Gaussian kernel GARCH models

Xibin (Bill) Zhang and Maxwell L. King

Department of Econometrics and Business Statistics
Faculty of Business and Economics

MONASH University

7 June 2013
Motivation

• A regression model is often characterised by a regression function and an error term, while the error density is often specified from a parametric family of distributions.

• The true error density is always unknown. Any specification of the error density can only be regarded as an approximation.

• Take the GARCH model as an example. Let $\mathbf{y} = (y_1, \cdots, y_n)'$ denote a vector of $n$ observations of an asset’s return. A GARCH(1,1) model is often formulated as

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

(1)

where $\varepsilon_t$, for $t = 1, 2, \cdots, n$, are independent.

• The assumption of conditional normality of $\varepsilon_t$ has contributed to early successes of GARCH models.

• However, enough evidence has shown that it is possible to reject the assumption of conditional normality.
• 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 distributions.

• Is it possible to estimate parameters without specifying an analytical form for the error density?

• A motivating idea behind this question is due to the importance for being able to estimate the response density.

• In the current literature, there exist two main directions for investigating this question in a GARCH model.

• One direction is the adaptive estimation investigated by Linton (1993, ET) and Drost and Klaassen (1997, JoE).

• The main focus of adaptive estimation is efficiency. They showed that under certain conditions, the estimates of parameters in a GARCH model are approximately efficient.
Engle and González-Rivera (1991, JBES) proposed a semiparametric GARCH model without assumptions on the analytical form of error density.

The error density was estimated nonparametrically based on residuals, which were obtained by applying either the QMLE (assuming conditional normality) or OLS.

The parameters of the semiparametric GARCH model were estimated by maximising the log–likelihood constructed through the estimated density of residuals.

Their Monte Carlo study showed that this semiparametric approach could improve the efficiency of parameter estimates up to 50% against QMLEs obtained under conditional normality.
Limitations and our aims

- Their likelihood is affected by initial parameter estimates.
- Their semiparametric estimation uses the data twice because residuals have to be pre-fitted to construct likelihood.
- Their derived semiparametric estimates of parameters would not be used again to improve the error density estimator.
- We aim to approximate the true error density by a mixture of $n$ normal densities, which has the form of kernel density estimator of the errors.
- We treat the bandwidth as an additional parameter and investigate the likelihood and posterior under this mixture density of errors.
A location–mixture of $n$ Gaussian densities

- We assume that the unknown density of $\varepsilon_t$ denoted as $f(\varepsilon_t)$, is approximated by a location–mixture density:

$$f(\varepsilon_t; h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \phi \left( \frac{\varepsilon_t - \varepsilon_i}{h} \right),$$

for $t = 1, 2, \ldots, n$, where $\phi(\cdot)$ is the Gaussian PDF, and the component Gaussian densities have a common variance $h^2$ and different means at $\varepsilon_i$, for $i = 1, 2, \ldots, n$.

- We call it the Gaussian kernel density of errors, where $h$ is called either the standard deviation or bandwidth. The resulting GARCH model is called the Gaussian kernel GARCH.

- From the view of kernel smoothing, this mixture error density is a kernel density estimator of errors with $h$ the bandwidth.
Our contribution

- We propose to approximate the unknown error density by a Gaussian kernel density of errors, based on which we are able to construct the likelihood.

- We choose prior densities for the two types of parameters and obtain the posterior of all parameters. MCMC is used to estimate these parameters.

- We use this Gaussian kernel GARCH model to compute value-at-risk (VaR).
Investigations related our work

- The validity of this mixture density as a density of the regression errors was investigated by Yuan and de Gooijer (2007, Scandinavian J of Statistics) in a class of nonlinear regression models.

- Our work is related to adaptive estimation discussed by Linton (1993) and Drost and Klaassen (1997) for (G)ARCH models. A conclusion from their work is the parameters are approximately adaptively estimable.

- In all these studies, bandwidth was chosen based on pre-fitted residuals, which were used as proxies of errors.

