Central European Journal of Economic Modelling and Econometrics

New Approach in Dealing with the Non-Negativity of the Conditional Variance in the Estimation of GARCH Model

Abdeljalil Settar[*], Nadia Idrissi Fatmi[†], Mohammed Badaoui[‡]

Submitted: 13.03.2020, Accepted: 22.08.2020

Abstract

In this paper, we propose a robust estimation of the conditional variance of the GARCH(1,1) model with respect to the non-negativity constraint against parameter sign. Conditions of second order stationary as well as the existence of moments are given for the new relaxed GARCH(1,1) model whose conditional variance is estimated deriving firstly the unconstrained estimation of the conditional variance from the GARCH(1,1) state space model, then, the robustification is implemented by the Kalman filter outcomes via density function truncation method. The GARCH(1,1) parameters are subsequently estimated by the quasi-maximum likelihood, using the simultaneous perturbation stochastic approximation, based, first, on the Gaussian distribution and, second, on the Student-t distribution. The proposed approach seems to be efficient in improving the accuracy of the quasi-maximum likelihood estimation of GARCH model parameters, in particular, with a prior boundedness information on volatility.

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

JEL Classification: C13, C15, C51, C61

[*]LIPIM, École Nationale des Sciences Appliquées (ENSA), Khouribga, Morocco; e-mail: abdeljalilsettar@gmail.com; ORCID: 0000-0002-9476-5180

[†]LIPIM, École Nationale des Sciences Appliquées (ENSA), Khouribga, Morocco; e-mail: nadidrissi200133@gmail.com; ORCID: 0000-0001-7423-3008

[‡]LIPIM, École Nationale des Sciences Appliquées (ENSA), Khouribga, Morocco and LaMSD, École Supérieure de Technologie (EST), Oujda, Morocco; e-mail: med.badaoui@gmail.com; ORCID: 0000-0002-5296-5117

1 Introduction

Since Bollerslev (1986) enlarged the Engel's ARCH model (1982), by the introduction of GARCH model, it has proven to be very useful in empirical works and financial surveys. GARCH model maintains the same time varying parametrization for the conditional variance with a more parsimonious form which allows the description of the data with less parameters.

However, to be well defined, GARCH process has been faced with the problem of non-negativity of its conditional variance. For this purpose, Bollerslev (1986) imposed the positivity on GARCH parameters as sufficient conditions but not necessary ensuring the non-negativity of the conditional variance. Based on the ARCH(∞) representation of the GARCH process, Nelson and Cao (1992) derived some necessary and sufficient conditions for a such non-negativity for GARCH(p, q) models with $p \leq 2$ and sufficient for $p > 2$. Tsai and Chan (2008) showed that for $p \geq 2$, the sufficient condition of Nelson and Cao (1992) is also necessary. Such approach requires an infinite number of inequality constraints on parameters, which can only be reduced in partial cases of the GARCH model orders, requiring further, a selection of start-up values that keep the conditional variance non-negative.

At the stage of estimation, especially for the quasi-maximum likelihood estimation, the non-negativity issue gives rise to additional difficulties related to the definedness of the likelihood function as well as the optimization procedure. Indeed, non-negativity constraints on parameters may be violated without using specified penalty function whose choice is much delicate in view of problems of slow convergence and non-smoothness that it creates. In fact, the common basis on which all previous investigations rely is to obtain the non-negativity of the conditional variance from a set of constraints on parameters. Thus, because of such a dependency, the GARCH model becomes very restrictive so that features like the random oscillatory in the conditional variance and the persistence of shocks on the volatility are not accurately captured. Moreover, the drawbacks of such approach appear clearly beyond the positivity of the conditional variance. Indeed, any additional information that can be available a priori on volatility, remains dependent on parameter constraints so that they should be modified a priori whenever the information changes.

Allal and Benmoumen (2011) together with Ossandón and Bahamonde (2013) are among others who have focused on the pre-estimation of the GARCH(1,1) conditional variance using the Kalman filter. All these contributions are based on the standard specification of GARCH model, whereby, the parameters are assumed to be non-negative in order to ensure the non-negativity of the conditional variance. Neverthless, no attention was given to the pre-estimation of this last regardless of the parameter sign.

In the present work, we propose a double step pre-estimation of the conditional variance generated by the GARCH model so that it would remains non-negative without positivity constraints on the model parameters. Primarily, following the relaxation of the GARCH in terms of positivity of the parameters, new sufficient

conditions of second order stationarity and existence of moments were established. Following these conditions, the first step consists of estimating the conditional variance by the standard Kalman filter through a state-space representation. As for the second step, the conditional variance estimates are robustified to be independent from the positivity restrictions on GARCH parameters. A constrained Kalman filter is implemented to this aim, avoiding to use any penalty method in the stage of parameter estimation. The robsutified estimate of the conditional variance allows to evaluate the likelihood function that we maximize using the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm to obtain a maximum likelihood estimates of the GARCH parameters. The proposed estimation involves two innovation distributions, namely, Gaussian and Student-t. Also, we advocate the usefulness of our approach in exploiting a priori available information on volatility, including boundedness information to make a statistical benchmark of the parameter estimates.
The results of our evaluation indicated that the constrained Kalman filter is advantageous for estimating the conditional variance with respect to the non-negativity constraint. Simulations demonstrated the effectiveness of the proposed approach in improving the accuracy of the quasi-maximum likelihood estimation, notably, for taking account other boundedness information on volatility beyond non-negativity.
The structure of the paper is as follows: In Section 2, we extend the condition of second order stationary as well as the sufficient condition of existence of moments for the GARCH(1,1) model without non-negativity constraints on parameters. In Section 3, GARCH state space representation is derived from which the conditional variance is robustified through a constrained Kalman filter. In Section 4, the estimated conditional variance is used to estimate the quasi-likelihood function that is optimized via the Simultaneous Perturbation Stochastic Approximation (SPSA). The performance of the proposed approach is evaluated through estimation simulations of GARCH(1,1) parameters in finite samples in Section 5. A conclusion is given in Section 6.

2 Model framework

Notations: For a matrix M, M' is the transpose of M. $Abs(M)$ is the matrix of same size as M, whose elements are the absolute values of the corresponding elements of M. The norm of a matrix M is defined by the sum of the absolute values of its elements. For any sequence of identically distributed random matrices $(M_t)_t$ and for any integer m, let $M^{(m)} = \mathbb{E}[\{Abs(M_1)\}^{\otimes m}]$, where $\otimes$ denotes the Kronecker product. For a random variable X, the m-norm of X is defined by $\|X\|_m = \{\mathbb{E}\|X\|^m\}^{1/m}$. L stands for the backshift operator ($LX_k = X_{k-1}$).

