Recursive estimation of GARCH processes

László Gerencsér†, Zsanett Orlovits‡ and Balázs Torma†

***Abstract*— ARCH processes and their extensions known as GARCH processes are widely accepted for modelling financial time series, in particular stochastic volatility processes. The off-line estimation of ARCH and GARCH processes have been analyzed under a variety of conditions in the literature. The main contribution of this paper is a rigorous convergence analysis of a recursive estimation method for GARCH processes with large stability margin under reasonable technical conditions. The main tool in the convergence analysis is an appropriate modification of the theory developed by Benveniste, Métivier and Priouret.**

I. Introduction

Analysis of financial data has received considerable attention in the literature during the last 20 years. Several models have been suggested to capture special features such as volatility clustering of financial time series. (This means that long periods of low volatility are followed by short periods of high volatility.) Similar phenomena has been experienced in telecommunication networks and EEG signals of epileptic patients. A minimum requirement for a mathematical model of these time-series is to ensure that the conditional variance of the observation process is time-varying. This is not true for linear processes. The first widely accepted non-linear stochastic volatility model, due to R. Engle [10], is the so-called ARCH (autoregressive conditional heteroscedastic) model. Here volatility is modelled as the output of a linear FIR (finite impulse response) system, combined with static non-linearities, driven by observed log-returns. In turn, log-returns are assumed to be defined as an i.i.d. process multiplied by the current volatility. Thus we get a stochastic non-linear feedback system, driven by an i.i.d. process. In the case of GARCH (generalized autoregressive conditional heteroscedastic) models, introduced by Bollerslev in [4], the linear FIR system is replaced by a general linear ARMA system. More precisely, (y_n), with $-\infty < n < +\infty$, is called a GARCH process of order (r, s) if it satisfies

$$y_n = \sigma_n \varepsilon_n,$$
$$(\sigma_n^2 - \gamma^*) = \sum_{i=1}^{r} \alpha_i^* (y_{n-i}^2 - \gamma^*) + \sum_{j=1}^{s} \beta_j^* (\sigma_{n-j}^2 - \gamma^*),$$

where σ_n^2 is the conditional variance of y_n given its own past up to time $(n-1)$, (ε_n) is an i.i.d. sequence of random variables with zero mean and unit variance, $\gamma^* = \mathrm{E} y_{n-i}^2 = \mathrm{E}\sigma_{n-j}^2 > 0$ and $\alpha_i^*, \beta_j^* \geq 0$, $i = 1, \ldots, r$, $j = 1, \ldots, s$ denotes the true, unknown parameters of the model. An alternative standard parametrization is obtained by defining

$$\alpha_0^* = \gamma^* \left(1 - \sum_{i=1}^{r} \alpha_i^* - \sum_{j=1}^{s} \beta_j^* \right).$$

The system parameter vector will then be defined as

$$\theta^* = (\alpha_0^*, \ldots, \alpha_r^*, \beta_1^*, \ldots, \beta_s^*).$$

Extensive experimental analysis have shown that even low order GARCH models, such as GARCH$(1, 1)$ fit well to financial data, see e.g. [10], [5], [22], [25].

The estimation, or identification, of the parameters of GARCH processes has attracted considerable attention recently, see e.g. [3], [18], [21], [23], [27]. All the cited works consider the so-called off-line estimation problem, when the collection of data and statistical analysis are separated in time. The weakest conditions for the strong consistency of an off-line quasi likelihood method has been given in Berkes et al., see [3].

However, given that financial time series are often sampled at high frequency, a more convenient, and less expensive approach would be to use an *on-line* or recursive method. Here, at time n, we use the estimate of the parameters at time $n-1$ and the observation at time n to update the estimated parameters at time point n. While there is an extensive literature on the recursive estimation of linear stochastic systems, see e.g. the books Ljung and Söderström [20] or Benveniste et. al. [2], the recursive estimation of GARCH processes has attracted little attention until recently. A recursive method for estimating the parameters of an ARCH process has been presented in Dahlhaus and Subba Rao [8]. For the recursive identification of GARCH models Aknouche and Guerbyenne [1] propose an algorithm based on the recursive least squares (LSQ) method. They show the consistency of the proposed algorithm only in some restricted sense under strong conditions on the underlying process. A recursive estimation method for GARCH processes, supported only by empirical evidence, is developed by Kierkegaard et al. [16].

The main contribution of this paper is a rigorous convergence analysis of a recursive estimation method for GARCH processes, with large stability margin, under reasonable technical conditions. The construction and the analysis is based on the theory of stochastic approximation with Markovian dynamics presented in Benveniste et. al [2], appropriately

† L. Gerencsér and B.Torma are with MTA SZTAKI, Computer and Automation Institute, Hungarian Academy of Sciences, Kende u. 13-17, 1111 Budapest, Hungary gerencser@sztaki.hu, torma@sztaki.hu

‡ Zs. Orlovits is with the Department of Differential Equations, Institute of Mathematics, Budapest University of Technology and Economics (BME), Műegyetem rkp. 3-9., 1111 Budapest, Hungary orlovits@math.bme.hu

ISBN 978-963-311-370-7

modified, in particular by applying a suitable resetting mechanism. We outline the proof of almost sure convergence and L_q-convergence up to certain q-s, using the results of [14]. Our results complement the results of [1], [8] in the sense that our algorithm will converge not only in almost sure sense but also in L_q up to certain q-s.

II. An overview of GARCH(r,s) processes

Let (y_n), with $-\infty < n < +\infty$, be a strictly stationary GARCH(r,s) process satisfying the equation

$$y_n = \sigma_n \varepsilon_n, \tag{1}$$

where σ_n^2 is the conditional variance of y_n given its own past up to time $(n-1)$, and (ε_n) is an i.i.d. sequence of random variables with zero mean and unit variances, and $\mathrm{E}\sigma_n^2 = \mathrm{E}y_n^2 < \infty$.

