Numerical estimation of log-GARCH model with relaxed moment assumptions

Abdeljalil Settar\textsuperscript{1}, Nadia Idrissi Fatmi\textsuperscript{1} and Mohammed Badaoui\textsuperscript{1,2}

\textsuperscript{1}LIPIM, École Nationale des Sciences Appliquées, Université Sultan Moulay Slimane, Khouribga, Morocco
\textsuperscript{2}LaMSD, École Supérieure de Technologie, Université Mohamed premier, Oujda, Morocco
E-mail: abdeljalilssettar@gmail.com

Abstract. This paper provides algorithms for the numerical estimation of the log-GARCH model parameters with no assumptions on the existence of the log-moment orders greater than one. Our approach is based on the quasi-maximum likelihood estimation combined with the information filter. The proposed estimation is employed for two aims. The first is to treat the zero returns considered as missing values through an EM imputation algorithm. The second is to compute the kurtosis of the log-GARCH process by the so-called, right and left measures. A Monte Carlo simulation is performed to investigate the potential of the proposed algorithms to improve the accuracy of the quasi-maximum likelihood for the parameter estimation and the treatment of zero returns as well as to check the robustness of the used kurtosis measures.

1. Introduction
The modeling of financial series has seen an extreme use of ARCH models \cite{1} as well as their GARCH generalization \cite{2}, due to their potential to cover the stylized facts of such time series. However, their drawbacks kept on appearing because of the restrictive conditions imposed on the parameters \cite{2,3,4}. The non-negativity of the conditional variance still among the problems poorly addressed by the volatility specification as given by the standard GARCH model.

The log-GARCH model is one of the GARCH extensions address this issue by modeling log-volatility rather than volatility so that the parameters are not a priori subject to non-negativity constraints. Early works introduced the log-GARCH model are those of Geweke \cite{5}, Pantula and Sastry \cite{6}, Milhøj \cite{7}, while the first generalization of log-GARCH is due to Sucarrat and Escribano \cite{8} whereby the so-called power log-GARCH models have been proposed as a general class of log-GARCH models. The standard form of the log-GARCH(\textit{p,q}) model for a process \(\varepsilon_t\) having a conditional variance \(\sigma_t^2\), is given as follows

\begin{equation}
\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim \text{iid}(0,1), \quad \text{Prob}(\eta_t = 0) = 0
\end{equation}

\begin{equation}
\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \log \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2
\end{equation}

Where \(\sigma_t > 0, \omega, \alpha_1, \ldots, \alpha_p \) and \(\beta_1, \ldots, \beta_q\) are real numbers.

In fact, the theoretical structure of the consistency and asymptotic normality of the QML estimation of the log-GARCH models has been established by Francq, Wintenberger and Zakoïan \cite{9} with an extension to the asymmetric case. In the same vein, Francq and Sucarrat \cite{10} proposed an exponential Chi-squared QML estimation for log-GARCH models via an ARMA
representation. This approach has led to the development of a more robust estimation of so-called log-GARCH-X models when covariates or other conditioning variables are added to the volatility equation and the conditional density is unknown [11]. An interesting part of the literature that overlaps with the log-GARCH models is the stochastic volatility models. Indeed, through such models, the log-GARCH can be estimated consistently via its ARMA representation [12] or using state space forms combined with Kalman filter [13].

The log-moment structure of the log-GARCH process has been deeply investigated by Francq, Wintenberger and Zakoïan [9] in a framework that involves symmetric and asymmetric cases. However, the finiteness of high moment orders is not empirically verified when looking at real financial returns which exhibit heavy tails so that fourth or even second moments may be infinite [14, 15, 16]. This explains for instance, the need of robust kurtosis measures that reflect the empirical tail properties of financial returns. Actually, an interesting attention has been given to the non-robustness of the conventional measure of kurtosis given by the fourth standardized moment. Its terrible disadvantages is its sensitiveness to the presence of a single outlier, let alone that financial returns are not compatible with a Gaussian distribution. Related works investigating of alternatives to the conventional kurtosis measure include Siddiqui and Raghunandanan, Hogg, Moors and Kim. With this mind, Fiori proposed a convenient framework for the construction of empirical estimators so called right and left kurtosis which exhibit suitable properties of consistency and asymptotic normality under mild moment requirements.

