

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/228766913>

# Bayesian estimation of bandwidths for a nonparametric regression model with a flexible error density

Article · January 2011

---

CITATION

1

---

READS

21

3 authors, including:



[Maxwell L. King](#)

Monash University (Australia)

156 PUBLICATIONS 2,401 CITATIONS

[SEE PROFILE](#)



[Han Lin Shang](#)

Australian National University

50 PUBLICATIONS 297 CITATIONS

[SEE PROFILE](#)



**MONASH** University

**Australia**

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Bayesian estimation of bandwidths for a  
nonparametric regression model with a flexible  
error density**

**Xibin Zhang, Maxwell L. King, Han Lin Shang**

**August 2011**

**Working Paper 10/11**

# Bayesian estimation of bandwidths for a nonparametric regression model with a flexible error density

Xibin Zhang, Maxwell L. King<sup>1</sup>, Han Lin Shang

Department of Econometrics and Business Statistics, Monash University

August 22, 2011

**Abstract:** We approximate the error density of a nonparametric regression model by a mixture of Gaussian densities with means being the individual error realizations and variance a constant parameter. We investigate the construction of a likelihood and posterior for bandwidth parameters under this Gaussian-component mixture density of errors in a nonparametric regression. A Markov chain Monte Carlo algorithm is presented to sample bandwidths for the kernel estimators of the regression function and error density. A simulation study shows that the proposed Gaussian-component mixture density of errors is clearly favored against wrong assumptions of the error density. We apply our sampling algorithm to a nonparametric regression model of the All Ordinaries daily return on the overnight FTSE and S&P 500 returns, where the error density is approximated by the proposed mixture density. With the estimated bandwidths, we estimate the density of the one-step-ahead point forecast of the All Ordinaries return, and therefore, a distribution-free value-at-risk is obtained. The proposed Gaussian-component mixture density of regression errors is also validated through the nonparametric regression involved in the state-price density estimation proposed by [Aït-Sahalia and Lo \(1998\)](#).

**Key words:** Bayes factors, Gaussian-component mixture density, Markov chain Monte Carlo, state-price density, value-at-risk.

**JEL Classification:** C11, C14, C15, G15

---

<sup>1</sup>Department of Econometrics and Business Statistics, Monash University, Wellington Road, Clayton, VIC 3800, Australia. Telephone: +61-3-99052449. Fax: +61-3-99020215. Email: max.king@monash.edu.

# 1 Introduction

Nonparametric regression has been widely used for exploring the relationship between a response variable and a set of explanatory variables without specifying a parametric form of such a relationship. A simple and commonly used estimator of the regression function is the Nadaraya-Watson (NW) estimator, whose performance is mainly determined by the choice of bandwidths. There exists a large body of literature on bandwidth selection for the NW estimator, such as the rule-of-thumb and cross-validation (CV) discussed by [Härdle \(1990\)](#), the plug-in method discussed by [Herrmann, Engel, Wand, and Gasser \(1995\)](#) and bootstrapping proposed by [Hall, Lahiri, and Polzehl \(1995\)](#). Even though the NW estimator does not require an assumption on the analytical form of the error density, one may have interest in the distribution of the response around the estimated mean. Such a distribution is characterized by the error density, estimation of which is a fundamental issue in statistical inference for any regression model. This importance was extensively discussed by [Efromovich \(2005\)](#), who developed a nonparametric approach to error-density estimation in a nonparametric regression model using residuals as proxies of errors.

A simple approach to the estimation of the error density is the kernel density estimator of residuals, whose performance is mainly determined by the choice of bandwidth. This density estimator depends on residuals fitted through the NW estimator of the regression function. Moreover, the resulting density estimator of the residuals provides no information for the purpose of choosing bandwidths in the NW regression estimator, although bandwidth selection in this situation depends on the error distribution (see for example, [Zhang, Brooks, and King, 2009](#)). Therefore, there is a lack of a data-driven procedure for choosing bandwidths for the two estimators simultaneously. This motivates the investigation of the paper.

Let  $\mathbf{y}$  denote the response and  $\mathbf{x} = (x_1, x_2, \dots, x_d)'$  a set of explanatory variables or regressors. Given observations  $(y_i, \mathbf{x}_i)$ , for  $i = 1, 2, \dots, n$ , the multivariate nonparametric regression model is expressed as

$$y_i = m(\mathbf{x}_i) + \varepsilon_i, \tag{1}$$

where  $\varepsilon_i$ , for  $i = 1, 2, \dots, n$ , are assumed to be independent and identically distributed (iid) with an unknown density denoted as  $f(\varepsilon)$ . It is also assumed that the errors are independent of the regressors. Let the NW estimator of the regression function be  $\widehat{m}(x; \mathbf{h})$  with  $\mathbf{h}$  a vector of bandwidths. In this paper, we assume that the unknown  $f(\varepsilon)$  is approximated by a mixture density given by

$$f(\varepsilon; b) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \phi\left(\frac{\varepsilon - \varepsilon_i}{b}\right), \quad (2)$$

where  $\phi(\cdot)$  is the probability density function (PDF) of the standard Gaussian distribution, and the component Gaussian densities have means at  $\varepsilon_i$ , for  $i = 1, 2, \dots, n$ , and a common variance  $b^2$ .

From the viewpoint of kernel smoothing, this error density is of the form of a kernel density estimator of the errors (rather than residuals) with  $\phi(\cdot)$  the kernel function and  $b$  the bandwidth. In the situation of density estimation based on direct observations, [Silverman \(1978\)](#) proved the strong uniform consistency of the kernel density estimator under the conditions that  $b \rightarrow 0$ ,  $(nb)^{-1} \ln n \rightarrow 0$  as  $n \rightarrow \infty$  and that  $f(\varepsilon)$  is uniformly continuous. Consequently, it is reasonable to expect that  $f(\varepsilon; b)$  approaches  $f(\varepsilon)$  as the sample size increases, even if  $f(\varepsilon)$  is of an unknown form. Throughout this paper, we call (2) either the Gaussian-component mixture (or mixture Gaussian) error density or the kernel-form error density, where  $b$  is referred to as either the standard deviation or bandwidth.

In this paper, we propose a re-parameterization to  $\mathbf{h}$  and  $b$  and treat the re-parameterized bandwidths as parameters. The main contribution of this paper is to construct an approximate likelihood and therefore, the posterior of re-parameterized bandwidths for the nonparametric regression model with its unknown error density approximated by the Gaussian-component mixture density given by (2). We aim to present a Bayesian sampling algorithm to estimate the re-parameterized bandwidths in the NW estimator and the mixture Gaussian error density in a nonparametric regression model. When the errors are assumed to follow a Gaussian distribution, [Zhang et al. \(2009\)](#) derived the posterior of  $\mathbf{h}$  for given  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ , where the likelihood of  $\mathbf{y}$  for given  $\mathbf{h}$  is the product of the Gaussian densities of  $y_i$  with its mean value approximated by the leave-one-out NW estimator denoted as  $\widehat{m}_i(x_i; \mathbf{h})$ , for  $i = 1, 2, \dots, n$ .

The innovation of our proposed investigation is to use the kernel-form error density given by (2) to replace the Gaussian error density discussed by [Zhang et al. \(2009\)](#).

The proposed investigation is motivated by the important roles that the bandwidths play in the NW estimator and the kernel-form error density. [Härdle \(1990\)](#) highlighted the importance of bandwidth in controlling the smoothness of the NW estimator. The bandwidth in the kernel density estimator of residuals plays the same role as the bandwidth in kernel density estimation based on directly observed data, where the latter issue has been extensively investigated in the literature (see for example, [Wand and Jones, 1995](#)). In our proposed approach, once the kernel function of the NW estimator is chosen, the performances of the NW estimator and the kernel-form error density are determined by the choices of the two types of bandwidths.

The investigation of error density estimation is also motivated by its practical applications, such as inference, prediction and model validation (see for example, [Efromovich, 2005](#); [Muhsal and Neumeier, 2010](#)). In financial economics, an important use of the estimated error density in modeling an asset return is to estimate the value-at-risk (VaR) for holding the asset. In such a model, any wrong specification of the error density may produce an inaccurate estimate of VaR 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 mean in any regression model. There is some existing research on estimation of the error density in nonparametric regression. [Efromovich \(2005\)](#) presented the so-called Efromovich-Pinsker estimator of the error density and showed that this estimator is asymptotically as accurate as an oracle that knows the underlying errors. [Cheng \(2004\)](#) showed that the kernel density estimator of residuals is uniformly, weakly and strongly consistent. When the regression function is estimated by the NW estimator and the error density is estimated by the kernel estimator of residuals, [Samb \(2010\)](#) proved the asymptotic normality of the bandwidths in both estimators and derived the optimal convergence rates of the two types of bandwidths. [Linton and Xiao \(2007\)](#) proposed a kernel estimator based on the local polynomial fitting for a nonparametric regression model with an unknown error density. They showed that their estimator is adaptive

and concluded that adaptive estimation is possible in local polynomial fitting, which includes the NW estimator as a special case.

In all these investigations, residuals were commonly used as proxies of errors, and the bandwidth for the kernel density estimator of residuals was pre-chosen. To our knowledge, there is no method that can simultaneously estimate the bandwidths for the NW estimator of the regression function and the kernel-form error density. This is the aim of this paper.

The assumption of the kernel-form error density was investigated by [Yuan and de Gooijer \(2007\)](#) for nonlinear regression models. The parameters were estimated by maximizing the likelihood with respect to parameters, where the likelihood was constructed through the kernel-form error density based on a bandwidth pre-chosen by the rule-of-thumb. They proved that under some regularity conditions, the maximum likelihood estimate of the vector of parameters is consistent, asymptotically normal and efficient. Nonetheless, their numerical maximization procedure depends on a pre-chosen bandwidth, which in turn depends on a pilot estimate of the parameter vector.

Under the Gaussian-component mixture density of the errors, we also propose to approximate the unknown mean function of  $y_i$  by  $\hat{m}_i(x_i; \mathbf{h})$ . The density of  $y_i$  is approximated by the mixture density given by (2) with plugged-in error realizations. Therefore, the likelihood of  $\mathbf{y}$  for given  $\mathbf{h}$  and  $\mathbf{b}$ , as well as the posterior of  $\mathbf{h}$  and  $\mathbf{b}$  for given  $\mathbf{y}$  can be approximately derived.

