The Korean Communications
in Statistics Vol. 11 No. 1, 2004
pp. 161-168

Posterior Inference in Single-Index Models

Chun Gun Park[1], Wan-Yeon Yang[2], Yeong-Hwa Kim[3]

Abstract

A single-index model is useful in fields which employ multidimensional regression models. Many methods have been developed in parametric and nonparametric approaches. In this paper, posterior inference is considered and a wavelet series is thought of as a function approximated to a true function in the single-index model.

The posterior inference needs a prior distribution for each parameter estimated. A prior distribution of each coefficient of the wavelet series is proposed as a hierarchical distribution. A direction β is assumed with a unit vector and affects estimate of the true function. Because of the constraint of the direction, a transformation, a spherical polar coordinate θ, of the direction is required. Since the posterior distribution of the direction is unknown, we apply a Metropolis-Hastings algorithm to generate random samples of the direction. Through a Monte Carlo simulation we investigate estimates of the true function and the direction.

Keywords : Single-index model, Wavelet series, Daubechies wavelet, Posterior inference, Hierarchical distribution, Metropolis-Hastings algorithm

1. Introduction

A single-index model is useful in a variety of fields such as econometrics, biometrics, and etc. with nonlinear regression models involving linear regression models (McCullagh and Nelder, 1983) and is defined as the form

$$y_i = g(X^t\beta) + \epsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

1) Researcher, Department of Statistics, Seoul National University, Seoul, 151-742, Korea
E-mail : txcgpark@snu.ac.kr

2) Associate Professor, Department of Applied Statistics, Kyungwon University, Sungnam, Korea
E-mail : ywy@kyungwon.ac.kr

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

where $X^i = (x_{i1},...,x_{ip})$, $\beta = (\beta_1,...,\beta_p)^T$ with $||\beta|| = 1$, g is an unknown function and ϵ_i is a random variable with zero mean and bounded variance.

This model (1) has been mainly analysed in parametric and nonparametric views. That is, kernel smoothing(Hardle, *et al.*, 1993), local linear methods (Carroll, *et al.*, 1997), average derivatives (Stoker 1986), and penalized splines (Yu and Ruppert, 2002) had been introduced. In these methods the assumption of the function g is a smoothing function.

To analyse the model (1) we consider the posterior inference which requires prior distributions of parameters estimated. When a wavelet series is approximated to a function from (1), we can assume that the function is any function in L^2 space[4]. A prior distribution of each coefficient of the wavelet series is defined as a hierarchical distribution.

The wavelet series is constructed based on two basis function, a scaling function $\phi(z)$ and a wavelet function $\psi(z)$ where z's must be equally spaced. From (1), however, $X^i\beta$'s shouldn't be equally spaced for all directions. To overcome this problem is to replace the unequally spaced values by equally spaced values and proceed as in the equally spaced case.

Because of the constraint of the direction, $||\beta|| = 1$, to implement the posterior inference the transformation, a sphere polar coordinate, of the direction should be necessary. There are two advantages for the transformation that one of them is reducing dimension and the other independency for all elements of the polar coordinate. The posterior distribution of the direction is not integrable and we propose a Metropolis-Hastings which require a proposal distribution. A normal distribution is thought of as the proposal distribution for the direction. Its variance and initial coefficients of the wavelet series may need to be chosen carefully.

Section 2 describes wavelet-based posterior inference for the single-index model. Section 3 presents a simulation study. Finally, Section 4 discusses the results.