It is worth noting that one of the weaknesses of the log-GARCH model is precisely its logarithmic specification which raises the problem of zero actual returns. Indeed, even with the assumption \( P(\eta_t = 0) = 0 \) as given in [1], zeros can be observed in the presence of some data issues, e.g. missing values or/and discreteness approximation error. As discussed by Sucarrat and Escribano, such a situation leads to asymptotically biased QML estimates. Francq, Wintenberger and Zakoïan rely on the assumption of zero probability of zero returns, namely \( P(\eta_t = 0) = 0 \) and multiply each ARCH coefficient \( \alpha_i \), for \( i = 1, \ldots, p \), by \( 1_{\{\eta_{t-i} = 0\}} \) so that the zeros of observed return \( \epsilon_{t-i} \) don’t occur in the recursion in the log-conditional variance equation. In another contribution using an ARMA representation of the log-volatility specification [2], Sucarrat and Escribano proposed a solution treats zeros as missing values using imputation by EM algorithm. Sucarrat and Grønneberg addresses the case when the assumption of zero probability of zero returns is omitted.

This work is focused on two axes. The first one deals with the log-GARCH parameter estimation by quasi-maximum likelihood without any assumptions on the existence of log-moments greater than one. To this end, the log-GARCH model is above all represented in a state space form. Afterwards, the log-conditional variance is estimated using the information filter under the stability condition ensuring the model stationary. Indeed, contrary to the standard Kalman filter, the information filter is useful for optimal estimation of the state variables in the absence of its initial covariance structure, which is effectively in agreement with the aim of estimating the log-conditional variance with no information about its covariance. Then, after plugging the estimated conditional variance, the quasi-likelihood associated with the log-GARCH model is evaluated and randomly maximized by Smoothed Functional algorithm. In the second axis, the proposed estimation method is used for two purposes. On the one hand, the presence of zero returns is treated as a missing value’s problem through imputation by the predicted estimation of log-GARCH observations using an EM algorithm. On the other hand, the conditional standard deviation estimate is used to standardize the log-GARCH residuals in order to compute the right and left measures of kurtosis.

A Monte Carlo study confirms that under mild conditions of moment existence, the QML estimation of the log-GARCH model based on the information filter has the potential to better fit such models or have at least the same estimation accuracy as existing methods. Furthermore,
our approach provides fitted conditional standard deviation which is able to reflect better the heavy tails behavior through the right and left measures of kurtosis. This is also the case in dealing with the problem of zero returns. Simulation results show that our approach is a considerable alternative for this kind of problem.

The rest of this paper proceeds as follows. Section 2 provides the state space representation of the log-GARCH\((p, q)\) model. Section 3 implements the information filter under mild conditions of moment finiteness. In Section 4, the quasi-maximum likelihood estimation is adopted combined with the smoothed functional algorithm as an optimization routine. Section 5 examines the performance of our approach in the parameter estimating, kurtosis measure and the treatment with the smoothed functional algorithm as an optimization routine. Section 6 concludes.

The following notations will be used throughout this paper. \(\mathcal{M}_{(k,l)}\) is the set of the matrices of size \((k,l)\) and \(0_{(k,l)}\) is its zero matrix. \(I_k\) is the identity matrix of \(\mathcal{M}_{(k,k)}\). \(t_{ek}\) is the first vector of the canonical base of \(\mathbb{R}^k\). For any matrix \(A\), \(Sp(A)\) stands for the set of eigenvalues of \(A\). For any sequence of random variables \((X_t)_{t \in \mathbb{Z}}\), \(X_t := (X_t, X_{t-1}, \ldots)\).

## 2. State space model

Taking the logarithm of \(\varepsilon^2_t\) in (1), the log-GARCH model is reformulated in the state space form as

\[
\begin{align*}
H_t &= \Omega + \Phi H_{t-1} + \nu_{t-1} \quad ; \quad \text{State equation} \\
Y_t &= e_r H_t + \xi_t \quad ; \quad \text{Measurement equation} \tag{3}
\end{align*}
\]

where

\[
\begin{align*}
H_t &= t (\log \sigma^2_t \ldots \log \sigma^2_{t-r+1}) \in \mathbb{R}^r, \\
Y_t &= \log \varepsilon^2_t \in \mathbb{R}, \\
\nu_t &= t \left( \sum_{i=1}^{p} \alpha_i \log \eta^2_{t-i+1}, 0_{(1,r-1)} \right) \in \mathbb{R}^r, \\
\xi_t &= \log \eta^2_t \in \mathbb{R}. \\
\Omega &= \left( \begin{array}{c} \omega \\ 0_{(r-1,1)} \end{array} \right) \in \mathbb{R}^r \quad \text{and} \quad \Phi := \left( \begin{array}{c} \phi_1 \\ \vdots \\ \phi_r \end{array} \right) \in \mathcal{M}_{r,r} \text{ with } \phi_i = \alpha_i + \beta_i \quad \text{such that} \quad \phi_i = \beta_i \quad \text{for } i > p \quad \text{and} \quad \phi_i = \alpha_i \quad \text{for } i < q.
\end{align*}
\]