In comparison to the Gaussian assumption of the error density discussed in [Zhang et al. \(2009\)](#), our assumption of the mixture Gaussian density of the errors in the same model is robust in terms of different specifications of the error density. In order to understand the benefit and loss that result from this robust assumption against other parametric assumptions, we conduct simulation studies by drawing samples from a nonparametric regression model, where the error densities are Gaussian, Student  $t$  and a mixture of two Gaussians, respectively.

We also investigate the nonparametric regression of the All Ordinaries daily return on the overnight FTSE and S&P 500 returns with the error density assumed to be Gaussian, Student  $t$  and mixture Gaussian, respectively. We compute the VaR under each estimated density, and find that the two parametric assumptions tend to underestimate the VaR in comparison

to the kernel-form error density. Our second application is motivated by the work of [Zhang et al. \(2009\)](#), where a security's state-price density (SPD) is estimated in a nonparametric regression model with Gaussian errors. In this paper, we assume that the unknown error density is approximated by the mixture Gaussian density given by (2).

The rest of this paper is organized as follows. In Section 2, we derive the posterior of the bandwidth parameters in the NW estimator and kernel-form error density. Section 3 presents simulations to evaluate the performance of Bayesian estimation of bandwidths under the Gaussian, Student  $t$  and mixture Gaussian error densities. In Section 4, we present an empirical investigation of the nonparametric relationship between stock index returns across three stock markets. Section 5 applies the proposed sampling procedure to estimate bandwidths in a nonparametric regression model involved in the SPD estimation. Section 6 concludes the paper.

## 2 Bayesian estimation of bandwidths

The bandwidths in the NW estimator of the regression function and the kernel-form error density estimator play an important role in controlling the smoothness of the regression function and the error density estimator. Because of that, bandwidths are also called smoothing parameters in the literature. In this paper, we treat these bandwidths as parameters. In the context of kernel density estimation based on direct observations, there exist some investigations involving similar treatment (see for example, [Brewer, 2000](#); [Gangopadhyay and Cheung, 2002](#); [de Lima and Atuncar, 2011](#)). In nonparametric and semiparametric regression models, bandwidths are also treated as parameters ([Härdle, Hall, and Ichimura, 1993](#); [Rothe, 2009](#), among others).

One may feel reluctant to treat bandwidths as parameters when the asymptotic properties of the NW estimator and the kernel estimator of the error density are under investigation, because both types of bandwidths approach zero as the sample size tends to infinity. Under the asymptotic mean integrated squared error (AMISE), [Samb \(2010\)](#) derived the optimal convergence rates for  $h$  and  $b$  in a nonparametric regression. Let  $\Delta_h$  and  $\Delta_b$  denote the



optimal rates for  $\mathbf{h}$  and  $b$ , respectively. Let  $\mathbf{h} = (h_1, h_2, \dots, h_d)'$  denote a vector of bandwidths used by the NW estimator. We re-parametrize  $b$  and the elements of  $\mathbf{h}$  as

$$b = \tau_0 n^{-\Delta_b}, \quad (3)$$

$$h_k = \tau_k n^{-\Delta_{h_k}}, \text{ for } k = 1, 2, \dots, d, \quad (4)$$

where  $\tau_i$ , for  $i = 0, 1, \dots, d$ , can certainly be treated as constant parameters. Nonetheless, we do not have to use such a re-parameterization in finite samples, where  $n^{-\Delta_{h_k}}$  and  $n^{-\Delta_b}$  are known constants. Let  $\boldsymbol{\tau}^2$  denote  $(\tau_0^2, \tau_1^2, \dots, \tau_d^2)$  throughout this paper, and strictly speaking, this is the vector of parameters in our proposed Bayesian sampling procedure.

Given observations denoted as  $(y_i, \mathbf{x}_i)$ , for  $i = 1, 2, \dots, n$ , we aim to construct the likelihood, as well as the posterior of the parameters. In Section 2.1, we briefly describe the construction of the likelihood and posterior under the assumption of the Gaussian errors. We then derive the likelihood and posterior of bandwidths under the assumption of mixture Gaussian error density in Section 2.2.

## 2.1 Gaussian error distribution

[Zhang et al. \(2009\)](#) considered the nonparametric regression model given by (1), where  $\varepsilon_i$ , for  $i = 1, 2, \dots, n$ , are iid and follow  $N(0, \sigma^2)$  with  $\sigma^2$  an unknown parameter. The model implies that

$$\frac{y_i - m(\mathbf{x}_i)}{\sigma} \sim N(0, 1).$$

As the analytical form of  $m(\mathbf{x}_i)$  is unknown, it is estimated by the leave-one-out NW estimator,

$$\hat{m}_i(\mathbf{x}_i; \mathbf{h}) = \frac{(n-1)^{-1} \sum_{j=1; j \neq i}^n K_{\mathbf{h}}(\mathbf{x}_i - \mathbf{x}_j) y_j}{(n-1)^{-1} \sum_{j=1; j \neq i}^n K_{\mathbf{h}}(\mathbf{x}_i - \mathbf{x}_j)}, \quad (5)$$

where  $K_{\mathbf{h}}(\mathbf{z}) = K(\mathbf{z} ./ \mathbf{h}) ./ \mathbf{h}$  with  $K(\cdot)$  being a kernel function and “./” division by elements. Let  $\mathbf{h}^2 = (h_1^2, h_2^2, \dots, h_d^2)'$ . Treating  $\sigma^2$  and the elements of  $\mathbf{h}^2$  as parameters, one can derive the likelihood of  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  as

$$L_g(\mathbf{y} | \mathbf{h}^2, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - \hat{m}_i(\mathbf{x}_i, \mathbf{h})]^2\right). \quad (6)$$

[Zhang et al. \(2009\)](#) derived the posterior of  $\mathbf{h}^2$  and  $\sigma^2$ , which is proportional to the product of (6) and pre-chosen priors of  $\mathbf{h}^2$  and  $\sigma^2$ . A posterior simulation algorithm was also presented for estimating  $\mathbf{h}^2$  and  $\sigma^2$ .

A limitation of this approach is that the distribution of the iid errors has to be specified. Any wrong assumption of the error density may lead to an inaccurate estimate of the vector of bandwidths. In what follows, we will investigate a robust specification of the error density.

## 2.2 Gaussian-component mixture error density

The aim of this paper is to investigate the construction of the likelihood and posterior in (1) with its unknown error density approximated by the Gaussian-component mixture density given by (2). If  $m(\mathbf{x})$  is known, this mixture density is a well-defined density function of the errors. Therefore, we can derive the density of the response variable as

$$y_i \sim f([y_i - m(\mathbf{x}_i)]; b) = \frac{1}{n} \sum_{j=1}^n \frac{1}{b} \phi\left(\frac{[y_i - m(\mathbf{x}_i)] - [y_j - m(\mathbf{x}_j)]}{b}\right), \quad (7)$$

for  $i = 1, 2, \dots, n$ . [Yuan and de Gooijer \(2007\)](#) and [Zhang and King \(2010\)](#) demonstrated the validity of this mixture density as a density of the regression errors in a class of nonlinear regression models and a family of univariate GARCH models, respectively. [Zhang and King \(2010\)](#) proposed using Bayesian sampling techniques to estimate bandwidths in their semiparametric GARCH model with its unknown error density approximated by a mixture Gaussian density, which is the same as (2). Consequently, the likelihood for their semiparametric GARCH model is well defined, and subsequent posterior simulation is meaningful.

The regression function in (1) is unknown, but can be estimated by the NW estimator for the purpose of constructing the likelihood. As a result, the realized errors or residuals are used as proxies of errors. We propose to plug-in the leave-one-out NW estimator of  $m(\mathbf{x})$  into (7). Therefore, the density of  $y_i$  is approximated by  $f(y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h}); b)$ , which is expressed as

$$\hat{f}([y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})]; b) = \frac{1}{n} \sum_{j=1}^n \frac{1}{b} \phi\left(\frac{[y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})] - [y_j - \hat{m}_j(\mathbf{x}_j; \mathbf{h})]}{b}\right). \quad (8)$$

In fact, the use of  $\hat{m}_i(\mathbf{x}_i; \mathbf{h})$  as an approximation to the mean of  $y_i$  was proposed by [Zhang et al. \(2009\)](#) in the nonparametric regression model with Gaussian errors. They constructed the

likelihood through the Gaussian density of  $y_i$  with its mean value approximated by  $\hat{m}_i(\mathbf{x}_i; \mathbf{h})$ , for  $i = 1, 2, \dots, n$ .

### 2.2.1 Likelihood

The likelihood of  $\mathbf{y}$  is essentially the product of the density of  $y_i$  given by (8), for  $i = 1, 2, \dots, n$ . However, it is impossible to estimate  $b$  by maximizing such a likelihood, because it contains at least one unwanted term  $\phi(0)/b$ . The likelihood approaches infinity as  $b$  tends to zero. An exclusion of the  $i$ th term only from the summation is not enough because when  $y_j - y_i = \hat{m}_j(\mathbf{x}_j; \mathbf{h}) - \hat{m}_i(\mathbf{x}_i; \mathbf{h})$ , for  $j \neq i$ , the  $j$ th term in the summation becomes  $\phi(0)/b$ . Nonetheless, a remedy to this problem is to exclude the  $j$ th term that makes  $[y_j - \hat{m}_j(\mathbf{x}_j; \mathbf{h})] = [y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})]$ , from the summation given by (8). Let

$$J_i = \{j : y_j - \hat{m}_j(\mathbf{x}_j; \mathbf{h}) \neq y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h}), \text{ for } j = 1, 2, \dots, n\}, \quad (9)$$

for  $i = 1, 2, \dots, n$ , and let  $n_i$  denote the number of terms excluded from the summation in (8).

The density of  $y_i$  is therefore, approximated as

$$\tilde{f}([y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})]; b) = \frac{1}{n - n_i} \sum_{j \in J_i} \frac{1}{b} \phi\left(\frac{[y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})] - [y_j - \hat{m}_j(\mathbf{x}_j; \mathbf{h})]}{b}\right), \quad (10)$$

where  $y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})$  is the  $i$ th residual denoted as  $\hat{\varepsilon}_i$ , for  $i = 1, 2, \dots, n$ .

There are a few issues that should be addressed. First, although  $\tilde{f}(\hat{\varepsilon}_i; b)$  has the form of the kernel density estimator of residuals excluding those with the same value as  $\hat{\varepsilon}_i$ , its functional form does not depend on the excluded residuals because  $\tilde{f}(\hat{\varepsilon}_i; b)$  can be rewritten as

$$\tilde{f}(\hat{\varepsilon}_i; b) = \frac{1}{(n - n_i)b} \left[ \sum_{j=1}^n \phi\left(\frac{\hat{\varepsilon}_i - \hat{\varepsilon}_j}{b}\right) - n_i \phi(0) \right],$$

whose functional form does not depend on  $\hat{\varepsilon}_i$ , for  $i = 1, 2, \dots, n$ .

