# Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density

Han Lin Shang[1,*]

*University of Southampton*

**Abstract**

Error density estimation in a nonparametric functional regression model with functional predictor and scalar response is considered. The unknown error density is approximated by a mixture of Gaussian densities with means being the individual residuals, and variance as a constant parameter. This proposed mixture error density has a form of a kernel density estimator of residuals, where the regression function is estimated by the functional Nadaraya-Watson estimator. A Bayesian bandwidth estimation procedure that can simultaneously estimate the bandwidths in the kernel-form error density and the functional Nadaraya-Watson estimator is proposed. A kernel likelihood and posterior for the bandwidth parameters are derived under the kernel-form error density. A series of simulation studies show that the proposed Bayesian estimation method performs on par with the functional cross validation for estimating the regression function, but it performs better than the likelihood cross validation for estimating the regression error density. The proposed Bayesian procedure is also applied to a nonparametric functional regression model, where the functional predictors are spectroscopy wavelengths and the scalar responses are fat/protein/moisture content, respectively.

*Keywords:* functional Nadaraya-Watson estimator; kernel density estimation; Markov chain Monte Carlo; mixture error density; spectroscopy.

*ESRC Centre for Population Change, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom. Tel: +44 (0)2380 595 796, Fax: +44 (0)2380 593 858, Email: H.Shang@soton.ac.uk

## 1. Introduction

Functional regression models describe the relationship between the predictor and response variables, where at least one variable is functional in nature. The first functional formulation of a linear model dates back to a discussion by Hastie and Mallows (1993), and it is later extended in detail by Ramsay and Silverman (2005). Since then, functional linear regression model has been further extended or modified to take into account possible nonlinear relationship, some of the regression models include the functional polynomial regression model (Yao and Müller, 2010; Horváth and Reeder, 2012), functional additive regression model (Müller and Yao, 2008; Febrero-Bande and González-Manteiga, 2013; Fan and James, 2013), and nonparametric functional regression model (Ferraty and Vieu, 2006; Ferraty, Van Keilegom, and Vieu, 2010). Due to the fast development in functional regression models, it has gained an increasing popularity in various fields of application, such as atmospheric radiation (Hlubinka and Prchal, 2007), chemometrics (Frank and Friedman, 1993; Ferraty and Vieu, 2002; Burba et al., 2009; Yao and Müller, 2010), climate variation forecasting (Shang and Hyndman, 2011), demographic modeling and forecasting (Hyndman and Ullah, 2007; Hyndman and Booth, 2008; Hyndman and Shang, 2009; Chiou and Müller, 2009), earthquake modeling (Quintela-del-Río et al., 2011), gene expression (Yao et al., 2005a; Chiou and Müller, 2007), health science (Harezlak, Coull, Laird, Magari, and Christiani, 2007), linguistics (Hastie et al., 1995; Malfait and Ramsay, 2003; Aston et al., 2010), medical research (Ratcliffe et al., 2002; Yao et al., 2005b; Erbas et al., 2007), ozone level prediction (Quintela-del-Río and Francisco-Fernández, 2011), and sulfur dioxide level prediction (Fernandez de Castro et al., 2005).

Despite the fast development in functional regression models for finding the relationship between predictor and response variables, error density estimation in functional regression models remains largely unexplored. However, the estimation of error density is important to understand the residual behavior and to assess the adequacy of error distribution assumption (see for example, Akritas and Van Keilegom, 2001; Cheng and Sun, 2008); the estimation of error density is also useful to test the symmetry of the residual distribution (see for example, Ahmad and Li, 1997; Dette et al., 2002; Neumeyer and Dette, 2007); the estimation of error density is important to statistical inference, prediction and model validation (see for example, Efromovich, 2005; Muhsal and Neumeyer, 2010); and the estimation of error density is also useful for the

estimation of the density of the response variable (see for example, Escanciano and Jacho-Chávez, 2012). In the realm of financial asset return, an important use of the estimated error density is to estimate value-at-risk for holding an asset. In such a model, any wrong specification of the error density may produce an inaccurate estimate of value-at-risk and make the asset holder unable to control risk. Therefore, being able to estimate the error density is as important as being able to estimate the regression function.

This motivates the investigation of a kernel-form error density for estimating unknown error density in a nonparametric functional regression model with functional predictors and scalar responses. This kernel-form error density depends on three parameters: 1) the type of semi-metric used to measure distances among functions, such as semi-metric based on second-order derivative; 2) residuals fitted through the functional Nadaraya-Watson (NW) estimator of the regression function; 3) bandwidth of residuals. Cheng (2002, 2004) studied weak and strong uniform consistency of such an error density estimator, while Samb (2011) established the optimal convergence rate of the kernel-form error density estimator in a multivariate framework. In this paper, we aim to develop a Bayesian bandwidth estimation procedure to simultaneously estimate the bandwidths in the functional NW estimator of the regression function and the kernel-form error density.

## 2. Bayesian bandwidth estimation

Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ be a vector of scalar responses, and $\boldsymbol{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n)^\top$ be a set of functional predictors. We consider a simple nonparametric functional regression model with homoscedastic errors. Given observations $(y_i, \mathcal{X}_i)_{i=1,2,\ldots,n}$, the model can be expressed as

$$y_i = m(\mathcal{X}_i) + \varepsilon_i, \tag{1}$$

where $m(\mathcal{X}_i) = \mathrm{E}(y|\mathcal{X})$ is the conditional mean, and $\varepsilon_i$ for $i = 1, 2, \ldots, n$ are assumed to be independent and identically distributed (iid) with an unknown error density, denoted as $f(\varepsilon)$. We assume that there is no correlation between the regression function and errors. In this paper, we investigate the problem of nonparametric estimation of the probability density function of the error term. As noted by Samb (2011), the difficulty of estimating error density is the fact that the regression error term is not observed and thus must be estimated.

There is a growing literature on the development of nonparametric functional estimators, such as functional NW estimator (Ferraty and Vieu, 2006), functional local linear estimator (Barrientos-Marin et al., 2010), functional $k$-nearest neighbour estimator (Burba et al., 2009), and distance-based local linear estimator (Boj et al., 2010). In this paper, we demonstrate the idea by using the functional NW estimator because of its simplicity and mathematical elegance. For a detailed exposition on functional NW estimator, consult Ferraty and Vieu (2006, Section 5.4).

In the functional NW estimator, the estimation accuracy of regression function is mainly determined by two parameters; type of semi-metric and bandwidth. While the semi-metric measures the distances among functions, the bandwidth measures the amount of smoothing. The optimal selections of these two parameters were two open questions given in Ferraty and Vieu (2006, p.193).

Since our simulated and real data are quite smooth, we chose a semi-metric based on second derivative. For a non-smooth functional data set, a semi-metric based on functional principal component analysis is advocated (see Ferraty and Vieu, 2006, Chapters 3 and 13 for detail on the choice of semi-metric from the practical and theoretical aspects, respectively). Having determined the type of semi-metric, the only unknown parameter in the functional NW estimator is the bandwidth (also known as smoothing parameter). As it is always the case in nonparametric estimation, the role of smoothing parameter is prominent. For example, the rates of convergence of the nonparametric functional estimator can be divided into two parts: a squared bias component which increases with the bandwidths, and a variance component which decreases with the bandwidths. Therefore, there is a need to select an optimal bandwidth in order to balance the trade-off between squared bias and variance.

In the literature of nonparametric functional data analysis, the bandwidth is commonly selected by a functional version of cross validation (CV) (see for example, Benhenni et al., 2007; Rachdi and Vieu, 2007). It is designed to assess the predictive performance of a model by an average of certain measures for the ability of predicting a subset of functions by a model fit, after deleting just these functions from the data set. Functional CV has the appealing feature that no estimation of the error variance is required. However, since residuals affect the estimation accuracy of regression function, functional CV may select sub-optimal bandwidths. This in turn leads to inferior estimation accuracy of regression function. As an alternative, we present a Bayesian

bandwidth estimation method that simultaneously estimates the bandwidths in the regression function and kernel-form error density.

### *2.1. Estimation of error density*

The unknown error density can be estimated differently, such as by using finite mixture models, log-spline approach or wavelet expansion (Schellhase and Kauermann, 2012). Here, we assume that the unknown error density $f(\varepsilon)$ can be approximated by a mixture of Gaussian densities (see also Roeder and Wasserman, 1997). Using such a mixture, any density on the real line can be approximated to within any preassigned accuracy in the $L_1$ norm (Ferguson, 1983). Concretely, the unknown error density is estimated by a location-mixture Gaussian density, given by

$$f(\varepsilon; b) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{b}\phi\left(\frac{\varepsilon - \varepsilon_j}{b}\right), \tag{2}$$

where $\phi(\cdot)$ is the probability density function of the standard Gaussian distribution, and the component Gaussian densities have means at $\varepsilon_j$, for $j = 1, 2, \ldots, n$, and a common standard deviation $b$. Note that our proposed kernel-form error density has only one bandwidth parameter to estimate, which is its main advantage over the scale-mixture Gaussian density. Although $m(\mathcal{X})$ is unknown, it can be estimated by the functional NW estimator. As a result, the density of $y_i$ is approximated by the estimated error density $\widehat{f}(\varepsilon; b_n)$, expressed as

$$f(\varepsilon; b) \approx \widehat{f}(\varepsilon; b_n) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{b_n}\phi\left(\frac{\varepsilon - \widehat{\varepsilon}_j}{b_n}\right), \tag{3}$$

where $b_n$ represents the estimate of residual bandwidth. As noted by Samb (2011), the kernel estimator $\widehat{f}(\varepsilon; b_n)$ is a feasible estimator in the sense that it does not depend on any unknown quantity, unlike (2).