Furthermore, in order to ensure the normality of the model (3) in conjunction with the normality of the model (1), several ways were proposed to derive/approximate the \(\log \eta^2_t\) distribution for the log-GARCH model (10) and also for some periodic autoregressive stochastic volatility representation (PAR-SV) (13). In this work, we proceed otherwise making use of the log-Normal approximation to the \(\chi^2\) distributions. Indeed, assuming that \(\eta_t \sim iid \mathcal{N}(0,1)\) implies that \(\eta^2_t \sim \chi^2_1\) with mean one and variance two. On the other side, assuming that \(\log \eta^2_t \sim iid \mathcal{N}(m_\xi, R)\) means that \(\eta^2_t \sim iid log - N(m_\xi, R)\), where \(m_\xi\) and \(R\) are related, by the method of moments, to the mean and variance of \(\chi^2_1\) distribution, by

\[
e^{m_\xi + \frac{R}{2}} = 1 \quad \text{and} \quad (e^R - 1)e^{2m_\xi + R} = 2,
\]

which yields

\[
m_\xi = -log\sqrt{3} \quad \text{and} \quad R = log 3,
\]

Therefore, \(\xi_t\) and \(\nu_{t-1}\) are uncorrelated and assumed Gaussian which are uncorrelated with \(H_0\), with respective parameters \((m_\xi, R)\) and \((m_{\nu}, Q)\) where
Thus, \((m_0, m_\xi, Q, R)\) can be seen as one among other parameterization of the state space model \((3)\) which ensure its normality and which differs according to the approximations that the \(\log \eta\) distribution undergo. In the sequel, the information filter is given as function of this parameterization. Investigating for the optimal parameterization of representation \((3)\) which provides the best information filtering beyond the goal of this paper.

3. Information filtering
Let \(\varepsilon_t\) be the log-GARCH\((p, q)\) model given by \((1)\). It is assumed that for all \(s > 1\) \(\mathbb{E}[\log \varepsilon_t^2]^s = \infty\). With this mind, we estimate the sample log-conditional variance using the information filter since it allows to initialize the inverse of the error covariance matrix

\[ P_t = \mathbb{E}\{(H_t - \mathbb{E}H_t)^\dagger (H_t - \mathbb{E}H_t)\} \]

rather than \(P_t\) being infinite. In the sequel, the inverse of \(P_t\) stands for the information matrix of the state space system \((3)\), denoted \(\mathcal{I}_t\). In practice, the information filter can be initialized as

\[ \mathcal{I}_0 = 0 \quad \text{and} \quad H_0 = \mathbb{E}H_1 = \frac{\omega + m_\xi \sum_{i=1}^p \alpha_i}{1 - \sum_{i=1}^r \phi_i} t(1 \ldots 1) \in \mathbb{R}^r, \]

since \(P_0\) is infinite and \(\mathbb{E}\log \sigma_t^2 < \infty\). Let \(\tilde{H}_t^+\) and \(\hat{H}_t^+\) denote respectively the predicted and the filtered estimates of \(H_t\) whereas \(\mathcal{I}_t\) and \(\mathcal{I}_t^+\) are their respective information matrices.

Following equations (5.1.1)-(5.1.10) in \cite{27}, from the observations \(Y_1, \ldots, Y_n\), the information filter equations applied to the model \((3)\) are obtained through the predicted information state and the filtered information state denoted respectively by \(\tilde{H}_t^+\) and \(\hat{H}_t^+\) for \(t = 1, \ldots, n\) as follows

\[
\begin{align*}
M_t &= \ t\Phi^{-1}\mathcal{I}_t^+\Phi^{-1}; \\
\mathcal{I}_t^- &= M_t - M_t(Q^{-1} + M_t)^{-1}M_t; \\
\mathcal{I}_t^+ &= \mathcal{I}_t^- + t e_{r} R^{-1} e_{r}; \\
K_t &= (\mathcal{I}_t^+)^{-1}t e_{r} R^{-1}; \\
\tilde{H}_t^+ &= \tilde{H}_t^- + t e_{r} R^{-1}(Y_t - m_\xi) ; \\
\hat{H}_t^+ &= (\mathcal{I}_t^+)^{-1}\tilde{H}_t^+; \\
N_t &= [I_r - M_t(Q^{-1} + M_t)^{-1}] t\Phi^{-1}; \\
\tilde{H}_t^- &= N_t [\tilde{H}_t^+ + \mathcal{I}_t^+\Phi^{-1}(\Omega + m_0)]; \\
\hat{H}_t^- &= (\mathcal{I}_t^-)^{-1}\tilde{H}_t^-.
\end{align*}
\]

However, the direct implementation of the information filter equations \((4)-(12)\) gives rise to two main problems. The first concerns the singularity of \(Q\) since \(Q^{-1}\) appears in the equations \((5)\) and \((10)\). To cope with that, we proceed as in \cite{27} whereby equation \((5)\) degenerates into \(\mathcal{I}_t = M_t\) whenever \(Q\) is singular. The second is related to the stability of the filter associated with the state space model \((3)\). Actually, the necessary and sufficient condition for stability is that all eigenvalues of \(\Phi\) should have modulus less than one \cite{28}. Recall that the stability condition ensures also the second order stationary of \(H_t\) since the system equations \((3)\) is time invariant (see Theorem 3.1, p.70 in \cite{29}). Then, under the stability condition, the log-GARCH parameters satisfy

\[ \sum_{i=1}^p \phi_i < 1. \]
Summing up, the final information filter is represented in the algorithm 3.

