Parametric inference and forecasting in continuously invertible volatility models

Olivier Wintenberger, Sixiang Cai

Parametric inference and forecasting in continuously invertible volatility models

Olivier Wintenberger
Centre De Recherche en Mathématiques de la Décision UMR CNRS 7534
Université de Paris-Dauphine, Place du Maréchal De Lattre De Tassigny,
75775 Paris Cedex 16, France
owintenb@ceremade.dauphine.fr

Sixiang Cai
Département de Mathématiques Université de Cergy-Pontoise
2, avenue Adolphe Chauvin
95302 Cergy-Pontoise, FRANCE
sixiang.cai@u-cergy.fr

Abstract

We introduce the notion of continuously invertible volatility models that relies on some Lyapunov condition and some regularity condition. We show that it is almost equivalent to the ability of the volatilities forecasting using the parametric inference approach based on the SRE given in [16]. Under very weak assumptions, we prove the strong consistency and the asymptotic normality of the parametric inference. Based on this parametric estimation, a natural strongly consistent forecast of the volatility is given. We apply successfully this approach to recover known results on univariate and multivariate GARCH type models and to the EGARCH(1,1) model. We prove the strong consistency of the forecasting as soon as the model is invertible and the asymptotic normality of the parametric inference as soon as the limiting variance exists. Finally, we give some encouraging empirical results of our approach on simulations and real data.

Keywords Invertibility, volatility models, parametric estimation, strong consistency, asymptotic normality, asymmetric GARCH, exponential GARCH, stochastic recurrence equation, stationarity.
Mathematical Subject Classification (2000) Primary 62F12; secondary 60H25, 62F10, 62M20, 62M10, 91B84.

1 Introduction

The paper focus on the volatility forecasting in volatility models of the general form $X_t = \Sigma_t^{1/2} \cdot Z_t$ where the innovations Z_t are normalized, centered iid variables. It is assumed that transformations of the volatilities satisfy some (possibly non-linear) Stochastic Recurrent Equation (SRE) of the form $(h(\Sigma_k))_{k\leq t} = \psi_t((h(\Sigma_k))_{k\leq t-1}, \theta_0)$. All the randomness of the process (X_t) comes from the innovations Z_t. It is assume that the expression of ψ_t with respect to $(Z_k)_{k\leq t-1}$ is known, but the parameter of interest θ_0 is unknown. In this parametric framework, including all the classical models of GARCH and EGARCH types, forecasting the volatility completely relies on the ability of inferring θ_0. The present paper give a general procedure based on SREs and its asymptotic properties that compute an approximation $\hat{\theta}_n$ as a measurable function of the observations $(X_t)_{1\leq t\leq n}$.

A SRE is said to be convergent if it exists a unique stationary non anticipative solution. The functional process (ψ_t) is generated by the filtration of the innovations $\sigma((Z_k)_{k\leq t-1})$. It is a well defined stationary ergodic process. The convergence of a SRE such as the one associated to the ψ_t leads to the existence of the stationary process (X_t). If the functions ψ_t are Lipschitz, sufficient Lyapunov conditions (also necessary in the linear case) are stated in [5] and [4]. However, the volatility process does not satisfy a unique SRE, see Remark 2.1. Optimal stationary conditions might use another SRE (think of the GARCH(p,q) model case). In the sequel we assume that the optimal SRE generated by the innovations is known and convergent. Then the process (X_t) is well defined, stationary and ergodic. Our first result in Proposition 2.1 asserts that, as any solutions of convergent SRE, (X_t) has also log-moments.

Assuming that the model to be continuously invertible, we study the estimator $\hat{\theta}_n$ from the procedure based on the SRE of [16]. Using only log-moments properties ensured by Proposition 2.1 and the locally uniform convergence of regular functions, we prove that $\hat{\theta}_n$ is strongly consistent as soon as the model is identifiable. We then give a natural strongly consistent forecaster of the volatility. The estimator $\hat{\theta}_n$ is moreover asymptotically normal as soon as the limiting variance exists. Modulo these weak and technical assumptions, we have the "equivalences"

Convergence of the SRE generated by the innovations $(Z_t, Z_{t-1}, \ldots)$	$\Longleftrightarrow$ (A)	stationarity, ergodicity and log-moments
Convergence of the SRE generated by the observations $(X_t, X_{t-1}, \ldots)$	$\Longleftrightarrow$ (B)	invertibility, forecasting and statistical inference

Let us detail the consequences of this tabular in the GARCH and EGARCH type models.
The equivalence (A) has been extensively studied in GARCH type models since the seminal papers of Nelson [12] and Bougerol and Picard [5] in the univariate case. For multivariate GARCH models the equivalence (A) has recently been studied by Francq and Zakoïan [8]. That the SRE generated by the observations converges for GARCH type models is straightforward, see [17] for discussions in the AGARCH case. Thus the equivalence (B) directly gives asymptotic properties of the statistical inference based on the SRE for these models. As the inference based on the SRE coincides with the QMLE, we recover existing results in Subsection 3.5: for GARCH(p,q) models we exactly recover the results of [9] that refine [2], for AGARCH(p,q) and CCC-GARCH(p,q) models we refine the results of [17] and [8].
In the EGARCH type models, the SRE generated by the innovations converges straightforwardly as it coincides with the one of an ARMA process. Thus it exists a stationary ergodic solution with finite moments, see [13]. The right implication (B) proved in Section 3 is new for that type of process as the inference based on the SRE differs from the QMLE (see Subsection 4.5 for details). It enables us to prove in Section 4 the asymptotic properties of the inference and the forecasting of the volatility for invertible EGARCH(1,1) models. For proving the asymptotic normality, we exhibit a necessary and sufficient condition for the existence of the asymptotic variance. Then we provide some encouraging empirical results of our approach in Section 5 on simulations and on real data. A first step on the reverse part of the equivalence (B) has been done recently by Sorokin [15]. He proves that forecast of the volatility based on the SRE is inconsistent for some non invertible models. Finally remark that statistical inference of θ_0 is possible without assuming the invertibility: it has been successfully done by Zaffaroni [19] using the approach of Whittle.

2 Stationarity and invertibility in volatility models

2.1 The general volatility model

All along the paper (Z_t) is a stationary ergodic sequence of real vectors called the innovations. Let us denote $\mathcal{F}_t$ the filtration generated by $(Z_t, Z_{t-1}, \ldots)$ and let us consider the general volatility model

$$X_t = \Sigma_t^{1/2} \cdot Z_t, \qquad \text{where} \qquad (h(\Sigma_k))_{k\leq t} = \psi_t((h(\Sigma_k))_{k\leq t-1}, \theta_0) \tag{1}$$

with an injective function h from the space of real matrices of size $k\times k$ to an auxiliary separable metric space F and $\psi_t(\cdot, \theta_0)$ is a $\mathcal{F}_{t-1}$ adapted random function from the space of the sequences of elements in the image of h to itself. Let us denote ℓ the inverse of h (from the image of h to the space of real matrices of size $k \times k$) and call it the link function.

2.2 Convergent SRE and stationarity

A first question regarding this very general model is wether or not a stationary solution exists. As the sequences of the transformed volatilities $(h(\Sigma_k))_{k\leq t}$ is a solution of a fixed point problem, we recall the following result due to Bougerol [4]. Let (E, d) be a complete separable metric space. A map $f : E \to E$ is a Lipschitz map if $\Lambda(f) = \sup_{(x,y)\in E^2} d(f(x), f(y))/d(x, y)$ is finite. For any sequence of random element in (E, d), (X_t) is said to be exponential almost sure convergence to 0 $X_t \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$ if for $X_t = o(e^{-Ct})$ a.s. for some $C > 0$.

Theorem 2.1. *Let (Ψ_t) be a stationary ergodic sequence of Lipschitz maps from E to E. Suppose that $\mathbb{E}[\log^+(d(\Psi_0(x), x))] < \infty$ for some $x \in E$, that $\mathbb{E}[\log^+ \Lambda(\Psi_0)] < \infty$ and that for some integer $r \geq 1$,*

$$\mathbb{E}[\log \Lambda(\Psi_0^{(r)})] = \mathbb{E}[\log \Lambda(\Psi_0 \circ \cdots \circ \Psi_{-r+1})] < 0.$$

Then the SRE $X_t = \Psi_t(X_{t-1})$ for all $t \in \mathbb{Z}$ is convergent: it admits a unique stationary solution $(Y_t)_{t\in\mathbb{Z}}$ which is ergodic and for any $y \in E$

$$Y_t = \lim_{m\to\infty} \Psi_t \circ \cdots \circ \Psi_{t-m}(y), \quad t \in \mathbb{Z}.$$

The Y_t are measurable with respect to the $\sigma(\Psi_{t-k}, k \geq 0)$ and

$$d(\tilde{Y}_t, Y_t) \xrightarrow{e.a.s.} 0, \quad t \to \infty$$

such that $\tilde{Y}_t = \Psi_t(\tilde{Y}_{t-1})$ for all $t > 0$.

Remark that the sufficient Lyapunov assumptions $\mathbb{E}[\log \Lambda(\Psi_0^{(r)})] < 0$ is necessary in the linear case, see [5]. The logarithmic moments of the solution of a convergent SRE is proved in the following result that seems to be new:

Proposition 2.1. *Under the assumptions of Theorem* (2.1) *the unique stationary solution also satisfies* $\mathbb{E}[\log^+(d(Y_0, x))] < \infty$.

Proof. First remark that $\log(1 + y + z) \leq \log(1 + y) + \log(1 + z)$ for any $y, z \geq 0$ and that $\mathbb{E}[\log(1+\Lambda(\Psi_0))] < \infty$ and $\mathbb{E}[\log(1+d(\Psi_0(x), x))] < \infty$ by assumption. Second, we remark that $\Lambda(f \circ g) \leq \Lambda(f)\Lambda(g)$ for any Lipschitz functions f and g entails that

$$\log(1 + d(x, \Psi_0 \circ \cdots \circ \Psi_{-m}(x))) \leq \log(1 + d(x, \Psi_0 \circ \cdots \circ \Psi_{1-m}(x))) + \log(1 + \Lambda(\Psi_0 \circ \cdots \circ \Psi_{1-m}) d(\Psi_{-m}(x), x)). \quad (2)$$

Third, we assert that the ergodic theorem in $\mathbb{L}^1$ leads to the existence of $\rho < 1$ such that

$$\frac{1}{n} \sum_{j=1}^{n} \Lambda(\Psi_{(j-1)r} \circ \cdots \circ \Psi_{1-jr}) \to \log(\rho) \qquad \text{in} \qquad \mathbb{L}^1.$$

Using again that $\Lambda(f \circ g) \leq \Lambda(f)\Lambda(g)$ for any Lipschitz functions f and g we infer $\overline{\lim}_{m\to\infty} m^{-1} \log(\Lambda(\Psi_0 \circ \cdots \circ \Psi_{-m})) \leq \log(\rho)$ in $\mathbb{L}^1$. Thus it exists a r. v. $M > 0$ such that $\mathbb{E}[|\log(M)|] < \infty$ and $m^{-1} \log(\Lambda(\Psi_0 \circ \cdots \circ \Psi_{-m})) - \log(\rho) \leq \log(M)$, i.e. $\Lambda(\Psi_0 \circ \cdots \circ \Psi_{-m}) \leq (M\rho)^m$ a.s. for all $m \geq 0$. Using this bound in (2) we get, writing $v_m = \log(1 + d(x, \Psi_0 \circ \cdots \circ \Psi_{-m}(x)))$:

$$v_m \leq v_{m-1} + \log(1 + (M\rho)^{m-1} d(\Psi_{-m}(x), x)).$$

As $\log(1 + yz) \leq y^\omega \log(1 + z)$ for any $0 < y < 1$, $z \geq 0$ and $0 < \omega < 1$ we get by a straightforward recurrence

$$v_m \leq v_0 + \sum_{j=0}^{m-1} \rho^{j\omega} \log(1 + M^j d(\Psi_0(x), x)).$$

As $\log(1+yz) \leq \log(1+y^2+z^2) \leq 2(\log(1+y)+\log(1+z))$ for any $y, z \geq 0$ we assert that $\mathbb{E}[\log(1+M^j d(\Psi_0(x), x))] \leq Cj$ for some $C > 0$ depending on the (finite) logarithmic moments of M and $d(\Psi_0(x), x))$. Finally, $\lim_{m\to\infty} v_m = \log(1 + d(Y_o, x))$ is dominated by an integrable r.v. and the result follows dominated integration. □

In order to apply Theorem 2.1 in our case, let us denote by E the separable metric space of the sequences of elements in the image of h. Equipped with the metric $\sum_{j\geq 1} 2^{-j} d(x_j, y_j)/(1 + d(x_j, y_j))$, the space E is complete. A sufficient condition for stationarity of (X_t) is that the SRE driven by (ψ_t) converges in E. It simply expresses as the Lyapunov condition $\mathbb{E}[\log \Lambda(\psi_0^{(r)})] < 0$ for some integer $r \geq 1$ and some logarithmic moments. This assumption of stationarity is sufficient but not optimal in many cases:

Remark 2.1. *The state space of the SRE* (1)*, denoted E, is in full generality a space of sequences. However in all classical models we can find a lag p such that $(h(\Sigma_k))_{t-p+1\leq k\leq t} = \psi_t((h(\Sigma_k))_{t-p\leq k\leq t-1}, \theta_0)$. The state space E is now the finite product of p spaces. It can be equipped by unbounded metrics such that $p^{-1}\sum_{j=1}^p d(x_j, y_j)$ or $\sqrt{\sum_{j=1}^p d^2(x_j, y_j)}$. The product metric has to be carefully chosen as it changes the value of the Lipschitz coefficients of the ϕ_t. Yet, even if the products spaces are embedded, the smallest possible lag p in the SRE yields the sharpest Lyapunov condition. Finally, if E has a finite dimension and if the condition of convergence of the SRE expresses in term of the top Lyapunov coefficient, one can choose any metric induced by any norm, see [4] for details.*

We prefer to work under the less explicit assumption

(ST) The process (X_t) satisfying (1) exists. It is a stationary, non anticipative and ergodic process with finite logarithmic moments.

