Statistics and Its Interface Volume 2 (2009) 71–81

Bayesian adaptive nonparametric M-regression

Colin Chen[*]

Nonparametric regression has been popularly used in curve fitting, signal denosing, and image processing. In such applications, the underlying functions (or signals) may vary irregularly, and it is very common that data are contaminated with outliers. Adaptive and robust techniques are needed to extract clean and accurate information. In this paper, we develop adaptive nonparametric M-regression with a Bayesian approach. This general approach fits M-regression using piecewise polynomial functions with an unknown number of knots at unknown locations, all treated as parameters to be inferred through Reversible Jump Markov Chain Monte Carlo (RJMCMC) of Green (1995, [9]). The Bayesian solution presented in this paper with computational details can be considered as an approximation to the general optimal solution for M-regression with free knots as described in Stone (2005, [22]). Numerical results show that the Bayesian approach performs well in various cases, especially with discontinuous underlying functions.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35.

Keywords and phrases: Least informative distribution, Monotone score function, M-regression, Noninformative prior, Reversible jump Markov chain Monte Carlo, Splines.

1. INTRODUCTION

In robust regression, M-regression is the regression method based on M-estimates. It has been popularly used to provide resistant results in the presence of outliers in linear regression (Huber 1981, [15]; Maronna et al. 2005, [18]). Very often, the linear relationship between the response variable and covariates is not enough and a nonparametric relationship is assumed to allow more flexibility. Classical nonparametric regression methods are usually not immune to outlier contamination. Nonparametric regression based on M-estimates has been developed by using either kernel functions (Härdle and Gasser 1984, [12]; Hall and Jones 1990, [10]) or spline functions (He and Shi 1994, [13]; Shi and Li 1995, [20]; Gao and Shi 2001, [8]). However, as in the classical least squares based nonparametric regression, bandwidth selection with the kernel method and smoothing parameter (or knot) choice with the spline method are usually difficult to achieve certain optimality in practice.

For the spline method, Stone (2005, [22]) studied the optimal knot selection problem with M-regression. Under some regularity conditions, Stone (2005, [22]) proved that a set of optimal knots exist and the corresponding spline estimate converges to the underlying function. Although this result is purely theoretical and does not provide how to find the optimal knots, the existence of optimal knots does encourage using adaptive knots to approximate the optimal knots.

In this paper, we develop adaptive nonparametric M-regression with a Bayesian method. The Bayesian method applies the reversible jump Markov chain Monte Carlo approach of Green (1995, [9]), which has the ability to travel across function spaces with different dimensions. This flexibility in addition to the free choice of a prior empowers the RJMCMC to fit various smoothing or non-smoothing functions with high accuracy as shown in the least squares case by Denison et al. (1998, [4]), DiMatteo et al. (2001, [6]), Hansen and Kooperberg (2002, [11]), and others. The method is adaptive in the sense that knots are adaptively selected according to some properly designed moving strategies in the fitting procedure based on data.

To fit Bayesian M-regression models, we use a well-defined likelihood function in robust statistics – the least informative distribution with both location (μ) and scale (σ) parameters

$$h(y|\mu,\sigma) = \frac{1}{\sigma C_H} e^{-\rho_H(\frac{y-\mu}{\sigma})}, \tag{1}$$

where

$$\rho_H(r) = \begin{cases} r^2/2 & \text{if } |r| \le H \\ H|r| - H^2/2 & \text{if } |r| > H \end{cases} \tag{2}$$

is the well-known Huber function with the tuning constant H and $C_H = \int_{-\infty}^{+\infty} e^{-\rho_H(r)} dr$. The least informative distribution was derived by minimizing the fisher information in the ϵ-contaminated neighborhood of the Gaussian distribution (Huber, 1981, Section 4.5, [16]). With this likelihood, we show that the Bayes factor is equivalent to a robust version of the Schwarz information criterion (RSIC) (Section 3).

Bayesian model averaging has been used for robust nonparametric regression. Peña and Redondas (2006, [19]) developed Bayesian local polynomial regression with random orders. By using mixtures of normal distributions, the Bayesian local polynomial regression method achieves certain robustness against outliers. The use of random order

[*] Currently with Fannie Mae, 3900 Wisconsin Avenue NW, Washington, DC 20016. Phone: 202-752-6570.

might relieve the burden of selecting an optimal length of the local widow. However, as with classical local polynomial regression, the Bayesian local polynomial regression method might miss the global trend of the underlying functions by picking up too much suspicious local features. It also performs poorly when edge-preserving fitting is required for discontinuous underlying functions (Chu et al. 1998, [3]).

RJMCMC has been implemented on least squares based nonparametric regression by Denison et al. (1998, [4]) and others.

The current work extends the RJMCMC implementation to general M-regression, which more focuses on robustness. We also implement the data dependent tuning constant (H) selection technique of Wang et al. (2007, [23]) to make our Bayesian robust nonparametric regression a more automatic procedure.

In Section 2 we introduce Bayesian nonparametric M-regression with the least informative likelihood and piecewise polynomial functions. In Section 3 we discuss the choice of priors and Bayesian factor approximation, which decide the knot-selection rule. Section 4 describes the moving strategies for RJMCMC. The complete algorithm is described in Section 5. Extensions to other ρ functions in M-regression and data dependent tuning constant selection are discussed in Section 6. Section 7 illustrates our approach with some examples. We also provide performance comparison and robustness analysis with numerical results in Section 7. Some discussions are given in Section 8.

2. BAYESIAN M-REGRESSION

Assume that (y_i, x_i), $i = 1, \ldots, n$, are independent bivariate observations from the pair of response-explanatory variables (Y, X). To describe the relationship between Y and X, a typical parametric model is

$$y_i = m(x_i, \beta) + \epsilon_i, \tag{3}$$

where β is a vector of unknown parameters, and ϵ_i, $i = 1, \ldots, n$, are i.i.d. zero-mean errors. The least squares estimator of β ($\hat{\beta}_{LS}$) minimizes the objective function

$$\sum_{i=1}^{n} (y_i - m(x_i, \beta))^2. \tag{4}$$

This estimator can be heavily affected even if a single observed y_i is contaminated with an extreme value. To obtain a resistant estimator (M-estimator $\hat{\beta}_M$), M-regression minimizes

$$\sum_{i=1}^{n} \rho_{\sigma H}(y_i - m(x_i, \beta)), \tag{5}$$

where $\rho_{\sigma H}(\cdot)$ is the Huber function in (2) with H replaced by σH. The Huber function replaces the square function with a less rapidly increasing function – the absolute value function when absolute standardized residuals $|\frac{y_i - m(x_i, \beta)}{\sigma}|$ exceed the tuning constant H. The robustness is obtained due to less contribution to (5) from outliers with large absolute standardized residuals.

It is well-known that the least squares estimator $\hat{\beta}_{LS}$ is also the maximum likelihood estimator if ϵ_i, $i = 1, \ldots, n$, follow the normal distribution with density $\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{r^2}{2\sigma^2}}$. Likewise, the M-estimator $\hat{\beta}_M$ is also the maximum likelihood estimator if ϵ_i, $i = 1, \ldots, n$, follow the least informative distribution with density in (1) and $\mu = 0$. Based on this distribution, the Bayesian M-regression assumes that the parameters β and σ are random with some priors and provides posterior inference on these parameters or functions of these parameters.

