Journal of the Korean Data & Information Science Society
2009, **20**(6), 1191–1197

한국데이터정보과학회지

On loss functions for model selection in wavelet based Bayesian method

Chun Gun Park[1]

Department of Applied Statistics, Konkuk University

Received 18 July 2009, revised 10 November 2009, accepted 22 November 2009

Abstract

Most Bayesian approaches to model selection of wavelet analysis have drawbacks that computational cost is expensive to obtain accuracy for the fitted unknown function. To overcome the drawback, this article introduces loss functions which are criteria for level dependent threshold selection in wavelet based Bayesian methods with arbitrary size and regular design points. We demonstrate the utility of these criteria by four test functions and real data.

Keywords: Bayesian methods, loss functions, model selection, posterior expected loss.

1. Introduction

Over the last decade the Bayesian approach has been successfully applied to wavelet estimators and wavelet shrinkage problems in wavelet regression, see Antoniadis and Sapatinas (2001) for extensive review. So far from the Bayesian wavelet shrinkage and wavelet thresholding estimators the focus has been on the posterior estimators and the selection of wavelet coefficients simultaneously. Recently block wavelet shrinkage and thresholding estimators have been studied in finite sample situations on an empirical Bayes approach to block shrinkage of wavelet coefficients (Abramovich *et al.*, 2002; Wang *et al.*, 2006). All Bayesian approaches to wavelet problems should require numerical techniques to estimating hyperparameters or wavelet coefficients by term by term or blocks.

Park *et al.* (2004) first introduced a simple and efficient approach to obtain posterior probabilities of a primary resolution and model selection in wavelet regression. The advantages of the proposed model selection method is avoiding expensive computational cost. To choose the primary resolution and wavelet thresholding, however, Bayes factor employed as a criterion could be strongly ad-hoc. The aim of this paper is to investigate some criteria for the remedy of the selection problem in wavelet domain.

In this paper, we will explore loss functions which can play a critical role from Bayesian model selection in wavelet regression at each resolution level which Park (2008) introduced using Bayes factor.

The organization of the paper is as follows. Section 2 describes designing loss functions for model selection. In Section 3, the posterior expected loss will be suggested for the choice of wavelet basis functions at each resolution level. Section 4 gives the results of the simulation study and an application. The conclusions are addressed in Section 5.

[1] Lecture, Department of Applied Statistics, Konkuk University, Seoul 143-701, Korea.
E-mail: chungunpark@daum.net

2. Loss functions

Let θ be a parameter of interest. A loss function is denoted by $L(\theta,\ \theta')$ which depends on θ and θ'. In our opinion, the following loss functions will be suggested for wavelet based Bayesian model selection:
0-1 loss function is

$$L_1(\theta, \theta') = [1 - \delta(\theta, \theta')] \tag{2.1}$$

where $\delta(\theta, \theta') = 1$, if $\theta = \theta'$ and zero, if otherwise.
Mean squared error loss function is

$$L_2(\theta, \theta') = (\theta - \theta')^2. \tag{2.2}$$

Weighted 0-1 loss function is

$$L_3(\theta, \theta') = [1 - \delta(\theta, \theta')] \times w(\theta). \tag{2.3}$$

These loss functions usually satisfy the following property, $L(\theta,\ \theta') = 0$ when $\theta = \theta'$. In this paper the weighted 0-1 loss function is focused on.

3. Posterior expected loss for wavelet based Bayesian method

3.1. Wavelet series

The data are assumed to be of the form

$$y_i = f(i/n) + \sigma\epsilon_i, \quad i = 1, \ldots, n,$$

where the ϵ's are independent and normally distributed with mean 0 and variance 1.

An orthogonal wavelet basis in $L_2(\mathbb{R})$ is a collection of functions obtained as translations and dilations of a scaling function ϕ and a wavelet function ψ (Daubechies, 1992). The unknown function f may represent wavelet series:

$$\widehat{f}_m(x) = \sum_{k\in\mathbb{Z}} s_{0,k}\phi_{0,k}(x) + \sum_{j\geq 0}^{m_0}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x) = \sum_{m=-1}^{m_0}\sum_{k=1}^{N(m)} \beta_{(m,k)}b_{(m,k)}. \tag{3.1}$$

where m_0 is a maximum resolution level, $N(m)$ is the number of wavelet bases for a resolution level, m, and the value $m = -1$ means the resolution level, zero, of smoothing part. Futhermore, β's and b's. are denoted by the wavelet coefficients and the wavelet bases, respectively.

