MATHEMATICAL MODELING AND COMPUTING, Vol. 8, No. 4, pp. 797–806 (2021)

Modeling Computing Mathematical

On the computational estimation of high order GARCH model

Settar A.[1], Fatmi N. I.[1], Badaoui M.[1,2]

[1] *IPIM, National Schools of Applied Sciences, Khouribga, Sultan Moulay Slimane University, Morocco*
[2] *LaMSD, Higher School of Technology, Oujda, Mohammed First University, Morocco*

(Received 23 May 2021; Accepted 7 June 2021)

To guarantee the non-negativity of the conditional variance of the GARCH process, it is sufficient to assume the non-negativity of its parameters. This condition was empirically violated besides rendering the GARCH model more restrictive. It was subsequently relaxed for some GARCH orders by necessary and sufficient constraints. In this paper, we generalized an approach for the QML estimation of the GARCH(p, q) parameters for all orders $p \geqslant 1$ and $q \geqslant 1$ using a constrained Kalman filter. Such an approach allows a relaxed QML estimation of the GARCH without the need to identify and/or apply the relaxed constraints to the parameters. The performance of our method is demonstrated through Monte Carlo simulations and empirical applications to real data.

Keywords: *GARCH, constrained Kalman filter, conditional variance, volatility, quasi-maximum likelihood.*

2010 MSC: 62-08, 62L20, 62M10, 60G12, 91B05 **DOI:** 10.23939/mmc2021.04.797

1. Introduction

We are not wrong in saying that the wide use of the GARCH(p, q) [1] for $p, q \in \{1, 2\}$ has hindered the study of computational proposals for the estimation of higher-order GARCH parameters, including the deal with the issue of the conditional variance non-negativity. For this last, Bollerslev [1] had imposed the non-negativity of the GARCH model parameters as sufficient condition avoiding the conditional variance to be negative. Bollerslev's GARCH(p, q) specification of a discrete process ε_t is given by

$$\begin{aligned} &\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim \text{iid}(0, 1), \\ &\text{Var}\left(\varepsilon_t \mid \varepsilon_u; u < t\right) := \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \end{aligned} \tag{1}$$

Bollerslev's assumptions are the next

$$\omega > 0, \quad \alpha_i \geqslant 0, \quad \text{for} \quad i = 1, \ldots, p \quad \text{and} \quad \beta_j \geqslant 0, \quad \text{for} \quad j = 1, \ldots, q.$$

However, several empirical studies have showed violations of Bollerslev's assumptions for some GARCH orders greater than 1, e.g., GARCH(2,1) for daily capital gains on the S&P500 from 1928 to 1984 [2] and GARCH(1,4) for exchange-rate movements in different world markets [3]. Indeed, in these works, some estimated parameters were significantly negative which questioned the prior enforcement of the parameter non-negativity.

Actually, the appearance of other extensions of the GARCH (EGARCH, log-GARCH, etc.) did not prevent researchers from investigating relaxation areas of the GARCH model related to the problem of non-negativity of the conditional variance. Nelson and Cao [4] and Tsai and Chan [5] derived necessary and sufficient constraints relaxing those of Bollerslev for $q < 2$. The same works deal briefly with $q \geqslant 3$ case giving sufficient conditions ensuring the almost sure non-negativity of σ_t^2. He and Terasvirta [7] studied the properties of the autocorrelation function of squared observations for

second order GARCH processes under Nelson and Cao's constraints and revealed a more flexibility in the shape of the correlation. Conrad and Karanasos [6] generalize the univariate case and derive a set of necessary and sufficient conditions for the non-negativity of the conditional variance vector of multivariate GARCH models.

In fact, the methodology developed in Nelson and Cao [4] and Tsai and Chan [5] does not seem to be practical from a computational viewpoint especially for $q \geqslant 3$ where the relaxed constraints degenerate to an infinite number of inequalities [4]. Nelson and Cao indeed do not argue to impose them in the estimation and advocate the use of other practical routines for higher-order GARCH models. Furthermore, in R program, all packages, e.g., fGarch [8] and rugarch [9] estimate the GARCH model under Bollerslev's standard assumptions.

Settar et al. [10] outlined a set of algorithms for the QML estimation of the GARCH(1,1) parameters using a constrained Kalman filter to enforce the non-negativity constraint of σ_t^2 or any other boundedness constraints on volatility with no need to identify or/and apply the associated relaxed constraints. Such an estimation was used to relax the QML estimation of the CGARCH model with GARCH(1,1) components against the relaxed non-negativity constraints [11]. In this work, we extend this approach to all orders $p \geqslant 1$ and $q \geqslant 1$ of GARCH(p,q) models. Such an extension consists in (i) determining the auto-covariance function of σ_t^2 which was not required for the GARCH(1,1) case and (ii) randomizing the non-negativity constraint of σ_t^2 used in Settar et al. [10] to be more accurate.

Thus, we construct in section 2 the extended state space representation of the GARCH(p,q), using it to compute the auto-covariance function of σ_t^2. The constrained Kalman filter is implemented in section 3 in order to estimate σ_t^2 with respect to a new random non-negativity constraint. Section 4 presents Monte Carlo simulations. A comparative application to real data is performed in section 5. Conclusion is given in section 6.

