



Duisenberg school of finance – Tinbergen Institute Discussion Paper

TI 12–055/2/DSF35

A New Semiparametric Volatility Model

Jiangyu Ji
André Lucas

Faculty of Economics and Business Administration, VU University Amsterdam.

** Duisenberg school of finance, Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

A New Semiparametric Volatility Model*

Jiangyu Ji^(a,b), André Lucas^(a,b)

(a) Department of Finance, VU University Amsterdam

(b) Duisenberg school of finance and Tinbergen Institute

May 21, 2012

Abstract

We propose a new semiparametric observation-driven volatility model where the form of the error density directly influences the volatility dynamics. This feature distinguishes our model from standard semiparametric GARCH models. The link between the estimated error density and the volatility dynamics follows from the application of the generalized autoregressive score framework of Creal, Koopman, and Lucas (2012). We provide simulated evidence for the estimation efficiency and forecast accuracy of the new model, particularly if errors are fat-tailed and possibly skewed. In an application to equity return data we find that the model also does well in density forecasting.

Keywords: volatility clustering, Generalized Autoregressive Score model, kernel density estimation, density forecast evaluation.

JEL codes: C10, C14, C22.

1 Introduction

We contribute to the literature on volatility modeling and return density forecasting by introducing a new semiparametric observation-driven time-varying volatility model. We estimate the error density using kernel density estimation techniques and use the resulting estimate to drive the volatility dynamics using the generalized autoregressive score dynamics of Creal, Koopman, and Lucas (2011, 2012). Unlike familiar semiparametric GARCH models, the new model provides a direct link between the form of the error distribution and the volatility dynamics. We show that this enhances the forecasting stability of the model.

Volatility clustering in financial return data is a well-established empirical fact. Starting with the seminal (G)ARCH papers by Engle (1982) and Bollerslev (1986), many variations

*Both authors thank the Dutch National Science Foundation (NWO) for financial support. Corresponding author: Jiangyu Ji, VU University Amsterdam, FEWEB/FIN, de Boelelaan 1105, 1081 HV Amsterdam, Netherlands, phone: +31 20 598 2895, email: j.ji@vu.nl.

and extensions of the GARCH class of models have been proposed in the literature, such as the GARCH-M model of Engle, Lilien, and Robins (1987), the EGARCH model of Nelson (1991), the GJR-GARCH model of Glosten, Jagannathan, and Runkle (1993), and the Threshold GARCH (TGARCH) model by Zakoian (1994). Interestingly, Quasi Maximum Likelihood Estimation (QMLE) of these models based on the (possibly incorrect) assumption of conditional normality for the error terms still yields \sqrt{n} -consistent estimators in many cases, with n denoting the sample size; see Weiss (1986), Lee and Hansen (1994), and Lumsdaine (1989). This is important, as the conditional normality assumption is often refuted for empirical data. The statistical efficiency loss, however, can be considerable. Therefore, many empirical studies assume non-Gaussian conditional distributions, for example a (skewed) Student t , see Baillie and Bollerslev (1989) and Bauwens and Laurent (2005).

To simultaneously avoid the potential efficiency loss and the risk of assuming a wrong parametric family of conditional distributions, semiparametric GARCH models have been proposed. Engle and Gonzalez-Rivera (1991) show that more efficient parameter estimates can be obtained by estimating the error density nonparametrically, though some efficiency loss remains. Drost and Klaassen (1997) and Sun and Stengos (2006) develop kernel-based estimators and establish the semiparametric efficiency bounds for parameter estimation. They characterize the conditions under which one can adaptively estimate a subset of the model's parameters, i.e., one can achieve the same asymptotic efficiency as if the true error density were known.

In all of the above models, the form of the conditional distribution has no impact on the specification of volatility dynamics. Creal, Koopman, and Lucas (2011) argue that if, for example, the error distribution is fat-tailed, we expect to see large (squared) observations from time to time. The occurrence of such observations should not automatically be attributed to an increase in volatility, as would be the case when the standard GARCH dynamics are used for volatility. Similarly, if the data are drawn from a skewed conditional distribution, one expects a large negative observations to convey a different signal about current volatility levels than large positive observations. Again, this implies a link between the shape of the error distribution and the specification of the volatility dynamics. No such direct link is embedded in standard GARCH or semiparametric GARCH models.

Our main contribution in this paper is to provide a semiparametric model for time-varying volatility in which the form of the error distribution is directly linked to the volatility dynamics. We do so by extending the current semiparametric GARCH framework to the Generalized Autoregressive Score (GAS) framework of Creal, Koopman, and Lucas (2011, 2012). At the same time, we extend the class of GAS models to the semiparametric context. A second

important contribution is that we use our semiparametric volatility model to perform out-of-sample forecasting. Most of the literature on semiparametric volatility modeling focuses on the in-sample behavior of the estimators. The out-of-sample forecasting performance of semiparametric methods, however, is equally important in most economic settings.

Within the class of GAS models, the volatility dynamics are driven by the score of the observation density with respect to the (log) volatility parameter. The interpretation is straightforward: at each time point the local likelihood fit of the model is improved using a Newton or gradient step. For the normal distribution, the GAS framework reduces to the standard GARCH model under appropriate parameterization and scaling choices. The GAS framework is, however, much more generally applicable. For example, under fat-tailed error distributions the GAS framework results in a model with interesting robustness properties, see the discussion in Creal, Koopman, and Lucas (2011). It is useful at this stage to note that the semiparametric nature of our model relates to the nonparametric estimation of the error distribution as in Drost and Klaassen (1997) and Sun and Stengos (2006), and not the nonparametric estimation of the volatility dynamics as in Pagan and Schwert (1990), Linton and Mammen (2005) and Yang (2006). The latter would be interesting in its own right, but is not pursued in the present paper.

Parameter estimation for our model is straightforward, since the model is defined in conditional terms similar to standard GARCH models. This implies that the likelihood function can be specified in closed form. Parameter estimation is carried out in two steps. First, we estimate the model by QMLE using a normal or Student's t quasi-likelihood. From this we obtain the standardized residuals and the kernel density estimate of the error density. Second, the kernel density estimate is used to re-estimate the model's parameters, resulting in the Semiparametric Maximum Likelihood Estimator (SMLE).

We investigate the new model's performance in an extensive simulation study using a variety of error distributions and volatility patterns. We find that the SMLM regains a substantial part of the efficiency loss due to the use of the QMLE. If the volatility transition dynamics are possibly mis-specified, we also find that the SMLM produces large improvements in in-sample estimation accuracy and out-of-sample forecast accuracy of the volatilities compared to QMLE. We also make a comparison between our semiparametric model and the semiparametric GARCH model in volatility forecasting.

In our empirical application, we estimate a model for IBM daily equity returns. We use the density forecast evaluation methods proposed by Diebold, Gunther, and Tay (1998) to evaluate the model's performance. The density forecast results show that our model satisfies all relevant

adequacy criteria, whereas the models based on other densities and volatility dynamics have more trouble in attaining statistical adquacy of the density forecasts. We conclude that the semiparametric GAS model provides a useful alternative to standard semiparametric GARCH models for in-sample fit and out-of-sample forecast accuracy.

The remainder of this paper is organized as follows. Section 2 introduces the model. Section 3 provides Monte Carlo evidence of the new model's performance. Section 4 presents our empirical application to equity return data and density forecast evaluation. Section 5 concludes.

2 The semiparametric volatility model

2.1 Semiparametric GARCH models

Let $y_t \in \mathbb{R}$ denote a time series of financial returns. The standard GARCH(1,1) model is given by

$$y_t = \mu + \xi_t = \mu + h_t^{1/2} \varepsilon_t, \quad \varepsilon_t \sim q(\varepsilon_t), \quad (1)$$

with $\mu \in \mathbb{R}$ the conditional mean, $q(\cdot)$ the density of the standardized error term ε_t , and $h_t \in \mathbb{R}^+$ the conditional variance satisfying the recursion

$$h_{t+1} = \omega \cdot (1 - \alpha - \beta) + \alpha \xi_t^2 + \beta h_t, \quad (2)$$

with parameters $\omega \in \mathbb{R}^+$, and $\alpha, \beta > 0$, and $\alpha + \beta < 1$. The standardized density $q(\cdot)$ does not depend on h_t . Note that we can easily replace the conditional mean μ by a non-constant conditional mean involving exogenous regressors or autoregressive moving average components. Moreover, we can include lags of ξ_t and h_t in (2).

If the distribution $q(\cdot)$ of ε_t is unknown and needs to be estimated, model (1)–(2) is called a semiparametric GARCH(1,1) model. Estimation is usually performed in two steps. First, one uses a Quasi Maximum Likelihood estimator (QMLE) based on a standard normality assumption for ε_t to estimate $(\mu, \omega, \alpha, \beta)$. Using the standardized residuals and nonparametric density estimation techniques, one can then obtain an estimate $\hat{q}(\cdot)$ of the error density. The estimated density is used to estimate the parameters in a second stage by semiparametric ML, assuming the density estimate $\hat{q}(\cdot)$ is the true density. Engle and Gonzalez-Rivera (1991) uses the discrete maximum penalized likelihood estimator (DMPLE) of Tapia and Thompson (1978) for density estimation, followed by a BHHH algorithm to maximize the semiparametric likelihood. Drost and Klaassen (1997) and Sun and Stengos (2006) use standard kernel density methods to estimate the error density and develop a Newton-Raphson algorithm to maximize

the semiparametric likelihood. Their estimator achieves the same asymptotic efficiency for (α, β) as the true MLE.

2.2 The semiparametric Generalized Autoregressive Score model

In model (1)–(2), the form of the error density has no impact on the volatility dynamics (2). Whatever the form of the error density $q(\cdot)$, future volatility levels h_{t+1} always react to the squared errors ξ_t^2 . The reaction pattern can be extended to allow for asymmetric reactions as in Nelson (1991) or Glosten, Jagannathan, and Runkle (1993), but this would still leave the volatility dynamics unrelated to the form of the error density. Following the arguments in Creal, Koopman, and Lucas (2011) as discussed in the introduction, such a strict delinkage may be counterintuitive.

In order to build a natural link between the error density and the volatility dynamics, we use the generalized autoregressive score (GAS) framework of Creal, Koopman, and Lucas (2011, 2012). The GAS framework nests many other successful time series models, such as the GARCH model, the autoregressive conditional duration (ACD) model of Engle and Russell (1998), and the multiplicative error model (MEM) of Engle and Gallo (2006). The GAS(1,1) specification that we use in our current volatility context is given by equation (1) and

$$f_{t+1} = \tilde{\omega} \cdot (1 - \tilde{\beta}) + \tilde{\alpha} s_t + \tilde{\beta} f_t, \quad (2')$$

where $h_t = \exp(f_t)$, $\tilde{\omega}, \tilde{\alpha} \in \mathbb{R}$, $|\tilde{\beta}| < 1$, and

$$s_t = \frac{\partial \log \left(h_t^{-1/2} \cdot q(h_t^{-1/2} \xi_t) \right)}{\partial f_t}. \quad (3)$$

Further lags of s_t and f_t can easily be included. Equation (2') has two main differences compared to (2): the use of the density score s_t rather than ξ_t , and the parameterization of the log-variance f_t rather than h_t itself. We discuss each of these in turn.

The use of the score s_t rather than ξ_t^2 is the main distinguishing feature of the GAS framework. As explained in Creal, Koopman, and Lucas (2012), the intuition is that a local steepest ascent step is taken in the time varying parameter f_t to improve the local log-density fit $p(y_t|h_t)$ of the model at time t , where $p(\cdot)$ denotes the observation density. As a result, the time-varying parameter automatically adapts to the most recent model mis-fit as measured by the score of the density. For example, if $q(\cdot)$ is the Gaussian density, this implies that f_t reacts to a linear function of ξ_t^2 as expected. If $q(\cdot)$ is the Student t distribution with ν degrees of freedom, Creal,

Koopman, and Lucas (2011) show that (2') reduces to

$$f_{t+1} = \tilde{\omega} + \tilde{\alpha}(1 + 3\nu^{-1}) \left(\frac{1 + \nu}{\nu - 2 + \xi_t^2/h_t} \cdot \frac{\xi_t^2}{h_t} - 1 \right) + \tilde{\beta}f_t, \quad (4)$$

which clearly shows that large standardized squared errors ξ_t^2/h_t for fat-tailed ($\nu < \infty$) densities automatically have less impact on the volatility dynamics due to the weight factor $(1 + \nu)/(\nu - 2 + \xi_t^2/h_t)$, which is decreasing in ξ_t^2/h_t . As argued before, this is intuitive: under fat-tailed densities we expect to see large values of ξ_t^2/h_t more often, and they should therefore not automatically be taken as a signal that the volatility has increased.

