considered that $\alpha = 0.1, 0.01$ and $\pi_0 = 0.9, 0.99$. The first two rows of this figure dedicate to the known σ^2 and the rest are considered by unknown σ^2. In the Bayesian approach, the probabilities of errors can be approximated by a Monte Carlo simulation.

If we have some evidence about prior, we can conclude that the Bayes method is accepted wrongly more than the proposed method, especially in small sample size.

3 An Application to Wavelet Thresholds

Consider the non-parametric regression model:

$$y_i = f(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the t_i's are equally spaced points, ε_i are Gaussian white noise with unknown variance σ_ε^2 and f is an unknown function to be recovered through the observations, y_i. The standard wavelet method to remove noise and estimate function f, is based on thresholding the discrete wavelet transform of observations and then transforming back to obtain the estimation. Wavelet thresholding (Donoho and Johnstone, 1994; Donoho et al., 1995) is a popular method for noise reduction.

The discrete wavelet transform (DWT) of $y = (y_1, \ldots, y_n)'$ using an orthonormal wavelet at level J can be written as follows:

$$\begin{aligned} &\hat{c}_{J,k} = c_{J,k} + \varepsilon_{J,k} \ k = 1, \ldots, 2^J, \\ &\hat{d}_{j,k} = d_{j,k} + \varepsilon_{j,k} \ k = 1, \ldots, 2^j, \qquad j = 1, \ldots, J, \end{aligned} \tag{5}$$

