Communications of the Korean Statistical Society Vol. 15, No. 3, 2008, pp. 379–386

The Choice of a Primary Resolution and Basis Functions in Wavelet Series for Random or Irregular Design Points Using Bayesian Methods

Chun Gun Park[1)]

Abstract

In this paper, the choice of a primary resolution and wavelet basis functions are introduced under random or irregular design points of which the sample size is free of a power of two. Most wavelet methods have used the number of the points as the primary resolution. However, it turns out that a proper primary resolution is much affected by the shape of an unknown function. The proposed methods are illustrated by some simulations.

Keywords: Wavelet series; primary resolution; wavelet basis functions; Bayesian methods.

1. Introduction

Consider a nonparametric regression

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $f(x) \in L_2(\mathbb{R})$ an unknown function, the x's are random or irregular design points satisfying $x_i < x_{i+1}$ and ε's independent and identically distributed random variables from the distribution $N(0, \sigma^2)$. The goal is to estimate the unknown function $f(\cdot)$. Many techniques have been developed to estimate $f(\cdot)$. Over the last decade, the nonparametric regression was dominated by wavelet shrinkage and wavelet thresholding estimators. To explain terminology, a shrinkage rule shrinks wavelet coefficients to zero and a thresholding rule sets to zero all coefficients below a certain level.

Among wavelet-based techniques, Bayesian approaches to a wavelet series have been studies as a technique on wavelet shrinkage and thresholding rules. These methods impose a prior model on the wavelet coefficients of the wavelet series approximating to the unknown function. The prior for each wavelet coefficient is typically a mixture of two non-degenerate distributions, or a mixture of a non-degenerate distribution and a point mass at zero, the latter of which represents sparseness of the wavelet series. For a recent comparative simulation study on wavelet estimators in nonparametric regression (see, Antoniadis *et al.*, 2001).

Under a fixed primary resolution, thresholding rules are applied to the coefficients whose resolution levels are equal to or finer than the primary resolution. The role of the

1) Researcher, Seoul Development Institute, Seocho-gu, Seoul 137-070, Korea.
Correspondence: chungunpark@daum.net

primary resolution parameter is similar to that of the usual bandwidth in linear smoother. The relationships among error variance, thresholding rules and primary resolutions were investigated by Hall and Patil (1995). Park *et al.* (2008) have recently proposed on selecting the primary resolution and wavelet basis functions under equally spaced points.

In this paper, the focus will be on the choice of the primary resolution and wavelet basis functions with random or irregular designs. Various attempts to tackling the problem of the irregular designs have been made: see, for instance, the interpolation method of Kovac and Silverman (2000); the binning method of Antoniadis *et al.* (1997); and the transformation method of Cai and Brown (1998), or its recent refinement by Maxim (2002) for a random design. See also Pensky and Vidakovic (2001).

The organization of the paper is as follows. Section 2 describes wavelets. In Section 3, Bayesian approaches to selecting the primary resolution and wavelet basis functions are proposed. Section 4 gives the results of the simulation studies. The conclusion is addressed in Section 5.

2. Wavelet Series

An orthogonal wavelet basis in $L(\mathbb{R})$ is a collection of functions obtained as translations and dilations of a scaling function ϕ and a wavelet function ψ (Daubechies, 1992). The wavelet series is one possible way to represent the unknown function. The form of the wavelet series is following:

$$f(x) = \sum_{k\in\mathbb{Z}} s_{J_0,k}\phi_{J_0,k}(x) + \sum_{j\geq J_0}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x), \tag{2.1}$$

for any integer.

The wavelet series with a primary resolution is to project the data from the (2.1) onto a sequence of the multiresolution analysis(MRA) of Mallat (1989). It can be expressed in terms of the scaling function only:

$$P_m f(x) = \sum_{k\in\mathbb{Z}} s_{J_0,k}\phi_{J_0,k}(x) + \sum_{j\geq J_0}^{m}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) = \sum_{k\in\mathbb{Z}} c_{(m,k)}\phi_{(m,k)}(x). \tag{2.2}$$

The MRA construction implies that $\lim_{m\to\infty} P_m f(x) = f(x)$.