- Bayesian semiparametric estimation was discussed by Koop (1992, Review of Economics & Statistics) for ARCH models, where the quasi likelihood was set up through a sequence of complicated polynomials.
Gaussian kernel density of errors

- Consider the GARCH(1,1) model

\[ y_t = \sigma_t \varepsilon_t, \]
\[ \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3} \]

where \( \omega > 0, \alpha \geq 0, \beta \geq 0 \) and \( \alpha + \beta < 1 \).

- Strictly speaking, conditional on information available at \( t - 1 \) denoted as \( I_{t-1} \), the density of \( \varepsilon_t \) denoted as \( f(\varepsilon_t) \), is unknown.

- When \( f(\varepsilon) \) is assumed to be known, the likelihood is

\[ \ell_0(y|\omega, \alpha, \beta) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f \left( \frac{y_t}{\sigma_t} \right). \]

- \( f(\varepsilon) \) could be the Gaussian, Student \( t \), and a mixture of several Gaussian densities.
A location–mixture density of $n$ Gaussian densities

- We propose a location–mixture density as the error density:

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

where the component Gaussian densities have a common variance $h^2$ and individual means at individual errors.

- In terms of kernel density estimation based on directly observed data, Silverman (1978, Annals of Stats) proved that a kernel density estimator approaches the underlying true density as $n \to \infty$.

- It is reasonable to expect $f(\varepsilon; h)$ approximates $f(\varepsilon)$ well for large $n$.

- $f(\varepsilon_t; h)$ differs from the kernel density estimator of residuals calculated through pre–estimated parameters. Our mixture density is defined conditional on model parameters:

$$f(\varepsilon_t; h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \phi \left( \frac{\varepsilon_t - y_i/\sigma_i}{h} \right), \quad (4)$$

where $\sigma_i^2 = \omega + \alpha y_{i-1}^2 + \beta \sigma_{i-1}^2$, for $i = 1, 2, \ldots, n.$
Benefit of the mixture error density (2)

- To construct likelihood, we use the leave–one–out:

\[
f(\varepsilon_t | \varepsilon(t); h) = \frac{1}{n - 1} \sum_{i=1 \atop i \neq t}^{n} \frac{1}{h} \phi \left( \frac{\varepsilon_t - \varepsilon_i}{h} \right). \tag{5}\]

- The density of \( y_t \) is approximated by

\[
f_Y (y_t | y(t); \theta) = \frac{1}{(n - 1)\sigma_t} \sum_{i=1 \atop i \neq t}^{n} \frac{1}{h} \phi \left( \frac{y_t/\sigma_t - y_i/\sigma_i}{h} \right). \tag{6}\]

- In the density function of \( y_t \) given by (6), \( h \) and \( \sigma_t \) always appear in the form of the product of the two:

\[
h^2 \sigma_t^2 = h^2 \omega + h^2 \alpha y_{t-1}^2 + \beta h^2 \sigma_{t-1}^2. \tag{7}\]
Diamond $h^2$ and $\omega$, as well as $h^2$ and $\alpha$, cannot be separately identified. If $\omega$ (or $\alpha$) is a constant, the other parameters can be identified.

Diamond In adaptive estimation for ARCH models, $\omega$ was restricted to be zero by Linton (1993) and one by Drost and Klaassen (1997).

Diamond As the unconditional variance of $y_t$ is $\omega/(1 - \alpha - \beta)$, we assume that $\omega = (1 - \alpha - \beta) s_y^2$. If $\{y_t\}$ is standardised, we have $\omega = (1 - \alpha - \beta)$, which was used by Engle and González-Rivera (1991).

Diamond The parameter vector is $\theta = (\sigma_0^2, \alpha, \beta, h^2)'$. The restrictions are $0 \leq \alpha < 1, 0 \leq \beta < 1$ and $0 < \alpha + \beta < 1$.