Definition 1. *Let η_t be a sequence of independent and identically distributed (i.i.d) random variables with mean zero and variance one. ε_t is called the generalized* *autoregressive conditionally heteroscedastic process or GARCH(p, q) model if*

$$\varepsilon_t = \sigma_t \eta_t \qquad (t \in \mathbb{Z}), \tag{1}$$

where σ_t^2 is a non-negative process under the non-negativity assumptions of ω, $(\alpha_i)_{1 \le i \le p}$ and $(\beta_j)_{1 \le j \le q}$ such that

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (t \in \mathbb{Z}). \tag{2}$$

It is worth noting that the main properties of GARCH model focus on the non-negativity assumptions on its parameters. Hence, at a first stage, we will be interested in providing the stationary property as well as the condition of existence of moments for the GARCH$(1, 1)$ model as given in (3) without any positivity restrictions on conditional variance coefficients, namely $\alpha_1 := \alpha$ and $\beta_1 := \beta$. The non-negativity of σ_t^2 will be studied independently in section (3.3).

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \ , \quad \eta_t \sim iid\,(0, 1) \\ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \ \text{ with } \omega > 0 \end{cases} \qquad (t \in \mathbb{Z}). \tag{3}$$

2.1 Stationarity

We shall focus on the second-order stationary of GARCH process. The corresponding condition has been established by Bollerslev (1986) depending on the positivity condition of the parameters. The following lemma aims at deriving a sufficient second-order stationary condition associated with the GARCH model (3).

Lemma 1. *Let $a_1, a_2, \ldots, a_s$ be a real numbers with $a_s \neq 0$. Then, all roots of $A(z) = 1 - \sum_{i=1}^{s} a_i z^i$ lie out of the unit circle if and only if $\sum_{i=1}^{s} |a_i| < 1$.*

Proposition 2 (**Second-order stationarity**)**.** *A process ε_t satisfying the GARCH(1,1) model given by (3) is second order stationary if*

$$|\alpha| + |\beta| < 1. \tag{4}$$

It must be stated that condition (4) coincides with the Bollerslev's weak stationary condition when parameters are assumed a priori non-negative. More generally, (4) holds for all α and β. Furthermore, (4) is sufficient to ensure that B's roots lie also out the unit circle using the same Lemma (1) and that $|\beta| < 1 - |\alpha| < 1$ (see the Appendix). Thus, from (1) and (23), we achieve:

$$\mathbb{E}\varepsilon_t^2 = \mathbb{E}\sigma_t^2 = \frac{\omega}{A(1)} = \frac{\omega}{1 - \alpha - \beta}.$$

2.2 Moment structure

The necessary and sufficient condition for the existence of the even order moment of the GARCH(1,1) model was provided by Bollerslev (1986), whereas the sufficient condition for the existence of the higher order moments of the GARCH(p, q) model was given by Ling (1999). Ling and McAlleer (2002) claim that Ling's condition is also necessary as given in the following Theorem.

Theorem 3 (**Ling and McAlleer 2002, Theorem 2.1**). *Under positivity assumptions of parameters and for any integer* m*, the necessary and sufficient condition for existence of* $2m - th$ *moment of GARCH process is* $\rho\left(\mathbb{E}A_t^{\otimes m}\right) = min\left\{\left|eigen\ values\ of\ \mathbb{E}A_t^{\otimes m}\right|\right\} < 1$ *where*

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

However, we are dealing here with the non-negativity constraints on parameters on which the theorem's proof given by Ling and McAlleer (2002) relies. Thus, in order to provide a sufficient condition ensuring the existence of $2m - th$ moment of the GARCH model (3), we refer to the Proposition 2.2 in Francq et al. (2013) whose proof was adapted to obtain the following Proposition.

Proposition 4. *Let* ε_t *be a GARCH(p, q) process with relaxed positivity conditions as in (3). A sufficient condition for* $\mathbb{E}\varepsilon_t^{2m} < \infty$ *is* $\rho(A^{(m)}) < 1$.

In special case of $m = 2$, $A_t^{(2)} = \mathbb{E}\left\{\left(\eta_t^2 \quad 1\right)'^{\otimes 2}\right\} \left(|\alpha| \quad |\beta|\right)^{\otimes 2}$ has one non-zero eigenvalue equals to its trace given by $\mu_4\alpha^2 + \beta^2 + 2|\alpha\beta|$. Hence, the fourth moment of ε_t exists if

$$\mu_4\alpha^2 + \beta^2 + 2|\alpha\beta| < 1, \tag{5}$$

where μ_4 stands for the fourth moment of η_t. Again, condition (5) is identical with Bollerslev's moment condition when α and β are non-negative. Moreover, for α and β verifying (4) and (5), a direct computations give

$$\mathbb{E}\varepsilon_t^4 = \mu_4\mathbb{E}\sigma_t^4 = \mu_4\mathbb{E}\left(\omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2\right)^2.$$

This yields

$$\mathbb{E}\varepsilon_t^4 = \frac{\omega^2(1 + \alpha + \beta)\mu_4}{(1 - \alpha - \beta)(1 - \mu_4\alpha^2 - \beta^2 - 2\alpha\beta)}. \tag{6}$$

3 Robust estimation of the conditional variance

3.1 State-Space representation of GARCH(1,1)

Under conditions (4) and (5), the GARCH(1,1) model given by (3) can be represented in a state space form based on the innovation $\nu_t = \varepsilon_t^2 - \sigma_t^2$ being a white noise of variance $\mathbb{E}\nu_t^2 = \mathbb{E}(\varepsilon_t^2 - \sigma_t^2)^2 = (\mu_4 - 1)\mathbb{E}\sigma_t^4$, assumed to be Gaussian. Such class of state space representations has been approached by Anderson and Moore (1979). Then, we suggest the following state space model

$$\sigma_t^2 = \omega + (\alpha+\beta)\sigma_{t-1}^2 + \alpha\nu_{t-1}, \tag{7}$$

$$\varepsilon_t^2 = \sigma_t^2 + \nu_t, \tag{8}$$

where (7) and (8) represent respectively the transition and the measurement equations associated respectively with the state variable σ_t^2, and the observations ε_t^2. Moreover σ_0^2 is assumed Gaussian independent of $(\nu_t)_{t\geq 0}$. Such assumption provides the Gaussian distribution of the process $\left\{\left(\varepsilon_t^2 \;\; \sigma_t^2\right)', t \geq 0\right\}$.