For detailed expositions of the statistical settings of wavelets (see, Härdle *et al.*, 1998; Vidakovic, 1999; Abramovich *et al.*, 2000; Antoniadis and Sapatinas, 2001).

3. The Proposed Bayesian Selection for Wavelet Series

3.1. The choice of a primary resolution

Given a primary resolution, the model (2.2) can be rewritten as

$$Y = W_m\beta_m + \varepsilon, \tag{3.1}$$

where $Y = (y_1,\ldots,y_n)^T$ is a response variable, W_m is the orthogonal matrix $n\times N(m)$ associated with the wavelet basis functions, $\{\phi_{J_0,k}\}$ and $\{\psi_{j,k}\}$ and $\beta_m = (\beta_{(m,1)},\cdots,$

$\beta_{(m,N(m))})^T$ is a $N(m) \times 1$ vector with the wavelet coefficients. Here, $N(m)$ denotes the number of wavelet basis functions at the primary resolution m. We assume the error vector ε follows $N(0, \sigma^2 I_n)$. Direct application of a Bayesian model selection model for the choice of a primary resolution has some drawbacks: (1) it tends to select a low primary resolution regardless of the unknown function as the sample size increases and (2) the computational cost is expensive since the number of the wavelet coefficients increases rapidly as the level of a primary resolution gets high.

To tackle these problems we use the following priors to obtain a posterior probability of the primary resolution parameter m:

1. Given σ^2 and m, the prior for $\beta_{(m,k)}$ is

$$P\left(\beta_{(m,k)}|\sigma^2, m\right) \propto \text{Constant}, \quad 1 \le k \le N(m). \tag{3.2}$$

2. The prior of m is noninformative, that is,

$$P(m) = \frac{1}{M}, \quad m = 0, 1, \ldots, M-1, \tag{3.3}$$

   for a large positive integer M. Thus, all of the resolution levels up to $M-1$ are equally likely a priori.

From (3.2), (3.3) and the orthonormal property $W_m^T W_m = I_n$, the posterior probability of a level m is given by

$$P(m|\sigma^2, Y) \propto \exp\left(\frac{1}{2\sigma^2} Y^T W_m W_m^T Y\right). \tag{3.4}$$

A detailed derivation is given in Appendix from Park *et al.* (2007). A level of primary resolution is then determined as follows. the posterior probability (3.4) increases as m increases. Therefore, given a large error variance which is a nuisance parameter, we use this criterion by selecting a low level m where

$$\mathrm{R}_{(j+1,j)} = \frac{P(m_{j+1}|Y, \sigma^2)}{P(m_j|Y, \sigma^2)}, \tag{3.5}$$

is closest to 1. Here m_j denotes a resolution at level j. However, in practice, we choose m of the case that the first hits a predetermined number δ which close to 1. Note that since the δ is required, the proposed method is not an automatic procedure for selecting the primary resolution level.

3.2. The selection of wavelet basis functions

The main idea is to select a subset among $N(m)$ basis functions that makes the most significant contribution to the posterior probability given by (3.1). There are $2^{N(m)}$ subsets to choose, which increases very rapidly as m gets large. The computation of posterior probabilities for all these subsets would be out of reach. An efficient way of tackling this problem is to rearrange the $N(m)$ wavelet basis functions in the order of their importance as we describe it below.

Given the primary resolution, we rewrite the linear model (3.1) in order to select the wavelet basis function as follows:

$$Y = W_m^* \alpha_m + \varepsilon. \tag{3.6}$$

Here, W_m^* is the resulting matrix obtained by exchanging the columns of W_m according to the size of the elements of $|W_m^T Y|$: put the values of $\phi_{(m,1)}$ in the first column for which $|\sum_{i=1}^n \phi_{(m,1)}(x_i) y_i|$ is the smallest among $|\sum_{i=1}^n \phi_{(m,j)}(x_i) y_i|$, $j = 1, \ldots, N(m)$ and those of $\phi_{(m,N(m))}$ in the last for which $|\sum_{i=1}^n \phi_{(m,N(m))}(x_i) y_i|$ is the largest among those. Furthermore, α_m denotes the coefficient vector obtained by rearranging the elements of β_m according to W_m^*.