A key ingredient of GARCH models is a feedback mechanism in which σ_n^2 is defined in terms of past values y_{n-i}^2 via the linear dynamics

$$(\sigma_n^2 - \gamma^*) = \sum_{i=1}^{r} \alpha_i^* (y_{n-i}^2 - \gamma^*) + \sum_{j=1}^{s} \beta_j^* (\sigma_{n-j}^2 - \gamma^*), \tag{2}$$

with $\gamma^* = \mathrm{E}y_{n-i}^2 = \mathrm{E}\sigma_{n-j}^2 > 0$ and $\alpha_i^*, \beta_j^* \geq 0$, $i = 1,\ldots,r$, $j = 1,\ldots,s$. Defining the polynomials

$$C(z^{-1}) = \sum_{i=1}^{r} \alpha_i^* z^{-i}, \quad D(z^{-1}) = 1 - \sum_{j=1}^{s} \beta_j^* z^{-j},$$

equation (2) can be written in a compact form as

$$D^*(z^{-1})(\sigma^2 - \gamma^*) = C^*(z^{-1})(y^2 - \gamma^*), \tag{3}$$

where z^{-1} is the backward shift operator. In the following we will assume that the polynomials C^* and D^* are stable and relative prime. Defining the state vector

$$X_n^* = (y_n^2, \ldots y_{n-r+1}^2, \sigma_n^2, \ldots, \sigma_{n-s+1}^2),$$

it is easy to see that X_n^* satisfies a linear state equation

$$X_{n+1}^* = A_{n+1}^* X_n^* + u_{n+1}^*, \quad n \in \mathbb{Z}, \tag{4}$$

with

$$A_n^* = \begin{pmatrix} \eta^* \varepsilon_n^2 & \xi^* \varepsilon_n^2 \\ \bar{S} & 0 \\ \eta^* & \xi^* \\ 0 & \bar{S} \end{pmatrix} \quad \text{and} \quad u_n^* = \begin{pmatrix} \alpha_0^* \varepsilon_n^2 \\ 0 \\ \alpha_0^* \\ 0 \end{pmatrix}, \tag{5}$$

where

$$\eta^* = (\alpha_1^*, \ldots, \alpha_r^*), \quad \xi^* = (\beta_1^*, \ldots, \beta_s^*),$$

and $\bar{S}$ is the shift matrix having 1-s in the sub-diagonal, and 0-s elsewhere. It is easy to see that X_n^* is a Markov process.

A necessary and sufficient condition for the existence of a second order stationary solution of (2) is that

$$\sum_{i=1}^{r} \alpha_i^* + \sum_{j=1}^{s} \beta_j^* < 1, \tag{6}$$

see [4], [24]. Necessity follows trivially by taking expectation in equation (2), and noting that $\gamma^* = \mathrm{E}y_{n-i}^2 = \mathrm{E}\sigma_{n-j}^2 > 0$. It is easy to see that a second order stationary solution is necessarily strictly stationary.

Conditions for the existence of a unique, strictly stationary and causal solution of (1), without restrictions on the moments, can be formulated in terms of a state-space representation, (24). Thus, consider the so-called generalized linear autoregressive model (24) without specifying the matrices $A_n^* = A_n$ and the vectors $u_n^* = u_n$, instead, assuming only the following condition:

Condition 2.1: Assume that (A_n, u_n) is an i.i.d. sequence of random matrices, and the σ-fields

$$\mathcal{F}_{n+}^A = \sigma\{A_i : i > n\}, \quad \mathcal{F}_{n-}^u = \sigma\{u_i : i \leq n\}$$

are independent for any n.

Theorem 2.1: Consider the generalized linear autoregressive model (24) satisfying Condition 2.1. Assume that that the model (24) is irreducible and that both $\mathrm{E}\log^+ \|A_0\|$ and $\mathrm{E}\log^+ \|u_0\|$ are finite. Then (24) has a strictly stationary solution if and only if the top-Lyapunov exponent

$$\lambda(\mathcal{A}) := \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\|A_n \ldots A_1\|$$

is strictly negative.

The above result is due to Bougerol and Picard, [6]. A simple application of the above theorem is when we consider a GARCH-process with the stability condition (6). Then $\mathrm{E}A$ is *sub row-stochastic*, hence, by the Perron-Frobenius theorem $\rho(\mathrm{E}A) < 1$, which in turn implies $\lambda(\mathcal{A}) \leq \log\rho(\mathrm{E}A) < 0$, (the latter inequality being not quite trivial), from which the existence of a strictly stationary and causal solution follows.

III. Off-line identification

The literature on the identification estimation of GARCH models is almost exclusively devoted to *off-line* quasi-maximum likelihood (ML) methods. The weakest condition under which consistency and asymptotic normality of the estimator hold are formulated by Berkes et al. [3]:

Condition 3.1: The system noise process (ε_n), an i.i.d. process with zero mean and unit variance, there exists some $\delta > 0$ such that

$$\mathrm{E}\left(|\varepsilon_n^2|^{1+\delta}\right) < \infty,$$

and the density of ε_n around zero is of the form

$$f_\varepsilon(t) = O(|t|^{-c}) \quad \text{with } 0 < c < 1.$$

For the identification of the parameters of GARCH processes we proceed similarly to ARMA processes. Write $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)^T$ and let $K \subset \mathbb{R}^{r+s+1}$ denote the set of θ-s such that the stability condition (6) holds and the polynomials C and D, defined by these θ-s, are stable. Let $K_0 \subset \mathrm{int}\, K$ be a compact domain such that $\theta^* \in \mathrm{int} K_0$. For a fixed tentative value of the system parameters, say θ, we *invert* the system

$$\bar{\sigma}_n^2(\theta) - \gamma = \sum_{i=1}^{r} \alpha_i (y_{n-i}^2 - \gamma) + \sum_{j=1}^{s} \beta_j (\bar{\sigma}_{n-j}^2(\theta) - \gamma) \tag{7}$$

to get the *frozen parameter* process $\bar{\sigma}_n^2(\theta)$, using the initial values

$$y_n = 0 \quad \text{and} \quad \bar{\sigma}_n^2(\theta) - \gamma = 0 \quad \text{for all } n \leq 0.$$