The second difference between (2) and (2') is the parameterization of the log variance $f_t = \log(h_t)$. Though this parameterization is not crucial and other parameterizations (such as $f_t = h_t$) are also possible, the current parameterization has a number of advantages. First, we need not to worry about the signs of the coefficients $\tilde{\omega}$, $\tilde{\alpha}$, and $\tilde{\beta}$, as the variance h_t is positive by construction. This brings the current specification close to the familiar EGARCH model of Nelson (1991). Second, the GAS specification in (2') only uses the score or gradient. Creal, Koopman, and Lucas (2012) argue that it is useful to scale the score by some function of the local curvature of the log-density in order to enhance model stability. Under the current parameterization, however, we obtain

$$s_t = -\frac{1}{2} - \frac{1}{2} \frac{q'(h_t^{-1/2}\xi_t)}{q(h_t^{-1/2}\xi_t)} \cdot \frac{\xi_t}{h_t^{1/2}},$$

with $q'(\cdot)$ denoting the derivative of $q(\cdot)$, such that

$$\mathbb{E}_{t-1} [s_t^2] = \frac{1}{4} \mathbb{E}_{t-1} \left[\left(1 + \frac{\varepsilon_t \cdot q'(\varepsilon_t)}{q(\varepsilon_t)} \right)^2 \right], \quad (5)$$

where $\mathbb{E}_{t-1}[\cdot]$ denotes a conditional expectation given all the data up to time $t-1$. As $\mathbb{E}_{t-1}[s_t] = 0$ for a correctly specified model, (5) equals the Fisher information matrix. Equation (5) then shows that the information is static and only depends on the (standardized) error density $q(\cdot)$. As a result, under the log parameterization $f_t = \log(h_t)$, information on the time-varying parameter is constant over time and scaling by (a function of) the Fisher information can be omitted or assumed to be included in the parameter $\tilde{\alpha}$.

Parameter estimation for the GAS model is straightforward. Given the observation driven nature of the GAS model, the likelihood can be expressed in closed form and can be computed efficiently through a prediction error decomposition. In particular, we obtain parameter estimates by maximizing the function

$$L = \sum_{t=1}^n -\frac{1}{2} \log h_t + \log q \left((y_t - \mu)/h_t^{1/2} \right), \quad (6)$$

where h_t is defined recursively through (2'). If the error density is assumed to be of a certain parametric form such as the standard normal or (skewed) Student's t distribution, we obtain a parametric model as in Creal, Koopman, and Lucas (2011). However, to avoid the introduction of an incorrect parametric specification, we estimate $q(\cdot)$ nonparametrically using a kernel density estimator as in Drost and Klaassen (1997) and Sun and Stengos (2006). To do so, we first estimate the model's parameters by assuming a specific parametric family for $q(\cdot)$ and interpret the estimator based on (6) as a Quasi Maximum Likelihood estimator (QMLE). The QMLE allows us to compute standardized residuals $\hat{\varepsilon}_t = (y_t - \hat{\mu})/\hat{h}_t^{1/2}$, which are the input of a kernel density estimator to estimate $q(\cdot)$. In particular, we use a standard kernel density estimator

$$\hat{q}(x) = \frac{1}{n \cdot b_n} \sum_{i=1}^n k\left(\frac{\hat{\varepsilon}_t - x}{b_n}\right), \quad (7)$$

where n is the sample size, b_n is a bandwidth or smoothing parameter, and $k(\cdot)$ is a kernel function. We use the standard normal kernel

$$k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}, \quad -\infty < v < \infty. \quad (8)$$

We rescale and relocate $\hat{q}(\cdot)$ such that it has mean zero and unit variance and use it in (6) to estimate the parameters $(\mu, \omega, \alpha, \beta)$. The resulting estimator is labeled the Semiparametric Maximum Likelihood Estimator (SMLE) for the semiparametric GAS volatility model. It is denoted as SMLE(1).

Once SMLE(1) is obtained, we can recalculate the standardized residuals and obtain a new kernel density estimate of $q(\cdot)$ to be used in (6). Maximizing this new loglikelihood function we get new estimates, which are denoted as SMLE(2). Similarly, SMLE(i) can be obtained for $i = 3, 4, \dots$, by repeating this process. In our simulation study below, we restrict our attention to $i = 1, 2, 3$. Further iterations typically did not produce valuable improvements in performance.

3 Monte carlo evidence

To study the model's behavior, we carry out two sets of simulation experiments. In the first set of simulation experiments, we use the model in Section 2.2 as the data generating process (DGP). We simulate return series and investigate the estimation efficiency and volatility forecast accuracy of the semiparametric approach compared to the MLE and the (Gaussian) QMLE model. In the second set of simulation experiments, we allow the model to be mis-specified by

simulating return series with deterministic as well as stochastic volatility patterns. This allows us to investigate whether the model's performance is also robust. The reported simulation results are a subset of a more elaborate set of (unreported) simulation experiments, the output of which is available as an online companion appendix to this paper.

3.1 Results under correct specification

We use the DGP (1)–(2') for four different choices of $q(\cdot)$, namely the standard normal, the Student's t distribution with $\nu = 3$ and $\nu = 5$ degrees of freedom, and a mixture of normals. The mixture of normals is chosen such that it is close (in a Kulback-Leibler sense) to a chi-squared distribution with 6 degrees of freedom, thus exhibiting substantial skewness. Each of the above densities is rescaled and relocated to have zero mean and unit variance. For each DGP, we simulate $B = 100$ return series of length $n = 2000$. The first 1000 observations are labeled the in-sample observations and are used to estimate the model's parameters. The last 1000 observations are called out-of-sample. We set $\mu = 0$, $\tilde{\omega} = 2$, $\tilde{\alpha} = 0.3$, and $\tilde{\beta} = 0.9$. Reasonable changes of the parameter values do not alter the conclusions below. Note that $\tilde{\beta}$ takes the role of $\alpha + \beta$ from the familiar GARCH model. It is therefore not a problem that $\tilde{\alpha} + \tilde{\beta} > 1$, see also (Blasques, Koopman, and Lucas 2012).

For each simulated sample, we compute the parameter estimates using four different statistical models: MLE, QMLE, QMLE- $t(\nu)$ and SMLE(1). We also iterate the semiparametric approach to obtain SMLE(2) and SMLE(3). As the simulation results for SMLE(2)–(3) are comparable to those of SMLE(1), we do not report them here. QMLE uses the normal distribution as a quasi-likelihood, while QMLE- $t(\nu)$ uses the Student's t density with ν degrees of freedom as a quasi-likelihood, where ν is estimated together with the other parameters.

The semiparametric estimators require the choice of the bandwidth parameter b_n . In our current setting, we use $b_n = 0.5$ as in Drost and Klaassen (1997). Other reasonable choices $0.3 \leq b_n \leq 0.8$ do not alter our conclusions. Note that the choice of the (quasi)-density does not only matter for the likelihood, but also for the volatility dynamics in (2').

The simulations allow us to investigate two dimensions of performance of the different statistical models and estimators. First, we evaluate the small-sample properties of the estimators for the parameters $(\mu, \tilde{\omega}, \tilde{\alpha}, \tilde{\beta})$. We are particularly interested to see which parameters are adaptively estimable in the sense that their finite-sample behavior in terms of bias and root mean squared error (RMSE) resembles that of the MLE. Second, we can evaluate and compare the forecast accuracy of the different models by comparing the simulated value of $h_t^{1/2}$ with its

fitted value $\hat{h}_t^{1/2}$. The RMSE for the volatility estimate can be used to distinguish the forecasting performance of the different models. We can do this in-sample based on the first 1000 simulated observations, as well as out-of-sample using the last 1000 observations.

For the semiparametric estimator SMLE(1), we use the QMLE- $t(\nu)$ estimate as our starting value. We do not use the MLE estimate, as ML estimation is not feasible in practice. Table 1 reports the results.

We concentrate our discussion on $\tilde{\alpha}$ and $\tilde{\beta}$, as $\tilde{\mu}$ and $\tilde{\omega}$ cannot be estimated adaptively, see Drost and Klaassen (1997). The RMSE obtained with the MLE serves as our efficiency bound for each parameter. For the normal distribution (N), the MLE equals the QMLE. The RMSE values for the (Q)MLE are close to those of the QMLE- $t(\nu)$ estimator. The largest (relative) efficiency loss appears for $\tilde{\alpha}$ and equals 10% relative to a benchmark RMSE of 0.046 for (Q)MLE. The relative RMSE of SMLE(1) for $\tilde{\beta}$ is also small. A larger increase in relative RMSE is seen for $\tilde{\alpha}$. We also note that though the relative increase of RMSE for SMLE(1) for $\tilde{\alpha}$ appears substantial, this relative increase is with respect to a low absolute level of the RMSE for the QMLE in this case. The RMSEs of the QMLE for the non-Gaussian distributions that are discussed next are considerably higher and show an advantage of the SMLE(1) compared to the QMLE.

For the fat-tailed distributions $t(3)$ and $t(5)$, the (Gaussian) QMLE performs poorly in terms of RMSE compared to the MLE. For $\tilde{\beta}$, the RMSE of the QMLE is about 6 and 3 times as high as that of the MLE for the $t(3)$ and $t(5)$ case, respectively. For $\tilde{\alpha}$ the factors are 3 and 2, respectively. The Student's t based QMLE does considerably better and almost attains the same level of efficiency as the MLE. The SMLE(1) also performs well. Though it is not as efficient as the MLE, it performs considerably better than the Gaussian QMLE. In particular, a large part of the RMSE increase in the QMLE compared to the MLE is recovered by the use of SMLE(1). The inferior performance of the QMLE should not come as a surprise: the QMLE model is doubly mis-specified in our case. Not only does it impose the incorrect conditionally Gaussian distributional assumption on ε_t . It also assumes the incorrect specification for the volatility dynamics in (2'). In particular, large values of $(y_t - \mu)^2$ receive a higher impact in the QMLE model than warranted under the true data generating process as specified by the volatility dynamics (4).

The simulation results for the mixture of normals is interesting due to the distribution's substantial skewness. The QMLE is inferior to the MLE for both $\tilde{\alpha}$ and $\tilde{\beta}$. Compared to the QMLE- $t(\nu)$ and SMLE(1), the results are mixed. The RMSEs for $\tilde{\beta}$ of the QMLE are higher, while those for $\tilde{\alpha}$ are lower.

We shed some more light on the differences between the different models by comparing their forecast accuracies. Except for the normal density, we see that the SMLE reduces part of the efficiency loss incurred by the QMLE compared to the MLE. The reductions in efficiency loss of the SMLE(1) compared to the QMLE range from around 55% for $t(3)$ and $MN(\chi^2(6))$ to around 80% for $t(5)$. Similar improvements are seen in-sample and out-of-sample.

3.2 Results under mis-specification: deterministic volatility

In Section 3.1, the set of statistical models embedded the true Data Generating Process (DGP). In reality, however, we do not know whether the statistical model encompasses the DGP. To see how the different models perform if the DGP lies outside the set of all statistical models considered, we conduct a second simulation experiment similar to Engle (2002). We choose a number of deterministic volatility patterns and estimate each of the statistical models discussed earlier. As our performance metric, we only choose the in-sample and out-of-sample RMSEs of the volatility. The RMSEs of the parameters are less useful here, as the statistical model never encompasses the DGP. The specific deterministic volatility patterns are presented in Figure 1. They include slowly moving patterns of volatility (sine), long-lived and short-lived structural changes in volatility, and regular bursts of volatility that gradually taper off.