Unlike other GARCH state space representations reported in the literature, the innovation model (7)–(8) has some particular properties as compared with others. First, it can be obtained from any other state space model. Indeed, Definition (3.2) in Anderson and Moore (1979) shows how any such model can be transformed on an innovation form. Furthermore, it is essentially unique. This can be checked through Theorem (3.2) in Anderson and Moore (1979) which prove the uniqueness of such representation. Thus, in regards to (7)-(8), the condition $\alpha \neq 0$ is necessarily required.

3.2 Kalman filtering

We consider the GARCH state space model (7)-(8) given under assumptions (4) and (5). Let $\widehat{\sigma}_{t^-}^2$ and $\widehat{\sigma}_t^2$ denote respectively the unconstrained filtered, and unconstrained predicted estimates of σ_t^2. P_{t^-} and P_t are their respective error covariances. Starting from initial conditions

$$\widehat{\sigma}_0^2 = \mathbb{E}\sigma_0^2 \text{ and } P_0 = \mathbb{E}\left(\sigma_0^2 - \mathbb{E}\sigma_0^2\right)^2.$$

Let's assume that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ have been observed. Then, for $t = 1, \ldots, n$, the recursive expressions (9)-(13) characterize the Kalman filter applied to the GARCH state space model (7)-(8).

Prediction equations

$$\widehat{\sigma}_{t^-}^2 = \omega + (\alpha+\beta)\widehat{\sigma}_{t-1}^2, \tag{9}$$

$$P_{t^-} = (\alpha+\beta)^2 P_{t-1} + \frac{2\alpha^2\omega^2(1+\alpha+\beta)}{(1-\alpha-\beta)(1-\mu_4\alpha^2-\beta^2-2\alpha\beta)}. \tag{10}$$

Kalman gain

$$K_t = P_{t^-}\left[P_{t^-} + \frac{2\omega^2(1+\alpha+\beta)}{(1-\alpha-\beta)(1-\mu_4\alpha^2-\beta^2-2\alpha\beta)}\right]^{-1}. \qquad (11)$$

Updating equations

$$\widehat{\sigma}_t^2 = \widehat{\sigma}_{t^-}^2 + K_t\left(\varepsilon_t^2 - \widehat{\sigma}_{t^-}^2\right), \qquad (12)$$

$$P_t = P_{t^-}(1-K_t). \qquad (13)$$

It is interesting to note that the Kalman filter applied to the innovation model (7)-(8) provides estimates of state σ_t^2 with zero error, i.e. $\widehat{\sigma}_{t^-}^2 = \sigma_t^2$ and the filter measurement innovation process $\widetilde{\varepsilon}_t^2 = \varepsilon_t^2 - \widehat{\varepsilon}_{t^-}^2$ is identical with the innovation model ν_t, which correspond to the definition of ν_t and can be seen using Theorem (3.3) in Anderson and Moore (1979) as well in a general frame.

3.3 The robustification

We propose a correction of the conditional variance estimated in (3.2) aimed in preventing it from being negative independently from the sign of parameter model using the probability density function (pdf) truncation method (see Simon 2006) as one of useful constrained Kalman filter for linear boundedness constraint on the state variable. It consists precisely of taking the probability density function computed by the Kalman filter (assumed to be Gaussian) and truncating it at the constraint boundaries. The constrained conditional variance estimate is then obtained as function of the expectation of the truncated probability density. As for the non-negativity constraint of σ_t^2, one can suppose that at each time t and for a constant N, empirically set, we have

$$\frac{1}{N} \leq \sigma_t^2 \leq N. \qquad (14)$$

Note that (14) is not the unique shape of constraints. Indeed, the boundaries of (14) can change according to the information available a priori on volatility. Moreover, the pdf truncation method allows to take account all shapes of boundedness as described in the following proceeding steps.

Proceeding steps

Let's start by initializing the constrained conditional variance such that

$$\widetilde{\sigma}_{t0}^2 = \widehat{\sigma}_{t^-}^2 \quad \textit{and} \quad \widetilde{P}_{t0} = \widehat{P}_{t^-}.$$

Now we generate a realization $\sigma_t^2 \sim \mathcal{N}\left(\widehat{\sigma}_{t^-}^2, \widehat{P}_{t^-}\right)$ and let's perform the following transformation

$$\Sigma_t = \frac{\sigma_t^2 - \widetilde{\sigma}_{t0}^2}{\sqrt{\widetilde{P}_{t0}}}. \qquad (15)$$

Therefore, inequality (14) is transformed to $l_t \leq \Sigma_t \leq u_t$, with

$$l_t = \frac{1 - N\widetilde{\sigma}^2_{t0}}{N\sqrt{\widetilde{P}_{t0}}} \quad \textit{and} \quad u_t = \frac{N - \widetilde{\sigma}^2_{t0}}{\sqrt{\widetilde{P}_{t0}}}.$$

We define Σ^*_t as the random variable having the pdf of Σ_t truncated and normalized between the limits l_t and u_t. Let μ_{Σ^*} be the truncated expectation of Σ^*_t. We take then the inverse of the transformation (15) to obtain the conditional variance estimate after enforcement of positivity constraint (14). This yields

$$\widetilde{\sigma}^2_{t^-} = \sqrt{\widetilde{P}_{t0}}\,\mu_{\Sigma^*} + \widetilde{\sigma}^2_{t0}. \tag{16}$$

It is interesting to note that the importance of the approach proposed above is in no way lessened by the ease with which it can be seen. Indeed, it should not only be seen as a way to avoid the non-negativity conditions on parameters as well as the limitations that they impose as discussed previously, but also as a statistical benchmark of the parameter estimation using the information on volatility. Simulation 2 in section (5) provides evidence for the improvement of the parameter estimation quality obtained in the presence of additional information on volatility.