We consider the following $N(m)$ linear model: for $s = 0, 1, \ldots, N(m) - 1$,

$$Y = W_{m(s)} \alpha_s + \varepsilon, \tag{3.7}$$

where $W_{m(s)}$ denotes the matrix obtained by removing the first s columns of W_m and α_s is the coefficient vector obtained by removing the first s elements of α_m. Selection of a subset of wavelet basis function is done by choosing s among the above $N(m)$ models. We use this idea in the same Bayesian framework as in the selection of a primary resolution.

Following the same way in the previous subsection, we take the following priors for α's and s

1. Given σ^2 and m, the prior for $\alpha_s = (\alpha_{s+1}, \ldots, \alpha_{N(m)})^T$ is

$$P(\alpha_k | \sigma^2, m) \propto \text{Constant}. \tag{3.8}$$

2. The prior of s is noninformative, that is,

$$P(s) = \frac{1}{N(m)}, \quad s = 0, 1, \ldots, N(m) - 1. \tag{3.9}$$

   Thus, all of the models have equally likely a priori.

Then, the probability of s is given by

$$P(s | \sigma^2, Y) \propto \exp\left(\frac{1}{2\sigma^2} Y^T W_{m(s)} W_{m(s)}^T Y \right). \tag{3.10}$$

A choice of s in (3.10) follows the same way as in the selection of a primary resolution.

4. Simulation Study

In this section, we report the results of simulations made for the choice of a primary resolution and wavelet basis functions based on our proposed method.

4.1. Experimental setup

All the data of the simulations is of the form (1.1). The sample sizes selected were $n = 100$ (intermediate sample size) and $n = 2000$ (large). For each sample size, we used two test functions to assess the choice of a primary resolution and wavelet basis functions,

Table 4.1: Formulae of the test functions

Test function	Formula
Cosine	$g_1(x) = 0.5\cos(2.2\pi/(3+8x)) + 0.5$
Doppler	$g_2(x) = 0.6\{\sqrt{x(1-x)}\sin(2.1\pi/(x+0.05)) + 0.5\} + 0.2$

Table 4.2: Primary resolution for the test functions

Test function	Error variance	Sample size	Primary resolution									
			J_0	Detail								
			0	0	1	2	3	4	5	6	7	8
Cosine	0.05^2	$n = 100$			100							
		$n = 2000$			100							
	0.1^2	$n = 100$			100							
		$n = 2000$			100							
Doppler	0.05^2	$n = 100$				100						
		$n = 2000$					96			4		
	0.1^2	$n = 100$				89		11				
		$n = 2000$					100					

which are listed in Table 4.1. The values of σ^2, variance of the noise, were chosen: $\sigma^2 = 0.05^2$ and 0.1^2. For the simulation of each test function, 100 sets of observations were generated and all sample points were determined according to a uniform distribution from the interval $(0, 1)$.

The first function is a *Cosine* function which is necessary to be a relatively lower primary resolution. The second function is a *Doppler* function which might need a higher primary resolution. The both of the functions might be not necessary all wavelet basis functions under a given primary resolution.

To report the results of the simulations, we used the MSE as the numerical measure, that is if given the primary resolution, $\hat{f}(t_i)$ is the empirical estimate function value at any point at which values of scaling and wavelet functions can be computed by interpolation or simply considering the value at the closest point in the grid, then

$$\mathrm{MSE}(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(t_i) - \hat{f}(t_i)\right\}^2.$$

Throughout the simulation study, Daubeachies's wavelet with vanishing moment 8 has been used.