The simulation results are presented in Table 2. We normalize each RMSE with respect to the in-sample RMSE of QMLE. Also note that even the MLE entry is mis-specified here: though it uses the correct distributional assumptions, the volatility dynamics in (2') are incorrect under the deterministic volatility patterns in Figure 1.

The general picture emerging from Table 2 is that for each density and most volatility patterns, the RMSEs of SMLE(1) are close to those of MLE. This reveals that in terms of in-sample fit and out-of-sample forecast accuracy, SMLE performs well. We particularly note that if the difference between QMLE and MLE is large, then SMLE(1) also performs considerably better than QMLE. Examples of this are provided by the fat-tailed $t(3)$ and $t(5)$ distributions. We conclude that the SMLE also performs well under possible model mis-specification: volatilities are estimated and forecast more accurately than using the standard QMLE approach for a variety of deterministic volatility patterns and error distributions.

3.3 Results under mis-specification: stochastic volatility

To conclude our simulations, we investigate the performance of the semiparametric GAS model if the data is generated by a stochastic volatility process. This may appear more reasonable

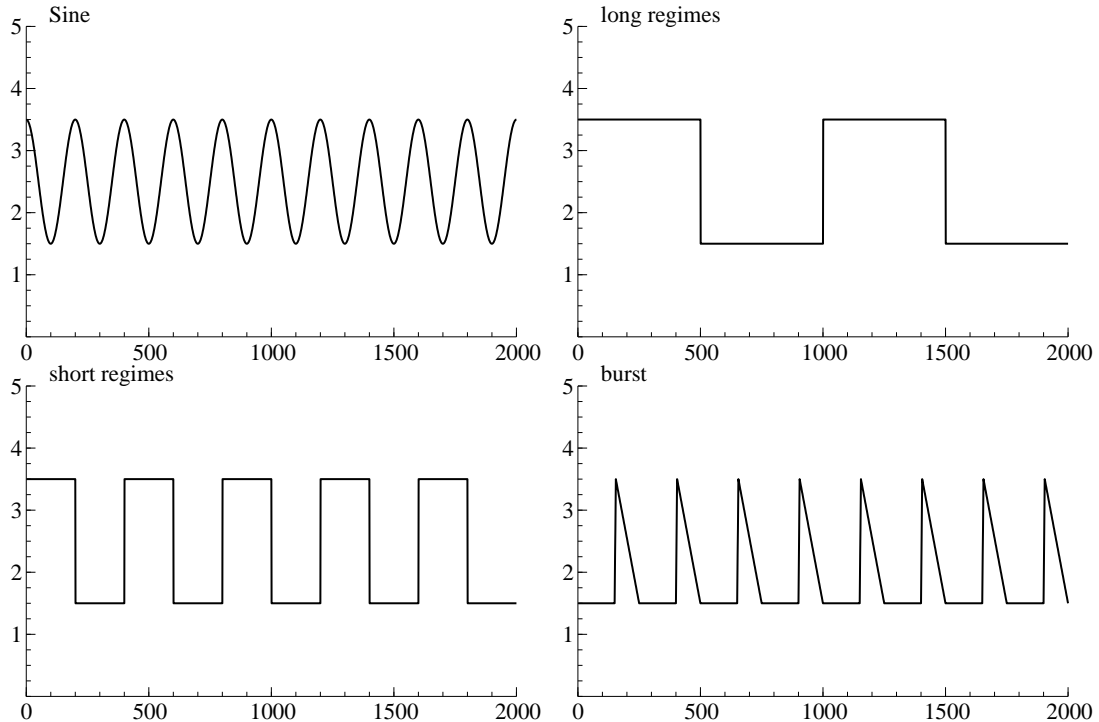


Figure 1: Deterministic volatility patterns

The specifications are as follows. Sine: $\sigma_t = 2.5 + \cos(2\pi t/200)$. Long regimes last 500 days and have $\sigma_t = 1.5, 3.5$. Short regimes have the same values for σ_t but last 200 days. Burst: $\sigma_t = 1.5$ for 150 days, then grows linearly to $\sigma_t = 3.5$ over 5 days, and then linearly decreases to $\sigma_t = 1.5$ over the next 95 days. Then the burst pattern restarts.

than the deterministic volatility patterns from Section 3.2. In addition, we benchmark the performance of the semiparametric GAS model to that of the semiparametric GARCH model. This allows us to obtain a better perspective on the value-added of the altered dynamics of volatility in equation (2').

The stochastic volatility (SV) DGP is specified as $y_t \sim p(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.01 + 0.98\alpha_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1^2)$, for $t = 1, \dots, n$. The distribution p can be any of the four distributions used in 3.1. We also performed simulations with a less persistent SV process, but the results are qualitatively similar and not reported here. They are available in the online appendix.

The results are reported in Table 3. For each distribution and estimator, we compute a test statistic to see whether over the 100 Monte Carlo simulations the average RMSE of the GAS specification is significantly lower than that of the GARCH specification. Note that even under the normal distribution the GAS and GARCH specifications do not coincide, as the former parameterizes the log volatility, whereas the latter parameterizes the variance itself.

The table shows that for the normal distribution, the performance of QMLE, MLE, and SMLE(1) are very similar, both in-sample and out-of-sample. Also the performance of the different dynamic specifications for the volatility does not seem to matter much. The GAS and GARCH models behave similarly. In particular, looking at the test whether GAS significantly outperforms GARCH, we see that none of the p-values indicate a rejection of the null that the two specifications have equal performance.

The picture changes substantially if we consider the fat-tailed distributions $t(3)$ and $t(5)$. We see that the QMLE- $t(\nu)$ and SMLE(1) improve upon QMLE for the GAS specification, but not so much for the GARCH specification. This is due to the fact that under the SV specification, the volatility is a smoothly evolving Gaussian process. The data, however, are drawn from a fat-tailed distribution. This implies that from time to time a large value is drawn, even though the volatility has not changed considerably. The volatility in the GARCH specification directly reacts to this through its dependence on squared lagged observations. This explains the limited improvement of the forecast accuracy for the GARCH specification. The GAS model, by contrast, accounts for the fat-tailed nature of the observation density and results in more moderate changes in the variance. This results in a better overall performance in-sample and out-of-sample compared to the QMLE. The improvement in RMSE is significant for both the QMLE- $t(\nu)$ and SMLE, as can be seen from the reported p-values.

A similar result is seen for the mixture of normals, particularly for SMLE. Though the semiparametric GARCH model improves on the QMLE here, the improvement is more substantial for the GAS SMLE. The reason for this is that the GAS SMLE exploits the fact that left and right-hand tail observations convey different information about the changes in volatility. The GARCH approach, however, exploits the asymmetric density properties for the likelihood, but not for the volatility dynamics. The latter still only depend on squared lagged observations. The result is a significant difference in in-sample and out-of-sample fit of the semiparametric GAS and GARCH specifications.

We conclude that our semiparametric volatility model also performs well under stochastic volatility DGPs: volatilities are still estimated comparatively well compared to the competing semiparametric GARCH models. The results are further corroborated in an (unreported) more comprehensive set of simulation experiments with more densities: if the true underlying density has fatter tails or more skewness, and if the underlying stochastic volatility process is more persistent, the gains of the semiparametric GAS specification over the semiparametric GARCH model are larger.

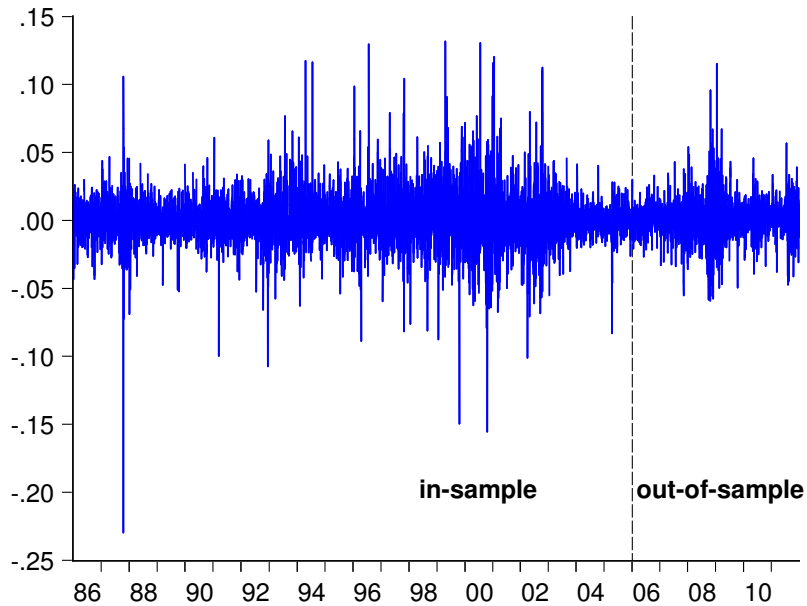


Figure 2: Daily IBM Returns

IBM returns, 01/02/1986 - 12/31/2010. In-sample period is 01/02/1986 - 12/30/2005. Out-of-sample period is 01/03/2006 - 12/30/2011.

4 Empirical application and density forecast evaluation

4.1 Estimation results

We apply the new semiparametric GAS volatility model to IBM daily returns from Jan 2, 1986 to Dec 30, 2011. We split the sample into an in-sample estimation period containing the data up to Dec 31, 2005 and a subsequent out-of-sample forecasting period. The latter is used for density forecast evaluation. The in-sample and out-of-sample periods contain 5047 and 1511 observations, respectively. Figure 2 presents the data.

The estimation results are presented in Table 4. We report the results for the score driven GAS models as well as for a number of GARCH models. For comparison, we report the results in terms of persistence parameter $\tilde{\beta}$ of the GAS specification and $\alpha + \beta$ of the GARCH specification; see also Creal, Koopman, and Lucas (2011).

The volatility is highly persistent across all specifications. All values of $\tilde{\beta}$ and $\alpha + \beta$ are close to one. The estimate of $\tilde{\omega}$ for the GAS models is the result of a slight reparameterization: we estimate the model by replacing $\tilde{\omega}$ by $\tilde{\omega}(1 - \tilde{\beta})$, such that $\tilde{\omega}$ is 2 times the unconditional expectation of the log daily volatility. A value of $\tilde{\omega} \approx -8$ then corresponds to annualized

volatility levels of around 28%.

The degrees of freedom parameter for the QMLE- $t(\nu)$ is estimated at about $\nu = 5$, indicating a substantial degree of (conditional) fat-tailedness of the returns. The small value of ν have a double impact: it changes the likelihood, and it changes the impact of incidental large observations on the volatility dynamics through equation (2'). As a result, the likelihood value increases by more than 424 points from QMLE-normal to QMLE- $t(\nu)$. To decompose this effect into its two components, we also estimate the model QMLE* - $t(\nu)$. This model has the Student's t likelihood, but the volatility dynamics from the QMLE-normal model. We see that the biggest effect (more than 369 points) is due to the change of the likelihood. A substantial part (of about 55 points), however, is due to the altered volatility dynamics. In terms of likelihood value, the switch from $t(\nu)$ to a non-parametrically estimated density is relatively small (about 6 points). Interestingly, however, the likelihood values of both QMLE- $t(\nu)$ and SMLE(1) are substantially higher than their GARCH counterparts GARCH- $t(\nu)$ and semiparametric-GARCH, respectively, the differences being around 23 to 25 points. As the models are non-nested, however, a strict likelihood comparison is unwarranted.

Table 4 only presents the estimates for the parametric part of the model. In particular, the estimate $\hat{q}(\cdot)$ of the error density is not shown. Note that in contrast to the semiparametric GARCH model, the error density is not a nuisance parameter in the GAS model any more. Rather, it is a crucial part of the model through its impact on the volatility dynamics in (2'). A different estimate of the error density results in a different news impact curve, i.e., a different functional relationship between values of ε_t and values of f_{t+1} . To obtain some more insight, we plot the estimated score functions. The results are presented in Figure 3.

Figure 3 shows that for the QMLE the impact has its familiar quadratic shape. Large positive and negative standardized residuals result in an identically large increase in the (log) volatility level f_t . For the QMLE- $t(\nu)$, the score function has an entirely different shape; see also Creal, Koopman, and Lucas (2011). In particular, large positive and negative standardized residuals have a bounded impact on future log volatility levels. This is due to the downweighting effect for the Student's t distribution in (4) as explained earlier.