Jaki and West (2008, p.989) and Jaki and West (2011) also proposed to approximate the error density by a kernel density estimator given in (3), and estimate parameters by maximizing the so-called kernel likelihood. The kernel likelihood of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ is essentially the product of the density given by (3). However, it is impossible to estimate $b$ by maximizing such a likelihood, because it contains at least one unwanted term $\phi(0)/b_n$. The

likelihood would approach infinity as $b_n$ tends to zero. To address this issue, a leave-one-out version of the kernel likelihood is used, and given by

$$\widehat{f}(\widehat{\varepsilon}_i; b_n) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{b_n} \phi\left(\frac{\widehat{\varepsilon}_i - \widehat{\varepsilon}_j}{b_n}\right),$$

where $\widehat{\varepsilon}_i = y_i - \widehat{m}(\mathcal{X}_i; h_n)$ is the $i^{\text{th}}$ residual for $i = 1, 2, \ldots, n$, and $h_n$ represents the bandwidth estimate in the functional NW estimator. Given $(h_n, b_n)$ and iid assumption of the errors, the kernel likelihood of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ can be approximated by

$$\widehat{L}(\boldsymbol{y}|h_n, b_n) = \prod_{i=1}^{n} \left[ \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{b_n} \phi\left(\frac{\widehat{\varepsilon}_i - \widehat{\varepsilon}_j}{b_n}\right) \right].$$

We now discuss the issue of prior density for the bandwidths. Let $\pi(h^2)$ and $\pi(b^2)$ be the prior of squared bandwidths $h$ and $b$. Since $h^2$ and $b^2$ play the same role as a variance parameter in the Gaussian density, we assume that the priors of $h^2$ and $b^2$ are inverse Gamma density, denoted as $\text{IG}(\alpha_h, \beta_h)$ and $\text{IG}(\alpha_b, \beta_b)$, respectively. Therefore, the prior densities of $h^2$ and $b^2$ are given by

$$\pi(h^2) = \frac{(\beta_h)^{\alpha_h}}{\Gamma(\alpha_h)} \left(\frac{1}{h^2}\right)^{\alpha_h+1} \exp\left(-\frac{\beta_h}{h^2}\right),$$
$$\pi(b^2) = \frac{(\beta_b)^{\alpha_b}}{\Gamma(\alpha_b)} \left(\frac{1}{b^2}\right)^{\alpha_b+1} \exp\left(-\frac{\beta_b}{b^2}\right),$$

where $\alpha_h = \alpha_b = 1.0$ and $\beta_h = \beta_b = 0.05$ are hyperparameters. Notice that IG(1,0.05) has previously been used as a prior density in Geweke (2010).

*2.1.1. Posterior sampler*

According to Bayes theorem, the posterior of $h_n^2$ and $b_n^2$ is approximated by (up to a normalizing constant)

$$\pi(h_n^2, b_n^2|\boldsymbol{y}) \propto \widehat{L}(\boldsymbol{y}|h_n^2, b_n^2)\pi(h^2)\pi(b^2), \tag{4}$$

where $\widehat{L}(\boldsymbol{y}|h_n^2, b_n^2)$ is the approximate likelihood function with squared bandwidths. Since we assume that there is no correlation between the regression function and error density in (1), the bandwidths of the regression function and error density are uncorrelated in (4).

In line with the Bayesian paradigm, statistical inference is drawn from the posterior, which can be analytically intractable especially in the case of multiple parameters. However, if we can sample the parameters from the posterior, statistical inference about the parameters can be obtained using the Monte Carlo method. The Markov chain Monte Carlo (MCMC) method provides a general mechanism to sample the parameters from its posterior density. In essence, the MCMC method sets up a Markov chain so that its stationary distribution is the same as the posterior density. As the Markov chain converges, the simulated realizations are treated as samples from the posterior. Because of its mathematical properties, the MCMC strategy has proved useful in many statistical applications and has many advantages over classical methods (see a survey article by Geweke, 1999). Gilks et al. (1996) presented a collection of papers on the application of MCMC algorithms, while Robert and Casella (2010) presented the theoretical underpinnings of MCMC methods.

From (4), we use the adaptive block random-walk Metropolis algorithm of Garthwaite et al. (2010) to sample $(h_n^2, b_n^2)$, the sampling algorithm is briefly described below. For simplicity of notation, I shall let $\boldsymbol{\theta}_n = (h_n^2, b_n^2)$ to represent a vector of the squared bandwidths.

---

Algorithm

---

Step 0 Specify a Gaussian proposal distribution, with an arbitrary starting point $\boldsymbol{\theta}_n^{(0)} \sim U(0, 1)$.

Step 1 At the $k^{\text{th}}$ iteration, the current state $\boldsymbol{\theta}_n^{(k)}$ is updated as $\boldsymbol{\theta}_n^{(k)} = \boldsymbol{\theta}_n^{(k-1)} + \tau^{(k-1)}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{I})$, and $\tau^{(k-1)}$ is an adaptive tuning parameter with an arbitrary initial value $\tau^{(0)}$.

Step 2 The updated $\boldsymbol{\theta}_n^{(k)}$ is accepted with probability $\min\left\{\frac{\pi\left(\boldsymbol{\theta}_n^{(k)}|\boldsymbol{y}\right)}{\pi\left(\boldsymbol{\theta}_n^{(k-1)}|\boldsymbol{y}\right)}, 1\right\}$, where $\pi$ symbolizes the posterior density.

Step 3 Using the stochastic search algorithm of Robbins and Monro (1951),

the tuning parameter is set

$$\tau^{(k)} = \begin{cases} \tau^{(k-1)} + c(1-p)/k & \text{if } \boldsymbol{\theta}_n^{(k)} \text{ is accepted} \\ \tau^{(k-1)} - cp/k & \text{if } \boldsymbol{\theta}_n^{(k)} \text{ is rejected} \end{cases},$$

where $c = \frac{\tau^{(k-1)}}{p(1-p)}$ is a fixed constant, and $p = 0.234$ is the optimal acceptance probability for drawing multiple parameters (Roberts and Rosenthal, 2009).

Step 4 Repeat Steps 1-3 for $M + N$ times, discard $\left(\theta_n^{(0)}, \theta_n^{(1)}, \ldots, \theta_n^{(M)}\right)$ for burn-in in order to let the effects of the transients wear off, estimate $\widehat{h}_n = \frac{\sum_{k=M+1}^{M+N} h_n^{(k)}}{N}$ and $\widehat{b}_n = \frac{\sum_{k=M+1}^{M+N} b_n^{(k)}}{N}$. The analytical form of the kernel-form error density can be derived based on $\widehat{h}_n$ and $\widehat{b}_n$. Note that a similar result can be obtained by taking the average of the kernel-form error densities for all iterations, but at the cost of slower computational speed.

---

*2.1.2. Diagnostic checking*

In the implementation of the MCMC algorithm, the sample path $\eta^{(i)} = \left[h_n^{(i)}\right]^2$ or $\eta^{(i)} = \left[b_n^{(i)}\right]^2$ for $i = 1, \ldots, N$ forms a Markov chain, whose stationary density is the posterior $\pi(\eta|\boldsymbol{y})$. The sample estimate is summarized by the ergodic averages in the form of

$$\bar{\eta} = \frac{1}{N} \sum_{i=1}^{N} \eta^{(i)}.$$

Roberts (1996) pointed out that most Markov chains produced in MCMC converge geometrically to the stationary distribution $\pi(\eta|\boldsymbol{y})$, and a main consequence of geometric convergence is the central limit theorem, i.e.,

$$\sqrt{N}[\bar{\eta} - \mathrm{E}_\pi(\eta)] \xrightarrow{D} N(0, \sigma^2), \tag{5}$$

where $\mathrm{E}_\pi(\cdot)$ denotes the expectation operator under $\pi(\eta|\boldsymbol{y})$. From (5), the sample average converges in distribution to the true posterior density. To assess the accuracy of ergodic average as an estimate of $\mathrm{E}_\pi(\eta)$, it is essential to estimate $\sigma^2$. One of the most commonly used methods for estimating $\sigma^2$ is the batch mean (Roberts, 1996).

To estimate $\sigma^2$ using the batch mean, the MCMC algorithm is run for $N = m \times n$ iterations, where $m$ is the number of batches and $n$ is the batch sample size. Thus, $\sigma^2$ can be estimated by

$$\widehat{\sigma}^2 = \frac{n}{m-1} \sum_{p=1}^{m} \left( \frac{1}{n} \sum_{i=(p-1)n+1}^{pn} [\eta^{(i)} - \bar{\eta}] \right)^2 ,$$

and the standard error (SE) of $\bar{\eta}$ can be estimated by $\sqrt{\widehat{\sigma}^2/N}$, which is known as the batch-mean SE (Roberts, 1996).