In view of Proposition 2.1, it does not cost to ask for the solution to have finite logarithmic moments.

2.3 The invertibility and the observably invertibility

Now that under **(ST)** the process (X_t) is stationary and ergodic, we investigate the question of invertibility as it is done in the literature. We want to emphasis that invertibility is a question of convergent SRE solved by Lyapunov condition. Following Tong [18], we say that a volatility model is invertible if the volatility can be expressed as a function of the past observed values:

Definition 2.1. *The model is invertible if the sequence of the volatilities (Σ_t) is adapted to the filtration $(\mathcal{G}_{t-1})$ generated by $(X_{t-1}, X_{t-2}, \cdots)$.*

It is natural to assume invertibility to be able to forecast the volatility. This notion of invertibility is very weak and consists on restricting the

underlying filtration $(\mathcal{F}_t)$ of the SRE to $(\mathcal{G}_{t-1})$. Indeed, under **(ST)** then $\mathcal{G}_t \subseteq \mathcal{F}_t$ is well defined and writing $Z_t = \Sigma_t^{-1} \cdot X_t$ in ψ_t we can rewrite the general model as

$$(h(\Sigma_k))_{k\leq t} = \phi_t((h(\Sigma_k))_{k\leq t-1}, \theta_0) \tag{3}$$

for some ergodic and stationary sequence (ϕ_t) adapted to $(\mathcal{G}_{t-1})$. Thus the invertibility is implied by Theorem 2.1 if the ϕ_t are Lipschitz maps such that for some $r > 0$,

$$\mathbb{E}[\log \Lambda(\phi_0(\cdot, \theta_0)^{(r)})] < 0. \tag{4}$$

The remark 2.1 also holds for the SRE driven by (ϕ_t); the state space E of the SRE and the product metric must be chosen carefully. The condition (4) (with the optimal choice of the state space E and its metric) is called the condition of invertibility.

Proposition 2.2. *Under* **(ST)** *and the condition of invertibility* (4)*, the general model* (3) *is invertible.*

Another notion of invertibility is the one introduced in [17]. We call it observably invertibility. Let us assume that there exists some approximations $\hat{\phi}_t$ of ϕ_t such that $\hat{\phi}_t$ is a measurable function of the observations $(X_{t-1}, X_{t-2}, \ldots, X_1)$.

Definition 2.2. *The model is observably invertible iff the solution of the approximative SRE*

$$(h(\hat{\Sigma}_k))_{k\leq t} = \hat{\phi}_t((h(\hat{\Sigma}_k))_{k\leq t-1}, \theta_0) \tag{5}$$

is convergent, i.e. $\|\hat{\Sigma}_t - \Sigma_t\| \to 0$ *in probability as* $t \to \infty$.

Remark that in general the approximative SRE does not fit the conditions of Theorem 2.1, in particular $(\hat{\phi}_t)$ is not necessarily stationary and ergodic. However, the useful Property below ensures that an invertible model is also observably invertible. It is a straightforward Corollary of Theorem 2.10 of [17] and our Proposition 2.1

Proposition 2.3. *Assume that* **(ST)** *and* (4) *holds, that the link function* ℓ *is continuous and that there exists* $x \in E$ *such that* $d(\hat{\phi}_t(x), \phi_t(x)) \xrightarrow{e.a.s.} 0$ *and* $\Lambda(\hat{\phi}_t(\cdot, \theta_0) - \phi_t(\cdot, \theta_0)) \xrightarrow{e.a.s.} 0$ *as* $t \to \infty$*, then the model is observably invertible.*

Remark 2.2. *Classical models satisfy an SRE for finite p lags of volatilities $(h(\Sigma_k))_{t-p+1\le k\le t} = \phi_t((h(\Sigma_k))_{t-p\le k\le t-1}, \theta_0)$ and for some ϕ_t generated only by a finite of past observation $(X_{t-1}, \dots, X_{t-q})$. In this context, the approximative SRE coincides with the initial ones, i.e. one can choose $\hat{\phi}_t = \phi_t$ for $t > q$. Conditions of Proposition 2.3 hold systematically; invertibility and observable invertibility are equivalent, i.e. induced by the same Lyapunov condition. As the initial values (for $0 \le t \le q$) in the SRE can be chosen arbitrarily from Theorem 2.1, with some abuse we will work in the sequel with $\hat{\phi}_t = \phi_t$ for $t \ge 1$.*

2.4 The continuous invertibility

We have seen that the existing invertibility notions can be expressed in term of Lyapunov conditions. We introduce the notion of continuous invertibility in term of a Lyapunov condition and some regularity on the model. Let us consider parametric functions of the model having continuous Lipschitz coefficients:

(CL) For any metric spaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, a function $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathcal{Z}$ satisfies (**CL**) if there exists a continuous function $\Lambda_f : \mathcal{Y} \mapsto \mathbb{R}^+$ such that $\Lambda(f(\cdot, y)) \le \Lambda_f(y)$ for all $y \in \mathcal{Y}$.

From Remark 2.2, let us consider the functional SRE of the form

$$(\hat{g}_k(\theta))_{t-p+1\le k\le t} = \phi_t((\hat{g}_k(\theta))_{t-p\le k\le t-1}, \theta), \qquad \forall \theta \in \Theta, \qquad \forall t \ge 1, \quad (6)$$

with any fixed initial value $(\hat{g}_k(\theta))_{1-p\le k\le 0}$. The continuous invertibility expresses as

(CI) Assume that the SRE (6) holds with continuous functions ϕ_t satisfying **(CL)** for stationary (Λ_{ϕ_t}) satisfying the Lyapunov condition $\mathbb{E}[\log \Lambda^{(r)}_{\phi_0}(\theta)] < 0$ for all $\theta \in \Theta$ and $\mathbb{E}[\sup_\Theta \log \Lambda^{(r)}_{\phi_0}(\theta)] < \infty$. Moreover there exists an $y \in E$ such that $\mathbb{E}[\sup_\Theta \log^+(d(\phi_0(y, \theta), y))] < \infty$.

The condition **(CI)** implies the standard invertible conditions for all $\theta \in \Theta$ and in particular for the unknown θ_0. It also implies the local uniform regularity of the solution $g_t(\cdot)$ of the functional SRE $(g_k(\cdot))_{t-p+1\le k\le t} = \phi_t((g_k(\cdot))_{t-p\le k\le t-1}, \cdot)$ for all $t \in \mathbb{Z}$.

Theorem 2.2. *Assume that (Σ_t^2) is a stationary and ergodic process such that* **(CI)** *holds. Then the functions $g_t(\cdot)$ are continuous for all $\theta \in \Theta$ and*

all $t \in \mathbb{Z}$. Moreover, for any $\theta \in \Theta$ there exists an $\epsilon > 0$ such that $\hat{g}_t(\theta)$ satisfying (6) *satisfies*

$$\lim \sup_{\theta' \in \overline{B}(\theta,\epsilon)\cap\Theta} d(\hat{g}_t(\theta'), g_t(\theta')) \xrightarrow{e.a.s.} 0. \tag{7}$$

Proof. For any $\rho > 0$, let us write $\Lambda_*^{(r)}(\theta,\rho) = \sup\{\Lambda_{\phi_0}^{(r)}(\theta'), \theta' \in \overline{B}(\theta,\rho)\cap\Theta\}$, where $\overline{B}(\theta,\rho)$ stands for the closed ball centered at θ with radius ρ. As $\mathbb{E}[\sup_\Theta \log \Lambda_{\phi_0}^{(r)}(\theta)] < \infty$, by the dominated convergence Theorem we obtain $\lim_{\rho\to 0} \mathbb{E}(\Lambda_*^{(r)}(\theta,\rho)) = \mathbb{E}(\lim_{\rho\to 0} \Lambda_*^{(r)}(\theta,\rho))$. But $\lim_{\rho\to 0} \Lambda_*^{(r)}(\theta,\rho) = \Lambda_{\phi_0}^{(r)}(\theta)$ by continuity of $\Lambda_{\phi_0}^{(r)}(\theta)$, we finally obtain

$$\lim_{\rho\to 0} \mathbb{E}(\Lambda_*^{(r)}(\theta,\rho)) = \mathbb{E}(\lim \sup_{\theta'\to\theta} \Lambda_{\phi_0}^{(r)}(\theta))) = \mathbb{E}(\Lambda_{\phi_0}^{(r)}(\theta)) < 0.$$

Thus, there exists an $\epsilon > 0$ such that $\mathbb{E}(\Lambda_*^{(r)}(\theta,\epsilon)) < 0$.

Let us now work on $C(\overline{B}(\theta,\epsilon)\cap\Theta)$, the complete metric space of continuous functions from $\overline{B}(\theta,\epsilon)\cap\Theta$ to $\mathbb{R}$ equipped with the supremum norm $d_\infty = \sup_{\overline{B}(\theta,\epsilon)\cap\Theta} d$. In this setting $(\hat{g}_t)$ satisfy a functional SRE $(\hat{g}_k)_{k\le t} = \phi_t((\hat{g}_k)_{k\le t-1})$ with Lipschitz constants satisfying

$$\begin{aligned}
\Lambda_\infty(\phi_t^{(r)}(\cdot)) &\le \sup_{s_1,s_2\in C(\overline{B}(\theta,\epsilon)\cap\Theta)} \frac{d_\infty(\phi_t^{(r)}(s_1), \phi_t^{(r)}(s_2))}{d_\infty(s_1,s_2)} \\
&\le \sup_{s_1,s_2\in C(\overline{B}(\theta,\epsilon)\cap\Theta)} \frac{\sup_{\overline{B}(\theta,\epsilon)\cap\Theta} d(\phi_t^{(r)}(s_1(\theta'),\theta'), \phi_t^{(r)}(s_2(\theta'),\theta')}{d_\infty(s_1,s_2)} \\
&\le \sup_{s_1,s_2\in C(\overline{B}(\theta,\epsilon)\cap\Theta)} \frac{\sup_{\overline{B}(\theta,\epsilon)\cap\Theta} \Lambda(\phi_t^{(r)}(\cdot,\theta'))d(s_1(\theta')s_2(\theta'))}{d_\infty(s_1,s_2)} \\
&\le \sup_{s_1,s_2\in C(\overline{B}(\theta,\epsilon)\cap\Theta)} \frac{\sup_{\overline{B}(\theta,\epsilon)\cap\Theta} \Lambda(\phi_t^{(r)}(\cdot,\theta'))d_\infty(s_1,s_2)}{d_\infty(s_1,s_2)} \\
&\le \sup_{\overline{B}(\theta,\epsilon)\cap\Theta} \Lambda(\phi_t^{(r)}(\cdot,\theta')) \le \sup_{\overline{B}(\theta,\epsilon)\cap\Theta} \Lambda_{\phi_t}^{(r)}(\theta') \le \Lambda_*^{(r)}(\theta,\epsilon)).
\end{aligned}$$

As $\mathbb{E}[\sup_{\overline{B}(\theta,\epsilon)\cap\Theta} \log^+(d(\phi_t(y,\theta'),y))] \le \mathbb{E}[\sup_\Theta \log^+(d(\phi_t(y,\theta),y))]$ is finite we can apply Theorem 2.1. By recurrence $\phi_t \circ \cdots \circ \phi_{t-m}(\zeta_0) \in C(\overline{B}(\theta,\epsilon)\cap\Theta)$ is continuous in θ and so is g_t as the convergence holds uniformly on $\overline{B}(\theta,\epsilon)\cap\Theta$. It is true for any $\theta \in \Theta$ and the result follows. □

3 Statistical inference under continuous invertibility

3.1 Statistical inference based on the SRE

Here we describe the approach of Straumann in [16]. Assume that (6) holds with θ_0 unknown and θ_0 belonging in a known compact set Θ. Consider

$$\hat{\theta}_n = \text{argmin}_{\theta\in\Theta}\hat{S}_n(\theta)$$

the M-estimator associated with the SRE criteria

$$n\hat{S}_n(\theta) = \sum_{t=1}^{n}\hat{s}_t(\theta) = \sum_{t=1}^{n} 2^{-1}\left(X_t^T\ell(\hat{g}_t(\theta))^{-1}X_t + \log(\det(\ell(\hat{g}_t(\theta))))\right). \quad (8)$$

where $(\hat{g}_t)$ is obtained from the approximative SRE (6).