The curves for the SMLE estimator look more like those of the Student's t than those of the Gaussian distribution. For standardized residuals between -2.5 and 2.5, the score functions behave like that of QMLE- $t(\nu)$. For larger absolute values of the standardized residuals, however, the values of the score functions remain at a finite level or even decrease. This means that large absolute innovations in the semiparametric model are more likely to be attributed to a draw from the tail area rather than to an increase in volatility. It is also interesting to see that the

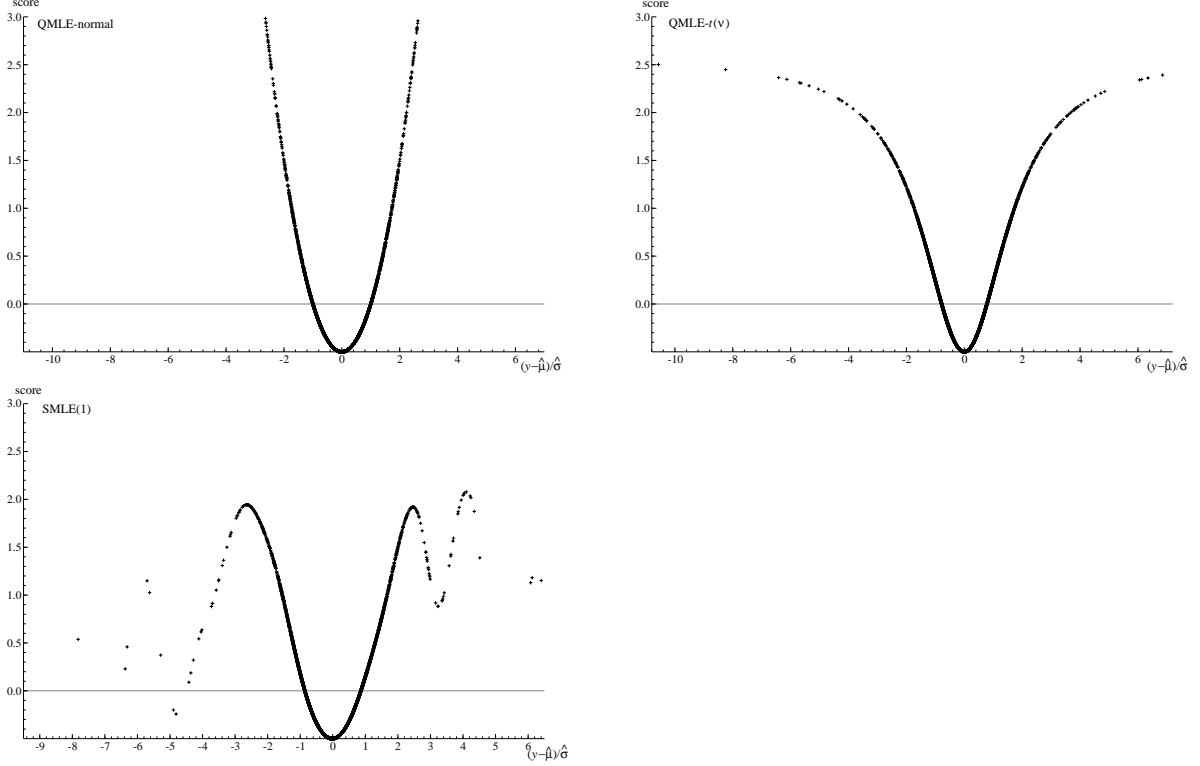


Figure 3: Plots of Score Functions

Plots of the score steps s_t in equations (2') and (3) as a function of $\hat{\varepsilon}_t = (y_t - \hat{\mu})/\hat{\sigma}_t$.

news impact curve decreases in $|\varepsilon_t|$ for large negative ε_t , whereas it appears to level off for large positive values. This indicates that the semiparametric approach recognizes the error density to be left-skewed. As a result, the volatility dynamics attribute less of the left-tail observations to local volatility increases and more to the skewed nature of the error distribution.

4.2 Density forecast evaluation

So far, we have concentrated on the models' performance in-sample. We now proceed to evaluate the performance of the different models out-of-sample. Our main tool to compare the different methods is the density forecast evaluation methods as proposed by Diebold, Gunther, and Tay (1998). Assume we have series of true densities $\{p(y_t|h_t)\}_{t=1}^m$ as well as density forecasts $\{\hat{p}(y_t|h_t)\}_{t=1}^m$, where m is the number of observations used in the evaluation. Although $\{p(y_t|h_t)\}_{t=1}^m$ is never observed, Diebold, Gunther, and Tay (1998) show that we can evaluate density forecasts by assessing the probability integral transforms z_t , with

$$z_t = \int_{-\infty}^{y_t} \hat{p}(u|h_t) du. \quad (9)$$

They prove that if a sequence of density forecasts $\{\hat{p}(y_t|h_t)\}_{t=1}^m$ coincides with $\{p(y_t|h_t)\}_{t=1}^m$, then under usual mild conditions the sequence of probability integral transforms z_t is independently and identically distributed (i.i.d.), each z_t being standard uniform between 0 and 1, i.e., $z_t \stackrel{\text{i.i.d.}}{\sim} U[0, 1]$. We can thus evaluate a density forecast by assessing whether the probability integral transforms z_t are i.i.d. $U[0, 1]$. Diebold, Gunther, and Tay (1998) suggest to test this using a histogram plot with confidence bounds for each of the bins underlying the histogram. They also suggest the use of correlograms of $(z_t - \bar{z})^i$ for $i = 1, \dots, 4$, where \bar{z} is the time-series average of z_t . Diebold, Gunther, and Tay (1998) argue that these values of i suffice to reveal dependence through the conditional mean, conditional variance, conditional skewness, or conditional kurtosis.

Figure 4 presents the histograms of $\{z_t\}_{t=1}^m$ and Figure 5 presents the correlograms of $(z - \bar{z})^2$ based on the out-of-sample part of the data only. We do not present correlograms of $(z - \bar{z})^i$ for $i = 1, 3$, because there is almost no significant serial correlation in both $(z - \bar{z})$ and $(z - \bar{z})^3$ for any of the models considered, whereas the correlograms for $(z - \bar{z})^4$ look very much like those of $(z - \bar{z})^2$. We present the results for QMLE-normal, QMLE- $t(\nu)$, SMLE(1) and semiparametric-GARCH.

For the Gaussian model, the upper-left panel in Figure 4 shows that the histogram of the probability integral transforms has a distinct, nonuniform ‘butterfly’ shape. There is a hump in the middle and a lower region in each tail. This indicates that too many of the realizations fall in the middle of the distribution and the tails of the model are relatively too thin compared to the realized data. This is exactly what we expect for financial returns. The horizontal lines in the histogram indicate the approximate 95% confidence intervals for the individual bin heights under the null that z_t is i.i.d. $U[0, 1]$.

The upper-left graph of Figure 5 shows the correlogram of $(z - \bar{z})^2$. The dashed lines are Bartlett’s approximate 95% confidence intervals under the null that z_t is i.i.d. There is some small but significant serial correlation in $(z - \bar{z})^2$ for QMLE-normal. This means that the Gaussian GAS model does not succeed in capturing the second order dynamics of the data, or that the volatility dynamics might still be mis-specified.

The upper-right graph of Figure 4 shows the histogram for the Student’s t based model. The histogram in this case is much closer to the uniform distribution, except for a slight peak in the right tail. In particular, the butterfly pattern has disappeared because we allow for fat tails and leptokurtosis. The peak in the right tail signals that the symmetric Student’s t model assigns too much tail mass to the right-hand tail. Put differently, the left tail should be fatter than the right-hand tail, which is commensurate with the left skew found in the impact curves

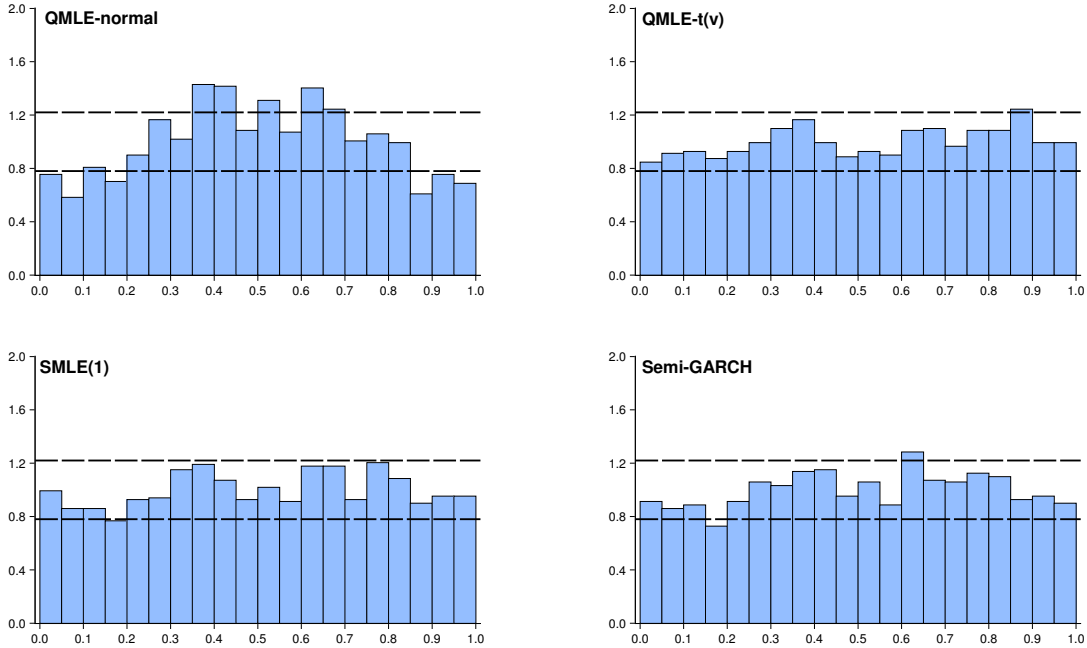


Figure 4: Density Forecast Evaluation: densities of z_t

The different panels plot the density estimates of the probability integral transforms z_t for four different statistical models: QMLE-normal(upper left), QMLE- $t(\nu)$ (upper right), SMLE(1)(lower left), and semiparametric-GARCH(lower right).

shown earlier. The upper-right graph of Figure 5 shows the correlogram for the Student's t based model. Interestingly, the correlogram of $(z_t - \bar{z})^2$ no longer shows any evidence of mis-specification in the volatility dynamics after accounting for the reduced impact of large observations on volatility levels through equation (4).

For the semiparametric model, the lower-left graph in Figure 4 reveals that the histogram of the probability integral transform remains entirely inside the pointwise confidence 95% band. As the nonparametric estimate of the error density accommodates both the fat tails and the skewness prevalent in the empirical data, we no longer detect any mis-specification here. Also the lower-left graph in Figure 5 shows no evidence of mis-specified conditional volatility dynamics.

