Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 2, pp. 467~476

Determinacy on a Maximum Resolution in Wavelet Series

Chun Gun Park[1] · Yeong-Hwa Kim[2] · Wan-Youn Yang[3]

Abstract

Recently, an approximation of a wavelet series has been developed in the analyses of an unknown function. Most of articles have been studied on thresholding and shrinkage methods for its wavelet coefficients based on (non)parametric and Bayesian methods when the sample size is considered as a maximum resolution in wavelet series. In this paper, regardless of the sample size, we are focusing only on the choice of a maximum resolution in wavelet series. We propose a Bayesian approach to the choice of a maximum resolution based on the linear combination of the wavelet basis functions.

Keywords : Daubechies wavelets, MCMC, Posterior mode, Resolution, Wavelet series

1. Introduction

Suppose that a nonparametric regression model is as follows

$$y_i = g(x_i) + \epsilon_i \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where a function, say g, is any function in L^2 space and the errors, say $\epsilon_i \sim N(0, \sigma^2)$, are independent variables, and $x_i \in [a, b]$ where $a, b \in \mathbb{R}$. The assumption of the function can be made weaker than that of traditional methods,

---

1) First Author : Researcher, Department of Statistics, Seoul National University, Seoul, Korea.

2) Corresponding Author : Assistant Professor, Department of Statistics, Chung-Ang University, Seoul, Korea.
E-mail : gogators@cau.ac.kr

3) Associate Professor, Department of Applied Statistics, Kyungwon University, Sungnam, Korea.

using a wavelet series as an approximation to the function

$$g(x) \approx \sum_{k \in Z} s_{0(k)} \phi_{0(k)}(x) + \sum_{j=0}^{m} \sum_{k \in Z} d_{j(k)} \psi_{j(k)}(x) \tag{2}$$

for m given (Müller and Vidakovic, 1999).

Suppose that the shape of the function is unknown. If the choice of a maximum resolution, say m from (2), depends on a sample size, then in Bayesian methods the estimation of the function has a critical drawback. For an example, if the function is smooth, that is, the smooth part (or a lower maximum resolution required) from (2) represents the function, and the sample size is so large, then the Bayesian methods cause substantially computational "burns". That is, there are so many basis functions which are negligent.

To select a maximum resolution, we apply the Bayesian method proposed by Smith and Kohn (1997) to a wavelet series and modify the prior distribution of the coefficients in the wavelet series. That is, by replacing the regression spline by the wavelet series which is a linear combination of two basis functions called the *father* wavelets, say ϕ, and the *mother* wavelets, say ψ (Daubechies, 1992), the form (2) is a linear model in the wavelet coefficients. We cannot, however, estimate the coefficients in according to least squares, because sometimes the design matrix which consists of the wavelet functions is singular and the estimates are poor. In this paper, we are focusing only on the choice of the maximum resolution based on the Bayesian method modified. This approach is useful for multidimensional cases in which unknown functions' shapes may be unknown.

Section 2 is devoted to explain maximum resolution selection and estimation for a function in a wavelet series. Simulated examples are showed in Section 3, and the conclusion is given in Section 4.

2. Maximum Resolution Selection in a Wavelet Series

2.1 Wavelet series

For suitable wavelet basis functions, ϕ and ψ, which are as follows

$$\phi_{j(k)}(x) = 2^{j/2}(2^j x - k) \text{ and } \psi_{(j)k}(x) = 2^{j/2}\psi(2^j x - k)$$

for dilation j and translation k, the wavelet series from (2) have the orthogonal property

$$s_{0(k)} = \int g(x)\phi_{0k}(x)dx \text{ and } d_{(j)k} = \int g(x)\psi_{(j)k}(x)dx.$$

Intuitively, the first part from (2) is used to represent a smoothing function and the other from (2) is related to localization.