Diamond The likelihood of $y = (y_1, y_2, \ldots, y_n)'$, for given $\theta$, is

\[
\ell(y|\theta) = \prod_{t=1}^{n} \left\{ \frac{1}{(n-1)\sigma_t} \sum_{i=1}^{n} \frac{1}{h} \phi \left( \frac{y_t/\sigma_t - y_i/\sigma_i}{h} \right) \right\}.
\] (8)

Diamond Conditional on model parameters, this likelihood is the one used by the likelihood cross-validation in choosing bandwidth for the kernel density estimator of standardised $y_i$, for $i = 1, 2, \ldots, n$. 
Prior choices and posterior

- The prior of $\sigma_0^2$ is the log normal with mean 0 and variance 1. The prior of $\beta$ is the uniform density on $(0, 1)$. The prior of $\alpha$ is the uniform density on $(0, 1 - \beta)$. The prior of $h^2$ is an inverse Gamma density given by

$$p(h^2) = \frac{b}{\Gamma(a)} \left( \frac{1}{h^2} \right)^{a+1} \exp \left\{ -\frac{b}{h^2} \right\},$$

where $a$ and $b$ are hyperparameters which are 1 and 0.05, respectively.

- The joint prior of $\theta$ denoted as $p(\theta)$, is the product of these marginal priors.

- The posterior of $\theta$ for given $y$ is proportional to the product of the joint prior of $\theta$ and the likelihood of $y$ for given $\theta$:

$$\pi(\theta|y) \propto p(\theta) \times \ell(y|\theta).$$

- We used the random–walk Metropolis algorithm to sample parameters, where the proposal density is the standard normal.
GJR GARCH models for equity returns and commodity returns

- Negative and positive returns may have different effect on volatilities.

\[ y_t = \sigma_t \varepsilon_t, \]
\[ \sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 I(y_{t-1} \geq 0) + \alpha_2 y_{t-1}^2 I(y_{t-1} < 0) \beta \sigma_{t-1}^2, \tag{9} \]

where \( \omega > 0, 0 \leq \alpha_1 < 1, 0 \leq \alpha_2 < 1, 0 \leq \beta < 1 \) and \( \max(\alpha_1, \alpha_2) + \beta < 1 \).

- For identification concerns, we assume that \( \omega = 1 - \max(\alpha_1, \alpha_2) - \beta \).

- When choosing priors, we assume that the priors of \( \alpha_1 \) and \( \alpha_2 \) are the uniform density on \( (0, 1 - \beta) \).
Data, models and results

- We applied the GARCH(1,1) model with its error density being the Gaussian kernel density to modelling daily returns of the S&P 500 index.

- Sample period: 3 JAN 2007 — 31 MAY 2013; and \( n = 1612 \).

- Estimate parameters in the Gaussian kernel GARCH(1,1).

- A competing model is the skewed \( t \)-GARCH(1,1) model of Hansen (1994, IER). This error density contains two parameters, which are \( \eta > 2 \) and \(-1 < \lambda < 1\). The prior of \( \eta \) is the density of \( N(7, 3^2) \) truncated at 2. The prior of \( \lambda \) is the uniform density on \( U(-1, 1) \).

- Simulation inefficiency factor (SIF) monitors convergence.

- The tuning parameters for the random–walk Metropolis were chosen, such that the resulting acceptance rates are 0.234 for univariate draws and 0.44 for multivariate draws.

