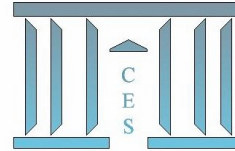# Documents de Travail du Centre d'Economie de la Sorbonne

CES Working Papers

Likelihood-Related Estimation Methods and Non-Gaussian GARCH Processes

Christophe CHORRO, Dominique GUEGAN, Florian IELPO

2010.67

Likelihood-Related Estimation Methods and Non-Gaussian GARCH Processes

Christophe Chorro[*]   Dominique Guégan[†]   Florian Ielpo[‡]

July 22, 2010

Abstract

This article discusses the finite distance properties of three likelihood-based estimation strategies for GARCH processes with non-Gaussian conditional distributions: (1) the maximum likelihood approach; (2) the Quasi Maximum Likelihood approach; (3) a multi-steps recursive estimation approach (REC). We first run a Monte Carlo test which shows that the recursive method may be the most relevant approach for estimation purposes. We then turn to a sample of SP500 returns. We confirm that the REC estimates are statistically dominating the parameters estimated by the two other competing methods. Regardless of the selected model, REC estimates deliver the more stable results.

Keywords: Maximum likelihood method, related-GARCH process, Recursive estimation method, Mixture of Gaussian distributions, Generalized hyperbolic distributions, SP500.

JEL classification: G13, C22.

---

[*]Centre d'Economie de la Sorbonne, Maison des Sciences Economiques, 106 bd de l'Hôpital, 75013 Paris, France. Email: cchorro@univ-paris1.fr.

[†]Corresponding author. PSE, Centre d'Economie de la Sorbonne, University Paris 1 Panthéon-Sorbonne, MSE, 106 bd de l'Hôpital, 75013 Paris, France. Email: dguegan@univ-paris1.fr.. Tel: +33 1 40 07 82 98.

[‡]Pictet Asset Management, Route des Acacias 60, 1211 Geneva, Switzerland and Centre d'Economie de la Sorbonne, Maison des Sciences Economiques, 106 bd de l'Hôpital, 75013 Paris, France. Email: florian.ielpo@ensae.org.

1 Introduction

Models mixing time varying volatility and non-Gaussian conditional distributions turned out to be a successful empirical approach to compute the price of contingent claims such as options. Duan (1995), Christoffersen *et al.* (2006), Badescu *et al.* (2008), and Chorro *et al.* (2010) provided insight in this direction. They have the common approach to be primarily interested in the historical dynamics of financial asset prices, from which the celebrated "risk-neutral" distribution is inferred. Hence the estimation of the parameters of such complicated processes from the times series of returns stand a good chance to have serious implications when it comes to the quality of the computed option prices. The design of a finite distance estimation strategy to obtain a precise measure of the true value of the parameters involved by these models is important. It can reveal however tricky as the parameters driving the volatility process and the conditional distribution often seem to play a similar role. For example, when dealing with Nelson (1991)'s EGARCH with a conditional distribution that is a Generalized Hyperbolic distribution, both the volatility structure and the conditional distribution have a parameter that explicitly controls the amount of skewness in the returns. Even though they have a similar statistical role, the associated interpretation is not the same. The EGARCH parameter controlling the skewness aims at modeling the volatility to returns spill-overs, also called "leverage effects" in equity markets. On the other hand, the parameter controlling the skewness in the GH distribution is used to model the structural tendency of financial markets to have negative returns of a larger scale than positive ones over the past hundred years. When estimating jointly these two sets of parameters, the mixing effects jeopardize the estimation results – and the interpretation obtained from these figures. Similar remarks can be made concerning the amount of kurtosis in the unconditional distribution of the returns. In this paper, we present Monte Carlo evidence that a recursive estimation method proves to behave fairly well when it comes to the estimation of these two distinct sets of parameters. This approach especially dominates the plain maximum likelihood in the case of volatility parameters estimation. A similar conclusion is reached when using this recursive and maximum likelihood estimation methods with a real data set of SP500 returns.

The estimation theory for the parameters of GARCH-like processes has been a lot developed and discussed. There exist theoretical results for maximum likelihood methods proving that the estimated parameters are asymptotically Gaussian; similar results for quasi-maximum likelihood methods are available as well as for recursive ones. All methods provide consistent estimates under specified conditions, and generally these conditions to get the theoretical convergence of the estimated parameters – whatever the form under which they are developed – are verified for nearly all the classical models and distributions used in practice. On this point, see Gourieroux *et al.* (1993), Lumsdaine (1996), Chernov and Ghysels (2000), Straumann (2005), and Fan *et al.* (2007). Not withstanding the fact these theoretical results are essential, finite distance estimates are known to depart from their asymptotic counterparts.

Our aim is to review and discuss the methods available to estimate jointly the volatility and the conditional distribution parameters, given the difficulties driven by these potential mixing effects between both sets of parameters. We investigate three different approaches: (1) the one step Maximum Likelihood strategy (ML), that estimates both the volatility and conditional density parameters in the meantime. (2) The Quasi Maximum Likelihood strategy (QML) that estimates separately these two groups of parameters. (3) A recursive approach (REC) that starts from the QML estimates and that iteratively maximizes the likelihood over either the volatility or the distribution parameters.

First, focusing on likelihood related approaches and assuming that the distribution of the residuals is known, we provide Monte Carlo experiments comparing them on an equal foot. We investigate the quality of the estimated parameters using three different methods of estimation, assuming that the models correspond to an EGARCH or an APARCH model (Nelson 1991, Ding and Granger, 1996), with a mixture of two Gaussian distributions (MN), Kon (1984), or Generalized Hyperbolic (GH) distributions for the residuals, Barndorff-Nielsen and Blaesid (1981). We restrict to EGARCH and APARCH models as they are consistent with time varying volatility, leverage effect and non linear effects usually found in the descriptive statistics performed

3

over financial returns. In the meantime, they contain other GARCH models as special cases. The GH distribution also contains the Gaussian distribution, the Student-t distribution or the NIG distribution amongst of others as special cases. The MN distribution contains the same number of parameters while providing a similar distributional fit. Instead of solely focusing on the quality of the overall fit, we distinguish the quality of the estimation of the volatility and of the distribution parameters. We also rise and answer questions concerning the convergence of the optimization algorithm used in these estimation strategies, an aspect often left apart in the literature. We show that the three estimation strategies provide convergence rates that are different. Finally, we also provide estimates for the time consumption of each method, showing that there is a trade-off between the quality of the estimates in terms of root mean square error and time required to perform the estimation. With the recursive algorithm used in this article, we obtain parameters closer to their true values that those yielded by the ML and QML methods – especially in the case of volatility – at the price of a lower convergence rate and a longer time to estimate these parameters. As asymptotic convergence rates of the different estimation strategies proposed here are the same, our focus is explicitly on the finite-distance behavior of estimates.

Second, we use the three methods to estimate the parameters of such time series models in the case of a real data set of financial returns. We focused on the four mixed models between EGARCH/APARCH volatility structures and the GH/MN distributions on a sample of returns on the SP500 index from 1998 to 2003. For a given model, we compare the quality of the joint distribution across the three types of estimates using a likelihood-based test. For all our cases, the pattern is the same: the ML estimates dominate the QML ones, but are outperformed by the recursive ones. The recursive estimates have the interesting properties to yield volatility news response functions (Engle and Ng, 1993) that are stable across the time series models, even with different estimated volatility structures. The same conclusion of stability arises when investigating the shape of the estimated conditional distributions for this particular sample. This is however not the case of the ML estimates, as they suffer from the previously mentioned mixing

4

effects between parameters having the same impact on the joint distribution of returns. Then, considering that the REC parameters are the best given our sample, we rank the different models in terms of goodness of fit. When no dominant model is found, a similar ranking based on ML parameters would have yield a clearly different and thus misleading result.

In the end, from both the Monte Carlo and real data set exercises, the conclusion seems to be that the choice of the estimation method may be key for two reasons: (1) There is a trade-off between the convergence rate, the estimation time and the quality of the fit obtained that clearly depends on the selected strategy; (2) From an economic point of view, the convergence of the estimation algorithm gives no guaranty at all that the model captures the right phenomena that drive the financial returns.

