Gaussian Kernel GARCH Models

XIBIN ZHANG, MAXWELL L. KING[1]

Department of Econometrics and Business Statistics, Monash University, Australia

June 19, 2014

ABSTRACT: This paper develops Bayesian sampling algorithms for parameter estimation in a GARCH model with a Gaussian kernel density for the errors. This study is motivated by the lack of robustness in GARCH models with a parametric assumption for the error density when used for error-density based inference such as value-at-risk estimation. A contribution is the construction of the likelihood and posterior and the derivation of the one-step-ahead posterior predictive density of asset returns. We also investigate the use of localized bandwidths in the Gaussian kernel error density. Applying this GARCH model to daily returns of 42 assets in stock, commodity and currency markets, we find that this GARCH model is favored against the GARCH model with a skewed Student t error density for all stock indices, two out of 11 currencies and nearly half of the commodities. This provides an empirical justification for the value of the proposed GARCH model.

KEYWORDS: Gaussian kernel error density, marginal likelihood, Markov chain Monte Carlo, posterior predictive density, value-at-risk.

[1]Address: Wellington Road, Clayton, Victoria 3800, Australia. Telephone: +61 3 99050082. Fax: +61 3 99055474. Email: *max.king@monash.edu*.

The autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986) have proven to be very useful in modelling volatilities of financial asset returns, and the assumption of conditional normality of the error term has contributed to early successes of GARCH models. Weiss (1986) and Bollerslev and Wooldridge (1992) showed that under this assumption, the quasi maximum likelihood estimator (QMLE) of the vector of parameters is consistent when the first two moments of the underlying GARCH process are correctly specified. However, the Gaussian QMLE suffers from efficiency loss when the conditional error density is non-Gaussian. Engle and González-Rivera (1991) investigated the efficiency loss through Monte Carlo simulations when the conditional error distribution is non-Gaussian. In the GARCH literature, evidence found by theoretical and empirical studies has shown that it is possible to reject the assumption of conditional normality (Singleton and Wingender, 1986; Bollerslev, 1986; Badrinath and Chatterjee, 1988, among others). This has motivated the investigation of other specifications of the conditional distribution of errors in GARCH models, such as the Student t and other heavy-tailed densities (see for example, Hall and Yao, 2003). In this paper, we investigate the estimation of parameters and error density in a GARCH model with an unknown error density.

A parametric distributional assumption of the error density has the benefit of simplicity in obtaining some theoretical results, but is likely to suffer from the problem of being a poor fit to the sample. Therefore, one may wish to improve the fit through a flexible distribution. Engle and González-Rivera (1991) highlighted the importance of investigating the issue of nonparametric estimation of the conditional density of errors at the same time the parameters are estimated. They proposed a semiparametric GARCH model without any assumption on the analytical form of the error density. The error density was estimated by the discrete maximum penalized likelihood estimator (DMPLE) of Tapia and Thompson (1978) based on residuals, which were calculated either by ordinary least squares or QMLE (under conditional normality). The parameters of the semiparametric GARCH model were then estimated by maximizing the log likelihood function constructed through the estimated error density based on initially derived residuals. Engle and González-Rivera's (1991) simulation results showed that this approach could improve the efficiency of parameter estimates by up to 50% compared to QMLEs obtained under conditional normality. However, their likelihood function is affected by initial parameter estimates, which might be inaccurate. Also, their parameter estimates are not used again to improve the accuracy of the error density estimator.

This paper aims to investigate how we can simultaneously estimate the parameters and conditional error density using information provided by the data without specifying the form of the error density. It is very attractive to impose minimal assumptions on the form of the error density in a GARCH model, because the resulting semiparametric model would gain robustness in terms of specifications of the error density (see for example, Durham and Geweke, 2014). In this situation, being able to estimate the error density is as important as estimating the GARCH parameters because any error-density-based inference would be robust with respect to the specification of the error density. Moreover, we can forecast the density of the underlying asset's return. Let $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$ be a vector of n observations of an asset's return. A GARCH(1,1) model is expressed as

$$\begin{aligned} y_t &= \sigma_t\, \varepsilon_t, \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{1}$$

where ε_t, for $t = 1, 2, \cdots, n$, are independent zero-mean errors. It is often assumed that $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$, and that conditional on information available at $t-1$ denoted as I_{t-1}, ε_t follows a known distribution. Strictly speaking, we will never know the true density of ε_t. To estimate parameters and make statistical inferences, the error density is usually assumed to be of a known form such as the standard Gaussian or Student t density. Any assumed density is only an approximation to the true unknown error density. In this paper, we assume that the unknown density of ε_t, denoted as $f(\varepsilon_t)$, is approximated by

$$\widetilde{f}(\varepsilon_t | h) = \frac{1}{n} \sum_{j=1}^{n} \phi \frac{1}{h} \left(\frac{\varepsilon_t - \varepsilon_j}{h} \right). \tag{2}$$

It is a location-mixture density of n Gaussian components, which have a common standard deviation h and different mean values at individual errors. Therefore, it is a well-defined probability density function characterized by h. From the view of kernel smoothing, $\widetilde{f}(\varepsilon_t|h)$ is the kernel error-density estimator based on errors, $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\}$. The performance of this kernel-form error density would be only second to that of an oracle who knows the true error density.

From a Bayesian's view, conditional on the parameters that characterize the GARCH model, this kernel-form error density can be used to construct the likelihood and therefore, the posterior. In the literature, the use of a scale-mixture density of several Gaussian densities as the error density in a regression model has been investigated, where the Gaussian components are usually assumed to have a zero mean and different variances (see Jensen and Maheu, 2013, among others). Therefore, this type of error density is at the cost of dramatically increasing the number of parameters. In contrast, our kernel-form error density places its locations at the individual realized errors and has only one parameter, which is the bandwidth.

Instead of using frequency approaches to investigate parameter estimation, we propose to derive an approximate posterior of the GARCH parameters and the bandwidths up to a normalizing constant, where the likelihood is approximated through the Gaussian kernel density of the errors. Bayesian sampling techniques have been used to estimate parameters of a GARCH model when the error density is specified (see for example, Bauwens and Lubrano, 1998; Nakatsuma, 2000; Vrontos, Dellaportas, and Politis, 2000). However, the posterior of the parameters, on which those sampling methods were developed, is unavailable when the error density is unknown. In this situation, Ausín, Galeano, and Ghosh (2014) developed a Bayesian semiparametric approach to parameter estimation in GARCH models with a class of scale mixtures of Gaussian distributions with a Dirichlet process prior. There exists an extensive literature on Bayesian nonparametrics on modeling an unknown distribution. See for example, Hjort, Holmes, Müller, and Walker (2010).

To deal with possible misspecification of the error density and impose inequality constraints on some parameters in the quasi likelihood, Koop (1994) presented Bayesian semiparametric ARCH models, where the quasi likelihood was constructed through a sequence of complicated polynomials.

Our proposed Gaussian kernel error density is different from the kernel density estimator of pre-fitted residuals, which is often used to construct a quasi likelihood for adaptive estimation of parameters in many models including (G)ARCH models investigated by Linton (1993) and Drost and Klaassen (1997). The conclusion drawn from their investigations is that (G)ARCH parameters are approximately adaptively estimable. This type of estimation is often conducted in a two-step procedure that uses the data twice. Di and Gangopadhyay (2011) presented a semiparametric maximum likelihood estimator of parameters in GARCH models. All those methods for the semiparametric GARCH model are based on pre-fitted residuals, and therefore are second-stage methods. In contrast, our Gaussian kernel error density depends on the errors rather than the pre-fitted residuals.

The rest of the paper is organized as follows. In Section 1, we discuss the validity and benefit of the Gaussian kernel error density in GARCH models, derive the likelihood and posterior, and discuss extensions including the use of localized bandwidths and the incorporation of asymmetric effect of past squared returns. Section 2 presents a simulation study to demonstrate the performance of the Gaussian kernel GARCH model in comparison to the skewed t GARCH model of Hansen (1994). In Section 3, we apply the Gaussian kernel asymmetric GARCH(1,1) model to the S&P 500 daily returns. Section 4 presents a comprehensive study of the performance of the Gaussian kernel GARCH model and skewed t GARCH model to daily returns of another nine stock indices, ten currency prices and 21 futures prices. Section 5 concludes the paper.

1 A GARCH MODEL WITH GAUSSIAN KERNEL ERROR DENSITY

1.1 A Mixture of Gaussian Densities

Let $\{x_1, x_2, \cdots, x_n\}$ denote a sample of independent observations drawn from an unknown probability density function $g(x|\kappa)$ with unbounded support, where κ is the parameter vector. In order to make statistical inference based on the sample, one has to make assumptions about the analytical form of $g(x|\kappa)$ based on some descriptive statistics such as the histogram of observations.

Strictly speaking, any specification of the true density is only an approximation to $g(x|\kappa)$. One such approximation is given by

$$\widetilde{g}(x|h) = \int_{-\infty}^{\infty} g(z|\kappa) K_h(x-z)\, dz, \tag{3}$$

where $K_h(z) = K(z/h)/h$ with $K(\cdot)$ being a kernel function. This is a convolution-form kernel estimator of $g(x|\kappa)$ proposed by Parzen (1962). As a random sample is already observed from $g(x|\kappa)$, this kernel estimator can be approximated as

$$\widetilde{g}(x|h) \approx \frac{1}{n} \sum_{i=1}^{n} K_h(x - x_i),$$

which is commonly known as the kernel estimator of $g(x|\kappa)$ in the literature. Silverman (1978) proved that under some regularity conditions, $\widetilde{g}(x|h)$ is strongly uniformly consistent.

In this paper, we investigate how we can use this mixture density of Gaussian components as an approximation to the unknown error density in a regression model. Zhang, King, and Shang (2014) justified the use of such Gaussian error density in a nonparametric regression model. A realization of this mixture density of the errors is equivalent to the kernel density estimator of pre-fitted residuals, which is employed to construct a quasi likelihood for adaptive estimation in the sense of Bickel (1982). Therefore, parameters can be estimated by maximizing the quasi likelihood. One main issue of adaptive estimation is the efficiency of the resulting parameter estimates when the sample size increases. It has been found that parameters can be asymptotically adaptively estimable for a range of parametric models. However, a major problem in adaptive estimation is that the bandwidth has to be pre-chosen based on pre-fitted residuals through initial estimates of parameters. Therefore, the sample is used twice, and the chosen bandwidth depends on inaccurate initial estimates of parameters.