- 3000 draws for burn–in, and 10,000 draws recorded.
### Table 1: Results from the Gaussian kernel GARCH(1,1) model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>95% Bayesian credible interval</th>
<th>Batch–mean standard mean</th>
<th>Standard deviation</th>
<th>SIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0^2$</td>
<td>0.7309</td>
<td>(0.0784, 3.0147)</td>
<td>0.051976</td>
<td>0.797843</td>
<td>42.4</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0031</td>
<td>(0.0002, 0.0124)</td>
<td>0.000254</td>
<td>0.003175</td>
<td>63.8</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.0825</td>
<td>(0.0499, 0.1192)</td>
<td>0.001156</td>
<td>0.018408</td>
<td>39.5</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9079</td>
<td>(0.8666, 0.9452)</td>
<td>0.001311</td>
<td>0.020818</td>
<td>39.7</td>
</tr>
<tr>
<td>$h$</td>
<td>0.3646</td>
<td>(0.2295, 0.4785)</td>
<td>0.003596</td>
<td>0.061975</td>
<td>33.7</td>
</tr>
<tr>
<td>LML</td>
<td></td>
<td>−2482.88</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Results from the skewed t GARCH(1,1) model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>95% Bayesian credible interval</th>
<th>Batch–mean standard mean</th>
<th>Standard deviation</th>
<th>SIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0^2$</td>
<td>0.3632</td>
<td>(0.0818, 0.9849)</td>
<td>0.002295</td>
<td>0.243678</td>
<td>4.4</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.0381</td>
<td>(0.0240, 0.0560)</td>
<td>0.000212</td>
<td>0.008400</td>
<td>31.9</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0368</td>
<td>(0.0138, 0.0655)</td>
<td>0.000345</td>
<td>0.013055</td>
<td>35.0</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1109</td>
<td>(0.0833, 0.1438)</td>
<td>0.000352</td>
<td>0.015435</td>
<td>26.1</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.8872</td>
<td>(0.8541, 0.9152)</td>
<td>0.000355</td>
<td>0.015577</td>
<td>25.9</td>
</tr>
<tr>
<td>$\eta$</td>
<td>8.0969</td>
<td>(5.9481, 11.056)</td>
<td>0.015074</td>
<td>1.313166</td>
<td>6.6</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>−0.1495</td>
<td>(−0.2047, −0.0934)</td>
<td>0.000275</td>
<td>0.028547</td>
<td>4.6</td>
</tr>
<tr>
<td>LML</td>
<td></td>
<td>−2498.59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The difference between two LMLs is 15.71.
Convergence performance of our sampler

- The SIF is approximately interpreted as the number of draws needed to derive independent draws.
- For example, a SIF value of 20 means that approximately, we should retain 1 draw for every 20 draws to obtain independent draws in this sampling procedure.
- All simulated chains under the mixture error density have achieved very reasonable convergence.
- The marginal likelihood derived under the mixture error density is obviously larger than that derived under the $t$ errors.
Bayes factor for model comparison

- Bayes factor is a ratio of the marginal likelihoods derived under a model of interest and its competing model.

- Let \( \theta \) denote the parameter vector under model \( \mathcal{A} \). The marginal likelihood under model \( \mathcal{A} \) is (Chib, 1995, JASA)

\[
    m_\mathcal{A}(y) = \frac{\ell_\mathcal{A}(y|\theta)p_\mathcal{A}(\theta)}{\pi_\mathcal{A}(\theta|y)}.
\]

\( \ell_\mathcal{A}(y|\theta) \) and \( p_\mathcal{A}(\theta) \) are likelihood and prior under model \( \mathcal{A} \).

- The Bayes factor of model \( \mathcal{A} \) against model \( \mathcal{B} \) is

\[
    BF = \frac{m_\mathcal{A}(y)}{m_\mathcal{B}(y)}.
\]

- \( 3 \leq BF \leq 20 \): \( \mathcal{A} \) is favored against \( \mathcal{B} \) with positive evidence.
- \( 20 < BF \leq 150 \): \( \mathcal{A} \) is favored against \( \mathcal{B} \) with strong evidence.
- \( BF > 150 \): \( \mathcal{A} \) is favored against \( \mathcal{B} \) with very strong evidence.
Density forecast of the one–day–ahead S&P 500 return

- VaR can be computed through the derived CDF.
- The derived VaR is distribution–free.
- In the left tail area, the CDF of skewed $t$–GARCH is below the CDF of Gaussian kernel GARCH.
Motivation for localised bandwidths