Remark 3.1. *This statistical procedure does not coincides with the Quasi Maximum Likelihood approach that is not always observable, see Subsection 4.5 for the EGARCH(1,1) example.*

3.2 Strong consistency of the parametric inference

From now assume that the innovations process (Z_t) is iid:

(IN) The Z_t are iid variables such that $\mathbb{E}[Z_1^T Z_1] =$ is the identity matrice.

Under **(ST)** and **(IN)**, $\mathbb{E}[X_t^T\cdot X_t \mid \mathcal{F}_0] = \Sigma_t$ in agreement with the definition of the volatility. The following assumption implies that the volatilities are invertible matrices:

(IV) The functions ℓ^{-1} and $\log(\det(\ell))$ are Lipschitz satisfying $\det(\ell(g_0(\theta))) \geq C(\theta)$ for some continuous function $C : \Theta \mapsto (0,\infty)$.

Remark 3.2. *The SRE criteria converges to the possibly degenerate limit*

$$S(\theta) = \mathbb{E}[s_0(\theta)] = 2^{-1}\mathbb{E}\left[X_0^T\ell(g_0(\theta))^{-1}X_0 + \log(\det[\ell(g_0(\theta))])\right]$$

Remark that $S(\theta_0) = 2^{-1}\mathbb{E}[Z_0^T Z_0 + \log(\det(\Sigma_0))]$ *is finite under* **(ST)** *and* **(IN)** *because the volatilities have finite log moments as solutions of an SRE, see Proposition 2.1. But* $S(\theta)$ *may be infinite for* $\theta \neq \theta_0$.

If the model is identifiable, the estimator $\hat{\theta}_n$ is strongly consistent:

Theorem 3.1. *Assume that (***ST***) and* **(CI)** *are satisfied on the compact set* Θ*. If* **(IN)** *and* **(IV)** *are satisfied and the model is identifiable, i.e.* $g_0(\theta) = h(\Sigma_0)$ *iff* $\theta = \theta_0$*, then* $\hat{\theta}_n \to \theta_0$ *a.s. for any* $\theta_0 \in \Theta$*.*

Proof. First, remark that with no loss of generality we can always restrict in (6) to $\det(\ell(\hat{\phi}_t(\cdot, \theta))) \geq C(\theta)$. We adapt the proof of Jeantheau [10] keeping the same notation $s_{*t}(\theta, \rho) = \inf\{s_t(\theta'), \theta' \in \overline{B}(\theta, \rho)\}$ and $\hat{s}_{*t}(\theta, \rho) = \inf\{\hat{s}_t(\theta'), \theta' \in \overline{B}(\theta, \rho)\}$. Let us recall Theorem 5.1 in Jeantheau [10]. The M-estimator associated with the loss (8) is strongly consistent under the hypothesis H1-H6:

H1 Θ is compact.

H2 $\hat{S}_n(\theta) \to S(\theta)$ a.s. under the stationary law P_{θ_0}.

H3 $S(\theta)$ admits a unique minimum for $\theta = \theta_0$ in Θ. Moreover for any $\theta_1 \neq \theta_0$ we have:

$$\liminf_{\theta \to \theta_1} S(\theta) > S(\theta_1).$$

H4 $\forall \theta \in \Theta$ and sufficiently small $\rho > 0$ the process $(\hat{s}_{*t}(\theta, \rho)))_t$ is ergodic.

H5 $\forall \theta \in \Theta$, $\mathbb{E}_{\theta_0}[s_{*1}(\theta, \rho)] > -\infty$.

H5 $\lim_{\rho \to 0} \mathbb{E}_{\theta_0}[s_{*1}(\theta, \rho)] = \mathbb{E}[s_{*1}(\theta)]$.

Let us check **H1-H6** in our case of study. **H1** is satisfied by assumption. **H2** is verified in two steps. First, by the e.a.s. convergence given by Theorem 2.2, arguments of [16] and the Lipschitz properties of ℓ^{-1} and $\log(\det(\ell))$ we obtain

$$\frac{1}{n} \sum_{t=1}^{n} \sup_{\overline{B}(\theta, \epsilon)} |\hat{s}_t(\theta') - s_t(\theta')| \to 0 \qquad P_{\theta_0} - a.s.$$

Second we use that (s_t) is an ergodic sequence. Using Proposition 1.1 of [10], $n^{-1} \sum_{t=1}^{n} s_t(\theta)$ converges P_{θ_0}-a.s. to $S(\theta)$ (taking values in $\mathbb{R} \cup \{+\infty\}$) as the s_t are bounded from below:

$$\frac{1}{n} \sum_{t=1}^{n} |s_t(\theta) - S(\theta)| \to 0 \qquad P_{\theta_0} - a.s.$$

Combining this two steps leads to **H2**. The first part of **H3** is checked similarly than in (ii) p.2474 of [17] and with the help of the Remark 3.2. Remark that S has a unique minimum iff $\mathbb{E}[\mathrm{Tr}(\Sigma_0 \cdot \ell(g_0(\theta))^{-1}) - \log(\det(\Sigma_0 \cdot \ell(g_0(\theta))^{-1}))]$ has a unique minimum. As this criteria is the integrand of a

sum of the $\lambda_i - \log(\lambda_i)$ where the λ_i are positive eigenvalues, we conclude under the identifiability condition from the property $x - \log(x) \geq 1$ for all $x > 0$ with equality iff $x = 1$. The second part is checked using that

$$\liminf_{\theta \to \theta_1} S(\theta) \geq \mathbb{E}[\liminf_{\theta \to \theta_1} s_0(\theta)] = \mathbb{E}[s_0(\theta_1)] = S(\theta_1)$$

where the first inequality was already used for proving Theorem 2.2 and the first equality comes from the local continuity of g_0 and ℓ. **H4** is satisfied from the ergodicity of $(\hat{s}_t)$. **H5** and **H6** follows from Theorem 2.2 that ensures the continuity of the function s_{*1} and by the lower bounded assumption on $\det(\ell)$, see Proposition 1.3 of [10]. □

3.3 Volatility forecasting

From the inference of θ_0, we deduce a natural forecast of the volatility $\hat{\Sigma}_t = \ell(\hat{g}_t(\hat{\theta}_t))$. It is strongly consistent:

Proposition 3.1. *Under the conditions of Theorem 3.1 then* $\|\hat{\Sigma}_t - \Sigma_t\| \to 0$ *a.s. as* $t \to \infty$.

Proof. It is a direct consequence of Theorems 2.2 and 3.1 that assert the a.s. convergence of $\hat{\theta}_t$ toward θ_0 and the local uniform convergence of $\hat{g}_t$ toward g_t. Remark that for t sufficiently large such that $\hat{\theta}_t \in \overline{B}(\theta, \epsilon)$, a ball where the uniform Lyapunov condition $\mathbb{E}[\log \Lambda_\infty(\phi_t(\cdot))] < 0$ is satisfied. Thus $\hat{g}_t(\hat{\theta}_t) - g_t(\theta_t) \to 0$ a.s. and by continuity of ℓ and g_t and from the identification $\Sigma_t = \ell(g_t(\theta_0))$ the result follows if $g_t(\hat{\theta}_t)$ converges to $g_t(\theta_0)$. For proving it, we use

$$d(g_t(\hat{\theta}_t), g_t(\theta_0)) \leq \Lambda_\infty(\phi_t(\cdot)) d(g_{t-1}(\hat{\theta}_t), g_{t-1}(\theta_0)) + w_t(\hat{\theta}_t)$$

where $w_t(\hat{\theta}_t) = d(\phi_t(g_{t-1}(\theta_0), \hat{\theta}_t), \phi_t(g_{t-1}(\theta_0), \theta_0))$. The RHS term satisfies an SRE of linear stationary maps satisfying the Lyapunov condition. We apply Theorem 2.1 as for any $\hat{\theta}_t$, by assumption $\mathbb{E} \log^+(w_0(\hat{\theta}_t))$ is uniformly bounded by $\mathbb{E}[\sup_\Theta \log^+(2d(\phi_t(y, \theta), y))] < \infty$. We get

$$d(g_t(\hat{\theta}_t), g_t(\theta_0)) \leq \sum_{i=0}^{\infty} \Lambda_\infty(\phi_t(\cdot)) \cdots \Lambda_\infty(\phi_{t-i+1}(\cdot)) w_{t-i}(\hat{\theta}_t).$$

Conditioning on $(\hat{\theta}_t)$, the upper bound is a stationary normally convergent

series of functions and

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{i=0}^{\infty}\Lambda_\infty(\phi_t(\cdot))\cdots\Lambda_\infty(\phi_{t-i+1}(\cdot))w_{t-i}(\hat{\theta}_t)\to 0\Big)\\
&=\mathbb{E}\Big[\mathbb{P}\Big(\sum_{i=0}^{\infty}\Lambda_\infty(\phi_t(\cdot))\cdots\Lambda_\infty(\phi_{t-i+1}(\cdot))w_{t-i}(\hat{\theta}_t)\to 0\mid(\hat{\theta}_t)\Big)\Big]\\
&=\mathbb{E}\Big[\mathbb{P}\Big(\sum_{i=0}^{\infty}\Lambda_\infty(\phi_0(\cdot))\cdots\Lambda_\infty(\phi_{-i+1}(\cdot))w_{-i}(\hat{\theta}_t)\to 0\mid(\hat{\theta}_t)\Big)\Big]\\
&=\mathbb{E}[1]=1,
\end{aligned}$$

the last inequalities following from the continuity of normally convergent series of functions, $\hat{\theta}_t \to \theta_0$ and $w_i(\hat{\theta}_t) \to w_i(\theta_0) = 0$ for all i a.s. as $t \to \infty$. □

3.4 Asymptotic normality of the parametric inference

Classical computations show easily that if the M-estmator $\hat{\theta}_n$ is asymptotically normal then the asymptotic variance is given by the expression

$$\mathbf{V} = \mathbf{P}^{-1}\mathbf{Q}\mathbf{P}^{-1}$$

with $\mathbf{P} = \mathbb{E}[\mathbb{H}s_0(\theta_0)]$ and $\mathbf{Q} = \mathbb{E}[\nabla s_0(\theta_0)\nabla s_0(\theta_0)^T]$. Let $\mathcal{V} = \overline{B}(\theta_0, \epsilon) \subset \Theta$ with $\theta_0 \in \overset{\circ}{\Theta}$ with $\epsilon > 0$ chosen in accordance with Theorem 2.2, i.e. such that $\mathbb{E}[\log(\sup_{\mathcal{V}} \Lambda_{\phi_0})] < 0$.

(AV) Assume that $\mathbb{E}(\|Z_0 Z_0^T\|^2) < \infty$ and that the functions ℓ and ϕ_t are 2-times continuously differentiable on the compact set Θ that coincides with the closure of its interior.

The next assumption is used to ensure that g_0 is 2-times differentiable.

(DL) The partial derivatives $\Phi_t = D_x(\phi_t)$, $= D_\theta(\phi_t)$, $= D^2_{x^2}(\phi_0)$, $D^2_{\theta,x}(\phi_0)$ or $D^2_{\theta^2}(\phi_0)$ satisfy **(CL)** for stationary (Λ_{Φ_t}) with $\mathbb{E}[\sup_{\mathcal{V}} \log(\Lambda_{\Phi_0})] < \infty$.

The following moments assumptions ensure the existence of $\mathbf{Q}$ and $\mathbf{P}$:

(MM) Assume that $\mathbb{E}[\|\nabla s_0(\theta_0)\|^2] < \infty$ and $\mathbb{E}[\|\mathbb{H}s_0(\theta_0)\|] < \infty$.

These moments assumptions holds only for $\theta = \theta_0$; they are simpler to verify than for $\theta \neq \theta_0$ due to the specific form of the SRE criteria derivatives, see Remark 3.2 and computations in [1]. The next assumption is classical and ensures to the existence of $\mathbf{P}^{-1}$:

(LI) The components of the vector $\nabla g_0(\theta_0)$ are linearly independent.

Finally, the last assumption is specific to the SRE approach. It ensures that $\nabla \hat{s}_t(\theta)$ is a good approximation of $\nabla s_t(\theta)$ uniformly on $\mathcal{V}$:

(LM) Assume that $y \to \nabla \ell^{-1}(y)$ and $y \to \nabla \log(\det(\ell(y)))$ are Lipschitz functions.