3.2. Posterior probability of model selection

From (2.1) - (2.3) and (3.1) let $B_m = (b_{-1}, b_0, b_1, \ldots, b_{m_0})$ be the matrix of wavelet basis functions, $\phi_{0,k}$ and $\psi_{j,k}$, $j = 0, 1, \ldots, m_0$ and $Y = (y_1, \ldots, y_n)^T$ be a response variable.

For notational convenience for model selection, W_m is the resulting matrix obtained by exchanging the columns of b_m according to the size of the elements of $b_m^T Y$ for each resolution level m: 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 $\sum_{i=1}^{n}\phi_{m,k}(x_i)y_i, \quad k=1,\ldots,N(m)$, 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 for each resolution level m. Furthermore, α_m denotes the coefficient vector obtained by rearranging the elements of β_m according to W_m.

From the above explanation, the model (2.1) can be rewritten as

$$Y = \sum_{m=-1}^{m_0} W_m\alpha_m + \sigma\epsilon, \tag{3.2}$$

where W_m is the collection of wavelet basis functions for a resolution level m and $\alpha_m = (\alpha_{(m,1)},\ldots,\alpha_{(m,N(m))})^T$ is the wavelet coefficients according to the order of the magnitudes b_m^TY.

Proposition 3.1 Assume that the error variance σ^2 and each resolution level m are given. Let the conditional prior of each wavelet coefficient be identically independent, improper and proportional to some constant and the prior of model selection, s, for each resolution level is equally likely chance. Then the posterior probability of model selection for each resolution level is

$$P(s\ |m,\sigma^2,Y) \propto \exp\left[\frac{1}{2\sigma^2}Y^TW_{m(s)}W_{m(s)}^TY\right], \quad 1\le s\le N(m), \tag{3.3}$$

where $W_{m(s)} = (W_{(m,1)}\cdots W_{(m,s)})$, $s=1,\ldots,N(m)$ and $m=-1,0,\ldots,m_0$. A detailed derivation is given in Appendix from Park *et al.* (2008).

3.3. Model selection with posterior expected loss

We use the loss functions and the posterior probability for the problem of model selection in wavelet based Bayesian methods. It is well known that we must choose a model s for each resolution level minimizing the posterior expected loss,

$$\arg\min(\widehat{s}) \quad \sum_{s=1}^{N(m)} L_i(s,\widehat{s})\cdot \Pr(s|m,\sigma^2,Y). \tag{3.4}$$

where the loss functions are defined by (2.1)-(2.3). From (2.3) the weighted values are on the penalty of the number of coefficients for each resolution level;

$$w(s) = (c_1 - l(s)) = (c_1 - (a\cdot s + d)). \tag{3.5}$$

From (3.5) if $l(s)=0$ for all s, the loss function is the 0-1 loss function. Then we set up $l(s)>0$ for all s, given $a>0$ and d. That means that as s increasing, $l(s)$ should be increasing. In the next section the explanation of the linear weighted values will be provided in details.

4. Simulation study and an application

4.1. Experiment setup

In this section, we report the results of simulations made for model selection based on our proposed posterior expected losses.

All the data of the simulations is of the form (2.1). The sample sizes selected were $n = 128$ and $n = 2048$. For each sample size, we used four test functions to assess model selection for wavelet based Bayesian method. The values of σ^2, variance of the noise, were chosen: $\sigma^2 = 0.05^2$.

The first test function is a Cosine function which is necessary to be some wavelet bases corresponding to a relatively lower resolution level. The second function is Heavisine function which may be represented with more the number of wavelet bases and resolution level then the Cosine function. The remaining functions are the Doppler function and the Bump function which might need a higher resolution level. All of the test functions might be not necessary all wavelet basis functions under a given resolution level. To select which wavelet coefficients, the posterior expected loss should be used.