The paper is organized as follows: Section 2 presents and develops our methodology. Section 3 is devoted to the analysis of the results of the simulations. Section 4 presents an application of the three estimation approaches on a real data set. Section 5 concludes.

2   Methodology and strategy

We assume that the dynamics of a stock price process $(S_t)_{t \in \{0,1,...,T\}}$ are characterized by the stationary process $(Y_t)_{t \in \{1,...,T\}}$ defined by:

$$Y_t = log(\frac{S_t}{S_{t-1}}) = \sqrt{h_t} z_t \tag{1}$$

where $(z_t)_{t \in \{1,...,T\}}$ are independent identically distributed random variables defined on the sample space $(\Omega, \mathcal{A}, \mathbb{P})$ and $(\mathcal{F}_t = \sigma(z_u; 1 \le u \le t))_{t \in \{1,...,T\}}$ is the associated information filtration. The volatility process $\sqrt{h_t}$ is driven by the set of parameters θ_V and the distribution of z_t by a set of parameters θ_D. The impact on the joint distribution of returns of both these sets of parameters is possibly the same and we wish to isolate the effect of each parameter in order to improve both the fit and the financial interpretation of these parameters.

5

Several estimation strategies are available to estimate these sets of parameters (there is no una-
nimity for one particular choice), and in the following, we consider three different approaches
and compare them through several criteria. We distinguish a direct method (estimating all the
parameters at the same time), a sequential method using two steps (beginning with the estima-
tion of the parameters of the volatility, then following with the distribution parameters) and a
multi-steps sequential method. Our strategy relies on simulations experiments. We specify the
algorithmes, and for each algorithm we recall the conditions under which the estimates converge.

We assume that we observe a data set $(Y_1, \cdots, Y_n)$. Conditionally to past observations, we as-
sume that the volatility of the returns is characterized by a set of parameters denoted θ_V, and
the conditional distribution of $(Y_t)_t$ is known, and characterized by the set of parameters θ_D. We
denote $\theta = (\theta_V, \theta_D) \in \Theta$ the p-vector of parameters to estimate, $\theta_0 \in \Theta$ being the true value of
the parameters. The different strategies that we investigate are the following ones:

1. The direct procedure based on Maximum Likelihood (ML) method. We estimate the vector
   of parameters $\theta = (\theta_V, \theta_D)$ using maximum likelihood method. For the previous given set
   of observations, the log likelihood is equal to:

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta), \tag{2}$$

   where $l_t(\theta)$ is the conditional log-likelihood given $\mathcal{F}_{t-1}$ computed for the model specified in
   equation (1). Assuming that the likelihood is computed using the initial density for z_t, the
   maximum likelihood estimate $\hat{\theta}_n$ of θ is such that $\hat{\theta}_n = \mathrm{argmin}_{\theta \in \Theta} L_n(\theta)$. This estimate is
   asymptotically Gaussian under mild conditions for z_t, say strict stationarity and ergodicity,
   and also the existence of a finite moment $E[z_t^{\nu+4}]$ for a small $\nu \in \mathbb{R}_+^*$:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow N(0, V_0), \tag{3}$$

   where the matrix V_0 classically depends on the estimate of the variance-covariance matrix
   of $\hat{\theta}_n$.

2. The two steps procedure based on the Quasi-Maximum Likelihood (QML) method.

6

(a) Assuming that the residuals $(z_t)_t$ in (1) follow a Gaussian distribution, we estimate the parameters θ_V using the maximum likelihood approach.

(b) Given the estimates $\hat{\theta}_V$, we use the residuals $(\hat{z}_t)_t$ to estimate the parameters θ_D by maximum likelihood method.

(c) Then, we obtain the maximum likelihood estimates $\hat{\theta} = (\hat{\theta}_V, \hat{\theta}_D)$.

Under very mild conditions on the distribution of z_t, Lee and Hansen (2004) prove the asymptotic normality of the estimates when Y_t is a GARCH(1,1) process. Straumann (2005) specifies the equivalent conditions for other classes of GARCH models. Then the same result as in (3) is obtained for $\hat{\theta}$, and the variance V_0 is equal to

$$V_0 = 4^{-1} E[z_1^4 - 1](E[h_1'(\theta_0)^\top h_1'(\theta_0)/\sigma^4])^{-1},$$

where h_1' denotes the vector of the first partial derivatives of h_1 with respect to θ, $\top$ represents the transpose operator, θ_0 the true value for θ, and σ^2 the variance of Y_1.

3. The multi-steps procedure denoted RECursive (REC) method based on an iterative approach starting from the QML one. The steps are the following:

(a) Assuming that the residuals $(z_t)_t$ follow a Gaussian law, we estimate the parameters θ_V using maximum likelihood approach.

(b) Given the estimates $\hat{\theta}_V$, we get the residuals $(\hat{z}_t)_t$, and we estimate its parameters θ_D using maximum likelihood method.

(c) Assuming that the residuals $(\hat{z}_t)_t$ follow now a given distribution, we re-estimate the volatility parameters by maximum likelihood method, and we get $\hat{\hat{\theta}}_V$.

(d) Given these previous estimates $\hat{\hat{\theta}}_V$, we build the new residuals $(\hat{\hat{z}}_t)$, and we re-estimate the parameters of the given distribution, obtaining $\hat{\hat{\theta}}_D$.

(e) And so on. We stop when $||\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}|| = \epsilon$, for a given small enough ϵ and k being the given number of iterations[1].

---

[1]Below, in both the Monte Carlo and the real data set experiments, we stopped at $k = 10$.

7

To provide the asymptotic normality of the last step estimate for θ, we follow Song *et al.* (2005) considering an additive decomposition of the log-likelihood function $L_n(\theta)$:

$$L_n(\theta) = L_{V,n}(\theta_V) + L_{D,n}(\theta_V, \theta_D),$$

where $L_{n,V}(\theta_V)$ corresponds to the log-likelihood computed using only the volatility function and $L_{D,n}(\theta_V, \theta_D)$ corresponds to the likelihood computed with all the parameters (conditional variance parameters and distribution parameters). After a finite number of iterations denoted k, the maximum likelihood estimate $\theta_n^{(k)}$ of θ is asymptotically Gaussian. To get this result, we assume regular conditions we recall now:

(a) L_n, $L_{V,n}$ and $L_{D,n}$ are twice continuously differentiable for $\theta \in \Theta$.

(b) $||I_V^{-1} I_D|| < 1$ where $I_V = -\frac{1}{n} E[L''_{V,n}(\theta_0)]$ and $I_D = -\frac{1}{n} E[L''_{D,n}(\theta_0)]$, $L''_{V,n}$ (resp. $L''_{D,n}$) being the Hessian matrix of $L_{V,n}$ (resp. $L_{D,n}$). This last condition is referred as the information dominance.

Then,

$$\sqrt{n}(\theta_n^k - \theta_0) \to N(0, \Sigma_k) \text{ as } n \to \infty, \tag{4}$$

with

$$\Sigma_k = A_k^T \Omega A_k,$$

where

$$A_k = \begin{pmatrix} (I_p - \tau^k)I^{-1} \\ \\ (I_p - \tau^{k-1})I^{-1} \end{pmatrix} \tag{5}$$

and

$$\Omega = \lim_{n \to \infty} \frac{1}{n} \begin{pmatrix} E[L'_{V,n}(L'_{V,n})^\top(\theta_0)] & E[L'_{V,n}(L'_{D,n})^\top(\theta_0)] \\ \\ E[L'_{D,n}(L'_{V,n})^\top(\theta_0)] & E[L'_{D,n}(L'_{D,n})^\top(\theta_0)] \end{pmatrix}. \tag{6}$$

I_p is the identity matrix of dimension $p \times p$, $\tau_k = -(I_V^{-1} I_D)^k$, and I^{-1} is the inverse of the Fisher information. It is interesting to note that condition (b) implies $(I_V^{-1} I_D)^k \to 0$ as $k \to \infty$ and that I_V is "larger" than I_D, meaning that $L_{V,n}$ contains more information on θ_0

than $L_{D,n}$. Consequently, the Hessian matrix of $L_{V,n}$ directs the movement of the updated values.