Then we compute the estimated driving noise $\bar{\varepsilon}_n(\theta)$ by the inverse equation

$$\bar{\varepsilon}_n(\theta) = \frac{y_n}{\bar{\sigma}_n(\theta)} \tag{8}$$

with $n \geq 0$. Assuming that ε_n is standard normal, the (conditional) log-likelihood function, modulo a constant, can be expressed via $\bar{\sigma}_n^2(\theta)$ as

$$L_N = \sum_{n=1}^{N} l_n = \sum_{n=1}^{N} -\frac{1}{2}\left(\log \bar{\sigma}_n^2(\theta) + \frac{y_n^2}{\bar{\sigma}_n^2(\theta)}\right). \tag{9}$$

The right hand side depends on θ^* via (y_n^2). To stress dependence of L_N on both θ and θ^*, we shall write $L_N = L_N(\theta, \theta^*)$. The same cost function is used also in the non-gaussian case. Then the general abstract estimation problem, assuming for a moment *stationary initialization*, is to find the solution of the non-linear algebraic equation

$$\mathrm{E}_{\theta^*} \frac{\partial}{\partial \theta} l_0(\theta, \theta^*) = 0.$$

The conditional quasi-maximum likelihood estimation $\hat{\theta}_N$ of θ^* is defined as the solution of the equation

$$\frac{\partial}{\partial \theta} L_N(\theta, \theta^*) = L_{\theta N}(\theta, \theta^*) = 0. \tag{10}$$

The differentiation here is taken in the almost sure sense. More exactly: $\hat{\theta}_N$ is a random vector such that $\hat{\theta}_N \in K$ for all ω, and if the equation (10) has a unique solution in K, then $\hat{\theta}_N$ is equal to this solution. By the measurable selection theorem such a random variable exists.

Define the asymptotic cost function, (a negative log-likelihood for the gaussian case) as

$$W(\theta, \theta^*) = \lim_{n \to \infty} \mathrm{E}\left[\frac{1}{2}\left(\log \bar{\sigma}_n^2(\theta) + \frac{y_n^2}{\bar{\sigma}_n^2(\theta)}\right)\right]. \tag{11}$$

Then the the asymptotic estimation problem is

$$\frac{\partial}{\partial \theta} W(\theta, \theta^*) = \lim_{n \to \infty} \mathrm{E} \frac{\bar{\sigma}_{\theta,n}(\theta)}{\bar{\sigma}_n^3(\theta)} \left(1 - \frac{y_n^2}{\bar{\sigma}_n^2(\theta)}\right) = 0. \tag{12}$$

The existence of the above limits can be easily proven, see e.g. [3]. Also note that, by Lemma 5.5 of [3], θ^* is the a unique solution of the asymptotic problem in K_0.

The on-line algorithm: For the solution of the above estimation problem the following stochastic approximation procedure is proposed: starting with some initial condition $\theta_0 \in K_0$ we define recursively

$$\theta_n = \theta_{n-1} - \frac{1}{n} \frac{\sigma_{\theta,n}}{\sigma_n} \left(1 - \frac{y_n^2}{\sigma_n^2}\right), \tag{13}$$

where σ_n and $\sigma_{\theta,n}$ denote the on-line estimates of $\bar{\sigma}_n(\theta_{n-1})$ and $\bar{\sigma}_{\theta,n}(\theta_{n-1})$, respectively. Thus, at time n, the volatility process σ is generated via the feedback equation

$$[D_{n-1}(z^{-1})(\sigma^2 - \gamma_{n-1})]_n = [C_{n-1}(z^{-1})(y^2 - \gamma_{n-1})]_n$$

with $D_{n-1} = D(z^{-1}, \theta_{n-1})$, and similarly for C_{n-1}.

The convergence properties of the above stochastic gradient methods can be improved, and the analysis can be simplified by using a stochastic Newton method. Note, that we have

$$R^* = \frac{\partial^2}{\partial \theta^2} W(\theta, \theta^*)|_{\theta=\theta^*} = \lim_{n \to \infty} 2\, \mathrm{E} \frac{\bar{\sigma}_{\theta,n}(\theta)\bar{\sigma}_{\theta,n}(\theta)^T}{\bar{\sigma}_n^2(\theta)}|_{\theta=\theta^*}.$$

Thus the *stochastic Newton* method would read:

$$\theta_n = \theta_{n-1} - \frac{1}{n} R_{n-1}^{-1} \frac{\sigma_{\theta,n}}{\sigma_n}\left(1 - \frac{y_n^2}{\sigma_n^2}\right), \tag{14}$$

$$R_n = R_{n-1} + \frac{1}{n}\left(2 \frac{\sigma_{\theta,n}\sigma_{\theta,n}^T}{\sigma_n^2} - R_{n-1}\right). \tag{15}$$

The analysis of algorithm (13) to be outlined in this paper, is based on the BMP-theory and its modification by a resetting mechanism, given in [14]. These will be summarized in the next section. The analysis is equally applicable to (14)-(15).