4 GARCH(1,1) parameter estimation

4.1 Quasi-likelihood estimate

Financial time series often exhibit excess kurtosis, so-called leptokurtic behavior which can not be taken account with Gaussian innovation assumption (Bollerslev 1987). The quasi-likelihood estimation methods for GARCH(1,1) has been proposed and well studied in literatures (e.g. Ling and McAleer (2003), Francq and Zakoian (2007)). We propose estimating the GARCH(1,1) parameters by quasi-maximum likelihood estimation using SPSA algorithm for two distributions of η_t, namely, Gaussian and Student-t.
Let ε_t be the GARCH(1,1) model defined by (3). Let's denote by $\theta = (\omega, \alpha, \beta)'$ the parameter vector satisfying only conditions (4) and (5). Define Θ as the subset of $\mathbb{R}^3$ such that

$$\Theta = \left\{\theta \in \mathbb{R}^3 / \ \omega > 0\ ,\ |\alpha| + |\beta| < 1\ ,\ \beta^2 + 2\,|\alpha\beta| + \mu_4\alpha^2 < 1\right\},$$

where $\mu_4 = 3$ for the standard Gaussian distribution and $\mu_4 = \dfrac{3\nu^2}{(\nu-2)(\nu-4)}$ for the student-t distribution with ν degrees of freedom.
From the observed data $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, the quasi log-likelihood and the log-likelihood

are estimated respectively for the Gaussian and Student-t distributions, for all $\theta \in \Theta$ by

$$\widehat{L}_n(\theta;\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n) = -\frac{n}{2}log(2\pi) - \frac{n}{2}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{\widetilde{\sigma}_{t^-}^2(\theta)} + log\,\widetilde{\sigma}_{t^-}^2(\theta)\right) \quad (17)$$

and

$$\widehat{L}_n(\theta;\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n) = n\,log\left[\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}\right] - \frac{n}{2}\times$$
$$\times\left[\frac{1}{n}\sum_{t=1}^{n} log\left((\nu-2)\,\widetilde{\sigma}_{t^-}^2(\theta)\right) + (\nu+1)\,log\left(1+\frac{\varepsilon_t^2}{(\nu-2)\,\widetilde{\sigma}_{t^-}^2(\theta)}\right)\right] \quad (18)$$

Thus, maximizing (17) and (18) is respectively equivalent to minimizing, with respect to $\theta \in \Theta$

$$\widehat{l}_n(\theta;\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n) = \frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{\widetilde{\sigma}_{t^-}^2(\theta)} + log(\widetilde{\sigma}_{t^-}^2(\theta)) \quad (19)$$

and

$$\widehat{l}_n(\theta;\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n) = \frac{1}{n}\sum_{t=1}^{n} log\left((\nu-2)\,\widetilde{\sigma}_{t^-}^2(\theta)\right) + (\nu+1)\times$$
$$\times\, log\left(1+\frac{\varepsilon_t^2}{(\nu-2)\,\widetilde{\sigma}_{t^-}^2(\theta)}\right). \quad (20)$$

Thus, both likelihoods are completely defined since $\widetilde{\sigma}_{t^-}^2$ is robustified so that its non-negativity is ensured without any positivity restrictions on the parameters α and β.

Remark. *A more general framework for estimating the GARCH model in the case where the distribution of innovation is unknown is contained in the works of Francq, Wintenberger, and Zakoïan (2013), Fan, Qi, and Xiu (2014), Zhu and Xie (2016).*

4.2 SPSA maximization

At this stage, we are interested in using an iterative algorithm for the maximization of l_n based on measurements of $\widehat{l}_n$ only, not on the measurements of its derivatives (gradient, hessian) as it is used in most gradient descent methods. Spall (1992,1998), invented the simultaneous perturbation stochastic approximation (SPSA) which rely on the approximation of the gradient using only two measurements of $\widehat{l}_n$ for a parameter vector of any dimension and exhibits fast convergence. It should be noted that the choice of constants and gain sequences are the same as those used in Spall (1998). The step-by-step summary below shows how SPSA was applied to minimize the $\widehat{l}_n$.

Step 1: Give initial $\theta_0 \in \Theta^\circ$ and non-negative coefficient $a = 0.16$, $c = 0.5$, $\lambda = 0.602$, $\gamma = 0.101$ and $A = 10\%$ of the iteration number. Θ° stands for the interior of Θ.

Step 2: Compute gain sequences $a_k = a(A + k + 1)^{-\lambda}$ and $c_k = c(k + 1)^{\gamma}$.

Step 3: Generate a 3-dimensional random perturbation vector Δ_k having Bernoulli distribution, ± 1-valued with probability of $\frac{1}{2}$.

Step 4: Check the existence of 4-th moment from (5) as well as $\omega > 0$.

Step 5: Obtain two measurements of the $\widehat{l}_n$ based on the simultaneous perturbation around the current $\widehat{\theta}_k$, as follows

$$y_k^{\pm}(\widehat{\theta}_k) = \widehat{l}_n(\widehat{\theta}_k \pm c_k \Delta_k) + \delta_k^{\pm}$$

where δ_k^- and δ_k^+ are two independent random vectors having Uniform distribution over the interval $[0, 1]$.

Step 6: Generate the simultaneous perturbation approximation of the gradient $\widehat{g}(\widehat{\theta}_k)$

$$\widehat{g}(\widehat{\theta}_k) = \frac{y_k^+(\widehat{\theta}_k) - y_k^-(\widehat{\theta}_k)}{2c_k}(\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, \Delta_{k3}^{-1})'$$

where Δ_{ki}, $i = 1, 2, 3$ is the ith component of Δ_k vector.

Step 7: Check the second order stationarity condition from (4).

Step 8: Use the recursive stochastic updating

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k - a_k \widehat{g}(\widehat{\theta}_k). \tag{21}$$

Step 9: Return to Step 2 with $k + 1$ replacing k.

Step 10: Terminate the algorithm if the sequence (θ_k) converges, or the maximum allowable number of iterations has been reached.

