On the treatment of zero returns in the estimation of log-GARCH model: Empirical study

Abdeljalil Settar\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: abdeljalilsettar@gmail.com

Abstract. This paper proposes a QML estimation of log-GARCH model in the presence of zero returns. Our approach consists in treating zeros within the data as missing values which are objects of imputation based on the EM algorithm and information filter. Empirical applications to the Crude Oil WTI, CAC40 and the FTSE100 return series are presented and show that through the proposed imputation, the log-GARCH model fits better the data compared to the traditional approach whereby the estimation is performed after removing the zeros from the data.

1. Introduction
In the financial literature, the occurrence of zero returns is often considered as a measure of liquidity in the sense that the more frequent these returns are, the lower the liquidity (Lesmond, Ogden and Trzcinka \cite{1}). Zero returns will therefore occur when traders do not consider the information available to them as important enough to cover trading costs. See Bekaert, Harvey and Lundblad \cite{2}, Goyenko, Holden and Trzcinka \cite{3}, Levine and Schmukler \cite{4} and Mazza \cite{5} for a more detailed discussion.

In continuous time finance framework, Bandi, Pirino and Reno \cite{6} reject the hypothesis that asset price traded in a frictionless market is a semi martingale which implies that returns must be lowered by a threshold with a high probability. They claim, on the contrary, that asset prices are stale in the sense that they show a large incidence of zero returns, or more generally, small returns. In this spirit, Kolokolov, Livieri and Pirino \cite{7} propose a non-parametric theory for statistical inferences on zero returns of high-frequency asset prices.

Moreover, the probability of an observed zero return may change and depend on market condition. Hausman, Lo and MacKinlay \cite{8} propose a probit framework allowing this probability to depend on some conditioning variables like volume, duration and past returns. For more details see Engle and Russell \cite{9}, Russell and Engle \cite{10}, Bien, Nolte and Pohlmeier \cite{11} and Rydberg and Shephard \cite{12}. Sucarrat and Grønneberg \cite{13} establish a new class of financial return model with time-varying conditional probability of zero return combined with standard models of return volatility, e.g. GARCH and Stochastic Volatility (SV).

Actually, when faced with a return series containing zeros, the log-GARCH model is by construction inappropriate to be used to fit such a data because of the logarithmic specification of volatility as a function of the past squared returns logarithm which produces ”log-zero returns”
terms. Indeed, the standard form of the log-GARCH\((p,q)\) model for a process \(\varepsilon_t\) having a conditional variance \(\sigma_t^2\), is given as follows (Geweke [14], Pantula [15] and Milhøj [16])

\[
\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim iid(0,1) \tag{1}
\]

\[
\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \log \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2 \tag{2}
\]

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

Even with the zero probability assumption, namely \(Prob(\eta_t = 0) = 0\), considered by Francq, Wintenberger and Zakoian [17] to avoid the theoretical presence of zero returns, they can be empirically observed because of some data issues, e.g. missing values or discrete approximation error which lead to asymptotically biased QML estimates (Sucarrat and Escribano [18]). Thus, in order to perform a QML estimation of log-GARCH parameters, Francq, Wintenberger and Zakoian [17] rely on the assumption of zero probability and multiply each ARCH coefficient \(\alpha_i\), for \(i = 1, \ldots, p\), by \(\{\eta_{t-i} = 0\}\) so that the zeros of observed return \(\varepsilon_{t-i}\) don’t occur in the recursion in the log-conditional variance equation (2). In another contribution to the QML estimation of log-GARCH, Sucarrat and Escribano [18] proposed an approach treats zeros as missing values using imputation by EM algorithm whereby the log-zero returns are replaced by the conditional expectation \(E(\log \varepsilon_{t}^{2}/\varepsilon_{t-1}^{2}, \varepsilon_{t-2}^{2}, \ldots)\) estimated through an ARMA representation.