IV. The BMP scheme

Following [2], we formulate the following general problem. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(X_n(\theta))$, with $\theta \in D \subset \mathbb{R}^d$, be an $\mathbb{R}^k$-valued Markov-chain over $(\Omega, \mathcal{F}, P)$ with transition kernel $\Pi_\theta(x, A)$, having a unique invariant measure μ_θ. The initial state $X_0(\theta)$ is assumed to have distribution μ_θ. Let H be a mapping from $D \times \mathbb{R}^k$ to $\mathbb{R}^d$. Then the basic estimation problem of the BMP-theory is to solve the equation

$$E_{\mu_\theta} H(\theta, X_n(\theta)) = 0,$$

using observed values of $H(\theta, X_n(\theta))$, or their computable approximations. We assume that a unique solution $\theta^* \in D$ exists. For the solution of the above problem the following stochastic approximation procedure is proposed: starting with some initial condition $\theta_0 = \xi$ define recursively

$$\theta_n = \theta_{n-1} + \frac{1}{n} H(\theta_{n-1}, X_n), \tag{16}$$

$$X_0 = x_0,$$

where $x_0 \in \mathbb{R}^k$ is a possibly random initial state, and X_n is a non-homogeneous Markov chain defined by

$$P(X_{n+1} \in A | \mathcal{F}_n) = \Pi_{\theta_n}(X_n, A).$$

Here $\mathcal{F}_n$ is the σ-field of events generated by the random variables $X_0, \ldots, X_n$, and A is any Borel subset of $\mathbb{R}^k$.

Convergence of stochastic approximation procedures given above in (16) has been studied under various sets of assumptions, see Benveniste et al. [2], Delyon [9]. A similar SA method with mixing rather than Markovian state dynamics has been studied in Gerencsér [12], [13].

A key tool in the analysis of the above method is a so-called ODE method, see below for details, in which the partial sums of the correction terms in (16) are approximated by the solution of an ordinary differential equation. These partial sums are first approximated by a suitable martingale using a standard device in the in the theory of a Markov chains, namely the Poisson equation. This is defined by

$$(I - \Pi_\theta)u = g$$

to be solved for u for a possibly large class of functions g, satisfying $\mathrm{E}_{\mu_\theta} g = 0$. A major observation of the BMP theory is that existence an uniqueness of the solution for a relatively small class of functions g, to be denoted by $Li(q)$, implies existence an uniqueness for a much larger class. A standard tool in this analysis is to prove that $(X_n(\theta))$ is geometrically ergodic for an appropriate class of functions.

V. Basic assumptions

In BMP theory we need a number of verifiable technical conditions, ensuring that $(X_n(\theta))$ is geometrically ergodic for an appropriate class of functions. We will formally state only one of these conditions which is critical when BMP theory is applied to the analysis of recursive estimation of GARCH processes. The condition below expresses one-step boundedness and r-step contraction of the Markovian dynamics.

Condition 5.1: Assume that for all $q \geq 1$ and for any compact subset $Q \subset D$ there exists a constant K such that for any $\theta \in Q$ and all $x \in \mathbb{R}^k$:

$$\int \Pi_\theta(x, dy)(1 + |y|^{q+1}) \leq K(1 + |x|^{q+1}).$$

Assume furthermore, that for all $q \geq 1$ and for any compact subset $Q \subset D$ there exist $r \in \mathbb{N}, 0 < \alpha < 1$ and $\beta \in \mathbb{R}$ such that for any $\theta \in Q$ and all $x \in \mathbb{R}^k$:

$$\int \Pi_\theta^r(x, dy)|y|^{q+1} \leq \alpha|x|^{q+1} + \beta.$$

To ensure geometric ergodicity we need also a condition implying that the Markov chain forgets its initial condition exponentially fast,. i.e. we need an upper bound on

$$|\Pi_\theta^n g(x) - \Pi_\theta^n g(x')| \tag{17}$$

for a class of functions g specified below. To ensure some kind of regularity of the solution $\mu_\theta(x)$ with respect to θ we need to impose regularity conditions both on Π_θ^n and on $H(\theta, x)$. In particular, we assume that the kernels Π_θ^n are Lipschitz-continuous, uniformly in n, with respect to θ, implying convenient upper bounds for

$$|\Pi_\theta^n g(x) - \Pi_{\theta'}^n g(x)| \tag{18}$$

for a class of functions g specified below.

The class of functions for which the standard inequalities of geometric ergodicity has to be established is the so-called class $Li(q)$, which defined as follows: for a measurable real-valued function g on $\mathbb{R}^k$ and any $q \geq 0$ define the norms

$$||g||_q := \text{ess}\sup_x \frac{|g(x)|}{1 + |x|^q},$$

and also

$$||\Delta g||_q = \sup_{x_1 \neq x_2} \frac{|g(x_1) - g(x_2)|}{|x_1 - x_2|(1 + |x_1|^q + |x_2|^q)}.$$

Then the class of test functions $Li(q)$ is defined as

$$Li(q) = \{\ g\ :\ ||\Delta g||_q < +\infty\}.$$

The updating function $H(\theta, x)$ will satisfy the following:

Condition 5.2: There exists $q \geq p \geq 1$ such that for any compact subset $Q \subset D$ there exists a constant K depending only on Q, such that for all $\theta, \theta' \in Q$ and any $x \in \mathbb{R}^k$:

$$||\Delta H(\theta, \cdot)||_p \leq K$$
$$|H(\theta, x)| \leq K(1 + |x|^{p+1})$$
$$|H(\theta, x) - H(\theta', x)| \leq K|\theta - \theta'|(1 + |x|^{p+1}).$$

VI. The resetting mechanism

Convergence of the estimator sequence (θ_n) with probability *strictly less than 1* has been proved in [2], in Theorem 13, p. 236 of Part II. A critical issue in the analysis of the BMP scheme is that θ_n may eventually leave its domain of definition. This is sometimes referred to as the boundedness problem. A simple remedy for this is to restrict the process to D, or to a compact truncation domain $D_0 \subset D$ containing θ^* in its interior, by stopping the process if θ_n would leave D_0. This situation is getting even worse, when an additional criterion for stopping is introduced: namely, when the process will be stopped, if the difference between two successive estimators exceeds a fixed threshold. A set of technical conditions and rigorous analysis of the effect of resettings has been first given by Gerencsér [12].