5 Simulation evidence

In this section, we will conduct three simulation studies. The two first simulations aime at evaluating the performance of the proposed robust estimation of the conditional variance in estimating the GARCH(1,1) parameters, whereas the last simulation tends to shed some light on the usefulness of our approach in improving the quality of the parameter estimates in the case of availability a priori of some information on volatility. The absolute error (AE), the mean absolute error (MAE) and the mean squared error (MSE) are used to measure the difference between the true and estimated parameters. Parameters together with the unconditional variance, denoted $\overline{\sigma}^2$, are represented together by the vector $\pi := (\theta', \overline{\sigma}^2)'$, where $\pi_0 := (\theta_0', \overline{\sigma}_0^2)'$ contains the true initial values. Referring to Carnero et al. (2012) and Bahamonde and Veiga (2016), including the unconditional variance for the results analysis is justifiable insofar as the accurate estimation of the unconditional variance is also a significant measure of the volatility estimation quality. Indeed, notice that the volatility estimation error can be written as

$$\begin{aligned}
\xi_t &= \widehat{\sigma}_t^2 - \sigma_t^2 = \\
&= (\widehat{\omega} - \omega) + (\widehat{\alpha} - \alpha)\varepsilon_{t-1}^2 + (\widehat{\beta} - \beta)\sigma_{t-1}^2 + \widehat{\beta}\xi_{t-1} = \\
&= \sum_{i=0}^{t-2} \left[(\widehat{\omega} - \omega) + (\widehat{\alpha} - \alpha)\varepsilon_{t-i-1}^2 + (\widehat{\beta} - \beta)\sigma_{t-i-1}^2\right] \widehat{\beta}^i + (\widehat{\sigma}_1 - \sigma_1)\widehat{\beta}^{t-1}.
\end{aligned}$$

It's, then, clear that the expected error depends, not only on the parameter biases, but also on the unconditional variance $\overline{\sigma}^2 = \mathbb{E}\sigma_1^2 = \omega/(1 - \alpha - \beta)$.
In the following simulation studies, we use notation QML (resp. ML) for the standard quasi-maximum likelihood estimation (resp. maximum likelihood estimation) with corresponding estimate $\widehat{\pi}$ and Q-CK for the estimation by the proposed algorithm with corresponding estimate $\widetilde{\pi}$. The number of replications used for all simulations is 1000.

Table 1: MAE and MSE of estimated parameters with Gaussian innovation assumption

		Q-CK			QML		
n	π_0	$\widetilde{\pi}$	MAE($\widetilde{\pi}$)	MSE($\widetilde{\pi}$)	$\widehat{\pi}$	MAE($\widehat{\pi}$)	MSE($\widehat{\pi}$)
500	1.5	1.5026	0.0079	0.0001	1.5696	0.3278	0.1698
	0.3	0.3056	0.0080	0.0001	0.3059	0.0515	0.0043
	0.2	0.1976	0.0089	0.0001	0.1752	0.1153	0.0192
	3	3.0271	0.0546	0.0068	3.0509	0.2711	0.1304
1000	1.5	1.4999	0.0070	$< 10^{-4}$	1.5135	0.2144	0.0780
	0.3	0.3050	0.0071	$< 10^{-4}$	0.3022	0.0448	0.0031
	0.2	0.1995	0.0073	$< 10^{-4}$	0.1852	0.0903	0.0129
	3	3.0295	0.0529	0.0068	2.9634	0.1682	0.0486
5000	1.5	1.5016	0.0081	0.0001	1.5075	0.1186	0.0227
	0.3	0.3063	0.0080	0.0001	0.3011	0.0205	0.0007
	0.2	0.2017	0.0080	0.0001	0.1969	0.0454	0.0033
	3	3.0554	0.0690	0.0122	3.0667	0.0909	0.0136

5.1 Simulation study: Gaussian distribution

We assume the true distribution of η_t as Gaussian with 0 means and variance 1. Three series of the GARCH(1,1) process of size $n \in \{500, 1000, 5000\}$, where $\pi_0 = (1.5, 0.3, 0.2, 3)$. Table (1) reports the Monte Carlo means, MAE and MSE of the parameter and the unconditional variance estimates. Thus, as reported in table (1), it is observed that both MAE and MSE of the Q-CK estimates are on the same scale for all sample sizes and all widely smaller than those obtained by QML estimation. Yet, the unconditional variance is satisfactorily estimated by Q-CK approach, and once again, the corresponding MAE and MSE are both less than those obtained by QML estimation. From the simulation results, it is expected that the proposed algorithm estimates better the volatility. For this purpose, we complete the evaluation of Q-CK estimation accuracy considering three trajectories generated by the same GARCH(1,1) given previously. Figure (1) shows that altogether, the simulated volatility of all generated series is completely close to the estimated volatility. In parallel, except for the size 5000, the outcomes displayed in table (2) confirm the accuracy of the Q-CK method for the estimation of the volatility in view of the smallest values of the MAE of the Q-CK estimates compared to those obtained by QML.

Table 2: MAE of the estimated Q-CK and QML volatilities, denoted respectively by $\widetilde{\sigma_t}$ and $\widehat{\sigma_t}$, with Gaussian innovation assumption

$(\omega_0, \alpha_0, \beta_0)$	n	MAE($\widetilde{\sigma_t}$)	MAE($\widehat{\sigma_t}$)
	500	0.0047	0.1022
(1.5, 0.3, 0.2)	1000	0.0030	0.0088
	5000	0.0081	0.0014

5.2 Simulation study: Student-t distribution

Now, we set the true distribution of η_t as Student-t with $\nu = 5$ for which the existence of the fourth moment is ensured. We generate three series of GARCH(1,1) process with true parameter values $\pi_0 = (1.2, 0.07, 0.04, 1.34)$, with the same sample sizes considered previously. Simulation results are reported in table (3) where it can be seen that the Q-CK parameter estimates, including the unconditional variance, still have smaller MAE and MSE compared with those estimated by ML estimation. Additionally, from figure (2), the estimated volatility using the Q-CK method is very close to the true volatility for all simulation sizes with the smallest values of MAE compared to the ML estimates, except for size 5000 as displayed in (4).
Broadly speaking, the comparison of the results of the two distribution cases leads to the conclusion that the Q-CK estimation based on the Gaussian distribution

outperforms its Student-t version in terms of MAE and MSE, while the accuracy of the unconditional variance remains stable for both cases.