- When the true error density has sufficient long tails, the leave-one-out kernel density estimator with its bandwidth selected under the Kullback–Leibler criterion, is likely to over-estimate the tails density.
- Consequently, the Gaussian kernel density of errors is likely to over-estimate the tail thickness.
- One may argue that this phenomenon is likely to be caused by the use of a global bandwidth. A remedy to this problem in that situation is to use variable bandwidths or localized bandwidths.
- 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.
Localised bandwidths

- We assume the underlying true error density is unimodal. Large absolute errors should be assigned relatively large bandwidths, while small absolute errors should be given relatively small bandwidths.

- We propose the following error density estimator:

\[
f_a(\varepsilon_t; h, h_{\varepsilon}) = \frac{1}{n-1} \sum_{i=1}^{n} \frac{1}{h (1 + h_{\varepsilon} |\varepsilon_i|)} \phi \left( \frac{\varepsilon_t - \varepsilon_i}{h (1 + h_{\varepsilon} |\varepsilon_i|)} \right),
\]

where \( h (1 + h_{\varepsilon} |\varepsilon_i|) \) is the bandwidth assigned to \( \varepsilon_i \). The vector of parameters is now \( \theta_a = (\sigma_0^2, \alpha, \beta, h, h_{\varepsilon})' \).

- The prior of \( h_{\varepsilon} \) is the uniform density on \((0, 1)\).
Bayesian estimate

- The parameter estimates are:
  \[ \sigma_0^2 = 0.3120; \quad \alpha_1 = 0.0033; \quad \alpha_2 = 0.1005; \]
  \[ \beta = 0.8912; \quad h = 0.4709; \quad h_\varepsilon = 0.6577. \]

- The sampler has converged very well.

- The log marginal likelihood under localised bandwidths (and global bandwidth) is -2470.49 (and -2482.88).

- The Bayes factor of the use of localised bandwidths against the use of global bandwidth is \( \exp(12.39) \). Therefore, the former is favored against the latter with very strong evidence.

- The use of localised bandwidths has increased the competitiveness of the Gaussian kernel GARCH model.
Application to other index returns

Table 3(1): Results from the Gaussian kernel GARCH

<table>
<thead>
<tr>
<th></th>
<th>DJIA</th>
<th>Nasdaq</th>
<th>NYSE</th>
<th>CAC</th>
<th>DAX</th>
<th>FTSE</th>
<th>AORD</th>
<th>HK</th>
<th>Nikkei</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0^2$</td>
<td>1.2004</td>
<td>0.9433</td>
<td>1.1224</td>
<td>0.5127</td>
<td>0.3719</td>
<td>0.4178</td>
<td>0.5349</td>
<td>1.3989</td>
<td>0.8966</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0026</td>
<td>0.0049</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.0026</td>
<td>0.0045</td>
<td>0.0191</td>
<td>0.0238</td>
<td>0.0738</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.0897</td>
<td>0.0835</td>
<td>0.0673</td>
<td>0.1099</td>
<td>0.0881</td>
<td>0.0929</td>
<td>0.0963</td>
<td>0.0895</td>
<td>0.0826</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9029</td>
<td>0.9052</td>
<td>0.9252</td>
<td>0.8795</td>
<td>0.9066</td>
<td>0.8993</td>
<td>0.8952</td>
<td>0.9003</td>
<td>0.9015</td>
</tr>
<tr>
<td>$h$</td>
<td>0.4005</td>
<td>0.2641</td>
<td>0.3067</td>
<td>0.2794</td>
<td>0.3699</td>
<td>0.3225</td>
<td>0.1410</td>
<td>0.2406</td>
<td>0.2305</td>
</tr>
<tr>
<td>LML</td>
<td>-2334.7</td>
<td>-2660.0</td>
<td>-2596.6</td>
<td>-2875.5</td>
<td>-2754.2</td>
<td>-2517.6</td>
<td>-2376.5</td>
<td>-2881.9</td>
<td>-2787.7</td>
</tr>
</tbody>
</table>