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 $\mathcal{M}_{(k,k)}$. The transpose of a matrix $A = (a_{ij})$ is A^T. $A_{i\bullet}$ and $A_{\bullet j}$ stand respectively for the i^{th} row and the j^{th} column of A. $A_{\times\bullet} := (A_{1\bullet}^T \dots A_{k\bullet}^T)^T \in \mathbb{R}^{kl}$ and $A_{\bullet\times} := (A_{\bullet 1} \dots A_{\bullet l})^T \in \mathbb{R}^{k^2}$. $A_{\bullet l}^{(s)} := (a_{sl} \dots a_{kl})^T$, $1 \leqslant s \leqslant k$. $A_{\bullet\times}^{(s)} := (A_{\bullet 1}^{(s)} \dots A_{\bullet k}^{(s)})^T \in \mathbb{R}^{(k-s+1)^2}$. ρ stands for the spectral radius of a matrix. "iid" means independent and identically distributed. The Kronecker product is denoted by $\otimes$. e_k^T is the first vector of the canonical base of $\mathbb{R}^k$. For any sequence of random variables $(X_t)_{t\in\mathbb{Z}}$, $\underline{X}_t := (X_t, \dots, X_{t-r+1})^T \in \mathbb{R}^r$.

2. Relaxed GARCH state space representation

Throughout this work, we use the state space representation in the form proposed by Hung [12], applying to the relaxed GARCH(p,q) model with iid Gaussian innovation η_t[1], under assumptions **(A1)** and **(A2)** giving respectively a sufficient conditions of the second order stationary and the existence of the fourth moment of the relaxed GARCH [10],

$$\textbf{(A1)} \quad \sum_{i=1}^{r} |\alpha_i| + |\beta_i| < 1 \qquad \text{and} \qquad \textbf{(A2)} \quad \rho(A_1^{(2)}) < 1,$$

where $r = \max(p,q)$ such that $\alpha_i = 0$ (resp. $\beta_j = 0$) if $i > p$ (resp. $j > q$) and

$$A_t = \begin{pmatrix} \alpha_1\eta_t^2 & \cdots & \alpha_p\eta_t^2 & \beta_1\eta_t^2 & \cdots & \beta_q\eta_t^2 \\ & I_{p-1} & 0_{(p-1,1)} & & 0_{(p-1,q)} & \\ \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ & 0_{(q-1,p)} & & & I_{q-1} & 0_{(q-1,1)} \end{pmatrix}.$$

[1]The choice of the Gaussian distribution is not restrictive since we work with the quasi-maximum likelihood for the parameter estimation.

Therefore, GARCH equations (1) can be rewritten in the state space form as:

$$\begin{cases} h_t = \underline{\omega} + \Lambda\, h_{t-1} + \Phi\, \underline{\nu}_{t-1}, \\ \varepsilon_t^2 = e_r\, h_t + \nu_t, \end{cases} \tag{2}$$

where $\nu_t \sim \text{iid}(0,\nu)$ which represents the linear white noise innovation of ε_t^2 given by $\nu_t = \varepsilon_t^2 - \sigma_t^2$. Here, the state vector is $h_t = (\sigma_t^2\ \sigma_{t-1}^2 \dots \sigma_{t-r+1}^2)^T \in \mathbb{R}^r$. The transition matrices are defined by

$$\underline{\omega} = \begin{pmatrix}\omega & 0_{(r-1,1)}\end{pmatrix}^T \in \mathbb{R}^r,\ \Lambda = \begin{pmatrix} \lambda_1 & \cdots & \lambda_r \\ I_{r-1} & & 0_{(r-1,1)} \end{pmatrix} \in \mathcal{M}_{(r,r)},\ \Phi = \begin{pmatrix} \alpha_1 & \cdots & \alpha_r \\ & 0_{(r-1,r-1)} & \end{pmatrix} \in \mathcal{M}_{(r,r)}.$$

2.1. Auto-covariance structure

Unlike the GARCH(1,1) case dealt by Settar et al., the application of the Kalman filter requires a knowledge of the covariance structure of h_t. For this purpose, we provide in the following a recursive method based on the state space model (2) to compute the auto-covariance function of h_t. At first, we express via the following lemma, $\mathbb{E}(\underline{\varepsilon}_t^2 \otimes \underline{\nu}_t)$ and $\mathbb{E}(\underline{\nu}_t \otimes \underline{\varepsilon}_t^2)$ as functions of the relaxed GARCH(p,q) parameters and ν. These quantities will be needed to provide the auto-covariance function of h_t given by the proposition 7.