4.2. Computation and simulation results

The motivation of the linear weighted equation is the distance from the posterior probabilities and some linear equation at each resolution level. We want to select some position such that less number of wavelet coefficients than that corresponding to maximum of the posterior probabilities at each resolution level. The posterior probability corresponding to choosing the position could be closer than maximum of the posterior probabilities at each resolution level. Now the values of $w(s)$ are determined for the resolution levels $m = -1, 0, \ldots, m_0$ by the following;

$$a = \frac{\Pr(s = N(m)|m, \sigma^2, Y) - \Pr(s = 1|m, \sigma^2, Y)}{N(m) - 1}$$
$$d = -a \cdot N(m) + \Pr(s = N(m)|m, \sigma^2, Y).$$

From (3.4) and the above weighted loss function we can select some wavelet basis functions at each resolution level and estimate wavelet coefficients corresponding them using MCMC proposed by Park *et al.* (2005).

To report the results of the simulations, we used the MSE as the numerical measure;

$$\text{MSE}(\widehat{f}) = \frac{1}{n}\sum_{i=1}^{n}\left[f(t_i) - \widehat{f}(t_i)\right]^2.$$

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

4.3. A real example

This section illustrates the proposed method using a real example. An electrical consumption example presented here involves a real-world signal electrical consumption measured

Table 4.1 Simulation results in term of integrated MSE for $\sigma = 0.05$

Sample size	Shrinkage methods	Test functions			
		Cosine	*Doppler*	*HeaviSine*	*Bumps*
$n = 128$	Loss & MCMC	5.9072e-4	3.4492e-003	1.1852e-003	1.5513e-003
	BAM	1.3924e-3	1.5831e-003	1.4139e-003	1.5058e-003
$n = 2,048$	Loss & MCMC	6.6327e-5	4.3541e-004	1.8660e-004	4.9589e-004
	BAM	1.4040e-4	2.7438e-004	1.9214e-004	3.6071e-004

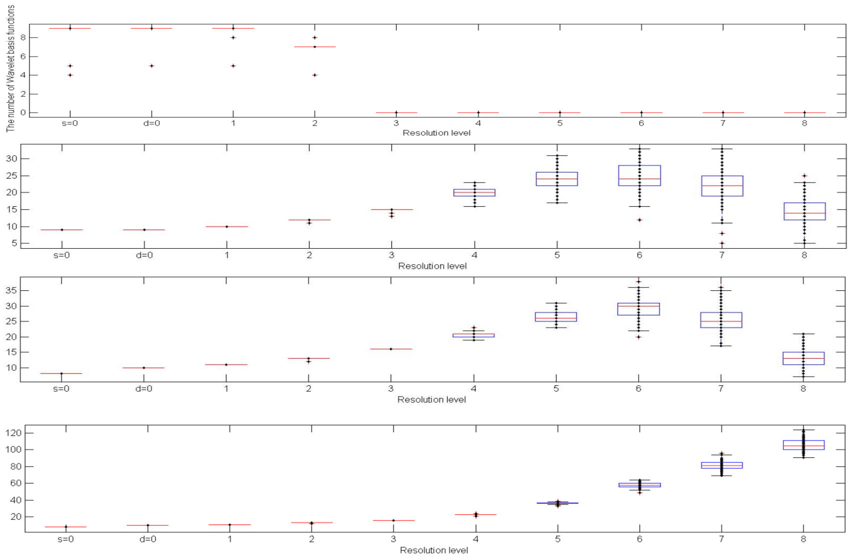

Figure 4.1 The number of wavelet basis functions at each resolution level for $n = 2048$ and from the top the listed functions are Cosine, Heavisine, Doppler, and Bump functions

minute by minute over the course of three days. The detailed information can be found in Misiti, *et al.* (1994).

In this real example, Daubeachies's wavelet with vanishing moment 8 has been used. Applying the proposed method to the data, we choose some wavelet bases at each resolution level. From Figure 4.3, the nature of wavelet regression, that is, the sparseness, is well shown.

5. Concluding remarks

For the problem of selecting wavelet basis functions, the wavelet based Bayesian method introduced by Park *et al.* (2008) did not consider wavelet selection at each resolution level and Park (2008) first introduce the model selection at each resolution level using Bayes factor which is very strongly ad-hoc.