3   Monte Carlo experiments

In this section, we review and detail the results of the Monte Carlo experiments developed to test the three estimation strategies presented earlier. In this article, we aim at comparing these different estimation strategies based on simulated processes. The selected processes mix a special time-varying volatility structure with a non-Gaussian distribution. In such a case, a single step maximum likelihood estimation is due to be perturbated by the similar role played by parameters governing the volatility structure and the conditional distribution. The selected volatility structure are the Nelson (1991)'s EGARCH and Granger and Ding (1996)'s APARCH models as discussed above. These models were chosen for their ability to generate both a time varying volatility and leverage effects. The conditional distribution that we retained are the Generalized Hyperbolic distribution (GH hereafter) and the Mixture of Gaussian distributions (MN hereafter). The GH distribution has been introduced by Barndorff-Nielsen and Blaesild (1981) and applied to finance in Eberlein and Prause (2002) and Chorro *et al.* (2010). The MN distribution has been applied to financial returns modeling in Kon (1984), Akgiray and Booth (1987), Tucker and Pond (1998) and Alexander and Lazar (2006). These two distributions are interesting candidates as they encompass both fat tails and asymmetry of the conditional distribution of financial returns. Here, we focus on the mixture of two Gaussian distributions as in such a case the number of parameters driving the distribution is five, that is the number of parameters characterizing the GH one.

In order to compare the estimation approaches, we propose to rely on the following criteria:

- The first criterion we rely on is the Root Mean Square Error (RMSE hereafter) of the estimated parameters. It is due to help us deciding upon which methodology yields estimated parameters that are the closest to the true values used to sample the process. Let θ_i^0 be the true value for the parameter i and $\hat{\theta}_i^{j,n}$ be the estimated parameter using the methodology

9

j, obtained from the nth simulated sample. This parameter can belong either to the volatility structure or to the conditional distribution one. With a total number of N simulations and I parameters, the total RMSE is

$$RMSE_j = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{I} \left(\frac{\theta_i^0 - \hat{\theta}_i^{j,n}}{\theta_i^0} \right)^2}. \tag{7}$$

$RMSE_j$ should be seen as a score obtained by the estimation approach j. Larger parameters are usually affected by larger estimation errors: in the criterion that we propose, the errors are weighted by the true value of the parameter, making the aggregation of the estimation errors possible.

In the end, we compute three different scores. The first one is the previous score computed over the total number of parameters that we refer to as the "Total RMSE". The second and third scores focus on specific parameters: we are interested in the differences obtained with each estimation strategy when it either comes to the volatility parameters or the distribution ones. We thus compute the previous RMSE criterion for these two different subsets of parameters. We refer to them as "Volatility RMSE" and "Distribution RMSE". These criteria are computed as follows:

$$\text{Volatility } RMSE_j = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \sum_{i=I_1}^{I_1'} \left(\frac{\theta_i^0 - \hat{\theta}_i^{j,n}}{\theta_i^0} \right)^2} \tag{8}$$

$$\text{Distribution } RMSE_j = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \sum_{i=I_2}^{I_2'} \left(\frac{\theta_i^0 - \hat{\theta}_i^{j,n}}{\theta_i^0} \right)^2}. \tag{9}$$

The parameters $i \in [I_1; I_1']$ are the volatility parameters and the parameters $i \in [I_2 : I_2']$ are the distribution ones.

– Maximizing a likelihood function over parameters usually requires to use a numerical optimizer. Most of these optimizers are gradient-based method, such as the Newton-Raphson algorithm. Such algorithm requires are certain number of iterations before it delivers the output of the likelihood maximization. In the meantime, not every optimization delivers successful convergence of the algorithm. Hence, we use as a second criterion the rate of convergence of the BFGS optimization algorithm used to estimate the parameters with

10

each methodology. This method is a quasi-Newton method also known as "variable metric algorithm". Interested readers in such aspects can refer to Nocedal and Wright (1999). For each estimation approach, we thus computed the percentage of convergence with respect to the total number of experiments.

– The third numerical aspect we are interested in is the time used by each method: computational burden is an aspect of numerical optimization that should not be left apart, as the feasibility of the approaches investigated here should be an important trigger for readers interested in our results. Hence, the third criterion used in this article is the average time required to perform the estimation of each simulated model using the different estimation strategies presented before.

The numerical features of the simulation strategy implemented here unfolds as follows: the total number of experiments with each strategy for each model is 2,000 and the sample size used each time is 1,500 observations[2]. The simulations are based on parameters selected so that to mimic the salient features of financial markets' returns. The parameters used are the following:

– For the APARCH model:

$$h_t^\delta = a_0 + a_1 \left(|z_{t-1}| - \gamma z_{t-1} \right)^\delta + b_1 h_{t-1}^\delta, \tag{10}$$

with $\delta = 1.2$, $a_0 = 0.04$, $a_1 = 0.3$, $b_1 = 0.6$ and $\gamma = 0.75$.

– For the EGARCH model:

$$\log(h_t) = a_0 + a_1|z_{t-1}| + \gamma z_{t-1} + b_1 \log(h_{t-1}), \tag{11}$$

with $a_0 = -0.4$, $a_1 = 0.1$, $b_1 = 0.96$ and $\gamma = -0.1$.

---

[2]The selection of the sample size is important here. A too large sample size would lead to results very close to the asymptotic ones. In the meantime, we are interested in using a sample size close enough to the reality of financial markets. Broadly speaking, economic and market cycles in the recent years had a length of 6 years, that is 1,500 trading days. It seems that to avoid structural breaks in the parameters, using more than one economic cycle remains a risky choice to perform the estimation of such GARCH-like models, hence our choice of focusing on a sample size of 1,500 observations.

– The MN density is given by:

$$f(x) = \phi f(x, \mu_1, \sigma_1) + (1 - \phi)f(x, \mu_2, \sigma_2), \tag{12}$$

where $f(., \mu_i, \sigma_i)$ is the density of a Gaussian random variable with expectation μ_i and standard deviation σ_i. The parameters selected are: $\phi = 0.23$, $\mu_1 = -0.4$, $\sigma_1 = 1.3$, $\mu_2 = 0.12$ and $\sigma_2 = 0.86$.

– The GH distribution is given by:

$$f(x) = \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}e^{\beta(x-\mu)}\frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}} \tag{13}$$

where $\alpha = 1.9$, $\beta = -0.55$, $\delta = 3.6$, $\mu = 0.55$, $\lambda = -5.5$ and $K(.)$ being a Bessel function of the third kind.

For each experiment, the starting values of the parameters are obtained by perturbing the true values stated above in the following way:

$$\theta_i^{j,n} = \theta_i^0 \left(\frac{1}{2} + u\right), \tag{14}$$

for a given estimation approach j and the nth replication of the Monte Carlo experience. u is a random variable which follows a uniform distribution over $[0 : 1]$. There are numbers of constraints required with these volatility structures and distributions. Once the parameters are perturbed, we check for their consistency with these constraints and discard those that do not match these requirements. What is more, we impose these constraints numerically within the optimization process. However, this is of little impact on the results as the starting point is selected to be close to the true value of the parameters. This would have a sharper influence on the results in the case of a real dataset, involving the difficult step of the initialization of the parameters without knowing them. We do not recall these constraints here as they are detailed in the relevant literature on these volatility processes and on these distributions.

The Monte Carlo exercise is performed as follows: for a given model, we generate a sample using the true value for the parameters. Then, we perturb these values to obtain a starting point to optimize the likelihood function using the previous sample. At the end of this process, we store

12

the relevant information and start again these steps. In the case of the volatility, we initialize the volatility process to its long term average as estimated from the sample using the method of moment estimator. The REC estimation approach is run 10 times: this number was selected on the ground of various information. First, the improvement over this tenth step was unclear during the initial trial-and-error study we did. Second, the average time required to obtain these 10 optimization steps is around 6 to 9 times the time required for the QML estimator to be run. We wanted to make the approaches broadly comparable, even from a time consumption perspective.

Now, we turn our attention to the detailed analysis of the results, specifing the outcome of this simulation-based assessment of the various estimation approaches detailed earlier:

– Case 1: the GH-APARCH model.