Second, from the view of kernel density estimation, the approximate density of  $y_i$  given by (10) is the kernel density estimator based on a reduced set of residuals excluding the observations whose residuals are the same as  $\hat{\varepsilon}_i$ . This implies that residuals are used as the proxies of the errors (see also [Efromovich, 2005](#)). Expressing it mathematically, the mixture

Gaussian density of the errors,  $f(\varepsilon_i; b)$ , is approximated as

$$f(\varepsilon_i; b) \approx \tilde{f}(\hat{\varepsilon}_i; b) = \frac{1}{n - n_i} \sum_{j \in J_i} \frac{1}{b} K\left(\frac{\hat{\varepsilon}_i - \hat{\varepsilon}_j}{b}\right),$$

for  $i = 1, 2, \dots, n$ .

Finally, using residuals as the proxies of the errors, we approximate the density of  $y_i$  by  $\tilde{f}(y_i - \hat{m}_i(\mathbf{x}_i); b)$ , for  $i = 1, 2, \dots, n$ , from which we are able to derive an approximation of the likelihood.

Given  $h^2$  and  $b^2$ , the likelihood of  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  is approximately

$$L(\mathbf{y} | h^2, b^2) = \prod_{i=1}^n \left\{ \frac{1}{n - n_i} \sum_{j \in J_i} \frac{1}{b} \phi\left(\frac{[y_i - \hat{m}_i(\mathbf{x}_i; \mathbf{h})] - [y_j - \hat{m}_j(\mathbf{x}_j; \mathbf{h})]}{b}\right) \right\},$$

where  $b$  and  $h$  can be re-parameterized according to (3) and (4) whenever it is necessary. In such a situation, this likelihood function is denoted as  $L(\mathbf{y} | \tau^2)$ .

## 2.2.2 Priors

We now discuss the issue of prior choices for the re-parameterized bandwidths. Let  $\pi(\tau_k^2)$  denote the prior of  $\tau_k^2$ , for  $k = 0, 1, \dots, d$ . As  $\tau_0^2 n^{-2\Delta_b}$  and  $\tau_k^2 n^{-2\Delta_h}$ , for  $k = 1, 2, \dots, d$ , which are respectively, the squared bandwidths for the kernel-form error density and the NW estimator, play the same role as a variance parameter, we assume that the priors of  $\tau_0^2 n^{-2\Delta_b}$  and  $\tau_k^2 n^{-2\Delta_h}$  are the inverse Gamma density denoted as  $\text{IG}(\alpha_b, \beta_b)$  and  $\text{IG}(\alpha_h, \beta_h)$ , respectively. Therefore, the prior of  $\tau_0^2$  is

$$\pi(\tau_0^2) = \frac{(\beta_b)^{\alpha_b}}{\Gamma(\alpha_b)} \left(\frac{1}{\tau_0^2 n^{-2\Delta_b}}\right)^{\alpha_b+1} \exp\left\{-\frac{\beta_b}{\tau_0^2 n^{-2\Delta_b}}\right\} n^{-2\Delta_b}, \quad (11)$$

where  $n^{-\Delta_b}$  is the optimal rate of  $b$  under the AMISE, and  $\alpha_b$  and  $\beta_b$  are hyperparameters.

The prior of  $\tau_k^2$  is

$$\pi(\tau_k^2) = \frac{(\beta_h)^{\alpha_h}}{\Gamma(\alpha_h)} \left(\frac{1}{\tau_k^2 n^{-2\Delta_h}}\right)^{\alpha_h+1} \exp\left\{-\frac{\beta_h}{\tau_k^2 n^{-2\Delta_h}}\right\} n^{-2\Delta_h}, \quad (12)$$

for  $k = 1, 2, \dots, d$ , where  $n^{-\Delta_h}$  is the optimal rate of  $h_k$  under the AMISE, and  $\alpha_h$  and  $\beta_h$  are hyperparameters.

### 2.2.3 Posterior

According to Bayes theorem, the posterior of  $\tau^2$  is approximately expressed as (up to a normalizing constant)

$$\pi(\tau^2|\mathbf{y}) \propto L(\mathbf{y}|\tau^2) \prod_{k=0}^d \pi(\tau_k^2), \quad (13)$$

from which we use the random-walk Metropolis algorithm to sample the elements of  $\tau^2$ . The sampling procedure is as follows.

**Step 1:** Initialize  $\tau^2$  with the initial value denoted as  $\tau_{(0)}^2$ .

**Step 2:** Update  $\tau^2$  using the random-walk Metropolis algorithm with updated values denoted as  $\tau_{(1)}^2$ .

**Step 3:** Repeat **Step 2** until the chain  $\{\tau_{(i)}^2 : i = 1, 2, \dots\}$  achieves a reasonable mixing performance.

When the sampling procedure is completed, the ergodic average of the sampled chain of  $\tau$  is used as the estimate of  $\tau$ , where  $\tau$  denotes  $(\tau_0, \tau_1, \dots, \tau_d)'$ . Therefore, the estimates of the element of  $b$  and  $h$  can be computed through (3) and (4), and the analytical form of the kernel-form error density can be derived based on the estimated  $b$  and  $h$ .

## 3 Monte Carlo simulation

The purposes of this simulation study are as follows. First, with one simulated sample, we illustrate the use and effectiveness of our Bayesian sampling algorithm for estimating the bandwidths in the NW regression estimator and the kernel estimator of the error density based on residuals. As our sampling method is built up under an unknown error density, this method is also compared with its parametric counterparts under the assumptions of Student  $t$  and Gaussian error densities, where Bayes factors derived from marginal likelihood are used to determine which assumption is favored against the others.

Second, we will generate 1,000 samples from a nonparametric regression model with its error densities assumed to be respectively, the Gaussian, Student  $t$  and a mixture of

two Gaussian densities. We estimate the bandwidths in the NW estimator of the regression function using the normal reference rule (NRR) discussed in [Scott \(1992\)](#), likelihood CV, bootstrapping and Bayesian sampling technique under each of the above three error-density assumptions. Their estimation accuracies are assessed by average squared errors (ASE) in Section 3.3. Furthermore, Bayesian methods also allow us to estimate either the bandwidth in the kernel density estimator of residuals under the mixture Gaussian error density, or the parameters  $\sigma$  and  $\nu$  in the parametric error-density assumptions.

Finally, three Bayesian sampling approaches are compared through 1,000 samples in Section 3.4, where Bayes factors are used for comparison purposes. We briefly describe Bayes factors as follows.

### 3.1 Bayes factors

In Bayesian inference, model selection is often conducted through the Bayes factor of the model of interest against a competing model, which reflects a summary of evidence provided by the data supporting the model as opposed to its competing model. The Bayes factor is defined as the ratio of the marginal likelihoods derived under the model of interest and its competing model, respectively.

The marginal likelihood is the expectation of the likelihood with respect to the prior of parameters. It is seldom calculated as the integral of the product of the likelihood and prior of parameters, but instead, is often computed numerically ([Gelfand and Dey, 1994](#); [Newton and Raftery, 1994](#); [Chib, 1995](#); [Kass and Raftery, 1995](#); [Geweke, 1999](#), among others). In this paper, we employed the method proposed by [Chib \(1995\)](#) to compute the marginal likelihood.

Let  $\theta$  denote the parameter vector and  $\mathbf{y}$  the data. [Chib \(1995\)](#) showed that the marginal likelihood under a model  $\mathcal{A}$  is expressed as

$$P_{\mathcal{A}}(\mathbf{y}) = \frac{\ell_{\mathcal{A}}(\mathbf{y}|\theta)\pi_{\mathcal{A}}(\theta)}{\pi_{\mathcal{A}}(\theta|\mathbf{y})}, \quad (14)$$

where  $\ell_{\mathcal{A}}(\mathbf{y}|\theta)$ ,  $\pi_{\mathcal{A}}(\theta)$  and  $\pi_{\mathcal{A}}(\theta|\mathbf{y})$  denote respectively, the likelihood, prior and posterior under model  $\mathcal{A}$ .  $P_{\mathcal{A}}(\mathbf{y})$  is usually computed at the posterior estimate of  $\theta$ . The numerator has a closed form and can be computed analytically. The denominator is the posterior of

$\theta$ , which is often replaced by its kernel density estimator based on the simulated chain of  $\theta$  through a posterior sampler.

The Bayes factor of model  $\mathcal{A}$  against model  $\mathcal{B}$  is defined as

$$\text{BF} = \frac{P_{\mathcal{A}}(\mathbf{y})}{P_{\mathcal{B}}(\mathbf{y})},$$

which is used to make a decision on whether  $\mathcal{A}$  is favored against  $\mathcal{B}$ , according to the [Jeffreys \(1961\)](#) scales modified by [Kass and Raftery \(1995\)](#). A Bayes factor value between 1 and 3 indicates that the evidence supporting  $\mathcal{A}$  against  $\mathcal{B}$  is not worth more than a bare mention. When the Bayes factor is between 3 and 20,  $\mathcal{A}$  is favored against  $\mathcal{B}$  with positive evidence; when the Bayes factor is between 20 and 150,  $\mathcal{A}$  is favored against  $\mathcal{B}$  with strong evidence; and when the Bayes factor is above 150,  $\mathcal{A}$  is favored against  $\mathcal{B}$  with very strong evidence.

### 3.2 Performance of the proposed Bayesian methods

Consider the relationships between  $\mathbf{y}$  and  $\mathbf{x} = (x_1, x_2, x_3)'$  given by

$$y_i = \sin(2\pi x_{1,i}) + 4(1 - x_{2,i})(1 + x_{2,i}) + \frac{2x_{3,i}}{1 + 0.8x_{3,i}^2} + \varepsilon_i, \quad (15)$$

for  $i = 1, 2, \dots, n$ . A sample of 1,000 observations was generated by drawing  $x_{1,i}$ ,  $x_{2,i}$  and  $x_{3,i}$  independently from the uniform density on  $(0, 1)$  and  $\varepsilon_i$  from the mixture of two Gaussian densities defined as  $0.7N(0, 0.7^2) + 0.3N(0, 1.5^2)$ , and calculating  $y_i$  based on (15), for  $i = 1, 2, \dots, 1000$ .