Lemma 1. *Let ε_t be a relaxed GARCH(p,q) process. For all $k \in [\![2,r]\!]$, let $\underline{\alpha}_{k-1} = (\alpha_{k-1} \dots \alpha_1)^T$ and $\Gamma_{k-1}, M \in \mathcal{M}_{(k-1,k-1)}$ defined by:*

$$\Gamma_{k-1} = \begin{pmatrix} 1 & -\lambda_1 & -\lambda_2 & . & . & -\lambda_{k-1} \\ 0 & 1 & -\lambda_1 & . & . & -\lambda_{k-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & -\lambda_1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \textit{and} \quad M = \begin{pmatrix} 1 & m_1^{(1)} & m_1^{(2)} & . & . & m_1^{(k-1)} \\ 0 & 1 & m_2^{(2)} & . & . & m_2^{(k-1)} \\ . & . & . & . & . & . \\ . & . & . & . & . & m_{k-1}^{(k-1)} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

with $m_i^{(k-1)}$ stands for the i^{th} component of $\Gamma_{k-1}^{-1}\underline{\alpha}_{k-1}$, for all $i \in [\![1,k]\!]$. Then:

$$\mathbb{E}(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1}) = M_{\times\bullet}\nu \quad \textit{and} \quad \mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2) = M_{\bullet\times}\nu.$$

Proof. Recall that the white noise ν_t is the linear innovation of ε_t^2. Hence, for all $i > k > 0$:

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k}) = 0 \tag{3}$$

It follows:

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-i}) = \mathbb{E}\left[\omega\nu_{t-i} + \sum_{j=1}^{r} \lambda_j \varepsilon_{t-i-j}^2 \nu_{t-i} + \nu_{t-i}^2 - \sum_{j=1}^{q} \beta_j \nu_{t-i-j}\nu_{t-i}\right] = \nu. \tag{4}$$

Consider the case $0 < i < k \leqslant r$. By taking the expectation of $\varepsilon_{t-i}^2\nu_{t-k}$ into the ARMA representation of ε_t^2 given by:

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{r}(\alpha_i + \beta_i)\varepsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{r}\beta_j\nu_{t-j},$$

one can obtain:

$$\mathbb{E}(\varepsilon_{t-i}^2\nu_{t-k}) = \begin{cases} \alpha_{k-i}\nu + \sum_{j=1}^{k-i-1}\lambda_j \mathbb{E}(\varepsilon_{t-i-j}^2\nu_{t-k}), & \text{if } 0 < i < k-1, \\ \alpha_1\nu, & \text{if } i = k-1. \end{cases} \tag{5}$$

Let's set $X_{i,k} = \mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k})$. Then, (3)–(5) become:

$$X_{i,k} = \begin{cases} \alpha_{k-i}\nu + \sum_{j=1}^{k-i-1} \lambda_j X_{i+j,k}, & \text{if } 0 < i < k-1, \\ \alpha_1 \nu, & \text{if } i = k-1, \\ \nu, & \text{if } i = k, \\ 0, & \text{if } i > k. \end{cases} \tag{6}$$

The first two equations of the system (6) can be expressed in matrix form as:

$$(X_{i,1} \dots X_{i,k-1})^T = \Gamma_{k-1}^{-1} \underline{\alpha}_{k-1} \nu.$$

This gives, by denoting $m_i^{(k-1)}$ the i^{th} component of $\Gamma_{k-1}^{-1}\underline{\alpha}_{k-1}$, that

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k}) = X_{i,k} = \begin{cases} m_i^{(k-1)}\nu, & \text{if } 0 < i < k, \\ \nu, & \text{if } i = k, \\ 0, & \text{if } i > k. \end{cases} \tag{7}$$

Let $\mathbf{X} = (X_{i,k})_{i,k=1,\dots,r} \in \mathcal{M}_{(r,r)}$ which can be written as $\mathbf{X} = M\nu$. Hence, two expectations $\mathbb{E}(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1})$ and $\mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2)$ result. ■

The following proposition gives the covariance matrix of h_t.

Proposition 7. Let Ω, P, Q, B and C be the matrices defined by:

- $\Omega \in \mathcal{M}_{(r^2,r^2)}$ is the zero matrix except at the positions $((k-1)r+k, (k-1)r+k)$ for $k = 1, \dots, r$ being ones;
- $P = 2(\Lambda \otimes \Phi)(M_{\times\bullet} e_{r^2} - \Omega) + \Lambda^{\otimes 2}$;
- $Q = \left\{(\Phi \otimes \Lambda)(M_{\bullet\times} e_{r^2} - \Omega) + \Phi^{\otimes 2}\right\}\Omega$;
- $B = (B_k)_k \in \mathbb{R}^{r^2}$ such that $B_k = \omega^2 \left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j\right)^{-1}$ for $k = 1, \dots, r$ and 0 elsewhere;
- $C = (C_k)_k \in \mathbb{R}^{r^2}$ such that $C_{(l-1)r+1} = \omega^2 \left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j\right)^{-1}$ for $l = 1, \dots, r$ and 0 elsewhere.

Assume that $(I_{r^2} - P - 2Q)$ is not singular. Then, the auto-covariance matrix of h_t is obtained through the following recursive system:

$$\begin{cases} \mathbb{E}(h_t^{\otimes 2}) = (I_{r^2} - P - 2Q)^{-1} \{B + (\Lambda \otimes I_r)C\}; \\ \mathbb{E}(h_t \otimes h_{t-s}) = B + (\Lambda \otimes I_r)\mathbb{E}(h_t \otimes h_{t-s+1}) + 2(\Phi \otimes I_r)(M_{\bullet\times}^{(s)} - g_k) e_{r^2} \Omega \mathbb{E}(h_t^{\otimes 2}), \ \forall s > 0; \\ \mathrm{Cov}\left(\sigma_t^2, \sigma_{t-s}^2\right) = e_{r^2}\mathbb{E}(h_t \otimes h_{t-s}) - \left(e_r \, E h_0\right)^2, \ \forall s \geqslant 0, \end{cases} \tag{8}$$

where $g_k \in \mathbb{R}^{r^2}$ with ones at the position $(s-1+k)(r-s+1)+k+1$ for $k \geqslant 0$ and 0 elsewhere.