As in typical nonparametric regression, the format of the function $m(x, \beta)$ is not known and assumed to be a nonparametric function $m(x)$, which belongs to the closure of a linear function space – here the piecewise polynomials

$$P_{k,l}(x) = \sum_{v=0}^{l} \beta_{v,0}(x - t_0)_+^v + \sum_{m=1}^{k} \sum_{v=l_0}^{l} \beta_{v,m}(x - t_m)_+^v, \tag{6}$$

where $(\cdot)_+$ and $(\cdot)_-$ represent the positive and negative part of the quantity, respectively. $t_i, i = 0, \ldots, k+1$, indexed in ascending order, are the knot points with the boundary knots $t_0 = \min\{x_i, i = 1, \ldots, n\}$ and $t_{k+1} = \max\{x_i, i = 1, \ldots, n\}$. Without loss of generality, we assume that x_i, $i = 1, \ldots, n$, are in ascending order. So, $t_0 = x_1$ and $t_{k+1} = x_n$. $l(\geq 0)$ is the order of the piecewise polynomials, and $l_0(\geq 0)$ controls the degree of continuity at the knots. Piecewise polynomials include splines. The special case with $l = l_0 = 3$ corresponds to the cubic splines.

Piecewise polynomials $P_{k,l}$ have been used by Denison et al. (1998, [4]) for least squares based curve fitting. In this paper, we assume that $m(x)$ is such a piecewise polynomial with l and l_0 pre-decided, while the coefficients $\beta = \{\beta_{v,m}, 0 \leq v \leq l, 1 \leq m \leq k\}$, the number of knots k, and their locations t_i are estimated with the data (y_i, x_i) by RJMCMC. Let $t = \{t_m, 1 \leq m \leq k\}$. (k, t, β, σ) represents the full vector of parameters in the model. The following section describes the prior specification with (k, t, β, σ).

3. PRIOR SPECIFICATION AND BAYESIAN FACTOR APPROXIMATION

We specify a prior for parameters (k, t, β, σ) hierarchically,

$$\pi_{k,t,\beta,\sigma}(k, t, \beta, \sigma) = \pi_{k,t}(k, t)\pi_\beta(\beta|k, t, \sigma)\pi_\sigma(\sigma). \tag{7}$$

First, we specify a prior for the model space, which is characterized by the first two parameters k and t. Then, we specify a prior for β. For the scale parameter σ, we use the noninformative prior $\pi_\sigma(\sigma) = 1$.

For the model space, a prior $\pi_{k,t}(k,t)$ can be further decomposed as

$$\pi_{k,t}(k,t) = \pi_k(k)\pi_t(t|k). \tag{8}$$

So, we first need to specify a prior $\pi_k(k)$ for the number of knots k. We explore several proposals in the literature. The first one is a simple Poisson distribution with mean γ suggested by Green (1995, [9]), and also used by Denison et al. (1998, [4]) and DiMatteo et al. (2001, [6]). The second one is the uniform prior on the set $k_{\min}, \ldots, k_{\max}$, which has been used by Smith and Kohn (1996, [21]) and Hansen and Kooperberg (2002, [11]). The third one is the geometric prior $\pi_k(k) \propto \exp(-ck)$ proposed by Hansen and Kooperberg (2002, [11]) based on the model selection criterion SIC.

With the number of knots k specified, the sequence of knots t_i, $i = 1, \ldots, k$, are considered order statistics from the uniform distribution with candidate knot sites $\{x_1, \ldots, x_n\}$ as the state space. We also consider the candidate knots from the continuous state space (x_1, x_n).

When the model space has been specified, the M-regression function

$$m(x) = \sum_{v=0}^{l} \beta_{v,0}(x-t_0)_+^v + \sum_{m=1}^{k}\sum_{v=l_0}^{l} \beta_{v,m}(x-t_m)_+^v \tag{9}$$

is specified through the coefficients β. Let z be the vector of the basis of piecewise polynomials evaluated at x, then

$$m(x) = z'\beta. \tag{10}$$

We use the noninformative prior $\pi_\beta(\beta|k,t,\sigma) = 1$ on $\boldsymbol{R}^d$ for β, where $d = l+1+k(l-l_0+1)$. Although without a closed form, the posterior of β derived from this noninformative prior is proper (Lemma 1 in the Appendix).

A key step in RJMCMC is to decide the accept/reject probability for moves from one model (k,t) to another model (k',t'). As shown by Green (1995, [9]) and Denison et al. (1998, [4]), the acceptance probability for our problem is

$$\alpha = \min\left\{1, \frac{p(y|k',t')}{p(y|k,t)} \frac{\pi_{k,t}(k',t')}{\pi_{k,t}(k,t)} \frac{q(k,t|k',t')}{q(k',t'|k,t)}\right\}, \tag{11}$$

where

$$\begin{aligned} p(y|k,t) \\ &= \int\int \prod_{i=1}^{n} h(y_i|m(x_i),\sigma)\pi_\beta(\beta|k,t,\sigma)\pi_\sigma(\sigma)d\beta d\sigma \end{aligned} \tag{12}$$

is the marginal likelihood of (k,t) and $q(k,t|k',t')$ is the proposal probability of the equilibrium distribution.

The prior ratio $\frac{\pi_{k,t}(k',t')}{\pi_{k,t}(k,t)}$ can be computed once the priors are specified. The proposal ratio $\frac{q(k,t|k',t')}{q(k',t'|k,t)}$ can be computed according to the moving strategy, which will be discussed in the next section. What left is to compute the marginal likelihood ratio $\frac{p(y|k',t')}{p(y|k,t)}$, which is also called the Bayesian factor. In the literature of mean curve fitting, Denison et al. (1998, [4]) used the conditional likelihood ratio evaluated at the maximum likelihood estimate of β. DiMatteo et al. (2001, [6]) pointed out that the conditional likelihood ratio incurs overfitting and penalty due to the uncertainty of the coefficient β should be considered. They showed the closed form of the marginal likelihood ratio in the Gaussian case and the SIC approximation in other cases of an exponential family. Hansen and Kooperberg (2002, [11]) also used the conditional likelihood ratio. However, they evaluated this ratio at a penalized smoothing estimator of β, which is a Bayesian solution with a partially improper prior. Kass and Wallstrom (2002, [17]) pointed out that the Hansen–Kooperberg method can be approximately Bayesian and won't overfit, if a proper smoothing parameter is chosen.

In our case, using a noninformative prior of β, we are able to get the approximation

$$\frac{p(y|k',t')}{p(y|k,t)} \approx n^{\frac{d-d'}{2}} \left(\frac{D(k,t)}{D(k',t')}\right)^{n/2}, \tag{13}$$

where $D(k,t) = \sum_{i=1}^{n} \rho_{\hat\sigma H}(y_i - z_i'\hat\beta_M(k,t))$, $d' = l+1+k'(l-l_0+1)$, $\hat\sigma$ is a scale estimate, and $\hat\beta_M(k,t)$ is the M-estimate for the model (k,t). A simple derivation of this approximation is given in the Appendix (Lemma 2).