The relationship between  $y_i$  and  $(x_{1,i}, x_{2,i}, x_{3,i})'$  was modeled by the nonparametric regression model given as

$$y_i = m(x_{1,i}, x_{2,i}, x_{3,i}) + \varepsilon_i, \quad (16)$$

where  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  are assumed to be iid.

As there are three regressors, the optimal rates of the bandwidths for the NW estimator of the regression function and kernel estimator of the error density are respectively,  $n^{-1/7}$  and  $n^{-3/17}$  (see [Samb, 2010](#)). Therefore, the bandwidths are re-parameterized as

$$b = \tau_0 n^{-3/17}, \text{ and } h_k = \tau_k n^{-1/7}, \text{ for } k = 1, 2, 3,$$

which are used throughout this section.

Assuming that the error density of (16) is unknown and is approximated by the mixture Gaussian density given by (2), we applied our Bayesian sampling algorithm to (16) using the generated sample. The hyperparameters were chosen as  $\alpha_h = \alpha_b = 1$  and  $\beta_h = \beta_b = 0.05$ . These values are often used as the parameter values of an inverse Gamma density when it is chosen as the prior of a variance parameter (see also [Geweke, 2009](#)). The burn-in period contains 1,000 draws, and the following 10,000 draws were recorded. Whenever the random-walk Metropolis algorithm was used, the acceptance rate was controlled to be between 0.2 and 0.3. The posterior means of the re-parameterized bandwidths are presented in the first panel of Table 1. The mixing performance of this posterior sampler is examined by the simulation inefficiency factor (SIF), which can be loosely interpreted as the number of draws needed so as to obtain independent draws from the simulated Markov chain. For example, a SIF value of 20 indicates that approximately, we would need to keep one draw for every 20 draws so as to derive independent draws (see for example, [Roberts, 1996](#); [Kim, Shephard, and Chib, 1998](#); [Tse, Zhang, and Yu, 2004](#)).

The standard deviation of the posterior mean is approximated by the batch-mean standard deviation. It becomes smaller and smaller with the number of simulation iterations increasing, if the sampler achieves a reasonable mixing performance. The SIF and batch-mean standard deviation were used to monitor the mixing performance. Table 1 presents the values of these two indicators, which show that the sampler has mixed very well.

Under the assumption of Gaussian errors, we used [Zhang et al.'s \(2009\)](#) simulation algorithm to sample the re-parameterized bandwidths and the variance parameter from their conditional posteriors, where the re-parameterization was carried out according to (4). The priors of the re-parameterized bandwidths are the same as those under the mixture Gaussian error density. The prior of  $\sigma^2$  is the inverse Gamma density with hyperparameters  $\alpha_\sigma = 1$  and  $\beta_\sigma = 0.05$ . The results are given in the second panel of Table 1.

Assuming the errors of (16) follow the Student  $t$  distribution with  $\nu$  degrees of freedom, we derived the likelihood and posterior in a similar way as those derived under Gaussian errors.



The priors of the re-parameterized bandwidths are the same as those under Gaussian errors, and the prior of  $\nu$  is the Gaussian density with mean 10 and variance  $5^2$ , which is truncated at 3 with the functional form given by

$$\pi(\nu) = \frac{1/5 \phi((\nu - 10)/5)}{1 - \Phi((3 - 10)/5)}, \quad (17)$$

where  $\Phi(\cdot)$  is the cumulative density function (CDF) of the standard Gaussian distribution. This prior is flat and restricts  $\nu$  to be greater than 3.

We used the random-walk Metropolis algorithm to sample the re-parameterized bandwidths and  $\nu$  from their posterior. The results are presented in the third panel of Table 1. Under these two parametric assumptions, the values of the SIF and batch-mean standard deviation indicate that both samplers have achieved a reasonable mixing performance.

For each error-density assumption of the nonparametric regression model given by (16), we computed the marginal likelihood given by (14) and the average squared error (ASE) defined as

$$\text{ASE}(\mathbf{h}) = \frac{1}{n} \sum_{i=1}^n [\hat{m}(\mathbf{x}_i, \mathbf{h}) - m(\mathbf{x}_i)]^2.$$

We found the following evidence from this simulation study. First, the marginal likelihood obtained under the mixture Gaussian error density is larger than that obtained under either the Gaussian or Student  $t$  error density. The Bayes factors of the mixture Gaussian error density are  $\exp(20.48)$  against the Student  $t$ , and  $\exp(32.79)$  against the Gaussian error densities, respectively. Therefore, the assumption of the mixture Gaussian error density is favored against its parametric counterparts with very strong evidence. Nonetheless, we cannot draw a conclusion only based on one simulated sample. In Section 3.3, we calculate the marginal likelihood under each assumption of the error density for 1,000 samples that are independently drawn.

Second, even though the estimated bandwidth vectors in the NW estimator under three assumptions of error density are almost the same, their associated ASEs are different. The ASE derived under the mixture Gaussian error density is 0.0661, which is smaller than that derived under either the Gaussian or Student  $t$  error density.

Third, we obtained an estimate of the bandwidth for the kernel error-density estimator, whose graph is plotted in the top panel of Figure 1, along with the graphs of the true density, the Gaussian and Student  $t$  error densities. Among these three error-density assumptions, the one derived under the mixture Gaussian error density is the closest to the true density. Moreover, we also plotted the corresponding CDF in the bottom panel of Figure 1. As the goodness of fit of the resulting NW estimator is only about 69.4% under each error-density assumption, we believe the error-density estimator obtained under the mixture Gaussian error density performs reasonably well.

### 3.3 Accuracy of the estimated bandwidth vectors

The accuracy of chosen bandwidths is measured by the ASE of the resulting NW estimator of the true regression function. In kernel density estimation of directly observed data, the NRR is often used for bandwidth selection (see for example, [Silverman, 1986](#); [Scott, 1992](#); [Bowman and Azzalini, 1997](#)). [Härdle and Müller \(2000\)](#) indicated that methods for bandwidth selection in nonparametric regression are the same as those for kernel density estimation. Therefore, we considered the NRR for bandwidth selection in the nonparametric regression model in this simulation study.

The likelihood CV for bandwidth selection has been extensively discussed (see for example, [Wahba and Wold, 1975](#); [Härdle and Marron, 1985](#); [Härdle and Müller, 2000](#)). The bootstrapping approach to bandwidth selection in nonparametric regression was presented by [Hall et al. \(1995\)](#), where two pilot bandwidths have to be specified before bootstrapping begins. The purpose of the first bandwidth is to generate a bootstrapping sample, while the second aims to obtain an initial estimate of the regression function. In our simulation study, the two pilot bandwidths were chosen using the NRR and likelihood CV.

For the purpose of generating samples, the error densities we considered are the Gaussian, Student  $t$  and a mixture of two Gaussian densities. We generated 1,000 samples through the regression function given by (15) based on each of the three densities of the errors. For each sample, we chose bandwidths for the NW estimator through the NRR, likelihood CV

and bootstrapping, and derived the estimates of the re-parameterized bandwidths through Bayesian sampling under each of the three assumptions of the error density. As with the Bayesian sampling method, we also estimated the re-parameterized bandwidth for the kernel error-density estimator, as well as  $\sigma$  and  $\nu$  under the Gaussian and Student  $t$  error densities, respectively.

We calculated the ASE of the NW estimator of the regression function with bandwidths estimated through the aforementioned methods. The mean and standard deviation of the 1,000 ASE values were calculated and tabulated in Table 2. When the errors were generated from  $N(0, 0.9^2)$ , the mean ASE derived through Bayesian sampling with any error-density assumption is smaller than that derived through either the NRR, likelihood CV or bootstrapping. In addition, different assumptions of the error density make no obvious difference in the resulting mean ASE values.

When errors were generated from the Student  $t$  density with 4 degrees of freedom, the mean ASE derived through Bayesian sampling with any error-density assumption is again smaller than that derived through either the NRR, likelihood CV or bootstrapping. The mean ASE obtained through Bayesian sampling under the Gaussian error density is slightly smaller than that obtained through the likelihood CV. Moreover, bootstrapping is clearly the worst performer among all four methods considered.

When the errors were generated from the mixture of two Gaussian densities defined as  $0.7N(0, 0.7^2) + 0.3N(0, 1.5^2)$ , the mean ASE derived through Bayesian sampling under any error-density assumption is clearly smaller than that derived through either the NRR, likelihood CV or bootstrapping. In addition, different assumptions of the error density lead to similar mean ASE values.

### **3.4 Bayesian comparison among error-density assumptions**

The benefit of the mixture Gaussian error density assumption is to gain robustness in terms of error-density specification, because this mixture density is a kernel density estimator of residuals, and has the capacity to approximate an unknown error density. In the nonpara-

metric regression model given by (1), the assumption of a mixture Gaussian error density does not outperform its parametric counterparts under correct error-density assumptions. However, this mixture density usually outperforms its parametric counterparts under wrong assumptions of the error density. We conducted a simulation study using the same 1,000 samples in Section 3.3 to illustrate this.

#### 3.4.1 Comparison when errors were simulated from Gaussian distribution

For those samples generated by simulating the errors of (15) from  $N(0, 0.9^2)$ , we applied the Bayesian sampling methods to estimate the parameters in (16), where the error density was assumed to be either the Gaussian with mean zero and variance  $\sigma^2$ , the Student  $t$  with  $\nu$  degrees of freedom or the mixture Gaussian density given by (8). The Bayesian sampling algorithm under each error-density assumption was applied to each of the 1,000 generated samples. Under each error-density assumption, we derived the estimates of all parameters and marginal likelihood for each sample. We also computed the mean and standard deviation of the 1,000 values of each estimated parameter. Table 3 presents a summary of these results, where the mean of the SIF values computed under each assumption of the error density is below 16. This finding indicates that the Bayesian sampling algorithms have mixed very well.

As the Gaussian error assumption is correct, we calculated the Bayes factors of the assumption of Gaussian error density against the assumptions of Student  $t$  and mixture Gaussian error densities. According to the [Jeffreys \(1961\)](#) scales modified by [Kass and Raftery \(1995\)](#), the assumption of Gaussian error density is favored against the Student  $t$  assumption with either strong or very strong evidence in 98.8% of simulated samples. In only 40.6% of simulated samples, the Gaussian assumption is favored with either strong or very strong evidence against the mixture Gaussian error density. Moreover, our assumption of a mixture Gaussian error density is favored with either positive, strong or very strong evidence against the correct assumption of the error density in 8.7% of simulated samples. Therefore, on average, the assumption of a mixture Gaussian error density is not favored against the correct assumption of the error density .