Algorithm 1 Information filtering

1: Initialization : $\hat{H}_0^{-} = \mathbb{E}H_0$ and $\mathcal{I}_0^{-} = 0$
2: for $t = 0 : n$ do
3: \hspace{1em} $\hat{H}_t^{-} = \mathcal{I}_t \hat{H}_t^{-}$
4: \hspace{1em} Compute $\mathcal{I}_t^{-}$ from (6)
5: \hspace{1em} Compute $\hat{H}_t^{+}$ from (8)
6: \hspace{1em} Compute $\hat{H}_t^{+}$ from (9)
7: \hspace{1em} Compute $M_t$ and $N_t$ respectively from (4) and (10)
8: \hspace{1em} if $\max|\text{Sp}(\Phi)| < 1$ then
9: \hspace{2em} if $\text{det}(Q) \neq 0$ then
10: \hspace{3em} Compute $\mathcal{I}_t^{-}$ from (5)
11: \hspace{3em} Compute $H_{t+1}^{-}$ from (11)
12: \hspace{2em} else
13: \hspace{3em} $\mathcal{I}_{t+1}^{-} = M_t$
14: \hspace{3em} $\hat{H}_{t+1}^{-} = t\Phi^{-1} \hat{H}_t^{+} + M_t(\Omega + m_\nu)$
15: \hspace{1em} end if
16: \hspace{1em} else
17: \hspace{2em} $\hat{H}_{t+1}^{-} = \hat{H}_t^{-}$
18: \hspace{1em} end if
19: end for

Now, given the predicted estimate $\hat{H}_t^{-}$, we should extract the conditional variance estimate, namely $\hat{\sigma}_t^2^{-} = \mathbb{E}(\sigma_t^2/\varepsilon_{t-1}^2)$, that will be plugged in the quasi-likelihood function. It is then obvious to estimate $\log(\sigma_t^2)$ by

$$e_r \hat{H}_t^{-} = \left( \log \sigma_t^2 \right)^{-} = \mathbb{E}(\log \sigma_t^2/\varepsilon_{t-1}^2). \quad (13)$$

But extracting the conditional variance estimate as $\exp(\log \sigma_t^2)^{-}$ is incorrect unless a multiplicative constant $C$ is introduced to minimize the bias due to using incorrectly the latter estimate. Then, as advocated by Aknouche [13], we take $C$ as the predicted variance of $\log \sigma_t^2$, namely, the inverse of the first diagonal element of the predicted information matrix, denoted $(\mathcal{I}_t^{-})_{11}^{-1}$. It follows that

$$\hat{\sigma}_t^2^{-} = (\mathcal{I}_t^{-})_{11}^{-1} \exp(\log \sigma_t^2)^{-}. \quad (14)$$

4. Estimating the log-GARCH parameters

We now estimate the log-GARCH($p, q$) parameters by quasi-maximum likelihood. Let $\varepsilon_1, \ldots, \varepsilon_n$ be observations of the log-GARCH process (1)-(2) of an unknown parameter vector $\theta = t(\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \in \Theta$, where

$$\Theta = \{ \theta \in \mathbb{R}^d/\max|\text{Sp}(\Phi(\theta))| < 1 \} \text{ with } d = p + q + 1.$$

Given the log-conditional variance estimate $(\log \sigma_t^2)^{-}$ as in (13) and the conditional variance estimate $\hat{\sigma}_t^2^{-}$ as in (14) and , a quasi-maximum likelihood estimate of $\theta$, denoted $\hat{\theta}_{QIF}$, is defined
as
\[ \hat{\theta}_{QIF} = \arg \min_{\theta \in \Theta} \hat{l}_n(\theta), \]
where
\[ \hat{l}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\epsilon_i^2}{\hat{\sigma}_i^2} + \left( \log \hat{\sigma}_i^2(\theta) \right)^{-}. \] (15)