In fact, the approach of Sucarrat and Escribano [18] gives rise to the investigation of other ways of exploiting the EM algorithm for the estimation of the log-GARCH parameters applied to return data containing zeros. In general, the EM algorithm (Dempster, Laird and Rubin [19], Roderick and Donald [20], Schafer [21], McLachlan and Krishnan [22]) is an iterative algorithm to compute maximum likelihood estimates in parametric models for incomplete data. It consists in starting from initial parameter value and repeating in turn two steps E (Expectation) and M (Maximization) until convergence. Furthermore, state space representations and associated Kalman filters, are adequate tools to facilitate the running of the EM algorithm. The theoretical framework of such a mixture can be found in Brockwell, Davis and Fienberg [23] with application to ARCH model in Bahamonde [24].

In this paper, we propose to deal with zero returns as missing values that are filled in by imputation method. The goal is to provide QML estimates of log-GARCH parameters through an EM algorithm. Indeed, we reformulate the log-GARCH model in state space representation in order to implement the information filter. This last allows i) to obtain the estimated conditional variance which is used to evaluate the quasi-likelihood, and ii) to compute the predicted estimate of measurements corresponding to the log-zero returns, considered as imputation routine.

The benefit of our approach lies in the use of the information filter whereby, there is no requirement for existence of log-moment returns in excess of the first order. In other words, we assume that no information about \(E[\log \varepsilon_{t}^{2}|s]\) is available for \(s > 1\). Worth note that such relaxation is not allowed in the case of EM algorithms based on the Kalman filter since this last requires the existence of the second order moment of the state variable for the initialization and recursive computation of the error covariance matrix. Empirical applications are conducted to evaluate the performance of the proposed algorithms in terms of QML estimate bias and goodness of fit with respect to different sample sizes and proportions of zeros.

The rest of this paper proceeds as follows. Section 2 provides a state space representation associated with the log-GARCH\((p,q)\) model. Section 3 implements the information filter under mild conditions of log-moment finiteness. In Section 4, the EM algorithm combining the information filter and QML estimation is presented. Section 5 examines the performance of our approach for three cases of return series. Finally, 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 $\mathcal{M}_{(k,k)}$. $t e_k$ 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 $A$.

2. State space representation

We are inspired by the state space representation proposed by Aknouche [25] and adapt it to the log-GARCH specification. Let’s then transform the model (1)-(2) into additive form by log-transforming the square of the equation (1). Thus, the log-GARCH($p,q$) can be reformulated as a state space model of state variable $H_t = (log \sigma_t^2 \ldots log \sigma_{t-r+1}^2) \in \mathbb{R}^r$, with $r = max(p,q)$, and measurement variable $Y_t = log \varepsilon_t^2 \in \mathbb{R}$, such that:

$$
\begin{align*}
H_t &= \Omega + \Phi H_{t-1} + \nu_{t-1} \\
Y_t &= e_r H_t + \xi_t
\end{align*}
$$

where $\Omega = t(\omega, 0_{(r-1,1)}) \in \mathbb{R}^r$ and $\Phi = \left( \begin{array}{c} \phi_1 \\ \vdots \\ \phi_r \end{array} \right) \in \mathbb{R}^{r \times r}$ are the transition matrices, with $\phi_i = \alpha_i + \beta_i$ such that $\phi_i = \beta_i$ for $i > p$ and $\phi_i = \alpha_i$ for $i > q$.

The state noise $\nu_{t-1}$ and the measurement noise $\xi_t$ are given by:

$$
\nu_t = t \left( \sum_{i=1}^{p} \alpha_i log \eta_{t-i+1}^2, 0_{(1,r-1)} \right) \text{ and } \xi_t = log \eta_t^2
$$

Clearly, $\nu_{t-1}$ and $\xi_t$ are uncorrelated as functions of $\eta_t$ being white noise.