For the semiparametric GARCH model, the lower-right graph in Figure 4 shows that the density of the probability integral transforms is close to the uniform distribution, except for a small peak in the middle. We expect some violations in the center of the distribution for the semiparametric-GARCH model under fat-tailed innovations. Due to the GARCH dynamics, a large observation increases volatility for a few periods, while such observations are from

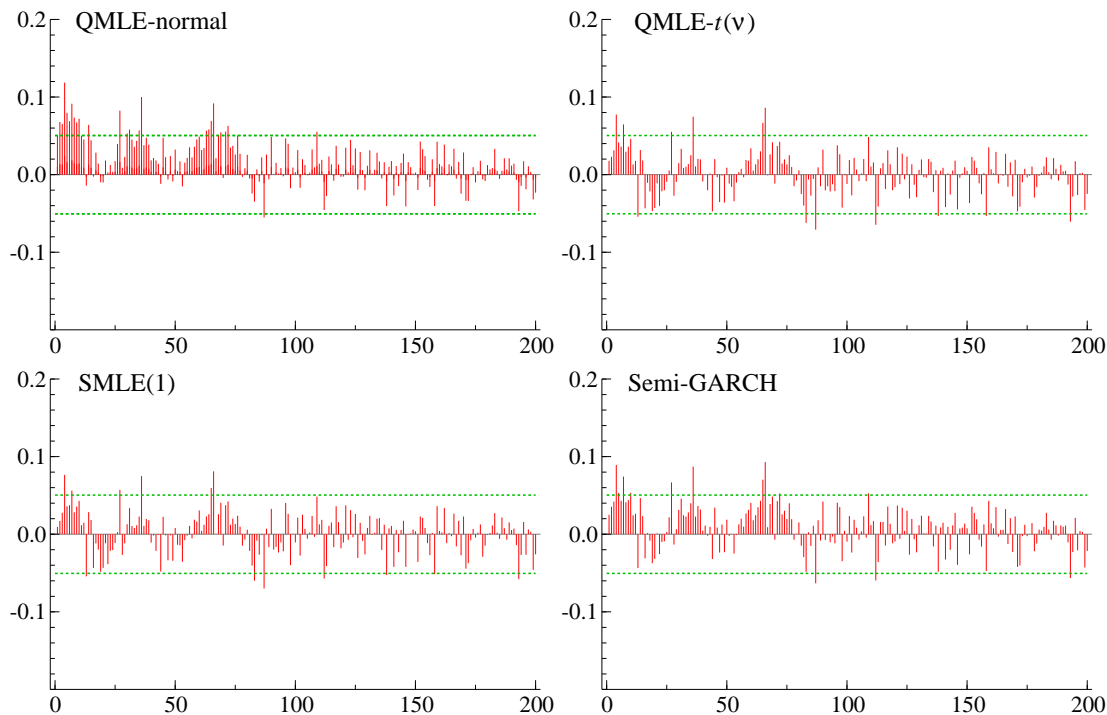


Figure 5: Density Forecast Evaluation: correlogram of $(z_t - \bar{z})^2$

The figure plots the autocorrelation functions of the squares of the demeaned probability integral transforms z_t . The different panels are for four different statistical models: QMLE-normal(upper left), QMLE- $t(\nu)$ (upper right), SMLE(1)(lower left), and Semi-GARCH(lower right).

time to time due to the fat-tailedness rather than to local increases in volatility. As a result, the volatility will be too large for a few periods, resulting in an (unduely) higher number of probability integral transforms z_t near the center. This is corroborated by the lower-right graph in Figure 5, which is similar to that of QMLE-normal. There is small but significant correlation in $(z_t - \bar{z})^2$, resulting from the volatility dynamics. We therefore conclude that the semiparametric volatility model proposed in this paper provides an improvement over the QMLE method as well as over the semiparametric-GARCH model, both in terms of in-sample model fit and out-of-sample density forecast evaluation.

5 Conclusion

We have proposed a new semiparametric observation-driven volatility model. As in the standard semiparametric GARCH models, we use kernel density estimation methods to estimate the error density. Unlike the semiparametric GARCH models, however, the density estimate also enters

the volatility dynamics directly via a Generalized Autoregressive Score (GAS) structure of the model. As a result, the volatility dynamics in the new model are much less sensitive to incidental influential observations and outliers.

We showed that it is easy to estimate the model by standard maximum likelihood methods. We conducted three simulation experiments. In the first one, we showed that the model regains parts of the estimation inefficiency of the standard (Gaussian) QMLE. In the second and third experiment, we used deterministic and stochastic volatility dynamics, respectively, to investigate the effect of model mis-specification on the performance of the different statistical models. We demonstrated that the new model gives in-sample estimation and out-of-sample forecasts of volatility levels of similar accuracy as when we would know the true error density. For stochastic volatilities, we also made a comparison of our model with the semiparametric-GARCH model. This revealed that our model outperforms the semiparametric-GARCH model when the density is fat-tailed and/or skewed.

In our empirical example, we illustrated the performance of the new model for stock return data. By plotting the score function, we saw that the new model has moderate increase in volatility following large absolute innovations. Volatility levels also reacted differently to large positive versus large negative innovations due to the estimated skewness of the distribution. When evaluating the model using density forecast evaluation, we showed that the model performs better than its competitors, both in terms of in-sample model fit and out-of-sample density forecast evaluation.

The new semiparametric volatility model with generalized autoregressive score dynamics thus performs well both in-sample and out-of-sample, and both for empirical and simulated data. It provides a useful alternative to standard semiparametric GARCH models. Moreover, the model is also easily extended to allow for more complicated dynamics, such as higher order or even fractional dynamics in either the mean or the volatility specification.

References

- Baillie, R. T., and T. Bollerslev. 1989. "The message in daily exchange rates: a conditional-variance tale." *Journal of Business and Economic Statistics* 7:297–305.
- Bauwens, L., and S. Laurent. 2005. "A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models." *Journal of Business and Economic Statistics* 23 (3): 346–354.
- Blasques, F., S.J. Koopman, and A. Lucas. 2012. "Stationarity and ergodicity of univariate generalized autoregressive score processes." *Tinbergen Institute Discussion Paper*.

- Bollerslev, T. 1986. "Generalized autoregressive conditional heteroscedasticity." *Journal of Econometrics* 31(3):307–327.
- Creal, D. D., S. J. Koopman, and A. Lucas. 2011. "A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations." *Journal of Business and Economic Statistics* 29:552 – 563.
- . 2012. "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics*, p. forthcoming.
- Diebold, F. X., T. Gunther, and A. Tay. 1998. "Evaluating Density Forecasts." *International Economic Review* 39:863–883.
- Drost, F. C., and C. A. J. Klaassen. 1997. "Efficient estimation in semiparametric GARCH models." *Journal of Econometrics* 81(1):193–221.
- Engle, R. F. 1982. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica* 50(4):987–1007.
- . 2002. "Dynamic Conditional Correlation." *Journal of Business and Economic Statistics* 20(3):339–350.
- Engle, R. F., and G. M. Gallo. 2006. "A multiple indicators model for volatility using intra-daily data." *Journal of Econometrics* 131:3–27.
- Engle, R. F., and G. Gonzalez-Rivera. 1991. "Semiparametric ARCH models." *Journal of Business and Economic Statistics* 9:345–359.
- Engle, R. F., D.M. Liliien, and R.P. Robins. 1987. "Estimating time varying risk premia in the term structure: the ARCH-M model." *Econometrica* 55:391–407.
- Engle, R. F., and J. R. Russell. 1998. "Autoregressive conditional duration: a new model for irregularly spaced transaction data." *Econometrica* 66(5):1127–1162.
- Glosten, L.R., R. Jagannathan, and D.E. Runkle. 1993. "On the relation between the expected value and the volatility of the normal excess return on stocks." *The Journal of Finance* 48:1779–1801.
- Lee, S.-W., and B. E. Hansen. 1994. "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator." *Econometric Theory* 10(1):29–52.
- Linton, O., and E. Mammen. 2005. "Estimating semiparametric GARCH(∞) model by kernel smoothing methods." *Econometrica* 73(3):771–836.
- Lumsdaine, R. L. 1989. "Asymptotic properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models." *Manuscript, Department of Economics, Harvard University*.
- Nelson, D.B. 1991. "Conditional heteroskedasticity in asset returns: a new approach." *Econometrica* 59:347–370.
- Pagan, A. R., and G. W. Schwert. 1990. "Alternative models for conditional stock volatility." *Journal of Econometrics* 45:267–290.
- Sun, Y., and T. Stengos. 2006. "Semiparametric efficient adaptive estimation of asymmetric GARCH models." *Journal of Econometrics* 133(1):373–386.
- Tapia, R. A., and J. R. Thompson. 1978. *Nonparametric Probability Density Estimation*. Baltimore: Johns Hopkins University Press.

- Weiss, A. A. 1986. "Asymptotic theory for ARCH models: estimation and testing." *Econometric Theory* 2(1):107–131.
- Yang, L. 2006. "Semiparametric GARCH model for foreign exchange volatility." *Journal of Econometrics* 130(2):365–384.
- Zakoian, J. 1994. "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control* 18:931–955.

Table 1: Simulation results: generalized autoregressive score dynamics

This table presents the relative Root Mean Square Error (RMSE) of parameter estimates with respect to QMLE and the relative median Root Mean Square Error of in-sample volatility estimates (RMSE in) and out-of-sample volatility forecasts (RMSE out) with respect to the in-sample RMSE of volatility estimates for the QMLE. Results are based on 100 Monte Carlo replications of time series of 2,000 observations (1,000 in-sample and 1,000 out-of-sample). The underlying data generating process is given by equations (1)–(2′). We consider four distributions: the standard normal (N), Student’s $t(3)$, $t(5)$, and a mixture of normals resembling a $\chi^2(6)$ distribution ($MN(\chi^2(6))$). The statistical models are the based on the true distribution (MLE), the normal distribution (QMLE), the Student’s t distribution with estimated degrees of freedom (QMLE- $t(\nu)$), and on the semiparametric approach (SMLE(1)). The SML(1) approach is initialized using ththe QMLE- $t(\nu)$ estimates.

		μ	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	RMSE in	RMSE out
N	MLE	1.000	1.000	1.000	1.000	1.000	1.025
	QMLE	1.000	1.000	1.000	1.000	1.000	1.025
	QMLE- $t(\nu)$	1.001	0.998	1.104	0.998	1.022	1.033
	SMLE(1)	1.029	1.197	1.434	1.023	1.803	1.994
t(3)	MLE	0.666	0.616	0.361	0.161	0.274	0.295
	QMLE	1.000	1.000	1.000	1.000	1.000	1.147
	QMLE- $t(\nu)$	0.668	1.018	0.357	0.165	0.410	0.411
	SMLE(1)	0.873	1.154	0.400	0.209	0.599	0.648
t(5)	MLE	0.764	0.852	0.448	0.335	0.298	0.294
	QMLE	1.000	1.000	1.000	1.000	1.000	1.021
	QMLE- $t(\nu)$	0.767	0.971	0.462	0.334	0.319	0.327
	SMLE(1)	0.942	1.137	0.499	0.480	0.443	0.519
$MN(\chi^2(6))$	MLE	0.585	0.977	0.744	0.748	0.279	0.274
	QMLE	1.000	1.000	1.000	1.000	1.000	1.014
	QMLE- $t(\nu)$	1.592	1.008	1.612	1.010	0.981	0.986
	SMLE(1)	0.751	1.115	1.367	0.776	0.589	0.587

Table 2: Simulation results: deterministic volatility patterns

This table presents the relative median of Root Mean Square Error of in-sample volatility estimates (RMSE in) and out-of-sample volatility forecasts (RMSE out) with respect to in-sample RMSE of the QMLE estimates over 100 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample). There are four distributions: normal, Student's $t(3)$, $t(5)$, and a mixture of normals $MN(\chi^2(6))$. The initial starting values for SMLE(1) are the QMLE estimates.

		sin		short regime		long regime		burst(1)	
		in	out	in	out	in	out	in	out
N	MLE	1.000	0.986	1.000	1.029	1.000	1.081	1.000	1.001
	QMLE	1.000	0.986	1.000	1.029	1.000	1.081	1.000	1.001
	QMLE- $t(\nu)$	0.981	0.965	0.994	1.009	0.984	1.082	1.010	1.024
	SMLE(1)	0.984	0.978	0.986	1.003	0.924	1.132	1.067	1.100
t(3)	MLE	0.683	0.680	0.586	0.574	0.595	0.714	0.778	0.779
	QMLE	1.000	1.038	1.000	1.046	1.000	1.283	1.000	1.023
	QMLE- $t(\nu)$	0.754	0.767	0.703	0.702	0.702	0.775	0.725	0.831
	SMLE(1)	0.778	0.799	0.670	0.647	0.754	0.979	0.887	0.911
t(5)	MLE	0.796	0.791	0.752	0.738	0.764	0.912	0.856	0.857
	QMLE	1.000	0.984	1.000	0.937	1.000	1.091	1.000	0.969
	QMLE- $t(\nu)$	0.815	0.820	0.776	0.779	0.812	0.955	0.875	0.888
	SMLE(1)	0.820	0.823	0.746	0.764	0.756	0.982	0.901	0.932
$MN(\chi^2(6))$	MLE	0.885	0.874	0.895	0.933	0.896	0.995	0.910	0.916
	QMLE	1.000	0.973	1.000	1.015	1.000	1.092	1.000	1.021
	QMLE- $t(\nu)$	0.990	0.953	0.998	0.998	1.016	1.113	1.049	1.049
	SMLE(1)	0.882	0.868	0.882	0.900	0.856	1.062	0.980	0.999

Table 3: Simulation results: stochastic volatility

This table presents the relative median of the Root Mean Square Error (RMSE) of in-sample volatility estimates (RMSE in) and out-of-sample volatility forecasts (RMSE out) with respect to the GAS model's in-sample RMSE of QMLE. For each density we use 100 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample). The stochastic volatility DGP SV is specified as $y_t \sim p(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.01 + 0.98\alpha_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1^2)$, for $t = 1, \dots, n$. Four distributions are used to generate y_t : normal, Student's $t(3)$, $t(5)$, and a mixture of normals $MN(\chi^2(6))$. The estimation methods are QMLE, QMLE- $t(\nu)$ and SMLE(1) for both GAS and GARCH specifications of the volatility dynamics, see equations (2) and (2'). A one-side test is carried out for the average RMSE for the GAS model being smaller than that of the GARCH specification using the 100 simulations. The p-values are reported. Significance is denoted as *** (1%), ** (5%), and * (10%).