4.2. The choice of the primary resolution and wavelet basis functions

In this simulation, an empirical study for assessing the choice of the primary resolution and wavelet bases in the wavelet series was conducted. Table 4.2 shows the results of selecting primary resolutions for each function with 100 repetitions. The proposed method estimates lower primary resolutions for the *Cosine* function and higher resolutions for the *Doppler* function. In addition, Figure 4.1 shows the posterior probabilities of primary resolutions and the ratios from (3.5) when the *Doppler*, $n = 2000$ and $\sigma^2 = 0.05^2$ were used. Here $m = 3$ was selected. Note that -1 in x-axis label of Figure 4.1 denotes

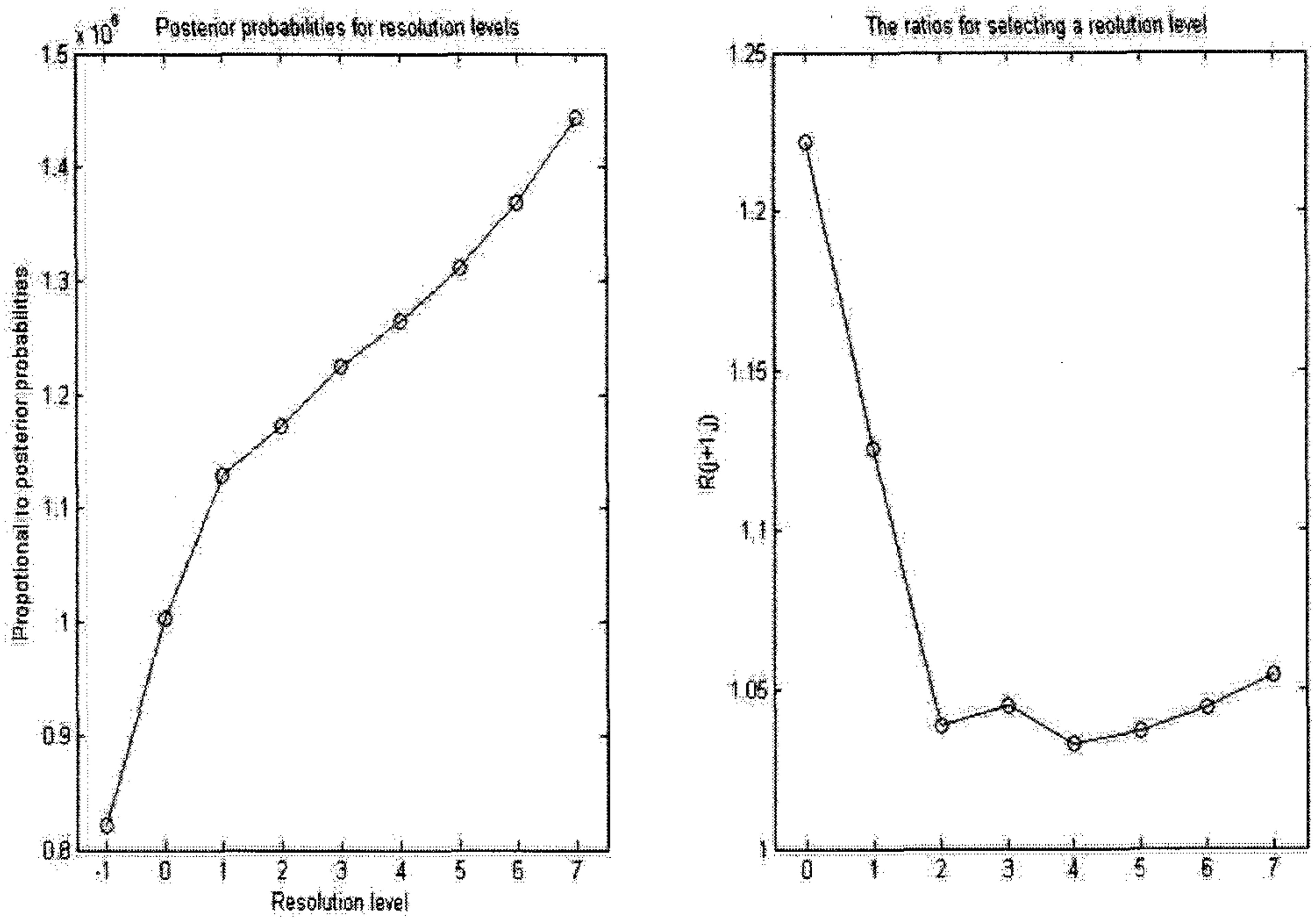

Figure 4.1: Proportional to posterior probabilities for primary resolution levels and ratios for a it Doppler function.