According to the Gaussian structure of the model (1)-(2) assuming that $\eta_t$ is Gaussian, the parameters $m_\nu := \mathbb{E} \nu_t$, $Q := \text{Cov}(\nu_t)$, $m_\xi := \mathbb{E} \xi_t$ and $R := \text{Cov}(\xi_t)$ are determined so that the noises $\nu_t$ and $\xi_t$ are Gaussian. Indeed, $\eta_t \sim iid \mathcal{N}(0,1)$ as given in (1) implies that $\eta_t^2 \sim \chi_1^2$ with mean one and variance two. Since $\xi_t = log \eta_t^2 \sim iid \mathcal{N}(m_\xi, R)$ means that $\eta_t^2 \sim iid log -N(m_\xi, R)$, one can relate $m_\xi$ and $R$ to the mean and the variance of $\chi_1^2$ distribution by the method of moments as follows:

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

which yields

$$m_\xi = -log \sqrt{3} \text{ and } R = log 3,$$

Hence, $\nu_t$ is also Gaussian of parameters

$$m_\nu = t \left( \sum_{i=1}^{p} \alpha_i, 0_{(1,r-1)} \right) \in \mathbb{R}^r \text{ and } Q = \left( \begin{array}{ccc} R \sum_{i=1}^{p} \alpha_i^2 & 0 & \cdots \\ 0_{(r-1,r)} & 0 \end{array} \right) \in \mathcal{M}_{r,r}.$$

Note that knowing of noise means, Kalman filter performs better than assuming that noises are zero mean which corresponds to the case where the true distribution of the noises is unknown. Thus, Kalman filter equations still hold involving the noise means.
3. Information filtering

It is well known that Kalman filter requires the existence of second order moment of the state variable for the initialization and recursive computation of the error covariance matrix. In order to relax the filtering conditions, we pre-estimate the log-conditional variance $H_t$ generated by the log-GARCH$(p,q)$ via the information filter allowing to initialize the inverse of the error covariance matrix $P_t = E \{ (H_t - \mathbb{E}H_t)^t(H_t - \mathbb{E}H_t) \}$ rather than $P_t$ being unknown.

Consequently, the inverse of $P_t$ represents the information matrix of the state space system (3) that we denote $I_t$. In practice, we consider that $P_0$ is infinite and that $E|H_t| < \infty$ to initialize the information filter as follows:

$$I_0 = 0 \quad \text{and} \quad H_0 = \mathbb{E}H_1 = \frac{\omega + m\xi \sum_{i=1}^{p} \alpha_i \phi}{1 - \sum_{i=1}^{r} \phi_i} (1 \ldots 1) \in \mathbb{R}^r$$

Let $\hat{H}_{t/-1}$ and $\hat{H}_{t/t}$ be respectively the predicted and filtered estimates of $H_t$. $I_{t/-1}$ and $I_{t/t}$ are their respective information matrices.

From observations $Y_1, \ldots, Y_n$, the information filter equations applied to the model (3) are obtained by inverting the Kalman filter equations (4), (6), (11), (13) and (14) associated with the same model through the predicted information state and the filtered information state noted respectively $\tilde{H}_{t/-1}$ and $\tilde{H}_{t/t}$.

Indeed, from the updating equation of the filtered error covariance, we have

$$P_{t/t} = (I_r - K_{t}e_r)P_{t/-1}$$

Then

$$K_{t}e_r P_{t/-1} = P_{t/-1} - P_{t/t}$$

Recall that the Kalman gain is given by :

$$K_t = P_{t/-1} t e_r (e_r P_{t/-1} t e_r + R)^{-1}$$

Then

$$K_t (e_r P_{t/-1} t e_r + R) = P_{t/-1} t e_r$$

By substituting (5) into (7), we get the Kalman gain as :

$$K_t = P_{t/t} t e_r R^{-1}$$

Let’s move on to the filtered information update equation.