The results for this model are presented in Table 1. For this case, the lowest Total RMSE is obtained with the final step of the recursive estimation approach, whose value is 2.66. It improves the QML starting point whose value is 3.65. The ML approach behaves poorly, yielding a Total RMSE equal to 4.69. The REC10 improvement stems mainly from the estimation of the volatility parameters, as the ML approach is the worst performing competitor, with a Volatility RMSE equal to 3.99, vs. 3.45 for the QML and 2.54 for the REC10. When it comes to the distribution's parameters estimation, the ML becomes the best competitor with a RMSE of 0.64, vs. 1.21 for the QML and 0.8 for REC10. In the end, the improvement over the volatility structure estimation is so important with the REC10 approach that it dominates the Total RMSE obtained. The biggest convergence rate achieved is obtain with the QML method. The ML only converges in 49% of our experiments and it shrinks down to 18% with REC10. Finally, the average time required to perform the estimation is 6 seconds for QML, 12 seconds for ML and 55 seconds for REC10.

– Case 2: the MN-APARCH model.

The results for this model are presented in Table 2. This case is very similar to the latter as the REC10 approach delivers the best Total RMSE (2.16), improving not only the

13

QML score (3.08), but also the ML case (4.69). Again, the main explanation stems from the volatility's parameters estimation as for the Volatility RMSE, the hierarchy between the approaches remains the same: ML obtains 4.67, QML 2.99 and REC10 2.09. The distribution's parameters are best estimated with the ML method (0.43), whereas the REC10 one comes second (0.55) while still improving the QML case (0.75). However, these figures remain remarkably close to each other. As in the previous case, the QML approach obtains the highest convergence rate (86%), that is higher than those of the ML approach (64%) or of the REC10 (21%). It seems that the quality of the estimation still comes at a convergence price along with an increase in the time required to perform the estimation: REC10 uses 61 seconds when the ML estimators are obtained after 14 seconds and the QML ones after 7 seconds.

– Case 3: the GH-EGARCH model.

The case of GH-EGARCH is presented in Table 3. The results obtained here are different from those obtained before: it appears that QML gets the lowest Total RMSE (3.12), while being very close to the ML estimates (3.14). The REC10 delivers a score of 3.64, which remains however close to the others. This is again the result of the estimation of the volatility parameters, as a similar ranking is obtained with the Volatility RMSE: QML yields 2.96, ML gets 3.04 and finally REC10 delivers a ratio of 3.58. Here again the scores remain close. The rate of convergence for REC10 is higher than in the previous cases: 54% of our estimations converged, vs. 89% in the ML case and 95% in the QML one. Finally, the QML approach is the fastest strategy, with an average of 5.97 seconds, when ML yields estimates after 9.82 seconds and REC10 after 62.14 seconds.

– Case 4: the MN-EGARCH model.

This final model's results are presented in Table 4. For such a case, the REC10 dominates again the Total RMSE ranking (1.38), whereas QML is the worst performing competitor (3.16), right behind ML (1.77). In this case, the volatility results explain most of this performance, as the REC10 shrinks the QML's Volatility RMSE from 3.02 to 1.15. ML gets a score of 1.68. The story for the Distribution RMSE is different again, as ML appears to

14

be the best competitor with a score of 0.55, when QML gets a score of 0.92 that is reduced by the REC10 to 0.76. The gap between ML and REC10 is thus small. The algorithm convergence rates are 90% with ML, 92% with QML and 55% with REC10. This result is thus similar to the one obtained in the GH-EGARCH. The time consumption in this case for the REC10 approach is smallest of all our experiments, as it requires only 42 seconds. The same comment applies for the ML whose average time needed to estimate the parameters is 8 seconds.

We can thus summarize the salient features of this exercise as follows: first, the REC10 approach always improves the QML one, whatever the model and the subset of parameters investigated. Second, the REC10 dominates 3 out of the 4 Total RMSE investigated, mostly because it provides estimates of the volatility parameters with smaller estimation errors. The results obtained in the case of the distribution's parameter do not provide such sharp conclusions. Even though the ML estimates remain the most precise ones, the difference between the REC10 estimates and the ML ones remain small in every case. A notable exception is obtained in the GH-EGARCH case, with a lower Distribution RMSE in the REC10 case than in the ML one. This overall nice performance of the REC10 approach comes at different costs: first, the rate of convergence is smaller. In the EGARCH cases the rate of convergence drops to 50% while in the APARCH case it falls to 20%. The APARCH model requires the estimation of a parameter that is the power to which we should model the volatility process. This kind of parameter is typically more difficult to estimate than the usual linear GARCH parameters. In the meantime, the time to estimate the parameters using the recursive approach is due to be longer than the QML one. It is around 6 to 7 times the QML estimation time. The interesting element there is that ML remains a faster method than the recursive one. Hence, the rise in estimation quality obtained with the REC10 method is obtained with a lesser rate of convergence and with a longer estimation time.

4   Empirical application

This section proposes an empirical application building on the results of the previous Monte Carlo study. We propose to estimate the various models proposed earlier using a data set of 1506

returns on the SP 500, the American equity index. The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes different market phases, both bull and bear, as it includes the 1998 market rally and the explosion of the technology bubble in 2001. Figure 1 charts the evolution of the index over the sample. Descriptive statistics are provided by Table 5. The SP500 returns are characterized by fat tails, as the excess kurtosis is positive. The skewness of the sample is close to 0: this is explained by the balance between the positive and negative momentum periods over the sample. 49.83% of the returns are negative in the sample. This remains however compatible with leverage effects.

We estimate the parameters of the models combining the volatility structures and the distributions used in our Monte Carlo experience and the three estimation strategies: ML, QML and REC. The starting values to perform these estimates are the same for each method and are set to be equal to the values used in the Monte Carlo experience. All estimations converged with these starting points, contrasting with the Monte Carlo results for these four recursive estimations. However the time consumption – that is not provided here – of the recursive method remains higher than the one necessary in the other cases. The estimated parameters for the EGARCH-GH, the EGARCH-MN, the APARCH-GH and the APARCH-MN models are respectively provided in Tables 6, 7, 8 and 9. The examination of the estimation tables yields the impression that the estimated parameters can be different, even when the sample remains the same. For example, focusing on the results of Table 7, the size of the leverage effect is clearly different across the estimation approaches. In the case of the EGARCH model, the returns-to-volatility spill-over effect is captured in the conditional variance equation:

$$\log(h_t) = a_0 + a_1|z_{t-1}| + \gamma z_{t-1} + b_1 \log(h_{t-1}). \tag{15}$$

Whenever z_t is positive, the volatility is increased by a factor equal to $a_1 + \gamma$. When it is negative, this factor becomes $a_1 - \gamma$. The leverage effect is characterized by the difference between these two figures. In the EGARCH-MN case, QML estimation yields a spread equal to 0.26, that becomes in the case of the ML strategy 0.1. Finally, the REC estimations point toward a difference equal to 0.38. Such a disparity between figures is too high to be trustworthy. The

question is now to detect which of these estimations provide the fairer results. We propose to answer this question using two different arguments. First, we propose to test which of these estimation strategies yield the best fit to the joint distribution of the sample considered here. Then, we propose to investigate the stability of the news impact curves obtained for each model and of the estimated conditional distribution.

We are not interested into deciding which of these estimation approaches delivers a set of parameters making that a given model is likely to be the true data generating process. We rather focus on the fact that one of these estimation approaches should be better than the remaining two others. To test such an hypothesis, we propose to use the joint density of the sample as a score to differentiate the estimation strategies. The test we use can be seen as an in-sample version of the test proposed for density forecast in Amisano and Giacomini (2007). Say we deal with a time series model for the log-returns whose conditional density at time t is $f_1(Y_t|\underline{Y_{t-1}}, \theta_1)$, where $\underline{Y_{t-1}} = (Y_1, ..., Y_{t-1})$ and θ_1 is the vector of parameters describing the shape of this conditional distribution and the volatility structure. We compare this model to another one defined by the conditional density $f_2(Y_t|\underline{Y_{t-1}}, \theta_2)$, with θ_2 being the parameters associated to this second model. The null hypothesis of the test is "models 1 and 2 provide a similar fit of the log-return's conditional distribution". The corresponding test statistic used is:

$$t_{1,2} = \frac{1}{n} \sum_{t=1}^{n} \left(\log f_1(Y_t|\underline{Y_{t-1}}, \theta_1) - \log f_2(Y_t|\underline{Y_{t-1}}, \theta_2) \right), \tag{16}$$

where n is the total number of observations available. Under the null hypothesis

$$\frac{t_{1,2}}{\hat{\sigma}_n} \sqrt{n} \underset{n \to +\infty}{\to} \mathcal{N}(0, 1), \tag{17}$$

where $\hat{\sigma}_n$ is a properly selected estimator for the statistic volatility. Here, as proposed in Amisano and Giacomini (2007), we use a Newey-West estimator, with a lag empirically retained to be large (around 25).