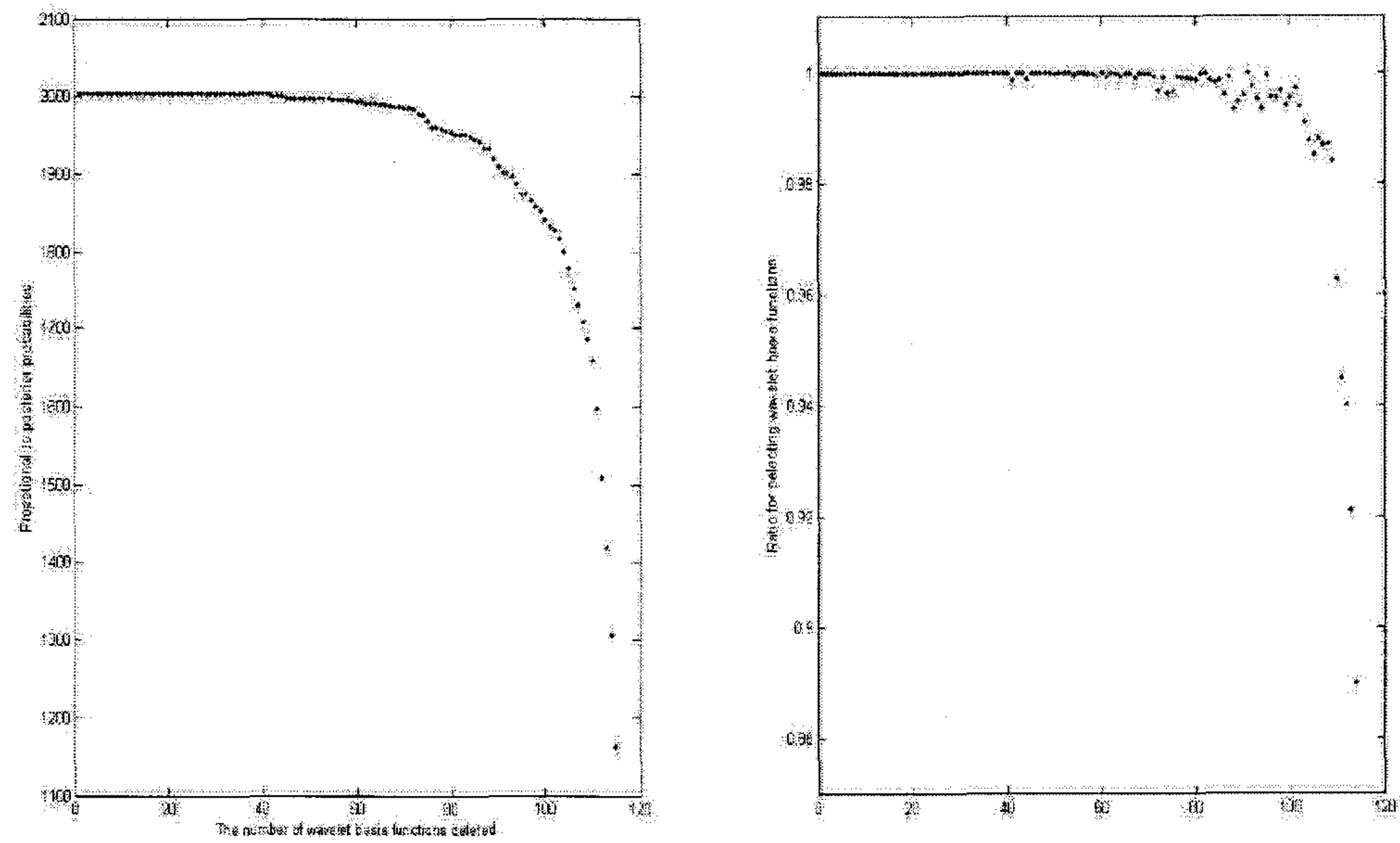

Figure 4.2: Proportional to posterior probabilities for models and ratios for a Doppler function.