### 3.4.2 Comparison when errors were simulated from Student $t$ distribution

We generated samples by simulating the errors from the Student  $t$  distribution with  $\nu = 4$ . The Bayesian sampling method was used to estimate all parameters in the nonparametric regression model given by (16), where we considered three assumptions of the error density. We found that the mean of the 1,000 SIF values of each parameter is below 17, which indicates a very reasonable mixing performance of all three samplers. We obtained the mean and standard deviation of the 1,000 values of each estimated parameter under three different assumptions of the error density. We also calculated the marginal likelihood under three assumptions of the error density using each simulated sample. As the Student  $t$  error density is the correct assumption, we calculated the Bayes factors of the Student  $t$  assumption of the error density against the other two assumptions, respectively. A summary of these results is presented in the second panel of Table 3.

According to the modified Jeffreys scales, the assumption of Student  $t$  errors is favored with either strong or very strong evidence against the Gaussian assumption in all simulated samples, and against the assumption of mixture Gaussian errors in 90.1% of simulated samples. However, the mixture Gaussian error-density assumption is favored with either positive, strong or very strong evidence against the correct assumption in 2.7% of simulated samples. This finding shows that on average, the assumption of mixture Gaussian error density is not favored against the correct assumption of the error density.

### 3.4.3 Comparison when errors were simulated from mixture Gaussian distribution

We simulated samples by drawing errors from a mixture of two Gaussian distributions defined as  $0.7N(0, 0.7^2) + 0.3N(0, 1.5^2)$ . The Bayesian sampling method under each of the three error-density assumptions, was applied to estimate parameters in the nonparametric regression model. According to Bayes factors, we derived the estimates of parameters and marginal likelihood for each sample under three assumptions of the error density. We made decisions on whether the assumption of the mixture Gaussian error density is favored against the other two assumptions for each simulated sample. A summary of these results is given in the third

panel of Table 3.

All three assumptions of the error density led to similar estimates of the bandwidth vector in the NW estimator. The mean of SIF values for each parameter under three assumptions of the error density is below 16. This indicates a very good mixing performance of all three samplers.

As the errors were generated from a mixture of two Gaussian densities, all three assumptions of the error density are inaccurate. The assumption of Gaussian errors is likely to lead to poorer performance than the other two, because the true error density exhibits heavy tails. According to the computed Bayes factors and the modified Jeffreys scales, the mixture Gaussian assumption of the error density is favored against the Gaussian assumption in 99.2% of simulated samples, and against the Student  $t$  assumption in 60.7% of simulated samples. In comparison, the Student  $t$  assumption is favored against the mixture Gaussian assumption in only 27.0% of simulated samples. Moreover, we found that the assumption of the mixture Gaussian error density is favored with either strong or very strong evidence against the assumption of the Student  $t$  error density in 49.2% of simulated samples. On the other hand, the latter assumption is favored against the former with the same strength of evidence in only 16.3% of simulated samples. Thus, we can conclude that when the error density is unknown, our proposed mixture Gaussian error density is often favored against its parametric counterparts with wrong error-density assumptions.

### 3.5 Sensitivity of prior choices

Current asymptotic results show that as the sample size tends to infinity, the two types of bandwidths approach zero asymptotically. When there are three or more regressors in the nonparametric regression model, Samb (2010) proved that the optimal rates are  $n^{-3/(2d+11)}$  for the bandwidth in the kernel density estimator of residuals and  $n^{-1/(d+4)}$  for the bandwidths in the NW estimator of the regression function, where  $d$  represents the number of regressors.

Therefore, we assume that the prior densities of  $b$  and  $h_i$  are the uniform densities given by

$$b \sim U(0, 2\sigma_\varepsilon n^{-3/17+1/10}), \quad (18)$$

$$h_k \sim U(0, 2\sigma_k n^{-1/7+1/10}), \quad (19)$$

where  $\sigma_\varepsilon$  is the standard deviation of residuals, and  $\sigma_k$  is the standard deviation of the  $k$ th regressor, for  $k = 1, 2, \dots, d$ . The optimal rates were scaled up by  $n^{1/10}$  to guarantee that the most appropriate values of bandwidths are within the intervals.

Replacing the prior densities in (13) with the uniform priors defined by (18) and (19), we derived the approximate posterior of  $\mathbf{h}$  and  $b$ , through which we sampled the parameters using the random-walk Metropolis algorithm for each of the 1,000 samples in Section 3.2. Moreover, we used the uniform priors given by (19) to derive the approximate posterior for the nonparametric regression model with either the Gaussian or Student  $t$  errors. We derived the estimate of the parameter vector by implementing the same posterior sampler for each sample. A summary of the simulation results is presented in Table 4.

Transforming the estimated  $\tau$  values reported in Table 3 into bandwidth values according to (4), we found that over the 1,000 generated samples, the means of the posterior estimates of the parameter vector obtained under the uniform priors are similar to those obtained under the inverse Gamma priors for each assumption of the error density. Therefore, the posterior estimate of the parameter vector obtained through the aforementioned posterior samplers is on average, insensitive to different prior choices of the bandwidth parameters. Nonetheless, the decisions on whether one assumption of the error density is favored against the others under the uniform priors are different from those under the inverse Gamma priors.

When the errors were generated from the Gaussian distribution, the assumption of Gaussian errors is favored against the assumption of Student  $t$  errors with either strong or very strong evidence in 99.7% of simulated samples under the uniform priors. This relative frequency is similar to the relative frequency of 98.8% under the inverse Gamma priors. On the other hand, the Gaussian assumption is favored against the mixture Gaussian error density with either strong or very strong evidence in 48.3% of simulated samples under the uniform priors, rather than 40.6% of simulated samples under the inverse Gamma priors.

When the errors were generated from the Student  $t$  distribution, the decisions on whether the Student  $t$  assumption is favored against the other two under the uniform priors are largely the same as those under the inverse Gamma priors.

When the errors were generated from the mixture of two Gaussian densities, the proposed mixture Gaussian error density is favored with either positive, strong or very strong evidence against the Gaussian and Student  $t$  error densities in respectively, 99.6% and 73.7% of the simulated samples, under the uniform priors. The latter value is clearly larger than its counterpart, which is 60.7%, under the inverse Gamma priors.

To summarize, on average, the estimated bandwidths are largely insensitive to prior choices. Nonetheless, when samples were generated from either the Gaussian or the mixture Gaussian error distributions, the assumption of a mixture Gaussian error density is more frequently favored against the Student  $t$  assumption under uniform priors than under inverse Gamma priors.

## **4 An application to nonparametric regression of stock returns**

To investigate the empirical relevance of the Bayesian bandwidth selectors, we applied them to the nonparametric regression of the All Ordinaries daily return on the overnight returns of FTSE and S&P 500. As the opening time for share trading in the Australian stock market is after the closing time of the previous-day trading in the UK and USA stock markets, such an investigation can reveal the relationship between the Australian stock market and the other two markets.

### **4.1 Data**

The data consist of the All Ordinaries, FTSE and S&P 500 closing indices collected from the 3rd January 2007 to the 30th March 2011, excluding non-trading days. The All Ordinaries daily return was matched to the overnight FTSE and S&P 500 returns. As a consequence, the daily returns of the FTSE and S&P 500 indices on the 30th March 2011 (local time) was not used for the purpose of estimating bandwidths. When one market experienced a non-



trading day, we also deleted the trading data in the other two markets on that day. The sample contains  $n = 1,022$  observations, from which we computed continuously compounded daily percentage returns. We employ the multivariate nonparametric regression model given by

$$y_i = m(x_{1,i}, x_{2,i}) + \varepsilon_i, \quad \text{for } i = 1, 2, \dots, n, \quad (20)$$

where  $y_i$  is the return of the All Ordinaries daily index, and  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  are assumed to be iid with their distributions assumed to be Gaussian, Student  $t$  and mixture Gaussian, respectively.

## 4.2 Bandwidth estimates under different error densities

As Bayesian sampling was used to estimate the re-parameterized bandwidths (and the parameter under each parametric assumption of the error density), we re-parameterized the bandwidths for the NW estimator of the regression function as  $h_k = \tau_k n^{-1/6}$ , for  $k = 1$  and  $2$ , under each of the three assumptions of the error density. The prior of  $\tau_k^2$  given by (12), was chosen by assuming that  $\tau_k^2 n^{-2/6}$  follows  $\text{IG}(\alpha_h, \beta_h)$  with  $\alpha_h = 1$  and  $\beta_h = 0.05$ , under each assumption of the error density.

When the error density of (20) are assumed Gaussian with mean zero and variance  $\sigma^2$ , we chose the prior of  $\sigma^2$  as  $\text{IG}(1, 0.05)$ . Note that our prior choices for  $\sigma^2$  and  $\tau_k^2$ , for  $k = 1$  and  $2$ , are different from those in Zhang et al. (2009). The random-walk Metropolis algorithm was used to sample  $\tau_1^2$  and  $\tau_2^2$  from their conditional posterior, while the Gibbs sampler was used to sample  $\sigma^2$  from its conditional posterior. We used the batch-mean standard deviation and SIF to examine the mixing performance of the sampling algorithm. Both the batch-mean standard deviation and SIF indicate that all simulated chains have achieved a reasonable mixing performance with the SIF values below 35. Table 5 presents the estimates of  $\sigma$  and the elements of  $\tau$ , and their associated statistics. During the MCMC iterations, we obtained the 95% Bayesian credible interval of each element of the bandwidth vector, which cannot be obtained through the NRR, likelihood CV and bootstrapping.

When the errors are assumed to follow the Student  $t$  distribution with  $\nu$  degrees of freedom, we assumed that the prior of  $\nu$  is a truncated Gaussian density given by (17) and used the random-walk Metropolis algorithm to sample  $\nu$  and the elements of  $\tau^2$ . Table 5 presents

the estimated  $(\nu, \tau_1, \tau_2)'$ , and its associated statistics. According to the batch-mean standard deviation and SIF, each simulated chain has mixed well. As the estimate of  $\nu$  is small, this indicates a slightly heavy-tailed behavior of the error density. As a consequence, it leads to a different distributional shape in comparison with the Gaussian error density.

When the errors are assumed to follow the mixture of Gaussian densities given by (8), the bandwidth for the kernel-form error density was re-parameterized as  $b = \tau_0 n^{-1/5}$ . The prior of  $\tau_0^2$  given by (11), was chosen by assuming that  $\tau_0^2 n^{-2/5}$  follows  $\text{IG}(\alpha_b, \beta_b)$  with  $\alpha_b = 1$  and  $\beta_b = 0.05$ . We used the random-walk Metropolis algorithm to sample  $\tau_k^2$ , for  $k = 0, 1$  and  $2$ , from their posterior given by (13). Table 5 presents the estimated  $\tau_k$ , for  $k = 0, 1$  and  $2$ , and their associated statistics. According to the batch-mean standard deviation and SIF values, all simulated chains have achieved a reasonable mixing performance.