1.2 Conditional Gaussian Kernel Density of Errors

Consider the GARCH(1,1) model given by (1), in which we assume that $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$. Strictly speaking, the true density of ε_t denoted as $f(\varepsilon_t)$, is unknown. We propose to approximate the unknown error density of (1) by

$$\widetilde{f}(\varepsilon_t|h) = \int_{-\infty}^{\infty} f(z) K_h(\varepsilon_t - z)\, dz, \tag{4}$$

where the kernel function is chosen as the density of the standard Gaussian distribution denoted as $\phi(\cdot)$. As $f(\cdot)$ is unknown, the convolution-form kernel estimator of $f(\cdot)$ given by (4) is practically intractable. Nonetheless, it can be approximated as

$$\widetilde{f}(\varepsilon_t|h) \approx \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h}\phi\left(\frac{\varepsilon_t - \varepsilon_j}{h}\right), \tag{5}$$

which we call the Gaussian kernel error density and is specified as the error density of (1). The common standard deviation of the Gaussian components is also called the bandwidth due to its smoothing role. In addition to the parameters that characterize the parametric component of the GARCH model, we treat h as a parameter. The resulting GARCH model is referred to as the Gaussian kernel GARCH model.

Remark 1: From the view of error density specification, the Gaussian kernel density is a well-defined density function because it is a location-mixture density of n Gaussian component densities. These component Gaussian densities have means at individual errors and the same variances. If these component densities were allowed a constant mean at zero, this mixture density would become a Gaussian density with a zero mean and constant variance. Moreover, if $\varepsilon \sim \widetilde{f}(\varepsilon|h)$, we have

$$E(\varepsilon) = \overline{\varepsilon}, \qquad Var(\varepsilon) = h^2 + s_\varepsilon^2,$$

where $\overline{\varepsilon} = 1/n\sum_{i=1}^{n}\varepsilon_i$ and $s_\varepsilon^2 = 1/n\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2$.

This Gaussian kernel density is defined conditional on model parameters and is rewritten as

$$\widetilde{f}(\varepsilon_t|h) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h}\phi\left(\frac{\varepsilon_t - y_i/\sigma_i}{h}\right), \tag{6}$$

where $\sigma_i^2 = \omega + \alpha y_{i-1}^2 + \beta\sigma_{i-1}^2$, for $i = 1, 2, \cdots, n$. From a Bayesian's view, this mixture density has a closed form conditional on two types of parameters: the model parameters and smoothing parameter. Both the likelihood and posterior can be constructed through this error density.

However, the kernel density estimator of residuals used by adaptive estimation relies on the residuals calculated through the pre-estimated parameters.

Remark 2: When using the density of ε_t to construct the likelihood of $\boldsymbol{y}$, we use the leave-one-out version of the Gaussian kernel density given by

$$\widetilde{f}\left(\varepsilon_t|\varepsilon_{(t)},h\right)=\frac{1}{n-1}\sum_{\substack{i=1\\i\neq t}}^{n}\frac{1}{h}\phi\left(\frac{\varepsilon_t-\varepsilon_j}{h}\right), \tag{7}$$

where $\varepsilon_{(t)}$ is $\varepsilon=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)'$ without ε_t, for $t=1,2,\cdots,n$. The purpose of leaving ε_t out of the summation in (5) or (6) is to exclude $\phi(0/h)/h$, which can be made arbitrarily large by allowing h to be arbitrarily small. Otherwise, a numerical maximization of the likelihood with respect to h, or any posterior simulator based on the resulting posterior, would encounter problems.

Remark 3: The functional form of $f(\varepsilon_t|\varepsilon_{(t)},h)$ does not depend on t because it can also be expressed as

$$\widetilde{f}\left(\varepsilon_t|\varepsilon_{(t)},h\right)=\frac{1}{(n-1)h}\left\{\sum_{j=1}^{n}\phi\left(\frac{\varepsilon_t-\varepsilon_j}{h}\right)-\phi(0)\right\}, \tag{8}$$

for $t=1,2,\cdots,n$.

Remark 4: The density of y_t is approximated by

$$f_Y\left(y_t|\boldsymbol{y}_{(t)},\theta\right)=\frac{1}{(n-1)\sigma_t}\sum_{\substack{i=1\\i\neq t}}^{n}\frac{1}{h}\,\phi\left(\frac{y_t/\sigma_t-y_i/\sigma_i}{h}\right), \tag{9}$$

which is actually the leave-one-out kernel density estimator of y_t through the transformation of standardization, for $t=1,2,\cdots,n$. A kernel density estimator of the direct observations of $\boldsymbol{y}$ is likely to be inappropriate because the return series $\{y_t:t=1,2,\cdots,n\}$ are heteroskedastic. However, scaling the returns by their conditional standard deviations, a reasonable approximation is to assume the standardized returns are independent and identically distributed.

1.3 Likelihood

Let $\theta_0=\left(\omega,\alpha,\beta,\sigma_0^2\right)'$ denote the vector of parameters of the GARCH(1,1) model given by (1). When $f(\varepsilon)$ is known, the likelihood of $\boldsymbol{y}$ for given θ_0 is

$$\ell_0\left(\boldsymbol{y}|\theta_0\right)=\prod_{t=1}^{n}\frac{1}{\sigma_t}f\left(y_t/\sigma_t\right).$$

When the analytical form of $f(\varepsilon)$ is unknown, we propose to approximate the density of y_t by (9), where h and σ_t always appear in the form of the product of the two. We found that

$$h^2\sigma_t^2 = h^2\omega + h^2\alpha y_{t-1}^2 + \beta h^2\sigma_{t-1}^2, \tag{10}$$

where h^2 and ω, as well as h^2 and α, cannot be separately identified. If ω is assumed to be a known constant, all the other parameters can be separately identified.

For adaptive estimation of ARCH models, ω was restricted to be zero by Linton (1993) and one by Drost and Klaassen (1997). In light of the fact that the unconditional variance of y_t is $\omega/(1-\alpha-\beta)$, we assume that $\omega = (1-\alpha-\beta)s_y^2$, where $s_y^2 = (n-1)^{-1}\sum_{i=1}^n (y_t - \overline{y})^2$ is the sample variance of y_t. When the return series is pre-standardized, ω would be assumed to be $(1-\alpha-\beta)$, which is what Engle and González-Rivera (1991) assumed for ω in their GARCH model.

The starting value of the conditional variance series, σ_0^2, is unknown and is treated as a parameter. Therefore, under the Gaussian kernel error density, the GARCH(1,1) model becomes

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \omega_0 + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2, \end{aligned} \tag{11}$$

where $\omega_0 = (1-\alpha-\beta)s_y^2$, and ε_t follow the Gaussian kernel density characterized by h. The parameter vector is $\theta = \left(\sigma_0^2, \alpha, \beta, h^2\right)'$, and restrictions on the parameter space are that $0 \le \alpha < 1$, $0 \le \beta < 1$ and $0 \le \alpha + \beta < 1$.

The likelihood of $\boldsymbol{y}$, for given θ, is approximately

$$\ell(\boldsymbol{y}|\theta) = \prod_{t=1}^n f_Y\left(y_t|\boldsymbol{y}_{(t)}, \theta\right) = \prod_{t=1}^n \left\{ \frac{1}{(n-1)\sigma_t} \sum_{\substack{i=1 \\ i\neq t}}^n \frac{1}{h_n} \phi\left(\frac{y_t/\sigma_t - y_i/\sigma_i}{h_n} \right) \right\}, \tag{12}$$

which is an approximate likelihood of $\boldsymbol{y}$ for given θ. Conditional on model parameters, this likelihood function is the one used by the likelihood cross-validation in choosing bandwidth for the kernel density estimator of the standardized y_i, for $i = 1, 2, \cdots, n$ (see for example, Bowman, Hall, and Titterington, 1984).

Remark 5: Observe that the approximate likelihood given by (12) has the form of a full conditional composite likelihood in the sense that the density of y_t is defined conditional on $\boldsymbol{y}_{(t)}$.

This feature has not been noted in the current literature even though the composite likelihood has been extensively investigated. See for example, Varin, Reid, and Firth (2011) and Mardia, Kent, Hughes, and Taylor (2009) for an overview of and discussion about composite likelihood.

Remark 6: The likelihood function given by (12) is related to the so-called kernel likelihood functions derived by Yuan and de Gooijer (2007) and Yuan (2009) for semiparametric regression models and by Grillenzoni (2009) for dynamic time-series regression models, where their likelihood functions were derived based on pre-fitted residuals. In contrast, our likelihood given by (12) is constructed based on a well-defined error density given by (7).

If the proposed Gaussian kernel error density is to replace the DMPLE density estimator in the semiparametric estimation procedure suggested by Engle and González-Rivera (1991), the implementation of their estimation method becomes an issue of choosing bandwidth and maximizing the quasi likelihood, which was constructed through the nonparametrically estimated error density, with respect to the parameters. It is possible to maximize the constructed likelihood with respect to the parameters and bandwidth. Therefore, their initial parameter estimates have no effect on the resulting parameter estimates that maximize the quasi likelihood. However, numerical optimization of the approximate likelihood sometimes encounters a convergence problem due to the Hessian matrix failing to invert. Nonetheless, we confine our investigation within posterior simulation.

1.4 Priors

The prior of each component of θ should be a proper prior because later on, we will compute the marginal likelihood and use Bayes factors for decision making. In the empirical finance literature, the magnitude of β reflects the persistency level of the volatility of the underlying asset returns. Many empirical studies reveal that the value of β is often close to one. Therefore, we assume that the prior of β is the density of the Beta distribution given as

$$p\left(\beta\right)=\frac{\Gamma\left(a_\beta+b_\beta\right)}{\Gamma\left(a_\beta\right)\Gamma\left(b_\beta\right)}\beta^{a_\beta-1}\left(1-\beta\right)^{b_\beta-1},$$

where a_β and b_β are hyperparameters, which we choose to set to 10 and 2, respectively. Such a choice reflects the empirical finding that the value of β is usually close to one. An alternative choice of prior for β is the uniform density on $(0,1)$, which also works very well.