Table 4.3: Simulation results in term of integrated MSE

Error variance	Model	Sample size	Function	
			Cosine	*Doppler*
0.05^2	Full	$n =$ 100	$8.37e-4$	$1.07e-2$
		$n =$ 2000	$6.86e-4$	$4.50e-3$
	Reduced	$n =$ 100	$9.85e-4$	$1.14e-2$
		$n =$ 2000	$6.83e-4$	$4.50e-3$
0.1^2	Full	$n =$ 100	$1.76e-3$	$2.08e-2$
		$n =$ 2000	$7.26e-3$	$4.82e-3$
	Reduced	$n =$ 100	$1.87e-3$	$2.12e-2$
		$n =$ 2000	$7.25e-3$	$4.82e-3$

the case of a wavelet series which is expanded only smoothing coefficients without detailed coefficients. Figure 4.2 shows that the probability does not change until 72 wavelet bases with the small magnitudes of the elements of W^TY are deleted.

To evaluate the choice of wavelet bases, two types of models were conducted: (1) a full model with all wavelet bases and (2) a reduced model with the selective wavelet bases. Given the predetermined primary resolutions. Table 4.3 shows the integrated MSE($\hat{f}$)'s values for the full model and the reduced model. As can be seen from Table 4.3, the reduced model displays very similar results even though the only selected wavelet bases are used for estimation.

5. Conclusion

The proposed Bayesian methods are a wavelet shrinkage method for selecting the primary resolution and wavelet basis functions. To implement wavelet thresholding, two steps are proposed: first step to choose a proper primary resolution level and the second to select wavelet basis functions that are corresponding to significant coefficients. This approach is simpler than exiting Bayesian methods for the wavelet shrinkage.

Acknowledgment

The author is most grateful to the reviewers for their most constructive comments.

References

Abramovich, F., Bailey, T. C. and Sapatinas, T. (2000). Wavelet analysis and its statistical applications, *The Statistician,* **49**, 1–29.

Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: A comparative simulation study, *Journal of Statistical Software,* **6**.

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing, *Statistics & Probability Letters,* **35**, 225–232.

Antoniadis, A. and Sapatinas, T. (2001). Wavelet shrinkage for natural exponential families with quadratic variance functions, *Biometrika,* **88**, 805–820.

Cai, T. and Brown, L. D. (1998). Wavelet shrinkage for nonequispaced samples, *The Annals of Statistics,* **26**, 1783–1799.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* (CBMS-NSF regional conference series in applied mathematics), SIAM: Society for industrial and applied mathmatics, Philadelphia.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators, *The Annals of statistics,* **23**, 905–928.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation and Statistical Applications,* (Lecture notes in statistics), **129,** Springer, New York.

Kovac, A. and Silverman, B. W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding, *Journal of the American Statistical Association,* **95**, 172–183.

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

Maxim, V. (2002). Denoising signals observed on a random design, Paper presented to the fifth AFA-SMAI conference on curves and surfaces.

Park, C. G., Oh, H. S. and Lee, H. (2008). Bayesian selection of primary resolution and wavelet basis functions for wavelet regression, *Computational Statistics,* **23**, 291–302.

Pensky, M. and Vidakovic, B. (2001). On non-equally spaced wavelet regression, *Annals of the Institute of Statistical Mathematics,* **53**, 681–690.

Vidakovic, B. (1999). Linear versus nonlinear rules for mixture normal priors, *Annals of Institute of Statistical Mathematics,* **51**, 111–124.

[Received February 2008, Accepted April 2008]