Proof. Applying the elementary properties of the Kronecker product to the state equation of the representation (2), we obtain

$$h_t^{\otimes 2} = \underline{\omega} \otimes h_t + \Lambda\, h_{t-1} \otimes \left(\underline{\omega} + \Lambda\, h_{t-1} + \Phi\, \underline{\nu}_{t-1}\right) + \Phi\, \underline{\nu}_{t-1} \otimes \left(\underline{\omega} + \Lambda\, h_{t-1} + \Phi\, \underline{\nu}_{t-1}\right).$$

Let set

$$\begin{aligned} T_1 &= \Lambda\, h_{t-1} \otimes \left(\underline{\omega} + \Lambda\, h_{t-1} + \Phi\, \underline{\nu}_{t-1}\right), \\ T_2 &= \Phi\, \underline{\nu}_{t-1} \otimes \left(\underline{\omega} + \Lambda\, h_{t-1} + \Phi\, \underline{\nu}_{t-1}\right). \end{aligned}$$

Then, on the one hand

$$\begin{aligned} T_1 &= (\Lambda \otimes I_r)\left(h_{t-1} \otimes \underline{\omega}\right) + \Lambda^{\otimes 2} h_{t-1}^{\otimes 2} + (\Lambda \otimes \Phi)\left(h_{t-1} \otimes \underline{\nu}_{t-1}\right) \\ &= (\Lambda \otimes I_r)\left(h_{t-1} \otimes \underline{\omega}\right) + \Lambda^{\otimes 2} h_{t-1}^{\otimes 2} + (\Lambda \otimes \Phi)\left(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1}\right) - (\Lambda \otimes \Phi)\, \underline{\nu}_{t-1}^{\otimes 2}. \end{aligned}$$

Since

$$\nu = \mathbb{E}\nu_t^2 = \mathbb{E}(\varepsilon_t^2 - \sigma_t^2)^2 = \mathbb{E}(\eta_t^2\sigma_t^2 - \sigma_t^2)^2 = \mathbb{E}(\eta_t^2 - 1)^2\mathbb{E}\sigma_t^4 = 2e_{r^2}\mathbb{E}(h_t^{\otimes 2}) \tag{9}$$

and $\mathbb{E}\underline{\nu}_t^{\otimes 2} = 2\Omega\mathbb{E}(h_t^{\otimes 2})$.

It follows

$$\begin{aligned}\mathbb{E}(T_1) &= (\Lambda \otimes I_r)\,\mathbb{E}\left(h_{t-1} \otimes \underline{\omega}\right) + \left\{2\,(\Lambda \otimes \Phi)\,(M_{\times\bullet}e_{r^2} - I_{r^2})\,\Omega + \Lambda^{\otimes 2}\right\}\mathbb{E}\left(h_t^{\otimes 2}\right)\\ &= (\Lambda \otimes I_r)\,C + P\mathbb{E}\left(h_t^{\otimes 2}\right).\end{aligned}$$

On the other hand

$$\begin{aligned}T_2 &= (\Phi \otimes I_r)\left(\underline{\nu}_{t-1} \otimes \underline{\omega}\right) + (\Phi \otimes \Lambda)\left(\underline{\nu}_{t-1} \otimes h_{t-1}\right) + \Phi^{\otimes 2}\underline{\nu}_{t-1}^{\otimes 2}\\ &= (\Phi \otimes I_r)\left(\underline{\nu}_{t-1} \otimes \underline{\omega}\right) + (\Phi \otimes \Lambda)\left(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2\right) - \left\{(\Phi \otimes \Lambda) - \Phi^{\otimes 2}\right\}\underline{\nu}_{t-1}^{\otimes 2}.\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}(T_2) &= 2\,(\Phi \otimes \Lambda)\,M_{\bullet\times}e_{r^2}\Omega\mathbb{E}\left(h_t^{\otimes 2}\right) + 2\left(\Phi^{\otimes 2} - \Phi \otimes \Lambda\right)\Omega\mathbb{E}\left(h_t^{\otimes 2}\right)\\ &= 2\left\{(\Phi \otimes \Lambda)\,(M_{\bullet\times}e_{r^2} - I_{r^2}) + \Phi^{\otimes 2}\right\}\Omega\mathbb{E}\left(h_t^{\otimes 2}\right)\\ &= Q\mathbb{E}\left(h_t^{\otimes 2}\right).\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}\left(h_t^{\otimes 2}\right) &= \mathbb{E}\left(\underline{\omega} \otimes h_t\right) + (\Lambda \otimes I_r)\,C + P\mathbb{E}\left(h_t^{\otimes 2}\right) + Q\mathbb{E}\left(h_t^{\otimes 2}\right)\\ &= B + (\Lambda \otimes I_r)\,C + P\mathbb{E}\left(h_t^{\otimes 2}\right) + Q\mathbb{E}\left(h_t^{\otimes 2}\right).\end{aligned}$$

Which implies

$$(I_{r^2} - P - Q)\,\mathbb{E}\left(h_t^{\otimes 2}\right) = B + (\Lambda \otimes I_r)\,C.$$

Since $(I_{r^2} - P - Q)$ is not singular, then $\mathbb{E}(h_t^{\otimes 2})$ follows.