2. Wavelet-based posterior inference

2.1 Wavelet series

A wavelet series is an good approximation to a square integrable function. We can represent the function of the single-index model as the wavelet series which is

$$g(X^i\beta) \approx \sum_{k \in Z} s_{J_0k}\phi_{J_0k}(X^i\beta) + \sum_{j=J_0}^{m}\sum_{k \in Z} d_{jk}\psi_{jk}(X^i\beta) \tag{2}$$

for any integer J_0 and an integer m given. Since all functions, ϕ's and ψ's, are orthogonal basis functions, in the literature all coefficients are defined as follows

4) A function r is in the L^2 space if r is square integrable, that is,

$$||r|| = \left(\int |r(x)|^2 dx\right)^{1/2} < \infty.$$

$$s_{J_0k} = \int g(z)\phi_{J_0k}(z)dz \text{ and } d_{jk} = \int g(z)\psi_{jk}(z)dz.$$

The functions, ϕ and ψ, use a dilation j and a translation k, that is,

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

From the approximation (2) the assumption of the resolution J_0 is zero (Müller and Vidakovic, 1999).

2.2 Likelihood function and prior distributions

To implement the posterior inference a likelihood function and prior distributions of parameters estimated are required. From (1) the errors ϵ's are independent and follow a normal distribution with mean zero and bounded variance σ^2. From (1) and (2) let $\Omega = (s_{0k}, d_{jk}, \beta, \sigma^2)$ and the likelihood function is

$$\begin{aligned} P(Y\,|\Omega, \text{ values given}) &= \prod_{i=1}^{n} P(y_i|\Omega, \text{ values given}) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} R_i^2(y_i)\right) \end{aligned} \tag{3}$$

where

$$R_i(y_i) = y_i - \sum_{k \in Z} s_{0k}\phi_{0k}(X^i\beta) - \sum_{j=0}^{m}\sum_{k \in Z} d_{jk}\psi_{jk}(X^i\beta).$$

Since the direction parameter is the unit vector and it is difficult to determine a sign, positive or negative, of the last element of the direction parameter in the posterior inference, we propose a spherical polar coordinate $\gamma = (\gamma_1, ..., \gamma_{p-1})$ for the unit vector which makes the posterior inference easier. The nature of the polar coordinate is independency for all elements of the polar coordinate and reducing dimension to $p-1$. This transformation is as follows

$$t_a(\gamma) = \beta_a = \sin(\gamma_a)\prod_{b=0}^{p-a}\cos(\gamma_{p-a}),\ a = 1, ..., p, \tag{4}$$

for $\sin(\gamma_0) = \cos(\gamma_p) = 1$, $0 < \theta_1 < 2\pi$, and $-\frac{\pi}{2} < \theta_l < \frac{\pi}{2}$, $l = 2, ..., p-1$. Based on the polar coordinate, that is, $X^i\beta = X^i T(\gamma)$ where $\beta = T(\gamma) = (t_1(\gamma), ..., t_p(\gamma))^T$, we can rewrite the likelihood function

$$P(Y \mid \Omega, \text{ values given}) = P(y_i \mid \Omega^*, \text{ values given}) \tag{5}$$

where $\Omega = (s_{0k}, d_{jk}, \beta, \sigma^2)$ and $\Omega^* = (s_{0k}, d_{jk}, \gamma, \sigma^2)$.

For notational simplicity, let '$\sim$' mean 'distributed as', let $N(\mu, \eta)$ denote a normal distribution with mean μ and variance η, let $IG(\alpha, \zeta)$ denote an inverse gamma distribution, and let $P(\cdot)$ generically denote a probability density function. We propose a prior distribution for each coefficient from (2) as a normal distribution which have mean zero and variance depending on its resolution level and a hyperparameter τ (Müller and Vidakovic, 1999)

$$s_{0k} | \tau \sim N(0, \tau) \text{ and } d_{jk} | \tau \sim N(0, \tau 2^{-1}). \tag{6}$$

The prior distributions of τ and σ^2 are as follows

$$\tau \sim IG(\alpha_\tau, \lambda_\tau) \text{ and } \sigma^2 \sim IG(\alpha_v, \lambda_v). \tag{7}$$

We define a prior distribution for the direction parameter γ from (4) as a joint uniform distribution

$$P(\gamma) = \frac{1}{2\pi}\left(\frac{1}{\pi}\right)^{p-2}, \tag{8}$$

where γ is $(p-1) \times 1$ and $0 < \gamma_1 < 2\pi$, $-\frac{\pi}{2} < \gamma_a < \frac{\pi}{2}$, $a = 2, \ldots, p-1$.