### 4.3 Comparison of different error-density assumptions

We computed the Bayes factors with marginal likelihoods calculated through (14), where the marginal likelihood derived under each assumption of the error density is shown in Table 5. The marginal likelihood under the mixture Gaussian error density is found to be slightly larger than that under the Student  $t$  errors, while the Bayes factor reveals no sufficient evidence to make a decision on whether the former assumption is favored against the latter due to narrow difference between the two marginal likelihood values. However, the assumption of the mixture Gaussian error density is favored against the assumption of Gaussian error density with very strong evidence.

Conditional on the overnight daily returns of the FTSE and S&P 500 indices on the 30th March 2011 (local time), we forecasted the daily return of the All Ordinaries index on the 31st March 2011 (local time) before the Australian stock market opened. Under each assumption of the error density, we derived the posterior estimates of the re-parameterized bandwidths for the NW estimator of the regression function, which was computed at the daily returns of FTSE and S&P 500 on the 30th March 2011 to derive a point forecast of the daily return of the All Ordinaries index on the 31st March 2011. We also produced such a point forecast

at each iteration of the MCMC simulation so as to obtain a posterior sample of the point forecast, based on which we derived the 95% Bayesian credible interval for the point forecast on the All Ordinaries index on the 31st March 2011. Under the three assumptions of the error density, the point forecasts are respectively, 0.2663%, 0.2562% and 0.2676%; and the corresponding credible intervals are respectively, (0.2449%, 0.2885%), (0.2381%, 0.2736%) and (0.2486%, 0.2905%).

#### **4.4 Error-density estimators under different assumptions**

Under the Gaussian-component mixture density of the errors, we are able to estimate the density of the one-step-ahead point forecast of the All Ordinaries daily return, and the derived density can be regarded as a density forecast. Using the estimated bandwidths and parameters given in Table 5, we derived the density of the point forecast of the All Ordinaries daily return on the 31st March 2011 under each of the three assumptions of the error density. The graphs of the three derived densities are presented in Figure 2. The three density graphs are obviously different to each other at their peaks and left tails. As the sample period covers the current global financial crisis, the heavy-tailed feature of stock index-return during this period is more evident than that during non-crisis periods. Consequently, the heavy-tailed mixture Gaussian and Student  $t$  densities are favored against the Gaussian according to Bayes factors. During the period of global financial crisis, global stock markets had experienced very frequent huge drops. Therefore, the estimated density of the point forecast of the All Ordinaries return obtained through the Gaussian-component mixture density of errors exhibits a thick left tail. This distributional feature cannot be revealed through the assumption of Student  $t$  errors.

Under each of the three estimated densities of the All Ordinaries daily return, we computed the one-day VaR for holding an investment on the All Ordinaries index on the 31st March 2011. At the 95% confidence level, the one-day VaRs are respectively, \$1.5668, \$1.5452 and \$1.4803 for every \$100 investment on the All Ordinaries index under the assumptions of the mixture Gaussian, Student  $t$  and Gaussian error densities. In light of the fact that many financial firms had experienced severe liquidity stress or even filed for bankruptcy during the global financial

crisis, we tend to believe that market risk had been underestimated. Therefore, we believe that in comparison to the mixture Gaussian error density, the assumption of Gaussian errors leads to an underestimated VaR.

## 5 An application to SPD estimation

[Aït-Sahalia and Lo \(1998\)](#) showed that in a dynamic equilibrium model, the price of a security is

$$P_t = \exp\{r_{t,\lambda}\lambda\} E_t^* \{Z(S_T)\} = \exp\{r_{t,\lambda}\lambda\} \int_{-\infty}^{\infty} Z(S_T) f_t^*(S_T) dS_T,$$

where  $T = t + \lambda$ ,  $\lambda$  is the length of time to maturity,  $r_{t,\lambda}$  is a constant risk-free interest rate between  $t$  and  $T$ ,  $E_t^*$  represents the expectation taken conditional on information available at date  $t$ ,  $S_T$  is the price of the security at date  $T$ ,  $Z(S_T)$  is the payoff of the security at the expiry date  $T$ , and  $f_t^*(S_T)$  is the date- $t$  SPD for the payoff of the security at date  $T$ . When an option is the security of interest, the SPD is the second-order derivative of a call-option pricing formula with respect to strike price calculated at  $S_T$ . [Aït-Sahalia and Lo \(1998\)](#) showed that the date- $t$  price of a call option, is a nonlinear function of  $(S_t, X_t, \lambda, r_{t,\lambda}, \delta_{t,\lambda})'$ , which can be estimated through the nonparametric regression technique, where  $\delta_{t,\lambda}$  is the dividend rate at date  $t$ .

In order to reduce the number of regressors, [Aït-Sahalia and Lo \(1998\)](#) assumed that the call-option pricing formula is given by the Black-Scholes (BS) formula except that the date- $t$  volatility denoted by  $\sigma_t$ , is estimated by the nonparametric regression of the implied volatility on  $\tilde{z}_t = (F_t, X, \delta)$ , where  $F_t$  is the futures price of the underlying asset. The kernel estimator of the regression function is

$$\hat{\sigma}_t(F_t, X, \lambda | \mathbf{h}) = \frac{n^{-1} \sum_{j=1}^n K_{\mathbf{h}}(\tilde{z}_t - \tilde{z}_j) \tilde{\sigma}_j}{n^{-1} \sum_{j=1}^n K_{\mathbf{h}}(\tilde{z}_t - \tilde{z}_j)},$$

where  $\tilde{\sigma}_j$  is the volatility implied by the price of the call option, and  $\mathbf{h}$  is a vector of bandwidths. According to [Aït-Sahalia and Lo \(1998\)](#) and [Huynh, Kervella, and Zheng \(2002\)](#), the SPD and

the risk measures of delta ( $\Delta$ ) and Gamma ( $\Gamma$ ) are expressed as

$$f_{BS,t}(S_T) = \frac{1}{S_T \sqrt{2\pi\sigma^2\lambda}} \exp\left\{-\frac{[\ln(S_T/S_t) - (r_{t,\lambda} - \delta_{t,\lambda} - \sigma^2/2)\lambda]^2}{2\sigma^2\lambda}\right\},$$

$$\Delta_{BS} = \Phi(d_1),$$

$$\Gamma_{BS} = \frac{\phi(d_1)}{S_t \sigma \sqrt{\lambda}}.$$

where  $d_1 = \{\ln(S_t/X) + (r_{t,\lambda} - \delta_{t,\lambda} + \sigma^2/2)\lambda\} / (\sigma\lambda^{1/2})$ .

[Zhang et al. \(2009\)](#) assumed that the errors of the nonparametric regression model of  $\tilde{\sigma}_t$  on  $\tilde{z}_t$  follow the Gaussian distribution with a zero mean and unknown variance. In this paper, we assume that the errors of this nonparametric regression model follow the mixture of Gaussian densities given by (2). We fitted the model to the S&P 500 index options data, which are the same as those investigated by [Aït-Sahalia and Lo \(1998\)](#) and [Zhang et al. \(2009\)](#). The sample period is from the 4th January to the 31st December 1993, and the sample size is  $n = 14,431$ . We applied the sampling procedure proposed in Section 2.2 to sample the re-parameterized bandwidths from their posterior. Table 6 presents the estimates of the re-parameterized bandwidths and some associated statistics. Transforming the estimated  $\tau$  values into bandwidths according to (4), we found that the bandwidth vector for the NW estimator is  $(4.9810, 4.7884, 13.8418)'$ . This is clearly different from  $(5.6243, 5.4831, 9.7509)'$  derived under the assumption of Gaussian error density. The bandwidth for the kernel-form error density is 0.2424.

The Bayes factor of the mixture Gaussian error density against the Gaussian error density is  $\exp(4210.13)$ , which is very strong evidence supporting the former. Using the bandwidth vector derived by [Aït-Sahalia and Lo \(1998\)](#), and the ones estimated through Bayesian sampling under both error densities, we plotted the graphs of the SPD, the risk measures  $\Delta$  and  $\Gamma$  at maturities of 2 and 10 days, respectively, in Figure 3. At the maturity of 2 days, the SPD and  $\Gamma$  produced through the bandwidth vector derived under the mixture Gaussian error density are respectively, different from those derived under the Gaussian error density. However, as the time to maturity increases to 10 days, both densities lead to similar estimates of the SPD and  $\Gamma$ .

Moreover, the SPD,  $\Delta$  and  $\Gamma$  derived through Bayesian sampling under each assumption of the error density are clearly different from those derived through the rule-of-thumb reported by [Ait-Sahalia and Lo \(1998\)](#). However, different assumptions of the error density lead to similar estimates of the SPD, risk measures  $\Delta$  and  $\Gamma$  when maturity is 25 days or more.

## 6 Conclusion

We have presented a Bayesian approach to the estimation of bandwidths for the Nadaraya-Watson regression estimator and kernel-form error density estimator in the nonparametric regression model. The unknown error density is approximated by the mixture of Gaussian densities centered at individual error realizations and scaled by a constant parameter or the bandwidth in the context of kernel smoothing. We have conducted a series of simulation studies, which reveal that our approach outperforms the normal reference rule, likelihood cross-validation and bootstrapping methods in estimating the bandwidths for the Nadaraya-Watson estimator (as measured by ASE). In comparison to the parametric assumption of either the Gaussian or Student  $t$  error density, our proposed mixture Gaussian error density or equivalently the kernel-form density estimator of residuals, is not favored against the correct error-density assumption, but is favored against wrong assumptions of the error density. Our Bayesian sampling procedure represents a data-driven solution to the problem of simultaneously estimating bandwidths for the kernel estimators of the regression function and error density.

Applying to the nonparametric regression of the All Ordinaries daily return on the overnight FTSE and S&P 500 returns, we have obtained the bandwidth estimates for the kernel estimator of the regression under the three error-density assumptions. Although the assumption of a Gaussian-component mixture error density performs on par with the assumption of the Student  $t$  error density, they both outperform the Gaussian error-density assumption. The density estimator of the All Ordinaries daily return obtained through the mixture Gaussian error density exhibits a more reasonable left-tail behavior than that obtained through either the Gaussian or Student  $t$  error density. Moreover, under the Gaussian-component mixture

error density, we are able to estimate the density of the one-step-ahead point forecast of the All Ordinaries return. Therefore, the resulting value-at-risk is distribution-free and gain robustness in terms of different specifications of the error density.