Table 3(2): Results from the skewed $t$–GARCH(1,1) model

<table>
<thead>
<tr>
<th></th>
<th>DJIA</th>
<th>Nasdaq</th>
<th>NYSE</th>
<th>CAC</th>
<th>DAX</th>
<th>FTSE</th>
<th>AORD</th>
<th>HK</th>
<th>Nikkei</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_0^2$</td>
<td>0.3794</td>
<td>1.0458</td>
<td>0.5920</td>
<td>0.7217</td>
<td>0.6753</td>
<td>0.6355</td>
<td>0.9405</td>
<td>1.6345</td>
<td>1.1553</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.0307</td>
<td>0.0625</td>
<td>0.0425</td>
<td>0.1001</td>
<td>0.0639</td>
<td>0.0581</td>
<td>0.0377</td>
<td>0.0460</td>
<td>0.0950</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0404</td>
<td>0.0326</td>
<td>0.0391</td>
<td>0.0190</td>
<td>0.0246</td>
<td>0.0270</td>
<td>0.0506</td>
<td>0.0562</td>
<td>0.0564</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1127</td>
<td>0.1215</td>
<td>0.1072</td>
<td>0.1292</td>
<td>0.1054</td>
<td>0.1251</td>
<td>0.1191</td>
<td>0.0943</td>
<td>0.1322</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.8856</td>
<td>0.8762</td>
<td>0.8905</td>
<td>0.8685</td>
<td>0.8924</td>
<td>0.8728</td>
<td>0.8784</td>
<td>0.9025</td>
<td>0.8620</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.1360</td>
<td>-0.1233</td>
<td>-0.1764</td>
<td>-0.0762</td>
<td>-0.0919</td>
<td>-0.1098</td>
<td>-0.1607</td>
<td>-0.0831</td>
<td>-0.1253</td>
</tr>
<tr>
<td>LML</td>
<td>-2348.4</td>
<td>-2682.0</td>
<td>-2610.0</td>
<td>-2899.0</td>
<td>-2769.9</td>
<td>-2534.3</td>
<td>-2397.5</td>
<td>-2895.6</td>
<td>-2788.2</td>
</tr>
</tbody>
</table>

The Gaussian kernel GARCH(1,1) model is favored against the skewed $t$ model with very strong evidence for 10 stock indices.
Table 4: Results from Gaussian kernel GARCH (localised bandwidth)