As for $\mathbb{E}(h_t \otimes h_{t-s})$, it is directly deducted by multiplying the state equation in (2) by h_{t-s} and noting that $\mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-s}^2) = M_{\bullet\times}^{(s)}\nu$ and that $\mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\nu}_{t-s}) = g_k\nu$. ∎

3. Conditional variance estimation

In this section, we start by estimating σ_t^2 without any non-negativity constraint (Prediction step). Afterwards, the obtained predicted estimates are truncated so that the non-negativity constraint is enforced (Robustification step).

3.1. Prediction

We extend the Kalman filter algorithm proposed in Settar et al. [10] for the GARCH(1,1) to the GARCH(p, q) where the error covariance matrices are computed through the recursive equations (8).

Let $\widehat{h}_{t|t}$, $\widehat{h}_{t|t-1}$, $P_{t|t}$ and $P_{t|t-1}$ be respectively the filtered and predicted estimates of h_t and their error covariance matrices. Note that the initial state of h_t is given by its mean:

$$\mathbb{E}h_0 = \frac{\omega}{1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j}(1 \dots 1)^T \in \mathbb{R}^r$$

and its covariance matrix $P_{0|0} = \mathbb{E}\left\{(h_0 - \mathbb{E}h_0)(h_0 - \mathbb{E}h_0)^T\right\}$ determined by the recursive equations (8). Then, one derives the Kalman filter equations for $t = 1, \dots, n$ as follows,

$$\widehat{h}_{t|t-1} = \underline{\omega} + \Lambda\widehat{h}_{t-1|t-1},$$

$$
\begin{aligned}
P_{t|t-1} &= \Lambda P_{t-1|t-1}\Lambda^T + \nu\Phi\Phi^T,\\
K_t &= P_{t|t-1}e_r^T\left(e_r P_{t|t-1}e_r^T + \nu\right)^{-1},\\
\widehat{h}_{t|t} &= \widehat{h}_{t|t-1} + K_t\big(\varepsilon_t^2 - e_r\widehat{h}_{t|t-1}\big),\\
P_{t|t} &= P_{t|t-1}\left(I_r - e_r^T K_t^T\right).
\end{aligned}
$$

Thus, the predicted conditional variance is extracted for $t = 1, \ldots, n$ as

$$\widehat{\sigma}^2_{t|t-1} = e_r\,\widehat{h}_{t|t-1}. \tag{10}$$

3.2. Robustification

This step consists in enforcing the non-negativity constraint of $\widehat{\sigma}_t^2$ without regard to the relaxed constraints on the model parameters. Settar et al. [10] used the probability density function truncation method [13] to truncate the density function of $\left\{\widehat{\sigma}^2_{t|t-1}\right\}_{t=1,\ldots,n}$ (assumed Gaussian), with respect to the non-negativity constraint $\frac{1}{N} \leqslant \sigma_t^2 \leqslant N$, for $t = 1, \ldots, n$ and some upper bound N empirically set.

However, because of the deterministic shape of such a constraint, the empirical choice of N related to the non-negativity of σ_t^2 remains arbitrary or even fallacious in the absence of a priori information on volatility boundedness. In order to control such a choice, we explore a random non-negativity constraint through a bound $N_{1-\tau}$ such that for a fixed confidence level $1-\tau$ and for all $t = 1, \ldots, n$, we set

$$1 - \tau = \mathbb{P}\left\{\frac{1}{N_{1-\tau}} \leqslant \sigma_t^2 \leqslant N_{1-\tau}\right\}. \tag{11}$$

Which implies

$$N_{1-\tau} = \sqrt{p_{t|t-1}}\; z_{1-\tau} + \widehat{\sigma}^2_{t|t-1},$$

where $z_{1-\tau}$ stands for $(1-\tau)^{th}$ quantile of the standard Gaussian distribution.

The conditional variance estimate after enforcement of constraint (11):

$$\widetilde{\sigma}^2_{t|t-1} = \sqrt{p_{t|t-1}}\;\mu + \widehat{\sigma}^2_{t|t-1}, \tag{12}$$

$p_{t|t-1}$ is the first diagonal element of $P_{t|t-1}$ and μ is the mean of the truncated Gaussian density between $l_t = \frac{1-N_{1-\tau}\widehat{\sigma}^2_{t|t-1}}{N_{1-\tau}\sqrt{p_{t|t-1}}}$ and $u_t = \frac{N_{1-\tau}-\widehat{\sigma}^2_{t|t-1}}{\sqrt{p_{t|t-1}}}$.