We have also employed the proposed Bayesian sampling algorithm to estimate bandwidths for the nonparametric regression model involved in the state-price density estimation, where the error density is approximated by the Gaussian-component mixture density of the errors. The assumption of such an error density is favored with very strong evidence against the assumption of Gaussian error density. Moreover, we have found that the state-price density, risk measures  $\Delta$  and  $\Gamma$  estimated under this mixture error density are different from the corresponding ones estimated under Gaussian assumption errors at short maturities of the underlying asset. This example confirms the usefulness of relaxing the Gaussian assumption of the error density to a kernel-form or equivalently the Gaussian-component mixture error density in a nonparametric regression model.

## **Acknowledgements**

We extend our sincere thanks to John Geweke and Peter Robinson for their very insightful comments on an earlier draft of this paper. Thanks also go to the Victorian Partnership for Advanced Computing (VPAC) for its quality computing facility. This research was supported under the Australian Research Council's *Discovery Projects* funding scheme (project number DP1095838).

## References

- [Aït-Sahalia, Y., Lo, A. W., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. The Journal of Finance 53 \(2\), 499–547.](#)
- [Bowman, A. W., Azzalini, A., 1997. Applied Smoothing Techniques for Data Analysis. Oxford University Press, London.](#)
- [Brewer, M. J., 2000. A Bayesian model for local smoothing in kernel density estimation. Statistics and Computing 10 \(4\), 299–309.](#)
- [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.](#)
- [Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90 \(432\), 1313–1321.](#)
- [de Lima, M. S., Atuncar, G. S., 2011. A Bayesian method to estimate the optimal bandwidth for multivariate kernel estimator. Journal of Nonparametric Statistics 23 \(1\), 137–148.](#)
- [Efromovich, S., 2005. Estimation of the density of regression errors. The Annals of Statistics 33 \(5\), 2194–2227.](#)
- [Gangopadhyay, A., Cheung, K., 2002. Bayesian approach to the choice of smoothing parameter in kernel density estimation. Journal of Nonparametric Statistics 14 \(6\), 655–664.](#)
- [Gelfand, A. E., Dey, D. K., 1994. Bayesian model choice: Asymptotics and exact calculations. Journal of the Royal Statistical Society, Series B 56 \(3\), 501–514.](#)
- [Geweke, J., 2009. Complete and Incomplete Econometric Models. Princeton University Press, New Jersey.](#)
- [Geweke, J. F., 1999. Using simulation methods for Bayesian econometric models: Inference, development, and communication. Econometric Reviews 18 \(1\), 1–73.](#)



- [Hall, P., Lahiri, S. N., Polzehl, J., 1995. On bandwidth choice in nonparametric regression with both short-and long-range dependent errors. The Annals of Statistics 23 \(6\), 1921–1936.](#)
- [Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.](#)
- [Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single-index models. The Annals of Statistics 21 \(1\), 157–178.](#)
- [Härdle, W., Marron, J. S., 1985. Optimal bandwidth selection in nonparametric regression function estimation. The Annals of Statistics 13 \(4\), 1465–1481.](#)
- Härdle, W., Müller, M., 2000. Multivariate and semiparametric kernel regression. In: Schimek, M. G. (Ed.), *Smoothing and Regression: Approaches, Computation, and Application*. John Wiley & Sons, New York, pp. 357–392.
- [Herrmann, E., Engel, J., Wand, M. P., Gasser, T., 1995. A bandwidth selector for bivariate kernel regression. Journal of the Royal Statistical Society, Series B 57 \(1\), 171–180.](#)
- [Huynh, K., Kervella, P., Zheng, J., 2002. Estimating state price densities with nonparametric regression. In: Härdle, W., Kleinow, T., Stahl, T. \(Eds.\), \*Applied Quantitative Finance\*. Springer Verlag, Heidelberg.](#)
- [Jeffreys, H., 1961. Theory of Probability. Oxford University Press, Oxford, U.K.](#)
- Kass, R. E., Raftery, A. E., 1995. Bayes factors. *Journal of the American Statistical Association* 90 (430), 773–795.
- Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65 (3), 361–393.
- [Linton, O., Xiao, Z., 2007. A nonparametric regression estimator that adapts to error distribution of unknown form. Econometric Theory 23 \(3\), 371–413.](#)
- [Muhsal, B., Neumeier, N., 2010. A note on residual-based empirical likelihood kernel density estimation. Electronic Journal of Statistics 4, 1386–1401.](#)

- Newton, M. A., Raftery, A. E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society, Series B 56 (1), 3–48.
- 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 & Hall, London, pp. 45–57.
- Rothe, C., 2009. Semiparametric estimation of binary response models with endogenous regressors. Journal of Econometrics 153 (1), 51–64.
- Samb, R., 2010. Nonparametric kernel estimation of the probability density function of regression errors using estimated residuals. working paper, Universite Pierre et Marie Curie, LSTA.  
URL <http://arxiv4.library.cornell.edu/abs/1010.0439>
- Scott, D. W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.
- Silverman, B. W., 1978. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. The Annals of Statistics 6 (1), 177–184.
- Silverman, B. W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York.
- 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.
- Wahba, G., Wold, S., 1975. A completely automatic French curve: Fitting spline functions by cross validation. Communications in Statistics — Theory and Methods 4 (1), 1–17.
- Wand, M. P., Jones, M. C., 1995. Kernel Smoothing. Chapman and Hall, New York.
- Yuan, A., de Gooijer, J. G., 2007. Semiparametric regression with kernel error model. Scandinavian Journal of Statistics 34 (4), 841–869.

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., 2010. Bayesian semiparametric GARCH models. *Bayes on the Beach*, 4–5 October, Gold Coast, Australia.

URL [www.ccdamc.org.au/utilities/news/beachbayes2010.jsp](http://www.ccdamc.org.au/utilities/news/beachbayes2010.jsp)

Table 1: Parameter estimates and their statistics for Bayesian bandwidth estimation under mixture Gaussian, Gaussian and Student  $t$  error densities for a simulated sample with sample size  $n = 1,000$ . LML refers to log marginal likelihood.

Error density	Parameter	Estimate	95% Bayesian credible interval	Standard deviation	Batch-mean standard dev	SIF
Mixture	$\tau_b$	0.8991	(0.6319, 1.2015)	0.1461	0.0034	5.45
	$\tau_1$	0.1904	(0.1591, 0.2275)	0.0176	0.0007	16.63
	$\tau_2$	0.2620	(0.2181, 0.3090)	0.0234	0.0007	9.88
	$\tau_3$	0.5263	(0.4160, 0.6578)	0.0604	0.0021	11.80
	LML	-1389.99				
	ASE	0.0515				
Gaussian	$\sigma$	0.9881	(0.9460, 1.0332)	0.0220	0.0002	1.24
	$\tau_1$	0.1908	(0.1553, 0.2322)	0.0194	0.0006	10.44
	$\tau_2$	0.2640	(0.2169, 0.3140)	0.0238	0.0008	10.01
	$\tau_3$	0.5733	(0.4239, 0.7411)	0.0792	0.0029	12.96
	LML	-1422.78				
	ASE	0.0523				
Student $t$	$\nu$	13.7022	(9.1922, 20.1306)	2.7330	0.0683	6.25
	$\tau_1$	0.2003	(0.1623, 0.2463)	0.0207	0.0006	9.40
	$\tau_2$	0.2683	(0.2171, 0.3185)	0.0270	0.0009	11.57
	$\tau_3$	0.5356	(0.3994, 0.7038)	0.0775	0.0033	18.32
	LML	-1410.47				
	ASE	0.0520				

Table 2: Mean and standard deviation (sd) of 1,000 ASE derived through each bandwidth estimation method based on 1,000 generated samples with errors simulated from three specified densities.

Source of simulated errors		NRR	CV	Bootstrap	Bayesian		
					Gaussian	Student $t$	Mixture
$N(0, 0.9^2)$	Mean	0.0721	0.0701	0.0900	0.0582	0.0591	0.0585
	sd	0.0089	0.0107	0.0161	0.0090	0.0092	0.0091
Student $t$ with $\nu=4$	Mean	0.1113	0.0963	0.2377	0.0950	0.0914	0.0930
	sd	0.0189	0.0192	0.1167	0.0195	0.0175	0.0183
$0.7N(0, 0.7^2) + 0.3N(0, 1.5^2)$	Mean	0.0794	0.0754	0.1136	0.0664	0.0663	0.0656
	sd	0.0103	0.0120	0.0244	0.0108	0.0107	0.0105

Table 3: The mean and standard deviation (given in parentheses) of the 1,000 estimates of parameters obtained through Bayesian sampling under three assumptions of the error density in the nonparametric regression model given by (16), where LML refers to log marginal likelihood. The priors of  $\tau_k^2$ , for  $k = 0, 1, 2$  and  $3$ , are given by (11) and (12).

Assumption of error density	Gaussian errors		Student $t$ errors		Mixture Gaussian errors	
	Gaussian	Student $t$	Gaussian	Student $t$	Gaussian	Student $t$
<u>Estimation</u>						
$\sigma/\nu/\tau_0$	0.9342 (0.0211)	21.3173 (0.6967)	0.9043 (0.1376)	3.9948 (0.3915)	1.4026 (0.4059)	0.9934 (0.1716)
$\tau_1$	0.2252 (0.0147)	0.2312 (0.0142)	0.2253 (0.0162)	0.2616 (0.0210)	0.2665 (0.0255)	0.2325 (0.0176)
$\tau_2$	0.2423 (0.0149)	0.2477 (0.0140)	0.2433 (0.0168)	0.2795 (0.0219)	0.2825 (0.0245)	0.2500 (0.0183)
$\tau_3$	0.5023 (0.0535)	0.5019 (0.0513)	0.5037 (0.0590)	0.5874 (0.1022)	0.5879 (0.0933)	0.5211 (0.0644)
LML	-1364.74 (22.32)	-1375.97 (18.19)	-1367.10 (22.34)	-1725.14 (32.14)	-1757.96 (58.89)	-1450.87 (28.61)
<u>Bayesian Decision</u>						
$\text{BF} \leq 1/150$	0.0%	0.0%	1.0%	0.0%	0.7%	10.8%
$1/150 < \text{BF} \leq 1/20$	0.1%	0.1%	1.7%	0.0%	0.6%	5.5%
$1/20 < \text{BF} \leq 1/3$	0.1%	0.1%	6.0%	0.0%	1.4%	9.7%
$1/3 < \text{BF} \leq 1$	0.0%	0.0%	8.3%	0.0%	1.9%	5.6%
$1 < \text{BF} \leq 3$	0.0%	0.0%	12.6%	0.0%	1.1%	7.7%
$3 < \text{BF} \leq 20$	1.0%	1.0%	29.8%	0.0%	4.2%	11.5%
$20 < \text{BF} \leq 150$	4.8%	4.8%	25.0%	0.0%	5.5%	12.0%
$\text{BF} > 150$	94.0%	94.0%	15.6%	100.0%	84.6%	37.2%

Table 4: The mean and standard deviation (given in parentheses) of the 1,000 estimates of parameters obtained through Bayesian sampling under three assumptions of the error density in the nonparametric regression model given by (16), where LML refers to log marginal likelihood. The priors of bandwidths,  $b$  and  $h_k$ , for  $k = 1, 2$  and  $3$ , are the uniform densities given by (18) and (19).