As the constraint imposed on α is $0 < \alpha < 1-\beta$. Therefore, its prior density is assumed to be uniform on $(0, 1-\beta)$.

The prior of h^2 is chosen to be the inverse Gamma density denoted as $\mathrm{IG}(a_h, b_h)$ with its density given by

$$p\left(h^2\right) = \frac{b_h^{a_h}}{\Gamma(a_h)}\left(\frac{1}{h^2}\right)^{a_h+1}\exp\left\{-\frac{b_h}{h^2}\right\}, \tag{13}$$

where a_h and b_h are hyperparameters, which can be chosen to be 1 and 0.05, respectively. The appropriateness of the inverse Gamma prior is due to the fact that the squared bandwidth h^2 is the variance parameter of the component densities. The prior of the variance of a Gaussian distribution is usually the inverse Gamma density (see for example, Geweke, 2009). An alternative choice of prior for h^2 is the exponential density given by

$$p\left(h^2\right) = \delta\exp\left(-\delta h^2\right),$$

where the hyperparameter δ is chosen to be 1 in this situation. If h is treated as a parameter, its prior density can be chosen as either the standard Cauchy density truncated above zero or the exponential density. In our experience, both priors work very well.

The prior of σ_0^2 is assumed to be the log normal density with mean zero and variance one. An alternative choice of prior for this parameter is the $\mathrm{IG}(1, 0.05)$ density. In our experience, the estimate of θ, as well as the error-density estimator, is insensitive to the prior choice of σ_0^2.

The joint prior of θ denoted as $p(\theta)$, is the product of the marginal priors of α, β, h^2 and σ_0^2.

1.5 Posterior of Parameters

The posterior of θ for given $\boldsymbol{y}$ is proportional to the product of the joint prior of θ and the likelihood of $\boldsymbol{y}$ given θ. In the re-parameterized GARCH model given by (11), the posterior of θ is (up to a normalizing constant)

$$\boldsymbol{\pi}(\theta|\boldsymbol{y}) \propto p(\theta)\times\ell(\boldsymbol{y}|\theta), \tag{14}$$

which is well explained in terms of conditional posteriors. Conditional on h^2, the Gaussian kernel density of the errors is well defined, and the posterior of $\left(\widetilde{\alpha},\beta,\sigma_0^2\right)'$ can be derived. Similarly, conditional on $\left(\widetilde{\alpha},\beta,\sigma_0^2\right)'$, we can compute the errors, or equivalently the standardized returns, and then derive the posterior of h^2 constructed through the Gaussian kernel error density.

We use the Markov chain Monte Carlo (MCMC) simulation technique to sample θ from its posterior given by (14). The random-walk Metropolis algorithm is used to simulate θ, and the sampling procedure is as follows.

Step 1: Randomly choose initial values denoted as $\theta^{(0)}$.

Step 2: Update θ using the random-walk Metropolis algorithm with the proposal density being the standard Gaussian and the acceptance probability computed through $\pi(\theta|\boldsymbol{y})$. Let $\theta^{(1)}$ denote the updated θ.

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

Upon completing the sampling procedure, we use the ergodic average of the sampled values of θ as an estimate of θ. The density of standardized returns can be derived through averaging. At each iteration, we calculate the density function at a number of grid points by plugging in the simulated value of θ. After completing all iterations, we take an average of the density value calculated at each grid point over all iterations. Alternatively, as the density of standardized returns has a closed form, the density can be estimated by plugging in the estimated θ.

The second step of the sampling procedure can also be implemented as follows. First, conditional on the current value of h^2, we update $(\alpha,\beta,\sigma_0^2)$ using the random-walk Metropolis algorithm with the acceptance probability computed through (14). This sampling algorithm is the same as that developed by Zhang and King (2008) for the GARCH(1,1) model with its Student t density replaced by the Gaussian kernel density. Second, conditional on the updated $(\alpha,\beta,\sigma_0^2)$, we sample h^2 from the posterior given by (14) using the random-walk Metropolis algorithm. This algorithm is the same as that proposed by Zhang, King, and Hyndman (2006) for kernel density estimation of directly observed data, which are now replaced by the standardized returns.

1.6 Asymmetric GARCH Model with the Gaussian Kernel Error Density

It is generally believed that the past negative and positive returns may have asymmetric effects on the current conditional variance. Glosten, Jagannathan, and Runkle (1993) introduced an asymmetric GARCH model:

$$y_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 I\left(y_{t-1} \geq 0\right) + \alpha_2 y_{t-1}^2 I\left(y_{t-1} < 0\right) + \beta\sigma_{t-1}^2, \tag{15}$$

where $I(\cdot)$ is an indicator function that equals one when its argument is true and zero otherwise. The restrictions on parameters are $\omega > 0$, $0 \leq \alpha_1 < 1$, $0 \leq \alpha_2 < 1$, $0 \leq \beta < 1$ and $\max\{\alpha_1, \alpha_2\} + \beta < 1$. The errors $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are usually assumed to be iid and follow a known parametric distribution such as the standard Gaussian or Student t distribution.

We assume that $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are iid and follow a distribution with the Gaussian kernel density given by (6). We also assume that $\omega = \left(1 - \max\{\alpha_1, \alpha_2\} - \beta\right) s_y^2$ for identification purposes. The likelihood of $\boldsymbol{y}$ for given parameters, denoted as $\ell_{\text{GJR}}\left(\boldsymbol{y}|\alpha_1, \alpha_2, \beta, h^2\right)$, has the same form as (12) with its conditional variance modelled by (15).

The priors of β and h^2 are assumed to be same as those given in Section 1.4. The priors of α_1 and α_2 are assumed to be the uniform density defined on $(0, 1-\beta)$. The joint prior of all parameters is denoted as $p_{\text{GJR}}\left(\alpha_1, \alpha_2, \beta, h^2\right)$. The posterior is proportional to the product of the joint prior and likelihood and is expressed as

$$\boldsymbol{\pi}_{\text{GJR}}\left(\alpha_1, \alpha_2, \beta, h^2|\boldsymbol{y}\right) \propto p_{\text{GJR}}\left(\alpha_1, \alpha_2, \beta, h^2\right) \times \ell_{\text{GJR}}\left(\boldsymbol{y}|\alpha_1, \alpha_2, \beta, h^2\right), \tag{16}$$

from which we sample parameters using the random-walk Metropolis algorithm. The sampling procedure is the same as the one described in Section 1.5.

1.7 Localized Bandwidths for the Gaussian Kernel Density of Errors

In Section 1.2, we proposed using the leave-one-out Gaussian kernel error density to approximate the unknown error density. In terms of kernel density estimation of directly observed data, the leave-one-out estimator is known to be heavily affected by extreme observations (see for example, Bowman, 1984). Consequently, when the true error density has sufficiently long tails, the leave-one-out Gaussian kernel density with its bandwidth selected under the Kullback-Leibler criterion, is likely to over estimate its tail density. One may argue that this phenomenon is likely to be caused by the use of a global bandwidth. A possible remedy is to use variable bandwidths or localized bandwidths (see for example, Silverman, 1986).

The approximate likelihood given by (12) was built up through the leave-one-out Gaussian kernel density. In the empirical finance literature, there is evidence that the density of the standardized errors is heavy-tailed. Thus, we have to be cautious about large standardized errors when the Gaussian kernel error density is used for constructing the posterior for the GARCH model. We now investigate the use of localized bandwidths in the Gaussian kernel error density.

Recent developments on kernel density estimation of directly observed data with adaptive or variable bandwidths suggest that small bandwidths should be assigned to the observations in the high-density region and larger bandwidths should be assigned to those in the low-density region. One of the key issues on the use of adaptive bandwidths is how we choose different bandwidths for different groups of observations. Brewer (2000) suggested assigning different bandwidths to different observations and obtaining posterior estimates of the bandwidths. As bandwidths are treated as parameters, we do not want lots of parameters in addition to the model parameters.

In light of the above intuitive idea of using variable bandwidths for kernel density estimation, we assume that the underlying true error density is unimodal. Therefore, large absolute errors should be assigned relatively large bandwidths, while small absolute errors should be assigned relatively small bandwidths. Thus, we propose the following error density:

$$f_{\text{local}}(\varepsilon_t|h,h_\varepsilon) = \frac{1}{n-1}\sum_{\substack{i=1\\ i\neq t}}^{n} \frac{1}{h(1+h_\varepsilon|\varepsilon_i|)}\,\phi\left(\frac{\varepsilon_t-\varepsilon_i}{h(1+h_\varepsilon|\varepsilon_i|)}\right), \tag{17}$$

where $h(1+h_\varepsilon|\varepsilon_i|)$ is the bandwidth assigned to ε_i, for $i=1,2,\cdots,n$, and the vector of parameters is now $\theta_{\text{local}} = \left(\sigma_0^2,\alpha,\beta,h^2,h_\varepsilon\right)'$. The meaning of this kernel-form error density is also clear. The density of ε_t is approximated by a mixture of $n-1$ Gaussian densities with their means being at the other errors and variances localized.

The approximate likelihood of $\boldsymbol{y}$ for given θ_{local} is:

$$\ell_{\text{local}}\left(\boldsymbol{y}|\theta_{\text{local}}\right) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_{\text{local}}\left(y_t/\sigma_t | h, h_\varepsilon\right), \tag{18}$$

where σ_t^2 can be modelled by either (11) or (15).

We assume the prior of h_ε is the uniform density on $(0,1)$. The priors of the other parameters are the same as those in the situation of using a global bandwidth discussed in Sections 1.2 and 1.6. The joint prior of θ_{local} denoted as $p_{\text{local}}(\theta_{\text{local}})$, is the product of these marginal priors. Therefore, the posterior of θ_{local} is (up to a normalizing constant)

$$\pi_{\text{local}}\left(\theta_{\text{local}}|\boldsymbol{y}\right) \propto p_{\text{local}}(\theta_{\text{local}}) \times \ell_{\text{local}}\left(\boldsymbol{y}|\theta_{\text{local}}\right), \tag{19}$$

from which we sample θ_{local} using the random-walk Metropolis algorithm.