Once this likelihood ratio is computed, the remaining work to compute the Metropolis-Hastings accept/reject probability α in RJMCMC is to compute the proposal ratio $\frac{q(k,t|k',t')}{q(k',t'|k,t)}$. This ratio acts as a symmetric correction for various moves. The following section describes a scheme that involves these moves and their corresponding proposal ratios.

4. MOVING STRATEGIES IN RJMCMC

Following the scheme of Green (1995, [9]) and Denison et al. (1998, [4]), we describe moves that involve knot addition, deletion, and relocation. For each move $(k,t) \longrightarrow (k',t')$, the potential destination models (k',t') form a subspace, which is called *allowable space* by Hansen and Kooperberg (2002, [11]). For the same type of moves, for example knot addition $(k' = k+1, t' = (t, t_{k+1}))$, the subspace is defined by all possible choices of the $(k+1)$st knot. Denote $M_k = (k,t)$ and $M_{k+1} = (k+1, (t, t_{k+1}))$.

Different ways to restrict the subspace provide different moving strategies in RJMCMC. Denison et al. (1998) and Hansen and Kooperberg (2002, [11]) chose the candidate knot uniformly from data points and require that it is at least n_{sep} data points away from current knots to avoid numerical instability. We call it the *discrete proposal*. On the contrary, DiMatteo et al. (2001, [6]) used *continuous proposal* distributions. This continuous strategy, which follows a locality heuristic observation by Zhou and Shen (2001,

[25]), attempts to place knots close to existing knots in order to catch sharp changes.

Denison et al. (1998, [4]) required $n_{sep} \geq l$ to avoid numerical instability. Using a different set of basis functions for piecewise polynomials, such as the Boor basis, rather than the explosive truncated power basis as in (4) significantly reduces the condition number of the design matrix, thus the numerical instability.

Our experiences suggest that, in M-regression, the discrete proposal works as well as or better than the continuous proposal, especially with middle or large data sets ($n \geq 200$). One explanation is that placing too many knots near a point, where data may form some suspicious patterns, would result in a chance of overfitting locally, inflating the design matrix and impairing the computational efficiency.

Following Denison et al. (1998, [4]), the probabilities of addition, deletion, and relocation steps of the RJMCMC sampler are

$$\begin{aligned} b_k &= c \min\{1, \pi_k(k+1)/\pi_k(k)\}, \\ d_k &= c \min\{1, \pi_k(k-1)/\pi_k(k)\}, \\ \eta_k &= 1 - b_k - d_k, \end{aligned}$$

where c is a constant in $(0, \frac{1}{2})$, which controls the rate of the dimension change among these moves. These probabilities ensure that $b_k \pi_k(k) = d_{k+1} \pi_k(k+1)$, which will be used to maintain detailed balance as requested by RJMCMC. With these probabilities, RJMCMC cycles among proposals of addition, deletion, and relocation.

Knot Addition. A candidate knot is uniformly selected from the allowable space. Assume that currently there are k knots from the n data points. Then, the allowable space has $n - Z(k)$ data points, where

$$Z(k) = 2(n_{sep} + 1) + k(2n_{sep} + 1). \tag{14}$$

In this case the jump probability is

$$q(M_{k+1}|M_k) = b_k \frac{n - Z(k)}{n}. \tag{15}$$

Knot Deletion. A knot is uniformly chosen from the existing set of knots and deleted. The jump probability from M_k to M_{k-1} is

$$q(M_{k-1}|M_k) = d_k \frac{1}{k}. \tag{16}$$

Knot Relocation. A knot t_{i^*} is uniformly chosen from the existing set of knots and relocated within the allowable intervals between its two neighbors. Relocation does not change the order of the knots. Let M_C be the current model and M_R be the model after relocation. The jump probability from M_R to M_C is

$$q(M_R|M_C) = \eta_k \frac{1}{k} \frac{n(t_{i^*}) - 2n_{sep}}{n(t_{i^*})}, \tag{17}$$

where $n(t_{i^*})$ is the number of data points between the two neighboring knots of t_{i^*}. Due to the symmetry, $q(M_R|M_C) = q(M_C|M_R)$.

5. THE RJMCMC ALGORITHM WITH M-REGRESSION

In this section we describe details of the RJMCMC algorithm for M-regression, especially the initialization of RJMCMC. Our experience shows that the initialization has impact on the performance of RJMCMC.

To set up an initial model configuration, we choose λ locations between x_1 and x_n, where λ could be the prespecified mean of the distribution of the number of knots. The λ locations take the values of $[hJ]$th observations of x_i, $i = 1, \ldots, n$, where $h = [\frac{n}{\lambda+1}]$ and $J = 1, \ldots, \lambda$. In this way, we evenly assign $h - 1$ observations between two neighboring knots. Compared with other initial knot assignment methods, this even observation assignment (EOA) is more nature for the implementation of our strategy to add a knot, which prefers certain symmetric distribution of observations between neighboring knots. Our experience shows that EOA performs better than other initial knot assignment methods, for example, evenly spacing on (x_1, x_n).

To implement a full Bayesian version of RJMCMC for M-regression, we need to draw β from its posterior distribution, which does not have a closed form in our case. Instead, we use the posterior mode $\hat{\beta}$, which is the M-estimate for the given model (k, t) and scale parameter σ. This improves computation efficiency, since the posterior mode $\hat{\beta}$ has already been computed when we compute the acceptance probability of (k, t) with the given σ.

With these details clarified, the algorithm of RJMCMC for M-regression is described as the following steps:

1. Sort the data by the independent variable. Then, normalize the independent variable to interval $[0, 1]$.
2. Assign initial knots according to the method described early in this section.
3. Run RJMCMC N_b iterations for the burn-in process from step (a) to (e).
   (a) Take knot steps: addition, deletion, relocation. This recommends a new model (k, t).
   (b) Compute the M-estimate $\hat{\beta}_M(k, t)$ for model (k, t).
   (c) Compute the acceptance probability α based on $\hat{\beta}_M(k, t)$.
   (d) Update the model according to the accept/reject scheme.
   (e) Draw σ with the Gibbs sampling method.
4. Run RJMCMC N_s iterations for the sampling process after the N_b iterations of burn-in. Within each iteration, in addition to the steps (a) to (e) in 3, sequentially run the following step (f):

Table 1. The Huber family for M estimation

Name	$\rho(r)$	$\psi(r)$	Range
A	$\begin{cases} A^2[1-\cos(r/A)] \\ 2A^2 \end{cases}$	$A\sin(r/A)$ 0	$\lvert r\rvert \leq \pi A$ $\lvert r\rvert > \pi A$
B	$\begin{cases} (B^2/2)[1-[1-(r/B)^2]^3] \\ (B^2/2) \end{cases}$	$3r[1-(r/B)^2]^2$ 0	$\lvert r\rvert \leq B$ $\lvert r\rvert > B$
T	$\begin{cases} r^2/2 \\ T^2/2 \end{cases}$	r 0	$\lvert r\rvert \leq T$ $\lvert r\rvert > T$
C	$(C^2/2)\log[1+(r/C)^2]$	$r[1+(r/C)^2]^{-1}$	
W	$(W^2/2)[1-\exp[-(r/W)^2]]$	$r\exp[-(r/W)^2]$	
H	$\begin{cases} r^2/2 \\ H\lvert r\rvert - H^2/2 \end{cases}$	r $H\mathrm{sign}(r)$	$\lvert r\rvert \leq H$ $\lvert r\rvert > H$
L	$L^2\log[\cosh(r/L)]$	$L\tanh(r/L)$	
F	$[\lvert r\rvert - F\log(1+\lvert r\rvert/F)]$	$r(1+\lvert r\rvert/F)^{-1}$	

(f) Using $\hat{\beta}_M(k,t)$, obtain the M-regression fit $\hat{m}(x)$, objective function value D, number of modes of $\hat{m}(x)$, and other interested summary statistics.