Equation (5) implies

$$I_r - K_{t}e_r = P_{t/t} P_{t/-1}$$

By substituting (8) into (9), we get

$$P_{t/-1}^-1 = P_{t/-1}^-1 + t e_r R^{-1} e_r$$

Hence,

$$I_{t/t} = I_{t/-1} + t e_r R^{-1} e_r$$
Concerning equation for updating the predicted information, we invert the update equation of filtered error covariance:

\[ P_{t/t-1} = \Phi P_{t/t}^t \Phi + Q \]  

using the matrix inverse formula:

\[ (A + B)^{-1} = A^{-1} - A^{-1}B(I + A^{-1}B)^{-1}A^{-1} \]

With

\[ A = \Phi P_{t/t}^t \Phi \quad \text{and} \quad B = Q \]

It follows

\[ P_{t/t-1}^{-1} = (\Phi P_{t/t}^t \Phi + Q)^{-1} \]

\[ = M_t - M_t Q(I_r + M_t Q)^{-1}M_t \]

With

\[ M_t = (\Phi P_{t-1/t-2}^t \Phi)^{-1} \]

Thus,

\[ \mathcal{I}_{t/t-1} = M_t - M_t (Q^{-1} + M_t)^{-1}M_t \]  

As for the predicted and filtered estimates of \( H_t \), the first can be derived from the update equation of the filtered state:

\[ \hat{H}_{t/t} = \hat{H}_{t/t-1} + K_t(Y_t - e_r \hat{H}_{t/t-1} - m_\xi) \]  

which implies

\[ \mathcal{I}_{t/t} \hat{H}_{t/t} = \mathcal{I}_{t/t} \hat{H}_{t/t-1} + \mathcal{I}_{t/t} K_t(Y_t - e_r \hat{H}_{t/t-1} - m_\xi) \]

\[ \hat{H}_{t/t} = \mathcal{I}_{t/t} \hat{H}_{t/t-1} + t e_r R^{-1}(Y_t - e_r \hat{H}_{t/t-1} - m_\xi), \quad \text{from [8]} \]

\[ \tilde{H}_{t/t} = \mathcal{I}_{t/t} \hat{H}_{t/t-1} + t e_r R^{-1}(Y_t - m_\xi), \quad \text{from [10]} \]

Therefore,

\[ \tilde{H}_{t/t} = \tilde{H}_{t/t-1} + t e_r R^{-1}(Y_t - m_\xi) \]

For the second one, we start from the update equation of the predicted state:

\[ \hat{H}_{t/t-1} = \Omega + \Phi \hat{H}_{t/t} + m_\nu \]

Then we get

\[ \mathcal{I}_{t/t-1} \hat{H}_{t/t-1} = (M_t - M_t (Q^{-1} + M_t)^{-1}M_t)(\Omega + \Phi \hat{H}_{t/t} + m_\nu), \quad \text{from [12]} \]

\[ \hat{H}_{t/t-1} = (I_r - M_t (Q^{-1} + M_t)^{-1}) \Phi^{-1} \mathcal{I}_{t/t} \Phi^{-1}(\Omega + \Phi \hat{H}_{t/t} + m_\nu) \]

\[ = (I_r - M_t (Q^{-1} + M_t)^{-1}) \Phi^{-1} \mathcal{I}_{t/t}(\Phi^{-1} \Omega + \hat{H}_{t/t} + \Phi^{-1} m_\nu) \]

with

\[ N_t = (I_r - M_t (Q^{-1} + M_t)^{-1}) \Phi^{-1} \]
Hence,
\[ \tilde{H}_{t/t-1} = N_t \tilde{H}_{t-1/t-1} + N_t \Phi^{-1} I_{t-1/t-1} \Phi^{-1}(\Omega + m_\nu) \]

Summing up, the final information filter is represented in the algorithm (3).

It is worth noting that \( Q \) is not necessarily invertible. To cope with that, we proceed as in Brown [26] whereby equation (12) degenerates into \( I_{t/t} = M_t \) whenever \( Q \) is singular.