However, it is important to note that the minimization routine of \( \hat{l}_n \) must take into account two of its main properties. First, there is some randomness in the measurements of \( \hat{l}_n \) arising from the pre-estimating of the log-conditional variance before to be plugged in \( \hat{l}_n \). Second, we are interested in implementing a search algorithm that sidesteps the use of the high derivatives of \( \hat{l}_n \) (e.g. gradient, Hessian) but instead, it uses its approximations in the parameter update equation. Such approximations led to a random decision made in the search direction as the algorithm iterates toward a minima. For these reasons, a stochastic search using the simultaneous perturbation algorithms is a wise choice to minimize \( \hat{l}_n \) [30]. In this work we focus on the smoothed functional algorithm [31, 32, 33], henceforth denoted SF, that belongs to the class of simultaneous perturbation methods. The idea is to approximate the gradient of \( \hat{l}_n \) by its convolution with a Uniform density that ensures a fast convergence of the algorithm. Thus, we can summarize the SF algorithm applied to \( \hat{l}_n \) as follows:

**Algorithm 2 SF algorithm**

1: Calibration: \((a, b, \lambda, \gamma) = (0.2, 1.5, 0.602, 0.101), A = 0.1 \times \text{Iter}, \) and some \( \epsilon > 0 \)
2: Initialization : \( \theta_0 \in \Theta \)
3: for \( k = 1 : \text{Iter} \) do
4: \( a_k = a(A + k + 1)^{-\lambda} \) and \( b_k = b(k + 1)^{-\gamma} \)
5: Draw \( \Delta_k \sim iid U([-1, 1]), \Delta_k \in \mathbb{R}^d \)
6: Draw \( \delta_{k,1}, \delta_{k,2} \sim iid U[0,1] \)
7: if \( \max|Sp(\Phi(\hat{\theta}_k))| < 1 \) then
8: \( y_{k,1}(\hat{\theta}_k) = \hat{l}_n(\hat{\theta}_k + b_k \Delta_k) + \delta_{k,1} \) and \( y_{k,2}(\hat{\theta}_k) = \hat{l}_n(\hat{\theta}_k - b_k \Delta_k) + \delta_{k,2} \)
9: \( \hat{g}(\hat{\theta}_k) = \frac{y_{k,1}(\hat{\theta}_k) - y_{k,2}(\hat{\theta}_k)}{2b_k} \Delta_k \)
10: \( \hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}(\hat{\theta}_k) \)
11: if \( \left| \hat{\theta}_k - \hat{\theta}_{k-1} \right| < \epsilon \) or \( k = \text{Iter} \) then
12: return \( \hat{\theta}_k \)
13: else \( k = k + 1 \)
14: end if
15: else \( k = k - 1 \) and return to (6)
16: end if
17: end for

Remark 4.1.

(i) Contrary to other simultaneous perturbation methods that provide local minima, the smoothing of the objective function helps the SF algorithm to converge to a global minimum or to a point close to it [33].

(ii) Other density functions than the Uniform can also be used for the perturbation random variables, e.g. the Gaussian and the Cauchy [34]. Moreover, Board [35] established general conditions on candidate density, to be used as a smoothing function.
(iii) A convergence analysis of the SF algorithm was investigated by Board [35] ensuring the convergence almost surely of the iterates to a neighborhood of the minima.

It’s worth noting that the information filter approach gives another way to deal with the problem of the presence of zero returns whose logarithms are replaced by the conditional expectation $E \left( \log \frac{\epsilon_t^2}{\hat{\epsilon}^2_{t-1}} \right)$ [25]. In the same vein, we see that the same conditional expectation can be obtained from the information filter in combination with the EM algorithm as the predicted measurement estimate by the information filter, namely $\hat{Y}_t^-$. Thus, if $Y_t^*$ denotes the logarithm of the zero log-GARCH observation found at stage $t$, the imputation process is implemented as follows

$$
Y_t^* = \begin{cases} 
Y_1 = \frac{\omega + m\xi \left( 1 - \sum_{j=1}^{q} \beta_j \right)}{1 - \sum_{i=1}^{r} \phi_i} \\
Y_t^* = \hat{Y}_t^- (\hat{\theta}_{t-1}) = e_r \Delta H_t \left( \hat{\theta}_{t-1} \right) + m\xi, \quad \text{for } t > 1 
\end{cases}
$$

(16)

Where $\hat{\theta}_{t-1}$ is the QIF estimate obtained from observations $\{Y_1, \ldots, Y_{t-1}\}$.

5. Monte Carlo experiment

In this section, we investigate the performance of the QIF estimation for the log-GARCH(1,1) model with an infinite second log-moment. We focus on two stages. The first assesses the finite sample performance of the QIF estimation compared to the QML estimation [9] and the exponential Chi-squared QML estimation [10]. The second targets to evaluate the robustness of the Right and Left kurtosis measures, henceforth called R-L measures, applied to the log-GARCH observations standardized by the conditional standard deviation derived from (14). We conduct a Monte Carlo simulation for a range of sample sizes $n \in \{500, 1000, 5000\}$ starting from a log-GARCH(1,1) model of parameter $\theta_0 = (1, 0.8, -0.5)$, with $\eta_i \sim iid N(0, 1)$. Notice that with this parameter setting, the process is stationary with an infinite second log-moment. For each size, we have generated 1000 replications of log-GARCH(1,1) series. Mean of estimates, denoted $\hat{\theta}_{QML}$ (QML), $\hat{\theta}_{ECQ}$ (exponential Chi-squared QML) and $\hat{\theta}_{QIF}$ (QIF), their standard deviations and their mean squared errors (MSE) over the 1000 replications are displayed in table [1]. By default, No zero returns were generated over all replications.