Table 3: MAE and MSE of estimated parameters with Student-t innovation assumption

		Q-CK			ML		
n	π_0	$\widetilde{\pi}$	MAE($\widetilde{\pi}$)	MSE($\widetilde{\pi}$)	$\widehat{\pi}$	MAE($\widehat{\pi}$)	MSE($\widehat{\pi}$)
500	1.2	1.2002	0.0241	0.0008	0.7438	0.4922	0.4173
	0.07	0.0704	0.0218	0.0007	0.0782	0.0521	0.0044
	0.04	0.0493	0.0260	0.0010	0.3788	0.3669	0.2567
	1.34	1.3669	0.0599	0.0063	1.3568	0.1726	0.0743
1000	1.2	1.1991	0.0240	0.0008	0.4744	0.2144	0.4091
	0.07	0.0734	0.0245	0.0010	0.0626	0.0401	0.0021
	0.04	0.0475	0.0235	0.0008	0.3772	0.3610	0.2537
	1.34	1.3676	0.0624	0.0061	1.3094	0.1522	0.0899
5000	1.2	1.2028	0.0248	0.0009	1.0873	0.1636	0.0709
	0.07	0.0743	0.0256	0.0010	0.0693	0.0184	0.0005
	0.04	0.0593	0.0291	0.0013	0.1261	0.1211	0.0425
	1.34	1.3917	0.0717	0.0077	1.3522	0.0390	0.0024

Table 4: MAE of the estimated Q-CK and ML volatilities, denoted respectively by $\widetilde{\sigma_t}$ and $\widehat{\sigma_t}$, with Student-t innovation assumption

$(\omega_0, \alpha_0, \beta_0)$	n	MAE($\widetilde{\sigma_t}$)	MAE($\widehat{\sigma_t}$)
$(1.2, 0.07, 0.04)$	500	0.0061	0.0205
	1000	0.0061	0.0112
	5000	0.0155	0.0035

5.3 Effect of additional information on volatility

In order to reflect one special situation when the conditional variance parametrization is enriched by additional information on volatility process, we carry out a simulation of sample size $n = 1000$ of a GARCH(1,1) process with true parameter values $\omega_0 = 1.5$, $\alpha_0 = 0.4$ and $\beta_0 = 0.1$. We assume that a prior information on volatility is available and given for all $t = 1, \dots, n$ as follows:

$$\frac{1}{N} + \sigma_{t,0} - 1 \le \sigma_t \le N + \sigma_{t,0} - 1, \tag{22}$$

where $\sigma_{t,0}$ is the true volatility of the generated GARCH(1,1) (in practice $\sigma_{t,0}$ can be replaced by the realized volatility estimate under stability assumption of volatility in several periods) and N is a constant greater than 1. Then, it is clear that σ_t approaches the true value $\sigma_{t,0}$ as $N \to 1$. In other words, more N approaches 1,

Figure 1: Volatility estimates with Gaussian innovation assumption using a logarithmic scale respectively for $n = 500, 1000$ and 5000. Black dashed line represents the original volatility, and the grey continuous line indicates the estimated Q-CK volatility

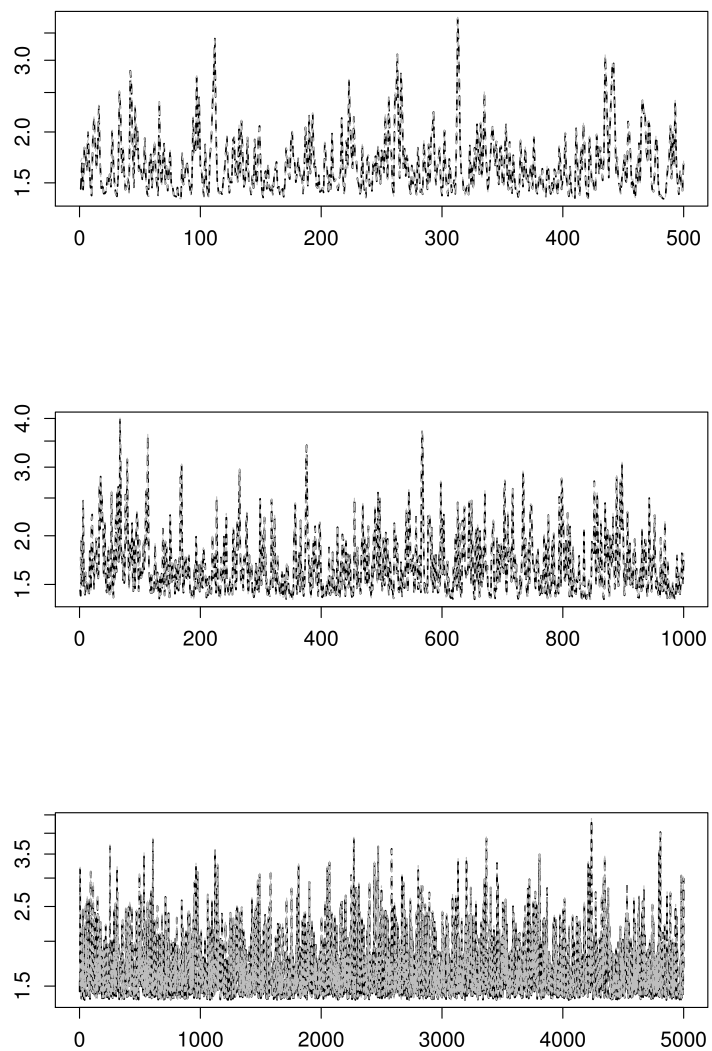
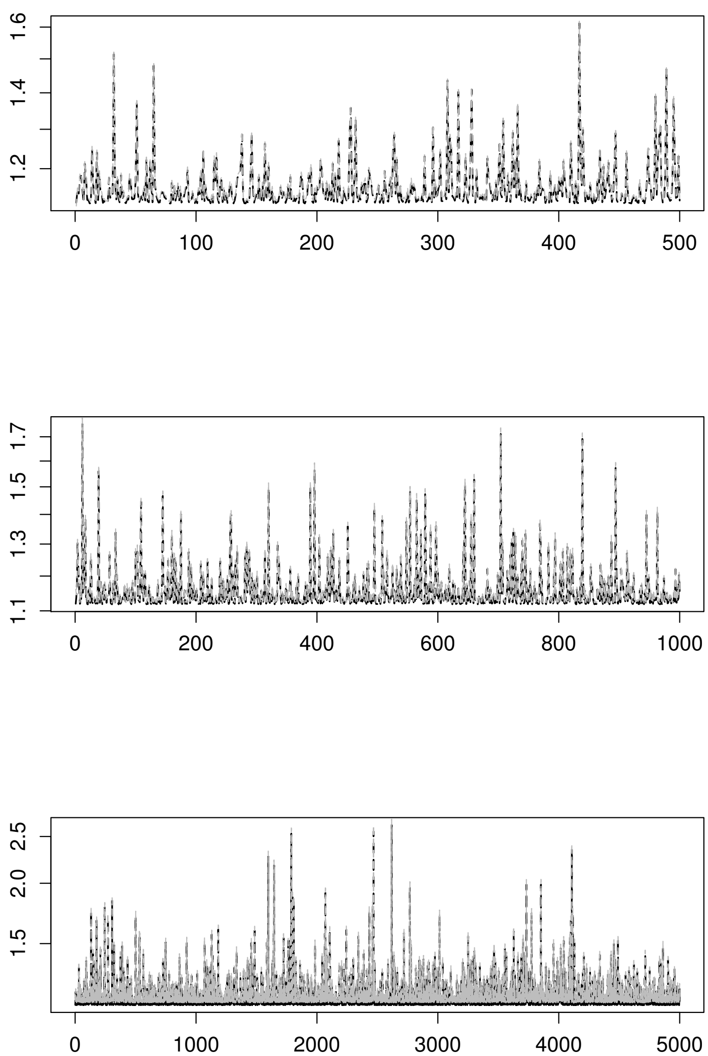