2.2 Range of translation parameters

We consider minimum phase Daubechies wavelets with compact support which ensure a finite range of values. The supports of the scaling function $\phi(x)$ and the wavelet function $\psi(x)$ are $[0, 2N-1]$ and $[-N, N-1]$, respectively, where N is the number of vanishing moments. We can easily calculate the supports of $\phi_{j(k)}(x)$, $\psi_{(j)k}(x)$ and the range of the translation parameter k for $\phi_{j(k)}(x)$ and $\psi_{(j)k}(x)$. The supports of $\phi_{j(k)}(x)$ and $\psi_{(j)k}(x)$ are

$$\left[\frac{k}{2^j}, \frac{2N-1+k}{2^j}\right] \text{ and } \left[\frac{1-N+k}{2^j}, \frac{N+k}{2^j}\right].$$

Thus, given $x \in [a, b]$, we can calculate the range of k based on the supports corresponding to the functions $\phi_{j(k)}(x)$ and $\psi_{(j)k}(x)$ intersecting $x \in [a, b]$,

$$[\lceil a2^j \rceil - 2N + 1, \lfloor b2^j \rfloor] \text{ and } [\lceil a2^j \rceil - N + 1, \lfloor b2^j \rfloor + N]$$

where $\lfloor x \rfloor = \max\{n \in \mathbf{Z}; n \le x\}$ and $\lceil x \rceil = \min\{n \in \mathbf{Z}; n \ge x\}$.

2.3 Bayesian approach to the choice of a maximum resolution

In this section we describe a Bayesian method based on the method proposed by Smith and Kohn (1997) for selecting a maximum resolution in a linear regression as it forms the basis of the wavelet approximation to the nonparametric function. For notational convenience, rewrite (2)

$$g(x|m) \approx \sum_{k \in Z} s_{0(k)}\phi_{0(k)}(x) + \sum_{j=0}^{m} \sum_{k \in Z} d_{j(k)}\psi_{j(k)}(x)H(m \ge 0) \qquad (3)$$

for $m = -1, 0, 1, \ldots, m_0$, where $H(\cdot)$ is an indicator function. Let M be a family of linear regression models all have the same dependent variables such that for $m \in M$

$$\underline{y} = \boldsymbol{W}_m \underline{\theta}_m + \underline{\epsilon} \tag{4}$$

where $\underline{y} = (y_1, \ldots, y_n)^T$ is an $n \times 1$ row vector of dependent variables, $\boldsymbol{W}_m$ is an $n \times q(m)$ design matrix of regression parameters, $\{\phi_{0(k)}\}$ and/or $\{\psi_{j(k)}\}$, $\theta_{-1} = \{s_{0(k)}\}$ or $\underline{\theta}_m = s_{0(k)}, d_{0(k)}, \ldots, d_{m(k)}$ is a $q(m) \times 1$ column vector, and the error vector $\underline{\epsilon} \sim N(\underline{0}, \sigma^2 \boldsymbol{I}_n)$. If, in variable selection, the design matrix consists of q independent variables, there are all 2^q possible subsets. In the problem of resolution selection, however, there are only $m_0 + 2$ possible subsets from (4).

Since the wavelet basis functions are orthornomal, we can make an assumption such that $\boldsymbol{I}_m = \boldsymbol{W}_m^T \boldsymbol{W}_m$. To choose a maximum resolution, we employ the following Bayesian procedure :

Step 1 . The prior distribution for $\underline{\theta}_m \mid \sigma^2, m$ is

$$\underline{\theta}_m \mid \sigma^2, m \sim N(\underline{0}, \sigma^2 \boldsymbol{I}_m) \ . \tag{5}$$

Step 2 . The prior distribution for $\sigma^2 \mid m$ is (Box and Tiao, 1973)

$$p(\sigma^2 \mid m) \ \propto \ \sigma^{-2} \ . \tag{6}$$

Step 3 . The noninformative prior distribution for m is

$$p(m) \ \propto \ \frac{1}{\#(M)} \tag{7}$$

where $\#(M)$ is the number of elements in the family M . From (5)-(7), and the likelihood function, the posterior probability of a model m is given by

$$\begin{aligned} p(m \mid \underline{y}) &\propto p(\underline{y} \mid m) p(m) \\ &\propto \int_{\sigma^2} \left[\int_{\underline{\theta}_m} p(\underline{y} \mid \underline{\theta}_m, \sigma^2, m)\, p(\underline{\theta}_m \mid \sigma^2, m)\, d\underline{\theta}_m \right] p(\sigma^2)\, d\sigma^2 \\ &\propto 2^{-q(m)/2} S(m)^{-n/2} \quad , \end{aligned} \tag{8}$$

where $S(m) = \underline{y}^T \underline{y} - \frac{1}{2} \underline{y}^T \boldsymbol{W}_m \boldsymbol{W}_m^T \underline{y}$.