2.3 MCMC methods

Before setting a MCMC scheme, we assume that Daubechies wavelets are used and the boundaries of the translations k of the coefficients s_{0k}, d_{jk} are $[L_0, U_0]$ and $[L_j, U_j]$, $j = 0, 1, \ldots, m$ (Park, 2003).

From (3) and (6)-(8) the joint distribution of the parameter space Ω^* is calculated. Given initial values for the parameter space Ω^*, each parameters is updated from each posterior distribution, one at time. A 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.
5. Generate γ from a normal proposal distribution with mean γ^*, a previous direction, and variance $\tilde{\nu}$ and compute the acceptance probability

$$\alpha(\gamma,\gamma^*) = \min\left(1, \frac{P(\gamma|w(-\gamma))}{P(\gamma^*|w(-\gamma^*))}\right).$$

Steps 1-4 are Gibbs samplers and Step 5 is Metropolis-Hastings (Metropolis, *et al.*, 1953). The choices of $\tilde{\nu}$ and initial values are described in Subsection 2.4.

2.4 Initialization

Before initializing all parameters except for the variance parameter σ^2, we need maximum range of $X\beta$. When Daubechies wavelets is applied to representing a wavelet series, we can calculate the ranges of the translation k for $\phi_{0k}(z)$ and $\psi_{jk}(z)$ based on the supports of the Daubechies wavelets and the maximum range of $X\beta$. (Park, 2003)

Since the direction parameter is affecting estimates of the coefficients of the wavelet series, it is important to obtain good initial values the direction parameter. Here we implement a preliminary step with 1000 iterations for several starting directions for choosing a good initial direction, that is, we obtain a direction from these directions by minimizing the residual sums of squares calculated based on the estimated mean function through 1000 iterations corresponding to each starting direction. For the initialization of the coefficients, we use the orthogonal property of the wavelet function as follows (Hart, 1997)

$$\begin{aligned}\hat{s}_{0k} &= \sum_{i=1}^{n} y_i (w_i - w_{i-1})\phi_{0k}\left(\frac{w_i + w_{i-1}}{2}\right) \\ &\approx \sum_{i=1}^{n} y_i \int_{w_{i-1}}^{w_i} \phi_{0,k}(z)dz\end{aligned}$$

and

$$\begin{aligned}\hat{d}_{jk} &= \sum_{i=1}^{n} y_i (w_i - w_{i-1})\psi_{jk}\left(\frac{w_i + w_{i-1}}{2}\right) \\ &\approx \sum_{i=1}^{n} y_i \int_{w_{i-1}}^{w_i} \psi_{0,k}(z)dz\end{aligned}$$

where $z_{(i)}$ is the ith smallest of the ordered $X\beta$, $w_0 = z_{(1)}$, $w_{(i)} = \frac{z(i)+z(i+1)}{2}$, $i = 1, ..., n-1$, $w_{(n)} = z_{(n)}$. The initial value of τ is generated from its prior. An initial choice of a variance of the normal proposal distribution for the direction parameter γ is

$$\hat{\sigma}_\gamma^2 = \frac{C}{n-1}\sum_{i=1}^{n}(y_i - \hat{g}_i)^2,$$

where C is arbitrary and $\hat{g}_i$ is the wavelet approximation for a starting direction. In practice, given an initial direction, the variance $\hat{\sigma}_\gamma^2$ of the MCMC is determined until a variance $\hat{\sigma}_\gamma^2$ is found that yields an acceptance rate around 60% for the Metropolis-Hastings steps.

The hyperparameters of τ and σ^2 are $\alpha_\gamma = \alpha_v = 1$ and $\lambda_\tau = \lambda_v = 1$ so that the mean of the prior of τ or σ^2 is infinite.