5. From the sampling process, obtain mean estimates of the M-regression function values $m(x)$ and means of the objective function value D and the number of modes of $m(x)$, respectively.

In the algorithm, several parameters need to be specified. Most of them should be specified problem-wise. For the number of iterations in the burn-in process, we require it large enough so that the mean of objective function values becomes stable. We recommend $N_b = 2000$ in practice. The number of iterations in the sampling process depends more on the requirement of accuracy of the summary statistics. We recommend $N_s = 5000$ in practice.

With the least informative likelihood in (1), the posterior of σ given (k,t) and β follows an inverse gamma distribution. The Gibbs sampler draws σ from this inverse gamma distribution.

The most computationally intensive part of the algorithm is computing the M-estimate $\hat{\beta}_M(k,t)$. We use the iterative reweighted least squares algorithm.

The final evaluation of the fitted M-regression function can be taken on all observed values or a grid of the independent variable. We measure the goodness-of-fit of the estimated Bayesian M-regression function based on the mean squared error on the observed values

$$mse = \frac{1}{n}\sum_{i=1}^{n}(\hat{m}(x_i) - m(x_i))^2. \tag{18}$$

6. EXTENSION OF THE ρ FUNCTION AND SELECTION OF THE TUNING CONSTANT

We have developed Bayesian M-regression based on the least informative distribution derived from the Huber function ρ_H in (2). In practice, other ρ functions have been used with M-regression. Table 1 presents the most commonly used ones, which constitute the Huber family. The first derivative of ρ, ψ, is the score function. The Huber family for M-estimation can be divided into three classes according to the score functions:

The *hard redescenders*: Scores A (Andrews et al. 1972, [1]), B (Tukey's bisquare), and T (Hinich and Talwar 1975, [14]) with $\psi(r) = 0$ for sufficiently large $|r|$.

The *soft redescenders*: Scores C (Cauchy or t-likelihood) and W (Dennis and Welsch 1976, [5]) with $\psi(r) \to 0$ as $r \to \pm\infty$.

The *monotone scores*: Scores H (Huber 1964, [15]), L (Logistic), and F (Fair 1974, [7]) with monotone ψ functions.

Some of the ρ functions are negative log-likelihood functions of well-known distributions, for example, the Cauchy and logistic distributions. It can be easily verified that all ρ functions with monotone or soft redescending scores are negative log-likelihood functions of proper distributions. Bayesian M-regression can be exercised similarly as what we have done with the least informative distribution.

However, ρ functions with hard redescending scores can not be negative log-likelihood functions of any distribution (Maronna et al. 2005, P. 29, [18]). This is due to the fact that these ρ functions have flat constant tails. To overcome this difficulty, instead of using proper likelihood functions, we introduce the pseudo-likelihood function. Corresponding to (1), with the location parameter μ and the scale parameter σ, we call

$$\frac{1}{\sigma} e^{-\rho(\frac{y-\mu}{\sigma})} \tag{19}$$

a pseudo-likelihood function.

With the noninformative priors for β and σ as used in Section 3, the M-estimate of β,

$$\hat{\beta}_M = \arg\min_{\beta} \sum_{i-1}^{n} \rho(y_i - m(x_i, \beta)), \tag{20}$$

where $\rho = \rho_{\sigma A}$ (or $\rho_{\sigma B}$, $\rho_{\sigma T}$), is taken as a posterior mode of β, while the posterior of σ has an inverse gamma distribution.

Although the marginal likelihood for (k, t) does not exist, we use

$$n^{\frac{d-d'}{2}} \left(\frac{D(k,t)}{D(k',t')} \right)^{\frac{n}{2}},$$

where $D(k,t) = \sum_{i=1}^{n} \rho_{\hat{\sigma}A}(y_i - z_i' \hat{\beta}_M)$, as an approximation to the Bayesian factor. When ρ is not convex, the solution $\hat{\beta}_M$ in (20) might not be unique and a good start value is needed to obtain a local solution. When ρ is convex, especially with a monotone score function, the solution for the M-estimate is usually unique and can be obtained with less computational complexity in practice. Although M-estimates from hard redescenders are more resistant to large outliers (Maronna et al. 2005, P. 39, [18]), experiences from our implementation with RJMCMC recommend M-estimates with monotone scores.

To compute M-estimates in (5) and (20), we need to specify the tuning constant. In M-regression, the tuning constant is usually chosen such that the M-estimate has a specified asymptotic efficiency (e.g., 0.85). Wang et al. (2007, [23]) recently proposed a data-driven method. Let $r_i = \frac{y_i - z_i' \hat{\beta}_{L1}}{\hat{\sigma}}$, where $\hat{\beta}_{L_1}$ is the L_1 (median) regression estimator and $\hat{\sigma}$ is the corresponding MAD (median absolute deviation) estimator of σ. Choose the tuning constant H such that the empirical efficiency

$$\begin{aligned} &\hat{\tau}(H) \\ &= \frac{\{\sum_{i=1}^{n} I(|r_i| \leq H)\}^2}{n \sum_{i=1}^{n} \{I(|r_i| \leq H)\psi^2(r_i) + H^2 I(|r_i| > H)\}} \end{aligned} \tag{21}$$

is maximized, where $I(\cdot)$ is the indicator function. Since $\hat{\tau}(H)$ is not a continuous function, a searching algorithm within an interval (e.g., (0, 3) by 0.1) was used to find the optimal H. Wang et al. (2007, [23]) show that H chosen in this way outperforms a fixed one chosen according to the asymptotic efficiency. Using the data-driven tuning constant makes our Bayesian M-regression a more automatic procedure.

7. NUMERICAL RESULTS

In this section, we present numerical results with the Bayesian M-regression method we developed with both fixed and data-driven tuning constants. We call the fixed tuning constant method Bayesian M-regression (BMR) and the data-driven tuning constant method automatic Bayesian M-regression (ABMR).

First, we use simulations to show how our method works on fitting various kinds of curves. Then, We compare our method with the Bayesian adaptive regression spline (BARS) method of DiMatteo et al. (2001, [6]) with or without outliers. Numeric results show that Bayesian M-regression performs competitively without outliers, but significantly better with outliers. Finally, as an application, we present how our method can be used on denoising image data with outliers.