Figure 2: Volatility estimates with Student-t innovation assumption using a logarithmic scale respectively for $n = 500, 1000$ and 5000. Black dashed line represents the original volatility, and the grey continuous line indicates the estimated Q-CK volatility

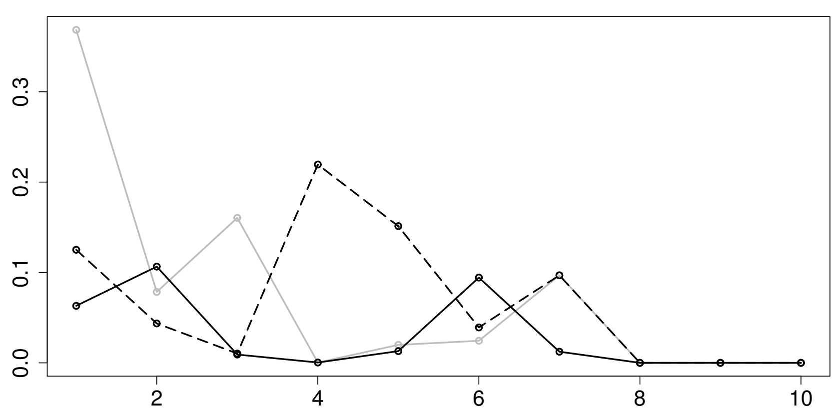


more the information on σ_t is accurate and, conversely, the higher a value of N, the less accurate this information. Thus, let $N = 1 + 10^{-i}$, for $i \in [\![1, 10]\!]$ and let's fit a GARCH(1,1) using the Q-CK algorithm. Table (5) as well as figure (3) provide evidence that the estimation accuracy is improved exploiting the available information (22). Indeed, the AE of parameter estimates fall to 0 as soon as $i \geq 8$. In other word, the more accurate the information on volatility, the more accurate the parameter estimates.

Table 5: AE of estimated parameters according to the information level i

i	$\widetilde{\omega}$	AE($\widetilde{\omega}$)	$\widetilde{\alpha}$	AE($\widetilde{\alpha}$)	$\widetilde{\beta}$	AE($\widetilde{\beta}$)
1	1.6251	0.1251	0.0315	0.3684	0.1630	0.0630
2	1.4564	0.0435	0.3215	0.0784	0.2064	0.1064
3	1.4897	0.0102	0.2395	0.1604	0.1091	0.0091
4	1.2805	0.2194	0.4003	0.0003	0.1003	0.0003
5	1.6512	0.1512	0.4198	0.0198	0.0870	0.0129
6	1.5391	0.0391	0.3755	0.0244	0.0056	0.0943
7	1.4031	0.0968	0.4968	0.0968	0.1123	0.0123
8	1.5000	0.0000	0.4000	0.0000	0.1000	0.0000
9	1.5000	0.0000	0.4000	0.0000	0.1000	0.0000
10	1.5000	0.0000	0.4000	0.0000	0.1000	0.0000

Figure 3: AE of Q-CK estimated parameters according to the information accuracy on volatility. Dashed line indicates the AE($\widetilde{\omega}$), grey and black continuous lines represent respectively the AE($\widetilde{\alpha}$) and AE($\widetilde{\beta}$)



6 Conclusions

The non-negativity constraint of the conditional variance plays an important part in the definedness of the GARCH model and the estimation performance as well because of its reliance on the GARCH parameters sign. Bearing this fact in mind, we propose, in this paper, a robust estimation of the conditional variance which aime at guaranting its non-negativity independently of the parameter sign so that the likelihood function would be well defined. The results of our evaluation indicated that the constrained Kalman filter is advantageous for estimating the conditional variance with respect to the non-negativity constraint. Simulations demonstrated the effectiveness of the proposed approach in improving the accuracy of the QML estimation, notably, for taking account other boundedness information beyond non-negativity.

References

[1] Allal J., Benmoumen M., (2011), Parameter Estimation for GARCH(1,1) Models Based on Kalman Filter, *Advances and Applications in Statistics* 25, 115–30.

[2] Anderson B., Moore J., (1979), *Optimal Filtering*, Prentice Hall, 230–238.

[3] Bahamonde N., Veiga H., (2016), A robust closed-form estimator for the GARCH (1,1) model, *Journal of Statistical Computation and Simulation* 86(8), 1605–1619.

[4] Bollerslev T., (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31(3), 307-327.

[5] Bollerslev T., (1987), A conditionally heteroskedastic time series model for speculative prices and rates of return, *The review of economics and statistics*, 542–547.

[6] Carnero M. A., Peña D., Ruiz E., (2012), Estimating GARCH volatility in the presence of outliers, *Economics Letters* 114(1), 86–90.

[7] Engle R. F., (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50(4), 987–1007.

[8] Fan J., Qi L., Xiu D., (2014), Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods, *Journal of Business & Economic Statistics* 32(2), 178–191.

[9] Francq C., Zakoïan J. M., (2007), Quasi-maximum likelihood estimation in garch processes when some coefficients are equal to zero, *Stochastic Processes and their Applications* 117(9), 1265–1284.

[10] Francq C., Wintenberger O., Zakoïan J. M., (2013), GARCH models without positivity constraints: Exponential or Log GARCH, *Journal of Econometrics* 177(1), 34–46.

[11] Ghalanos A. (2014), Rugarch: Univariate GARCH models. R package version 1.4-0, accessed 16 January 2019, available at: `https://cran.r-project.org/web/packages/rugarch/rugarch.pdf`.