Theorem 3.2. *Under the assumptions of Theorem 3.1,* **(AV)**, **(DL)**, **(MM)**, **(LI)** *and* **(LM)** *then the asymptotic variance is well defined* $\mathbf{V}$ *is well defined and the statistical inference is asymptotically normal, i.e.* $\sqrt{n}(\hat{\theta}_n - \hat{\theta}_0) \to \mathcal{N}(0, \mathbf{V})$ *in distribution for any* $\theta_0 \in \overset{\circ}{\Theta}$ *with* $\mathbf{V}$ *that is invertible.*

Proof. First, remark that from **(DL)** and Proposition 2.1 applied to $\sup_{\mathcal{V}} g_0(\theta)$ then $\mathbb{E}[\sup_{\mathcal{V}} \log^+(\|\Phi_0(\theta)\|)] < \infty$ for $\Phi_0(\theta) = D_\theta(\phi_0)(g_0(\theta), \theta)$ or $D^2_{x^2}(\phi_0)(g_0(\theta), \theta)$ or $D^2_{\theta,x}(\phi_0)(g_0(\theta), \theta)$ or $D^2_{\theta^2}(\phi_0)(g_0(\theta), \theta)$. Using the existence of these logarithmic moments and the relation $\mathbb{E}[\log(\sup_{\mathcal{V}} \Lambda_{\phi_0})] < 0$, we apply recursively the Theorem 2.1 land prove the existence of continuous first and second derivatives of $(g_t(\theta))$ on $\mathcal{V}$ as solutions of functional SRE. The asymptotic normality follows from a Taylor development on the first partial derivatives of S_n (see Section 5 of [1] for more details):

$$\nabla_i S_n(\hat{\theta}_n) - \nabla_i S_n(\theta_0) = \mathbb{H} S_n(\tilde{\theta}_{n,i})(\hat{\theta}_n - \theta_0)$$

Then the asymptotic normality follows from the sufficient conditions:

1. $n^{-1/2} \nabla S_n(\theta_0) \to \mathcal{N}(0, \mathbf{Q})$,
2. $\|n^{-1} \mathbb{H} S_n(\tilde{\theta}_n) - \mathbf{P}\|$ converges a.s. to 0 for any sequence $(\tilde{\theta}_n)$ converging a.s. to θ_0 and $\mathbf{P}$ is invertible,
3. $n^{-1/2} \|\nabla \hat{S}_n(\hat{\theta}_n) - \nabla S_n(\hat{\theta}_n)\|$ converges a.s. to 0.

Due to its specific expression and that (Z_t) is a normalized difference of martingales sequence with finite moments of order 4, $(\nabla S_n(\theta_0))$ is a martingale, see [1] for detailed computations. Under **(MM)**, the CLT for differences of martingale applied to $(\nabla S_n(\theta_0))$ leads to the first condition. The first part of the second condition are derived from similar arguments than in the proof of Proposition 3.1 and an application of the Cesaro mean theorem ensuring that $n^{-1} \|\mathbb{H} S_n(\tilde{\theta}_n) - \sum_{t=1}^n \mathbb{H} s_t(\theta_0)\| \to 0$ a.s. The ergodic Theorem on $(\mathbb{H} s_t(\theta_0))$ with **(MM)** leads to $\|n^{-1} \mathbb{H} S_n(\tilde{\theta}_n) - \mathbf{P}\| \to 0$ a.s. The fact that $\mathbf{P}$ is invertible follows from **(LI)**, see [1] for detailed computations. Finally the

third condition is obtained by applying Theorem 2.10 of [17] on the SRE satisfied by (∇g_t) and its approximative SRE satisfied by $(\nabla \hat{g}_t)$ uniformly on $\mathcal{V}$. Remark that the condition $\mathbb{E}[\log(\sup_{\mathcal{V}} \|\nabla g_0\|)] < \infty$ of Theorem 2.10 of [17] is automatically satisfied using the Proposition 2.1 on the functional SRE of ∇g_t and the local uniform norm on $\mathcal{V}$. Thus $\sup_{\mathcal{V}} \|\nabla \hat{g}_t - \nabla g_t\| \xrightarrow{\text{e.a.s.}} 0$ as $t \to \infty$ and Lipschitz conditions on $\nabla \ell^{-1}$ and $\nabla \log(\det(\ell))$ in **(LM)** and arguments similar than in [16] leads to the desired result. □

3.5 A first application to GARCH type models

In the GARCH type models, the stationarity assumption **(ST)** is crucial, whereas the continuous invertibility condition **(CI)** is automatically satisfied due to the form of the model. Sufficient conditions for **(ST)** have been extensively studied in the literature and they can be necessary in linear cases. The asymptotic properties of the QMLE (that coincides with the SRE based inference in these cases) follow from Theorems 3.1 and 3.2. Thus, we recover and slightly refine existing results in the AGARCH and CCC-GARCH models (we refer the reader to [16] and [8] respectively for details in these both cases).

First, let us consider the univariate AGARCH(p, q) model introduced in [7], [20] and studied in [16]:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|X_{t-i}| - \gamma X_{t-i})^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \qquad t \in \mathbb{Z},$$

where $\alpha_0 > 0$, α_i, $\beta_j \geq 0$ and $|\gamma| \leq 1$ (it coincides with the GARCH(p, q) model if $\gamma = 0$. Then we derive the strong consistency and the asymptotic normality directly from our Theorems 3.1 and 3.2. The conditions we obtained coincides with these of Theorem 5.5 and Theorem 8.1 of [17] except that the technical (and superfluous) condition (8.1) of [17] is not needed in our approach.

Second, let us consider the multivariate CCC-GARCH(p, q) model introduced by [3], first studied in [11] and refined in [8]

$$Diag(\Sigma_t^2) = A_0 + \sum_{i=1}^{q} A_i Diag(X_{t-i} X_{t-1}^T) + \sum_{i=1}^{p} B_i Diag(\Sigma_{t-i}^2)$$

and $(\Sigma_t^2)_{i,j} = \rho_{i,j} \sqrt{(\Sigma_t^2)_{i,i} (\Sigma_t^2)_{j,j}}$ for all (i, j), where $Diag(M)$ is the vector of the diagonal elements of M. A necessary and sufficient conditions for

(ST) is given in term of top Lyapunov condition in [8]. We recover the strong consistency and the asymptotic normality of [8] directly from our Theorems 3.1 and 3.2 except that we do not need the superfluous condition $\forall\theta \in \Theta, \quad |B_\theta(z)| = 0 \Longrightarrow |z| > 1$ assumed in **A2** of [8]. It is due to our local uniform approach of the SRE that improves the classical uniform approach used in [8]. The inequality (4.10) of [8] does not always holds in our context.

4 Application to the EGARCH$(1,1)$ model

4.1 Definition of the model

Let (Z_t) be an iid sequence of random variables not concentrated on two points such that $\mathbb{E}(Z_0^2) = 1$. The EGARCH$(1,1)$ model introduced by Nelson [13] is an AR(1) model on $\log\sigma_t^2$

$$X_t = \sigma_t Z_t \quad \text{with} \quad \log\sigma_t^2 = \alpha_0 + \beta_0 \log\sigma_{t-1}^2 + W_{t-1}(\theta_0)$$

where $W_t(\theta_0) = \gamma_0 Z_t + \delta_0 |Z_t|$ are the innovations of this AR(1) model. Let $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \delta_0)$ be the unknown parameter. Assume that $\theta_0 \in \Theta$ where Θ is a compact subset of $\mathbb{R}\times]-1,1[\times\{(\delta,\gamma) \in \mathbb{R}^2;\ \delta \geq |\gamma|\}$ (see the next subsection for a discussion on Θ). The restriction on the parameter β_0 ($|\beta_0| < 1$) is classical (see [17]) and sufficient for **(ST)** and thus for the existence of a stationary solution. Then we have a MA(∞) representation for the logarithm of the squared volatility:

$$\log\sigma_t^2 = \alpha_0(1-\beta_0)^{-1} + \sum_{k=1}^{\infty} \beta_0^{k-1} W_{t-k}(\theta_0). \tag{9}$$

The moments assumptions on Z_t ensures that the process $(\log\sigma_t^2)$ is ergodic, strongly and weakly stationary. Then the volatilities process (σ_t^2) is also ergodic and strongly stationary. However, it does not necessarily have finite variance.

Remark 4.1. *The EGARCH$(1,1)$ model takes into account some stylized facts such as the asymmetry in the squared volatility: if $Z_t > 0$, then $\log\sigma_t^2 = \alpha_0 + \beta_0\log\sigma_{t-1}^2 + (\gamma_0+\delta_0)Z_{t-1}$ and $\log\sigma_t^2 = \alpha_0 + \beta_0\log\sigma_{t-1}^2 + (\gamma_0-\delta_0)Z_{t-1}$ otherwise. So conditioning on the sign of the innovation, Z_t, the change of the log-volatility $\log(\sigma_{t+1}^2/\sigma_t^2)$ is asymmetric.*

4.2 Invertibility

The invertibility of the stationary solution of the EGARCH$(1,1)$ model does not hold in general. Straumann and Mikosch gives a partial answer to this important question in [17]: as $(\log \sigma_t^2)$ satisfies the SRE

$$\log \sigma_t^2 = \alpha_0 + \beta_0 \log \sigma_{t-1}^2 + W_{t-1}(\theta_0) \exp\Big(- \frac{\log \sigma_{t-1}^2}{2} \Big), \quad \text{for all } t \in \mathbb{Z},$$

if the above SRE admits a unique stationary solution, then the model is invertible. Keeping the notation of our Section 2, the function h is here the logarithmic function and the SRE (5) holds with (ϕ_t) defined by

$$\phi_t(\cdot;\theta) : s \mapsto \alpha + \beta s + (\gamma X_{t-1} + \delta\,|X_{t-1}|) \exp(-s/2) \tag{10}$$

We check that the ϕ_t are random functions generated by $\mathcal{G}_{t-1}$. As in [17] we restrict $\phi_t(\cdot;\theta)$ on the complete separable metric space $[\alpha/(1-\beta), \infty)$ equipped with $d(x,y) = |x-y|$. Then for any $\theta_0 \in \Theta$, as **(ST)** is satisfied, $(\phi_t(\cdot;\theta_0))$ is a stationary ergodic sequence of Lipschitz maps from $[\alpha_0/(1-\beta_0), \infty)$ to $[\alpha_0/(1-\beta_0), \infty)$ with the Lipschitz coefficient

$$\Lambda(\phi_t(\cdot,\theta_0)) \leq \max\{|\beta_0|, |2^{-1}(\gamma_0 X_{t-1} + \delta_0 |X_{t-1}|) \exp(-2^{-1}\alpha_0(1-\beta_0)) - \beta_0|\}.$$

Thus a sufficient condition for the invertibility condition (4) (with $r=1$) is

$$\mathbb{E}[\log(\max\{|\beta_0|, |2^{-1}(\gamma_0 X_{t-1} + \delta_0 |X_{t-1}|) \exp(-2^{-1}\alpha_0/(1-\beta_0)) - \beta_0|\})] < 0. \tag{11}$$

4.3 Condition on the compact set Θ

Let us detail in the sequel the compact sets Θ such that the relation (11) is satisfied for any $\theta_0 \in \Theta$. From the MA(∞) representation (9) of $\log \sigma_t^2$ we rewrite the condition (11) as

$$\mathbb{E}\Big[\log \Big(\max \Big\{ |\beta|, \Big| 2^{-1} \exp\Big(2^{-1} \sum_{k=0}^{\infty} \beta^k (\gamma Z_{-k-1} + \delta\,|Z_{-k-1}|) \Big) \times (\gamma Z_0 + \delta\,|Z_0|) - \beta \Big| \Big\} \Big) \Big] < 0 \tag{12}$$

which does not depend on α and is easier to check. Using the Monte Carlo algorithm, assuming the Z_t to be normally distributed, we report in Figure

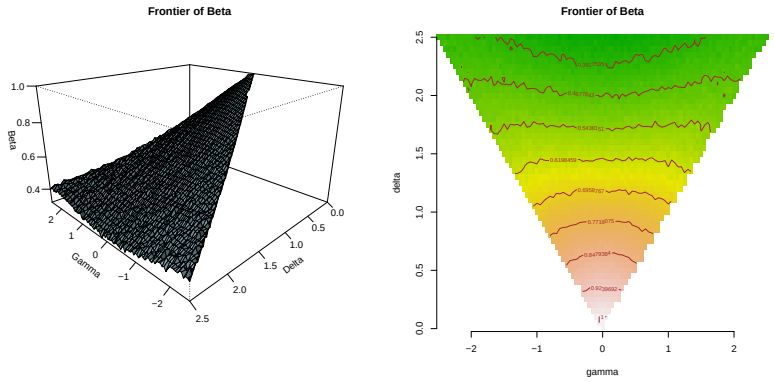


Figure 1: Perspective and contour plots of the admissible region for continuous invertibility.

1 the largest values of β that satisfies the condition (12) on a grid of values of γ and δ.