2 MONTE CARLO SIMULATION STUDY

2.1 Generate Samples

An argument in favor of the proposed specification of Gaussian kernel error density is that even though the true distribution is parametric, the flexible kernel model should perform reasonably well. Presumably, there would be an efficiency loss, for which the robustness would compensate, because we never know what the true distribution is in practice. Thus, we would like to study this issue through simulations.

We generate samples through the asymmetric GARCH(1,1) model given by (15), where the error density is the skewed Student t density proposed by Hansen (1994). This density function, which will be briefly discussed in Section 2.2, is characterized by two parameters, η and λ. To generate samples, we choose the parameter values as $\omega = 0.03, \alpha_1 = 0.01, \alpha_2 = 0.08, \beta = 0.9, \eta = 9$ and $\lambda = -0.2$. These values are very close to those estimated based on a sample of a stock index return series. We generated 1,000 samples, each of which has a size of 1,000.

2.2 A Competing Model

For comparison purposes, we choose the competing model as a GARCH model with a skewed Student t error density of Hansen (1994). The error density of the GARCH model given by (15) is

$$g\left(\varepsilon|\eta,\lambda\right)=\begin{cases} b_s c_s\left\{1+\frac{1}{\eta-2}\left[\left(b_s\varepsilon+a_s\right)/\left(1-\lambda\right)\right]^2\right\}^{-(\eta+1)/2} & \text{if } \varepsilon<-a_s/b_s \\ b_s c_s\left\{1+\frac{1}{\eta-2}\left[\left(b_s\varepsilon+a_s\right)/\left(1+\lambda\right)\right]^2\right\}^{-(\eta+1)/2} & \text{if } \varepsilon\geq-a_s/b_s \end{cases}, \tag{20}$$

where $\eta>2$, $-1<\lambda<1$, and

$$\begin{aligned} a_s &= 4\lambda c_s\left(\eta-2\right)/\left(\eta-1\right), \\ b_s &= \sqrt{1+3\lambda^2-a_s^2}, \\ c_s &= \Gamma((\eta+1)/2)\big/\left[\sqrt{\pi(\eta-2)}\,\Gamma(\eta/2)\right]. \end{aligned}$$

Hansen (1994) proved that the function given by (20) is a proper probability density function with a mean zero and variance one. He also proposed to allow λ and η to be time-varying in the following sense:

$$\begin{aligned} \eta_t &= \eta_a+\eta_b y_{t-1}+\eta_c y_{t-1}^2, \\ \lambda_t &= \lambda_a+\lambda_b y_{t-1}+\lambda_c y_{t-1}^2, \end{aligned}$$

where both η_t and λ_t should satisfy the above restrictions. Therefore, a logistic type transformation is applied to η_t and λ_t:

$$z_t = L_z+(U_z-L_z)\big/(1+\exp(-z_t)),$$

where z_t is either η_t or λ_t, and L_z and U_z are the corresponding lower and upper bounds.

The likelihood of $\boldsymbol{y}$ for given $\left(\omega,\alpha_1,\alpha_2,\beta,\eta,\lambda,\sigma_0^2\right)'$ is

$$\ell_{\text{skewt}}\left(\boldsymbol{y}|\omega,\alpha_1,\alpha_2,\beta,\eta,\lambda,\sigma_0^2\right)=\prod_{t=1}^{n} g\left(y_t/\sigma_t|\eta,\lambda\right)/\sigma_t^2, \tag{21}$$

where σ_t^2 is given by (15).

The prior of ω is the uniform density on $(0,1)$, and the priors of α_1, α_2, β and σ_0^2 are assumed to be the same as those discussed in Section 1.6. We assume that the prior of η is the density of $N(7,3^2)$ truncated at 2 from below to guarantee that η is greater than 2. The prior of λ is the uniform density defined on $(-1,1)$. When time-varying parameters are used in the skewed t distribution, the priors of η_a, η_b, η_c, λ_a, λ_b and λ_c are the density of $N(0,3^2)$. The joint prior of all parameters is the product of the marginal priors and is denoted as $p_{\text{skewt}}\left(\omega,\alpha_1,\alpha_2,\beta,\eta,\lambda,\sigma_0^2\right)$. Therefore, by Bayes theorem, the posterior of $\boldsymbol{y}$ is

$$\pi_{\text{skewt}}\left(\omega,\alpha_1,\alpha_2,\beta,\eta,\lambda,\sigma_0^2|\boldsymbol{y}\right) \propto p_{\text{skewt}}\left(\omega,\alpha_1,\alpha_2,\beta,\eta,\lambda,\sigma_0^2\right) \times \ell_{\text{skewt}}\left(\boldsymbol{y}|\omega,\alpha_1,\alpha_2,\beta,\eta,\lambda,\sigma_0^2\right),$$

from which we used the random-walk Metropolis algorithm to sample parameters.

2.3 Bayes Factors for Model Comparison

The Bayes factor for a model denoted as $\mathscr{A}_0$, against a competing model $\mathscr{A}_1$, is defined by (Spiegelhalter and Smith, 1982)

$$B_{01} = m(\boldsymbol{y}|\mathscr{A}_0)/m(\boldsymbol{y}|\mathscr{A}_1),$$

where $m(\boldsymbol{y}|\mathscr{A}_0)$ and $m(\boldsymbol{y}|\mathscr{A}_1)$ are marginal likelihoods derived under $\mathscr{A}_0$ and $\mathscr{A}_1$, respectively. Marginal likelihood is the expectation of the likelihood under the prior density of parameters and is often intractable in Bayesian inference. Nonetheless, there are several methods to numerically approximate the marginal likelihood (Newton and Raftery, 1994; Chib, 1995; Geweke, 1999; Chib and Jeliazkov, 2001, among others).

Let $\theta_{\mathscr{A}}$ denote a vector of parameters under the model $\mathscr{A}$, which can be either $\mathscr{A}_0$ or $\mathscr{A}_1$. The marginal likelihood is

$$m(\boldsymbol{y}|\mathscr{A}) = \int \ell_{\mathscr{A}}(\boldsymbol{y}|\theta_{\mathscr{A}})p_{\mathscr{A}}(\theta_{\mathscr{A}})d\theta_{\mathscr{A}}, \tag{22}$$

where $\ell_{\mathscr{A}}(\boldsymbol{y}|\theta_{\mathscr{A}})$ is the likelihood of $\boldsymbol{y}$, and $p_{\mathscr{A}}(\theta_{\mathscr{A}})$ is the prior of $\theta_{\mathscr{A}}$. Chib (1995) showed that the marginal likelihood under a model $\mathscr{A}$ is expressed as

$$\widetilde{m}_{\mathscr{A}}(\boldsymbol{y}) = \frac{\ell_{\mathscr{A}}(\boldsymbol{y}|\theta)p_{\mathscr{A}}(\theta)}{\pi_{\mathscr{A}}(\theta|\boldsymbol{y})}, \tag{23}$$

where $\pi_{\mathscr{A}}(\theta|\boldsymbol{y})$ is the posterior under model $\mathscr{A}$. $\widetilde{m}_{\mathscr{A}}(\boldsymbol{y})$ is usually computed at the posterior mean of the simulated chain of θ. The numerator has a closed form, while the denominator is often approximated by its kernel density estimator based on the simulated chain of θ.

Let $\widetilde{m}(\boldsymbol{y}|\mathscr{A}_0)$ and $\widetilde{m}(\boldsymbol{y}|\mathscr{A}_1)$ denote the marginal likelihoods derived under the Gaussian kernel GARCH(1,1) and skewed t GARCH(1,1) models, respectively. The Bayes factor of the former model against the latter is

$$\widetilde{BF}_{01} = \widetilde{m}(\boldsymbol{y}|\mathscr{A}_0) / \widetilde{m}(\boldsymbol{y}|\mathscr{A}_1).$$

According to Jeffreys' (1961) scales modified by Kass and Raftery (1995), when a Bayes factor value is between 1 and 3, the evidence supporting $\mathscr{A}$ against $\mathscr{B}$ is "not worth more than a bare mention". A Bayes factor value between 3 and 20 indicates that $\mathscr{A}$ is favored against $\mathscr{B}$ with positive evidence. When the Bayes factor is between 20 and 150, $\mathscr{A}$ is favored against $\mathscr{B}$ with strong evidence; and when the Bayes factor is greater than 150, $\mathscr{A}$ is favored against $\mathscr{B}$ with very strong evidence.

2.4 Simulation Results

For each generated sample, we estimated parameters of the Gaussian kernel asymmetric GARCH(1,1) model by implementing the proposed sampling algorithms, where both global and localized bandwidths were considered. We also estimated parameters of the GARCH(1,1) model with its error density being assumed to be the skewed Student t density. The marginal likelihood under each model was calculated, and the Bayes factor of the Gaussian kernel GARCH model with a global bandwidth (and localized bandwidths) against its competitor, the skewed t GARCH model, was then calculated.

Table 1 presents relative frequencies of simulated samples falling in different categories of Bayes factors. The Gaussian kernel GARCH model with a global bandwidth is favored against the skewed t GARCH model with very strong evidence in 2.7% of the simulated samples, strong evidence in 6.3% of the simulated samples, and positive evidence in 10.1% of the simulated samples. Therefore, the Gaussian kernel GARCH model is favored against its parametric competitor, the GARCH model with the correct assumption of error density, with at least positive evidence in 19.1% of the simulated samples. In contrast, the parametric GARCH model, from which the samples are generated, is favored against the Gaussian kernel GARCH model in 63.6% of the simulated samples. In the other 17.3% of simulated samples, neither is favored against the other.

For the Gaussian kernel GARCH model, the use of localized bandwidths tends to increase the marginal likelihood and improve the competitiveness of the model. According to the computed Bayes factors, the use of localized bandwidths is favored against the use of a global bandwidth with at least positive evidence in 59.1% of the simulated samples, while the latter is favored against the former in only 7.6% of the simulated samples. In the other 33.3% of the simulated samples, neither is favored against the other.