		SV				p-value	
		GAS		GARCH		one-side test	
		in	out	in	out	in	out
N	QMLE	1.000	1.092	1.001	1.078	0.458	0.526
	QMLE- $t(\nu)$	0.982	1.055	0.995	1.078	0.409	0.395
	SMLE(1)	1.002	1.091	1.008	1.083	0.717	0.742
t(3)	QMLE	1.000	1.197	1.033	1.153	0.766	0.996
	QMLE- $t(\nu)$	0.873	0.913	0.996	1.142	0.001***	0.000***
	SMLE(1)	0.953	1.053	1.018	1.142	0.000***	0.002***
t(5)	QMLE	1.000	1.115	0.981	1.058	0.849	0.993
	QMLE- $t(\nu)$	0.900	0.971	0.986	1.071	0.000***	0.000***
	SMLE(1)	0.942	1.006	0.974	1.058	0.011**	0.170
$\chi^2(6)$	QMLE	1.000	1.036	0.995	1.042	0.517	0.634
	QMLE- $t(\nu)$	1.002	1.029	1.026	1.076	0.057*	0.155
	SMLE(1)	0.938	0.974	0.985	1.024	0.029**	0.021**

Table 4: Empirical Estimation Results

Empirical results based on IBM daily stock return series between January 1986 and December 2011. We estimate seven different models: QMLE-norm(al), QMLE- $t(\nu)$, QMLE*- $t(\nu)$, SMLE(1), GARCH, GARCH- $t(\nu)$ and semi(parametric)-GARCH. The model QMLE* - $t(\nu)$ has the same likelihood as QMLE- $t(\nu)$, but the same volatility dynamics as QMLE-normal to isolate the effect of a change in likelihood and a change in volatility dynamics. The starting values for SMLE(1) and semiparametric-GARCH are QMLE- $t(\nu)$ and GARCH- $t(\nu)$, respectively.

	QMLE normal	QMLE* $t(\nu)$	QMLE $t(\nu)$	SMLE(1) semi		GARCH normal	GARCH $t(\nu)$	GARCH semi
μ	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	μ	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)
$\tilde{\omega}$	-8.098 (0.059)	-8.239 (0.112)	-8.182 (0.176)	-7.897 (0.149)	ω	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)
$\tilde{\alpha}$	0.059 (0.007)	0.016 (0.002)	0.131 (0.019)	0.142 (0.010)	α	0.078 (0.009)	0.033 (0.006)	0.046 (0.006)
$\tilde{\beta}$	0.985 (0.002)	0.993 (0.002)	0.994 (0.002)	0.991 (0.002)	$\alpha + \beta$	0.993 (0.002)	0.995 (0.002)	0.994 (0.002)
ν		4.743 (0.300)	5.174 (0.345)		ν		4.992 (0.302)	
log-lik	13241.5	13610.9	13665.6	13671.5		13324.5	13642.5	13644.9

Appendix: additional simulation results

Table A1: Simulation results(QMLE as the initial starting value for SMLE(1))

This table presents the average parameter estimates (Mean), the Root Mean Square Error (RMSE) of parameter estimates and the median of Root Mean Square Error of in-sample volatility estimation (RMSE in) and out-of-sample volatility forecast (RMSE out) over 100 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample) for seven distributions and six different estimation methods. The distributions used are standard normal, a balanced mixture of two standard normals with means 2 and -2, Student's $t(3)$, $t(5)$, $t(8)$, $\chi^2(6)$ and $\chi^2(12)$. The estimation methods are QMLE, QMLE- $t(\nu)$, MLE, SMLE(1)-(3). The initial starting values for SMLE(1) are QMLE- $t(\nu)$ estimates. The boldface numbers in column (μ), (ω), (α), (β) show the model with smaller RMSE between QMLE and SMLE(1). The boldface numbers in the last two columns show the model with smaller RMSE in and RMSE out between QMLE and SMLE(1).

			μ	ω	α	β	RMSE in	RMSE out	
N	QMLE	Mean	0.000	2.000	0.300	0.900			
		RMSE	0.075	0.096	0.046	0.035			
	QMLE- $t(\nu)$	Mean	-0.012	1.997	0.312	0.889	0.109	0.110	
		RMSE	0.075	0.096	0.051	0.035			
	SMLE(1)	Mean	-0.009	2.038	0.326	0.888	0.198	0.214	
		RMSE	0.077	0.115	0.065	0.036			
	SMLE(2)	Mean	-0.010	2.034	0.327	0.887	0.189	0.212	
		RMSE	0.077	0.109	0.065	0.035			
	SMLE(3)	Mean	-0.010	2.033	0.326	0.888	0.192	0.213	
		RMSE	0.077	0.109	0.064	0.035			
	MN	QMLE	Mean	0.000	2.001	0.832	0.894	0.219	0.222
			RMSE	0.053	0.148	0.535	0.019		
QMLE- $t(\nu)$		Mean	0.000	2.003	0.836	0.894	0.223	0.227	
		RMSE	0.053	0.148	0.539	0.019			
MLE		Mean	-0.001	1.996	0.300	0.894	0.079	0.080	
		RMSE	0.030	0.148	0.018	0.018			
SMLE(1)		Mean	0.000	1.969	0.489	0.893	0.184	0.183	
		RMSE	0.058	0.166	0.192	0.018			
SMLE(2)		Mean	0.000	1.969	0.475	0.894	0.181	0.182	
		RMSE	0.059	0.166	0.179	0.018			
SMLE(3)		Mean	0.000	1.971	0.476	0.894	0.181	0.182	
		RMSE	0.059	0.172	0.179	0.018			
t(3)	QMLE	Mean	-0.003	1.966	0.094	0.632	0.553	0.649	
		RMSE	0.094	0.227	0.278	0.482			
	QMLE- $t(\nu)$	Mean	-0.003	1.994	0.299	0.877	0.221	0.224	
		RMSE	0.062	0.207	0.095	0.073			
	MLE	Mean	-0.003	1.994	0.300	0.878	0.147	0.157	
		RMSE	0.062	0.121	0.096	0.070			
	SMLE(1)	Mean	-0.002	1.995	0.245	0.739	0.366	0.381	
		RMSE	0.084	0.284	0.192	0.406			
	SMLE(2)	Mean	-0.005	2.035	0.259	0.738	0.336	0.370	
		RMSE	0.082	0.303	0.181	0.431			
	SMLE(3)	Mean	-0.006	2.042	0.271	0.741	0.345	0.367	
		RMSE	0.083	0.320	0.170	0.429			
t(5)	QMLE	Mean	-0.011	2.013	0.153	0.833	0.440	0.450	
		RMSE	0.098	0.131	0.158	0.167			

Table A1: Simulation results(QMLE as the initial starting value for SMLE(1))(continue)

			μ	ω	α	β	RMSE in	RMSE out	
			0.000	2.000	0.300	0.900			
t(5)	QMLE- $t(\nu)$	Mean	-0.008	1.998	0.302	0.882	0.140	0.144	
		RMSE	0.075	0.127	0.073	0.056			
	MLE	Mean	-0.007	1.996	0.302	0.882	0.131	0.130	
		RMSE	0.075	0.111	0.071	0.056			
	SMLE(1)	Mean	-0.005	2.035	0.282	0.856	0.218	0.241	
		RMSE	0.091	0.148	0.091	0.151			
	SMLE(2)	Mean	-0.006	2.044	0.288	0.855	0.199	0.239	
		RMSE	0.092	0.151	0.083	0.160			
	SMLE(3)	Mean	-0.006	2.044	0.290	0.855	0.195	0.240	
		RMSE	0.092	0.154	0.083	0.161			
t(8)	QMLE	Mean	-0.006	1.989	0.191	0.887	0.285	0.320	
		RMSE	0.082	0.120	0.120	0.049			
	QMLE- $t(\nu)$	Mean	-0.008	1.981	0.290	0.896	0.129	0.124	
		RMSE	0.079	0.117	0.061	0.037			
	MLE	Mean	-0.008	1.981	0.291	0.895	0.117	0.115	
		RMSE	0.079	0.110	0.062	0.038			
	SMLE(1)	Mean	-0.006	2.021	0.281	0.893	0.191	0.213	
		RMSE	0.080	0.125	0.065	0.041			
	SMLE(2)	Mean	-0.005	2.023	0.283	0.892	0.192	0.217	
		RMSE	0.080	0.127	0.062	0.041			
	SMLE(3)	Mean	-0.005	2.023	0.284	0.892	0.189	0.217	
		RMSE	0.081	0.129	0.062	0.041			
	$\chi^2(6)$	QMLE	Mean	0.085	2.007	0.275	0.886	0.416	0.421
			RMSE	0.123	0.134	0.054	0.042		
		QMLE- $t(\nu)$	Mean	-0.142	2.020	0.352	0.888	0.408	0.410
			RMSE	0.196	0.135	0.087	0.043		
MLE		Mean	-0.013	1.981	0.301	0.894	0.116	0.114	
		RMSE	0.072	0.131	0.040	0.032			
SMLE(1)		Mean	-0.011	1.968	0.348	0.893	0.246	0.247	
		RMSE	0.092	0.153	0.073	0.032			
SMLE(2)		Mean	0.012	1.984	0.344	0.892	0.231	0.237	
		RMSE	0.089	0.149	0.069	0.033			
SMLE(3)		Mean	0.012	1.982	0.343	0.892	0.230	0.238	
		RMSE	0.089	0.146	0.067	0.032			
$\chi^2(12)$		QMLE	Mean	0.062	2.000	0.280	0.886	0.342	0.345
			RMSE	0.106	0.130	0.054	0.042		
		QMLE- $t(\nu)$	Mean	-0.058	2.002	0.333	0.887	0.332	0.334
			RMSE	0.123	0.130	0.074	0.043		
	MLE	Mean	-0.011	1.985	0.303	0.892	0.120	0.113	
		RMSE	0.080	0.128	0.044	0.035			
	SMLE(1)	Mean	-0.006	1.999	0.338	0.891	0.228	0.228	
		RMSE	0.089	0.145	0.072	0.036			
	SMLE(2)	Mean	0.006	2.005	0.336	0.890	0.221	0.230	
		RMSE	0.088	0.149	0.071	0.036			
	SMLE(3)	Mean	0.006	2.004	0.336	0.890	0.220	0.224	
		RMSE	0.087	0.143	0.069	0.036			

Table A2: Simulation results(QMLE- $t(\nu)$ as the initial starting value for SMLE(1))

This table presents the average parameter estimates (Mean), the Root Mean Square Error (RMSE) of parameter estimates and the median of Root Mean Square Error of in-sample volatility estimation (RMSE in) and out-of-sample volatility forecast (RMSE out) over 100 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample) for seven distributions and six different estimation methods. The distributions used are standard normal, a balanced mixture of two standard normals with means 2 and -2, Student's $t(3)$, $t(5)$, $t(8)$, $\chi^2(6)$ and $\chi^2(12)$. The estimation methods are QMLE, QMLE- $t(\nu)$, MLE, SMLE(1)-(3). The initial starting values for SMLE(1) are QMLE- $t(\nu)$ estimates. The boldface numbers in column (μ), (ω), (α), (β) show the model with smaller RMSE between QMLE and SMLE(1). The boldface numbers in the last two columns show the model with smaller RMSE in and RMSE out between QMLE and SMLE(1).