In the same manner, we rely on the consistency of the R-L measures of kurtosis to evaluate them by a simulation study. Thus, from the return residuals standardized by the conditional standard deviation derived from (14) given by $\hat{e}_t^* = \frac{-\tilde{e}_t}{\sqrt{\hat{\sigma}_t^2}}$, the computational forms of the R-L kurtosis [24], respectively denoted by $\hat{R}_{QIF}$ and $\hat{L}_{QIF}$, are to be

$$
\hat{R}_{QIF} = \frac{n^{-1} \sum_{i=k+1}^{n} \left( \frac{4i-2}{n} - 3 \right) \hat{e}_t^* - \left( \frac{k}{n} - 0.5 \right) (n-1) \hat{e}_t^*}{n^{-1} \sum_{i=k+1}^{n} \hat{e}_t^* + \left( \frac{k}{n} - 0.5 \right) \hat{e}_t^* - 0.5 F_n^{-1}(0.5)}
$$

and

$$
\hat{L}_{QIF} = \frac{n^{-1} \sum_{k+1}^{n} \left( \frac{4i-2}{n} - 3 \right) \hat{e}_t^* - \left( \frac{k}{n} - 0.5 \right) (n-3) \hat{e}_t^*}{0.5 F_n^{-1}(0.5) - n^{-1} \sum_{i=1}^{k} \hat{e}_t^* + \left( \frac{k}{n} - 0.5 \right) \hat{e}_t^*}
$$

where $k = \lfloor n/2 \rfloor$, $\hat{e}_t^*$ is the order statistic related to the residuals $\tilde{e}_t^*$ and $F_n^{-1}(0.5)$ stands for its empirical Median.
The R-L measures are then denoted \( \tilde{K}_{QIF} = (\hat{R}_{QIF}, \hat{L}_{QIF}) \). The analogue R-L measures for the QML and ECQ estimations are denoted respectively by \( \tilde{K}_{QML} = (\hat{R}_{QML}, \hat{L}_{QML}) \) and \( \tilde{K}_{ECQ} = (\hat{R}_{ECQ}, \hat{L}_{ECQ}) \) and they are computed on the basis of the associated residual returns standardized by their respective conditional standard deviations.

Mean of the R-L kurtosis estimates \( \tilde{K} \), their standard deviations and their mean absolute errors (MAE) over the same 1000 replications are reported in table (2). Notice that at this stage, the MAE is used instead of MSE in order to avoid the effect of outliers in the evaluation of the kurtosis estimates.

The first broad reading of table (1) shows that the results of QIF and ECQ estimations are very close to each other in terms of MSE, compared to the QML estimation. Indeed, for the both estimates, the MSE values are simultaneously in the order of \( 10^{-4} \) with a small superiority of the first for \( n = 500, 1000 \). More precisely, for all sample sizes \( n \), it can be seen that the intercept \( \omega \) is accurately estimated by QIF which arises from the double-step estimation used in the ECQ estimation (the intercept \( \omega \) is estimated in the second step after estimating the other parameters \([10]\)). However, for a large sample size, i.e. \( n = 5000 \), the three estimates seem relatively more close to each other. Furthermore, among all estimates, the small values of standard deviations of \( \hat{\theta}_{QIF} \) indicate that estimates over all replications are dispersed closely to \( \hat{\theta}_{QIF} \) and widely for \( \hat{\theta}_{QML} \) and \( \hat{\theta}_{ECQ} \).

### Table 1. Finite sample properties of the log-GARCH(1,1) fitted by QML, ECQ and QIF estimation. The estimated standard deviations are displayed in brackets. * represents the smallest MSE.