With the use of localized bandwidths, the Gaussian kernel GARCH model is favored against the skewed t GARCH model with very strong evidence in 2.5% of the simulated samples, strong evidence in 5.1% of the simulated samples, and positive evidence in 14.1% of the simulated samples. Thus, the Gaussian kernel GARCH model is favored against its competitor with at least positive evidence in 21.7% of the simulated samples. Moreover, with localized bandwidths, the relative frequency of simulated samples, where the skewed t GARCH is favored against the Gaussian kernel GARCH model with at least positive evidence, is reduced from 63.6% to 47.0%.

To conclude, the Gaussian kernel GARCH model is favored against the correct model, from which the samples are generated, in 21.7% of the simulated samples, and neither of the two models is favored against the other in another 31.3% of the simulated samples. This demonstrates the competitiveness of the proposed Gaussian kernel GARCH model. Although the model cannot compete against the correct model in 47.0% of simulated samples, the robustness of our model compensates for the loss because in practice, the true error density is never known.

3 MODELLING DAILY RETURNS OF THE S&P 500 INDEX

3.1 Gaussian Kernel Error Densities

In this section, we use the proposed sampling algorithm to estimate parameters of the GARCH(1,1) model of daily continuously compounded returns of the S&P 500 closing index, where the errors are assumed to be iid and follow the Gaussian kernel error density. The sample period is from 03/01/2007 to 31/05/2013 with 1,613 observations. The starting value of the return series, which is known as y_0, is the first observation in the sample. Thus, the actual sample size is $n = 1,612$.

We discarded 3000 iterations in the burn-in period and recorded 10,000 iterations therafter. The acceptance rate was controlled to be around 0.234 (see for example, Garthwaite, Fan, and Sisson, 2010). We calculated the batch-mean standard deviation and simulation inefficiency factor (SIF) for each parameter in each model. The batch-mean standard deviation is an approximation to the standard deviation of the posterior average of the simulated chain. If the mixing performance is reasonably good, the batch-mean standard deviation will decrease at a reasonable speed as the number of iterations increases (see Roberts, 1996, among others). The SIF is approximately interpreted as the number of draws needed to derive independent draws, because the simulated chain is a Markov chain (Kim, Shepherd, and Chib, 1998, among others). For example, a SIF value of 20 means that approximately, we should retain one draw for every 20 draws to obtain independent draws in this sampling procedure. According to our experience, a sampler usually achieves reasonable mixing performances when its SIF values are below 100.

We applied our sampling algorithms to the Gaussian kernel asymmetric GARCH(1,1) models of S&P 500 daily returns with respectively, a global bandwidth and localized bandwidths. Table 2 presents the estimate of each parameter, together with its 95% Bayesian credible interval, batch-mean standard deviation and SIF for each type of bandwidth. The batch-mean standard deviation and SIF were used to monitor the mixing performance, and both indicate very good mixing performance of each simulated chain.

In terms of the question of which type of bandwidth should be used, we found that the Bayes factor of the Gaussian kernel asymmetric model with localized bandwidths against the same model with a global bandwidth is 6.6×10^5, which shows very strong evidence supporting the former model against the latter.

3.2 Skewed Student t Error Densities

With the same sample of S&P 500 daily returns, we applied the sampling procedure discussed in Section 2.2 to the asymmetric GARCH(1,1) model with skewed t error densities, where the two parameters, η and λ, in the error density were considered respectively, constants and time-varying. Table 3 presents a summary of the derived results. The batch-mean standard deviation and SIF indicate that the sampling algorithm has achieved very good mixing performance no matter whether the skewed t parameters are constants or time-varying.

The log marginal likelihood of the skewed t GARCH with η and λ being constants is almost the same as that of the same model with two parameters being time-varying. One may also notice that the averaged log likelihood value is increased to -2472.17 by 12.51 as a consequence of allowing the two parameters to be time-varying. However, the improvement is at the cost of increasing the number of parameters.

3.3 Gaussian Kernel Density or Skewed Student t Density?

The Bayes factor of the Gaussian kernel asymmetric GARCH(1,1) model with a global bandwidth against the skewed t GARCH(1,1) model with constant parameters is 6.65×10^6, indicating that the former model is favored against the latter with very strong evidence. The use of localized bandwidths improves the competitiveness of the Gaussian kernel GARCH model, and its Bayes factor against the skewed t GARCH model is 1.6×10^{12}. This is very strong evidence favoring the Gaussian kernel GARCH model against its competitors. Thus, in terms of modelling S&P 500 daily returns, the Gaussian kernel error density with localized bandwidths is the best choice.

3.4 Posterior Predictive Densities of S&P 500 Daily Returns

It is of interest to estimate the density of the one-day-ahead return, y_{n+1}. Under the Gaussian kernel asymmetric GARCH model, the density of y_{n+1} conditional on I_n, is estimated as

$$f_Y\left(y_{n+1}|\boldsymbol{y},\theta\right) = \frac{1}{(n-1)\sigma_{n+1}} \sum_{i=1}^{n} \frac{1}{h}\,\phi\left(\frac{y_{n+1}/\sigma_{n+1} - y_i/\sigma_i}{h}\right), \tag{24}$$

where $\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 I(y_{t-1} \geq 0) + \alpha_2 y_{t-1}^2 I(y_{t-1} < 0) + \beta\sigma_{t-1}^2$ with $\omega = \left(1 - \max\{\alpha_1, \alpha_2\} - \beta\right) s_y^2$. Note that the density function can use either a global bandwidth or localized bandwidths.

The posterior predictive density requires the parameters to be integrated out over the posterior density and can be approximated numerically by averaging the sampling distribution through posterior draws of the parameters. Therefore, we have

$$f_Y\left(y_{n+1}|\boldsymbol{y}\right) = \int f_Y\left(y_{n+1}|\boldsymbol{y},\theta\right)\pi\left(\theta|\boldsymbol{y}\right)d\theta \approx \frac{1}{M}\sum_{i=1}^{M} f_Y\left(y_{n+1}|\boldsymbol{y},\theta^{(i)}\right),$$

where $\theta^{(i)}$, $i = 1,2,\cdots,M$, are draws from $\pi\left(\theta|\boldsymbol{y}\right)$.

At each iteration of the proposed sampling procedure, we calculated the predictive density of y_{n+1} according to (24) at 10,000 grid points, while these values were saved in a row of a matrix with M rows and 10,000 columns. Upon completing the sampling procedure, we calculated the average of these density functions derived at all iterations. The posterior predictive cumulative density function (CDF) was derived in the same way. The posterior predictive Gaussian kernel density and CDF of y_{n+1} are plotted in Figure 1, where in each graph, the blue solid and black dashed lines represent respectively, the posterior predictive density functions and CDFs obtained through a global bandwidth and localized bandwidths. The use of localized bandwidths leads to an obviously higher peak of the posterior predictive Gaussian kernel density than the use of a global bandwidth.

Let θ_{skewt} denote the vector of parameters updated at each iteration of the sampling procedure described in Section 2.2. The density of y_{n+1} conditional on I_n, is derived by plugging-in the updated parameters into the skewed Student t density and is given as

$$\widetilde{f}_Y\left(y_{n+1}|\theta_{\text{skewt}}\right) = f_{\text{skewt}}\left(y_{n+1}/\boldsymbol{y},\sigma_{n+1}|\theta_{\text{skewt}}\right)/\sigma_{n+1},$$

where $f_{\text{skewt}}(\cdot|\theta_{\text{skewt}})$ is the skewed Student t density with its parameter vector given by θ_{skewt}. The posterior predictive density of y_{n+1} is approximated by averaging $\widetilde{f}_Y\left(y_{n+1}|\boldsymbol{y},\theta_{\text{skewt}}\right)$.

The posterior predictive skewed Student t density (with constant η and λ) and its CDF are shown by the red dotted lines of Figure 1. The posterior predictive Gaussian kernel densities of y_{n+1} have obviously thicker left tails and higher peaks than their skewed Student t counterpart.

One direct application of the posterior predictive density and CDF of the one-day-ahead return is to calculate the conditional VaR. At a given confidence level denoted as $100(1-\tau)\%$ with $\tau \in (0,1)$, the VaR of an investment is defined as a threshold value, such that the probability that the maximum expected loss on this investment over a specified time horizon exceeds this value is no more than τ (see for example, Jorion, 1997). The VaR has been a widely used risk measure to control huge losses of a financial position by investment institutions.

The VaR for holding an asset is often estimated through the distribution of the asset return. When this distribution is modeled conditional on time-varying volatilities, the resulting VaR is referred to as the conditional VaR. For example, GARCH models are often used to estimate the conditional VaR. For a given sample $\{y_1, y_2, \cdots, y_n\}$, the conditional VaR with $100(1-\tau)\%$ confidence is defined as

$$y_\tau = \inf\left\{y : P\left(y_{n+1} \le y | y_0, y_1, \cdots, y_n\right) \ge \tau\right\}, \tag{25}$$

where the value of τ is often chosen as either 5% or 1%.

At the 95% confidence level and for every \$100 investment on the S&P 500 index, the one-day conditional VaRs are respectively, \$1.7155 (or \$1.9054) and \$1.4772 under the Gaussian kernel asymmetric GARCH(1,1) models with a global bandwidth (or localized bandwidths) and asymmetric skewed t GARCH(1,1) model. This indicates that the skewed t assumption is likely to underestimate the conditional VaR in comparison with its Gaussian kernel counterpart. However, we should not make such a conclusion based on one sample only.

4 GAUSSIAN KERNEL GARCH MODELS OF OTHER ASSET RETURNS

To demonstrate the usefulness of the proposed Gaussian kernel GARCH model, we considered daily return series of another 9 stock indices, 21 futures prices and 11 currency prices. All stock data were downloaded from yahoo.com, price data of futures front contracts were downloaded from quandl.com, and currency data were downloaded from the website of the Federal Reserve under its H.10 data category. The sample periods are from 03/01/2007 to 31/05/2013 for stock indices, and from 03/01/2007 to 17/05/2013 for futures and currency assets.