7.1 Simulations

We simulate data from three underlying functions on $(0, 1)$:

Wave:
$f(x) = 4(x - .5) + 2\exp(-256(x - .5)^2)$,
Doppler:
$f(x) = 4(.2x(1 - .2x))^{\frac{1}{2}} \sin(\pi(1 + \epsilon)/(.2x + \epsilon)),\ \epsilon = .05$,
Block:
$f(x) = \sum h_j K(x_j - x),\ K(x) = (1 + sgn(x))/2$,

where $x_j = \{0.1, 0.4, 0.5, 0.75, 0.8\}$, $h_j = \{2, -2, 4, -1, 1\}$. The first two curves are continuous. *Wave* has a single mode, while *Doppler* has multiple modes. *Block* is piecewise constant. Similar functions have been used by Denison et al. (1998) and others to check their curve fitting techniques.

First, we generate the data uniformly on $(0, 1)$ from these curves and the additive Gaussian noise.

$$\begin{aligned} &\textit{Wave:} && y = f(x) + N(0, 0.4),\ n = 200, \\ &\textit{Doppler:} && y = f(x) + N(0, 0.1),\ n = 512, \\ &\textit{Block:} && y = f(x) + N(0, 0.4),\ n = 200, \end{aligned}$$

where n is the sample size. Seven outliers with value 10 are randomly added to each curve.

Figures 1, 2, and 3 show the true and fitted curves of a single run using ABMR and BARS with linear piecewise polynomials ($l = l_0 = 1$) for the two continuous curves and piecewise constants ($l = l_0 = 0$) for the *Block* function. We used 2000 iterations in the burn-in process and 5000 iterations in the sampling process. We also ran BMR with a fixed tuning constant 1.25. Fitted curves (not shown) are very close to those of ABMR with a slightly larger *mse*.

Figure 1 shows that BARS largely mis-fits the *Wave* curve. The fitted curve is driven flat by the seven randomly distributed outliers. However, ABMR demonstrates an almost perfect fit. Figure 2 shows fits for *Doppler*. ABMR

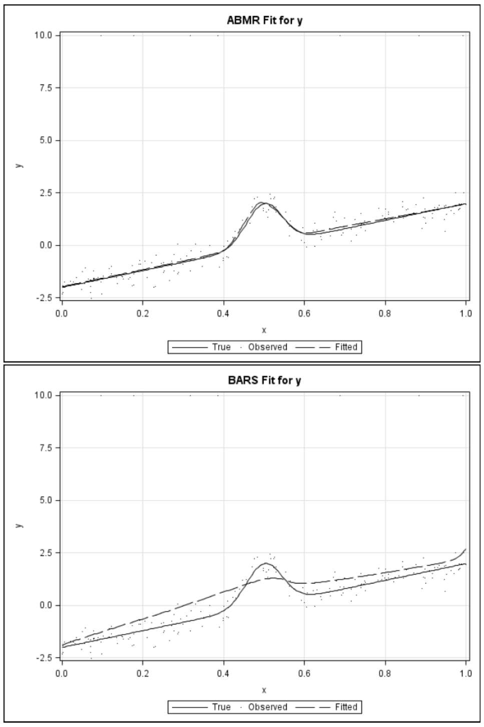

Figure 1. ABMR and BARS fits for $Wave$ *with outliers.*

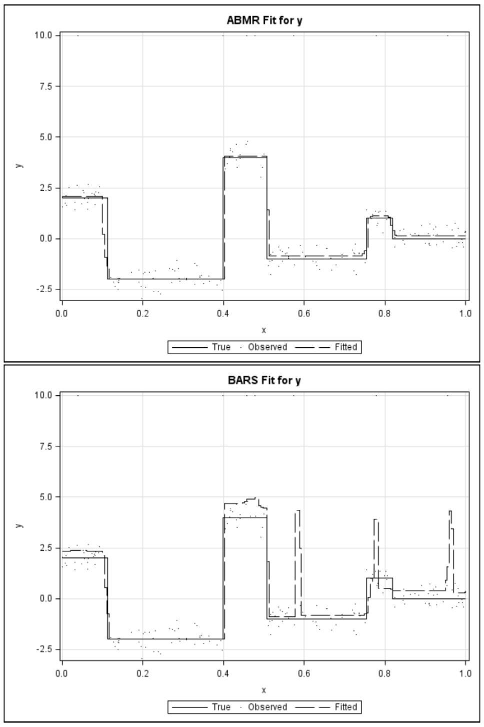

Figure 3. ABMR and BARS fits for $Block$ *with outliers.*

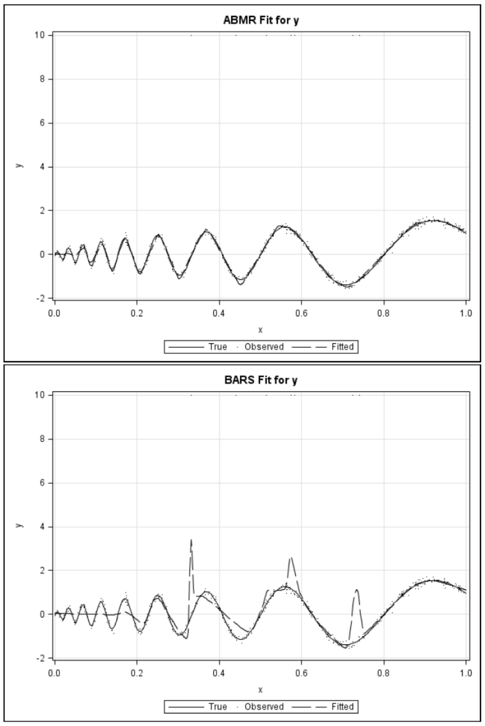

Figure 2. ABMR and BARS fits for $Doppler$ *with outliers.*

ignores the distraction from outliers and fits the multi-mode curve well, except near the origin, where signal is relatively weak compared to the noise. The fit by BARS presents jumps towards the outliers and misses the target between these jumps. However, BARS fits well at the outlier-free tail. Figures 3 shows fits for the discontinuous block function, which is usually more difficult for classical curve fit techniques. Bayesian methods based on RJMCMC has demonstrated advantages (Denison et al. 1998, [4]). With outliers, our AMBR method fits those blocks well. Again, BARS presents sharp jumps towards outliers and misfits some blocks. It fits well on outlier-free ranges.

For a systematic numerical comparison, the mean squared error mse defined in (18) is used for the measure. We run simulations with 10 repeats for both the case without outliers and the case with 3% outliers (with value 10). For BMR, we use a common tuning constant 1.25 for all cases. Further more, to show how the signal to noise ratio affects the performance, we run simulations with different standard errors. For $Wave$ and $Block$, we use standard errors 0.2, 0.4, and 0.8 for the Gaussian noise. For $Doppler$, we use 0.1, 0.2, and 0.4.