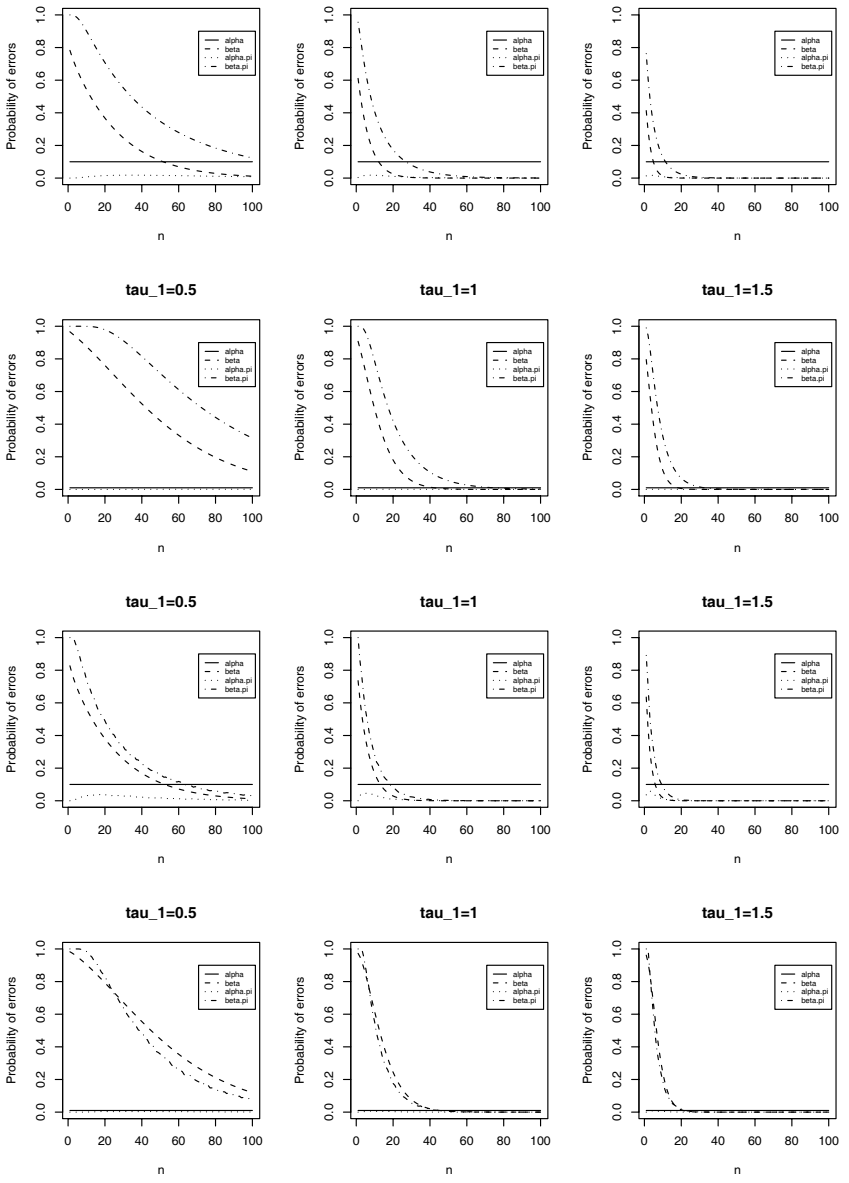

Figure 2. Comparing the probabilities of errors in the proposed and the Bayesian methods of hyperparameter determination, we assume that $\alpha = 0.1, 0.01$ and $\pi_0 = 0.9, 0.09$ for different $\tau_1 = 0.5, 1, 1.5$ when $\tau_0 = 0$ and $\sigma_0^2 = 1$. Top two rows: σ^2 known. Bottom two rows: σ^2 unknown.

where $\hat{d}_{j,k}$ and $\hat{c}_{j,k}$ are the empirical wavelet and scaling coefficients of noisy data y and $d_{j,k}$ and $c_{j,k}$ are the wavelet and scaling coefficients of function F, respectively and $\varepsilon_{j,k}$ are Gaussian white noise with the same variance as ε_i due to the orthogonality of DWT. The empirical scaling coefficients $\hat{c}_{j,k}$s are kept intact because they contain important components about the underlying function.

According to the equations (5), we have $\hat{d}_{j,k} \sim \mathcal{N}(d_{j,k}, \sigma_j^2)$. To obtain Bayesian shrinkage, we should consider suitable prior for the wavelet coefficients of f, $d_{j,k}$. Vidakovic and Ruggeri (2001) and Johnstone and Silverman (2005) have considered a particular mixture prior to this problem. Under this prior the d_{jk} are independently distributed with

$$\pi(d_{j,k}) = \epsilon_j \gamma(d_{j,k}) + (1 - \epsilon_j)\delta(0),$$

a mixture of Dirac mass function at zero and a density γ. They apply posterior mean, posterior median or maximum a posteriori for wavelet shrinkage estimation. In the proposed framework, we apply hypotheses testing to choose the best prior and thresholding the wavelet coefficients. We suppose two different heavy tailed distributions which are double exponential (Laplace), γ_1, and quasi-Cauchy, γ_2 as follows:

$$\begin{aligned} \gamma_1(d_{j,k}) &= \frac{1}{2} e^{-|d_{j,k}|}, \\ \gamma_2(d_{j,k}) &= \frac{1}{\sqrt{2\pi}} \left\{ 1 - |d_{j,k}| \frac{1 - \Phi(|d_{j,k}|)}{\phi(d_{j,k})} \right\}. \end{aligned} \tag{6}$$

Afterwards, we investigate the testing hypothesis to achieve the best prior. We could neglect the effect of sampling variability in the estimation of the noise variance at level j by defining $z_k = \frac{\hat{d}_{jk}}{\sigma_j}$ and apply the approach to this sequence. The estimated wavelet coefficients of the discrete wavelet transform of the sequence $f(t_i)$ are then given by:

$$d^*_{jk} = \sigma_j \hat{f}\left(d_{jk}; \frac{\hat{d}_{jk}}{\sigma_j} \right).$$

The three hypotheses are performed as follows:

$$\begin{cases} H_0 : \pi(d_{j,k}) = \delta(0) \\ H_1 : \pi(d_{j,k}) = \gamma_1(d_{j,k}) \end{cases}, \qquad \begin{cases} H_0 : \pi(d_{j,k}) = \delta(0) \\ H_1 : \pi(d_{j,k}) = \gamma_2(d_{j,k}) \end{cases}$$

$$\text{and} \qquad \begin{cases} H_0 : \pi(d_{j,k}) = (1-\epsilon_j)\delta(0) + \epsilon_j \gamma_1(d_{j,k}) \\ H_1 : \pi(d_{j,k}) = (1-\epsilon_j)\delta(0) + \epsilon_j \gamma_2(d_{j,k}) \end{cases} \tag{7}$$

where $\gamma_1(d_{j,k})$ and $\gamma_2(d_{j,k})$ are given in (6), these hypotheses are equivalent to the following hypotheses, respectively.