When modelling the return series of a stock index or futures price, we used the asymmetric GARCH model with a Gaussian kernel error density. For currency return series, there is no clear evidence of an asymmetric effect of past returns and thus the GARCH model given by (1) was used to model each currency's return series. We implemented the proposed sampling algorithms on the Gaussian kernel GARCH model of each asset's return series, where both a global bandwidth and localized bandwidths are considered.

As a competing model, the skewed Student t GARCH(1,1) model was estimated for each return series using the sampling algorithm discussed in Section 2.2. We calculated the log marginal likelihood at the posterior estimates of parameters, as well as the averaged log likelihood, under the Gaussian kernel GARCH model and its competitor, respectively.

4.1 Model Comparison for Stock Indices

The stock indices in our study are the Dow Jones industrial average (DJIA), Nasadaq-100, NYSE composite, CAC, DAX, FTSE 100, All Ordinaries (AORD), Hang Seng (HS) and Nikkei 225. Table 4 presents the estimates of parameters, their associated SIF values, log marginal likelihood and averaged log likelihood derived under the Gaussian kernel GARCH(1,1) model of each asset's return series, where these results were obtained using respectively, a global bandwidth and localized bandwidths.

The Bayes factors of the Gaussian kernel GARCH model with localized bandwidths against the same model with a global bandwidth are respectively, 2.51×10^7, 3905.0, 149.9 and 137.0 for DJIA, NYSE, DAX and Nikkei. Therefore, the use of localized bandwidths is favored against the use of a global bandwidth with either very strong or strong evidence. The Bayes factor of the former against the latter is 17.8 for Nasdaq, indicating that the use of localized bandwidths is favored against its counterpart with positive evidence. However, for the daily return series of CAC, FTSE, AORD and HS, the Bayes factors indicate that the use of localized bandwidths does not improve the model's competitiveness in comparison with the use of a global bandwidth.

The results obtained via the asymmetric GARCH(1,1) model with skewed Student t error and allowing the two parameters, η and λ, to be constants as well as time-varying are presented in right-hand-side panel of Table 4. The skewed t model with time-varying η and λ is favored against the same model with constant η and λ for Nasdaq, DAX and HS with positive evidence according to the corresponding Bayes factors. Thus, it is sometimes possible to improve the competitiveness of the skewed t model by allowing the two parameters to be time-varying. However, the skewed t model with constant η and λ is favored against its time-varying counterpart for NYSE and AORD with either strong or very strong evidence, and for DJIA and Nikkei with positive evidence. This indicates that the model's competitiveness is worsened by allowing the two parameters to be time-varying. The inclusion of time-varying η and λ into the skewed t model neither improves nor worsens the competitiveness of the model for CAC and FTSE.

In terms of the comparison between the Gaussian kernel GARCH and the skewed t GARCH models, the former model with a global bandwidth is favored against the latter with very strong evidence for all stock indices, except for the Nikkei. Nonetheless, allowing the bandwidth to be localized, the Gaussian kernel model for the Nikkei is favored against its skewed t counterpart with very strong evidence. This demonstrates the clear competitiveness of the Gaussian kernel GARCH model of stock index returns over its skewed Student t counterpart.

4.2 Model Comparison for Currency Return Series

The symmetric Gaussian kernel GARCH(1,1) model and its skewed t counterpart were applied to currency return series, where we considered 11 currencies, namely the Australian dollar (denoted as AUS), Canadian dollar (denoted as CAN), Danish Krone, Euro, Japanese Yen, Norwegian Krone, New Zealand dollar (denoted as NZ), Singapore dollar, Swedish Krona, Swiss Franc and UK Pound. Table 5 presents the averaged log likelihood and log marginal likelihood values obtained through each model.

Under the Gaussian kernel symmetric GARCH(1,1) model, the marginal likelihood is increased through the use of localized bandwidths rather than a global bandwidth for each asset. The use of localized bandwidths is favored against the use of a global bandwidth with either very strong or strong evidence for the Australian dollar, Japanese Yen, Norwegian Krone, New Zealand dollar, Singapore dollar and Swiss Franc, and with positive evidence for the Euro. Therefore, for these eight currencies, the competitiveness of the Gaussian kernel GARCH model has been clearly improved by using localized bandwidths. However, the improvement of the model's competitiveness is merely marginal for the Canadian dollar, Danish Krone and UK Pound.

Should the skewed t GARCH model use constant or time-varying η and λ? According to Bayes factors, the use of time-varying parameters is favored against the use of constant parameters with either very strong or strong evidence for the Canadian dollar, Danish Krone, Norwegian Krone, Singapore dollar, Swedish Krona, Swiss Franc and UK Pound. The use of constant parameters is favored against its counterpart with either strong or positive evidence for the Australian dollar, Euro and New Zealand dollar. The marginal likelihood under the skewed t model for the Japanese Yen is largely unchanged by allowing the two parameters to be time-varying.

In terms of comparison between the Gaussian kernel GARCH model and skewed t GARCH model, the former model with localized bandwidths is favored against the latter with at least positive evidence for the Australian dollar and Japanese Yen. Meanwhile, the skewed t model is favored with very strong evidence against its Gaussian kernel counterpart for the other currencies except for the Euro and New Zealand dollar, for which neither model is favored against the other.

4.3 Model Comparison for Futures Return Series

We applied the Gaussian kernel asymmetric GARCH(1,1) model and its skewed t counterpart to the return series of 21 futures contracts' closing prices. The futures assets in our study include gold, silver, copper, platinum, palladium in the category of metals; corn, wheat, soybean meal, soybean oil, soybeans and oats in the category of grains; sugar, coffee, cocoa, cotton and orange juice in category of softs; live cattle, lean hogs, feeder cattle and lumber in the category of other agriculturals; and heating oil in the energy category.

The averaged log likelihood and log marginal likelihood derived under the Gaussian kernel GARCH(1,1) model with a global bandwidth as well as localized bandwidths are presented in the left-hand-side panel of Table 6. The use of localized bandwidths for the Gaussian kernel model increases the model's competitiveness. The marginal likelihood values derived through localized bandwidths are larger than those derived through a global bandwidth for 19 out of 21 futures assets. For the other two assets, which are platinum and lumber, the use of localized bandwidths leads to slightly smaller marginal likelihood values than the use of a global bandwidth. The model with localized bandwidths is favored against the same model with a global bandwidth for 18 futures assets with at least positive evidence. This shows that the use of localized bandwidths clearly increases the competitiveness of the Gaussian kernel asymmetric GARCH model. Only for the platinum futures contract, the use of a global bandwidth is favored with positive evidence against the use of localized bandwidths.

The right-hand-side panel of Table 6 presents the averaged log likelihood and log marginal likelihood derived under the skewed t GARCH(1,1) model. The model with constant η and λ is favored against the same model with time-varying η and λ for 20 futures assets with very strong evidence. This finding indicates that the competitiveness of the skewed t GARCH model is worsened by allowing η and λ to be time-varying. The only exception is the return series of lean hogs prices, for which the competitiveness of the skewed t model is neither improved nor worsened by allowing the two parameters to be time-varying.

The Gaussian kernel asymmetric GARCH model with a global bandwidth is favored against its skewed t counterpart for return series of platinum, wheat, soybean oil, oats, sugar, cocoa, and lumber with at least positive evidence. Meanwhile, with a global bandwidth, the Gaussian kernel model cannot compete with its skewed t counterpart for gold, lean hogs and feeder cattle. However, by allowing localized bandwidths, the Gaussian kernel model is favored against its skewed t counterpart with very strong evidence for each of these three futures assets.

The skewed t GARCH model is favored against its Gaussian kernel counterpart with at least positive evidence for the return series of seven futures assets, which are silver, palladium, corn, soybean meal, cotton, live cattle and heating oil.

To summarize, the Gaussian kernel asymmetric GARCH model with either a global bandwidth or localized bandwidths is favored against the skewed Student t GARCH model for ten futures assets, while the latter model is favored against the former for seven futures assets. Neither of these two models is favored against for the other four futures assets, which are copper, soybeans, coffee and orange juice.

For all 41 assets considered in this section, we summarized parameter estimates and their associated SIF values derived under each model and reported them in tables, which are available at http://users.monash.edu/~xzhang/section5.tables.pdf.

5 CONCLUSION

We have presented a Bayesian sampling approach to parameter estimation for a GARCH model with an unknown error density, which we propose to approximate by a mixture of n Gaussian component densities centered at individual errors and scaled by a standard deviation parameter. This mixture density has the form of a kernel density estimator of the errors with Gaussian kernels and bandwidth being the standard deviation. Assuming an inverse Gamma prior of the bandwidth parameter and noninformative priors of model parameters, we have derived an approximate posterior of both types of parameters. The random-walk Metropolis algorithm has been used to sample these parameters simultaneously during MCMC iterations. To address the concern about the performance of a global bandwidth in the Gaussian kernel error density, we considered the use of localized bandwidths and derived the posterior of all parameters. Most importantly, the Gaussian kernel error density allows us to estimate the density of the one-step out-of-sample return, which can be used to compute value-at-risk.

A simulation study was conducted, where samples were simulated from the GARCH model with a skewed Student t error density. In 53% of the simulated samples, the Gaussian kernel error density performs better, or at least no worse, than the correct assumption of the error density. This demonstrates the robustness of the Gaussian kernel error density, which compensates for the loss of accuracy, because the true error density is always unknown in practice.

Applying the Gaussian kernel asymmetric GARCH model and its competitor, the asymmetric skewed Student t GARCH model, to a range of daily return series of stock indices, daily returns of closing prices of futures front contracts and daily returns of currency prices, we find that the Gaussian kernel GARCH model is favored against its competitor for all ten stock indices considered, for two out of 11 currencies, and for ten out of 21 futures assets. Neither model is favored against the other for two currencies and four futures assets. This provides a clear empirical justification for the symmetric and asymmetric GARCH models with Gaussian kernel error densities proposed in this paper.

