doi: 10.1111/j.1467-9469.2006.00495.x

Linear-representation Based Estimation of Stochastic Volatility Models

CHRISTIAN FRANCQ

GREMARS, Université Lille 3

JEAN-MICHEL ZAKOÏAN

CREST and GREMARS, Université Lille 3

ABSTRACT. A new way of estimating stochastic volatility models is developed. The method is based on the existence of autoregressive moving average (ARMA) representations for powers of the log-squared observations. These representations allow to build a criterion obtained by weighting the sums of squared innovations corresponding to the different ARMA models. The estimator obtained by minimizing the criterion with respect to the parameters of interest is shown to be consistent and asymptotically normal. Monte-Carlo experiments illustrate the finite sample properties of the estimator. The method has potential applications to other non-linear time-series models.

Key words: autoregressive moving average, conditional heteroskedasticity, consistency and asymptotic normality, non-linear least squares, stochastic volatility

1. Introduction

During the last 20 years, interest in models with stochastic variance has considerably increased in both the econometric and the finance literatures. The autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) proposed by Engle (1982) and Bollerslev (1986), have found widespread applications in macroeconomics and finance. This is due to their ability to capture many stylized facts of financial time series and to their tractability as far as inference and prediction are concerned. The maximum likelihood estimator is obtained straightforwardly and a sound asymptotic theory is now available in the literature (see e.g. Lee & Hansen, 1994; Lumsdaine, 1996; Linton, 1997; Berkes *et al.*, 2003; Francq & Zakoïan, 2004). However, the GARCH assumption that the volatility is driven by past observable variables can become a constraint. Consequently, a number of competing volatility models have been introduced. In particular, the so-called stochastic volatility (SV) models are alternatives to GARCH models that are built by analogy with the time-varying volatility in diffusions (see Hull & White, 1987; Melino & Turnbull, 1990; Hoffmann, 2002). In these models both the mean and volatility equations have separate error terms. The estimation of SV models is far from being straightforward as the volatility is a dynamic latent variable, which precludes standard estimation methods. Until the beginning of the 1990s, the literature on inference for SV models was brief.

The situation has tremendously changed in the recent years and there are currently various approaches to estimating SV models. Parameter estimation can be performed by the generalized method of moments (GMM) type of estimators (see e.g. Taylor, 1986; Andersen & Sørensen, 1996). Another popular approach is the so-called quasi-maximum likelihood (QML) which uses a linear state-space form of the model on which the usual Kalman filter is applied. This method has been advocated by Nelson (1988), Harvey *et al.* (1994) and Ruiz (1994) among others. An approach closely related to the QML uses a mixture of Gaussian distributions to approximate the non-Gaussian error in the observation equation of the state-space form (see Kim *et al.*, 1998; Mahieu & Schotman, 1998). Following Jacquier *et al.* (1994), the Bayesian approach has been adopted by several authors who have used different Markov-chain Monte Carlo (MCMC) algorithms to resolve the computational problem (see Chib *et al.*, 2002). Other papers have proposed to approximate the marginal likelihood of the observable process. The simulated maximum likelihood method relies on simulation of the latent volatility conditional on available information, see Danielsson (1994); another approach is to calculate the likelihood directly using recursive numerical integration procedures, as suggested by Fridman & Harris (1998). Comprehensive reviews on the estimation of SV models can be found in Taylor (1994), Ghysels *et al.* (1996) and Shephard (1996). Several of the papers just mentioned have attempted to compare the merits of some of the different approaches for estimating SV models. It is difficult to draw definitive conclusions because the (asymptotic) distributions of the estimators are, in general, not available. Moreover, efficiency is not the only basis to gauge the different procedures. Efficiency gains are generally obtained at the cost of important computational burdens. Finally, robustness to the distributional assumptions should be an important aspect in comparing the different estimation methods.

In this paper, we propose a new method for estimating SV models. In carrying out this work, our main aim was to derive a method which would be simple to implement, flexible with respect to the distributional assumptions and for which a sound asymptotic theory could be established. The proposed estimator meets these requirements and, fortunately, it appears to work quite well in simulation experiments. Our approach is based on the existence of autoregressive moving average (ARMA) representations for powers of the log-squared observed process. We propose an estimation procedure consisting in the maximization, with respect to the SV parameter, of a sum of weighted ARMA squared residuals.