2.4 The posterior mode for the choice of a maximum resolution

Since the number of models for resolutions is not large (less than 20), that is, a model with a maximum resolution m $(-1 \le m \le m_0)$ is including the submodel with the resolution $m-1$, use of the Gibbs sampler for the variable m provided by Smith and Kohn (1997) is not required. So from (8) we can simply calculate all posterior probabilities for the family M. We select a maximum resolution corresponding to a posterior mode from the posterior probabilities.

2.5 Estimation for a function

To estimate a function, we propose an MCMC algorithm. Given a maximum resolution, the priors of all parameters of interest are same as the above priors except for the prior of each coefficient which is a normal distribution with zero mean and variance τ. Proposed MCMC scheme is as follows:

1. Generate σ^2 from the complete inverse gamma conditional distribution.
2. Generate τ from the complete inverse gamma conditional distribution.
3. Generate s_{0k} from the complete normal conditional distribution.
4. Generate d_{jk} from the complete normal conditional distribution.

All Steps are Gibbs samplers and initial values are obtained by random values from each prior distribution.

3. Simulated Examples

We apply our Bayesian approach to a smooth function $g(x)=\cos(2.2\pi/3+8x)$ for $x \in [0,1]$ with two different sample sizes, 50 and 2,000 , and two different standard deviations, 0.02 and 0.5 in the simulation study. Also, to implement Monte Carlo simulation, we generate 1,000 dataset.

Simulation results are given in [Table 1], [Table 2], and [Figure 1]-[Figure 8].

4. Concluding Remarks

When running the MCMC algorithm to the cosine function of the above example given a maximum resolution, the estimated posterior functions of the wavelet series are approximated to the function well regardless of maximum resolutions, $m=0$ or 1 because the cosine function is a smoothing function. It is more efficient for selecting a resolution than for using a sample size as a maximum resolution in MCMC.

In this paper, however, the results of an irregular function with higher maximum resolution which are not presented were underestimated since the penalty part of the equation (8) for the number of coefficients, that is, as resolution increasing, the number of coefficients is increasing.

We conclude that the maximum resolution is affected by the shape of a function which will be estimated regardless of sample sizes.

References

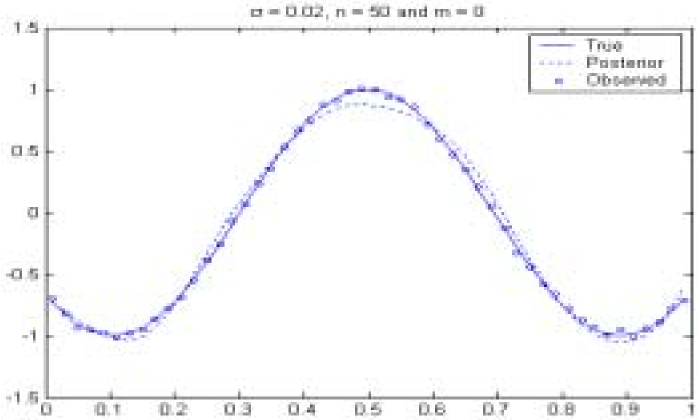
1. Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley.
2. Daubechies, I. (1988). *Ten Lectures on Wavelets*, CBMS-NSF, Series in Applied Mathematics, No. 61, Philadelphia: SIAM.
3. Müller, P. and Vidakovic, B. (1999). "MCMC Methods in Wavelet Shrinkage: Non-Equally Spaced Regression, Density and Spectral Density Estimation", in *Bayesian Inference in Wavelet Based Models*, eds. P. Müller and B. Vidakovic, New York: Springer-Verlag.
4. Smith, M. and Kohn, R. (1997). "A Bayesian Approach to Nonparametric Bivariate Regression", *Journal of the American Statistical Associate*, 92, 1522-1535.

[Table 1] For 1,000 dataset with $g(x_i) = \cos(2.2\pi/3 + 8x_i),\ i = 1, \dots, n$ where $x_i = (i - 0.5)/n$ and $\epsilon_i \sim iid\ \ N(0, \sigma^2)$.