			μ	ω	α	β	RMSE in	RMSE out	
N	QMLE	Mean	0.000	2.000	0.300	0.900			
		RMSE	0.075	0.096	0.046	0.035			
	QMLE- $t(\nu)$	Mean	-0.012	1.997	0.312	0.889	0.109	0.110	
		RMSE	0.075	0.096	0.051	0.035			
	SMLE(1)	Mean	-0.009	2.038	0.326	0.888	0.198	0.214	
		RMSE	0.077	0.115	0.066	0.036			
	SMLE(2)	Mean	-0.010	2.034	0.327	0.887	0.189	0.212	
		RMSE	0.077	0.109	0.065	0.035			
	SMLE(3)	Mean	-0.010	2.033	0.326	0.888	0.192	0.213	
		RMSE	0.077	0.109	0.064	0.035			
	MN	QMLE	Mean	0.000	2.001	0.832	0.894	0.219	0.222
			RMSE	0.053	0.148	0.535	0.019		
QMLE- $t(\nu)$		Mean	0.000	2.003	0.836	0.894	0.223	0.227	
		RMSE	0.053	0.148	0.539	0.019			
MLE		Mean	-0.001	1.996	0.300	0.894	0.079	0.080	
		RMSE	0.030	0.148	0.018	0.018			
SMLE(1)		Mean	0.000	1.969	0.490	0.893	0.184	0.183	
		RMSE	0.058	0.167	0.193	0.018			
SMLE(2)		Mean	0.000	1.969	0.475	0.894	0.181	0.182	
		RMSE	0.059	0.166	0.179	0.018			
SMLE(3)		Mean	0.000	1.972	0.476	0.894	0.181	0.182	
		RMSE	0.059	0.172	0.179	0.018			
t(3)	QMLE	Mean	-0.003	1.966	0.094	0.632	0.553	0.629	
		RMSE	0.094	0.227	0.278	0.482			
	QMLE- $t(\nu)$	Mean	-0.003	1.994	0.299	0.877	0.221	0.224	
		RMSE	0.062	0.207	0.095	0.073			
	MLE	Mean	-0.003	1.994	0.300	0.878	0.147	0.157	
		RMSE	0.062	0.121	0.096	0.070			
	SMLE(1)	Mean	-0.006	1.999	0.263	0.868	0.326	0.349	
		RMSE	0.081	0.244	0.110	0.097			
	SMLE(2)	Mean	-0.008	2.035	0.264	0.867	0.315	0.344	
		RMSE	0.079	0.283	0.116	0.104			
	SMLE(3)	Mean	-0.008	2.048	0.268	0.867	0.320	0.342	
		RMSE	0.081	0.310	0.122	0.105			
t(5)	QMLE	Mean	-0.011	2.013	0.153	0.833	0.440	0.450	
		RMSE	0.098	0.131	0.158	0.167			

Table A2: Simulation results(QMLE- $t(\nu)$ as the initial starting value for SMLE(1))(continue)

			μ	ω	α	β	RMSE in	RMSE out	
			0.000	2.000	0.300	0.900			
t(5)	QMLE- $t(\nu)$	Mean	-0.008	1.998	0.302	0.882	0.140	0.144	
		RMSE	0.075	0.127	0.073	0.056			
	MLE	Mean	-0.007	1.996	0.302	0.882	0.131	0.130	
		RMSE	0.075	0.111	0.071	0.056			
	SMLE(1)	Mean	-0.007	2.040	0.291	0.874	0.195	0.229	
		RMSE	0.092	0.149	0.079	0.080			
	SMLE(2)	Mean	-0.007	2.044	0.292	0.871	0.196	0.233	
		RMSE	0.091	0.153	0.081	0.088			
	SMLE(3)	Mean	-0.007	2.044	0.293	0.871	0.197	0.235	
		RMSE	0.091	0.155	0.082	0.091			
	t(8)	QMLE	Mean	-0.006	1.989	0.191	0.887	0.285	0.320
			RMSE	0.082	0.120	0.120	0.049		
QMLE- $t(\nu)$		Mean	-0.008	1.981	0.290	0.896	0.129	0.124	
		RMSE	0.079	0.117	0.061	0.037			
MLE		Mean	-0.008	1.981	0.291	0.895	0.117	0.115	
		RMSE	0.079	0.110	0.062	0.038			
SMLE(1)		Mean	-0.005	2.021	0.285	0.892	0.188	0.214	
		RMSE	0.080	0.124	0.062	0.041			
SMLE(2)		Mean	-0.005	2.023	0.285	0.892	0.190	0.217	
		RMSE	0.080	0.128	0.062	0.041			
SMLE(3)		Mean	-0.005	2.024	0.285	0.892	0.190	0.217	
		RMSE	0.081	0.128	0.062	0.041			
$\chi^2(6)$		QMLE	Mean	0.085	2.007	0.275	0.886	0.416	0.421
			RMSE	0.123	0.134	0.054	0.042		
		QMLE- $t(\nu)$	Mean	-0.142	2.020	0.352	0.888	0.408	0.410
			RMSE	0.196	0.135	0.087	0.043		
		MLE	Mean	-0.013	1.981	0.301	0.894	0.116	0.114
			RMSE	0.072	0.131	0.040	0.032		
	SMLE(1)	Mean	-0.013	1.967	0.351	0.893	0.245	0.244	
		RMSE	0.092	0.150	0.074	0.033			
	SMLE(2)	Mean	0.007	1.977	0.343	0.893	0.232	0.237	
		RMSE	0.089	0.143	0.067	0.032			
	SMLE(3)	Mean	0.008	1.977	0.342	0.893	0.229	0.239	
		RMSE	0.090	0.143	0.066	0.032			
	$\chi^2(6)$	QMLE	Mean	0.062	2.000	0.280	0.886	0.342	0.345
			RMSE	0.106	0.130	0.054	0.042		
		QMLE- $t(\nu)$	Mean	-0.058	2.002	0.333	0.887	0.332	0.334
			RMSE	0.123	0.130	0.074	0.043		
		MLE	Mean	-0.011	1.985	0.303	0.892	0.120	0.113
			RMSE	0.080	0.128	0.044	0.035		
SMLE(1)		Mean	-0.008	1.997	0.340	0.892	0.225	0.238	
		RMSE	0.091	0.142	0.072	0.037			
SMLE(2)		Mean	0.004	2.001	0.337	0.891	0.221	0.230	
		RMSE	0.088	0.139	0.070	0.036			
SMLE(3)		Mean	0.004	2.002	0.337	0.891	0.220	0.230	
		RMSE	0.087	0.139	0.071	0.036			

Table A3: Simulation results: deterministic volatility patterns

This table presents the median of Root Mean Square Error of in-sample volatility estimation (RMSE in) and out-of-sample volatility forecast (RMSE out) over 100 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample) for seven distributions and six different estimation methods. The distributions used are normal, a balanced mixture of two standard normals with means 2 and -2, Student's $t(3)$, $t(5)$, $t(8)$, $\chi^2(6)$ and $\chi^2(12)$. The estimation methods are QMLE, QMLE- $t(\nu)$, MLE, and SMLE(1)-(3). The initial starting values for SMLE(1) are QMLE estimates.

		Sin		short regime		long regime	
RMSE		in	out	in	out	in	out
N	MLE(QMLE)	0.437	0.431	0.538	0.554	0.396	0.429
	QMLE- $t(\nu)$	0.429	0.422	0.535	0.543	0.390	0.429
	SMLE(1)	0.430	0.428	0.530	0.540	0.366	0.449
	SMLE(2)	0.429	0.427	0.530	0.542	0.366	0.449
	SMLE(3)	0.430	0.427	0.530	0.541	0.366	0.449
MN	QMLE	0.328	0.318	0.406	0.422	0.298	0.329
	QMLE- $t(\nu)$	0.329	0.318	0.408	0.422	0.298	0.331
	MLE	0.334	0.327	0.421	0.453	0.317	0.346
	SMLE(1)	0.321	0.315	0.413	0.425	0.296	0.357
	SMLE(2)	0.321	0.315	0.412	0.425	0.296	0.357
	SMLE(3)	0.321	0.315	0.412	0.425	0.296	0.357
t(3)	QMLE	0.721	0.749	0.996	1.042	0.703	0.902
	QMLE- $t(\nu)$	0.544	0.553	0.700	0.699	0.494	0.545
	MLE	0.493	0.491	0.584	0.571	0.418	0.502
	SMLE(1)	0.561	0.577	0.668	0.644	0.530	0.688
	SMLE(2)	0.553	0.565	0.658	0.635	0.499	0.682
	SMLE(3)	0.555	0.571	0.649	0.624	0.496	0.671
t(5)	QMLE	0.588	0.578	0.763	0.715	0.525	0.572
	QMLE- $t(\nu)$	0.479	0.482	0.592	0.595	0.426	0.501
	MLE	0.468	0.465	0.574	0.564	0.401	0.479
	SMLE(1)	0.482	0.484	0.569	0.583	0.397	0.515
	SMLE(2)	0.483	0.482	0.571	0.578	0.388	0.516
	SMLE(3)	0.484	0.481	0.574	0.577	0.390	0.515
t(8)	QMLE	0.514	0.514	0.638	0.631	0.462	0.497
	QMLE- $t(\nu)$	0.472	0.455	0.561	0.566	0.409	0.468
	MLE	0.462	0.449	0.547	0.541	0.398	0.458
	SMLE(1)	0.469	0.464	0.555	0.560	0.371	0.476
	SMLE(2)	0.470	0.463	0.553	0.558	0.374	0.477
	SMLE(3)	0.469	0.464	0.555	0.560	0.374	0.478
$\chi^2(6)$	QMLE	0.471	0.458	0.570	0.578	0.418	0.456
	QMLE- $t(\nu)$	0.466	0.448	0.569	0.571	0.424	0.465
	MLE	0.416	0.411	0.510	0.532	0.375	0.416
	SMLE(1)	0.415	0.409	0.503	0.513	0.358	0.444
	SMLE(2)	0.415	0.407	0.501	0.512	0.358	0.439
	SMLE(3)	0.415	0.407	0.500	0.512	0.358	0.440
$\chi^2(12)$	QMLE	0.460	0.451	0.561	0.578	0.408	0.448
	QMLE- $t(\nu)$	0.449	0.436	0.551	0.555	0.401	0.452
	MLE	0.456	0.450	0.562	0.585	0.431	0.466
	SMLE(1)	0.425	0.414	0.511	0.531	0.357	0.451
	SMLE(2)	0.422	0.416	0.509	0.531	0.359	0.451
	SMLE(3)	0.421	0.416	0.509	0.532	0.359	0.451