Our investigation is only focused on the GARCH(1,1) specification. The proposed Gaussian kernel error density can be used to replace any parametric assumption of the error density in any parametric GARCH models as long as parameter constraints can be imposed. Moreover, the proposed Bayesian sampling algorithm can be modified accordingly with no increased difficulty.

Acknowledgements

We extend our sincere thanks to John Geweke for enlightening discussion on an early draft of this paper. We greatly appreciate comments from Cathy Chen, John Maheu, Christian Robert, Scott Sisson and Aman Ullah. Thanks also go to Jiti Gao, Hsein Kew and Farshid Vahid-Araghi for *ad hoc* discussion. This research was supported under the Australian Research Council's *Discovery Projects* funding scheme (project numbers DP1095838 and DP130104229).

References

Ausín, M. C., P. Galeano, and P. Ghosh. 2014. A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *European Journal of Operational Research* 232:350–358.

Badrinath, S. G., and S. Chatterjee. 1988. On measuring skewness and elongation in common stock return distributions: The case of the market index. *Journal of Business* 61:451–472.

Bauwens, L., and M. Lubrano. 1998. Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal* 1:23–46.

Bickel, P. J. 1982. On adaptive estimation. *The Annals of Statistics* 10:647–671.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.

Bollerslev, T., and J. M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11:143–172.

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

Bowman, A. W., P. Hall, and D. M. Titterington. 1984. Cross-validation in nonparametric estimation of probabilities and probability densities. *Biometrika* 71:341–351.

Brewer, M. J. 2000. A Bayesian model for local smoothing in kernel density estimation. *Statistics and Computing* 10:299–309.

Chib, S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90:1313–1321.

Chib, S., and I. Jeliazkov. 2001. Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association* 96:270–281.

Di, J., and A. Gangopadhyay. 2011. On the efficiency of a semi-parametric GARCH model. *Econometrics Journal* 14:257–277.

Drost, F. C., and C. A. J. Klaassen. 1997. Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81:193–221.

Durham, G., and J. Geweke. 2014. Improving asset price prediction when all models are false. *Journal of Financial Econometrics* 12:278–306.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1007.

Engle, R. F., and G. González-Rivera. 1991. Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9:345–359.

Garthwaite, P. H., Y. Fan, and S. A. Sisson. 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. 1999. Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews* 18:1–73.

Geweke, J. 2009. *Complete and Incomplete Econometric Models*. Princeton University Press, New Jersey.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48:1779–1801.

Grillenzoni, C. 2009. Kernel likelihood inference for time series. *Scandinavian Journal of Statistics* 36:127–140.

Hall, P., and Q. Yao. 2003. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71:285–317.

Hansen, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review* 6:705–730.

Hjort, N. L., C. Holmes, P. Müller, and S. G. Walker. 2010. *Bayesian Nonparametrics.* Cambridge University Press, Cambridge, U.K.

Jeffreys, H. 1961. *Theory of Probability.* Oxford University Press, Oxford, U.K.

Jensen, M. J., and J. M. Maheu. 2013. Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics* 176:3–17.

Jorion, P. 1997. *Value at Risk: The New Benchmark for Controlling Market Risk.* McGraw-Hill, New York.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–795.

Kim, S., N. Shepherd, and S. Chib. 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65:361–393.

Koop, G. 1994. Bayesian semi-nonparametric ARCH models. *The Review of Economics and Statistics* 76:176–181.

Linton, O. 1993. Adaptive estimation in ARCH models. *Econometric Theory* 9:539–569.

Mardia, K. V., J. T. Kent, G. Hughes, and C. C. Taylor. 2009. Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika* 96:975–982.

Nakatsuma, T. 2000. Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. *Journal of Econometrics* 95:57–69.

Newton, M. A., and A. E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society,* Series B 56:3–48.

Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33:1065–1076.

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

Silverman, B. W. 1978. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics* 6:177–184.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Singleton, J. C., and J. Wingender. 1986. Skewness persistence in common stock returns. *Journal of Financial and Quantitative Analysis* 21:335–341.

Spiegelhalter, D. J., and A. F. M. Smith. 1982. Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society,* Series B 44:377–387.

Tapia, R. A., and J. R. Thompson. 1978. *Nonparametric Probability Density Estimation*. Johns Hopkins University Press, Baltimore.

Varin, C., N. Reid, and D. Firth. 2011. An overview of composite likelihood methods. *Statistica Sinica* 21:5–42.

Vrontos, I. D., P. Dellaportas, and D. N. Politis. 2000. Full Bayesian inference for GARCH and EGARCH models. *Journal of Business and Economic Statistics* 18:187–198.

Weiss, A. A. 1986. Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2:107–131.

Yuan, A. 2009. Semiparametric inference with kernel likelihood. *Journal of Nonparametric Statistics* 21:207–228.

Yuan, A., and J. G. de Gooijer. 2007. Semiparametric regression with kernel error model. *Scandinavian Journal of Statistics* 34:841–869.

Zhang, X., and M. L. King. 2008. Box-Cox stochastic volatility models with heavy-tails and correlated errors. *Journal of Empirical Finance* 15:549–566.

Zhang, X., M. L. King, and R. J. Hyndman. 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics & Data Analysis* 50:3009–3031.

Zhang, X., M. L. King, and H. L. Shang. 2014. A sampling algorithm for bandwidth estimation in a nonparametric regression model with a flexible error density. *Computational Statistics and Data Analysis* 78:218–234.

Table 1: Relative frequencies of simulated samples falling in different categories of Bayes factors based on results from Gaussian kernel GARCH(1,1) models of S&P 500 daily returns.

Category of Bayes factors	Global bandwidth	localized bandwidths	Localized bandwidths against global bandwidth
$(150, \infty)$	2.7%	2.5%	19.3%
$(20, 150]$	6.3%	5.1%	14.4%
$(3, 20]$	10.1%	14.1%	25.4%
$(1/20, 1/3]$	18.6%	25.9%	7.4%
$(1/150, 1/20]$	12.9%	12.0%	0.2%
$(0, 1/150]$	32.1%	9.1%	0.0%

Table 2: Results from Bayesian estimation of the Gaussian kernel GARCH(1,1) model of S&P 500 daily returns. LL represents averaged log likelihood, and LML represents log marginal likelihood.

Parameters	Mean	95% Bayesian credible interval	Batch-mean standard dev.	Standard deviation	SIF
Global bandwidth					
σ_0^2	0.630914	(0.0830, 2.5468)	0.033031	0.653635	25.5
α_1	0.003089	(0.0001, 0.0113)	0.000237	0.003122	57.7
α_2	0.084255	(0.0534, 0.1213)	0.000933	0.017995	26.9
β	0.905672	(0.8641, 0.9412)	0.001049	0.020343	26.6
h	0.370311	(0.2462, 0.4725)	0.002788	0.057566	23.5
LL	−2475.79				
LML	−2482.68				
Localized bandwidths					
σ_0^2	0.311810	(0.0615, 0.9095)	0.004954	0.227820	4.7
α_1	0.003333	(0.0001, 0.0120)	0.000213	0.003190	44.8
α_2	0.100950	(0.0735, 0.1301)	0.000489	0.014508	11.4
β	0.890691	(0.8583, 0.9206)	0.000539	0.015834	11.6
h	0.230680	(0.1515, 0.3370)	0.002392	0.047543	25.3
h_ϵ	0.639238	(0.2866, 0.9649)	0.009720	0.180288	29.1
LL	−2460.78				
LML	−2469.28				

Table 3: Results from Bayesian estimation of the skewed Student t GARCH(1,1) model of S&P 500 daily returns. LL represents averaged log likelihood. LML represents log marginal likelihood.

Parameters	Mean	95% Bayesian credible interval	Batch-mean standard dev.	Standard deviation	SIF
Constant η and λ					
σ_0^2	0.363243	(0.0818, 0.9849)	0.002295	0.243678	4.4
ω	0.038075	(0.0240, 0.0560)	0.000212	0.008400	31.9
α_1	0.036799	(0.0138, 0.0655)	0.000345	0.013055	35.0
α_2	0.110905	(0.0833, 0.1438)	0.000352	0.015435	26.1
β	0.887206	(0.8541, 0.9152)	0.000355	0.015577	25.9
η	8.096940	(5.9481, 11.056)	0.015074	1.313166	6.6
λ	−0.149503	(−0.2047, −0.0934)	0.000275	0.028547	4.6
LL	−2484.68				
LML	−2498.59				
Time-varying η and λ					
σ_0^2	0.412586	(0.0902, 1.1578)	0.002491	0.285808	3.8
ω	0.042601	(0.0264, 0.0641)	0.000239	0.009451	31.9
α_1	0.064249	(0.0321, 0.1073)	0.000361	0.018871	18.3
α_2	0.127057	(0.0948, 0.1666)	0.000457	0.018354	31.0
β	0.870868	(0.8310, 0.9036)	0.000464	0.018562	31.3
η_a	−0.627567	(−1.5725, 1.1508)	0.023096	0.697102	54.9
η_b	−1.747564	(−4.0902, −0.5877)	0.037173	0.960488	74.9
η_c	−0.137599	(−0.7844, 0.8110)	0.015311	0.424880	64.9
λ_a	−0.320770	(−0.4639, −0.1806)	0.000941	0.071274	8.7
λ_b	0.042849	(−0.0800, 0.1621)	0.000691	0.061708	6.3
λ_c	0.002644	(−0.0758, 0.0748)	0.000621	0.038005	13.3
LL	−2472.17				
LML	−2498.53				

Table 4: Averaged log likelihood (LL) and log marginal likelihood (LML) derived from the Gaussian kernel asymmetric GARCH(1,1) model and skewed Student t GARCH(1,1) model of daily stock index returns. The largest LML in each row is in blue color.