Graphically, it seems that the most restrictive condition on β is when $\gamma = 0$ when (γ, δ) is far away from $(0, 0)$. Remark however that $|\beta|$ is never constrained to 0 as for $\beta = \gamma = 0$ the condition (12) becomes $\mathbb{E}[\log(\delta|Z_0|/2)] < \mathbb{E}[\delta|Z_0|/2]$ that is always satisfied as $\log(x) \leq x - 1$ for all $x > 0$. We then conjecture that the admissible set for θ is unbounded. Finally, remark that when $\beta = 0$, the EGARCH$(1, 1)$ model degenerates to a sub-model $Y_t = \exp(\alpha + \gamma Z_{t-1} + \delta\,|Z_{t-1}|)Z_t$ for which the admissible parameters set is well known, see [16] for details.

4.4 Asymptotic properties of the parametric inference and forecast

With the reasonable choice of Θ made in the last subsection, we know that **(ST)** is satisfied. Moreover, **(CL)** is also automatically satisfied as

$$(\Lambda_{\phi_t}(\theta)) = (\max(|\beta|, |2^{-1}(\gamma X_{t-1} + \delta|X_{t-1}|)\exp(-2^{-1}\alpha/(1-\beta)) - \beta|))$$

are stationary continuous functions of θ. Thus, as a corollary of Theorem 3.1 and Proposition 3.1, we get

Corollary 4.1. *If Θ satisfying the condition (12) and $\theta_0 \in \Theta$, then $\hat{\theta}_n \to \theta_0$ and $\hat{\sigma}_n^2 - \sigma_n^2 \to 0$ a.s. as $n \to \infty$ with $\hat{\sigma}_t^2 = \exp(\hat{g}_t(\hat{\theta}_n))$.*

Proof. The condition **(CI)** follows from $\mathbb{E}[\log \Lambda(\phi_t(,\theta))] < 0$ by assumption of Θ and $\mathbb{E}[\sup_\Theta \log \Lambda(\phi_t(,\theta))] < \infty$ since $\mathbb{E}\log|X_{t-1}| = E(\log\sigma + \log|Z_{t-1}|) < \infty$ as $\log\sigma_t^2$ has a MA(∞) representation (9) and Z is integrable. Moreover as $\log^+(d(\phi_0(0,\theta),0)) = \log^+|\alpha + (\gamma X_{-1} + \delta|X_{-1}|)|$ then $\mathbb{E}[\sup_\Theta \log^+(d(\phi_0(y,\theta),y)] < \infty$ for $y = 0$.
In the EGARCH(1,1) model the link function is the exponential function $\ell(x) = \exp(x)$ and since we have $\log\sigma_t^2 \geq \alpha/(1-\beta)$, $1/\ell(x) = \exp(-x)$ is a Lipschitz function ($\log(\det(\ell)) = id$ is also a Lipschitz function). Moreover the volatility process (σ_t^2) is bounded from below by $C(\theta) = \exp(\alpha/(1-\beta))$. Finally, the identifiability condition $g_0(\theta) = h(\theta_0)$ iff $\theta = \theta_0$ is checked in Section 5.1 of [17]. □

As a corollary of Theorem 3.2 we get the asymptotic normality of the inference in the EGARCH(1,1) model. It holds under the following necessary and sufficient condition of the existence of the asymptotic variance $\mathbf{V}$:

(MM') The innovation satisfy $\mathbb{E}[Z_0^4] < \infty$ and $\beta^2 - \delta\mathbb{E}|Z_0| + (\delta^2+\gamma^2)/4 < 1$.

Corollary 4.2. *Assume that Θ is well chosen as in Corollary 4.1 and that* **(MM')** *holds then $\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathcal{N}(0, \mathbf{V})$ in law with an invertible matrix $\mathbf{V}$ explicitly given in the Appendix 3.*

Proof. By definition, (ϕ_t) is 2-times continuously differentiable and simple computations give $D_x(\phi_t)(x,\theta) = \beta - 2^{-1}(\gamma X_{t-1} + \delta|X_{t-1}|)\exp(-x/2)$, $D_\theta(\phi_t)(x,\theta) = (1, x, X_{t-1}\exp(-x/2), |X_{t-1}|\exp(-x/2))^T$, $D^2_{x^2}(\phi_t)(x,\theta) = 4^{-1}(\gamma X_{t-1} + \delta|X_{t-1}|)\exp(-x/2)$, $D^2_{x,\theta}(\phi_t)(x,\theta) = (0, 1, 2^{-1}X_{t-1}\exp(-x/2), 2^{-1}|X_{t-1}|\exp(-x/2))^T$ and $D^2_{\theta^2}(\phi_t)(x,\theta) = 0$. Moreover, as the link function is $\ell(x) = \exp(x)$ is also 2-times continuously differentiable the last assertion of the condition **(AV)** of Theorem 3.2 holds. The fact that **(MM)** holds under the conditions $\mathbb{E}[Z_0^4] < \infty$ and $\beta^2 - \delta\mathbb{E}|Z_0| + (\delta^2+\gamma^2)/4 < 1$ is technical and postponed to the Appendix 1. The fact that **(LI)** holds if Z_0 is not concentrated on two points is classical and proved in the Appendix 2 for the sake of completeness. Assumption **(DL)** is satisfied from the expressions of the derivatives (that are Lipschitz functions) and as all the logarithmic moments are finite due to $\mathbb{E}[\log(X_{t-1}^2)] < \infty$. Finally **(LM)** is automatically satisfied due to the specific expression of the link function. □

4.5 The asymptotic variance V

Computations give $\mathbf{V} = (\mathbb{E}Z_0^4 - 1)\mathbf{B}^{-1}$ with $\mathbf{B} = \mathbb{E}[\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T]$. Lemma 5.1 in Appendix 1 states that the conditions $\beta^2 - \delta\mathbb{E}|Z_0| + (\delta^2 +$

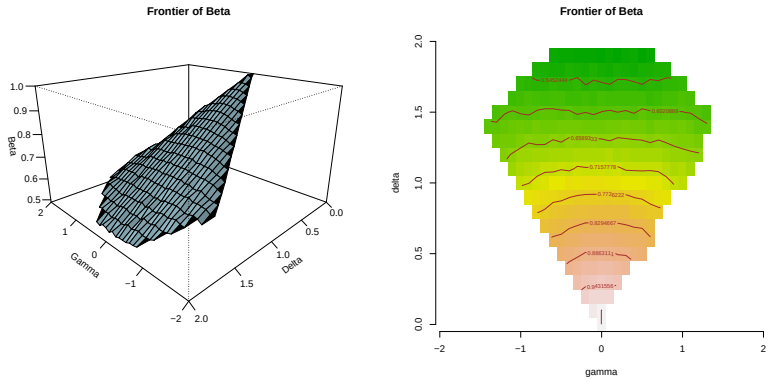


Figure 2: Perspective and contour plots of the admissible region for the asymptotic normality.

$\gamma^2)/4 < 1$ and $\mathbb{E}[Z_0^2] = 1$ are necessary and sufficient for the existence of $\mathbf{B}$. Assuming the Z_t to be normally distributed, we report in Figure 2 the largest values of β that satisfies the conditions (12) and $\beta^2 - \delta\mathbb{E}|Z_0| + (\delta^2+\gamma^2)/4 < 1$ on a grid of values of γ and δ such that $\gamma^2+\delta^2 \leq 4$. The additional condition does not affect much the constraint of β on this region of (γ, δ). However, for $\gamma^2+\delta^2 > 4$, the condition of existence of $\mathbf{V}$ imposes to exclude also small values of β starting from $\beta = 0$. Then the resulting admissible region is now bounded and seems to be convex.

The technical explicit computation of $\mathbf{B}$ is given in the Appendix 3. Remark that the quasi likelihood (computable if the observations Z_t were observable) is equal to $2^{-1}\sum_{t=1}^{n}(X_t^2/\hat{h}_t(\theta) + \log(\hat{h}_t(\theta)))$ where

$$\hat{h}_t(\theta) = \alpha + \beta\hat{h}_{t-1}(\theta) + (\gamma Z_{t-1} + \delta|Z_{t-1}|), \quad t > 1, \quad \text{and} \quad \hat{h}_1(\theta) = \zeta(\theta) \text{ fixed.}$$

The likelihood $2^{-1}\sum_{t=1}^{n}(X_t^2/h_t(\theta) + \log(h_t(\theta)))$ coincides with SRE criteria $S_n(\theta)$ at the point θ_0 as $h_t(\theta_0) = g_t(\theta_0) = \sigma_t^2$. However, the two criteria differ for $\theta \neq \theta_0$. To see it, from the MA(∞) representation (9) we can easily compute the partial derivative of h_0 with respect to β:

$$\frac{\partial h_0}{\partial \beta}(\theta) = \frac{\alpha\beta}{(1-\beta)^2} + \sum_{k\geq 1} k\beta^{k-1}(\gamma Z_{-k-1} + \delta|Z_{-k-1}|).$$

The value of this partial derivative is different from $\partial h_0/\partial\beta(\theta)$, even for $\theta = \theta_0$. Thus the estimator given by our SRE approach and the QMLE is

different. Moreover, the asymptotic variances of the two estimators differ and even for gaussian innovations (Z_t) our estimator $\hat{\theta}_n$ is not asymptotically efficient in the EGARCH$(1,1)$ case.

5 Numerical parametric inferences and forecasting

5.1 Monte Carlo analysis of the estimation risk

Let us study by Monte Carlo algorithm on 1000 replications of the risk for estimating a parameter $\theta_0 \in \Theta$, see Table 1. In accordance with our

Table 1: Statistical inference and normal approximation

	T=	512			1024			2048		
θ	θ_0	mean	rmse	napp	mean	rmse	napp	mean	rmse	napp
α	-0.399	-.381	.127	.059	-.393	.041	.042	-.396	.030	.030
β	.9	.874	.170	.023	.897	.017	.016	.899	.012	.011
γ	-.3	-.300	.057	.045	-.301	.033	.032	-.299	.023	.023
δ	.5	.488	.097	.075	.492	.052	.053	.496	.038	.038

theoretical results, the mean value of the estimators $\hat{\theta}_n$ over replications are closer the larger is the sample size ($T = 512$, 1024 or 2048). The columns "rmse" give the empirical Root Mean Square Error (RMSE) computed on the 1000 replications. The columns "napp" give the normal approximation of the RMSE. From the exact computation of $\mathbf{V}$ in Appendix 3 we compute the value of the asymptotic variance at the point θ_0 and divide the corresponding standard deviation by $\sqrt{T}$. From Table 1, the normal approximation seems to hold for T larger than 1024.

5.2 Estimation of the asymptotic covariance matrix

Assuming that $\mathbb{E}[Z_0^4] = 3$ in the EGARCH(1,1) model, we have two ways of estimating the asymptotic covariance matrix of our estimator $\hat{\theta}_n$. The first one relies on the plug-in $\mathbf{V}^{TH}(\hat{\theta}_n)$ in the explicit formula $\theta \to \mathbf{V}^{TH}(\theta)$ given in the Appendix 3. The second one relies on the SRE satisfied by $\hat{g}_t(\theta)$:

$$\nabla \hat{g}_t(\theta) = (1, \hat{g}_{t-1}(\theta), X_{t-1}\exp(-\hat{g}_{t-1}(\theta)/2), |X_{t-1}|\exp(-\hat{g}_{t-1}(\theta)/2))^T + (\beta - (\gamma X_{t-1} + \delta|X_{t-1}|)\exp(-\hat{g}_{t-1}(\theta)/2)/2)\nabla \hat{g}_{t-1}(\theta).$$

Running the SRE over $1 \leq t \leq n$ we obtain n values $\nabla \hat{g}_t(\theta)$, then we approximate $\mathbf{B}$ by $n^{-1}\sum_{t=1}^{n} \nabla \hat{g}_t(\theta) \nabla \hat{g}_t(\theta)^T$ and finally an approximation of $\mathbf{V} = 2\mathbf{B}^{-1}$ by inverting numerically the approximation of $\mathbf{B}$.