Table A3: Simulation results: deterministic volatility patterns(continue)

		burst(1)		burst(2)		burst(3)	
RMSE		in	out	in	out	in	out
N	MLE(QMLE)	0.376	0.377	0.353	0.355	0.355	0.352
	QMLE- $t(\nu)$	0.380	0.385	0.358	0.359	0.358	0.349
	SMLE(1)	0.401	0.414	0.376	0.378	0.366	0.365
	SMLE(2)	0.401	0.413	0.374	0.377	0.365	0.364
	SMLE(3)	0.400	0.413	0.375	0.378	0.366	0.364
MN	QMLE	0.293	0.289	0.273	0.269	0.264	0.264
	QMLE- $t(\nu)$	0.294	0.290	0.275	0.270	0.265	0.264
	MLE	0.286	0.284	0.266	0.266	0.260	0.262
	SMLE(1)	0.307	0.313	0.280	0.278	0.272	0.270
	SMLE(2)	0.307	0.313	0.280	0.277	0.272	0.270
	SMLE(3)	0.307	0.313	0.280	0.277	0.272	0.270
t(3)	QMLE	0.592	0.606	0.588	0.603	0.588	0.627
	QMLE- $t(\nu)$	0.429	0.492	0.476	0.472	0.467	0.462
	MLE	0.461	0.461	0.433	0.437	0.424	0.425
	SMLE(1)	0.525	0.539	0.508	0.534	0.500	0.517
	SMLE(2)	0.520	0.534	0.490	0.519	0.483	0.507
	SMLE(3)	0.513	0.523	0.486	0.513	0.482	0.501
t(5)	QMLE	0.505	0.490	0.485	0.475	0.485	0.465
	QMLE- $t(\nu)$	0.442	0.449	0.416	0.423	0.407	0.412
	MLE	0.433	0.433	0.408	0.411	0.404	0.405
	SMLE(1)	0.455	0.471	0.428	0.441	0.428	0.432
	SMLE(2)	0.450	0.471	0.426	0.440	0.422	0.430
	SMLE(3)	0.450	0.470	0.428	0.438	0.420	0.429
t(8)	QMLE	0.442	0.442	0.422	0.426	0.414	0.413
	QMLE- $t(\nu)$	0.422	0.417	0.401	0.386	0.388	0.381
	MLE	0.417	0.412	0.397	0.384	0.386	0.375
	SMLE(1)	0.439	0.439	0.411	0.410	0.397	0.400
	SMLE(2)	0.434	0.440	0.410	0.409	0.397	0.398
	SMLE(3)	0.434	0.439	0.409	0.409	0.398	0.398
$\chi^2(6)$	QMLE	0.400	0.408	0.380	0.381	0.375	0.377
	QMLE- $t(\nu)$	0.419	0.419	0.390	0.391	0.384	0.378
	MLE	0.363	0.366	0.344	0.343	0.336	0.339
	SMLE(1)	0.392	0.399	0.362	0.364	0.356	0.358
	SMLE(2)	0.390	0.396	0.361	0.363	0.354	0.356
	SMLE(3)	0.391	0.399	0.359	0.362	0.353	0.356
$\chi^2(12)$	QMLE	0.395	0.404	0.374	0.375	0.368	0.372
	QMLE- $t(\nu)$	0.406	0.404	0.381	0.378	0.369	0.369
	MLE	0.389	0.394	0.370	0.368	0.357	0.366
	SMLE(1)	0.398	0.409	0.368	0.371	0.357	0.361
	SMLE(2)	0.396	0.407	0.369	0.369	0.357	0.361
	SMLE(3)	0.396	0.407	0.369	0.368	0.357	0.360

Table A4: Simulation results: stochastic volatility

This table presents the median of Root Mean Square Error of in-sample volatility estimation (RMSE in) and out-of-sample volatility forecast (RMSE out) over 100 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample) for two stochastic volatility models. The first stochastic volatility DGP SV(1) is specified as: $y_t \sim p(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.01 + 0.98\alpha_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1^2)$, for $t = 1, \dots, n$. The second stochastic volatility DGP SV(2) is specified as: $y_t \sim p(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.05 + 0.9\alpha_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1^2)$, for $t = 1, \dots, n$. For both models, seven distributions are used to generate y_t . They are normal, a balanced mixture of two standard normals with means 2 and -2, Student's $t(3)$, $t(5)$, $t(8)$, $\chi^2(6)$ and $\chi^2(12)$. The estimation methods are QMLE, QMLE- $t(\nu)$ and SMLE(1)-(3) for both GAS and GARCH models, respectively. The initial starting values for SMLE(1) are QMLE estimates. A one-side test is carried out and p-values are given. The H_0 is $RMSE_{GAS} \geq RMSE_{GARCH}$. *** means the null is rejected at 95% significance level; and ** means the null is rejected at 90% significance level.

		SV(1)				p-value	
		GAS		GARCH		one-side test	
		in	out	in	out	in	out
N	QMLE	0.221	0.241	0.221	0.238	0.458	0.526
	QMLE- $t(\nu)$	0.217	0.233	0.220	0.238	0.409	0.395
	SMLE(1)	0.221	0.241	0.223	0.239	0.717	0.742
	SMLE(2)	0.221	0.241	0.223	0.240	0.712	0.748
	SMLE(3)	0.221	0.242	0.223	0.240	0.710	0.749
MN	QMLE	0.184	0.190	0.187	0.190	0.206	0.531
	QMLE- $t(\nu)$	0.185	0.188	0.187	0.190	0.431	0.520
	SMLE(1)	0.185	0.189	0.187	0.192	0.233	0.700
	SMLE(2)	0.186	0.189	0.187	0.192	0.231	0.697
	SMLE(3)	0.186	0.189	0.187	0.192	0.231	0.697
$t(3)$	QMLE	0.310	0.371	0.320	0.357	0.766	0.996
	QMLE- $t(\nu)$	0.271	0.283	0.309	0.354	0.001***	0.000***
	SMLE(1)	0.295	0.326	0.315	0.354	0.000***	0.002***
	SMLE(2)	0.288	0.318	0.314	0.352	0.000***	0.002***
	SMLE(3)	0.289	0.318	0.313	0.352	0.000***	0.001***
$t(5)$	QMLE	0.265	0.296	0.260	0.280	0.849	0.993
	QMLE- $t(\nu)$	0.239	0.257	0.261	0.284	0.000***	0.000***
	SMLE(1)	0.250	0.266	0.258	0.280	0.011***	0.170
	SMLE(2)	0.250	0.268	0.258	0.280	0.009***	0.171
	SMLE(3)	0.250	0.266	0.258	0.280	0.008***	0.160
$t(8)$	QMLE	0.247	0.267	0.246	0.259	0.732	0.975
	QMLE- $t(\nu)$	0.234	0.247	0.237	0.255	0.059**	0.031***
	SMLE(1)	0.234	0.259	0.246	0.260	0.018***	0.263
	SMLE(2)	0.235	0.258	0.246	0.260	0.016***	0.246
	SMLE(3)	0.235	0.258	0.246	0.260	0.017***	0.245
$\chi^2(6)$	QMLE	0.237	0.246	0.236	0.247	0.517	0.634
	QMLE- $t(\nu)$	0.238	0.244	0.243	0.255	0.057**	0.155
	SMLE(1)	0.223	0.231	0.234	0.243	0.029***	0.021***
	SMLE(2)	0.223	0.232	0.234	0.243	0.016***	0.016***
	SMLE(3)	0.223	0.232	0.234	0.243	0.017***	0.015***
$\chi^2(12)$	QMLE	0.233	0.242	0.231	0.244	0.557	0.646
	QMLE- $t(\nu)$	0.229	0.238	0.235	0.247	0.139	0.189
	SMLE(1)	0.225	0.234	0.232	0.240	0.198	0.145
	SMLE(2)	0.224	0.234	0.232	0.240	0.156	0.129
	SMLE(3)	0.224	0.234	0.232	0.241	0.157	0.127

Table A4: Simulation results: stochastic volatility(continue)

		SV(2)				p-value		
		GAS		GARCH		one-side test		
		RMSE	in	out	in	out	in	out
N	QMLE	0.144	0.150	0.144	0.150	0.650	0.407	
	QMLE- $t(v)$	0.145	0.150	0.147	0.150	0.258	0.484	
	SMLE(1)	0.144	0.148	0.145	0.149	0.739	0.461	
	SMLE(2)	0.144	0.148	0.145	0.149	0.748	0.464	
	SMLE(3)	0.144	0.148	0.145	0.149	0.748	0.464	
MN	QMLE	0.134	0.137	0.135	0.138	0.210	0.614	
	QMLE- $t(v)$	0.136	0.137	0.136	0.137	0.722	0.638	
	SMLE(1)	0.139	0.141	0.140	0.139	0.225	0.775	
	SMLE(2)	0.138	0.141	0.139	0.139	0.225	0.772	
	SMLE(3)	0.138	0.141	0.139	0.139	0.225	0.772	
$t(3)$	QMLE	0.180	0.193	0.198	0.206	0.843	0.856	
	QMLE- $t(v)$	0.168	0.171	0.174	0.181	0.093**	0.033***	
	SMLE(1)	0.177	0.189	0.208	0.205	0.004***	0.021***	
	SMLE(2)	0.175	0.186	0.207	0.200	0.001***	0.004***	
	SMLE(3)	0.176	0.186	0.207	0.200	0.003***	0.016***	
$t(5)$	QMLE	0.154	0.160	0.161	0.165	0.265	0.095**	
	QMLE- $t(v)$	0.153	0.159	0.154	0.161	0.112	0.076**	
	SMLE(1)	0.152	0.157	0.159	0.165	0.008***	0.019***	
	SMLE(2)	0.153	0.158	0.158	0.165	0.011***	0.031***	
	SMLE(3)	0.153	0.158	0.158	0.165	0.013***	0.040***	
$t(8)$	QMLE	0.151	0.157	0.149	0.155	0.699	0.868	
	QMLE- $t(v)$	0.150	0.156	0.152	0.153	0.664	0.708	
	SMLE(1)	0.151	0.156	0.150	0.154	0.407	0.428	
	SMLE(2)	0.151	0.156	0.151	0.155	0.431	0.444	
	SMLE(3)	0.151	0.156	0.151	0.155	0.442	0.448	
$\chi^2(6)$	QMLE	0.148	0.151	0.147	0.150	0.394	0.645	
	QMLE- $t(v)$	0.153	0.155	0.151	0.152	0.661	0.606	
	SMLE(1)	0.150	0.150	0.151	0.151	0.288	0.504	
	SMLE(2)	0.150	0.150	0.151	0.150	0.298	0.516	
	SMLE(3)	0.150	0.150	0.151	0.150	0.301	0.518	
$\chi^2(12)$	QMLE	0.147	0.151	0.147	0.149	0.435	0.666	
	QMLE- $t(v)$	0.152	0.153	0.146	0.150	0.770	0.798	
	SMLE(1)	0.147	0.149	0.147	0.148	0.338	0.550	
	SMLE(2)	0.147	0.149	0.147	0.148	0.348	0.563	
	SMLE(3)	0.146	0.149	0.147	0.148	0.348	0.564	

Table A5: Empirical Estimation Results

Empirical results based on IBM daily stock return series between January 1986 and December 2011. We estimate the new model by eleven different estimation methods, QMLE-normal, EGARCH- $t(\nu)$, QMLE- $t(\nu)$, SMLE(1), SMLE(2), SMLE(3), GARCH, GARCH- $t(\nu)$, Semi-GARCH, EGARCH, Semi-EGARCH. The initial starting values for SMLE(1) are QMLE- $t(\nu)$ estimates. The initial starting values for Semi-GARCH are GARCH- $t(\nu)$ estimates. The initial starting values for Semi-EGARCH are EGARCH- $t(\nu)$ estimates.

	QMLE-normal	EGARCH- $t(\nu)$	QMLE- $t(\nu)$	SMLE(1)	SMLE(2)	SMLE(3)	GARCH
μ	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)
ω	-8.098 (0.059)	-8.239 (0.112)	-8.182 (0.176)	-7.897 (0.149)	-7.878 (0.144)	-7.836 (0.150)	0.001 (0.000)
α	0.059 (0.007)	0.016 (0.002)	0.131 (0.019)	0.142 (0.010)	0.169 (0.013)	0.172 (0.016)	0.078 (0.009)
β	0.985 (0.002)	0.993 (0.002)	0.994 (0.002)	0.991 (0.002)	0.988 (0.003)	0.989 (0.003)	0.993 (0.002)
ν		4.743 (0.300)	5.174 (0.345)				
log-lik	13241.5	13610.9	13665.6	13671.5	13669.5	13666.2	13324.5
	GARCH- $t(\nu)$	Semi-GARCH	EGARCH	Semi-EGARCH			
μ	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)			
ω	0.000 (0.000)	0.000 (0.000)	-8.098 (0.059)	-8.173 (0.063)			
α	0.033 (0.006)	0.046 (0.006)	0.030 (0.003)	0.020 (0.002)			
β	0.995 (0.002)	0.994 (0.002)	0.985 (0.002)	0.990 (0.002)			
ν	4.992 (0.302)						
log-lik	13642.5	13644.9	13241.5	13609.0			