Apart from the batch-mean SE, one may also compute the SE ($\tilde{\sigma}$) based on the sample path using the formula

$$\tilde{\sigma} = \left\{ \frac{1}{N-1} \sum_{i=1}^{N} [\eta^{(i)} - \bar{\eta}]^2 \right\}^{\frac{1}{2}} .$$

Kim et al. (1998), Meyer and Yu (2000) and Tse et al. (2004) noted that the mixing performance of the sample paths can be measured by simulation inefficiency factor (SIF), which is also known as the integrated autocorrelation time by Berg (2005). It is estimated as the sample mean from a sampler that draws iid observations from the posterior distribution, SIF is given by $\widehat{\sigma}^2/\tilde{\sigma}^2$. In the following analyses, the burn-in period is taken as $M = 1,000$ iterations and the number of recorded iterations after the burn-in period is $N = 10,000$ iterations. The number of batches is $m = 200$, and there are $n = 50$ draws within each batch.

*2.1.3. Adaptive estimation of error density*

In kernel density estimation of directly observed data, it has been observed that the leave-one-out estimator is heavily affected by extreme observations in the sample (see for example, Bowman, 1984; Zhang and King, 2011). When the true error density has sufficient long tails, the leave-one-out kernel density estimator with its bandwidth estimated under the Kullback-Leibler criterion, is likely to overestimate the tails of the density. Such a phenomenon is likely to be caused by the use of a global bandwidth. A solution to this problem is to use localized bandwidth (see for example, Zhang and King, 2011, in the context of GARCH model).

The idea of the localized bandwidth is to assign small bandwidths to the observations in the high density region, and large bandwidths to the

observations in the low density region. One key issue is to choose different bandwidths for different groups of observations. Following the work by Zhang and King (2011), large absolute errors should be assigned relatively large bandwidths, while small absolute errors should be assigned relatively small bandwidths. Differing from Zhang and King (2011), we extend the localized bandwidth idea from the multivariate to functional setting, where the functional NW estimator is used instead of the multivariate NW estimator. The localized error density estimator can be expressed by

$$\widehat{f}\left(\widehat{\varepsilon}_i;\tau,\tau_\varepsilon\right)=\frac{1}{n-1}\sum_{\substack{j=1\\j\neq i}}^{n}\frac{1}{\tau\left(1+\tau_\varepsilon|\widehat{\varepsilon}_j|\right)}\phi\left(\frac{\widehat{\varepsilon}_i-\widehat{\varepsilon}_j}{\tau(1+\tau_\varepsilon|\widehat{\varepsilon}_j|)}\right), \tag{6}$$

where $\tau(1+\tau_\varepsilon|\widehat{\varepsilon}_j|)$ is the bandwidth assigned to $\widehat{\varepsilon}_j$, for $j=1,2,\ldots,n$, and the vector of parameters is now $(h_n,b_n,\tau,\tau_\varepsilon)$. The error density given in (6) can be interpreted as a mixture of $n-1$ Gaussian densities with their means being at the other errors and variances localized.

*2.1.4. Two-stage cross validation*

Despite the rapid development in estimating regression function, there is little work on the estimation of error density in functional regression models. Nonetheless, in econometrics literature, Engle and González-Rivera (1991) proposed a two-stage estimation procedure; the first stage uses the quasi-maximum likelihood estimator to obtain the residuals, from which error density is constructed in the second stage using a nonparametric density estimator. In statistics literature, Samb (2011) put forward a two-stage bandwidth estimator, where the regression function is estimated by the NW estimator in a multivariate regression model, and the regression error density is estimated by a univariate kernel error density. Here, we aim to extend Samb's (2011) work to a nonparametric functional regression model, and also consider a two-stage CV method as a competing method for separately estimating regression function and regression error density. Concretely, we apply the functional CV for selecting the bandwidth in the functional NW estimator and obtain a vector of real-valued residuals; based on these residuals, we then apply the likelihood CV (Bowman, 1984) to obtain the optimal bandwidth for a univariate kernel error density estimator. The asymptotic optimality of the bandwidth selected by the likelihood CV has been studied by Hall (1987) and van der Laan et al. (2004).

## 3. Simulation study

The main goal of this section is to illustrate the methodology through simulated data. One way to do that consists in comparing the true error density $f(\varepsilon)$ with the estimated one $\widehat{f}(\varepsilon)$. To measure the discrepancy between $f(\varepsilon)$ and $\widehat{f}(\varepsilon)$, we use integrated squared error (ISE) criterion, defined by $\int_a^b [f(\varepsilon) - \widehat{f}(\varepsilon)]^2 d\varepsilon$ for $\varepsilon \in [a, b]$. In practice, ISE can be approximated at 1001 grid points bounded between an interval, such as $[-5, 5]$. This can be given by

$$\text{ISE} \approx \frac{1}{100} \sum_{i=1}^{1001} \left[ f\left(-5 + \frac{(i-1)}{100}\right) - \widehat{f}\left(-5 + \frac{(i-1)}{100}\right) \right]^2. \tag{7}$$

Then, we estimate the mean integrated squared error MISE=$\text{E}\{\int_a^b [f(\varepsilon) - \widehat{f}(\varepsilon)]^2 d\varepsilon\}$ by the average of these integrated squared errors in (7) over 100 replications.

*Building the simulated samples.* First of all, we build simulated discretized curves:

$$\mathcal{X}_i(t_j) = a_i \cos(2t_j) + b_i \sin(4t_j) + c_i(t_j^2 - \pi t_j + \frac{2}{9}\pi^2), \quad i = 1, 2, \ldots, n, \tag{8}$$

where $0 \leq t_1 \leq t_2 \cdots \leq t_{100} \leq \pi$ are equispaced points, $a_i, b_i, c_i$ are independently drawn from a uniform distribution on $[0, 1]$, and $n$ represents the sample size. The functional form of (8) is taken from Ferraty et al. (2010). Figure 1 presents the simulated curves for one replication.

Once the curves are defined, we then simulate a nonparametric functional regression model to compute the responses in the following steps:

- construct a regression function operator $m$, which performs the mapping from function-valued space to real-valued space. Two functional operators were considered and they are expressed as

$$\begin{cases} \text{Model 1:} & m(\mathcal{X}_i) = 10 \times (a_i^2 - b_i^2); \\ \text{Model 2:} & m(\mathcal{X}_i) = \int_0^\pi t cos(t) \left(\mathcal{X}_i'(t)\right)^2 dt. \end{cases}$$

- generate $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, which are independent centered Gaussian of variance equal to 0.1, 0.5, or 0.9 times the empirical variance of $\{m(\mathcal{X}_1), m(\mathcal{X}_2), \ldots, m(\mathcal{X}_n)\}$ (i.e., signal-to-noise ratio $(\xi) = 0.1$, 0.5, or 0.9)

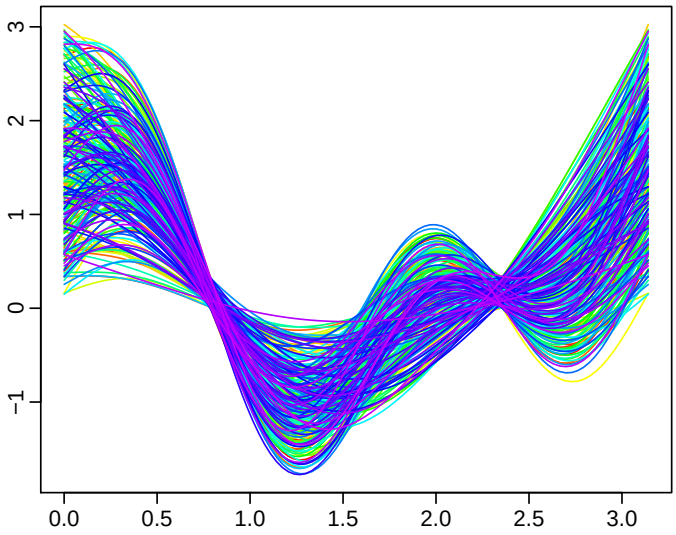



Figure 1: 250 simulated curves.

- compute the corresponding responses: $y_i = m(\mathcal{X}_i) + \varepsilon_i$, for $i = 1, 2, \ldots, n$.

*Estimating the regression function.* For a fixed curve $\mathcal{X}$ and a fixed bandwidth $h$, we compute the in-sample discrepancy between $m(\mathcal{X})$ and $\widehat{m}(\mathcal{X})$. To do that, we use the following Monte-Carlo scheme:

- build 100 replications: $\{(\mathcal{X}_i^s, y_i^s)_{i=1,\ldots,n}\}_{s=1,\ldots,100}$;
- compute 100 estimates $\{m(\mathcal{X}) - \widehat{m}_h^s(\mathcal{X})\}_{s=1,\ldots,100}$, where $\widehat{m}_h^s$ is the functional NW estimator of the regression function computed over the $s^{\text{th}}$ replication;
- obtain the mean squared error (MSE) by averaging over 100 replications of the squared errors.