| \( n \) | \( \theta_0 \) | \( \hat{\theta}_{QML} \) | MSE | \( \hat{\theta}_{ECQ} \) | MSE | \( \hat{\theta}_{QIF} \) | MSE |
|---|---|---|---|---|---|---|---|
| 500 | 1 | 1.0068 (0.1115) | 0.0123 | 0.9888 (0.1037) | 0.0107 | 1.0635 (0.0312) | 0.0050* |
| | 0.8 | 0.8039 (0.0435) | 0.0018 | 0.7971 (0.0302) | 0.0009 | 0.8110 (0.0258) | 0.0007* |
| | −0.5 | −0.5025 (0.0548) | 0.0029 | −0.4961 (0.0333) | 0.0011* | −0.5376 (0.0449) | 0.0020 |
| 1000 | 1 | 1.0011 (0.0782) | 0.0060 | 0.9990 (0.0721) | 0.0051 | 1.0013 (0.0269) | 0.0007* |
| | 0.8 | 0.8004 (0.0332) | 0.0010 | 0.8017 (0.0202) | 0.0004* | 0.7942 (0.0199) | 0.0004* |
| | −0.5 | −0.4945 (0.0432) | 0.0018 | −0.4961 (0.0246) | 0.0006 | −0.5013 (0.0218) | 0.0004* |
| 5000 | 1 | 1.0068 (0.0339) | 0.0011 | 1.0069 (0.0297) | 0.0009 | 0.9993 (0.0271) | 0.0007* |
| | 0.8 | 0.7984 (0.0141) | 0.0001 | 0.7990 (0.0096) | 0.0001* | 0.7944 (0.0120) | 0.0001 |
| | −0.5 | −0.5003 (0.0167) | 0.0002* | −0.5003 (0.0213) | 0.0006 | −0.5013 (0.0263) | 0.0006 |

Concerning finite properties of the R-L kurtosis estimates, the smallest values of the MAE for all samples confirm that the \( \tilde{K}_{QIF} \) outperforms the both estimates \( \tilde{K}_{QML} \) and \( \tilde{K}_{ECQ} \). Especially, the performance of \( \tilde{K}_{QIF} \) improves markedly for a large sample size, namely \( n = 5000 \) according to the MAE values that reaches the order of \( 10^{-3} \). More interestingly, knowing that for a Gaussian distribution, the R-L kurtosis measures are both closes to \( \sqrt{2} - 1 \simeq 0.4142 \), it is clear that the estimated \( \tilde{K}_{QIF} \) measures are robust to the behaviour of heavy tails arising from the non-finiteness of the fourth moment. On the contrary, the values of \( \tilde{K}_{QML} \) and \( \tilde{K}_{ECQ} \) remain close to those of a Gaussian distribution and significantly different from the real values corresponding to the simulated series.

To study the properties of the imputation method proposed in \([16]\) we run a Monte Carlo
Table 2. Finite sample properties of the Right and Left measures of kurtosis computed by QML, ECQ and QIF estimation. The estimated standard deviations are displayed in brackets. * represents the smallest MAE.

| n    | $\hat{K}_0$ | $\hat{K}_{QML}$ (MAE) | $\hat{K}_{ECQ}$ (MAE) | $\hat{K}_{QIF}$ (MAE) |
|------|--------------|------------------------|------------------------|------------------------|
| 500  | 0.5242       | 0.4171 (0.0169)        | 0.4159 (0.0170)        | 0.5403 (0.0203)        |
|      | 0.5367       | 0.4120 (0.0158)        | 0.4108 (0.0157)        | 0.5413 (0.0260)        |
| 1000 | 0.5116       | 0.4153 (0.0129)        | 0.4145 (0.0128)        | 0.5425 (0.0164)        |
|      | 0.5484       | 0.4151 (0.0104)        | 0.4146 (0.0103)        | 0.5407 (0.0184)        |
| 5000 | 0.5467       | 0.4127 (0.0055)        | 0.4126 (0.0055)        | 0.5393 (0.0067)        |
|      | 0.5424       | 0.4144 (0.0049)        | 0.4143 (0.0049)        | 0.5394 (0.0084)        |

Table 3. Finite sample properties of the log-GARCH(1,1) fitted by QML, ECQ and QIF estimation for different non-zero probability of zero returns.

| $\pi_0$ | $\theta_0$ | $\hat{\theta}_{QML}$ Bias | $\hat{\theta}_{ECQ}$ Bias | $\hat{\theta}_{QIF}$ Bias |
|---------|-------------|----------------------------|-----------------------------|-----------------------------|
| 0.05    | 1           | 1.2156 0.2156              | 1.1853 0.1853               | 1.1821 0.1821               |
|         | 0.8         | 0.7770 −0.0230             | 0.7610 −0.0390              | 0.7309 −0.0691              |
|         | −0.5        | −0.5081 −0.0081            | −0.4637 0.0363              | −0.4813 0.0187              |
| 0.1     | 1           | 1.2018 0.2018              | 1.1458 0.1458               | 1.1461 0.1461               |
|         | 0.8         | 0.7593 −0.0407             | 0.7483 −0.0517              | 0.7482 −0.0518              |
|         | −0.5        | −0.4922 0.0078             | −0.5001 −0.0001             | −0.4997 0.0003              |
| 0.2     | 1           | 1.3513 0.3513              | 1.2152 0.2152               | 1.2150 0.2152               |
|         | 0.8         | 0.7069 −0.0931             | 0.6344 −0.1656              | 0.6340 −0.1660              |
|         | −0.5        | −0.3538 0.1462             | −0.3683 0.1317              | −0.3680 0.1320              |