Now, the quasi-log likelihood is well defined for all $\theta \in \Theta$ as follows

$$L_n(\theta; \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{\widetilde{\sigma}^2_{t|t-1}(\theta)} + \log(\widetilde{\sigma}^2_{t|t-1}(\theta))\right), \tag{13}$$

where $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T$ is the parameter vector and Θ is a subset of $\mathbb{R}^{*+} \times \mathbb{R}^{p+q}$ satisfying assumptions **(A1)**–**(A2)**. Thus, maximizing (13) with respect to $\theta \in \Theta$ is equivalent to minimizing the following criterion:

$$\widehat{l}_n(\theta; \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{\widetilde{\sigma}^2_{t|t-1}(\theta)} + \log\left(\widetilde{\sigma}^2_{t|t-1}(\theta)\right). \tag{14}$$

Remark 1. The conditional variance extraction as given by (10) allows returning to the univariate dimension of σ_t^2. Then, the constrained Kalman filter is applied with a single non-negativity constraint, that is (11) for all $p \geqslant 1$ and $q \geqslant 1$.

4. Monte Carlo experiment

In this section, we conduct two Monte Carlo simulations to assess the performance of our QML estimation based on the minimization of (14) compared to the standard QML estimation, in terms of estimating the relaxed GARCH parameters.

For each sample size $n \in \{500, 1000, 5000\}$, splitted into 1000 replications, we simulated two processes GARCH(3,1) and GARCH(2,3) respectively with parameter vectors $(0.01, 0.1, 0.2, 0.1, 0.4)^T$ and $(0.01, 0.1, 0.2, 0.1, 0.4, 0.1)^T$. The generated data is afterword estimated by our method with $\tau = 0.005$ and by standard QML. Note that both GARCH processes have highly persistent volatility, of 0.8 for the GARCH(3,1) and 0.9 for the GARCH(2,3). The mean squared error (MSE) is used for comparison purposes. Further, we minimize the $\widehat{l}_n$ criterion using the *optim.* command in R [14].

Tables 1 and 2 show that all parameters are better estimated by our method based on the relaxed GARCH, compared to the standard QML estimation since it leads to a decrease in the MSE. Note also that the low values of the standard deviations (SE) of $\widetilde{\theta}$ compared to those of $\widehat{\theta}$ indicate that the estimates over all replications are less dispersed around the former. In other words, the proposed algorithm is more stable around $\widetilde{\theta}$. Moreover, it is clear that for all samples, our method captured the high persistence of the volatility that characterizes the simulated GARCH processes, by recording values $\sum_{i=1}^{p} \widetilde{\alpha}_i + \sum_{j=1}^{q} \widetilde{\beta}_j$ that are close to the true persistence values set during the simulations compared to the values resulting from the standard QML estimation.

Table 1. Finite sample properties of the r-GARCH(3.1) and GARCH(3.1) parameter estimates. Estimated standard deviations are shown in brackets. Smallest MSEs are marked by $^{(*)}$.

n	parameters	values	$\widetilde{\theta}$ (SE)	MSE	$\widehat{\theta}$ (SE)	MSE
500	ω	0.01	0.0102 (0.0223)	0.0005*	0.0964 (0.4935)	0.2511
	α_1	0.1	0.0995 (0.0574)	0.0033*	0.1823 (0.2245)	0.0572
	α_2	0.2	0.1953 (0.0753)	0.0057*	0.1330 (0.2757)	0.0805
	α_3	0.1	0.0963 (0.0478)	0.0023*	0.0413 (0.1332)	0.0212
	β_1	0.4	0.4002 (0.0888)	0.0079*	0.5678 (0.3222)	0.1320
1000	ω	0.01	0.0097 (0.0932)	0.0087*	0.0478 (0.2589)	0.0685
	α_1	0.1	0.0968 (0.1009)	0.0102*	0.1034 (0.1956)	0.0383
	α_2	0.2	0.1950 (0.0942)	0.0089*	0.1955 (0.2887)	0.0834
	α_3	0.1	0.0921 (0.1026)	0.0106*	0.1416 (0.1251)	0.0174
	β_1	0.4	0.3974 (0.0989)	0.0098*	0.2654 (0.2865)	0.1002
5000	ω	0.01	0.0092 (0.0830)	0.0069*	0.0172 (0.1573)	0.0248
	α_1	0.1	0.1013 (0.0888)	0.0079*	0.0829 (0.1323)	0.0178
	α_2	0.2	0.1926 (0.0945)	0.0090*	0.2123 (0.2935)	0.0863
	α_3	0.1	0.0960 (0.0853)	0.0073*	0.1222 (0.1170)	0.0142
	β_1	0.4	0.3981 (0.0830)	0.0069*	0.3599 (0.2935)	0.0878

Table 2. Finite sample properties of the r-GARCH(2,3) and GARCH(2,3) parameter estimates. Estimated standard deviations are shown in brackets. Smallest MSEs are marked by $^{(*)}$.