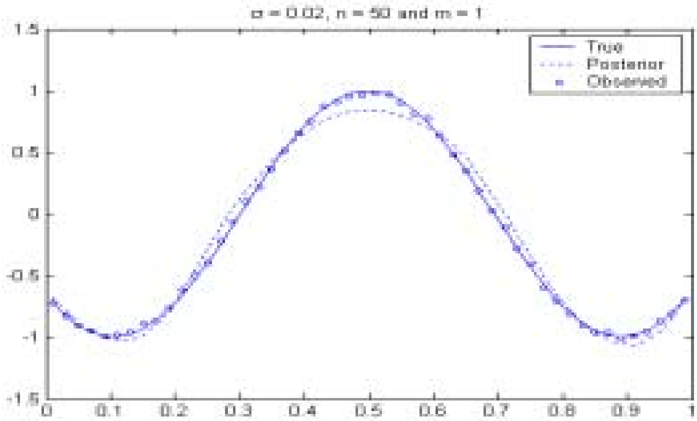
n	σ	Resolution (m)											
		-1	0	1	2	3	4	5	6	7	8	9	10
50	0.02	0	1000	0	0	0	0	0	0	0	0	0	0
	0.5	0	752	247	0	1	0	0	0	0	0	0	0
2000	0.02	0	1000	0	0	0	0	0	0	0	0	0	0
	0.5	0	1000	0	0	0	0	0	0	0	0	0	0

[Table 2] For one dataset estimation with $g(x_i) = \cos(2.2\pi/3 + 8x_i),\ i = 1, \dots, n$ where $x_i = (i - 0.5)/n$ and $\epsilon_i \sim iid\ \ N(0, \sigma^2)$.

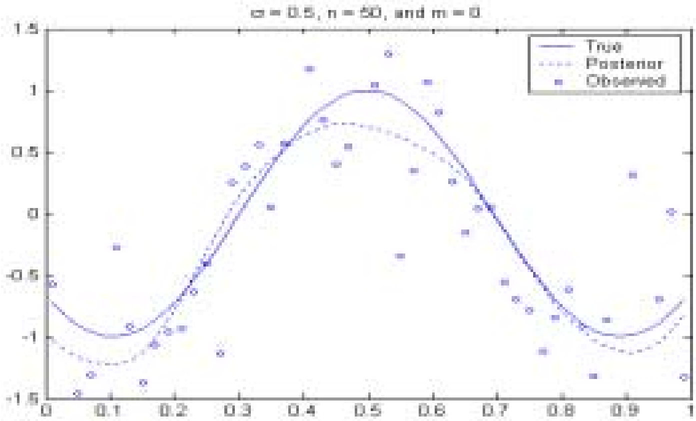
n	σ	mse $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	
		$m = 0$	$m = 1$
50	0.02	0.0050	0.0975
	0.5	0.1830	0.3515
2000	0.02	0.0054	0.0056
	0.5	0.2648	0.2750



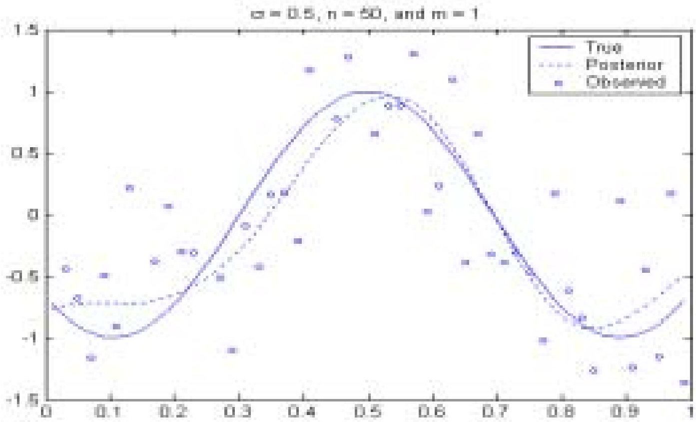
[Figure 1] Estimated posterior mean function : $m=0$, $n=50$, $\sigma=0.02$.