In this paper we propose loss function which is used to select wavelet basis functions at each resolution level. The estimation of wavelet coefficients is using MCMC (Park *et al.*,

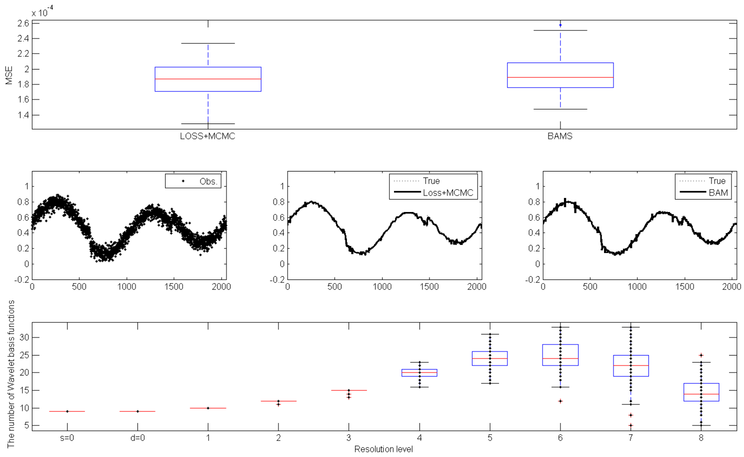

Figure 4.2 Some results for the Heavisine with $n = 2048$.

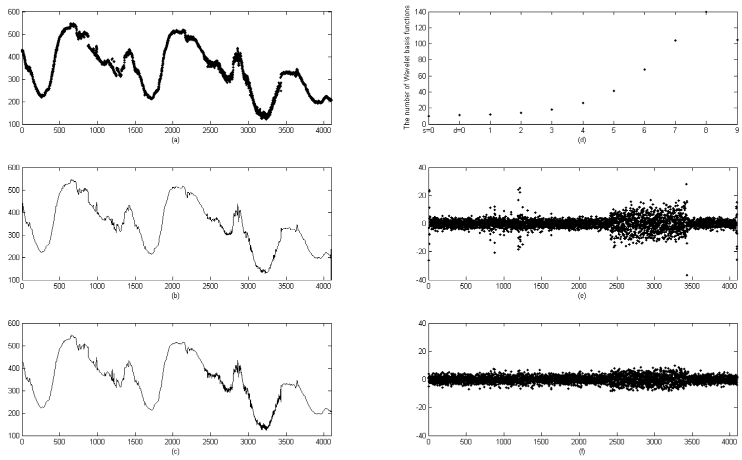

Figure 4.3 The electrical consumption. (a) observed data, (b) fitting data for Loss & MCMC (c) fitting data for BAM (d) the number of wavelet basis functions for Loss & MCMC (e) residual plots for Loss & MCMC (f) residual plots for Loss & MCMC.

2005). The advantage of the proposed method is reducing computational cost and easy to figure out which wavelet basis functions being significant at each resolution level and less ad-hoc than Bayes factor. We can make more accurate the estimation of the coefficients corresponding to the given wavelet basis functions.

Further research on model selection and the estimation of the wavelet coefficients is necessary to explore loss functions for the choice of wavelet basis functions at each resolution level and the priors of wavelet series which make the fitted function more accurate.

References

Abramovich, F., Besbeas, P. and Sapatinas, T. (2002). Empirical Bayes approach to block wavelet function estimation. *Computational Statistics and Data Analysis*, **39**, 435-451.

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

Daubechies, I. (1992). *Ten lectures on wavelets, CBMS-NSF*, Series in Applied Mathematics, No. 61, SIAM, Philadelphia.

Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J. M. (1994). *Wavelet toolbox for use with MATLAB*, The Math Works Incorporation.

Park, C. G. (2008). A note on a Bayesian approach to the choice of wavelet basis functions at each resolution level. *Journal of Korean Data & Information Science Society*, **19**, 1465-1476.

Park, C. G., Hart, Jeffrey D., Vannucci, M. (2005). Bayesian methods for wavelet series in single-index models. *Journal of Computational and Graphical Statistics*, **14**, 770-794.

Park, C. G., Kim, Y. H. and Yang, W. Y. (2004). Determinacy on a maximum resolution in wavelet series. *Journal of Korean Data & Information Science Society*, **15**, 467-476.

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

Wang, X. and Wood, Andrew, T. A. (2006). Empirical Bayes block shrinkage of wavelet coefficients via the noncentral χ^2 distribution. *Biometrika*, **93**, 705-722.