3. Simulation study

We estimate a cosine function which is approximated to only the smooth part in the wavelet series (1). To implement posterior inferences, we generate 20 datasets and based on Section (2.4) we repeat the MCMC scheme twice with the first 1000 iterations to initialize the direction parameter θ and the second 10000 iterations including a burn-in periods (1000 or 2000). The estimates of parameters of interest are posterior means calculated from their samples.

(Example) The number of observations is $n = 200$ and the direction parameter is 2×1. The function is $g(z_i) = \cos(2\pi z_i/5)$ where $z_i = X^t\beta$ and each covariate is independently generated from normal distribution with mean zero and variance 1.5^2 or 3^2, respectively. Table 1 shows the result and Figures 1 show the estimates of the mean function.

4. Conclusion

Since the wavelet series has advantage of localization, we can make weaker the assumption of the function. To estimate parameters of interest, posterior inferences are proposed. The transformation of the direction parameter which is proposed as polar coordinates is very useful to generate samples from the posterior distribution of the direction parameter. The simulation study which treat three different directions for a smoothing function shows that for the direction a variance of a normal proposal distribution is carefully selected.

For a maximum resolution m fixed all wavelet coefficients are estimated. In further study the selection of a maximum resolution and shrinkage or thesholding rules are of interest in the single-index model.

References

[1] Carroll, R. J., Fan, J., Gijbels, I., and Wand, M.P. (1997). "Generalized Partially Linear Single-Index Models", *Journal of the American Statistical Association,* 92,477-489.

[2] Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests,* New York: Springer-Verlag.

[3] McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models,* London: Chapman and Hall.

[4] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). "Equations of State Calculations by Fast Computing Machines", *Journal of Chemical Physics,* 21, 1087-1091.

[5] Park, C. G. (2003). MCMC Methods for Wavelet Representations in Single Index Models, Ph.D. dissertation, Dept. Statistics, Texas A&M University, in Appendix.

[6] Stoker, T. M. (1986). "Consistent Estimation of Scaled Coefficients" *Econometrica,* 54, 1461-1481.

[7] Yu, Y. and Ruppert, D. (2002). "Penalized Spline Estimation for Partially Linear Single-Index Models", *Journal of the American Statistical Association,* 97, 1042-1054.

[8] Müller, P. and Vidakovic, B. (1999). "MCMC Methods inWavelet 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.

[Received November 2003, Accepted April 2004]

Table 1. Simulation results for $y_i = \cos(2\pi z_i/5) + \epsilon_i,\ i = 1, \ldots, n,$ where $z_i = X^t T(\gamma)$ and $\epsilon_i \sim_{iid} N(0, \sigma^2)$.

True values (σ, γ)		direction γ		function $g(z) = \cos(z)$	
		bias*	mse**	bias***	mse****
0.2	0.3	$2.58e-3$	$6.66e-6$	$-2.26e-3$	$2.05e-3$
	2.5	$8.22e-4$	$6.75e-7$	$-2.80e-3$	$1.85e-3$
0.5	0.3	$9.51e-4$	$9.04e-7$	$-1.11e-2$	$8.33e-3$
	2.5	$-1.66e-3$	$2.77e-6$	$-1.27e-2$	$8.11e-3$

* bias= $\frac{1}{100}\sum_{j=1}^{100}(\gamma - \hat{\gamma}_j)$ where $\hat{\gamma}_j$ is estimated for each dataset.

** mse= $\frac{1}{100}\sum_{j=1}^{100}(\gamma - \hat{\gamma}_j)^2$

*** bias= $\frac{1}{100}\frac{1}{n}\sum_{j=1}^{100}\sum_{i=1}^{n}(g(z_i) - \hat{g}_j(z_i))$

where $\hat{g}_j(z_i)$ is estimated for each dataset.

**** mse= $\frac{1}{100}\frac{1}{n}\sum_{j=1}^{100}\sum_{i=1}^{n}(g(z_i) - \hat{g}_j(z_i))^2$

Figure 1 For the true direction $\gamma = 0.3$ and $\sigma = 0.2$ and 0.5, posterior estimated mean function with a true function (solid) and a estimated function (dotted line).