<table>
<thead>
<tr>
<th></th>
<th>DJIA</th>
<th>Nasdaq</th>
<th>NYSE</th>
<th>CAC</th>
<th>DAX</th>
<th>FTSE</th>
<th>AORD</th>
<th>HK</th>
<th>Nikkei</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_0$</td>
<td>0.2643</td>
<td>0.7324</td>
<td>0.4403</td>
<td>0.5339</td>
<td>0.3970</td>
<td>0.4218</td>
<td>0.5419</td>
<td>1.3543</td>
<td>0.8140</td>
</tr>
<tr>
<td></td>
<td>(6.1)</td>
<td>(5.3)</td>
<td>(4.8)</td>
<td>(3.7)</td>
<td>(5.9)</td>
<td>(4.4)</td>
<td>(5.5)</td>
<td>(3.3)</td>
<td>(4.3)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0029</td>
<td>0.0045</td>
<td>0.0039</td>
<td>0.0026</td>
<td>0.0027</td>
<td>0.0035</td>
<td>0.0221</td>
<td>0.0244</td>
<td>0.0532</td>
</tr>
<tr>
<td></td>
<td>(4.3)</td>
<td>(5.2)</td>
<td>(36.1)</td>
<td>(50.9)</td>
<td>(5.0)</td>
<td>(4.5)</td>
<td>(46.9)</td>
<td>(23.6)</td>
<td>(28.7)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.1055</td>
<td>0.0951</td>
<td>0.0848</td>
<td>0.1049</td>
<td>0.0822</td>
<td>0.0994</td>
<td>0.0991</td>
<td>0.0903</td>
<td>0.1007</td>
</tr>
<tr>
<td></td>
<td>(2.8)</td>
<td>(8.2)</td>
<td>(17.1)</td>
<td>(13.1)</td>
<td>(8.0)</td>
<td>(7.2)</td>
<td>(33.4)</td>
<td>(17.6)</td>
<td>(14.7)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.8875</td>
<td>0.8954</td>
<td>0.9073</td>
<td>0.8853</td>
<td>0.9115</td>
<td>0.8915</td>
<td>0.8921</td>
<td>0.9003</td>
<td>0.8840</td>
</tr>
<tr>
<td></td>
<td>(2.9)</td>
<td>(8.2)</td>
<td>(16.2)</td>
<td>(12.8)</td>
<td>(8.7)</td>
<td>(7.7)</td>
<td>(33.9)</td>
<td>(18.2)</td>
<td>(14.5)</td>
</tr>
<tr>
<td>$h$</td>
<td>0.4645</td>
<td>0.3747</td>
<td>0.4298</td>
<td>0.4011</td>
<td>0.4050</td>
<td>0.4391</td>
<td>0.2322</td>
<td>0.4042</td>
<td>0.4008</td>
</tr>
<tr>
<td></td>
<td>(22.1)</td>
<td>(32.8)</td>
<td>(27.3)</td>
<td>(28.3)</td>
<td>(19.3)</td>
<td>(20.6)</td>
<td>(39.3)</td>
<td>(32.0)</td>
<td>(32.5)</td>
</tr>
<tr>
<td>$h_\epsilon$</td>
<td>0.6799</td>
<td>0.6440</td>
<td>0.6472</td>
<td>0.3908</td>
<td>0.5966</td>
<td>0.4841</td>
<td>0.1518</td>
<td>0.4577</td>
<td>0.6760</td>
</tr>
<tr>
<td></td>
<td>(21.3)</td>
<td>(31.6)</td>
<td>(33.6)</td>
<td>(33.2)</td>
<td>(17.9)</td>
<td>(20.7)</td>
<td>(40.4)</td>
<td>(42.7)</td>
<td>(48.9)</td>
</tr>
<tr>
<td>LML</td>
<td>-2319.1</td>
<td>-2657.6</td>
<td>-2589.3</td>
<td>-2876.7</td>
<td>-2749.8</td>
<td>-2517.9</td>
<td>-2376.4</td>
<td>-2882.4</td>
<td>-2784.1</td>
</tr>
</tbody>
</table>

Table 5: A summary of log marginal likelihoods

<table>
<thead>
<tr>
<th></th>
<th>DJIA</th>
<th>Nasdaq</th>
<th>NYSE</th>
<th>CAC</th>
<th>DAX</th>
<th>FTSE</th>
<th>AORD</th>
<th>HK</th>
<th>Nikkei</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewed</td>
<td>-2348.4</td>
<td>-2682.0</td>
<td>-2610.0</td>
<td>-2899.0</td>
<td>-2769.9</td>
<td>-2534.3</td>
<td>-2397.5</td>
<td>-2895.6</td>
<td>-2788.2</td>
</tr>
<tr>
<td>Global</td>
<td>-2334.7</td>
<td>-2660.0</td>
<td>-2596.6</td>
<td>-2875.5</td>
<td>-2754.2</td>
<td>-2517.6</td>
<td>-2376.5</td>
<td>-2881.9</td>
<td>-2787.7</td>
</tr>
<tr>
<td>Local</td>
<td>-2319.1</td>
<td>-2657.6</td>
<td>-2589.3</td>
<td>-2876.7</td>
<td>-2749.8</td>
<td>-2517.9</td>
<td>-2376.4</td>
<td>-2882.4</td>
<td>-2784.1</td>
</tr>
</tbody>
</table>
Empirical findings