A BMP-scheme modified by an appropriate time-varying resetting mechanism has been proposed and analyzed by Delyon [9]. Gerencsér and Mátyás [14] use a fixed, fairly arbitrary truncation domain and a fixed threshold to prevent the estimates from making large jumps.

In the following we present the modified stochastic approximation algorithm introduced by Gerencsér and Mátyás in [14], using a suitable resetting mechanism, which is shown to converge with *probability 1* to θ^*, under reasonable technical conditions.

Since the updating function H is defined on $D \times \mathbb{R}^k$, the SA algorithm (16) makes sense only if $\theta_{n-1} \in D$. Hence we will require that the estimator lie in a compact truncation domain $D_0 \subseteq D$, with properties specified below. In addition we would like to prevent the estimates from making large jumps. Therefore we have to modify our algorithm to enforce these boundedness conditions. Let us choose a fixed small number $\varepsilon > 0$. Let $\tau_0 \equiv 0$ and for $i \geq 1$ define recursively the stopping times

$$\tau_i := \tau_i^e \wedge \tau_i^j,$$

where

$$\tau_i^e = \inf\{k > \tau_{i-1} : \theta_k \notin D_0\}$$
$$\tau_i^j = \inf\{k > \tau_{i-1} : |\theta_k - \theta_{k-1}| > \varepsilon\}.$$

Here $x \wedge y = \min(x, y)$. τ_i is thus the first time after τ_{i-1} at which the algorithm either leaves D_0 or a jump of magnitude at least ε occurs. At τ_i we re-initialize both the parameter value and the state vector of the algorithm: we set

$$\theta_{\tau_i} := \xi_0 \quad \text{and} \quad X_{\tau_i} := x_0.$$

To formalize the resetting procedure let θ_{i-} denote the value of θ computed at time i by (16) and define the set

$$B_i = \{\omega \mid \theta_{i-} \notin D_0 \text{ or } |\theta_{i-} - \theta_{i-1}| > \varepsilon\}.$$

The algorithm with resetting:

$$\theta_i = \theta_{i-1} + (1-\chi_{B_i})\frac{1}{i}H(\theta_{i-1}, X_i) + \chi_{B_i}(\xi_0 - \theta_{i-1}), \quad (19)$$

with $\theta_0 = \xi_0$, where X_i follows the dynamics

$$X_i = (1-\chi_{B_i})f(X_{i-1}, \theta_{i-1}, U_i) + \chi_{B_i}x_0, \qquad X_0 = x_0,$$

realizing the Markovian dynamics, with a Borel-measurable mapping f from $\mathbb{R}^k \times D \times [0,1]$ to $\mathbb{R}^k$, and a random variable U uniformly distributed on $[0,1]$.

VII. Convergence

The convergence of the sequence (θ_n) is analyzed using a so-called ODE-method. The ODE-method is stated in various forms in Kushner and Clark [17], Ljung [19], Benveniste et al. [2]. For SA procedures with resetting see Gerencsér [12], [13]. The ODE method indicates that the (θ_n) is closely related to the solution of the associated ODE

$$\frac{d}{ds}\bar{\theta}_s = h(\bar{\theta}_s), \quad \bar{\theta}_0 = \xi. \quad (20)$$

Let $\bar{\theta}(t, s, \xi)$ denote the general solution. To ensure convergence of (θ_n) we have to require asymptotic stability of the associated ODE. To ensure that the parameter sequence (θ_n) is not bounced back and forth by resetting we need to impose some conditions on the shape of the truncation domain D_0 and on the position of the initial value ξ relative to D_0. Condition 7.1 below is taken from [13].

Condition 7.1: Let $D_0 \subset D$ be a compact truncation domain. Assume that for any $\xi \in D_0$, $\bar{\theta}(t, 0, \xi) \in D$ is defined for any $t \geq 0$, and for some $\theta^* \in \operatorname{int} D_0$ we have

$$\lim_{t\to\infty} \bar{\theta}(t, 0, \xi) = \theta^*$$

for any initial value $\xi \in D_0$. Assume furthermore, that we have an initial estimate ξ_0 such that for all $t \geq 0$ we have

$$\bar{\theta}(t, 0, \xi_0) \in \mathrm{int D_0}.$$

The theorem below is the main result of [14]:

Theorem 7.1: Consider algorithm (19) and assume that Condition 5.1, 5.2 and 7.1 hold, and certain technical conditions expressed via (17) and (18) hold. Let ε, the limiting rate of θ_n, be sufficiently small. Then

$$\lim_n \theta_n = \theta^* \quad \text{w.p.1},$$

and also in L_q, for any $q \geq 1$.

Remark 7.1: Conditions 5.1 can be *relaxed*: instead of assuming it for all $q \geq 1$, it is sufficient to require that it holds for *some* q such that

$$q > 2(p+1),$$

where p is the exponent, characterizing the growth rate of H, see Condition 5.2. Then L_q-convergence holds with this particular q.