experiment similar to those in the previous experiments with $n = 10000$ split into 100 replications. Let $\pi_0$ denotes the proportion of zero returns which we set at 5%, 10%, and 20% of the sample size. Table 3 displays the results and figure 1 represents the finite sample bias graphically. Visual inspection of figure 1 shows that except for $\alpha$ estimates, the finite sample biases of $\omega$ estimates and $\beta$ estimates increase somewhat as the zero probability increases which is expected since actual observations are lost by treating zeros as missing values. More specifically, we note that, given the close values of $\hat{\omega}_{ECQ}$ and $\hat{\omega}_{QIF}$ biases, it can be said that the two corresponding approaches have kept the same level of volatility, unlike $\hat{\omega}_{QML}$, where the effect of zero returns is more significant. For $\alpha$, all three imputation methods decreased the bias with a lower level for the ECQ and the QIF imputations. Thus, the persistence of the shock is not strongly impacted. For $\beta$, the bias for all methods increases slightly compared to the other parameters while keeping the same level with the lowest slopes.
6. Conclusion
Under mild conditions of moment existence, we have proposed a quasi-maximum likelihood estimation of the log-GARCH model via the information filter (QIF) with the smoothed functional as an optimization routine. Our method was used for the imputation of zero returns considered as missing values through an EM algorithm while the fitted conditional standard deviations were used to standardize the log-GARCH residuals in order to compute the right and left measures of kurtosis. In finite samples, the QIF estimation has the potential to better fit the log-GARCH model and also to deal with the problem of zero returns, or to have at least the same estimation accuracy as existing methods. Furthermore, our approach provides fitted conditional standard deviation which combined with the right and left measures of kurtosis, is able to reflect better the heavy tails behavior.

References
[1] Engle R F 1982 Econometrica: Journal of the Econometric Society 987–1007
[2] Bollerslev T 1986 Journal of econometrics 31 307–327
[3] Nelson D B and Cao C Q 1992 Journal of Business & Economic Statistics 10 229–235
[4] Tsai H and Chan K S 2008 Econometric Theory 24 823–828
[5] Geweke J 1986 Econometric Reviews 5 57–61
[6] Pantula S G 1986 Econometric Reviews 5 79–97
[7] Milhøj A 1987 A multiplicative parameterization of ARCH models
[8] Sucarrat G and Escribano A 2010
[9] Francq C, Wintenberger O and Zakoian J M 2013 Journal of Econometrics 177 34–46
[10] Francq C and Sucarrat G 2013
[11] Sucarrat G, Grønneberg S and Escribano A 2016 Computational statistics & data analysis 100 582–594
[12] Francq C and Zakoian J M 2006 Scandinavian Journal of Statistics 33 785–806
[13] Aknouche A 2017 Statistical Inference for Stochastic Processes 20 139–177
[14] Anatolyev S 2019 Dependence Modeling 7 1–23
[15] Bai J and Ng S 2005 Journal of Business & Economic Statistics 23 49–60
[16] Cont R 2001
[17] Siddiqui M and Raghunandan K 1967 Journal of the American Statistical Association 62 950–953
[18] Hogg R V 1972 Journal of the American Statistical Association 67 422–424
[19] Hogg R V 1974 Journal of the American Statistical Association 69 909–923
[20] Moors J 1988 Journal of the Royal Statistical Society: Series D (The Statistician) 37 25–32
[21] Kim T H and White H 2004 Finance Research Letters 1 56–73
[22] Fiori A M 2008 Communications in Statistics—Theory and Methods 37 2665–2680
[23] Fiori A M and Beltrami D 2012 Asymptotic estimation of right and left kurtosis measures, with applications to finance *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society* (Citeseer)
[24] Fiori A M and Beltrami D 2014 *Stat* 3 95–108
[25] Sucarrat G and Escribano A 2013
[26] Sucarrat G and Grønneberg S 2016
[27] Brown R G 1983 *Introduction to random signal analysis and Kalman filtering* vol 8 (Wiley New York)
[28] Harvey A C 1990 *Forecasting, structural time series models and the Kalman filter* (Cambridge university press)
[29] Anderson B D and Moore J B 2012 *Optimal filtering* (Courier Corporation)
[30] Spall J C 1998 *IEEE Transactions on aerospace and electronic systems* 34 817–823
[31] Chin D C 1997 *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 27 244–249
[32] Katkovnik V Y and OY K 1972 *Automation and Remote Control* 33 1321–1326
[33] Styblinski M and Tang T S 1990 *Neural Networks* 3 467–483
[34] Rubinstein R Y and Kroese D P 2016 *Simulation and the Monte Carlo method* vol 10 (John Wiley & Sons)
[35] Board S A 2013