n	parameters	values	$\widetilde{\theta}$ (SE)	MSE	$\widehat{\theta}$ (SE)	MSE
500	ω	0.01	0.0106 (0.0727)	0.0053*	0.0062 (0.0859)	0.0074
	α_1	0.1	0.0998 (0.0989)	0.0098*	0.0885 (0.2221)	0.0495
	α_2	0.2	0.1987 (0.0574)	0.0033*	0.0129 (0.0222)	0.0355
	β_1	0.1	0.1006 (0.0299)	0.0009*	0.1223 (0.2729)	0.0750
	β_2	0.4	0.4001 (0.0299)	0.0009*	0.1563 (0.2970)	0.1476
	β_3	0.1	0.1005 (0.0264)	0.0007*	0.1890 (0.3973)	0.1658
1000	ω	0.01	0.0119 (0.0754)	0.0057*	0.0814 (0.0938)	0.0139
	α_1	0.1	0.0989 (0.0632)	0.0040*	0.0901 (0.1212)	0.0148
	α_2	0.2	0.1993 (0.0678)	0.0046*	0.0149 (0.3882)	0.1850
	β_1	0.1	0.1009 (0.0519)	0.0027*	0.0757 (0.3667)	0.1351
	β_2	0.4	0.4303 (0.0942)	0.0098*	0.1430 (0.2373)	0.1224
	β_3	0.1	0.0991 (0.0399)	0.0016*	0.1643 (0.2320)	0.0580
5000	ω	0.01	0.0116 (0.0647)	0.0042	0.0180 (0.0472)	0.0023*
	α_1	0.1	0.0987 (0.0479)	0.0023*	0.0977 (0.0573)	0.0033
	α_2	0.2	0.1980 (0.0871)	0.0076*	0.0102 (0.3920)	0.1897
	β_1	0.1	0.0995 (0.0399)	0.0016*	0.0256 (0.4359)	0.1956
	β_2	0.4	0.3987 (0.0734)	0.0054	0.3990 (0.0479)	0.0023*
	β_3	0.1	0.0992 (0.0399)	0.0016*	0.1405 (0.0957)	0.0108

5. Empirical application

In this section, we conduct an empirical application of our estimation in order to compare its performance to the relaxed QML estimation established by Nelson and Cao in terms of the optimums obtained by each approach. The application covers three real series used by Baillie and Bollerslev [15] and reused by Nelson and Cao [4], namely the exchange rate of the German mark/dollar, the French franc/dollar and the Japanese yen/dollar from June 1, 1973 to January 28, 1985, each with a size of 2920 observations. These series as well as the corresponding return series are presented respectively in figures 1–3. We denote by (s_t) the exchange rate series and by (r_t) the corresponding log-return series in percent given by $r_t = 100\log(s_t/s_{t-1})$, $t = 1, \ldots, 2920$.

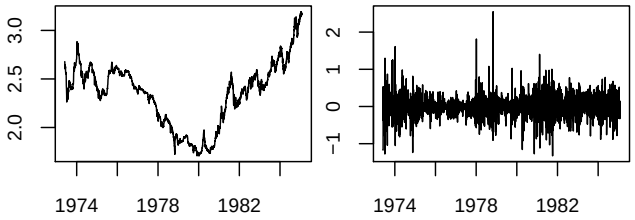

Fig. 1. Daily mark/dollar exchange rate (left) and the corresponding return series (right) from June 1, 1973 to January 28, 1985.

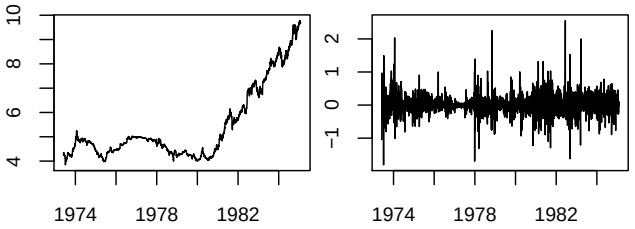

Fig. 2. Daily franc/dollar exchange rate (left) and the corresponding return series (right) from June 1, 1973 to January 28, 1985.

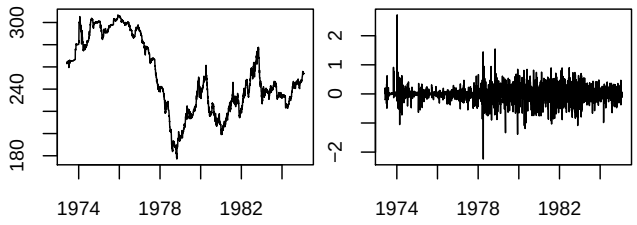

Fig. 3. Daily exchange rate of the yen/dollar (left) and the corresponding return series (right) from June 1, 1973 to January 28, 1985.

We follow the fitting of (r_t) adopted by Nelson and Cao [4] as a non-centred GARCH with an order selected by the AIC criterion, i.e.

$$r_t = \delta + \varepsilon_t, \quad (\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \varepsilon_1) \sim N(0, \sigma_t^2),$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

Table 3 reports the model parameter estimates obtained by both the QCK method and the approach of Nelson and Cao denoted $\widehat{\theta}^{NC}$. It is easy to check that $\widetilde{\theta}$'s values satisfy the inequalities of Nelson and Cao (See Theorem 1 and 2 in [4]) and relax accordingly Bollerslev's non-negativity constraints. This is indeed the case of $\widetilde{\alpha}_3$ in the fitting of the series of returns of the yen/dollar exchange rate. Actually, the advantage of using our method appears with the obtained values of log-likelihood. Indeed, it is clear that $\widetilde{\theta}$ maximize the log-likelihood compared to $\widehat{\theta}^{NC}$, which shows the potential of our method to reach a better optimum.