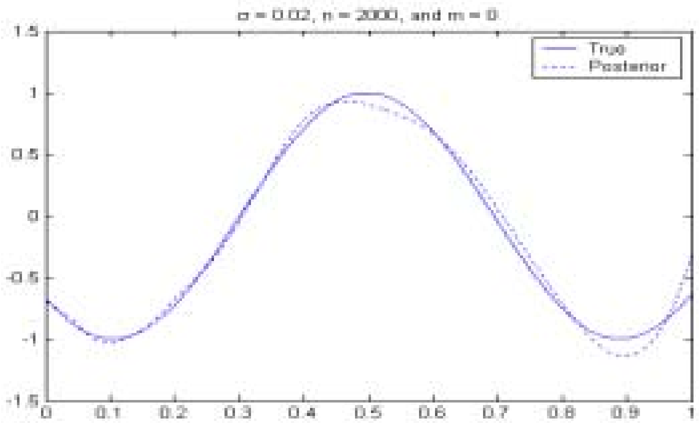
[Figure 2] Estimated posterior mean function : $m=1$, $n=50$, $\sigma=0.02$



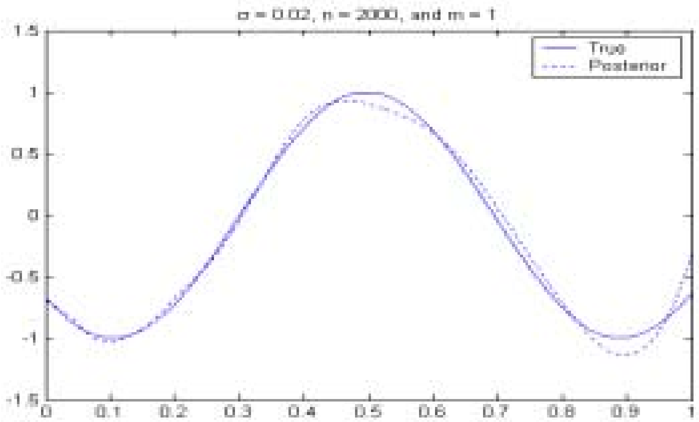
[Figure 3] Estimated posterior mean function : $m = 0$, $n = 50$, $\sigma = 0.5$



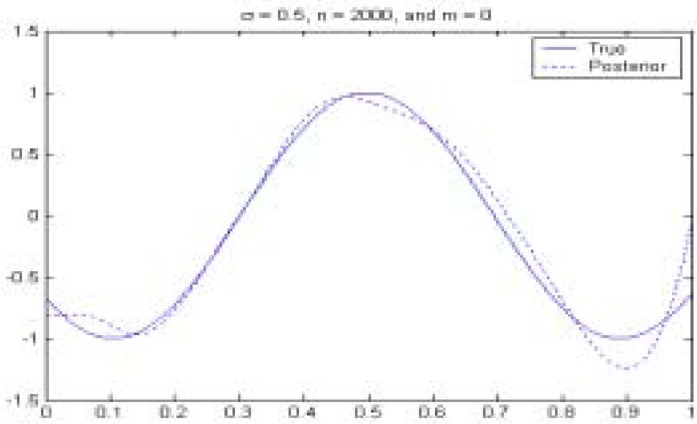
[Figure 4] Estimated posterior mean function : $m = 1$, $n = 50$, $\sigma = 0.5$



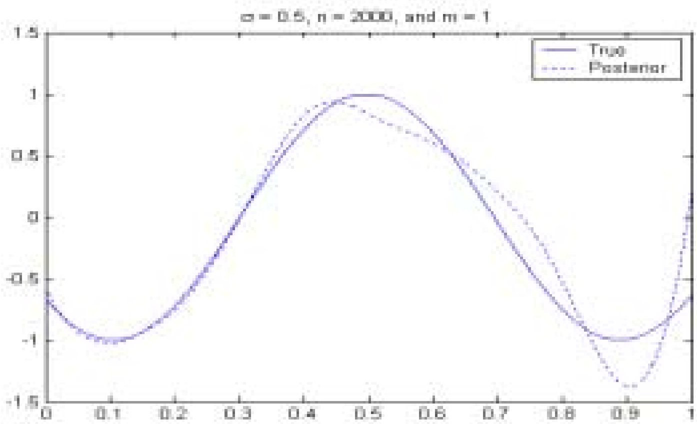
[Figure 5] Estimated posterior mean function : $m=0$, $n=2000$, $\sigma=0.02$



[Figure 6] Estimated posterior mean function : $m=1$, $n=2000$, $\sigma=0.02$



[Figure 7] Estimated posterior mean function : $m = 0$, $n = 2000$, $\sigma = 0.5$



[Figure 8] Estimated posterior mean function : $m = 1$, $n = 2000$, $\sigma = 0.5$

[received date : Dec. 2003, accepted date : Apr. 2004]