Source of errors	Gaussian errors			Student $t$ errors			Mixture Gaussian errors		
	Assumption of error density			Assumption of error density			Assumption of error density		
Estimation	Gaussian	Student $t$	Mixture	Gaussian	Student $t$	Mixture	Gaussian	Student $t$	Mixture
$\sigma / \nu / b$	0.9327 (0.0211)	21.3664 (0.6947)	0.2755 (0.0459)	1.4449 (0.0989)	3.9964 (0.3929)	0.4194 (0.1186)	1.0403 (0.0317)	12.1217 (1.9712)	0.3013 (0.0522)
$h_1$	0.0703 (0.0081)	0.0715 (0.0080)	0.0703 (0.0088)	0.0853 (0.0135)	0.0845 (0.0110)	0.0845 (0.0129)	0.0739 (0.0096)	0.0745 (0.0087)	0.0736 (0.0091)
$h_2$	0.0788 (0.0084)	0.0802 (0.0082)	0.0793 (0.0093)	0.0926 (0.0133)	0.0941 (0.0118)	0.0935 (0.0134)	0.0826 (0.0096)	0.0836 (0.0087)	0.0826 (0.0096)
$h_3$	0.1993 (0.0253)	0.2112 (0.0314)	0.1998 (0.0278)	0.2392 (0.0433)	0.2685 (0.0620)	0.2386 (0.0410)	0.2074 (0.0303)	0.2294 (0.0437)	0.2067 (0.0294)
LML	-1358.34 (22.67)	-1369.40 (18.37)	-1360.98 (22.67)	-1794.40 (60.63)	-1721.75 (32.41)	-1754.65 (58.82)	-1467.34 (30.54)	-1447.77 (25.90)	-1445.62 (28.95)
<b>Bayesian decision</b>									
$\text{BF} \leq 1/150$	0.0%	0.0%	0.6%	0.0%	0.0%	0.1%	0.0%	6.4%	0.0%
$1/150 < \text{BF} \leq 1/20$	0.0%	0.0%	2.1%	0.0%	0.0%	0.6%	0.1%	4.6%	0.1%
$1/20 < \text{BF} \leq 1/3$	0.0%	0.0%	4.3%	0.0%	0.0%	0.9%	0.0%	5.6%	0.0%
$1/3 < \text{BF} \leq 1$	0.2%	0.2%	6.3%	0.0%	0.0%	0.9%	0.2%	3.8%	0.2%
$1 < \text{BF} \leq 3$	0.0%	0.0%	11.4%	0.0%	0.0%	1.2%	0.1%	5.9%	0.1%
$3 < \text{BF} \leq 20$	0.1%	0.1%	27.0%	0.0%	0.0%	2.7%	0.2%	12.1%	0.2%
$20 < \text{BF} \leq 150$	0.8%	0.8%	31.7%	0.0%	0.0%	4.1%	0.8%	11.8%	0.8%
$\text{BF} > 150$	98.9%	98.9%	16.6%	100.0%	89.5%	89.5%	98.6%	49.8%	98.6%

Table 5: Estimated parameters and their associated statistics of Bayesian bandwidth estimation with the mixture Gaussian, Gaussian and Student  $t$  error densities. LML represents the log marginal likelihood.

Error density	Parameter	Estimate	95% Bayesian credible interval	Standard deviation	Batch-mean standard dev	SIF
Mixture	$\tau_0$	1.2286	(0.8503, 1.6091)	0.1978	0.0085	18.48
	$\tau_1$	1.5341	(1.1036, 1.8758)	0.2081	0.0110	27.83
	$\tau_2$	2.3616	(1.9774, 2.8225)	0.2169	0.0090	17.18
	LML	-1484.43				
Gaussian	$\sigma$	1.0560	(1.0127, 1.1013)	0.0233	0.0002	0.81
	$\tau_1$	1.9804	(1.7134, 2.2812)	0.1445	0.0067	21.47
	$\tau_2$	2.5214	(2.1956, 2.9128)	0.1880	0.0089	22.49
	LML	-1514.80				
Student $t$	$\nu$	9.9630	(7.1845, 14.3219)	1.7992	0.0443	6.06
	$\tau_1$	1.9116	(1.6346, 2.2549)	0.1630	0.0066	16.48
	$\tau_2$	2.2941	(1.9626, 2.6650)	0.1856	0.0080	18.58
	LML	-1485.41				

Table 6: Estimates of the re-parameterized bandwidths and their statistics: S&P500 index options data.

Parameters	Mean	95% Bayesian credible interval	Standard deviation	Batch-mean standard error	SIF
$\tau_0$	1.3140	(1.2357, 1.4023)	0.0433	0.0010	5.88
$\tau_1$	19.5661	(18.9537, 20.2791)	0.3450	0.0110	10.19
$\tau_2$	18.8094	(17.7145, 19.8321)	0.5279	0.0237	20.09
$\tau_3$	54.3724	(52.3348, 56.7741)	1.0790	0.0381	12.48

Figure 1: Graphs of the true and estimated densities of regression errors.

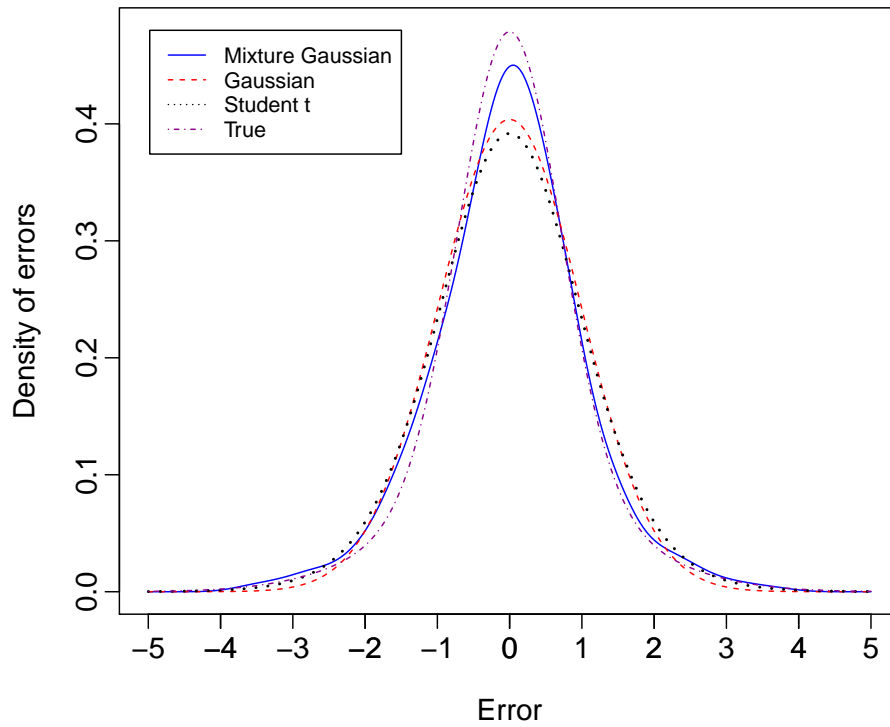


Figure 2: Estimated densities of the forecasted All Ordinaries return on the 31st March 2011.

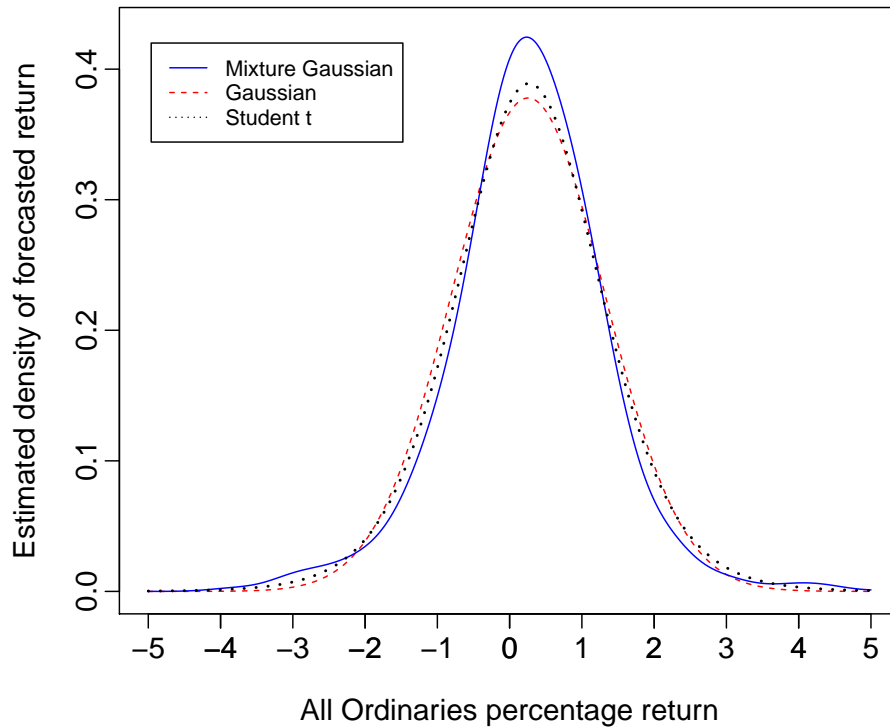




Figure 3: Graphs of the estimated state-price density, risk measures  $\Delta$  and  $\Gamma$  based on S&P 500 index options data. The first column is for the maturity of 2 days, and the second column is for the maturity of 10 days. AL denotes graphs obtained through the bandwidth vector provided by [Aït-Sahalia and Lo \(1998\)](#).