VIII. Application to GARCH processes

In this section we outline the proof of convergence for the recursive algorithm for GARCH processes, (13), modified with resetting. Recall that the inverse system defining the mapping from y to $\bar{\varepsilon}$ is defined by, following (7) and (8),

$$\bar{\sigma}_n^2(\theta) - \gamma = \sum_{i=1}^{r} \alpha_i (y_{n-i}^2 - \gamma) + \sum_{j=1}^{s} \beta_j (\bar{\sigma}_{n-j}^2(\theta) - \gamma) \quad (21)$$

with initial values $y_n = 0$ and $\bar{\sigma}_n^2(\theta) - \gamma = 0$ for all $n \leq 0$, or briefly by

$$D(z^{-1})(\bar{\sigma}^2 - \gamma) = C(z^{-1})(y^2 - \gamma) \quad (22)$$

and by

$$\bar{\varepsilon}_n(\theta) = \frac{y_n}{\bar{\sigma}_n(\theta)} \quad \text{for} \quad n \geq 0. \quad (23)$$

Extending the state vector X_n^* by

$$\bar{Z}_n(\theta) = (\bar{\sigma}_n^2(\theta), \ldots, \bar{\sigma}_{n-s+1}^2(\theta))^T,$$

it is easy to see that $\bar{Z}_n(\theta)$ follows the dynamics

$$\bar{Z}_{n+1}(\theta) = B(\theta)\begin{pmatrix} Y_n \\ \bar{Z}_n(\theta) \end{pmatrix} + v(\theta), \quad n \geq 0, \quad (24)$$

with

$$B(\theta) = \begin{pmatrix} \eta & \xi \\ 0 & \bar{S} \end{pmatrix} \quad \text{and} \quad v(\theta) = \begin{pmatrix} \alpha_0 \\ 0 \end{pmatrix}, \quad (25)$$

where

$$\eta = (\alpha_1, \ldots, \alpha_r), \quad \xi = (\beta_1, \ldots, \beta_s),$$

and

$$Y_n = (y_n^2, \ldots, y_{n-r+1}^2)^T,$$

Thus the extended state-vector

$$\bar{X}_n^e(\theta) = \begin{pmatrix} X_n^* \\ \bar{Z}_n(\theta) \end{pmatrix}$$

will follow a linear dynamics with a state-transition matrix of the following triangular structure:

$$A_n^e(\theta) = \begin{pmatrix} A_n^* & 0 \\ x & \xi \\ x & \bar{S} \end{pmatrix}$$

and with input

$$w_n^e(\theta) = \begin{pmatrix} u_n^* \\ v(\theta) \end{pmatrix}.$$

Note that the $(2,2)$ block of $A_n^e(\theta)$, i.e.

$$\begin{pmatrix} x & \xi \\ x & \bar{S} \end{pmatrix}$$

is stable due to the assumed stability of $D(z^{-1})$.

It is easy to see that $(\bar{X}_n^e(\theta))$ is a (parameter-dependent) Markov process. The state vector $\bar{X}_n^e(\theta)$ will be further extended by its derivative with respect to θ, denoted by $\bar{X}_{\theta,n+1}^e(\theta)$. Differentiating the state-equation for $\bar{X}_n^e(\theta)$ with respect to θ_i, and then collecting these equation for all i-s, we get that $\bar{X}^e_{\theta,n}(\theta)$ follows a linear dynamics with state transition matrix

$$\tilde{A}^e_n(\theta) = \mathrm{diag}(A^e_n(\theta), \ldots, A^e_n(\theta)).$$

Defining the extended state-vector $\bar{\psi}_n(\theta)$ as

$$\bar{\psi}_n(\theta) = \begin{pmatrix} \bar{X}^e_n(\theta) \\ \bar{X}^e_{\theta,n}(\theta) \end{pmatrix}$$

we get that $\bar{\psi}_n(\theta)$ follows a linear dynamics with a *block-triangular* state transition matrix

$$P_n(\theta) = \begin{pmatrix} A^e_n(\theta) & 0 \\ x & \tilde{A}^e_n(\theta) \end{pmatrix}.$$

It is obvious that $\bar{\psi}_n(\theta)$ is (a parameter dependent) Markov process. Since the asymptotic estimation problem (12) can be formulated in terms of $\bar{\psi}_n(\theta)$, the BMP theory is applicable, and we end up with the SA procedure (13). To prove the convergence of (13), suitably modified with resetting, we need to verify the conditions given in [2], or [14]. In this paper we will focus on a single condition, formulated exactly above, namely Condition 5.1.

To simplify the notations we drop the dependence on θ. The main step in the verification of Condition 5.1 is to prove that the so-called q-th mean Lyapufnov exponent associated with the i.i.d. sequence of matrices $\mathcal{P} = (P_n)$, and defined as

$$\lambda_q(\mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\|P_n \ldots P_1\|^q,$$

is strictly negative. A sufficient condition for the negativity of $\lambda_q(\mathcal{P})$ is formulated by Feigin and Tweedie [11]: namely

$$\rho[\mathrm{E}(P_0)^{\otimes q}] < 1, \tag{26}$$

implies that $\lambda_q(\mathcal{P}) < 0$. Here $\rho(\cdot)$ denotes the spectral radius, and $M^{\otimes q}$ denotes the q-th Kronecker power of the matrix M. See also [15] for a slightly simplified proof. The verification of (26) can be significantly simplified by exploiting the *block-triangular* structure of P_n. Namely, we have the following theorem, see [15]:

Theorem 8.1: Let A be a random matrix having a block-triangular structure

$$A = \begin{pmatrix} A_1 & 0 \\ B & A_2 \end{pmatrix}$$

with A_1 and A_2 being square matrices. Assume that $A \in L_q(\Omega, \mathcal{F}, P)$ for some integer $q \geq 2$, even or odd. Then

$$\rho[\ \mathrm{E}(A^{\otimes q})\ ]\ = \max\{\rho\ [\ \mathrm{E}(A_1^{\otimes q})\ ]\ ; \rho\ [\ \mathrm{E}(A_2^{\otimes q}\ )\ ]\}. \tag{27}$$

Corollary 8.1: Let q be an integer, even or odd, and let $D(z^{-1})$ be stable. Then

$$\rho[\mathrm{E}(A_0^*)^{\otimes q}] < 1 \quad \text{implies} \quad \rho[\mathrm{E}(P_0)^{\otimes q}] < 1.$$

If q is even, then it follows that

$$\lambda_q = \lambda_q(\mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\|P_n \ldots P_1\|^q < 0.$$

It follows that for any $\varepsilon > 0$ we have

$$\mathrm{E}\|P_n \ldots P_1\|^q \leq C e^{(\lambda_q+\varepsilon)n}$$

with some $C = C(\varepsilon) > 0$.