Table 1 presents MSE for the functional NW estimator with bandwidths selected by functional CV, Bayesian methods with global bandwidth and localized bandwidth for both models. For a small sample size ($n = 50$), functional CV performs the best in estimating the regression function, resulting in smallest overall MSE and the standard deviation (sd) of the squared errors. As sample size $n$ increases, the difference in estimation accuracy between the functional CV and Bayesian methods is marginal. In some cases, the Bayesian methods with the global and localized bandwidths perform

better than the functional CV. As the signal-to-noise ratio ($\xi$) increases, the regression function becomes harder to estimate accurately for all the methods.

| | Functional CV | | | Bayesian | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Global bandwidth | | | Local bandwidth | | |
| $\xi$ \ $n$ | 50 | 250 | 1000 | 50 | 250 | 1000 | 50 | 250 | 1000 |
| Model 1 | | | | | | | | | |
| $\xi = 0.1$ | 2.9532 | 0.9807 | 0.6299 | 3.0224 | 0.9514 | 0.3847 | 3.0217 | 0.9510 | 0.3847 |
| | (0.6050) | (0.1101) | (0.0541) | (0.6447) | (0.1045) | (0.0343) | (0.6441) | (0.1038) | (0.0343) |
| $\xi = 0.5$ | 4.6516 | 1.6410 | 0.7924 | 4.7784 | 1.6576 | 0.7357 | 4.7791 | 1.6575 | 0.7364 |
| | (1.0667) | (0.2496) | (0.1053) | (1.1396) | (0.2513) | (0.1039) | (1.1376) | (0.2515) | (0.1041) |
| $\xi = 0.9$ | 6.0889 | 2.1334 | 0.9629 | 6.2438 | 2.1509 | 0.9715 | 6.2347 | 2.1504 | 0.9715 |
| | (1.6021) | (0.3597) | (0.1557) | (1.7739) | (0.3658) | (0.1609) | (1.7616) | (0.3659) | (0.1608) |
| Model 2 | | | | | | | | | |
| $\xi = 0.1$ | 16.3007 | 5.7414 | 4.2912 | 16.6050 | 4.9041 | 1.8208 | 16.5489 | 4.9019 | 1.8205 |
| | (3.2026) | (0.4822) | (0.2290) | (3.3316) | (0.5900) | (0.1362) | (3.3183) | (0.5850) | (0.1361) |
| $\xi = 0.5$ | 20.9213 | 7.5627 | 4.6785 | 21.4262 | 7.5131 | 3.1856 | 21.3503 | 7.5137 | 3.1854 |
| | (3.9197) | (0.8952) | (0.4056) | (4.2405) | (0.9140) | (0.3221) | (4.1461) | (0.9114) | (0.3220) |
| $\xi = 0.9$ | 24.7496 | 9.4166 | 5.0762 | 25.0736 | 9.5005 | 4.1242 | 25.0492 | 9.5005 | 4.1240 |
| | (4.9474) | (1.3196) | (0.5408) | (5.1441) | (1.3119) | (0.4653) | (5.0993) | (1.3119) | (0.4650) |

Table 1: MSE comparison between the functional CV and Bayesian methods for estimating the regression function. The number in parenthesis represents the sample sd of the squared errors. The red colored text represents the minimal MSE, while the blue colored text represents the minimal sd of the squared errors.

*Estimating the error density.* With a set of residuals and a fixed residual bandwidth, one can apply a univariate kernel density estimator and compute the discrepancy between $f(\varepsilon)$ and $\widehat{f}(\varepsilon)$. To do that, one uses the following Monte-Carlo scheme:

- compute 100 replications of residuals $\{y_i^s - \widehat{m}_h^s(\mathcal{X}_i)\}_{s=1,\ldots,100}$;
- apply a univariate kernel density to estimate error density, where the residual bandwidths are estimated by either the likelihood CV (Bowman, 1984) or the Bayesian methods for 100 replications;
- for $s = 1, 2, \ldots, 100$, compute the MISE between the true error density $f^s(\varepsilon)$ and estimated error density $\widehat{f}^s(\varepsilon)$;

- obtain the overall discrepancy by averaging over 100 replications of discrepancy.

Table 2 presents MISE for the kernel-form error density with bandwidth estimated by likelihood CV, Bayesian methods with global bandwidth and localized bandwidth. The Bayesian methods perform uniformly better than the likelihood CV, which is the second-stage of the two-stage CV. Between the two Bayesian methods, there is an advantage in using the localized bandwidth over the global bandwidth, especially for small sample size.

| | likelihood CV | | | Bayesian | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Global bandwidth | | | Local bandwidth | | |
| $\xi$ \ $n$ | 50 | 250 | 1000 | 50 | 250 | 1000 | 50 | 250 | 1000 |
| Model 1 | | | | | | | | | |
| $\xi = 0.1$ | 0.0651 | 0.0186 | 0.0091 | 0.0413 | 0.0094 | 0.0030 | 0.0392 | 0.0090 | 0.0029 |
| | (0.0129) | (0.0069) | (0.0044) | (0.0111) | (0.0027) | (0.0008) | (0.0105) | (0.0025) | (0.0007) |
| $\xi = 0.5$ | 0.1360 | 0.1138 | 0.0965 | 0.0086 | 0.0019 | 0.0006 | 0.0078 | 0.0018 | 0.0006 |
| | (0.0126) | (0.0130) | (0.0128) | (0.0037) | (0.0008) | (0.0002) | (0.0037) | (0.0008) | (0.0002) |
| $\xi = 0.9$ | 0.1576 | 0.1430 | 0.1296 | 0.0051 | 0.0011 | 0.0003 | 0.0046 | 0.0011 | 0.0003 |
| | (0.0119) | (0.0129) | (0.0129) | (0.0027) | (0.0006) | (0.0001) | (0.0025) | (0.0006) | (0.0002) |
| Model 2 | | | | | | | | | |
| $\xi = 0.1$ | 0.1573 | 0.1074 | 0.0830 | 0.0451 | 0.0138 | 0.0037 | 0.0438 | 0.0130 | 0.0035 |
| | (0.0151) | (0.0099) | (0.0049) | (0.0091) | (0.0026) | (0.0006) | (0.0090) | (0.0022) | (0.0006) |
| $\xi = 0.5$ | 0.1792 | 0.1606 | 0.1477 | 0.0062 | 0.0016 | 0.0005 | 0.0061 | 0.0016 | 0.0005 |
| | (0.0147) | (0.0078) | (0.0045) | (0.0019) | (0.0005) | (0.0001) | (0.0019) | (0.0005) | (0.0001) |
| $\xi = 0.9$ | 0.1878 | 0.1765 | 0.1658 | 0.0030 | 0.0008 | 0.0003 | 0.0029 | 0.0008 | 0.0002 |
| | (0.0140) | (0.0074) | (0.0046) | (0.0010) | (0.0003) | (0.0001) | (0.0010) | (0.0003) | (0.0001) |

Table 2: MISE comparison between the likelihood CV and Bayesian methods for estimating the error density. The number in parenthesis represents the sample sd. The red colored text represents the minimal MISE, while the blue colored text represents the minimal sd of the ISE.

*Diagnostic check of Markov chains.* As a demonstration with one replication, we plot the MCMC sample paths of the parameters on the left panel of Figure 2, and the ACFs of these sample paths on the right panel of Figure 2.

Under model 1 with Gaussian error density and signal-to-noise ratio of 0.1, these plots show that the sample paths are mixed reasonably well. Table 3 summarizes the ergodic averages, 95% Bayesian confidence intervals (CIs), SE, batch mean SE, and SIF values.

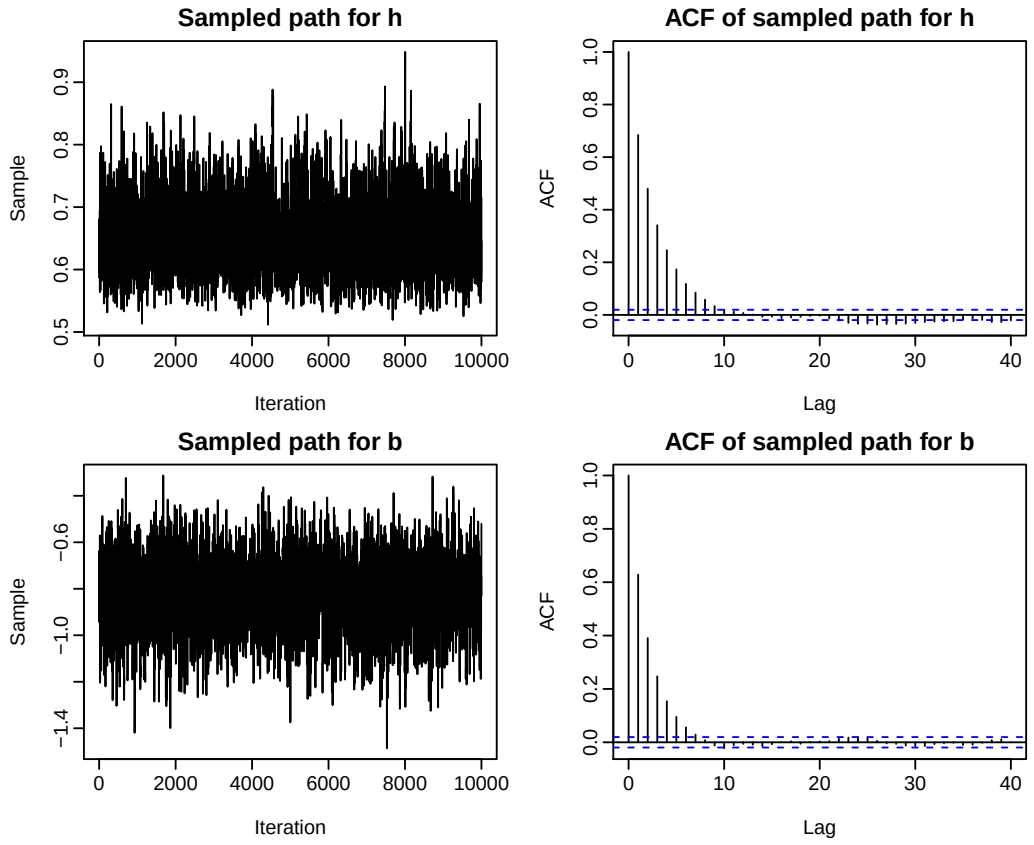



Figure 2: MCMC sample paths and ACF of the sample paths, for model 1 with Gaussian error density and signal-to-noise ratio of 0.1.