[12] Ling S., (1999), On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model, *Journal of Applied Probability* 36(3), 688–705.

[13] Ling S., McAleer M., (2002), Necessary and sufficient moment conditions for the GARCH (r,s) and asymmetric power GARCH (r,s) models, *Econometric Theory* 18(3), 722–729.

[14] Ling S., McAleer M., (2003), Asymptotic theory for a vector arma-garch model, *Econometric Theory* 19(2), 280–310.

[15] Luethi D., Erb P., & Otziger S., (2018), FKF: Fast Kalman Filter. R package version 0.1.5, accessed 16 July 2018, available at: `https://http://cran.univ-paris1.fr/web/packages/FKF/FKF.pdf`.

[16] Nelson D. B., Cao C. Q., (1992), Inequality Constraints in the Univariate GARCH Model, *Journal of Business & Economic Statistics* 10(2), 229–235.

[17] Ossandón S., Bahamonde N., (2013), A new nonlinear formulation for GARCH models, *Comptes Rendus Mathematique* 351(5-6), 235–239.

[18] Simon D., (2006), *Optimal state estimation*, John Wiley & Sons, 218–223.

[19] Spall J. C., (1992), Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation, *IEEE Transactions on Automatic Control* 37(3), 332–341.

[20] Spall J. C., (1998), Implementation of the simultaneous perturbation algorithm for stochastic optimization, *IEEE Transactions on aerospace and electronic systems* 34(3), 817–823.

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

[22] Zhu X., Xie L., (2016), Adaptive quasi-maximum likelihood estimation of GARCH models with Student'st likelihood, *Communications in Statistics-Theory and Methods* 45(20), 6102-6111.

Appendix: Proofs

Proof of Lemma 1. First, assuming that $\sum_{i=1}^{s}|a_i| \geq 1$. Since $A(0) = 1$ and $A(1) \leq 0$, we have then at least one solution of $A(z) = 0$ in the interval $]0, 1]$. This contradicts the given condition.
On the other side, for any $|z| \leq 1$, we have

$$|A(z)| = \left|1 - \sum_{i=1}^{s} a_i z^i\right| \geq \left|1 - \left|\sum_{i=1}^{s} a_i z^i\right|\right| \geq 1 - \left|\sum_{i=1}^{s} a_i z^i\right| \geq 1 - \sum_{i=1}^{s}|a_i||z^i| \geq 1 - \sum_{i=1}^{s}|a_i| > 0.$$

Which allows to conclude. □

Proof of Proposition 2. Let's start from the ARMA(1,1) representation of ε_t^2, which we rewrite in the polynomial form as

$$A(L)\varepsilon_t^2 = \omega + B(L)\nu_t, \tag{23}$$

where

$$A(L) = 1 - (\alpha + \beta)L \quad \textit{and} \quad B(L) = 1 - \beta L.$$

Then, it can be easily seen from lemma (1) that ε_t^2 is invertible since $|\alpha + \beta| \leq |\alpha| + |\beta| < 1$. Thus, since ν_t is a white noise, the Proposition follows.

□

Proof of Proposition 4. Let's start from the Markovian representation of GARCH(p, q) model:

$$Z_t = b_t + A_t Z_{t-1},$$

where

$$Z_t = \left(\varepsilon_t^2, \ldots, \varepsilon_{t-p+1}^2, \sigma_t^2, \ldots, \sigma_{t-q+1}^2\right)' \in \mathbb{R}^{p+q}$$

and

$$b_t = (\omega\eta_t, 0 \ldots, 0, \omega, 0 \ldots, 0)' \in \mathbb{R}^{p+q}.$$

Let's rewrite Z_t as $Z_t = \sum_{l=0}^{\infty} Z_{t,l}$ where for all $l > 0$:

$$A_{t,l} = A_t A_{t-1} \ldots A_{t-l+1} \text{ and } Z_{t,l} = A_{t,l} b_{t-l}$$

with the convention that $A_{t,0} = I_{p+q}$ and $Z_{t,0} = b_t$. Thus, Z_t is almost surely defined without any restriction on the model coefficients. Let's denote $\widetilde{A}_{t,l} = \prod_{j=0}^{l-1} Abs(A_{t-j})$. Then, using the independence between the matrices in the product $A_t A_{t-1} \dots A_{t-l+1} b_{t-l}$, since A_{t-i} is a function of η_{t-i} and assuming a priori that σ_t^2 is non-negative, we get:

$$\begin{aligned}
\| Z_t \|_m = \| Abs(Z_t) \|_m \leq \sum_{l=0}^{\infty} \| \widetilde{A}_{t,l} Abs(b_{t-l}) \|_m = \\
= \sum_{l=0}^{\infty} \left\{ \mathbb{E} \| \widetilde{A}_{t,l} Abs(b_{t-l}) \|^m \right\}^{1/m} = \\
= \sum_{l=0}^{\infty} \left\{ \mathbb{E} \| \widetilde{A}_{t,l}^{\otimes m} Abs(b_{t-l})^{\otimes m} \| \right\}^{1/m} = \\
= \sum_{l=0}^{\infty} \left\{ \| \mathbb{E} \widetilde{A}_{t,l}^{\otimes m} Abs(b_{t-l})^{\otimes m} \| \right\}^{1/m}
\end{aligned}$$

since $\widetilde{A}_{t,l}^{\otimes m} Abs(b_{t-l})^{\otimes m}$ is non-negative, then:

$$\begin{aligned}
\| Z_t \|_m \leq \sum_{l=0}^{\infty} \left\{ \| \mathbb{E} \prod_{j=0}^{l-1} Abs(A_{t-j})^{\otimes m} Abs(b_{t-l})^{\otimes m} \| \right\}^{1/m} = \\
= \sum_{l=0}^{\infty} \left\{ \| \prod_{j=0}^{l-1} \mathbb{E} Abs(A_{t-j})^{\otimes m} \, \mathbb{E} Abs(b_{t-l})^{\otimes m} \| \right\}^{1/m} = \\
= \| b_1^{(m)} \|^{1/m} \sum_{l=0}^{\infty} \| (A^{(m)})^l \|^{1/m} .
\end{aligned}$$

Thus, $\rho(A^{(m)}) < 1$ entails $\| (A^{(m)})^l \| \to 0$, as $l \to \infty$, and then the sum converges. This completes the proof. □