$$\begin{cases} H_0 : m_0(\hat{d}_{j,k}) = \frac{1}{2}\exp\{\sqrt{2}|\hat{d}_{j,k}|\} \\ H_1 : m_1(\hat{d}_{j,k}) = \exp(-|\hat{d}_{j,k}|) - \frac{1}{\sqrt{2}}\exp(\sqrt{2}|\hat{d}_{j,k}|), \end{cases}$$

$$\begin{cases} H_0 : m_0(\hat{d}_{j,k}) = \frac{1}{\sqrt{2\pi}}\left\{1 - |\hat{d}_{j,k}|\frac{1-\Phi(|\hat{d}_{j,k}|)}{\phi(\hat{d}_{j,k})}\right\} \\ H_1 : m_1(\hat{d}_{j,k}) = (2\pi)^{-\frac{1}{2}}\hat{d}_{j,k}^{-2}\left\{1 - \exp\left(-\frac{\hat{d}_{j,k}^2}{2}\right)\right\} \end{cases}$$

and

$$\begin{cases} H_0 : m_0(\hat{d}_{j,k}) = (1-\epsilon_j)\frac{1}{2}e^{-|\hat{d}_{j,k}|} + \epsilon_j g_1(\hat{d}_{j,k}) \\ H_1 : m_1(\hat{d}_{j,k}) = (1-\epsilon_j)(2\pi)^{-\frac{1}{2}}\left\{1 - |\hat{d}_{j,k}|\frac{1-\Phi(|\hat{d}_{j,k}|)}{\phi(|\hat{d}_{j,k}|)}\right\} + \epsilon_j g_2(\hat{d}_{j,k}) \end{cases} \tag{8}$$

where g_1 and g_2 are

$$g_1(\hat{d}_{j,k}) = \frac{\exp\{-|\hat{d}_{j,k}|\}}{\exp\{-\sqrt{2}|\hat{d}_{j,k}|\}},$$
$$g_2(\hat{d}_{j,k}) = (2\pi)^{-\frac{1}{2}}\hat{d}_{j,k}^{-2}\left\{1 - \exp\left(-\frac{\hat{d}_{j,k}^2}{2}\right)\right\}.$$

Therefore, by the Neyman-Pearson lemma, these tests are obtained as follows

$$\frac{m_1(\hat{d}_{j,k})}{m_0(\hat{d}_{j,k})} > C_0.$$

We can deduce from that in these three tests,

$$|\hat{d}_{j,k}| > C_1,$$

where $C_1 > 0$ is a constant determined by test size

$$P(|\hat{d}_{j,k}| > C_1; H_0) = \alpha. \tag{9}$$

So the test function can be written as follows

$$\phi^*(\hat{d}_{j,k}) = \begin{cases} 1 & \text{if } |\hat{d}_{j,k}| > C_1 \\ 0 & \text{if } |\hat{d}_{j,k}| < C_1 \end{cases}.$$

We reject H_0 in (7) if $|\hat{d}_{j,k}|$ is greater than a threshold. When the hypothesis H_0 is not rejected in (7) or equivalently in (8), $|\hat{d}_{j,k}|$ comes from noise and it is discarded otherwise it is retained (see Vidakovic, 1999, p. 202). This is exactly ordinary shrinkage or threshold rule when the threshold is selected based on α and observation.

4 Comparing Methods

In this section, we apply the proposed method to select threshold in the wavelet denoising context of the simulated examples built from well known signals introduced in Antoniadis et al. (2001). The following three methods are compared:

1. Laplace(median): Empirical Bayesian method with a Laplace prior and compute the median of posterior as an estimator proposed in Johnstone and Silverman (2005).

2. Quasi-Cauchy(median): Empirical Bayesian method with quasi-Cauchy prior and compute the median of posterior as an estimator proposed in Johnstone and Silverman (2005).

3. FDR: The False Discovery Rate method is derived from the principle of controlling the false discovery rate in simultaneous testing hypothesis and has been studied in Benjamini and Hochberg (1995).

4. PEB test: Using Parametric Empirical Bayes Test, hard thresholding is performed using obtained threshold in equation (9).