This paper is organized as follows. Section 2 introduces the canonical SV model and discusses stationarity and identifiability issues. Section 3 establishes the existence of ARMA representations for powers of the log-squared observations. Section 4 presents estimators based on one or several representations and derives their asymptotic properties. In section 5, two technical constraints (compactness of the parameter space and knowledge of moments for the latent processes) are relaxed for the estimator based on one representation. Results from sampling experiments are presented in section 6. Conclusions are given in section 7. The proofs are relegated to an appendix.

2. The canonical SV model

2.1. Definition

Consider the process (ε_t) satisfying the simple SV model: $\forall t \in \mathbb{Z}$,

$$\begin{cases} \varepsilon_t = \sqrt{h_t}\eta_t \\ \ln h_t = \omega + \beta \ln h_{t-1} + \sigma v_t, \end{cases} \tag{1}$$

where ω, β, σ are real coefficients, (η_t) and (v_t) are two independent sequences of independent and identically distributed (i.i.d.) random variables with zero mean, $\mathrm{var}(v_t) = 1$. Model (1) is the canonical SV model and has been considered by most of the authors mentioned in the introduction, among many others. Under additional normality assumptions on the two noises, researchers have found this model useful to approximate stochastic diffusion models. However, at this point, we do not need to make this assumption. In this model, the parameter β plays the role of the persistence coefficient in the standard GARCH(1,1) model, while σ can be interpreted as the volatility of the unobserved log-volatility h_t and ω as a constant scaling factor (see Kim *et al.*, 1998).

Assume that $P(\eta_t = 0) = 0$, which is implied by assumption A2 below. By taking the logarithm of the square of ε_t, the first equation in (1) becomes:

$$\ln \varepsilon_t^2 = \ln h_t + \ln \eta_t^2 \tag{2}$$

as suggested by Nelson (1988) among others.

Remark 1. This transformation entails no information loss when the distribution of η_t is symmetric. Indeed, in this case we have, for any $x \geq 0$:

$$\begin{aligned} P(\varepsilon_t \leq x) &= P(h_t\eta_t^2 \leq x^2, \eta_t > 0) + P(\eta_t < 0) \\ &= E\{P(h_t\eta_t^2 \leq x^2, \eta_t > 0 \mid h_t)\} + P(\eta_t < 0) \\ &= E\{P(h_t\eta_t^2 \leq x^2 \mid \eta_t > 0, h_t)P(\eta_t > 0 \mid h_t)\} + P(\eta_t < 0) \\ &= E\left\{P(h_t\eta_t^2 \leq x^2 \mid h_t)\tfrac{1}{2}\right\} + \tfrac{1}{2} = \tfrac{1}{2}P(\ln \varepsilon_t^2 \leq \ln x^2) + \tfrac{1}{2} \end{aligned}$$

where the fourth equality follows from the independence between η_t and (h_t) and from the symmetry of the law of η_t. The same arguments show that

$$P(\varepsilon_t \leq -x) = \tfrac{1}{2}P(\ln \varepsilon_t^2 \geq \ln x^2),$$

for $x \geq 0$. Thus there is a one-to-one relation between the law of ε_t and that of $\ln \varepsilon_t^2$. In addition the law of ϵ_t is symmetric. Similarly, for any $n \geq 1$, one can show that there is a one-to-one relation between the law of $(\varepsilon_t, \ldots, \varepsilon_{t+n})$ and that of $(\ln \varepsilon_t^2, \ldots, \ln \varepsilon_{t+n}^2)$. When the distribution of η_t is not symmetric, the fourth equality of the previous computation fails.

Let $X_t = \ln \varepsilon_t^2$, $Y_t = \ln h_t$, $Z_t = X_t - Y_t = \ln \eta_t^2$. For any second-order stationary process (W_t) let $\mu_W = E(W_t)$ and $\sigma_W^2 = \mathrm{var}(W_t)$.