Table 10 provide the results of this test, displaying the test statistics as presented in equation (17). We propose three tests: the first one compares the ML vs. the QML approach; the second

17

one compares the REC vs. the QML one and the third one compares the REC to the ML estimates. The core results can be summarized as follows: (1) ML and REC dominates the QML results, in a stronger manner than in our Monte Carlo simulations. The elaborated estimation strategies yield estimates that reach a higher value for the density. (2) The REC estimates outperform the ML ones: they lead to a log-likelihood that is statistically higher for each of the models investigated here. The global argument is thus that the REC approach advocated here improves the two remaining ones: the estimates obtained stand a greater chance to be closer to the true set of parameters that makes the likelihood function maximum.

Our second argument is related to Figure 2 and Figure 3. The first one provides the news impact curve obtained for each model with the different sets of estimated parameters. This concept introduced in the financial econometrics literature in Engle and Ng (1993) measures the contribution of past returns to volatility. In a general setting, Figure 2 charts the volatility on date t as a function of the return on date $t-1$. As this function clearly depends on the model used, we do not provide these computations here. Defining a plausible support for the returns' value and setting the past volatility to its long term average, we compute the expected value for the next volatility. Investigating Figure 2, we clearly obtain a high variability between the different shapes obtained. Our main finding here is that the REC approach – unlike its competitors – provides a shape that is stable across the models: the scale of values obtained for this news impact curves is the same across the four charts. These results can be directly extended to the estimated conditional distribution presented in Figure 3: REC is the only approach that yields a stable shape for the distribution. If this phenomenon is not exactly a surprise when it comes to the QML method, the case of ML is more striking. Examining the tables 6, 7, 8 and 9, the problem that arises with the ML estimation strategy seems to be partly related to the relation between leverage effects and skewness. The mixing effects advocated earlier seems to have a disruptive effect, leading to the instability in the estimated news response function and in the conditional distribution. A similar comment can be made when it comes to the kurtosis implied by the conditional distribution. From Table 6 for example, the estimated value for α in the GH

distribution is clearly different with the ML approach, as it reaches 4.22 whereas it is equal to 1.47 with REC and 1.2 with QML. Interestingly, when comparing the GH parameters from the Table 6 and Table 8, they are found to be quite stable in the REC case and very unstable in the ML one.

This instability could lead to a wrong model selection. Using the density test presented earlier, we propose to test which model should be favored for the data set at hand and for a given estimation strategy. For example, in the case of the REC estimation method, we test the different combinations of estimated models that mix the EGARCH/APARCH volatility structures to the MN/GH conditional distributions. The test results are presented in Table 11. Given the transitivity of this test statistics, it is possible to rank the different models and thus decide on the model(s) that should be favored. In the case of the QML estimates, as all the statistics are between -1.96 and 1.96, there is no best estimated model. This means that all the models can be considered as equivalent, as no model can be favored over the others. In the case of the ML estimates, the EGARCH GH outperforms all its competitors, as when testing it goodness of fit versus the other models, it obtains a test statistics that is always greater than 1.96. The conclusion obtained with the REC estimates is different: most models provide an equivalent fit, but for the EGARCH MN that dominates the EGARCH GH one. Hence, depending on the estimation strategy, the ranking obtained is different and misleading conclusions can be reached. This is interesting as focusing on the EGARCH GH case, the ML estimates lead to a quite symmetric news impact curve (see Figure 2), as most of the asymmetry in returns is captured in the conditional distribution. Indeed, when investigating Table 6, when β – that is the parameter controlling asymmetry in the conditional distribution – is estimated to be equal to -0.74 with the REC approach, with a twice as big value in the ML case (-1.62). Hence, the conditional distribution is more skewed to the left, as the probability to observe sharply negative returns is higher. Given the test results discussed above and the stability of the results obtained with the REC estimation approach, the estimation results obtained with the latter method are apparently the most trustworthy with this specific dataset.

19

5 Conclusion

This article raises the question of the impact of jointly estimating the parameters driving the volatility and the conditional distribution of financial asset returns in the case of conditionally non-Gaussian models. We compare three types of estimation methods: the standard maximum likelihood aims at estimating all the parameters in the meantime and is the asymptotically most efficient estimation approach. The QML and the recursive approaches are based on a separated estimation for the volatility and the distribution parameters. If the QML is found to be dominated by ML in a Monte Carlo exercise, a similar conclusion cannot be reached in the case of the recursive estimates. The recursive approach often dominates the ML one, especially when it comes to estimating the volatility parameters. Moving one step forward and using a real data set of SP500 returns, this article shows how the recursively-obtained estimates provide both a higher goodness of fit for a given model and stable results when it comes to volatility effects or to conditional distribution across models. As discussed in the introduction, the estimation problems encountered both with the QML and ML strategies could jeopardize the conclusions they lead to.

This article focuses only on maximum likelihood related approaches, ignoring other well know methods that lie outside the scope of our work such as the Generalized Method of Moment or Bayesian estimation approaches (e.g. Ardia, 2008). In the special case of GARCH-like models, the volatility being a function of past returns, such methods are rarely used to perform the estimation of such models. In the case of stochastic volatility – i.e. when the volatility dynamics involves a disturbance on its own – an extension of the spectrum of the methods compared would be necessary. Finally, this article's aim is not to provide the perfect estimation method but to explicitly compare three estimation approaches that may be the most common ones in the past thirty years of financial econometrics literature.

References

[1] Akgiray, V., Booth, G. G., 1987. Compound Distribution Models of Stock Returns: An Empirical Comparison. The Journal of Financial Research 10, 269-280.

[2] Alexander, C., Lazar, E., 2006. Normal Mixture GARCH(1,1): Applications to Exchange Rate Modeling. Journal of Applied Econometrics 21, 307-336.

[3] Amisano, G., Giacomini, R., 2007. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business & Economic Statistics, American Statistical Association 25, 177-190, April.

[4] Ardia, D., 2008. Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications. Lecture Notes in Economics and Mathematical Systems. Springer. Fribourg.

[5] Badescu, A., Kulperger, R., Lazar, E., 2008. Option Valuation with Normal Mixture GARCH Models. Studies in Nonlinear Dynamics and Econometrics 12 (2), 1580-1580.

[6] Barndorff-Nielsen, O.E., Blaesild, P., 1981. Hyperbolic Distributions and Ramifications: Contributions to Theory and Applications, in: Taillie C, Patil GP, Baldessari BA (Eds), Statistical distributions in scientific work, Vol. 4. Reidel, Dordrecht, pp. 19-44.

[7] Chernov, M., Ghysels, E., 2000. A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation. Journal of Financial Economics 56, 407 - 458.

[8] Christoffersen, P., Heston, S., Jacobs, K., 2006. Option Valuation with Conditional Skewness. Journal of Econometrics 131, 253 - 284.

[9] Chorro, C., Guégan, D., Ielpo, F., 2010. Martingalized Historical approach for Option Pricing. Finance Research Letters 7(1), 24-28.

[10] Chorro, C., Guégan, D., Ielpo, F., 2010. Option Pricing for GARCH-type Models with Generalized Hyperbolic Innovations. *To appear in* Quantitative Finance.

21

[11] Ding Z., Granger, C.W.J., Engle, R.F., 1993. A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance 1, 83-106.

[12] Duan, J.C., (1995). The GARCH Option Pricing Model. Mathematical Finance, 5, 13-32.

[13] Eberlein, E., Prause, K., 2002. The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures, in: Geman, H., Madan, D., Pliska, S., Vorst, T. (Eds), Mathematical Finance-Bachelier Congress 2000. Springer Verlag, pp. 245-267.

[14] Engle, R. F., Ng, S., 1993. Measuring and Testing the Impact of News on Volatility. Journal of Finance, American Finance Association 48(5), 1749-78.