Stock index	Gaussian kernel asymmetric GARCH				Skewed t GARCH			
	Global bandwidth		Localized bandwidths		Constant η and λ		Time-varying η and λ	
	LL	LML	LL	LML	LL	LML	LL	LML
DJIA	−2328.78	−2334.70	−2308.48	−2317.66	−2334.34	−2348.43	−2321.77	−2350.20
Nasdaq	−2652.85	−2660.02	−2650.06	−2657.14	−2669.39	−2681.99	−2655.93	−2680.66
NYSE	−2588.95	−2596.59	−2580.53	−2588.32	−2596.80	−2610.04	−2588.23	−2613.08
CAC	−2867.35	−2875.54	−2868.09	−2875.68	−2886.71	−2899.04	−2875.05	−2899.35
DAX	−2745.83	−2754.23	−2741.42	−2749.22	−2757.27	−2769.90	−2743.47	−2767.81
FTSE	−2509.84	−2517.56	−2509.24	−2516.86	−2521.07	−2534.33	−2508.95	−2534.40
AORD	−2367.81	−2376.49	−2371.07	−2379.79	−2383.33	−2397.52	−2381.78	−2403.76
HS	−2875.60	−2881.85	−2875.05	−2881.63	−2882.55	−2895.55	−2869.19	−2894.25
Nikkei	−2781.04	−2787.74	−2776.99	−2782.82	−2776.02	−2788.20	−2766.25	−2790.47

Table 5: Averaged log likelihood (LL) and log marginal likelihood (LML) derived from the Gaussian kernel symmetric GARCH(1,1) model and skewed Student t GARCH(1,1) model of daily currency price returns. The largest LML in each row is in blue color.

Currency	Gaussian kernel symmetric GARCH				Skewed t GARCH			
	Global bandwidth		Localized bandwidths		Constant η and λ		Time-varying η and λ	
	LL	LML	LL	LML	LL	LML	LL	LML
AUS	−1479.61	−1484.41	−1475.00	−1480.22	−1474.57	−1485.65	−1464.11	−1486.79
CAN	−1619.19	−1625.29	−1618.35	−1624.62	−1615.61	−1628.36	−1578.73	−1602.56
Denmark	−1152.41	−1160.13	−1151.96	−1159.29	−1147.57	−1161.25	−1118.26	−1146.18
Euro	−811.58	−816.63	−808.39	−813.19	−802.12	−813.64	−798.14	−822.92
Yen	−2723.63	−2729.61	−2718.83	−2725.18	−2716.63	−2727.69	−2700.31	−2728.07
Norway	−1583.32	−1589.50	−1564.00	−1569.83	−1550.78	−1561.99	−1518.92	−1543.99
NZ	−1074.26	−1083.50	−1047.41	−1053.86	−1042.51	−1054.19	−1034.05	−1057.81
Singapore	−1674.91	−1687.74	−1668.44	−1677.86	−1662.48	−1675.54	−1600.87	−1630.34
Swedish	−1634.76	−1640.37	−1631.49	−1637.30	−1629.94	−1642.48	−1591.30	−1614.85
Swiss	−1738.82	−1747.35	−1717.99	−1725.04	−1702.86	−1714.83	−1662.14	−1686.50
UK	−1930.36	−1937.21	−1929.44	−1936.37	−1932.89	−1945.39	−1906.74	−1930.27

Table 6: Averaged log likelihood (LL) and log marginal likelihood (LML) derived from the Gaussian kernel asymmetric GARCH(1,1) model and skewed Student t GARCH(1,1) model of futures asset daily returns. The largest LML in each row is in blue color.

Futures	Gaussian kernel GARCH				Skewed t GARCH			
	Global bandwidth		Localized bandwidths		Constant η and λ		Time-varying η and λ	
	LL	LML	LL	LML	LL	LML	LL	LML
Gold	−2627.64	−2635.92	−2619.86	−2625.66	−2622.51	−2633.82	−2619.80	−2645.50
Silver	−3479.36	−3483.21	−3473.27	−3477.47	−3465.51	−3475.54	−3460.60	−3486.24
Copper	−3354.21	−3361.96	−3354.71	−3361.17	−3350.76	−3362.23	−3350.77	−3375.94
Platinum	−2870.26	−2877.51	−2871.17	−2878.80	−2875.24	−2886.57	−2873.19	−2895.73
Palladium	−3423.18	−3429.40	−3419.37	−3424.92	−3412.64	−3423.65	−3410.09	−3433.89
Corn	−3451.48	−3459.11	−3450.04	−3455.18	−3444.06	−3453.01	−3448.25	−3468.56
Wheat	−2904.35	−2910.35	−2901.28	−2907.39	−2904.51	−2916.59	−2903.41	−2925.23
Soybean meal	−3273.50	−3282.20	−3258.84	−3265.00	−3249.10	−3259.85	−3248.80	−3274.35
Soybean oil	−2904.34	−2910.34	−2901.24	−2907.27	−2904.48	−2916.47	−2903.46	−2925.25
Soybeans	−3040.21	−3051.38	−3037.78	−3044.92	−3032.14	−3044.22	−3024.35	−3049.39
Oats	−3462.33	−3467.54	−3457.84	−3462.43	−3459.72	−3469.06	−3458.52	−3481.58
Sugar	−3550.04	−3556.08	−3546.61	−3552.27	−3545.58	−3557.98	−3539.95	−3567.25
Coffee	−3264.06	−3267.03	−3253.58	−3258.19	−3248.58	−3258.25	−3246.06	−3269.87
Cocoa	−3325.29	−3331.93	−3325.00	−3330.89	−3324.11	−3335.97	−3324.92	−3350.84
Cotton	−3377.39	−3384.01	−3357.52	−3363.74	−3352.16	−3362.56	−3363.06	−3388.34
Orange juice	−3523.98	−3530.28	−3511.72	−3516.78	−3506.98	−3516.75	−3502.45	−3526.07
Live cattle	−2157.82	−2165.84	−2132.10	−2138.03	−2120.16	−2132.07	−2119.33	−2146.24
Lean hogs	−2894.64	−2909.44	−2875.60	−2887.15	−2881.94	−2892.91	−2871.49	−2893.60
Feeder cattle	−1848.35	−1859.02	−1845.40	−1851.48	−1846.28	−1859.79	−1842.46	−1869.98
Lumber	−3491.89	−3501.93	−3493.16	−3502.05	−3512.64	−3525.25	−3504.90	−3531.37
Heating oil	−3262.10	−3269.74	−3252.68	−3259.51	−3242.36	−3255.45	−3239.38	−3263.29

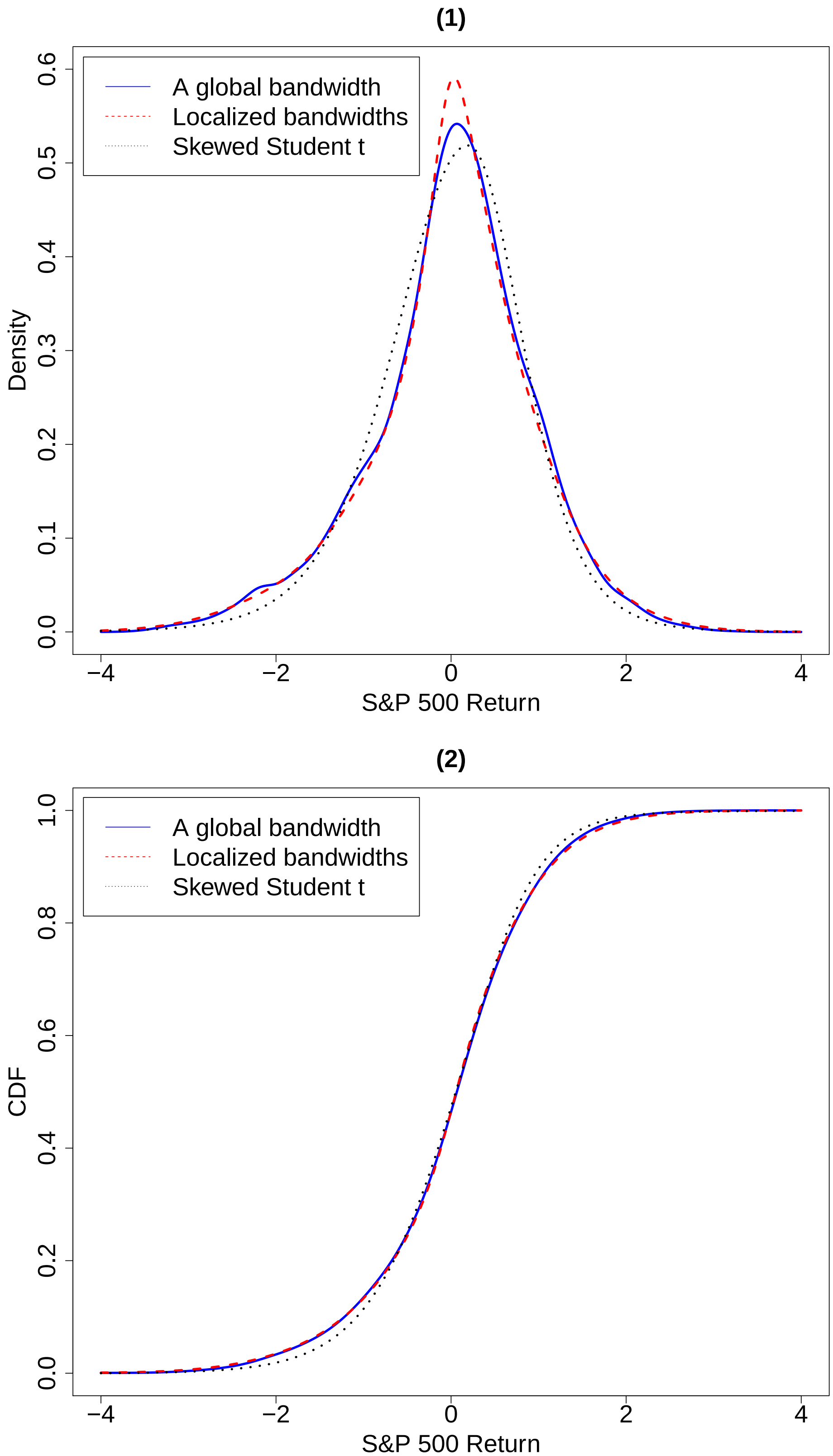

Figure 1: The posterior predictive densities and CDFs of the one-step out-of-sample return under the Gaussian kernel GARCH(1,1) and skewed Student t GARCH(1,1) models for the S&P 500 index: (1) density of y_{n+1}; and (2) CDF of y_{n+1}.