| Prior density: | $\text{IG}(\alpha = 1, \beta = 0.05)$ | | | | |
|---|---|---|---|---|---|
| Parameter | Mean | Bayesian CIs | SE | Batch-mean SE | SIF |
| $h_n$ | 1.9151 | (1.7375, 2.1936) | 0.0187 | 0.0039 | 4.76 |
| $b_n$ | 0.4419 | (0.3175, 0.5908) | 0.0924 | 0.0253 | 3.65 |

Table 3: MCMC results of the bandwidth estimation under the prior density of $\text{IG}(\alpha = 1, \beta = 0.05)$, for model 1 with Gaussian error density and signal-to-noise ratio of 0.1.

Using the coda package (Plummer et al., 2006), we further checked the convergence of Markov chain with Geweke's (1992) convergence diagnostic test and Heidelberger and Welch's (1983) convergence diagnostic test. Our Markov chains pass both tests for all 100 replications.

*Sensitivity analysis to the prior choice.* To examine the robustness of the results with respect to the choice of the priors, we change the priors in two ways. First, we keep the same prior distributions as before but alter the choice of hyperparameters. The results are very similar, as shown in Table 4. Second, we change the prior distributions from Inverse Gamma distribution to Cauchy distribution. The use of Cauchy prior for bandwidth estimation has been studied by Zhang et al. (2009). The MCMC results for the same sample reported in Table 3 are summarized in Table 4.

In comparison with the results shown in Table 3, the SIF values are comparable, suggesting that the mixing performance is not affected much by the different selections of prior density. There is also no obvious difference in the ergodic averages and 95% Bayesian CIs under both sets of priors, suggesting that the posterior distribution is robust to the change of priors.

| Prior density: | IG($\alpha = 5, \beta = 0.25$) | | | | |
|---|---|---|---|---|---|
| Parameter | Mean | Bayesian CIs | SE | Batch-mean SE | SIF |
| $h_n$ | 1.8538 | (1.7220, 2.1006) | 0.0134 | 0.0026 | 5.11 |
| $b_n$ | 0.3807 | (0.2731, 0.5271) | 0.1238 | 0.0288 | 4.30 |
| Prior density: | Cauchy($x_0 = 0, \gamma = 1$) | | | | |
| Parameter | Mean | Bayesian CIs | SE | Batch-mean SE | SIF |
| $h_n$ | 1.9399 | (1.7461 2.2412) | 0.0227 | 0.0043 | 5.28 |
| $b_n$ | 0.4795 | (0.3523 0.6363) | 0.1073 | 0.0239 | 4.50 |

Table 4: MCMC results of the bandwidth estimation under the different selections of prior density, for model 1 with Gaussian error density and signal-to-noise ratio of 0.1.

## 4. Application to food quality control

Let us consider a food quality control application, previously studied by Ferraty and Vieu (2006) and Aneiros-Pérez and Vieu (2006), among many others. The data set was obtained from `http://lib.stat.cmu.edu/datasets/tecator`. Each food sample contains finely chopped pure meat

with different percentages of the fat, protein and moisture contents. For each unit $i$ (among 215 pieces of finely chopped meat), we observe one spectrometric curve, denoted by $\mathcal{X}_i$, which corresponds to the absorbance measured at a grid of 100 wavelengths (i.e., $\mathcal{X}_i = (\mathcal{X}_i(t_1), \ldots, \mathcal{X}_i(t_{100}))$). For each unit $i$, we also observe its fat/protein/moisture content $y_i \in R$ obtained by analytical chemical processing. The data set contains the pairs $(y_i, \mathcal{X}_i)_{i=1,\ldots,215}$. Given a new spectrometric curve $\mathcal{X}$, our task is to predict the corresponding fat/protein/moisture content. As pointed out by Ferraty and Vieu (2006), the motivation is that obtaining a spectrometric curve is less time and cost consuming than the analytic chemistry needed for determining the fat/protein/moisture content. A graphical display of spectrometric curves is shown in Figure 3.

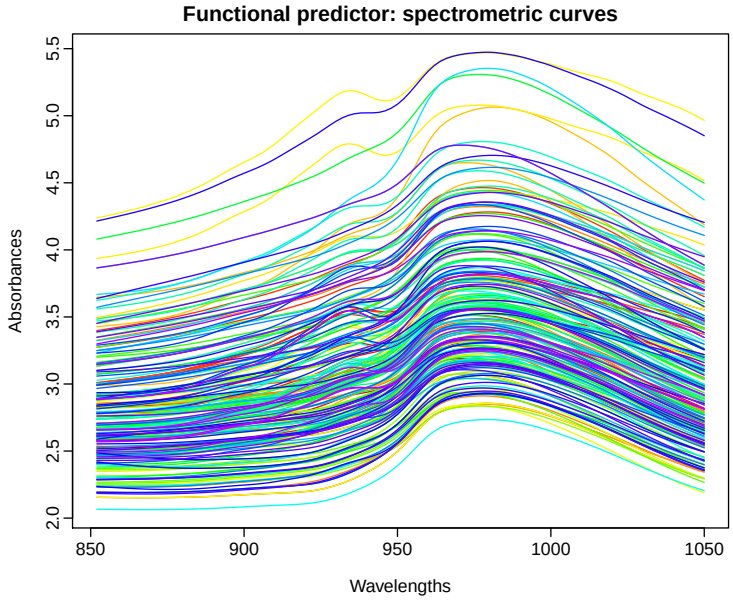



Figure 3: A Graphical display of spectrometric curves.

The first step is to study the relationship between the spectrometric curves and the corresponding fat/protein/moisture content, respectively. We use the nonparametric functional NW estimator in this paper. To assess the out-of-sample accuracy of the nonparametric functional estimator, we split the original samples into two subsamples (see also Ferraty and Vieu, 2006, p.105). The first one is called learning sample, which contains the first 160 units $\{(\mathcal{X}_i, y_i)_{i=1,\ldots,160}\}$. The second one is called testing sample, which contains the

last 55 units $\{(\mathcal{X}_i, y_i)_{i=161,\ldots,215}\}$. The learning sample allows us to build the functional NW estimator with optimal bandwidth, where the learning sample $(\mathcal{X}_i, y_i)_{i=1,\ldots,160}$ is used. To measure the prediction quality, we evaluate the functional NW estimator at the testing sample $(\mathcal{X}_{161}, \ldots, \mathcal{X}_{215})$, from which we predict responses $(y_{161}, \ldots, y_{215})$.

For comparison, we also computed the nonparametric kernel regression using the **funopare.kernel.cv** function provided in the **npfda** package. To measure the performance of each functional prediction method, we consider

(i) the distribution of the error $\text{se}_i = (y_i - \widehat{y}_i)$, for $i = 161, \ldots, 215$,

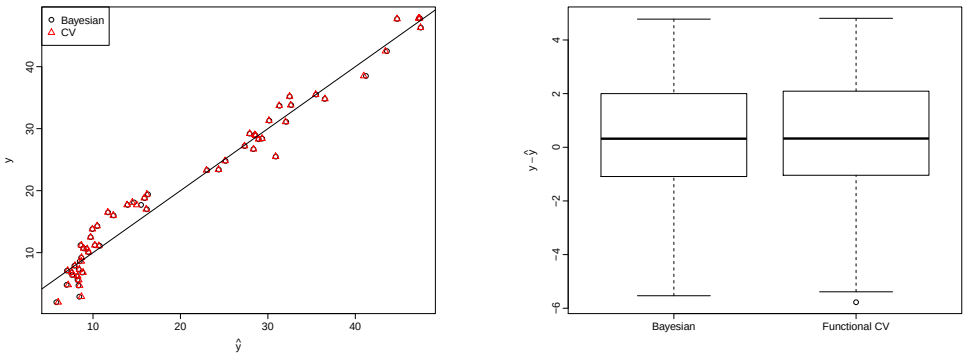
(ii) the empirical mean square prediction errors: $\text{MSPE} = \frac{1}{55}\sum_{i=161}^{215} \text{se}_i^2$.