[15] Fan, Y., Pastorello, S., Renault, E., 2007. Maximisation by Parts in Extremum Estimation. WP university of North Carolina, NC.

[16] Gouriéroux, C., Monfort, A., Renault E., 1993. Indirect Inference. Journal of Applied Econometrics 8, S85 - S118.

[17] Ghysels, E., Harvey, A., Renault E., 1996. Stochastic Volatility, in: Maddala, G.S., Rao, C.R. (Eds) Handbook of statistics, Vol. 14. Elsevier, Amsterdam.

[18] Kon, S. J., 1984. Models of Stock Returns: A Comparison. The Journal of Finance 39, 147-165.

[19] Lee, S.W., Hansen, B.C., 1994. Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. Econometric theory 10, 29 - 52.

[20] Lumsdaine, R.L., 1996. Consistency and Asymptotic Normality of the Quasi Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. Econometrica 64, 575 - 596.

[21] Nelson, D.B., (1991). Conditional Heteroskedasticity in Asset Returns. Econometrica, 59, 347-370.

[22] Nocedal, J., Wright, S. J., 1999. Numerical Optimization. Springer, Berlin.
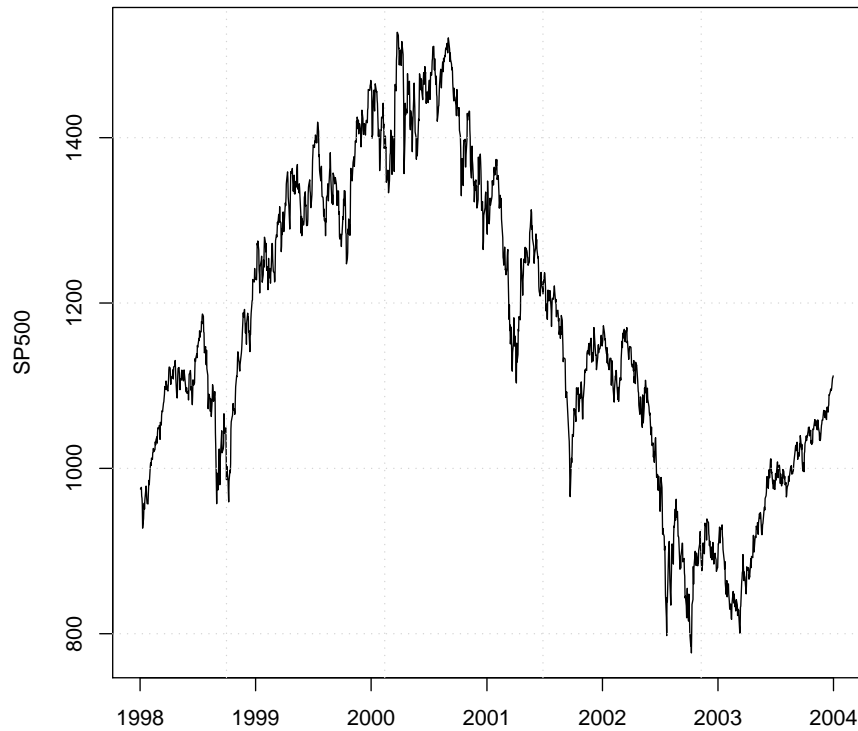
Figure 1: **SP500 prices. Data source: Bloomberg.**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data

[23] Song, P.X, Fan, Y., Kalfeisch, D., 2005. Maximization by Parts in Likelihood Inference. Journal of the American Statistical Association 100, 1145 - 1158.

[24] Straumann, D., 2005. Estimation in Conditionally Heteroscedastic Time Series Models. Lecture Notes in Statisitcs, 181. Springer, Berlin.

[25] Tucker, A. L., Pond, L., 1998. The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes. Review of Economics and Statistics 11, 638-647.

Tables and figures

23

Method	% of convergence	Time	Distribution RMSE	Volatility RMSE	Total RMSE
ML	0.49	12.02	0.64	3.99	4.04
QML	0.87	6.16	1.21	3.45	3.65
REC2	0.57	21.81	2.08	11.47	11.66
REC3	0.42	27.1	2.23	10.62	10.86
REC4	0.33	31.9	2.15	9.51	9.75
REC5	0.27	36.28	1.98	8.26	8.49
REC6	0.23	40.18	1.66	6.9	7.09
REC7	0.21	43.63	1.53	5.89	6.08
REC8	0.19	47.08	1.19	4.29	4.45
REC9	0.18	50.77	1.15	3.91	4.08
REC10	0.18	54.49	0.8	2.54	2.66

Table 1: **Comparison of the estimation methodology across the three criteria in the case of the GH-APARCH model**

The model specifications are the following:

– For the APARCH model:

$$h_t^\delta = a_0 + a_1 \left(|z_{t-1}| - \gamma z_{t-1}\right)^\delta + b_1 h_{t-1}^\delta, \tag{18}$$

with $\delta = 1.2$, $a_0 = 0.04$, $a_1 = 0.3$, $b_1 = 0.6$ and $\gamma = 0.75$.

– The GH distribution is given by:

$$f(x) = \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi} K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}} \tag{19}$$

where $\alpha = 1.9$, $\beta = -0.55$, $\delta = 3.6$, $\mu = 0.55$ and $\lambda = -5.5$.

The number of simulations is 2000 and the sample size is 1500. ML stands for maximum likelihood, QML for Quasi-Maximum Likelihood and RECx for the recursive algorithm at step x. The column "Time" is specified in terms of seconds. "% of convergence" presents the percentage of convergence obtained at the end of the Monte Carlo experiments. Distribution RMSE presents the quantity computed from equation (9), "Volatility RMSE" presents the quantity computed from equation (8) and "Total RMSE" presents the quantity computed from equation (7).

24

Method	% of convergence	Time	Distribution RMSE	Volatility RMSE	Total RMSE
ML	0.64	13.99	0.43	4.67	4.69
QML	0.86	6.67	0.75	2.99	3.08
REC2	0.49	25.43	1.03	10.81	10.86
REC3	0.34	29.68	1.02	9.08	9.14
REC4	0.29	34.05	0.94	7.91	7.97
REC5	0.26	38.18	0.84	6.47	6.52
REC6	0.24	42.32	0.75	5.25	5.3
REC7	0.22	46.46	0.68	4.06	4.12
REC8	0.22	50.87	0.62	3.4	3.45
REC9	0.21	55.55	0.58	2.67	2.73
REC10	0.21	60.54	0.55	2.09	2.16

Table 2: **Comparison of the estimation methodology across the three criteria in the case of the MN-APARCH model**

The model specifications are the following:

– For the APARCH model:

$$h_t^\delta = a_0 + a_1 \left(|z_{t-1}| - \gamma z_{t-1}\right)^\delta + b_1 h_{t-1}^\delta, \tag{20}$$

with $\delta = 1.2$, $a_0 = 0.04$, $a_1 = 0.3$, $b_1 = 0.6$ and $\gamma = 0.75$.

– The MN density is given by:

$$f(x) = \phi f(x, \mu_1, \sigma_1) + (1 - \phi) f(x, \mu_2, \sigma_2), \tag{21}$$

where $f(., \mu_i, \sigma_i)$ is the density of a Gaussian random variable with expectation μ_i and standard deviation σ_i. The parameters selected are: $\phi = 0.23$, $\mu_1 = -0.4$, $\sigma_1 = 1.3$, $\mu_2 = 0.12$ and $\sigma_2 = 0.86$.

The number of simulations is 2000 and the sample size is 1500. ML stands for maximum likelihood, QML for Quasi-Maximum Likelihood and RECx for the recursive algorithm at step x. The column "Time" is specified in terms of seconds. "% of convergence" presents the percentage of convergence obtained at the end of the Monte Carlo experiments. Distribution RMSE presents the quantity computed from equation (9), "Volatility RMSE" presents the quantity computed from equation (8) and "Total RMSE" presents the quantity computed from equation (7).