Table 3. Estimated GARCH models for daily exchange rates from June 1, 1973 to January 28, 1985. The last row gives the log-likelihoods.

	mark/dollar		franc/dollar		yen/dollar	
	GARCH(2,2)		GARCH(2,1)		GARCH(3,1)	
	$\widetilde{\theta}$	$\widehat{\theta}^{NC}$	$\widetilde{\theta}$	$\widehat{\theta}^{NC}$	$\widetilde{\theta}$	$\widehat{\theta}^{NC}$
δ	0.0028	−0.0009	0.0116	−0.0002	−0.0006	0.0016
ω	0.0182	0.0186	0.0021	0.0079	0.0111	0.0002
α_1	0.0576	0.0573	0.1680	0.1024	0.0088	0.1888
α_2	0.2265	0.2262	0.2449	0.1444	0.1988	0.0752
α_3					−0.1011	−0.2344
β_1	0.3829	0.3833	0.5846	0.7735	0.4011	0.9730
β_2	0.3096	0.3100				
log-Lik	−1462.94	−2447.24	−1459.63	−2356.21	−1659.94	−2086.27

6. Conclusion

In this work, we generalized an approach based on the constrained Kalman filter allowing to relax the QML estimation of the GARCH(p, q) parameters for all $p \geqslant 1$ and $q \geqslant 1$. A practical method for computing the auto-covariance function of the conditional variance based on the state-space representation was given and proved. We showed through a Monte Carlo experiment that our method is better for parameter and volatility estimation of high-order GARCH when compared to the QML estimation applied to the GARCH model under Bollerslev's conditions. The empirical study of three real series, namely the daily exchange rate returns of the mark/dollar, the franc/dollar and the yen/dollar showed the performance of our estimation method to relax the non-negativity assumption of parameters as well as maximizing the likelihood function compared to the QML estimation.

[1] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics. **31** (3), 307–327 (1986).

[2] French K. R., Schwert G. W., Stambaugh R. F. Expected stock returns and volatility. Journal of financial Economics. **19** (1), 3–29 (1987).

[3] Engle R. F., Ito T., Lin W. L. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. Working paper No. 2609 (1988).

[4] Nelson D. B., Cao C. Q. Inequality constraints in the univariate GARCH model. Journal of Business & Economic Statistics. **10** (2), 229–235 (1992).

[5] Tsai H., Chan K. S. A note on inequality constraints in the GARCH model. Econometric Theory. **24** (3), 823–828 (2008).

[6] Conrad C., Karanasos M. Negative volatility spillovers in the unrestricted ECCC-GARCH model. Econometric Theory. **26** (3), 838–862 (2010).

[7] He C., Teräsvirta T. Properties of the Autocorrelation Function of Squared Observations for Second-order Garch Processes Under Two Sets of Parameter Constraints. Journal of Time Series Analysis. **20**, 23–30 (1999).

[8] Wuertz D., Wuertz M. D., Team R. C. The fGarch Package. Wuertz, Diethelm, Maintainer Diethelm Wuertz, and Rmetrics Core Team (2013).

[9] Ghalanos A., Ghalanos M. A., Rcpp L. Package 'rugarch' (2019).

[10] Settar A., Fatmi N. I., Badaoui M. New Approach in Dealing with the Non-Negativity of the Conditional Variance in the Estimation of GARCH Model. Central European Journal of Economic Modelling and Econometrics. **13** (1), 55–74 (2021).

[11] Settar A., Fatmi N. I., Badaoui M. Quasi-maximum likelihood estimation of the Component-GARCH model using the stochastic approximation algorithm with application to the S&P 500. Mathematical Modeling and Computing. **8** (3), 379–390 (2021).

[12] Hung J. C. Robust Kalman filter based on a fuzzy GARCH model to forecast volatility using particle swarm optimization. Soft Computing. **19**, 2861–2869 (2015).

[13] Simon D. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley & Sons (2006).

[14] Steenbergen M. R. Maximum likelihood programming in R. University of North Carolina. Chapel Hill (2006).

[15] Baillie R. T., Bollerslev T. The message in daily exchange rates: a conditional-variance tale. Journal of Business & Economic Statistics. **20** (1), 60–68 (1989).

Про обчислювальну оцінку моделі GARCH високого порядку

Сеттар А.[1], Фатмі Н. І.[1], Бадауї М.[1,2]

[1] *IPIM, Національна школа прикладних наук Хорібги, Університет Султана Мулея Слімана, Марокко*
[2] *LaMSD, Вища технологічна школа Уджди, Перший університет Мохаммеда, Марокко*

Щоб гарантувати невід'ємність умовної дисперсії процесу GARCH, достатньо припустити невід'ємність її параметрів. Ця умова була емпірично порушена, що зробило модель GARCH більш обмеженою. Після цього ця умова була послаблена для деяких виборів необхідних та достатніх обмежень. У цій роботі узагальнено підхід для оцінки QML параметрів GARCH(p, q) для всіх порядків $p \geqslant 1$ та $q \geqslant 1$, використовуючи обмежений фільтр Калмана. Такий підхід дозволяє послаблену оцінку QML для GARCH без необхідності виявляти та/або застосовувати послаблені обмеження на параметри. Ефективність запропонованого методу демонструється за допомогою моделювання Монте–Карло та емпіричних застосувань до реальних даних.

Ключові слова: *GARCH, обмежений фільтр Калмана, умовна дисперсія, волатильність, квазімаксимальна ймовірність.*