While criterion (i) gives an indication how well each hold-out observation is predicted, criterion (ii) provides an overall error measure. The two different models used and the corresponding values of MSPE are shown in Table 5. As measured by the MSPE, there is a slight improvement in prediction accuracy for the functional NW estimator with the bandwidth selected by the Bayesian method over the functional CV method. For example, Figure 4 shows the ability of the estimated model to predict the fat content.

| | Response variable | | |
|---|---|---|---|
| Method | Fat | Protein | Moisture |
| Bayesian | 5.3097 | 2.5313 | 4.1125 |
| | (7.1241) | (7.6973) | (5.9090) |
| Functional CV | 5.3679 | 2.5417 | 4.3186 |
| | (7.4324) | (7.6842) | (6.3178) |

Table 5: Out-of-sample MSPE for the functional NW estimator with the bandwidth estimated by the Bayesian bandwidth estimation method and functional CV. The number in parenthesis represents the sample sd. The red colored text represents the minimal MSPE, while the blue colored text represents the minimal sd of the squared prediction errors.

We are also interested in computing the prediction interval nonparametrically. To this end, we first compute the cumulative density function (cdf) of the error distribution, over a set of grid points within a range, such as between -8 and 8; we then take the inverse of the cdf and find two grid points that are closest to the 2.5% and 97.5% quantiles; the 95% prediction interval of the holdout samples is obtained by adding the two grid points to the point forecast. For instance, the point forecasts of the fat content are shown as black dots in Figure 5, while the 95% prediction intervals are shown as red




(a) Plot of predicted values vs. holdout samples.

(b) Boxplot of the differences between the holdout samples and predicted values.

Figure 4: Graphical display of the criterion (i), using the functional NW estimator with two different bandwidth estimation methods.

parentheses in Figure 5.

## 5. Conclusions and some open questions

We propose a Bayesian approach to select optimal bandwidths in a nonparametric functional regression model with homoscedastic errors and unknown error density. Through a series of simulation, the Bayesian approach performs on par with the functional CV for estimating the regression function, but it is more superior to likelihood CV for estimating error density. Illustrated by a spectroscopy data set, the Bayesian bandwidth estimation approach allows the construction of nonparametric prediction interval for measuring the prediction uncertainty.

As pointed out by the two referees, there are many ways in which the proposed methodology can be extended, and we briefly mention a few at this point.

1. Apply the proposed methodology and sampling algorithm to other functional data sets, such as the ozone level prediction data studied in Quintela-del-Río and Francisco-Fernández (2011).

2. Consider other functional regression estimators, such as functional local linear kernel estimator of Benhenni et al. (2007) or $k$-nearest neighbour kernel estimator of Burba et al. (2009). The functional local linear kernel

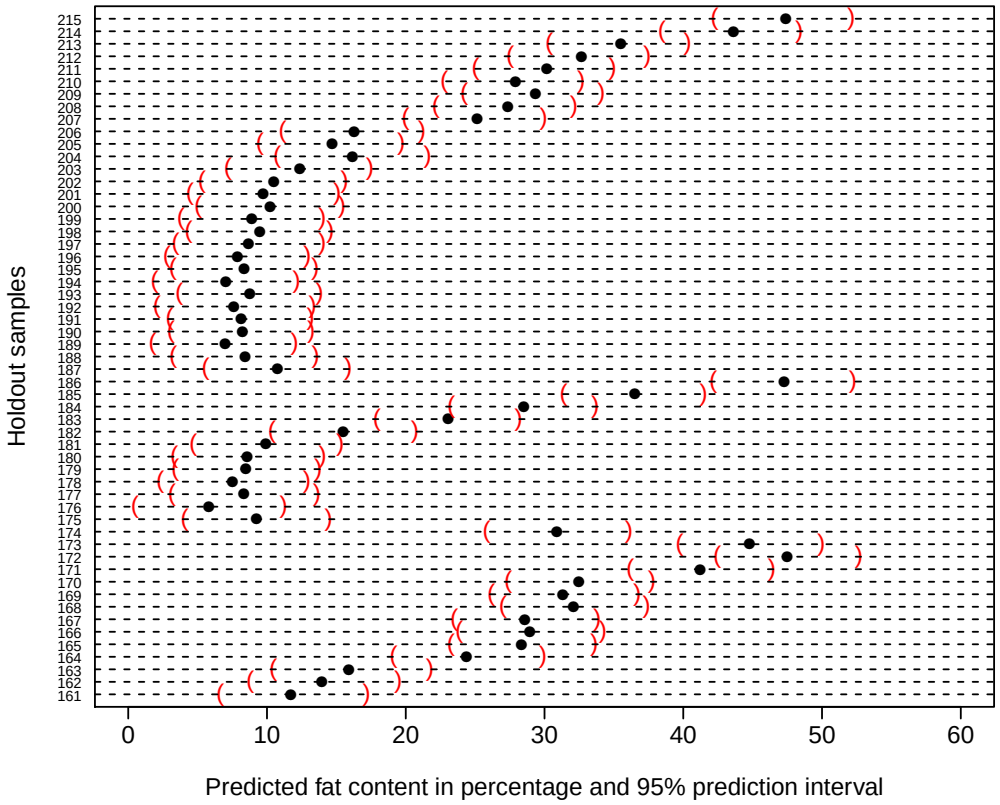



Figure 5: Plot of predicted fat contents in percentage and the 95% prediction intervals. The point forecasts of the fat content are shown as black dots, while the 95% prediction intervals are shown as red parentheses.

estimator can improve the estimation accuracy of the regression function by using a high-order kernel. The $k$-nearest neighbour kernel estimator takes into account the local structure of the data and gives better predictions when the functional data are heterogeneously concentrated.

3. Consider other bandwidth estimation methods for the kernel-form error density, such as the iterative methods proposed by Müller and Wang (1990) and Jones et al. (1991), which are based on relevant estimation of mean integrated square error.

4. Extend to nonparametric functional regression model with heteroscedastic errors. The covariate-dependent variance can be modeled by another

kernel density estimator.

5. Extend to nonparametric functional regression model with functional responses (see for example, Ferraty et al., 2011, 2012).

6. Extend to nonparametric functional regression model with dependent functional data, where the functional predictors are lagged values of the functional responses (see for example, Besse et al., 2000; Quintela-del-Río and Francisco-Fernández, 2011).

7. Extend to nonparametric functional regression model with autoregressive errors (see for example, Dabo-Niang and Guillas, 2010).

8. Extend to nonparametric functional regression model with mixed types of (function-valued, continuous real-valued and discrete-valued) regressors.

9. Extend to other functional regression models, such as functional single index regression model, functional additive regression model, and semi-functional partial linear regression model.

**Acknowledgements**

The author acknowledges the useful comments by the conference participants, especially Professor Germán Aneiros-Pérez for a valuable suggestion, at the ERCIM2012 in Oviedo, Spain. The author also thanks Professors Rob Hyndman and Donald Poskitt for introducing him to functional data analysis, and Professors Xibin Zhang and Maxwell King for introducing him to Bayesian bandwidth estimation. Special thanks also go to the Editor, an Associate Editor and two referees for their insightful suggestions and comments, which led to a much improved manuscript.

# APPENDIX: Monte-Carlo simulation results for more error densities

As a sequel of the simulation study, we also investigate the MSE and MISE of five other error densities, as shown in Tables (.6) and (.7). The first four error densities are simulated from mixtures of Gaussian densities selected from Marron and Wand (1992), while the last one is simulated from a non-Gaussian error density. The five error densities are listed below,

(1) outlier density with the functional form $\frac{1}{10}N(0,1) + \frac{9}{10}N\left(0, (\frac{1}{10})^2\right)$,
(2) separate bimodal density with the functional form $\frac{1}{2}N(-\frac{3}{2}, \frac{1}{2}) + \frac{1}{2}N(\frac{3}{2}, \frac{1}{2})$,
(3) skewed bimodal density with the functional form $\frac{3}{4}N(0,1) + \frac{1}{4}N\left(\frac{3}{2}, (\frac{1}{3})^2\right)$,
(4) claw density with the functional form $\frac{1}{2}N(0,1) + \sum_{l=0}^{4} \frac{1}{10}N\left(l/2 - 1, (\frac{1}{10})^2\right)$,
(5) student-$t$ density with five degree of freedom.