Method	% of convergence	Time	Distribution RMSE	Volatility RMSE	Total RMSE
ML	0.89	9.82	0.76	3.04	3.14
QML	0.95	5.97	0.99	2.96	3.12
REC2	0.84	14.34	1.15	5.12	5.25
REC3	0.67	26.32	1.01	8.26	8.32
REC4	0.62	33.56	0.91	6.13	6.2
REC5	0.59	47.04	0.85	7.93	7.98
REC6	0.57	49.47	0.77	9.43	9.47
REC7	0.56	52.05	0.73	7.35	7.38
REC8	0.55	54.95	0.7	7.37	7.41
REC9	0.55	58.09	0.68	4.64	4.69
REC10	0.54	62.14	0.66	3.58	3.64

Table 3: **Comparison of the estimation methodology across the three criteria in the case of the GH-EGARCH model**

The model specifications are the following:

– For the EGARCH model:

$$\log(h_t) = a_0 + a_1|z_{t-1}| + \gamma z_{t-1} + b_1 \log(h_{t-1}), \tag{22}$$

with $a_0 = -0.4$, $a_1 = 0.1$, $b_1 = 0.96$ and $\gamma = -0.1$.

– The GH distribution is given by:

$$f(x) = \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}} \tag{23}$$

where $\alpha = 1.9$, $\beta = -0.55$, $\delta = 3.6$, $\mu = 0.55$ and $\lambda = -5.5$.

The number of simulations is 2000 and the sample size is 1500. ML stands for maximum likelihood, QML for Quasi-Maximum Likelihood and RECx for the recursive algorithm at step x. The column "Time" is specified in terms of seconds. "% of convergence" presents the percentage of convergence obtained at the end of the Monte Carlo experiments. Distribution RMSE presents the quantity computed from equation (9), "Volatility RMSE" presents the quantity computed from equation (8) and "Total RMSE" presents the quantity computed from equation (7).

Method	% of convergence	Time	Distribution RMSE	Volatility RMSE	Total RMSE
ML	0.9	7.91	0.55	1.68	1.77
QML	0.92	7.15	0.92	3.02	3.16
REC2	0.69	8.5	3.4	3.14	4.63
REC3	0.6	14.05	3.26	10.44	10.94
REC4	0.58	19.38	2.63	10.51	10.84
REC5	0.57	21.27	2.3	10.24	10.49
REC6	0.56	25.91	1.96	8.72	8.94
REC7	0.55	28.85	1.31	8.73	8.83
REC8	0.55	33.26	1.13	8.64	8.71
REC9	0.55	37.67	0.94	1.36	1.65
REC10	0.55	42.03	0.76	1.15	1.38

Table 4: **Comparison of the estimation methodology across the three criteria in the case of the MN-EGARCH model**

The model specifications are the following:

– For the EGARCH model:

$$\log(h_t) = a_0 + a_1|z_{t-1}| + \gamma z_{t-1} + b_1 \log(h_{t-1}), \qquad (24)$$

with $a_0 = -0.4$, $a_1 = 0.1$, $b_1 = 0.96$ and $\gamma = -0.1$.

– The MN density is given by:

$$f(x) = \phi f(x, \mu_1, \sigma_1) + (1 - \phi)f(x, \mu_2, \sigma_2), \qquad (25)$$

where $f(., \mu_i, \sigma_i)$ is the density of a Gaussian random variable with expectation μ_i and standard deviation σ_i. The parameters selected are: $\phi = 0.23$, $\mu_1 = -0.4$, $\sigma_1 = 1.3$, $\mu_2 = 0.12$ and $\sigma_2 = 0.86$.

The number of simulations is 2000 and the sample size is 1500. ML stands for maximum likelihood, QML for Quasi-Maximum Likelihood and RECx for the recursive algorithm at step x. The column "Time" is specified in terms of seconds. "% of convergence" presents the percentage of convergence obtained at the end of the Monte Carlo experiments. Distribution RMSE presents the quantity computed from equation (9), "Volatility RMSE" presents the quantity computed from equation (8) and "Total RMSE" presents the quantity computed from equation (7).

	Average	Standard Dev.	Skewness	Kurtosis	Length
SP 500	0.02	0.21	0.01	1.68	1507

Table 5: **Descriptive statistics for the SP500 returns**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. The average and standard error statistics are annualized.

	α	β	δ	μ	λ	a_0	a_1	γ	b_1
QML	1.20	-0.60	4.42	0.57	-10.60	-0.32	-0.13	0.08	0.97
ML	4.22	-1.62	2.85	0.62	-9.59	-1.12	-0.05	1.13	0.91
REC	1.47	-0.74	1.84	0.16	-8.55	-0.33	-0.19	0.17	0.96

Table 6: **Estimated parameters for the EGARCH - GH model using the SP500 dataset**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. The model specifications are the following:

- For the EGARCH model:

$$\log(h_t) = a_0 + a_1|z_{t-1}| + \gamma z_{t-1} + b_1 \log(h_{t-1}), \tag{26}$$

- The GH distribution is given by:

$$f(x) = \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}} \tag{27}$$

28

	ϕ	μ_1	σ_1	μ_2	σ_2	a_0	a_1	γ	b_1
QML	0.004	-3.67	0.74	0.010	0.97	-0.32	-0.13	0.08	0.97
ML	0.002	-2.79	0.29	-0.204	0.61	-0.39	0.07	0.19	0.96
REC	0.003	-1.97	0.09	0.003	0.46	-0.33	-0.17	0.18	0.96

Table 7: **Estimated parameters for the EGARCH - MN model using the SP500 dataset**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. The model specifications are the following:

– For the EGARCH model:

$$\log(h_t) = a_0 + a_1|z_{t-1}| + \gamma z_{t-1} + b_1 \log(h_{t-1}),\tag{28}$$

– The MN density is given by:

$$f(x) = \phi f(x, \mu_1, \sigma_1) + (1-\phi)f(x, \mu_2, \sigma_2),\tag{29}$$

where $f(., \mu_i, \sigma_i)$ is the density of a Gaussian random variable with expectation μ_i and standard deviation σ_i.

	α	β	$\delta_{\text{distrib.}}$	μ	λ	a_0	a_1	b_1	γ	δ_{vol}
QML	1.08	-0.540	4.23	0.51	-9.80	0.0002	0.09	0.896	0.59	1.22
ML	1.08	-0.536	4.18	0.53	-9.80	0.0001	0.12	0.896	0.60	1.68
REC	1.27	-0.635	1.77	0.14	-7.99	0.0006	0.11	0.909	0.98	1.24

Table 8: **Estimated parameters for the APARCH - GH model using the SP500 dataset**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. The model specifications are the following:

– For the APARCH model:

$$h_t^\delta = a_0 + a_1 \left(|z_{t-1}| - \gamma z_{t-1}\right)^\delta + b_1 h_{t-1}^\delta,\tag{30}$$

– The GH distribution is given by:

$$f(x) = \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}}\tag{31}$$

29

	ϕ	μ_1	σ_1	μ_2	σ_2	a_0	a_1	b_1	γ	δ_{vol}
QML	0.9899	0.019	0.96	-2.562	1.164	0.00015	0.09	0.896	0.5924	1.22
ML	0.9999	0.018	0.96	-2.560	1.173	0.00019	0.21	0.908	0.5918	1.33
REC	0.9964	0.005	0.46	-1.859	0.199	0.00067	0.12	0.913	0.6549	1.16

Table 9: **Estimated parameters for the APARCH - MN model using the SP500 dataset**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. The model specifications are the following:

– For the APARCH model:

$$h_t^\delta = a_0 + a_1 \left(|z_{t-1}| - \gamma z_{t-1}\right)^\delta + b_1 h_{t-1}^\delta, \tag{32}$$

– The MN density is given by:

$$f(x) = \phi f(x, \mu_1, \sigma_1) + (1 - \phi)f(x, \mu_2, \sigma_2), \tag{33}$$

where $f(., \mu_i, \sigma_i)$ is the density of a Gaussian random variable with expectation μ_i and standard deviation σ_i.