2.2. *Stationarity and identifiability assumptions*

It is clear from the multiplicative form of first equation in model (1) that an identifiability condition is required. As in GARCH models we could set $\mathrm{var}(\eta_t) = 1$, without loss of generality. As we are going to work with $Z_t = \ln \eta_t^2$, it is more convenient, and no more restrictive, to set an arbitrary value μ for EZ_t. Throughout this paper, we make the following assumptions.

$$\mathrm{A1}: |\beta| < 1, \quad A2: \sigma > 0, \quad Z_t \in L^2 \quad \text{and} \quad EZ_t = \mu.$$

The scaling constant μ can be chosen equal to 0, e.g. or to $-1.27 = E \ln u^2$, where $u \sim \mathcal{N}(0, 1)$. Assumption A1 ensures the existence of a strictly stationary and ergodic solution to model (1) given by:

$$\varepsilon_t = \exp\left\{\frac{\omega}{2(1-\beta)} + \frac{\sigma}{2}\sum_{i=0}^{\infty}\beta^i v_{t-i}\right\}\eta_t, \quad t \in \mathbb{Z}. \tag{3}$$

Assumption A2 is made for identifiability purposes. First, notice that if $\sigma = 0$, (3) reduces to $\varepsilon_t = \exp\left\{\frac{\omega}{2(1-\beta)}\right\}\eta_t$, so that β and ω are not identifiable. Secondly, the other parts of assumption A2 allow us to identify ω, β and, if $\beta \neq 0$, σ^2 and $\sigma_Z^2 = \mathrm{var}(Z_t)$, from the second-order structure of (X_t). Indeed, as (Y_t) satisfies the strong AR(1) model

$$Y_t = \omega + \beta Y_{t-1} + \sigma v_t \tag{4}$$

and from the independence between (Y_t) and (Z_t), (X_t) is a second-order stationary process with mean and autocovariance function given by

$$\mu_X = \frac{\omega}{1-\beta} + \mu, \quad \gamma_X(0) = \frac{\sigma^2}{1-\beta^2} + \sigma_Z^2, \quad \gamma_X(k) = \beta^k \frac{\sigma^2}{1-\beta^2}, \quad k > 0. \tag{5}$$

When $\gamma_X(1) \neq 0$, the parameter β is identified from $\gamma_X(2)/\gamma_X(1)$. Then we deduce σ^2 from $\gamma_X(1)$, and σ_Z^2 from $\gamma_X(0)$. When $\gamma_X(1) = 0$ we have $\beta = 0$ and σ and σ_Z are not both identifiable. Finally, in any case, ω is obtained from μ_X, as μ is fixed.

3. ARMA representations

Based on (2), ARMA representations can be derived for the powers of $\ln \varepsilon_t^2$, which we now examine. For ease of presentation, we first consider a special case.

3.1. For $\ln \varepsilon_t^2$

As, from (5), $\gamma_X(k) = \beta\gamma_X(k-1), \quad \forall k > 1$, (X_t) satisfies an ARMA(1,1) representation of the form (see lemma 1 below)

$$X_t - \mu_X = \beta(X_{t-1} - \mu_X) + u_t - \alpha u_{t-1}, \tag{6}$$

where (u_t) is a white noise with variance σ_u^2, and $|\alpha| < 1$. When $\beta \neq 0$, we have, by (5), $\gamma_X(1) \neq \beta\gamma_X(0)$ because $\sigma_Z^2 > 0$, so X_t is not an AR(1), i.e. $\alpha \neq 0$. By (6) we obtain

$$\gamma_X(0) = \beta\gamma_X(1) + \sigma_u^2\{1 + \alpha(\alpha - \beta)\} \quad \text{and} \quad \gamma_X(1) = -\alpha\sigma_u^2 + \beta\gamma_X(0).$$

Therefore, we have

$$\frac{1 + \alpha(\alpha - \beta)}{\alpha} = \frac{\gamma_X(0) - \beta\gamma_X(1)}{\beta\gamma_X(0) - \gamma_X(1)} = \frac{\sigma^2 + \sigma_Z^2}{\beta\sigma_Z^2} \tag{7}$$

and the solution belonging to the unit disk is given by

$$\alpha = \alpha(\beta, \sigma; \sigma_Z) = \frac{(1+\beta^2)\sigma_Z^2 + \sigma^2 - \{(1+\beta)^2\sigma_Z^2 + \sigma^2\}^{1/2}\{(1-\beta)^2\sigma_Z^2 + \sigma^2\}^{1/2}}{2\beta\sigma_Z^2}. \tag{8}$$

Remark 2. If $\beta = 0$, then (X_t) is actually an i.i.d. process, so it is a white noise and we have $\alpha = 0$. Notice that $\alpha(\beta, \sigma; \sigma_Z)$ is continuous because $\lim_{\beta \to 0} \alpha(\beta, \sigma; \sigma_Z) = 0$.