In addition, we stabilize the filter assuming that all eigenvalues of \( \Phi \) are of modulus less than 1 as a necessary and sufficient condition (Harvey [27]). This is equivalent to:

\[ \max |\text{Sp}(\Phi)| < 1 \] (14)

Note that (14) ensures also the second order stationary of \( H_t \) since the system equations (3) is time invariant (see Theorem 3.1, p.70 in Anderson and Moore [28]). Then, under (14), the log-GARCH parameters satisfy
\[ \sum_{i=1}^{r} \phi_i < 1. \]

---

**Algorithm 1** Information filtering

1: Initialization : \( \hat{H}_{0/1} = \mathbb{E} H_0 \) and \( I_{0/1} = 0 \)
2: for \( t = 0 : n \) do
3: \( \tilde{H}_{t+1/t} = I_{t+1/t} \Phi^{-1}(\Omega + m_\nu) \)
4: \( \tilde{H}_{t+1/t} = M_t \)
5: if \( \max |\text{Sp}(\Phi)| < 1 \) then
6: if \( \det(Q) \neq 0 \) then
7: \( \tilde{H}_{t+1/t} = N_t \Phi^{-1} I_{t+1/t} \Phi^{-1}(\Omega + m_\nu) \)
8: else
9: \( \tilde{H}_{t+1/t} = M_t \)
10: end if
11: \( \tilde{H}_{t+1/t} = N_t \Phi^{-1} I_{t+1/t} \Phi^{-1}(\Omega + m_\nu) \)
12: end if
13: \( \tilde{H}_{t+1/t} = \tilde{H}_{t+1/t} \)
14: if \( \max |\text{Sp}(\Phi)| < 1 \) then
15: \( \tilde{H}_{t+1/t} = \tilde{H}_{t+1/t} \)
16: end if
17: end for

Thus, given the predicted estimate \( \hat{H}_{t/t-1} \), the log conditional-variance estimate, denoted \( \hat{H}_{t/t-1}^{(1)} \), is easily derived as
\[ \hat{H}_{t/t-1}^{(1)} = e_r \hat{H}_{t/t-1} = \mathbb{E}(\log \sigma_t^2 / \varepsilon_t^{2}) \] (15)

Then, we extract the conditional variance estimate \( \hat{\sigma}_t^2 \) from \( \hat{H}_{t/t-1}^{(1)} \) as
\[
\hat{\sigma}^2_{t/t-1} = \left( \mathcal{I}^{(1)}_{t/t-1} \right)^{-1} \exp \hat{H}^{(1)}_{t/t-1}
\]  
\text{(16)}

where \( \left( \mathcal{I}^{(1)}_{t/t-1} \right)^{-1} \) is the inverse of the first diagonal element of the predicted information matrix as the predicted variance of \( \log \sigma^2_t \) introduced to minimize the bias between \( \hat{\sigma}^2_{t/t-1} \) and \( \exp \hat{H}^{(1)}_{t/t-1} \) (see Aknouche [25]).

4. Estimating the log-GARCH parameters in presence of zero returns

We now deal with the problem of presence of zero returns in data to be fitted by log-GARCH\((p,q)\). The proposed estimation of model parameters consists in combining the information filter \(3\) with the EM (Expectation-Maximization) algorithm based on the state space model \(3\).

Let \( \varepsilon_1, \ldots, \varepsilon_n \) be observations of the log-GARCH process \(1\) of an unknown parameter vector \( \theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \in \Theta \), where

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

Let \( \{\varepsilon_{i_1}, \varepsilon_{i_2}, \ldots, \varepsilon_{i_m} / 1 \leq i_1 \leq \ldots \leq i_m \leq n\} \) be the set of zero returns contained in the data. Then, let \( \mathcal{Y}_m \) denotes the sequence of "log-zero returns" to be imputed, that is:

\[
\mathcal{Y}_m = \{Y_{i_1}, \ldots, Y_{i_m}\} \quad \text{with} \quad i_1 > 1
\]

and

\[
\mathcal{Y}_{n_1} = \{Y_1, Y_2, \ldots, Y_{i_1-1}^*\} \quad \text{with} \quad n_1 = i_1 - 1
\]

Thus, given an initial value \( \theta_0 \) and from the observation set \( \mathcal{Y}_{n_1} \), we apply the EM algorithm using the information filter outputs as follows:

- **Step E (Expectation):** Compute from the information filter \(3\) the predicted measurement \( \hat{Y}_{i_1/i_1-1}(\hat{\theta}_1) \) which will be the imputed value in \( Y_{i_1} \), that is:

\[
Y_{i_1}^* = \hat{Y}_{i_1/i_1-1}(\hat{\theta}_1) = \hat{H}^{(1)}_{i_1/i_1-1} + m \xi
\]

- **Step M (Maximization):** Maximize the log-likelihood function \( \hat{L}_{n_1+1}(\theta) \) relying on the observation set \( \mathcal{Y}_{n_1} \cup \{Y_{i_1}^*\} \). Thus, maximizing

\[
\hat{L}_{n_1+1}(\theta; \mathcal{Y}_{n_1}, Y_{i_1}^*) = -\frac{n_1 + 1}{2} \log(2\pi) - \frac{n_1 + 1}{2} \left( \frac{1}{n_1 + 1} \sum_{t=1}^{n_1+1} \frac{\varepsilon_t^2}{\hat{\sigma}^2_{t/t-1}(\theta)} + \hat{H}^{(1)}_{i_1/i_1-1}(\theta) \right)
\]

means minimizing the criterion \( \hat{l}_{n_1+1}(\theta) \) given by:

\[
\hat{l}_{n_1+1}(\theta; \mathcal{Y}_{n_1}, Y_{i_1}^*) = \frac{1}{n_1 + 1} \sum_{t=1}^{n_1+1} \frac{\varepsilon_t^2}{\hat{\sigma}^2_{t/t-1}(\theta)} + \hat{H}^{(1)}_{i_1/i_1-1}(\theta)
\]

which provides a first iterated parameter estimate:

\[
\hat{\theta}_1 = \arg \min_{\theta \in \Theta} \hat{l}_{n_1+1}(\theta)
\]
Then, we set $\mathcal{Y}_{n_2} = \{Y_1, Y_2, \ldots, Y^*_1, \ldots, Y_{i_2 - 1}\}$ with $n_2 = i_2 - 1$, and we continue in an ascending way the imputation of the remainder of the values to be by running recursively steps E and M until the imputation of the last log-zero observation $Y_{i_m}$. Finally, the parameter estimate is obtained using observations $\mathcal{Y}_{n_m+1} \cup \{Y_{i_m+1}, \ldots, Y_n\}$ as:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{l}_n(\theta)$$

As for the special case where the first observation $Y_1$ corresponds to a zero return ($i_1 = 1$), it is imputed by the theoretical mean $EY_1$:

$$Y^*_1 = \frac{\omega + m\xi \left(1 - \sum_{j=1}^{q} \beta_j\right)}{1 - \sum_{i=1}^{q} \phi_i} t(1, \ldots, 1)$$

We summarize the previous treatment in the following algorithm.

**Algorithm 2** EM algorithm

1: Initialization : $\theta_0 \in \Theta$
2: $\mathcal{Y}_n = \{Y_1, Y_2, \ldots, Y_n\}$
3: if $i_1 = 1$ then
4: $Y_1 = Y^*_1 = \frac{\omega + m\xi \left(1 - \sum_{j=1}^{q} \beta_j\right)}{1 - \sum_{i=1}^{q} \phi_i} t(1, \ldots, 1)$
5: else
6: for $j = 1 \rightarrow m$ do
7: $\mathcal{Y}_{nj} = \mathcal{Y}_n [1 : i_j - 1]$ and $n_j = Card \mathcal{Y}_{nj}$
8: Compute $\hat{H}^{(1)}_{i_j/i_{j-1}}(\mathcal{Y}_{nj}; \theta_{j-1})$ and $\hat{\sigma}^{2}_{i_j/i_{j-1}}(\mathcal{Y}_{nj}; \theta_{j-1})$
9: Deduce $\hat{Y}_{i_j/i_{j-1}} = \hat{H}^{(1)}_{i_j/i_{j-1}} + m\xi$
10: $Y^*_{i_j} = \hat{Y}_{i_j/i_{j-1}}$
11: Estimate $\hat{l}_{n_j+1}(\theta; \mathcal{Y}_{nj}, Y^*_{i_j})$
12: $\hat{\theta}_j = \arg \min_{\theta \in \Theta} \hat{l}_{n_j+1}(\theta)$
13: end for
14: $\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{l}_n(\theta; \mathcal{Y}_{n_m+1} \cup \{Y_{i_m+1}, \ldots, Y_n\})$
15: end if