We compare the risk of the two estimation procedure associated with the Riemannian distance for symmetric positives definitive matrix defined by

$$d(A,B) = \sqrt{\sum_{k=1}^{4} \log^2 \nu_k(AB^{-1})}$$

where $\nu_1(AB^{-1}), \ldots, \nu_4(AB^{-1})$ are the eigenvalues of the matrix AB^{-1}. To estimate the risk we sample randomly 100 parameters $\theta_1, \ldots, \theta_{100}$ in a compact set satisfying the conditions for the convergence of the SRE and the existence of $\mathbf{B}$. For each θ_k we simulate a path of $n = 512$, 1024 and 2048 observations and obtain the estimated values $\hat{\theta}_k$. Then we calculate the errors $d(\mathbf{V}^{TH}(\theta_k), \mathbf{V}^{SRE}(\theta_k))$, $d(\mathbf{V}^{TH}(\theta_k), \mathbf{V}^{TH}(\hat{\theta}_k))$,$d(\mathbf{V}^{TH}(\theta_k), \mathbf{V}^{SRE}(\hat{\theta}_k))$. We report the means of these distances in Table 2

Table 2:

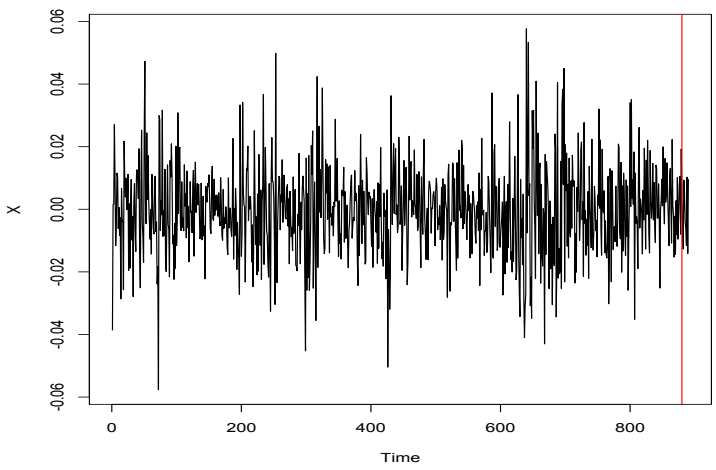
n	$d(\mathbf{V}^{TH}(\theta_k), \mathbf{V}^{SRE}(\theta_k))$	$d(\mathbf{V}^{TH}(\theta_k), \mathbf{V}^{TH}(\hat{\theta}_k))$	$d(\mathbf{V}^{TH}(\theta_k), \mathbf{V}^{SRE}(\hat{\theta}_k))$
512	.074	.788	.924
1024	.065	.767	.780
2048	.064	.426	.457

The table show that it is very safe to use $\mathbf{V}^{SRE}(\hat{\theta}_k)$. This might suggests that for general volatility models, we can safely use $\mathbf{V}^{SRE}(\hat{\theta}_k)$ if the explicit formula for $\mathbf{V}$ is not known.

5.3 Forecasting the volatility on real data

We investigate the filtering and the forecasting problem of the daily log-return of the Standard & Poor's 500 data from Jan 4th, 2000 to Jul 22th, 2003 ($n = 890$). More precisely, we consider the process (X_t) of the log difference of the daily close prices. The period has been chosen in order that the stationary condition might be reasonable (it is a period of high volatility). To test the stationarity, we perform a KPSS test and an augmented Dickey-Fuller on the logarithm of the X_t^2 (plotted in Figure 3) that give

Figure 3: The daily log-return of the Standard & Poor's 500 data from Jan 4th, 2000 to Jul 22th, 2003.



p-values 0.098 and less than 0.01 respectively. The unit root hypothesis can thus be reasonably rejected.

From now we assume that the general volatility model (1) is satisfied by the real data, i.e. $X_t = \sigma_t Z_t$ with a known parametric form ψ_t. Below we investigate some of the more classical models: the GARCH(1,1) model,the GARCH(1,1) model with Student innovations, the APGARCH(1,1) model and the EGARCH(1,1) model. Let us consider the classical in sample and out of sample procedures with two steps: for the in sample procedure, first we estimate $\hat{\theta}_n$ on the 890 first observations and then we investigate the performance of the natural forecast $\hat{\sigma}_t = \ell(\hat{g}_t(\hat{\theta}_n))$ of the volatility σ_t for $1 \leq t \leq 890$. For the out of sample procedure, first we estimate $\hat{\theta}_n$ on the 880 first observations and then we investigate the performance of the natural forecast $\hat{\sigma}_t = \ell(\hat{g}_t(\hat{\theta}_n))$ of the volatility σ_t for $881 \leq t \leq 890$ (the red line in figure 3 splits the training and validation data sets). One difficulty of evaluating the forecasting performance is that we could not observe the true volatility process. However, with the 5 mins high frequency prices data in hand, we bypass the problem: the Realized Volatility (RV) is used as a consistent proxy of the conditional volatility. To measure the forecasting performance we use the quasi likelihood (QLIK) criteria. As noticed by [14], it is robust with respect to unbiased proxy. Moreover, we claim that it is

more relevant than the Mean Square Error criteria as it does not involve moments of order larger than one of the volatilities. For the sake of completeness, we also give the QLIK criteria for forecasting the proxies obtained by 2 out of sample procedures, the Riskmetrics one and the rolling volatility one.

To sum up, we list the parametric forms of the volatilities in the different models that are investigated here:

- GARCH(1, 1): $\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2$,
- APGARCH(1, 1): $\sigma_t^\delta = \omega + \alpha(|X_{t-1}| - \gamma X_{t-1})^\delta + \beta\sigma_{t-1}^\delta$,
- Rolling Volatility (60days): the moving average $1/60\sum_{i=1}^{60} X_{t-i}^2$,
- Riskmetrics (Exponentially weighted moving average model)

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)X_{t-1}^2 \qquad \text{where} \qquad \lambda = .94,$$

- EGARCH(1, 1): $\log\sigma_t^2 = \alpha + \beta\log\sigma_{t-1}^2 + (\gamma X_{t-1} + \delta|X_{t-1}|)/\sigma_t$.

Denoting $(r_{1,m,t}, \ldots, r_{m,m,t})$ the m intra day returns, we define the daily RV with m-frequency as

$$\mathrm{RV}_t^{(m)} = \sum_{i=1}^{i=m} r_{i,m,t}^2.$$

Finally, the QLIK criteria of the forecasting of the proxies RV by $\hat{\sigma}^2$ is

$$\mathrm{QLIK}(\hat{\sigma}^2) = \sum_{t=1}^{n} \log(\hat{\sigma}_t^2) + \frac{RV_t}{\hat{\sigma}_t^2}.$$

We report the in sample performances in Table 3, the out of sample ones in Table 4. The best forecast for each proxy is bolded. The performance of the EGARCH(1,1) is close to the best one in each case.

Applying our approach in the EGARCH(1,1) model, we find the estimations $\hat{\alpha}_n = -0.312$, $\hat{\beta}_n = 0.976$, $\hat{\gamma}_n = -0.122$ and $\hat{\delta}_n = 0.122$ in the in sample procedure, $\hat{\alpha}_{n'} = -0.324$, $\hat{\beta}_{n'} = 0.974$, $\hat{\gamma}_n = -0.123$ and $\hat{\delta}_n = 0.123$. These values are at the frontier of the linear constraint $\delta \geq |\gamma|$ and satisfy the Lyapunov condition and the condition of existence of $\mathbf{V}$. The EGARCH(1,1)

Table 3: In sample performances of the different forecasts

QLIK	X_t^2	5min RV	15min RV	65min RV
GARCH(1,1)	-7.438	-7.517	-7.592	-7.607
GARCH(1,1) Student	-7.439	-7.516	-7.589	-7.604
APGARCH(1,1)	**-7.489**	-7.528	-7.613	-7.618
Rolling Volatility	-7.267	-7.332	-7.414	-7.455
Riskmetrics	-7.429	-7.510	-7.583	-7.597
EGARCH(1,1)	-7.487	**-7.537**	**-7.619**	**-7.626**

Table 4: Out of sample performances of the different forecasts

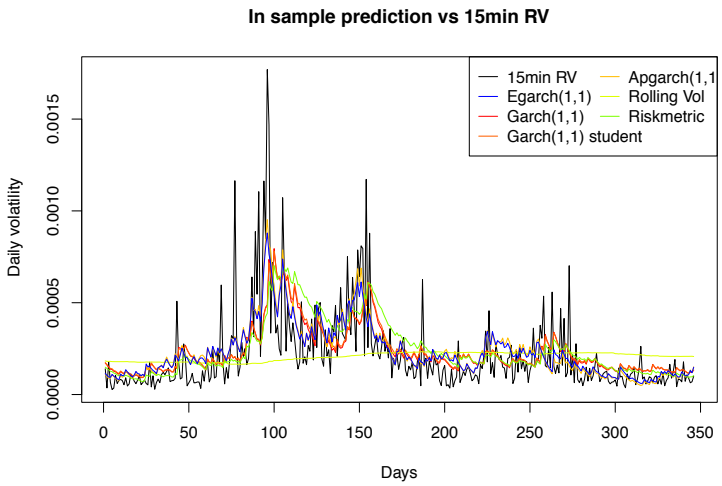
QLIK	X_t^2	5min RV	15min RV	65min RV
GARCH(1,1)	**-8.285**	-8.153	**-8.192**	**-8.170**
GARCH(1,1) Student	-8.226	-8.131	-8.153	-8.111
APGARCH(1,1)	-8.226	-8.130	-8.152	-8.111
Rolling Volatility	-8.195	-8.096	-8.128	-8.111
Riskmetrics	-8.053	-7.978	-7.998	-7.977
EGARCH(1,1)	-8.272	**-8.155**	-8.184	-8.135

Table 5: Confident intervals given by the normal approximation in the EGARCH type model $\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \eta X_{t-1}^- / \sigma_t$

In-S.	Value	Confident interval
α	-.312	[-.450, -.175]
β	.976	[.962, .989]
η	.243	[.171, .315]

Out of S.	value	Confident interval
α	-.324	[-.464, -.185]
β	.974	[.961,.988]
η	.246	[.172,.320]

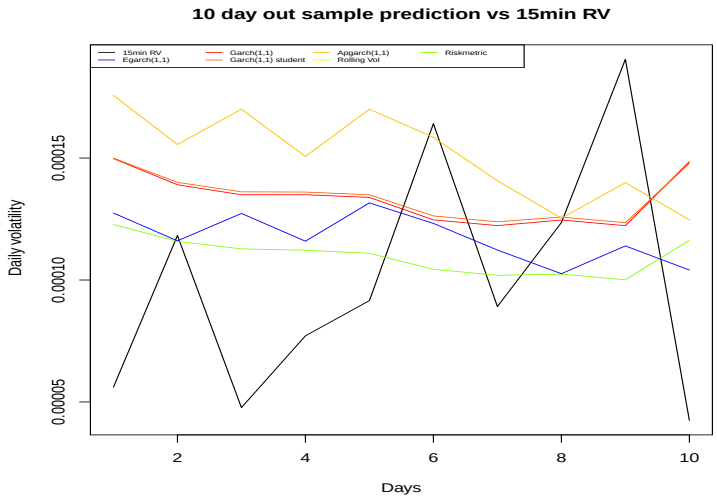
Figure 4:



model degenerates to the following model: $\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \eta X_{t-1}^- / \sigma_t$. where $\eta = \delta - \gamma$. It is a parsimonious model with only 3 parameters and the constraint $\eta \geq 0$ ensures the positivity of the innovations $\eta X_{t-1}^- / \sigma_t$ of the AR(1) model. As our estimations are not on the frontier $\eta = 0$, the asymptotic normality holds. Plugging our estimations in the explicit formula of the asymptotic variance given in Appendix 3 provides the 95% confident intervals for (α, γ, η) that are reported in Table 5. Remark that the behavior of the model $\log \sigma_t^2 = \alpha + \beta \log \sigma_{t-1}^2 + \eta X_{t-1}^- / \sigma_t$ differs completely wether the observations are positive or negative. It is in accordance with the plot in Figure 3: isolated extremes have negative values. It explains why the non symmetric AGARCH and EGARCH models have better in sample performance, see Table 3. For the out of sample performance, the EGARCH and the GARCH models have the best forecasting performances, see Table 4.

We draw the 15 min realized volatilities and the in sample forecasts for $540 \leq t \leq 890$ in Figure 4 (we only consider the last observations where the recurrent formula given by Riskmetrics should be the best) and the out of sample forecasts for $881 \leq t \leq 890$ in Figure 5. Graphically, the forecasts of the EGARCH model are satisfactory because it follows the global fluctuations of the realized volatilities. It is very close to the forecasts of the APGARCH model: their fluctuations are similar and their values are close (the EGARCH forecasts are slightly smaller). But as the APGARCH model

Figure 5:



relies on 5 unknown coefficients, we prefer to work with the more parsimonious EGARCH model (that degenerates to a form with only 3 unknown parameters).

Acknowledgments

We are grateful to Romain Allez and Vincent Vargas of the CFM and the CEREMADE for providing us the high volatilities data.

References

[1] J.-M. Bardet and O. Wintenberger. Asymptotic normality of the quasi maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.*, 37:2730–2759, 2009.

[2] L. Berkes I., Horvath and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9:201–227, 2003.

[3] T.P. Bollerslev. Modeling the coherence in short run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, 72:498–505, 1990.

[4] P. Bougerol. Kalman filtering with random coefficients and contractions. *SIAM J. Control and Optimization*, 31:942–959, 1993.

[5] P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, 52:115–127, 1992.

[6] A. Demos and D. Kyriakopoulou. Asymptotic expansions of the QMLEs in the EGARCH(1,1) model. preprint, 2009.

[7] Z. Ding, C. W. J. Granger, and R. Engle. A long memory property of stock market returns and a new model. *J. Empirical Finance*, 1:83?106, 1993.

[8] C. Francq and Zakoïan J.M. QML estimation of a class of multivariate asymmetric GARCH models. forthcoming in Econometric Theory, 2011.

[9] C. Francq and J.-M. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10:605–637, 2004.

[10] T. Jeantheau. *Modèles autorégressifs à erreur conditionellement hétéroscédastique*. PhD thesis, Université Paris VII, 1993.