We are now ready to state our main theorem:

Theorem 8.2: Let $D^*(z^{-1})$ be stable, and let ε, the limiting rate of θ_n, be sufficiently small. Let the truncation domain D_0 be such that for all $\theta \in D_0$ the corresponding polynomial $D(z^{-1})$ is stable, and let Condition 7.1 be satisfied. Assume further that with some even $q \geq 6$

$$\rho[\mathrm{E}(A_0^*)^{\otimes q}] < 1 \tag{28}$$

is satisfied. Then the estimator sequence θ_n given by (13), and modified by a resetting mechanism converges to θ^* w.p.1, and also in L_q, with rate

$$\mathrm{E}^{1/q}|\theta_n - \theta^*|^q = O(n^{-(\alpha \wedge \frac{1}{2})}),$$

where $-\alpha < 0$ is the Lyapunov-exponent of the associated ODE.

Remark 8.1: The Lyapunov-exponent of the associated ODE is given by the maximum of the real-parts of the eigenvalues of the Jacobian-matrix of the right hand side at $\theta = \theta^*$. An analogous theorem is valid for the stochastic Newton method. In this case $\alpha = -1$, and the convergence rate in L_q is $O(n^{-\frac{1}{2}})$.

Remark 8.2: The verification of (28) is not easy. A rough upper bound for $\rho[\mathrm{E}(A_0^*)^{\otimes q}]$ is given by the following inequality: assuming

$$\sum_{i=1}^r \alpha_i^* + \sum_{j=1}^s \beta_j^* < 1,$$

and setting $A := \mathrm{E}(A_0)$, we have, for any even q,

$$\rho[\mathrm{E}(A_0)^{\otimes q}] \leq \rho[\mathrm{E}(\bar{\varepsilon}_0^2 \cdot I)^{\otimes q} \cdot A^{\otimes q}] \leq \mathrm{E}(\bar{\varepsilon}_0^{2q})\rho(A)^q. \tag{29}$$

Here we used the special structure of the A_n, and the properties of the Kronecker product. Thus $\rho[\mathrm{E}(A_0)^{\otimes q}] < 1$ is satisfied if

$$\rho(A) < \frac{1}{\mathrm{E}^{1/q}(\bar{\varepsilon}_0^{2q})}. \tag{30}$$

Remark 8.3: Note that, by the Perron-Frobenius theorem, the stability condition implies that $\rho(A) < 1$. Thus condition (30) can be interpreted as requiring that A has a large stability margin. Since $\mathrm{E}^{1/q}(\bar{\varepsilon}_0^{2q}) > 1$, this restricts the range of applicability of Theorem 8.2. The extent of this restriction will be discussed below. However, experimental results show that the recursive GARCH algorithm given by (13), and modified by a resetting mechanism works excellently even for as *small stability margins* as 2%. For the verification of (30) a good upper bound for $\rho(A)$ is given in Theorem 1 of Stefanescu [26].

Finally the reason for $q \geq 6$: the correction term in our SA algorithm (13) is defined via the function

$$H = H(\theta; y^2, \sigma^2, y_\theta^2, \sigma_\theta^2) = \frac{\sigma_\theta}{\sigma}\left[1 - \frac{y^2}{\sigma^2}\right].$$

Thus it is straightforward to see that Condition 5.2 holds with $p = 1$. The condition that $q > 2(p+1)$, and q even brings us to the condition $q \geq 6$.

IX. Simulation results

Numerical studies indicate that the accuracy of the estimator depends strongly on the stability margin

$$1 - \alpha_1^* - \beta_1^*.$$

Figure 1 shows the trajectory of θ_n for simulated observations generated by a GARCH(1,1) model with parameters

$$\theta^* = \begin{pmatrix} 0.6 \\ 0.2 \\ 0.7 \end{pmatrix},$$

driven by gaussian noise. The stability margin is now 0.1. The convergence is excellent in this case.

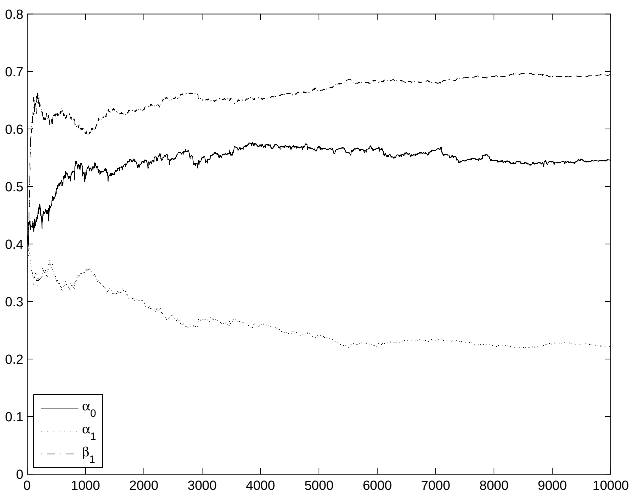

Fig. 1. Convergence of $\hat{\theta}$.

The inverse of the estimated Fisher information of the above model is

$$\hat{I}^{-1} = \begin{pmatrix} 29.5458 & 2.8821 & -7.8397 \\ 2.8821 & 1.4024 & -1.6001 \\ -7.8397 & -1.6001 & 2.8507 \end{pmatrix}$$

with eigenvalues

$$\lambda = \begin{pmatrix} 0.0903 \\ 1.6957 \\ 32.0127 \end{pmatrix}.$$

The relatively high value in the $(1,1)$ position indicates that the estimation of the parameter α_0^* is less accurate than the estimation of α_1^* and β_1^*. The condition number, denoted by κ is moderately high $\kappa = 353.69$.