5. Empirical illustrations

In this section, we illustrate the performance of our method in terms of parameter estimation and goodness of fit for three daily return series, namely the Crude Oil WTI (19 Sep. 1983- 23 Aug. 1999), CAC40 (02 Janv. 2005- 31 Dec. 2011) and the FTSE100 (02 Janv. 2005- 31 Dec. 2011) plotted in Figure 1. All used series are available in the website https://www.investing.com.
Figure 1: Daily financial log-returns in percents.
As displayed in Table 1, where 0s, $p_0$, $s^2$, $s^4$ and ARCH test stand respectively for the number of zeros in the sample of size $n$, the corresponding proportion, the sample second moment, the sample fourth moment, and the Lagrange Multiplier test statistics of first order. Clearly, all series exhibit the usual properties of excess kurtosis and significant presence of ARCH effect at first order. Moreover, we consider different proportion of zeros, namely 1.47%, 2.23% and 5% respectively for Crude Oil WTI, CAC40 and FTSE100. Note that zero returns include actual observed zeros and holidays.

Table 1: Descriptive statistics of financial returns.

| Series        | n  | 0s | $p_0$ | $s^2$  | $s^4$  | ARCH test ($p$-value) |
|---------------|----|----|-------|--------|--------|-----------------------|
| Crude Oil WTI | 5818 | 86 | 1.47% | 0.8093 | 31.4116 | 0.0000                |
| CAC40         | 2555 | 57 | 2.23% | 1.4    | 21.1892 | 0.0000                |
| FTSE100       | 2555 | 128| 5%    | 1.0570 | 14.9132 | 0.0000                |

For each zero proportion, we compare our method, denoted EM-QML with the method of Sucarrat and Escribano [18], denoted Cex-QML, as well as the method of Francq, Wintenberger and Zakoian [17], denoted QML, where the zero returns are removed and the data are bound together. Table 2 displays the estimated log-GARCH(1,1) model for each series applied to the log-returns in percent $r_t$ using the following conditional variance specifications:

$$
\begin{align*}
EM-QML, Cex-QML : & \quad \log \sigma_t^2 = \omega + \alpha_1 \log \sigma_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2 \\
QML : & \quad \log \sigma_t^2 = \omega + \alpha_1 \mathbb{I}_{\{\eta_{t-1}=0\}} \log \sigma_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2
\end{align*}
$$

(17) (18)

optim. command in R is used for the likelihood maximization. (see Steenbergen [29])

For all series, the EM-QML specification fits the data best according to the Bayesian information criterion (BIC) having also the higher (quasi) log-likelihood. Thus, the EM-QML imputation outperforms the Cex-QML method so that the imputed log-zero returns using our method are the most compatible with the data they belong to. Furthermore, our method seems more suitable that the QML method whereby the zeros are removed from data before performing the parameter estimation.

On the other side, the smallest estimate differences come from Crude Oil WTI (difference at the third decimal) with zero proportion equal to 1.47% whereas the biggest come from FTSE100 with zero proportion equal to 5%. For these cases, one can see that the number of zeros is one of causes producing the estimation bias. This is not a general rule, it is enough to review the empirical illustrations reported in Sucarrat and Escribano [18].
Table 2: Empirical estimates of log-GARCH(1,1) specifications of the three daily financial returns. The estimated standard deviations are displayed in brackets.