[11] T. Jeantheau. Strong consistency of estimation for multivariate arch models. *Econometric Theory*, 14:70–86, 1998.

[12] D. B. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990.

[13] D. B. Nelson. Conditional heteroskedasticity in asset returns : A new approach. *Econometrica*, 59:347–370, 1991.

[14] A. J. Patton. Volatility forecast comparison using imperfect volatility proxies. *J. Econometrics*, 160(1):246 – 256, 2011.

[15] A. Sorokin. Non-invertibility in some heteroscedastic models. Arxiv preprint #1104.3318, 2011.

[16] D. Straumann. *Estimation in conditionally heteroscedastic time series models.*, volume 181 of *Lectures Notes in Statistics*. Springer, New York, 2005.

[17] D. Straumann and T. Mikosch. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equation approach. *Ann. Statist.*, 34:2449–2495, 2006.

[18] H Tong. *Non-Linear Time Series, A Dynamical System Approach*, volume 6 of *Oxford Statistical Science Series*. Oxford Science Publications, Oxford, 1993.

[19] P. Zaffaroni. Whittle estimation of EGARCH and other exponential volatility models. *J. Econometrics*, 151:190–200, 2009.

[20] J.-M. Zakoïan. Threshold heteroscedastic models. *J. Econom. Dynam. Control*, 18:931–955, 1994.

Appendix 1: checking the assumption (MM)

Similar computations have been done in [6]. Remember that $\mathbf{V} = \mathbf{P}^{-1}\mathbf{Q}\mathbf{P}^{-1}$ with $\mathbf{P} = \mathbb{E}[\mathbb{H}s_0(\theta_0)]$ and $\mathbf{Q} = \mathbb{E}[\nabla s_0(\theta_0)\nabla s_0(\theta_0)^T]$. Let us first prove the three identities $\mathbf{P} = 2^{-1}\mathbf{B}$, $\mathbf{Q} = 4^{-1}(\mathbb{E}Z_0^4 - 1)\mathbf{B}$ and thus $\mathbf{V} = (\mathbb{E}Z_0^4 - 1)\mathbf{B}^{-1}$ with $\mathbf{B} = \mathbb{E}\Big[\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T\Big]$. For the first identity, we compute

$$\begin{aligned}\mathbf{P} &= 2^{-1}\mathbb{E}\left[(\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T Z_0^2 + \mathbb{H}g_t(\theta_0)(1 - Z_0^2)\right] \\ &= 2^{-1}\mathbb{E}[\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T] = 2^{-1}\mathbf{B}.\end{aligned}$$

For the second identity, we compute

$$\begin{aligned}\mathbf{Q} &= \mathbb{E}\left[\frac{1}{4}\mathbb{E}\left[\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T(1 - \mathbb{Z}_t^2)^2\right]|\mathcal{F}_{t-1}\right] \\ &= 4^{-1}\mathbb{E}[(1 - \mathbb{Z}_0^2)^2]\mathbb{E}[\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T] = 4^{-1}(\mathbb{E}Z_0^4 - 1)\mathbf{B}\end{aligned}$$

and the third identity follows the first ones. Thus, for checking the assumption **(MM)**, it is enough to check that diagonal coefficients $\mathbf{B}_{ii}$ are well defined when $\mathbb{E}(Z_0^4) < \infty$. Let us denote $W_t = \gamma Z_t + \delta|Z_t|$, $U_t = (1, \log\sigma_t^2, Z_t, |Z_t|)$ and $V_t = \beta - \frac{1}{2}(\gamma Z_t + \delta\,|Z_t|)$. Then $(\nabla g_t(\theta_0))$ is the solution of the linear SRE

$$\nabla g_t(\theta_0) = U_{t-1} + V_{t-1}\nabla g_{t-1}(\theta_0) = \sum_{l=1}^{\infty}\left(U_{t-l}\prod_{k=1}^{l-1}V_{t-k}\right).$$

Using the convention $\prod_{k=1}^{0} V_{t-k} = 1$, we obtain the expression

$$\nabla g_t(\theta_0) = \sum_{l=1}^{\infty}\left(U_{t-l}\prod_{k=1}^{l-1}V_{t-k}\right).$$

More precisely, we have the expressions: then

$$\mathbf{B}_{11} = \mathbb{E}\left(\frac{\partial g_t(\theta_0)}{\partial \theta_1}\right)^2 = \mathbb{E}\left[\sum_{l=1}^{\infty}\prod_{k=1}^{l-1} V_{t-k}\right]^2,$$

$$\mathbf{B}_{22} = \mathbb{E}\left(\frac{\partial g_t(\theta_0)}{\partial \theta_2}\right)^2 = \mathbb{E}\left[\sum_{l=1}^{\infty}\log\sigma_{t-l}^2\prod_{k=1}^{l-1} V_{t-k}\right]^2,$$

$$\mathbf{B}_{33} = \mathbb{E}\left(\frac{\partial g_t(\theta_0)}{\partial \theta_3}\right)^2 = \mathbb{E}\left[\sum_{l=1}^{\infty} Z_{t-l}\prod_{k=1}^{l-1} V_{t-k}\right]^2,$$

$$\mathbf{B}_{44} = \mathbb{E}\left(\frac{\partial g_t(\theta_0)}{\partial \theta_i}\right)^2 = \mathbb{E}\left[\sum_{l=1}^{\infty}|Z_{t-l}|\prod_{k=1}^{l-1} V_{t-k}\right]^2.$$

To prove that condition **(MM)** is satisfied, i.e. that $\sum_{i=1}^{4}\mathbf{B}_{ii} < \infty$, we use the following Lemma

Lemma 5.1. $\sum_{i=1}^{4}\mathbf{B}_{ii} < \infty$ *iff* $\mathbb{E}V_0^2 < 1$ *iff* $\beta^2 - \delta\mathbb{E}|Z_0| + (\delta^2+\gamma^2)/4 < 1$.

Proof. That the first coefficient B_{11} is finite comes easily:

$$\begin{aligned}
\mathbf{B}_{11} = \mathbb{E}(\sum_{l=1}^{\infty}\prod_{k=1}^{l-1} V_{t-k})^2 &= \mathbb{E}(\sum_{l=1}^{\infty}\sum_{l'=1}^{\infty}\prod_{k=1}^{l-1} V_{t-k}\prod_{k'=1}^{l'-1} V_{t-k'}) \\
&= \mathbb{E}(2\sum_{l\geq 1}^{\infty}(\prod_{k=1}^{l-1} V_{t-k})^2\sum_{l'>l}^{\infty}\prod_{k'=l}^{l'-1} V_{t-k'} + \mathbb{E}\sum_{l=1}^{\infty}(\prod_{k=1}^{l-1} {V_{t-k}}^2) \\
&= 2\sum_{l\geq 1}^{\infty}(\mathbb{E}V_0^2)^{l-1}\frac{\mathbb{E}V_0}{1-\mathbb{E}V_0} + \frac{1}{1-\mathbb{E}V_0^2} \\
&= 2\frac{1}{1-\mathbb{E}V_0^2}\times\frac{\mathbb{E}V_0}{1-\mathbb{E}V_0} + \frac{1}{1-\mathbb{E}V_0^2}.
\end{aligned}$$

For the second coefficient $\mathbf{B}_{22}$, it is more complicated. We need some preliminary work. We know that $W_t = \gamma Z_t + \delta|Z_t| = 2(\beta - V_t)$, and

$$\log\sigma_t^2 = \frac{\alpha}{1-\beta} + \sum_{k=1}^{\infty}\beta^{k-1}W_{t-k} = \frac{\alpha+2\beta}{1-\beta} - 2\sum_{k=1}^{\infty}\beta^{k-1}V_{t-k}$$

so, we decompose $\mathbf{B}_{22}$ into three parts,

$$\mathbf{B}_{22} = \mathbb{E}\left[\sum_{l=1}^{\infty} \log \sigma_{t-l}^2 \prod_{k=1}^{l-1} V_{t-k}\right]^2$$

$$= \mathbb{E}\left[\sum_{l=1}^{\infty}\left(\frac{\alpha+2\beta}{1-\beta} - 2\sum_{k=1}^{\infty}\beta^{k-1}V_{t-l-k}\right)\prod_{k'=1}^{l-1} V_{t-k'}\right]^2$$

$$= (\frac{\alpha+2\beta}{1-\beta})^2\mathbb{E}\left[\sum_{l=1}^{\infty}\prod_{k'=1}^{l-1} V_{t-k'}\right]^2 + 4\mathbb{E}\left[\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\beta^{k-1}V_{t-l-k}\prod_{k'=1}^{l-1} V_{t-k'}\right]^2$$

$$-4\times\frac{\alpha+2\beta}{1-\beta}\mathbb{E}\left[\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\beta^{k-1}V_{t-l-k}\prod_{k'=1}^{l-1} V_{t-k'}\right].$$

That the first term of the sum is finite is already known. For the last term, it is straightforward from $\mathbb{E}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\beta^{k-1}V_{t-l-k}\prod_{k'=1}^{l-1}V_{t-k'} = (1-\beta)^{-1}\mathbb{E}V_0/(1-\mathbb{E}V_0)$. For the second term of the sum, we need an expansion

$$\left[\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\beta^{k-1}V_{t-l-k}\prod_{k'=1}^{l-1} V_{t-k'}\right]^2$$

$$= 2\times\sum_{1\le l<l'<\infty 1}^{\infty}\sum_{p,q=1}^{\infty}\beta^{p+q-2}V_{t-l-p}V_{t-l'-q}\prod_{p'=1}^{l-1}V_{t-p'}^2\prod_{q'=l}^{l'-1}V_{t-q'}$$

$$+\sum_{l=1}^{\infty}\sum_{p,q=1}^{\infty}\beta^{p+q-2}V_{t-l-p}V_{t-l-q}\prod_{p'=1}^{l-1}V_{t-p'}^2$$

$$= 4\times\sum_{1\le l<l'<\infty}\sum_{1\le p<q<\infty}\beta^{p+q-2}V_{t-l-p}V_{t-l'-q}\prod_{p'=1}^{l-1}V_{t-p'}^2\prod_{q'=l}^{l'-1}V_{t-q'}$$

$$+2\times\sum_{1\le l<l'<\infty}\sum_{p=1}^{\infty}\beta^{2p-2}V_{t-l-p}V_{t-l'-p}\prod_{p'=1}^{l-1}V_{t-p'}^2\prod_{q'=l}^{l'-1}V_{t-q'}$$

$$+2\sum_{l=1}^{\infty}\sum_{1\le p<q<\infty}\beta^{p+q-2}V_{t-l-p}V_{t-l-q}\prod_{p'=1}^{l-1}V_{t-p'}^2 + \sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\beta^{2p-2}V_{t-l-p}^2\prod_{p'=1}^{l-1}V_{t-p'}^2$$

and in expectation we obtain a bounded term if $\mathbb{E}V_0^2 < 1$:

$$\mathbb{E}\left[\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\beta^{k-1}V_{t-l-k}\prod_{k'=1}^{l-1}V_{t-k'}\right]^2$$

$$= 4\times\frac{\mathbb{E}V_0^2}{1-\mathbb{E}V_0^2}\left[\frac{\beta}{(1-\beta)(1-\beta^2)}\frac{\mathbb{E}V_0}{1-\mathbb{E}V_0}-\frac{1}{(1-\beta)(1-\beta^2)}\frac{\mathbb{E}V_0\beta}{1-\beta^2\mathbb{E}V_0}\right]$$

$$+4\times\frac{\beta(\mathbb{E}V_0)^3}{(1-\beta)(1-\beta^2)(1-\beta^2\mathbb{E}V_0)}\frac{1}{1-\mathbb{E}V_0^2}$$

$$+2\times\frac{1}{1-\beta^2}\frac{\mathbb{E}V_0^2}{1-\mathbb{E}V_0^2}\left[\frac{\mathbb{E}V_0}{1-\mathbb{E}V_0}-\frac{\mathbb{E}V_0}{1-\beta^2\mathbb{E}V_0}\right]$$

$$+2\frac{1}{1-\mathbb{E}V_0^2}\frac{(\mathbb{E}V_0)^3}{(1-\beta^2)(1-\beta^2\mathbb{E}V_0)}$$

$$+2\frac{1}{1-\mathbb{E}V_0^2}(\mathbb{E}V_0)^2\frac{\beta}{(1-\beta)(1-\beta^2)}+\frac{\mathbb{E}V_0^2}{1-\mathbb{E}V_0^2}\frac{1}{1-\beta^2}.$$

That $\mathbf{B}_{33}$ is finite under $\mathbb{E}V_0^2<1$ comes from

$$\mathbf{B}_{33}=\mathbb{E}\left[\sum_{l=1}^{\infty}Z_{t-l}\prod_{k=1}^{l-1}V_{t-k}\right]^2$$

$$=2\mathbb{E}\sum_{l=1}^{\infty}\sum_{l'>l}^{\infty}Z_{t-l}Z_{t-l'}\prod_{k=1}^{l-1}V_{t-k}\prod_{k'=1}^{l'-1}V_{t-k'}+\mathbb{E}\sum_{l=1}^{\infty}Z_{t-l}^2(\prod_{k=1}^{l-1}V_{t-k})^2$$

$$=\mathbb{E}Z_0^2\sum_{l=1}^{\infty}(\mathbb{E}V_0^2)^{l-1}=\frac{\mathbb{E}Z_0^2}{1-\mathbb{E}V_0^2}.$$