| | | Functional CV | | | Bayesian | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Global bandwidth | | | Local bandwidth | | |
| $f(\varepsilon)$ \ $n$ | | 50 | 250 | 1000 | 50 | 250 | 1000 | 50 | 250 | 1000 |
| Model 1 | | | | | | | | | | |
| (1) | | 2.4967 | 0.8093 | 0.5941 | 2.5652 | 0.6995 | 0.2377 | 2.5585 | 0.6963 | 0.2385 |
| | | (0.6026) | (0.0918) | (0.0343) | (0.6492) | (0.0966) | (0.0218) | (0.6454) | (0.0936) | (0.0227) |
| (2) | | 3.1318 | 1.0780 | 0.6480 | 3.1898 | 1.0918 | 0.4482 | 3.1889 | 1.0922 | 0.4481 |
| | | (0.6604) | (0.1481) | (0.0573) | (0.7118) | (0.1538) | (0.0482) | (0.7100) | (0.1537) | (0.0482) |
| (3) | | 2.8841 | 1.0132 | 0.7068 | 2.9625 | 0.9640 | 0.4538 | 2.9654 | 0.9639 | 0.4539 |
| | | (0.6332) | (0.1221) | (0.0228) | (0.7147) | (0.1090) | (0.0375) | (0.7100) | (0.1090) | (0.0375) |
| (4) | | 2.6737 | 0.8826 | 0.6116 | 2.7316 | 0.8038 | 0.3029 | 2.7583 | 0.8035 | 0.3025 |
| | | (0.5988) | (0.0998) | (0.0404) | (0.6627) | (0.0920) | (0.0225) | (0.6139) | (0.0919) | (0.0217) |
| (5) | | 2.8832 | 0.9794 | 0.6300 | 2.9275 | 0.9403 | 0.3805 | 2.9222 | 0.9405 | 0.3798 |
| | | (0.6401) | (0.1257) | (0.0530) | (0.6873) | (0.1185) | (0.0365) | (0.6804) | (0.1187) | (0.0361) |
| Model 2 | | | | | | | | | | |
| (1) | | 14.9589 | 5.2936 | 4.2154 | 15.2968 | 4.0553 | 1.2843 | 15.2504 | 4.0514 | 1.2817 |
| | | (3.1366) | (0.4287) | (0.1657) | (3.6445) | (0.6043) | (0.1325) | (3.5283) | (0.6049) | (0.1326) |
| (2) | | 15.7507 | 5.5577 | 4.2869 | 16.1353 | 4.5648 | 1.6307 | 16.0842 | 4.5638 | 1.6311 |
| | | (3.1294) | (0.4826) | (0.2080) | (3.4815) | (0.6346) | (0.1119) | (3.4706) | (0.6349) | (0.1120) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| (3) | 15.6360 | 5.6672 | 4.5017 | 15.6004 | 4.3772 | 1.6443 | 15.9955 | 4.5193 | 1.6441 |
| | (3.2057) | (0.4667) | (0.1815) | (3.5566) | (0.6335) | (0.1255) | (3.6205) | (0.6265) | (0.1254) |
| (4) | 15.1832 | 5.3566 | 4.2229 | 15.5061 | 4.1939 | 1.3778 | 15.4894 | 4.1921 | 1.3768 |
| | (3.2171) | (0.4455) | (0.1718) | (3.4921) | (0.5999) | (0.1276) | (3.4965) | (0.5985) | (0.1276) |
| (5) | 15.3797 | 5.4495 | 4.2601 | 15.7671 | 4.3475 | 1.5098 | 15.7402 | 4.3491 | 1.5087 |
| | (3.0393) | (0.4689) | (0.1858) | (3.3367) | (0.5992) | (0.1203) | (3.2584) | (0.6001) | (0.1203) |

Table .6: MSE comparison between the functional CV and Bayesian methods for estimating the regression function. The number in parenthesis represents the sample sd of the squared errors. The red colored text represents the minimal MSE, while the blue colored text represents the minimal sd of the squared errors.

| | likelihood CV | | | Bayesian | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Global bandwidth | | | Local bandwidth | | |
| $f(\varepsilon)$ \ $n$ | 50 | 250 | 1000 | 50 | 250 | 1000 | 50 | 250 | 1000 |
| Model 1 | | | | | | | | | |
| (1) | 2.0776 | 1.7451 | 1.8712 | 2.0637 | 1.6605 | 1.1770 | 2.0453 | 1.5902 | 1.0830 |
| | (0.0371) | (0.0456) | (0.0636) | (0.0667) | (0.0523) | (0.0483) | (0.0710) | (0.0775) | (0.0514) |
| (2) | 0.1557 | 0.1069 | 0.0598 | 0.1350 | 0.0837 | 0.0394 | 0.1339 | 0.0829 | 0.0394 |
| | (0.0088) | (0.0066) | (0.0054) | (0.0132) | (0.0097) | (0.0039) | (0.0120) | (0.0097) | (0.0037) |
| (3) | 0.1021 | 0.0435 | 0.0365 | 0.0798 | 0.0387 | 0.0285 | 0.0774 | 0.0382 | 0.0285 |
| | (0.0129) | (0.0032) | (0.0014) | (0.0129) | (0.0027) | (0.0019) | (0.0124) | (0.0027) | (0.0018) |
| (4) | 0.1770 | 0.0838 | 0.0659 | 0.1418 | 0.0757 | 0.0554 | 0.1466 | 0.0739 | 0.0546 |
| | (0.0173) | (0.0063) | (0.0017) | (0.0173) | (0.0050) | (0.0015) | (0.0152) | (0.0047) | (0.0013) |
| (5) | 0.0615 | 0.0242 | 0.0146 | 0.0608 | 0.0228 | 0.0096 | 0.0590 | 0.0180 | 0.0044 |
| | (0.0179) | (0.0087) | (0.0062) | (0.0176) | (0.0084) | (0.0046) | (0.0172) | (0.0075) | (0.0019) |
| Model 2 | | | | | | | | | |
| (1) | 2.2867 | 2.1694 | 2.1466 | 2.2421 | 2.0898 | 1.8438 | 2.2397 | 2.0579 | 1.7960 |
| | (0.0099) | (0.0113) | (0.0098) | (0.0238) | (0.0226) | (0.0249) | (0.0240) | (0.0400) | (0.0360) |
| (2) | 0.2197 | 0.1630 | 0.1426 | 0.1838 | 0.1366 | 0.0958 | 0.1819 | 0.1366 | 0.0950 |
| | (0.0086) | (0.0064) | (0.0015) | (0.0100) | (0.0057) | (0.0049) | (0.0098) | (0.0061) | (0.0049) |
| (3) | 0.1966 | 0.1144 | 0.0792 | 0.1451 | 0.0480 | 0.0360 | 0.1498 | 0.0666 | 0.0352 |
| | (0.0088) | (0.0092) | (0.0035) | (0.0144) | (0.0095) | (0.0021) | (0.0140) | (0.0058) | (0.0020) |
| (4) | 0.3024 | 0.2102 | 0.1755 | 0.2584 | 0.1555 | 0.0888 | 0.2562 | 0.1472 | 0.0862 |
| | (0.0096) | (0.0098) | (0.0039) | (0.0172) | (0.0098) | (0.0041) | (0.0177) | (0.0109) | (0.0038) |
| (5) | 0.1877 | 0.1145 | 0.0913 | 0.1432 | 0.0688 | 0.0283 | 0.1417 | 0.0650 | 0.0203 |
| | (0.0130) | (0.0122) | (0.0105) | (0.0199) | (0.0093) | (0.0064) | (0.0190) | (0.0114) | (0.0051) |

Table .7: MISE comparison between the likelihood CV and Bayesian methods for estimating the error density. The number in parenthesis represents the sample sd of the ISE. The red colored text represents the minimal MISE, while the blue colored text represents the minimal sd of the ISE.

## References


Ahmad, I., Li, Q., 1997. Testing symmetry of an unknown density function by kernel method. Journal of Nonparametric Statistics 7 (3), 279–293.

Akritas, M. G., Van Keilegom, I., 2001. Non-parametric estimation of the residual distribution. Scandinavian Journal of Statistics 28 (3), 549–567.

Aneiros-Pérez, G., Vieu, P., 2006. Semi-functional partial linear regression. Statistics and Probability Letters 76 (11), 1102–1110.

Aston, J. A. D., Chiou, J.-M., Evans, J. P., 2010. Linguistic pitch analysis using functional principal component mixed effect models. Journal of the Royal Statistical Society: Series C 59 (2), 297–317.

Barrientos-Marin, J., Ferraty, F., Vieu, P., 2010. Locally modelled regression and functional data. Journal of Nonparametric Statistics 22 (5), 617–632.

Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P., 2007. Local smoothing regression with functional data. Computational Statistics 22 (3), 353–369.

Berg, B. A., 2005. Introduction to Markov chain Monte Carlo simulations and their statistical analysis. In: Kendall, W. S., Liang, F., Wang, J.-S. (Eds.), Markov Chain Monte Carlo: Innovation and Applications. World Scientific Publishing Co., Singapore.

Besse, P. C., Cardot, H., Stephenson, D. B., 2000. Autoregressive forecasting of some functional climatic variations. Scandinavian Journal of Statistics 27 (4), 673–687.

Boj, E., Delicado, P., Fortiana, J., 2010. Distance-based local linear regression for functional predictors. Computational Statistics & Data Analysis 54 (2), 429–437.

Bowman, A. W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71 (2), 353–360.

Burba, F., Ferraty, F., Vieu, P., 2009. $k$-nearest neighbour method in functional nonparametric regression. Journal of Nonparametric Statistics 21 (4), 453–469.

Cheng, F., 2002. Consistency of error density and distribution function estimators in nonparametric regression. Statistics and Probability Letters 59 (3), 257–270.

Cheng, F., 2004. Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression. Journal of Statistical Planning and Inference 119 (1), 95–107.

Cheng, F., Sun, S., 2008. A goodness-of-fit test of the errors in nonlinear autoregressive time series models. Statistics and Probability Letters 78 (1), 50–59.

Chiou, J.-M., Müller, H.-G., 2007. Diagnostics for functional regression via residual processes. Computational Statistics & Data Analysis 51 (10), 4849–4863.

Chiou, J.-M., Müller, H.-G., 2009. Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. Journal of the American Statistical Association 104 (486), 572–585.

Dabo-Niang, S., Guillas, S., 2010. Functional semiparametric partially linear model with autoregressive errors. Journal of Multivariate Analysis 101, 307–315.

Dette, H., Kusi-Appiah, S., Neumeyer, N., 2002. Testing symmetry in nonparametric regression models. Nonparametric Statistics 14 (5), 477–494.