This observation is reinforced by Figure 2. Here the diagonals of the inverse of the estimated Fisher information matrix and the condition number are plotted against the stability margin. The accuracy of the estimate of α_0 decreases, while the accuracy of α_1, β_1 increases with decreasing stability margins. The condition number also increases with decreasing stability margins. The estimation problem becomes badly ill-conditioned for stability margins less that 2%, i.e. when $\alpha_1^* + \beta_1^* > 0.98$.

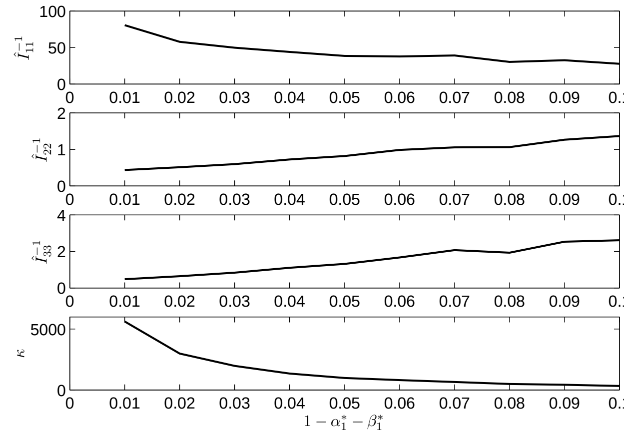

Fig. 2. The diagonals of the inverse of the Fisher information matrix as functions of the stability margin.

References

[1] A. Aknouche and H. Guerbyenne (2006). Recursive estimation of GARCH models, Communication in Statistics-Simulation and Computation 35: 925-938.

[2] A. Benveniste, M. Métivier and P. Priouret (1990). *Adaptive Algorithms and Stochastic Approximations*, Springer Verlag, Berlin.

[3] I. Berkes, L. Horváth and P.S. Kokoszka (2003). GARCH processes: structure and estimation, *Bernoulli* 9: 201-207.

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

[5] T. Bollerslev, R.Y. Chou and K.F. Kroner (1992). ARCH models in finance: A review of the theory and evidence, *Journal of Econometrics* 52:5-59.

[6] P. Bougerol, N. Picard (1992). Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics* 52: 115-127

[7] P. Bougerol, N. Picard (1992). Strict stationarity of generalized autoregressive processes, *The Annals of Probability* 20(4): 1714-1730.

[8] R. Dahlhaus and S. Subba Rao (2007). A recursive online algorithm for the estimation of time-varying ARCH parameters, *Bernoulli* 13(2): 389-422.

[9] B. Delyon (2000). Stochastic approximation with decreasing gain: Convergence and asymptotic theory, technical report, Université de Rennes.

[10] R.F. Engle (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation, *Econometrica* 50:987-1008

[11] P.D. Feigin, R.L. Tweedie (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments, *Journal of Time Series Analysis* 6(1): 1-14.

[12] L. Gerencsér (1992). Rate of convergence of recursive estimators, *SIAM Journal Control and Optimization* 30(5): 1200–1227.

[13] L. Gerencsér (2006). A representation theorem for the error of recursive estimators, *SIAM Journal on Control and Optimization* 44(6): 2123-2188 .

[14] L. Gerencsér and Z. Mátyás (2007). Almost sure convergence of the BMP scheme with resetting, In: Proc. of the European Control Conference, ECC'07, Kos, Greece, July 2-5, 2007.

[15] L. Gerencsér, Zs. Orlovits (2008). L_q-stability of products of block-triangular stationary random matrices, *Acta Scientiarum Mathematicarum (Szeged)* 74: 927-944.

[16] J.L. Kierkegaard, J.N. Nielsen, L. Jensen and H. Madsen (2000). Estimating GARCH models using recursive methods, http://citeseer.comp.nus.edu.sg/261917.html

[17] H.J. Kushner and D.S. Clark (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Optimization*, Springer-Verlag, Berlin, New York.

[18] S. Lee and B. Hansen (1994). Asymptotic theory for the GARCH$(1,1)$ quasi-maximum likelihood estimator, *Econometric Theory* 10:29-52.

[19] L. Ljung (1977). Analysis of recursive stochastic algorithms, *IEEE Trans. on Automatic Control* 22:551-575

[20] L. Ljung and T. Söderström (1987). *Theory and practice of recursive identification*, MIT Press.

[21] R. Lumsdaine (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH$(1, 1)$ and covariance stationary GARCH$(1, 1)$ models, *Econometric Theory* 64:575-596.

[22] T. Mikosch and C. Starica (2004). Changes of structure in financial time series and the GARCH model, *Revstat Statistical Journal* 2(1):41–73.

[23] T. Mikosch and D. Straumann (2006). Quasi-maximum likelihood estimation in heteroskedastic time series: a stochastic recurrence equation approach, *Annals of Statistics* 34(5):2449–2495.

[24] A. Milhøj (1985). The moment structure of ARCH processes, *Scandinavian Journal of Statistics* 12:281-292.

[25] N. Shephard (1996). Statistical aspects of ARCH and stochastic volatility, In: Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (Eds.) *Likelihood, Time Series with Econometric and other Applications.* Chapman and Hall, London.

[26] D. Stefanescu (2005). New bounds for positive roots of polynomials, *J. Univ. Comp. Sc.* 11:2125–2131.

[27] A.A. Weiss (1986). Asymptotic Theory for ARCH Models: Estimation and Testing, *Econometric Theory* 2:107-131.