| Series      | Method | n   | ω    | α    | β    | Log-Lik | BIC   |
|-------------|--------|-----|------|------|------|---------|-------|
| Crude Oil WTI | EM-QML | 5818 | 0.0802 | 0.0546 | 0.9373 | -4595.472 | 9216.495 |
|             | Cex-QML | 5818 | 0.0832 | 0.0576 | 0.9333 | -6296.561 | 12618.67 |
|             | QML     | 5732 | 0.0852 | 0.0580 | 0.9336 | -4782.213 | 9589.244 |
| CAC40       | EM-QML | 2555 | 1.3471 | 0.3568 | 0.1405 | -2344.437 | 4712.412 |
|             | Cex-QML | 2555 | 0.0701 | 0.0345 | 0.9487 | -3571.9 | 7167.337 |
|             | QML     | 2498 | 1.3321 | 0.3524 | 0.1506 | -481.519 | 8986.508 |
| FTSE100     | EM-QML | 2555 | 1.0704 | 0.3746 | 0.0913 | -2349.862 | 4723.261 |
|             | Cex-QML | 2555 | 0.0641 | 0.0346 | 0.9564 | -3099.961 | 6223.459 |
|             | QML     | 2427 | 1.0330 | 0.3648 | 0.1142 | -3768.716 | 7560.815 |

6. Conclusion
In this work, we presented a method of imputation of zeros observed in the return series in order to render the log-GARCH model used to fit such data. Our method combined the information filter used to obtain a predicted estimation of the "log-zeros" without constraint on the existence of log-moment greater than one, and the EM algorithm for the QML estimation of parameters. Empirical applications confirmed the benefit to use our imputation method instead of removing zeros from data, notably in the case whereby the estimation bias increases with the proportion of zeros.

References
[1] Lesmond D A, Ogden J P and Trzcinka C A 1999 The review of financial studies 12 1113–1141
[2] Bekaert G, Harvey C R and Lundblad C 2007 The review of financial studies 20 1783–1831
[3] Goyenko R Y, Holden C W and Trzcinka C A 2009 Journal of financial Economics 92 153–181
[4] Levine R and Schmukler S L 2006 Review of Finance 10 153–187
[5] Mazza P 2015 Finance 36 7–36
[6] Bandi F M, Pirino D and Reno R 2017 Econometrica 85 1793–1846
[7] Kolokolov A, Livieri G and Pirino D 2020 Journal of Econometrics
[8] Hausman J A, Lo A W and MacKinlay A C 1992 Journal of financial economics 31 319–379
[9] Engle R F and Russell J R 1998 Econometrics 1127–1162
[10] Russell J R and Engle R F 2005 Journal of Business & Economic Statistics 23 166–180
[11] Bien K, Nolte I and Pohlmeier W 2011 Journal of Applied Econometrics 26 669–707
[12] Rydberg T H and Shephard N 2003 Journal of Financial Econometrics 1 2–25
[13] Sucarrat G and Grønneberg S 2016
[14] Geweke J 1986 Econometric Reviews 5 57–61
[15] Pantula S G 1986 Econometric Reviews 5 79–97
[16] Milhøj A 1987 A multiplicative parameterization of ARCH models
[17] Francq C, Wintenberger O and Zakoian J M 2013 *Journal of Econometrics* **177** 34–46
[18] Sucarrat G and Escribano A 2013
[19] Dempster A P, Laird N M and Rubin D B 1977 *Journal of the Royal Statistical Society: Series B (Methodological)* **39** 1–22
[20] Little Roderick J and Rubin Donald B 1987 *Hoboken, NJ: Wiley* **65**
[21] Schafer J L 1997 *Analysis of incomplete multivariate data* (CRC press)
[22] McLachlan G J and Krishnan T 1997 The em algorithm and extensions
[23] Brockwell P J, Davis R A and Fienberg S E 1991 *Time series: theory and methods: theory and methods* (Springer Science & Business Media)
[24] BAHAMONDE N
[25] Aknouche A 2017 *Statistical Inference for Stochastic Processes* **20** 139–177
[26] Brown R G 1983 *Introduction to random signal analysis and Kalman filtering* vol 8 (Wiley New York)
[27] Harvey A C 1990 *Forecasting, structural time series models and the Kalman filter* (Cambridge university press)
[28] Anderson B D and Moore J B 2012 *Optimal filtering* (Courier Corporation)
[29] Steenbergen M R 2006 *University of North Carolina, Chapel Hill*