That the last coefficient is also finite comes form the computation

$$\mathbf{B}_{44} = \mathbb{E}\left[\sum_{l=1}^{\infty} |Z_{t-l}| \prod_{k=1}^{l-1} V_{t-k}\right]^2 =$$

$$2\mathbb{E}\sum_{l=1}^{\infty}\sum_{l'>l}^{\infty} |Z_{t-l}|\,|Z_{t-l'}| \prod_{k=1}^{l-1} V_{t-k} \prod_{k'=1}^{l'-1} V_{t-k'} + \mathbb{E}\sum_{l=1}^{\infty} Z_{t-l}^2 (\prod_{k=1}^{l-1} V_{t-k})^2$$

$$= 2\sum_{l=1}^{\infty}\sum_{l'>l}^{\infty} \mathbb{E}\,|Z_{t-l'}|\,\mathbb{E}\left(\prod_{k=1}^{l-1} V_{t-k}^2\right)\mathbb{E}\left(|Z_{t-l}| \prod_{k'=l}^{l'-1} V_{t-k'}\right) + \frac{\mathbb{E}Z_0^2}{1-\mathbb{E}V_0^2}$$

$$= 2\sum_{l=1}^{\infty}\sum_{l'>l}^{\infty} (\mathbb{E}\,|Z_0|)(\mathbb{E}V_0^2)^{l-1}(\mathbb{E}\,|Z_0|\,V_0)\mathbb{E}V_0^{l'-l-1} + \frac{\mathbb{E}Z_0^2}{1-\mathbb{E}V_0^2}$$

$$= \frac{2\mathbb{E}\,|Z_0|\,(\mathbb{E}\,|Z_0|\,V_0)}{(1-\mathbb{E}V_0)(1-\mathbb{E}V_0^2)} + \frac{\mathbb{E}Z_0^2}{1-\mathbb{E}V_0^2}.$$

□

Appendix 2: checking the assumption (LI)

Let $x_0 \in \mathbb{R}^4$ be a vector such that $\nabla g_0(\theta_0)^T x_0 = 0$ a.s. Since $(\nabla g_t(\theta_0))_{t\in\mathbb{Z}}$ is stationary, then we also have $\nabla g_1(\theta_0)^T x_0 = 0$ a.s. We know that $\nabla g_1(\theta_0) = U_0(\theta_0) + V_0(\theta_0)\nabla g_0(\theta_0)$, then we deduce

$$U_0(\theta_0)^T x_0 = \begin{pmatrix} 1 \\ \log\sigma_0^2 \\ Z_0 \\ |Z_0| \end{pmatrix}^T x_0 = 0 \text{ a.s.}$$

which is impossible for $x_0 \neq 0$ if Z_0 is not concentrated on two points, see Lemma 8.2 of [17] for more details.

Appendix 3: exact computation of V

Remember that $\mathbf{V} = (\mathbb{E}Z_0^4 - 1)\mathbf{B}^{-1}$ with $\mathbf{B} = \mathbb{E}\nabla g_t(\theta_0)(\nabla g_t(\theta_0))^T$. According the notation of Appendix 1, denote $W_t = \gamma Z_t + \delta|Z_t|$, $U_t = (1, \log\sigma_t^2, Z_t, |Z_t|)$, $V_t = \beta - \frac{1}{2}(\gamma Z_t + \delta\,|Z_t|)$ and $\mathbb{E}\nabla g_t(\theta_0) = \mathbf{G}$. Remark that $\mathbb{E}W_t = \delta\mathbb{E}\,|Z_0|$, $\mathbb{E}W_t^2 = (\gamma^2+\delta^2)\mathbb{E}Z_0^2$, $\mathbb{E}V_t = \beta - \frac{1}{2}\delta\mathbb{E}\,|Z_t|$, $\mathbb{E}V_tZ_t = -\frac{1}{2}\gamma\mathbb{E}Z_0^2$, $\mathbb{E}V_t\,|Z_t| =$

$\beta\mathbb{E}\left|Z_t\right| - \frac{1}{2}\delta\mathbb{E}Z_0^2$, $\mathbb{E}V_t^2 = \beta^2 - \beta\delta\mathbb{E}\left|Z_t\right| + \frac{1}{4}(\gamma^2+\delta^2)\mathbb{E}Z_0^2$ and $\mathbb{E}\log\sigma_t^2 = \frac{\alpha+\delta\mathbb{E}|Z_0|}{1-\beta}$. We also have

$$\begin{aligned}\mathbb{E}(\log\sigma_t^2)^2 &= \mathrm{Var}\left(\log\sigma_t^2\right) + \left(\mathbb{E}\log\sigma_t^2\right)^2 \\ &= \sum_{k=1}^{\infty}\beta^{2(k-1)}\mathrm{Var}\left(W_{t-k}\right) + \frac{(\alpha+\delta\mathbb{E}|Z_0|)^2}{(1-\beta)^2} \\ &= \frac{\gamma^2+\delta^2-(\delta\mathbb{E}|Z_0|)^2}{1-\beta^2} + \frac{(\alpha+\delta\mathbb{E}|Z_0|)^2}{(1-\beta)^2}\end{aligned}$$

and $\mathbb{E}V_tU_t = (\mathbb{E}V_0, \mathbb{E}V_0\mathbb{E}\log\sigma_0^2, -\frac{1}{2}\gamma, \beta\mathbb{E}\left|Z_t\right| - \frac{1}{2})^T$. Since $\mathbb{E}\left\|\nabla g_t(\theta_0)\right\|_2^2 < \infty$ from Appendix 1, taking expectation on both side of the equation

$$\mathbf{G} = \mathbb{E}\nabla g_t(\theta_0) = \mathbb{E}U_{t-1} + \mathbb{E}V_{t-1}\mathbb{E}\nabla g_{t-1}(\theta_0) = \mathbb{E}U_{t-1} + \mathbb{E}V_{t-1}\mathbf{G}$$

so $\mathbf{G} = \mathbb{E}U_0/(1-\mathbb{E}V_0) = (1-\beta+\frac{1}{2}\delta\mathbb{E}|Z_0|)^{-1}(1, \mathbb{E}\log\sigma_t^2, 0, \mathbb{E}|Z_0|)^T$. Using again the SRE, we have

$$\begin{aligned}\nabla g_t(\theta_0)\left(\nabla g_t(\theta_0)\right)^T &= \left[U_{t-1} + V_{t-1}\nabla g_{t-1}(\theta_0)\right]\left[U_{t-1} + V_{t-1}\nabla g_t(\theta_0)\right]^T \\ &= U_{t-1}U_{t-1}^T + V_{t-1}\left[U_{t-1}(\nabla g_{t-1}(\theta_0))^T + \nabla g_{t-1}(\theta_0)U_{t-1}^T\right] \\ &\quad + V_{t-1}^2\left[\nabla g_{t-1}(\theta_0)\left(\nabla g_{t-1}(\theta_0)\right)^T\right]\end{aligned}$$

so

$$\mathbf{B} = \mathbb{E}U_{t-1}U_{t-1}^T + \mathbb{E}\left(V_{t-1}\left[U_{t-1}(\nabla g_{t-1}(\theta_0))^T + \nabla g_{t-1}(\theta_0)U_{t-1}^T\right]\right) + \mathbb{E}V_{t-1}^2\mathbf{B}$$

and $\mathbf{B} = (1-\mathbb{E}V_0^2)^{-1}(\mathbb{E}U_0U_0^T + \mathbf{F})$ where $\mathbf{F} = \mathbb{E}\left[V_0\left[U_0(\nabla g_0(\theta_0))^T + \nabla g_0(\theta_0)U_0^T\right]\right]$. As we have

$$\mathbb{E}V_tU_t(\nabla g_t(\theta_0))^T = \mathbb{E}\begin{pmatrix} V_t \\ V_t\log\sigma_t^2 \\ V_tZ_t \\ V_t\left|Z_t\right| \end{pmatrix}\left(\nabla g_t(\theta_0)\right)^T = \begin{pmatrix} \mathbb{E}V_t\mathbf{G}^T \\ \mathbb{E}V\mathbb{E}\log\sigma_t^2(\nabla g_{t-1}(\theta_0))^T \\ \mathbb{E}V_tZ_t\mathbf{G}^T \\ \mathbb{E}V_t\left|Z_t\right|\mathbf{G}^T \end{pmatrix}$$

it remains to calculate

$$\begin{aligned}\mathbb{E}\left[\nabla g_t(\theta_0)\log\sigma_t^2\right] &= \mathbb{E}\nabla g_t(\theta_0)\left(\frac{\alpha+2\beta}{1-\beta} - 2\sum_{k=1}^{\infty}\beta^{k-1}V_{t-k}\right) \\ &= \frac{\alpha+2\beta}{1-\beta}\mathbf{G} - 2\mathbb{E}\sum_{l=1}^{\infty}\left(U_{t-l}\prod_{k_1=1}^{l-1}V_{t-k_1}\right)\sum_{k_2=1}^{\infty}\beta^{k_2-1}V_{t-k_2}\end{aligned}$$

where

$$\mathbb{E}\sum_{l=1}^{\infty}\left(U_{t-l}\prod_{k_1=1}^{l-1}V_{t-k_1}\right)\sum_{k_2=1}^{\infty}\beta^{k_2-1}V_{t-k_2}$$

$$=\mathbb{E}U_{t-1}V_{t-1}+\mathbb{E}\sum_{l=2}^{\infty}\left(U_{t-l}\prod_{k_1=1}^{l-1}V_{t-k_1}\right)\sum_{k_2=1}^{l-1}\beta^{k_2-1}V_{t-k_2}$$

$$+\mathbb{E}\sum_{l=2}^{\infty}\left(U_{t-l}\prod_{k_1=1}^{l-1}V_{t-k_1}\right)\beta^{l-1}V_{t-l}+\mathbf{A}$$

$$=\sum_{l=2}^{\infty}\mathbb{E}U_0(\mathbb{E}V_0)^{l-2}\mathbb{E}V_0^2\frac{1-\beta^{l-1}}{1-\beta}+\sum_{l=1}^{\infty}\mathbb{E}V_0U_0(\mathbb{E}V_0)^{l-1}\beta^{l-1}+\mathbf{A}$$

$$=\frac{\mathbb{E}U_0\mathbb{E}V_0^2}{1-\beta}(\frac{1}{1-\mathbb{E}V_0}-\frac{\beta}{1-\beta\mathbb{E}V_0})+\frac{\mathbb{E}V_0U_0}{1-\beta\mathbb{E}V_0}+\mathbf{A}.$$

Now we treat the last term. Remark that U_t and V_{t-1} are independent except for their second coordinates, then for $j\neq 2$:

$$\mathbf{A}_j=\mathbb{E}\sum_{l=1}^{\infty}\left(U_{t-l,j}\prod_{k_1=1}^{l-1}V_{t-k_1}\right)\sum_{k_2=l+1}^{\infty}\beta^{k_2-1}V_{t-k_2}$$

$$=\sum_{l=1}^{\infty}\left(\mathbb{E}U_{0,j}(\mathbb{E}V_0)^{\ell-1}\right)\sum_{k_2=l+1}^{\infty}\beta^{k_2-1}\mathbb{E}V_0=\frac{\beta\mathbb{E}U_{0,j}\mathbb{E}V_0}{(1-\beta)(1-\beta\mathbb{E}V_0)}.$$

For $j=2$, we get

$$\mathbf{A}_2=\mathbb{E}\sum_{l=1}^{\infty}\left(\log\sigma_{t-l}^2\prod_{k_1=1}^{l-1}V_{t-k_1}\right)\sum_{k_2=l+1}^{\infty}\beta^{k_2-1}V_{t-k_2}$$

$$=\sum_{l=1}^{\infty}(\mathbb{E}V_0)^{\ell-1}\mathbb{E}\left(\log\sigma_{t-l}^2\sum_{k_2=l+1}^{\infty}\beta^{k_2-1}V_{t-k_2}\right).$$

Remembering that

$$\log\sigma_t^2=\frac{\alpha+2\beta}{1-\beta}-2\sum_{k=1}^{\infty}\beta^{k-1}V_{t-k}\Leftrightarrow\sum_{k_2=l+1}^{\infty}\beta^{k_2-1}V_{t-k_2}=\frac{\beta^l}{2}\left(\frac{\alpha+2\beta}{1-\beta}-\log\sigma_{t-l}^2\right)$$

we finally obtain

$$\mathbf{A}_2=\frac{\beta}{2(1-\beta EV_0)}\left(\frac{\alpha+2\beta}{1-\beta}\mathbb{E}\log\sigma_0^2-\mathbb{E}(\log\sigma_0^2)^2\right).$$