Table 2 shows the average mean squared errors followed by the standard errors of 10 repeats for the three curves. Without outliers, for all different curves, ABMR, BMR, and BARS are competitive. The standard error of the noise σ, which decides the signal to noise ratio, plays the dominate factor for the changes of mse.

Table 2. MSE of fitted curves without outliers

Function	Method	σ		
Wave		0.2	0.4	0.8
	BARS	.0024(.0011)	.0095(.0033)	.0407(.0079)
	BMR	.0027(.0011)	.0116(.0037)	.0410(.0093)
	ABMR	.0028(.0012)	.0116(.0036)	.0414(.0096)
Doppler		0.1	0.2	0.4
	BARS	.0017(.0002)	.0051(.0005)	.0169(.0032)
	BMR	.0018(.0002)	.0056(.0005)	.0181(.0025)
	ABMR	.0018(.0002).	.0053(.0005)	.0182(.0024)
Block		0.2	0.4	0.8
	BARS	.0241(.0452)	.0404(.0424)	.0615(.0236)
	BMR	.0211(.0350)	.0464(.0502)	.0677(.0320)
	ABMR	.0182(.0253)	.0390(.0399)	.0628(.0248)

Table 3. MSE of fitted curves with outliers

Function	Method	σ		
Wave		0.2	0.4	0.8
	BARS	.2923(.1298)	.3184(.0694)	.5048(.3859)
	BMR	.0059(.0024)	.0094(.0036)	.0349(.0153)
	ABMR	.0028(.0009)	.0084(.0223)	.0334(.0168)
Doppler		0.1	0.2	0.4
	BARS	.2510(.0725)	.2975(.0755)	.3935(.1491)
	BMR	.0226(.0050)	.0222(.0031)	.0353(.0047)
	ABMR	.0121(.0031)	.0149(.0032)	.0322(.0052)
Block		0.2	0.4	0.8
	BARS	1.046(.4514)	1.1739(.5094)	1.1449(.4619)
	BMR	.0478(.0531)	.0646(.0620)	.0917(.0596)
	ABMR	.0270(.0586)	.0756(.0703)	.0863(.0774)

Table 3 shows the average mean squared errors followed by the standard errors of 10 repeats for the three curves with 3% outliers with value 10. One can see that the averages of *mse* of BARS are at least 10 times larger than that of BMR or ABMR across all simulations. For *Wave* with $\sigma = 0.2$, the averaged *mse* of BARS is more than 100 times of that of ABMR. Between BMR and ABMR, ABMR has smaller averages of *mse* for almost all cases, except for *Block* with $\sigma = 0.4$. Although σ (or the signal to noise ratio) plays a role in the changes of *mse*, it is not as significant as in the case without outliers. Outliers plays a larger role in the changes of *mse*, especially for BARS.

7.2 Applications with image data

In image processing, denoising is an important step. We use a simulated image data set similar to the one used by Chu et al. (1998) to illustrate that our automatic Bayesian M-regression method could be a better performer in denoising with outliers.

As pointed out by Chu et al. (1998, [3]), jumps or edges between regions happen frequently in images. We take a similar underlying function used by Chu et al. (1998, [3]), which is a step (block) function with a sharp jump near 0.65. Figure 4 displays the simulated one-dimensional image data together with the fitted curves. On the left, the curve fitted with our automatic Bayesian M-regression method is indicated by the dashed line. The fitted curve catches all of the jumps accurately. On the right, the curve fitted with BARS also catches almost all of the jumps, except for the smallest one near 0.88.

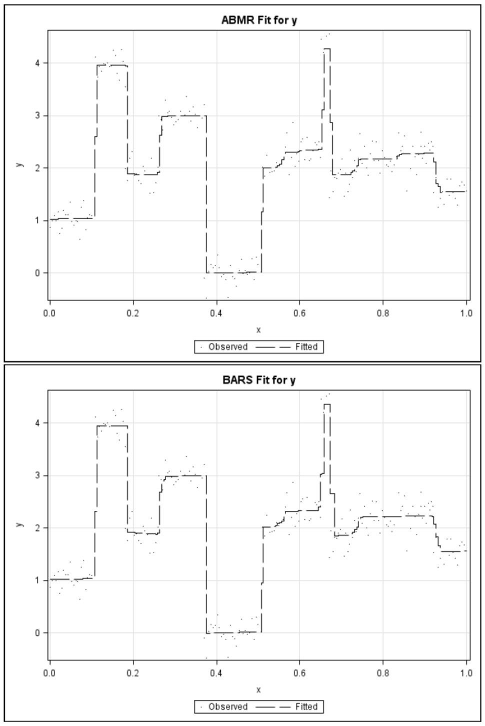

Figure 4. ABMR and BARS fits for image data.

To demonstrate robustness of Bayesian M-regression curve fit, we contaminate the data with one outlier near 0.4 with a value 4.25 and another outlier near 0.6 with a value 4.5. Figure 5 shows that the two outliers do not change the curve fitted by ABMR too much. The only noticeable change is the jump near 0.88, which now becomes less clear. However, the two outliers change the BARS fit significantly. Not only appear two large false jumps towards the outliers, the BARS fit also misses the target with the highest jump near 0.65.

To further test the strength of resistance to outliers, we enlarge the second outlier with a value of 20.5. Figure 6 (upper) shows the fitted curve by ABMR. The jump near 0.56 becomes less clear and the jump near 0.88 is gone. We also show the fitted curve (lower) by using the ρ function B (Table 1) with the bisquare score function. The bisquare fit presents a little clearer jump near 0.56 and some variation near 0.88. However, it took about twice the computing time used by the Huber function.

In summary, we can conclude that Bayesian M-regression is less affected by individual outliers (not clouded together). The Huber score function is fast and good enough for such outliers in practice.

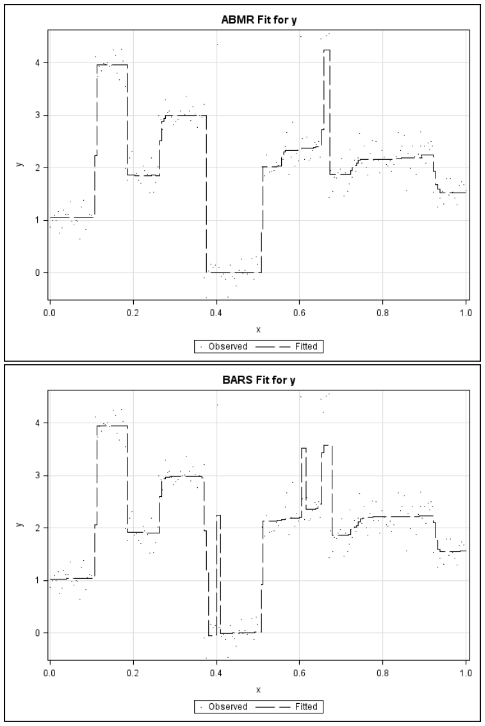

Figure 5. ABMR and BARS fits for image data with outliers.