The performance of each method is measured by the error sum of square (ESS) over 100 runs:

$$\text{ESS} = \frac{1}{100\hat{\sigma}_{\tilde{\varepsilon}}^2} \sum_{k=1}^{100} \sum_{n=1}^{N} \{f^k(n) - \breve{y}_n^k\}^2.$$

The performance results are computed for four standard samples, bumps (bmp), blocks (blk), heavysin (hea) and doppler (dop), describe in Johnstone and Silverman (2005) which are present in Table 1, respectively. In all methods, the variance should be estimated in each case. We see that, for heavysin and Doppler signals with high noise, the proposed method has the best performance and for the bumps signal has the worst performance. In the PEBT method we could use a prior for the variance with fixed mean (recall that variance has an exponential distribution with parameter 1). However, for the other methods, the variance should be estimated in each case. Also, we can use Generalized Likelihood Ratio Test (GLRT) to find threshold in the previous section when the hyperparameter of variance prior is unknown. We can conclude that in all four signals the proposed method work at least as good as the FDR method except in bump signal.

Table 1. Performance of the three methods: Average over 100 replications of summed squared errors over 1024 points for various models and methods.

	High noise				Low noise			
Method	bmp	blk	dop	hea	bmp	blk	dop	hea
Laplace (median)	171	176	93	41	**212**	164	**109**	57
Quasi-Cauchy (median)	177	185	97	40	221	169	115	56
FDR (0.05)	**169**	**173**	93	39	223	**163**	110	53
PEB test	173	178	**90**	**38**	220	166	110	**52**

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

References

Abramovich, F. and Benjamini, Y. (1996). Adaptive Thresholding of Wavelet Coefficients. *Computational Statistics and Data Analysis*, **22**, 351-361.

Abramovich, F., Sapatinas, T. and Silverman, B.W. (1998). Wavelet Thresholding Via A Bayesian Approach. *Journal of Royal Statistical Society, Series B*, **60**, 725-749.

Aminghafari, M. and Mohammadpour, A. (2006). An Alternative Approach to the Parametric Empirical Bayes Selection of Wavelet Thresholding. *AIP Conference Proceeding of Bayesian Inference and Maximum Entropy Methods in Science and Engineering. 26th International Workshop, Paris*, edited by Mohammad-Djafari, A., **CP872**, 433-440.

Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study. *Journal of Statistical Software*, **6**, 1-83.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of Royal Statistical Society, Series B*, **57**, 289-300.

Chipman, H.A., Kolaczyk, E.D. and McCulloch, R.E. (1997). Adaptive Bayesian Wavelet Shrinkage. *Journal of American Statistical Association*, **92**, 1413-1421.

Clyde, M., Parmigiani, G. and Vidakovic, B. (1998). Multiple Shrinkage and Subset Selection in Wavelets. *Biometrika*, **85**, 391-401.

Donoho, D.L. and Johnstone, I.M. (1994). Ideal Spatial adaptation by Wavelet Shrinkage. *Biometrika*, **81**, 425-455.

Donoho, D.L., Johnstone, I.M., Kerkyacharian G. and Picard, D. (1995). Wavelet Shrinkage: Asymptopia? *Journal of Royal Statistical Society, Series B*, **57**, 301-369.

Fergusen, S.T. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

Johnstone, I.M. and Silverman, B.W. (2005). Empirical Bayes Selection of Wavelet Thresholds. *The Annals of Statistics*, **33**, 1700-1752.

Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Wiley, New York.

Martiz, S.J. and Lwin, T. (1989). *Empirical Bayes Methods*, 2nd ed. Chapman and Hall, London.

Ogden, R.T. and Parzen, E. (1996). Change-Point Approach to Data Analytic Wavelet Thresholding. *Statistics and Computing*, **6**, 93-99.

Robert, P.C. (2001). *The Bayesian Choice*, 2nd ed. Springer-Verlag, New York.

Silverman, B.W. (1999). Wavelets in Statistics: Beyond the Standard Assumptions. *Royal of Society London Philosophy Transaction, Series A: A Mathematical Physics English Society*, **357**, 2459-2473.

Vidakovic, B. and Ruggeri, F. (2001). BAMS Method: Theory and Simulations. *Sankhyā, Series B*, **63**, 234-249.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley, New York.

Mona Shokripour
Department of Statistics,
Faculty of Mathematical Science,
Shahid Beheshti University,
Tehran, Iran.
email: *m_shokripour@sbu.ac.ir*

Adel Mohammadpour
Department of Statistics,
Faculty of Mathematics & Computer Science,
Amirkabir University of Technology,
Tehran, Iran.
email: *adel@aut.ac.ir*

Mina Aminghafari
Department of Statistics,
Faculty of Mathematics & Computer Science,
Amirkabir University of Technology,
Tehran, Iran.
email: *aminghafari@aut.ac.ir*