Models	ML vs. QML	REC vs. QML	REC vs. ML
EGARCH - GH	40.27	83.13	29.74
APARCH - GH	19	75.84	15.55
EGARCH - MN	43.32	86.67	34.22
APARCH - MN	19.02	91.1	14.92

Table 10: **Density tests for the estimated models**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. When comparing the accuracy of model 1 (with parameters θ_1) vs. model 2 (with parameters θ_2) to fit the joint distribution of a given sample, the test statistic is computed as follows:

$$t_{1,2} = \frac{1}{n} \sum_{t=1}^{n} \left(\log f_1(Y_t | \underline{Y_{t-1}}, \theta_1) - \log f_2(Y_t | \underline{Y_{t-1}}, \theta_2) \right), \tag{34}$$

with $f_.(.)$ the selected conditional density. The test reads as follows: in the EGARCH-GH case, the column "ML vs. QML" uses the estimated parameters by ML as model 1 and the parameters estimated by QML as model 2. The test statistics value is 40.27: this value being outside the $[-1.96 : 1.96]$ 5% interval confidence, the null hypothesis that both models are equivalent is strongly rejected. The positivity of this statistics indicates that model 1 is favored over model 2.

column 5c**QML**				
	EGARCH GH	APARCH GH	EGARCH MN	APARCH MN
EGARCH GH	–	0.57	-1.07	0.18
APARCH GH	–	–	-1.10	-0.84
EGARCH MN	–	–	–	0.86
APARCH MN	–	–	–	–
ML				
	EGARCH GH	APARCH GH	EGARCH MN	APARCH MN
EGARCH GH	–	21.20	2.42	21.12
APARCH GH	–	–	-15.86	0.91
EGARCH MN	–	–	–	15.66
APARCH MN	–	–	–	–
REC				
	EGARCH GH	APARCH GH	EGARCH MN	APARCH MN
EGARCH GH	–	-0.09	-3.88	-0.08
APARCH GH	–	–	-0.81	0.03
EGARCH MN	–	–	–	0.95
APARCH MN	–	–	–	–

Table 11: **Density tests testing model domination**

The data set starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The returns computed are logarithmic returns. ML stands for maximum likelihood, QML stands for Quasi Maximum Likelihood and REC stands for Recursive Likelihood. When comparing the accuracy of model 1 (with parameters θ_1) vs. model 2 (with parameters θ_2) to fit the joint distribution of a given sample, the test statistic is computed as follows:

$$t_{1,2} = \frac{1}{n} \sum_{t=1}^{n} \left(\log f_1(Y_t | \underline{Y_{t-1}}, \theta_1) - \log f_2(Y_t | \underline{Y_{t-1}}, \theta_2) \right), \tag{35}$$

with $f_.(.)$ the selected conditional density. The test reads as follows: in case of the ML estimation approach, when the EGARCH-GH is model 1 and the EGARCH-MN is model 2, the test statistics value is 2.42: this value being outside the $[-1.96 : 1.96]$ 5% interval confidence, the null hypothesis that both models are equivalent is strongly rejected. The positivity of this statistics indicates that model 1 is favored over model 2.
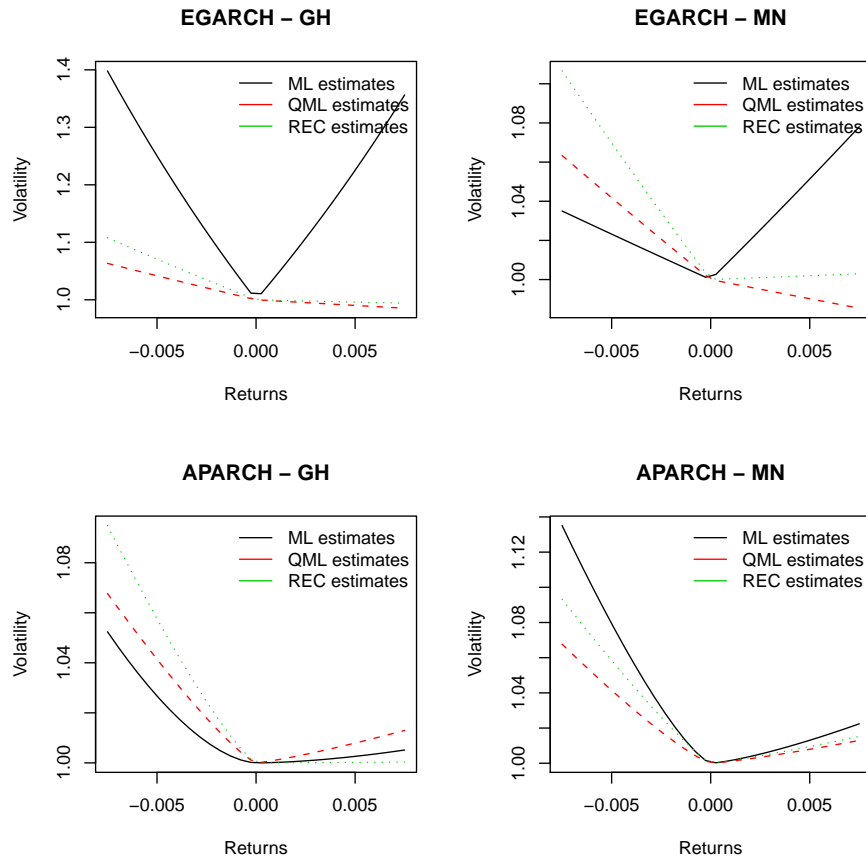
31

Figure 2: **Volatility News Impact curve**

The data set used starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The figure presents the News Impact curve when the initial level of volatility is set to the average of the sample at hand. The parameters used are the parameters estimated with the three maximum likelihood approaches. The top-left figure is for the EGARCH-GH model, the top-right is for the EGARCH-MN, the bottom-left figure is for the APARCH-GH model and the bottom-right figure is for the APARCH-MN. The returns value is expressed in terms of daily values.
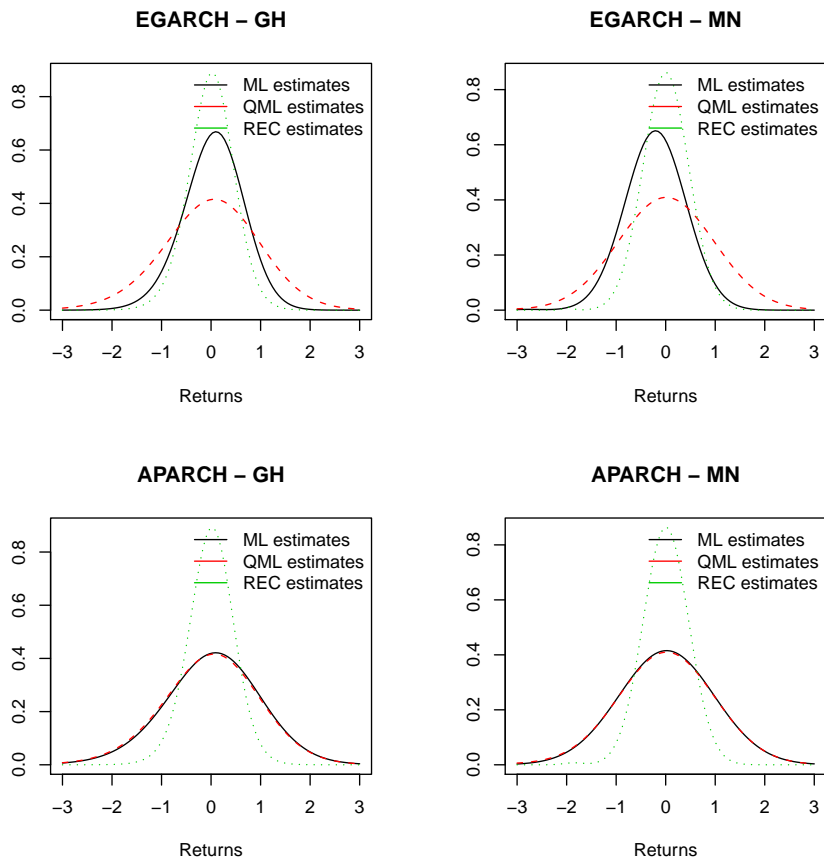
32

Figure 3: **Estimated conditional densities**

The data set used starts on January, 2nd 1998 and ends on December, 31st 2003. It includes 1506 data points. The figure presents the estimated conditional densities obtained from the four models estimated using the three estimation approaches investigated in this paper. The top-left figure is for the EGARCH-GH model, the top-right is for the EGARCH-MN, the bottom-left figure is for the APARCH-GH model and the bottom-right figure is for the APARCH-MN.

33