Efromovich, S., 2005. Estimation of the density of regression errors. The Annals of Statistics 33 (5), 2194–2227.

Engle, R. F., González-Rivera, G., 1991. Semiparametric ARCH models. Journal of Business & Economic Statistics 9 (4), 345–359.

Erbas, B., Hyndman, R. J., Gertig, D. M., 2007. Forecasting age-specific breast cancer mortality using functional data models. Statistics in Medicine 26 (2), 458–470.

Escanciano, J. C., Jacho-Chávez, D. T., 2012. $\sqrt{n}$ uniformly consistent density estimation in nonparametric regression models. Journal of Econometrics 167 (2), 305–316.

Fan, Y., James, G., 2013. Functional additive regression. Working paper, University of Southern California,
URL: http://www-bcf.usc.edu/~gareth/research/FAR.pdf.

Febrero-Bande, M., González-Manteiga, W., 2013. Generalized additive models for functional data. Test in press.

Ferguson, T. S., 1983. Bayesian density estimation by mixtures of normal distributions. In: Rizvi, M. H., Rustagi, J., Siegmund, D. (Eds.), Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday. Academic Press, New York.

Fernandez de Castro, B., Guillas, S., Gonzalez-Manteiga, W., 2005. Functional samples and bootstrap for predicting sulfur dioxide levels. Technometrics 47 (2), 212–222.

Ferraty, F., Laksaci, A., Tadj, A., Vieu, P., 2011. Kernel regression with functional response. Electronic Journal of Statistics 5, 159–171.

Ferraty, F., Van Keilegom, I., Vieu, P., 2010. On the validity of the bootstrap in non-parametric functional regression. Scandinavian Journal of Statistics 37 (2), 286–306.

Ferraty, F., Van Keilegom, I., Vieu, P., 2012. Regression when both response and predictor are functions. Journal of Multivariate Analysis 109, 10–28.

Ferraty, F., Vieu, P., 2002. The functional nonparametric model and application to spectrometric data. Computational Statistics 17 (4), 545–564.

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York.

Frank, I. E., Friedman, J. H., 1993. A statistical view of some chemometrics regression tools (with discussion). Technometrics 35 (2), 109–148.

Garthwaite, P. H., Fan, Y., Sisson, S. A., 2010. Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. Working paper, University of New South Wales,
URL: http://arxiv.org/pdf/1006.3690v1.pdf.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J. M., Berger, J., Dawid, A. P., Smith, J. F. M. (Eds.), Bayesian Statistics. Clarendon Press, Oxford, pp. 169–193.

Geweke, J., 1999. Using simulation methods for Bayesian econometric models: inference, development, and communication (with discussion). Econometric Reviews 18 (1), 1–73.

Geweke, J., 2010. Complete and Incomplete Econometric Models. Princeton University Press, Princeton, USA.

Gilks, W. R., Richardson, S., Spiegelhatler, D. J., 1996. Introducing Markov chain Monte Carlo. In: Markov Chain Monte Carlo in Practice. Chapman and Hall, London, pp. 45–57.

Hall, P., 1987. On Kullback-Leibler loss and density estimation. The Annals of Statistics 15 (4), 1491–1519.

Harezlak, J., Coull, B. A., Laird, N. M., Magari, S. R., Christiani, D. C., 2007. Penalized solutions to functional regression problems. Computational Statistics & Data Analysis 51 (10), 4911–4925.

Hastie, T., Buja, A., Tibshirani, R., 1995. Penalized discriminant analysis. The Annals of Statistics 23 (1), 73–102.

Hastie, T., Mallows, C., 1993. A statistical view of some chemometrics regression tools (discussion). Technometrics 35 (2), 140–143.

Heidelberger, P., Welch, P. D., 1983. Simulation run length control in the presence of an initial transient. Operations Research 31 (6), 1109–1144.

Hlubinka, D., Prchal, L., 2007. Changes in atmospheric radiation from the statistical point of view. Computational Statistics & Data Analysis 51 (10), 4926–4941.

Horváth, L., Reeder, R., 2012. A test of significance in functional quadratic regression. In: Horváth, L., Kokoszka, P. (Eds.), Inference for Functional Data with Applications. Springer, pp. 225–232.

Hyndman, R. J., Booth, H., 2008. Stochastic population forecasts using functional data models for mortality, fertility and migration. International Journal of Forecasting 24 (3), 323–342.

Hyndman, R. J., Shang, H. L., 2009. Forecasting functional time series (with discussion). Journal of the Korean Statistical Society 38 (3), 199–221.

Hyndman, R. J., Ullah, M. S., 2007. Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis 51 (10), 4942–4956.

Jaki, T., West, R. W., 2008. Maximum kernel likelihood estimation. Journal of Computational and Graphical Statistics 17 (4), 976–993.

Jaki, T., West, R. W., 2011. Symmetric maximum kernel likelihood estimation. Journal of Statistical Computation and Simulation 81 (2), 193–206.

Jones, M. C., Marron, J. S., Park, B. U., 1991. A simple root-$n$ bandwidth selector. The Annals of Statistics 19 (4), 1919–1932.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65 (3), 361–393.

Malfait, N., Ramsay, J. O., 2003. The historical functional linear model. The Canadian Journal of Statistics/La Revue Canadienne de Statistique 31 (2), 115–128.

Marron, J. S., Wand, M. P., 1992. Exact mean integrated squared error. The Annals of Statistics 20 (2), 712–736.

Meyer, R., Yu, J., 2000. BUGS for a Bayesian analysis of stochastic volatility models. Econometrics Journal 3 (2), 198–215.

Muhsal, B., Neumeyer, N., 2010. A note on residual-based empirical likelihood kernel density estimation. Electronic Journal of Statistics 4, 1386–1401.

Müller, H.-G., Wang, J. L., 1990. Locally adaptive hazard smoothing. Probability Theory and Related Fields 85 (4), 523–538.

Müller, H.-G., Yao, F., 2008. Functional additive models. Journal of the American Statistical Association 103 (484), 1534–1544.

Neumeyer, N., Dette, H., 2007. Testing for symmetric error distribution in nonparametric regression models. Statistica Sinica 17 (2), 775–795.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6 (1), 7–11.

Quintela-del-Río, A., Ferraty, F., Vieu, P., 2011. Analysis of time of occurrence of earthquakes: a functional data approach. Mathematical Geosciences 43 (6), 695–719.

Quintela-del-Río, A., Francisco-Fernández, M., 2011. Nonparametric functional data estimation applied to ozone data: prediction and extreme value analysis. Chemosphere 82 (6), 800–806.

Rachdi, M., Vieu, P., 2007. Nonparametric regression for functional data: automatic smoothing parameter selection. Journal of Statistical Planning and Inference 137 (9), 2784–2801.

Ramsay, J. O., Silverman, B. W., 2005. Functional Data Analysis, 2nd Edition. Springer, New York.

Ratcliffe, S. J., Leader, L. R., Heller, G. Z., 2002. Functional data analysis with application to periodically stimulated foetal heart rate data. I: functional regression. Statistics in Medicine 21 (8), 1103–1114.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Annals of Mathematical Statistics 22 (3), 400–407.

Robert, C. P., Casella, G., 2010. Monte Carlo Statistical Methods. Springer, New York.

Roberts, G. O., 1996. Markov chain concepts related to sampling algorithms. In: Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman and Hall, London, pp. 45–57.

Roberts, G. O., Rosenthal, J. S., 2009. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18 (2), 349–367.

Roeder, K., Wasserman, L., 1997. Practical Bayesian density estimation using mixtures of normals. Journal of the American Statistical Association 92 (439), 894–902.

Samb, R., 2011. Nonparametric estimation of the density of regression errors. Comptes Rendus Mathematique 349 (23-24), 1281–1285.

Schellhase, C., Kauermann, G., 2012. Density estimation and comparison with a penalized mixture approach. Computational Statistics 27 (4), 757–777.

Shang, H. L., Hyndman, R. J., 2011. Nonparametric time series forecasting with dynamic updating. Mathematics and Computers in Simulation 81 (7), 1310–1324.

Tse, Y. K., Zhang, X., Yu, J., 2004. Estimation of hyperbolic diffusion using the Markov chain Monte Carlo method. Quantitative Finance 4 (2), 158–169.

van der Laan, M. J., Dudoit, S., Keles, S., 2004. Asymptotic optimality of likelihood-based cross-validation. Statistical Applications in Genetics and Molecular Biology 3 (1), Article 4.

Yao, F., Müller, H.-G., 2010. Functional quadratic regression. Biometrika 97 (1), 49–64.

Yao, F., Müller, H.-G., Wang, J.-L., 2005a. Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association 100 (470), 577–590.

Yao, F., Müller, H.-G., Wang, J.-L., 2005b. Functional linear regression analysis for longitudinal data. The Annals of Statistics 33 (6), 2873–2903.

Zhang, X., Brooks, R. D., King, M. L., 2009. A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation. Journal of Econometrics 153 (1), 21–32.

Zhang, X., King, M. L., 2011. Bayesian semiparametric GARCH models. Working paper 24, Department of Econometrics & Business Statistics, URL: http://www.buseco.monash.edu/ebs/pubs/wpapers/2011/wp24-11.pdf.