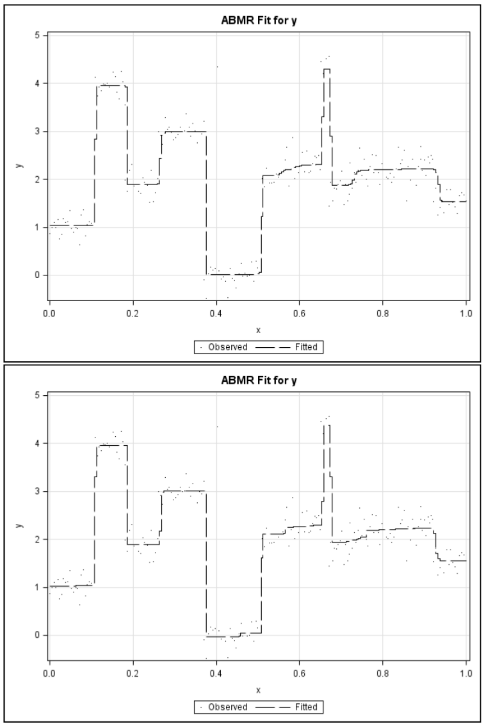

Figure 6. ABMR fits for image data with enlarged outliers (upper: Huber, lower: Bisquare).

8. DISCUSSIONS

We present a Bayesian approach to fit nonparametric M-regression models. The approach is automatic in the sense that the number of knots and their locations for the nonparametric piecewise polynomial model, as well as the tuning constants, are selected automatically in the fitting procedure based on data. Our approach can be considered as a natural extension of Bayesian least squares based curve fit with free knots to M-regression models. Numerical results show that our approach is competitive in accuracy and robustness for fitting nonparametric models.

The M-regression models discussed in the current paper have symmetric error distributions. For models with asymmetric error distributions or heterogeneous error distributions, M-regression may not be suitable. Some alternatives, for examples, the Bayesian adaptive quantile regression models proposed in Chen and Yu (2008, [2]), should be considered.

It is well-known that Bayesian modeling based on MCMC with large data sets has the drawback of slow computing speed. Our automatic Bayesian M-regression approach needs to compute M-estimates for each selected model. The computation involved could be extensive if a long chain is requested for MCMC. However our implementation of the iteratively reweighted least squares algorithm for computing M-estimates achieves reasonable computing efficiency. For data sets with several thousands of observations, we are able to get a stable chain of length 5,000 within several minutes with a Dell GX620 Desktop, which runs a 3.2 GHz Pentium(R) 4 processor with 2 GB RAM.

In this paper, we focus on one-dimensional fitting. For multi-dimensional fitting, additive models can be used. Applications of our approach on multi-dimensional fitting for M-regression models are under investigation.

APPENDIX A. PROOFS

Lemma 1. *For the least informative distribution in (1), given a sample $\{(y_i, x_i), i = 1, \dots, n\}$ and a piecewise polynomial fit $\mu = m(x) = z'\beta$, the posterior of β with the noninformative prior $\pi_\beta(\beta|k, t, \sigma) = 1$ is proper.*

Proof. The joint likelihood of β and σ is

$$\begin{aligned} l(\beta,\sigma) &= C_H^{-n}\sigma^{-n}e^{-\sum_{i=1}^n \rho_H(\frac{y_i - z_i'\beta}{\sigma})} \\ &= C_H^{-n}\sigma^{-n}e^{-\frac{1}{\sigma^2}\sum_{i=1}^n \rho_{\sigma H}(y_i - z_i'\beta)}. \end{aligned}$$

Let $v_i = y_i - z_i'\beta$, $r_i = \frac{v_i}{\sigma}$,

$$s_i(\beta) = \begin{cases} -1 & \text{if } r_i \le -H \\ 0 & \text{if } -H < r_i < H \\ 1 & \text{if } r_i \ge H, \end{cases}$$

and $w_i(\beta) = 1 - s_i^2(\beta)$, then $\rho_{\sigma H}(y_i - z_i'\beta) = \frac{1}{2}w_i v_i^2 + H s_i v_i -$

$\frac{H^2}{2}s_i^2$ and

$$D(\beta) = \sum_{i=1}^{n} \rho_{\sigma H}(y_i - z_i'\beta) = \frac{1}{2}v'Wv + g'(s)v + c(s), \tag{22}$$

where $s = (s_1, \ldots, s_n)'$, $v = (v_1, \ldots, v_n)'$, $g(s) = Hs$, $c(s) = -\frac{H^2}{2}\sum_{i=1}^{n} s_i^2$, and W is the $n \times n$ diagonal matrix with diagonal elements w_i.

$D(\beta)$ is continuous and piecewise quadratic on $\boldsymbol{R}^d$. Its gradient is given by

$$D^{(1)}(\beta) = -Z'[Wv + g(s)]$$

and for $\beta \notin \Omega = U_{i=1}^{n}\{\beta : |r_i(\beta)| = H\}$, the Hessian exists and is given by

$$D^{(2)}(\beta) = Z'WZ,$$

where $Z = (z_1, \ldots, z_n)'$ is the design matrix for the piecewise polynomial fit.

Assume that one component of β, $|\beta_{v,m}| \to \infty$ and the corresponding coefficients in $z_i, i = 1, \ldots, n$, are not zero while other components of β are fixed, then $|r_i| \to \infty$, $|s_i| = 1$, and $w_i = 0$, $i = 1, \ldots, n$. From (22), $D(\beta)$ is an L_1 function of $\beta_{v,m}$ when $\beta_{v,m}$ exceeds a threshold. The above argument applies to all components of β. Let $\Delta(\beta)$ $= \min\{|\beta_{v,m}|, 0 \le v \le l, 1 \le m \le k\}$. $D(\beta)$ is convex and there exists $\Delta_0 > 0$ such that $D(\beta)$ is an L_1 function of β when $\Delta(\beta) > \Delta_0$ and piecewise quadratic when $\Delta(\beta) \le \Delta_0$. So, given σ,

$$\int_{\boldsymbol{R}^d} e^{-\frac{D(\beta)}{\sigma^2}} d\beta = \int_{\Delta(\beta)>\Delta_0} e^{-\frac{D(\beta)}{\sigma^2}} d\beta + \int_{\Delta(\beta)\le\Delta_0} e^{-\frac{D(\beta)}{\sigma^2}} d\beta. \tag{23}$$

As a special case of Yu and Moyeed (2001, [24]) (Theorem 1, the median case), the first part of the right-hand side of (23) is finite. Since $D(\beta)$ is continuous on $\boldsymbol{R}^d$, the second part of the right-hand side of (23) is also finite. So, the posterior of β has a density. □

The following result for Bayesian factor approximation is under special conditions. However, we found that these conditions are commonly true in practice.

Lemma 2. *Define $\Delta(\beta)$ and Δ_0 as in Lemma 1. Assume $D^{(2)}(\beta)$ be full rank in $\{\beta : \Delta(\beta) < \Delta_0\}$ as $\Delta_0 \to \infty$. For $k' < k$,*

$$\frac{p(y|k',t')}{p(y|k,t)} = \left[n^{\frac{d-d'}{2}}\left(\frac{D(k,t)}{D(k',t')}\right)^{n/2}\right](1 + o(1)),$$

where $D(k,t) = \sum_{i=1}^{n}\rho_{\hat\sigma H}(y_i - z_i'\hat\beta_M(k,t))$, $d' = l + 1 + k'(l - l_0 + 1) < d = l + 1 + k(l - l_0 + 1)$, $\hat\sigma$ is a consistent scale estimate, and $\hat\beta_M(k,t)$ is the M-estimate for the model (k,t).