- The Gaussian kernel GARCH model is favored against the skewed $t$-GARCH for all 10 indices.
- The use of localised bandwidths increases the competitiveness against its competitor, the skewed $t$-GARCH model. This is evidenced by an increased marginal likelihood for each index.
Application to 24 futures return series

Table 6: A summary of log marginal likelihoods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Silver</th>
<th>Copper</th>
<th>Platinum</th>
<th>Palladium</th>
<th>Corn</th>
<th>Wheat</th>
<th>Soybn meal</th>
<th>Soybn</th>
<th>Soybns oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skew</td>
<td>-3475.5</td>
<td>-3362.2</td>
<td>-2886.6</td>
<td>-3423.7</td>
<td>-3453.0</td>
<td>-2916.6</td>
<td>-3259.9</td>
<td>-2916.5</td>
<td>-3044.2</td>
</tr>
<tr>
<td>Global</td>
<td>-3483.2</td>
<td>-3362.0</td>
<td>-2877.5</td>
<td>-3429.1</td>
<td>-3459.1</td>
<td>-2910.4</td>
<td>-3282.2</td>
<td>-2910.3</td>
<td>-3051.4</td>
</tr>
<tr>
<td>Local</td>
<td>-3478.4</td>
<td>-3362.2</td>
<td>-2879.3</td>
<td>-3426.0</td>
<td>-3456.2</td>
<td>-2907.8</td>
<td>-3265.8</td>
<td>-2907.9</td>
<td>-3045.9</td>
</tr>
</tbody>
</table>

Table 6 (continue): A summary of log marginal likelihoods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sugar</th>
<th>Coffee</th>
<th>Cocoa</th>
<th>Cotton</th>
<th>Orange juice</th>
<th>Live cattle</th>
<th>Lean hogs</th>
<th>Feeder cattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skew</td>
<td>-3558.0</td>
<td>-3335.9</td>
<td>-3336.0</td>
<td>-3362.6</td>
<td>-3516.8</td>
<td>-2132.1</td>
<td>-2892.9</td>
<td>-1859.8</td>
</tr>
<tr>
<td>Global</td>
<td>-3556.1</td>
<td>-3267.0</td>
<td>-3311.9</td>
<td>-3384.0</td>
<td>-3530.3</td>
<td>-2165.8</td>
<td>-2909.4</td>
<td>-1859.0</td>
</tr>
<tr>
<td>Local</td>
<td>-3552.9</td>
<td>-3259.3</td>
<td>-3331.5</td>
<td>-3364.8</td>
<td>-3518.1</td>
<td>-2139.7</td>
<td>-2887.8</td>
<td>-1852.7</td>
</tr>
</tbody>
</table>

Table 6 (continue): A summary of log marginal likelihoods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Oats</th>
<th>Heating</th>
<th>Gold oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skew</td>
<td>-3469.1</td>
<td>-3255.5</td>
<td>-2633.82</td>
</tr>
<tr>
<td>Global</td>
<td>-3467.5</td>
<td>-3269.8</td>
<td>-2635.92</td>
</tr>
<tr>
<td>Local</td>
<td>-3452.3</td>
<td>-3259.9</td>
<td>-2626.26</td>
</tr>
</tbody>
</table>
Conclusion

- We approximate unknown error density by a Gaussian kernel density of errors.
- We derived the likelihood and posterior for all parameters. MCMC is conducted for estimation.
- This error density allows for forecasting the response density.
- The use of localised bandwidths increases the competitiveness of the resulting GARCH model against its competitors.
- The Gaussian kernel GARCH is favored against the skewed $t$–GARCH for equity returns.
- Of all 24 futures return series, the Gaussian kernel GARCH is favored against the skewed $t$–GARCH for 9 series, while the latter is favored against the former for 6 series.
- The Gaussian kernel GARCH is favored against the skewed $t$–GARCH for 2 out of 10 currency return series.

Current work: VaR and backtesting