Proof. For simple notations, denote $\hat\beta_M(k,t)$ as $\hat\beta_M$. Because $\hat\beta_M$ is the M-estimate, so $D^{(1)}(\hat\beta_M) = 0$. From Lemma 1,

$$D(\beta) = \frac{1}{2}(\beta - \hat\beta_M)'D^{(2)}(\beta)(\beta - \hat\beta_M) + D(\hat\beta_M) \quad \text{a.s.},$$

then

$$p(y|k,t) = \int_{\boldsymbol{R}^+}\int_{\boldsymbol{R}^d}(\sigma C_H)^{-n}e^{-\frac{D(\beta)}{\sigma^2}}d\beta d\sigma = C_H^{-n}$$
$$\times\int_{\boldsymbol{R}^+}\left[\sigma^{-n}e^{-\frac{D(\hat\beta_M)}{\sigma^2}}\int_{\boldsymbol{R}^d}e^{-\frac{1}{2\sigma^2}(\beta-\hat\beta_M)'D^{(2)}(\beta)(\beta-\hat\beta_M)}d\beta\right]d\sigma.$$

$D(\hat\beta_M)$ depends on σ only through the sign vector s. Let $\hat\sigma$ be a consistent estimate of σ, i.e., $\hat\sigma \to \sigma$ as $n \to \infty$. So, $D(k,t) = D(\hat\beta_M)(\hat\sigma) = D(\hat\beta_M)$ as $\hat\sigma \to \sigma$.

Since $\sigma^{-n}e^{-\frac{D(k,t)}{\sigma^2}}$ is bounded (for fixed n), so

$$C_H^{-n}\int_{\boldsymbol{R}^+}\left[\sigma^{-n}e^{-\frac{D(k,t)}{\sigma^2}}\int_{\boldsymbol{R}^d}e^{-\frac{1}{2\sigma^2}(\beta-\hat\beta_M)'D^{(2)}(\beta)(\beta-\hat\beta_M)}d\beta\right]d\sigma$$
$$= p(y|k,t)$$

as $\hat\sigma \to \sigma$ and we have

$$C_H^{-n}\int_{\boldsymbol{R}^+}\left[\sigma^{-n}e^{-\frac{D(k,t)}{\sigma^2}}\int_{\boldsymbol{R}^d}e^{-\frac{1}{2\sigma^2}(\beta-\hat\beta_M)'D^{(2)}(\beta)(\beta-\hat\beta_M)}d\beta\right]d\sigma$$
$$= p(y|k,t)(1 + o(1)).$$

$\int_{\boldsymbol{R}^d}e^{-\frac{1}{2\sigma^2}(\beta-\hat\beta_M)'D^{(2)}(\beta)(\beta-\hat\beta_M)}d\beta$ is approximated by its integration on the finite range $\{\beta : \Delta(\beta) < \Delta_0\}$. Under the full rank condition of $D^{(2)}(\beta)$ in $\{\beta : \Delta(\beta) < \Delta_0\}$ and by integration of an inverse gamma distribution of σ with the left-hand side in the above equation, we prove the lemma. □

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Editors and the referee for their very critical and constructive comments and suggestions, which greatly improved the article.

Received 3 January 2008

REFERENCES

[1] Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W., and Tukey, J. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton: Princeton University Press. MR0331595

[2] Chen, C. and Yu, K. (2008). Automatic Bayesian quantile regression curve fitting. Revised.

[3] Chu, C. K., Glad, I. K., Godtliebsen, F., and Marron, J. S. (1998). Edge-preserving smoothers for image processing (with discussion). *Journal of the American Statistical Association* **93** 526–541. MR1631321

[4] Denison, D., Mallick, B., and Smith, A. (1998). Automatic Bayesian curve fitting. *J. Rol. Statist. Soc. B.* **60** 333–350. MR1616029

[5] Dennis, J. E. and Welsch, R. E. (1976). Techniques for nonlinear least squares and robust regression, *Proc. Amer. Statist. Assoc. Statist. Comp. Section* 83–87.

[6] DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve fitting with free-knot splines. *Biometrika* **88** 1055–1073. MR1872219

[7] Fair, R. C. (1974). On the robust estimation of econometric models. *Ann. Econ. Social Measurement* **3** 667–678.

[8] Gao, J. T. and Shi, P. D. (1997). M-type smoothing splines in nonparametric and semiparametric regression models. *Statistica Sinica* **7** 1155–1169. MR1488663

[9] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810

[10] Hall, P. and Jones, M. (1990). Adaptive M-estimation in nonparametric regression. *Ann. Statist.* **18** 1712–1728. MR1074431

[11] Hansen, M. H. and Kooperberg, C. (2002). Spline adaptation in extended linear models (with discussion). *Statistical Science* **17** 2–51. MR1910073

[12] Härdle, W. and Gasser, T. (1984). Robust nonparametric function fitting. *J. Rol. Statist. Soc. B.* **46** 42–51. MR0745214

[13] He, X. and Shi, P. D. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics* **3** 299–308. MR1291551

[14] Hinich, M. J. and Talwar, P. P. (1975). A Simple Method for Robust Regression, *Journal of the American Statistical Association* **70** 113–119.

[15] Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.* **35** 73–101. MR0161415

[16] Huber, P. J. (1981). *Robust Statistics*, New York: Wiley. MR0606374

[17] Kass, R. E. and Wallstrom, G. L. (2002). Comment on: Spline adaptation in extended linear models by Mark H. Hansen and Charles Kooperberg. *Statistical Science* **18** 2–51. MR1910073

[18] Maronna, R. A., Martin, R. D., and Yohai, V. J. (2005). *Robust Statistics – Theory and Methods*, New York: Wiley. MR2238141

[19] Peña, D. and Redondas, D. (2006). Bayesian curve estimation by model averaging. *Computational Statistics and Data Analysis* **50** 688–709. MR2207002

[20] Shi, P. D. and Li, G. Y. (1995). Global convergence rates of B-spline M-estimates in nonparametric regression. *Statistica Sinica* **5** 303–318. MR1329300

[21] Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *J. Econometrics* **75** 317–343.

[22] Stone, C. J. (2005). Nonparametric M-regression with free knots. *J. Statist. Plann. Inference* **130** 183–206. MR2128004

[23] Wang, Y., Lin, X., Zhu, M., and Bai, Z. (2007). Robust estimation using the Huber function with a data-dependent tuning constant. *Journal of Computational and Graphical Statistics* **16** 468–481. MR2370950

[24] Yu K. and Moyeed R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54** 437–447. MR1861390

[25] Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association* **96** 247–259. MR1952735

Colin Chen
SAS Institute Inc.
Cary, NC 27513
USA
E-mail address: chennorthc@yahoo.com