Remark 3. If $\beta \neq 0$ and $\sigma \neq 0$ we get $(\alpha - \beta)(\beta - 1/\alpha) = \sigma^2/\sigma_Z^2 > 0$, using (7). Thus, either $0 < \alpha < \beta < 1/\alpha$ or $0 > \alpha > \beta > 1/\alpha$. In particular $|\alpha| < |\beta|$, which shows that X_t is an exact ARMA(1,1).

Example 1. If (η_t) is $\mathcal{N}(0,1)$ distributed, $\mu_Z = -1.270 (= \mu)$, $\sigma_Z^2 = \pi^2/2 = 4.935$, $\sigma_{Z^2}^2 = 263.484$ (Abramovitz & Stegun, 1970, pp. 260 and 943). Moreover, if (η_t) is $\mathcal{N}(0,1)$ distributed and if $\omega = -1$, $\beta = 0.9$, $\sigma = 0.4$, then (X_t) satisfies the following ARMA(1,1) representation

$$(X_t - 11.27) - 0.9(X_{t-1} - 11.27) = u_t - 0.81u_{t-1},$$

where (u_t) is a white noise with variance $\sigma_u^2 = 5.52$.

The existence of the ARMA(1,1) representation for the log-squared observable process is well known; for example, it was mentioned by Ruiz (1994) and it was used by Breidt & Carriquiry (1996) to obtain QML estimates. However, some care should be exercised in interpreting this ARMA representation as the white noise (u_t) is only a sequence of uncorrelated variables. To see why (u_t) is not a martingale difference, it is sufficient to compute, e.g.

$$E\{u_t(X_{t-1}-\mu_X)^2\} = \frac{\sigma^3\alpha\{1+\beta^2(1-\alpha)\}}{(1-\beta^3)(1-\alpha\beta^2)}E(v_t^3)+(\alpha-\beta)E(Z_t-\mu_Z)^3$$

which is not equal to zero in general, even when the distribution of (v_t) is symmetric. Details on the computation can be provided by the authors.

3.2. For powers of $\ln \varepsilon_t^2$

In this section, we will show that ARMA models also hold for vectors of powers of X_t. The inferential consequences of this result will be clear in section 4.

We will use the following characterization of ARMA representations through the autocovariance function. This result can be proved by a straightforward adaptation of Brockwell & Davis (1991, Proposition 3.2.1 and Remark p. 90).

Lemma 1
If (W_t) *is any vector-valued stationary process, with autocovariance function* $\Gamma_W(\cdot)$ *satisfying an equation of the form:* $\Gamma_W(h)+\sum_{i=1}^{p_0}\Phi_i\Gamma_W(h-i)=0, \quad \forall h>q_0$, *then* (W_t) *is a vector* $ARMA(p_0,q_0)$ *process, i.e.*

$$W_t+\sum_{i=1}^{p_0}\Phi_i W_{t-i} = C+u_t+\sum_{i=1}^{q_0}\Psi_i u_{t-i},$$

where (u_t) *is a white noise process, which is not independent in general,* p_0, q_0 *are integers,* C *is a constant vector and* $\Phi_1,\ldots,\Phi_{p_0},\Psi_1,\ldots,\Psi_{q_0}$ *are constant matrices.*

Let M denote any positive integer and suppose that v_t and Z_t admit moments up to order $2M$. Let $\mathbf{X}_t^{(M)}=(X_t,\ldots,X_t^M)'$ denote the vector-valued process consisting of the first M powers X_t^m for $1\le m\le M$. The key result of this paper is the following.

Theorem 1
Let (ε_t) *be a solution of model (1) with* $\sigma^2_{Z^M}<\infty$ *and* $\sigma^2_{v^M}<\infty$. *Then*

(i) $(\mathbf{X}_t^{(M)})$ *is solution of an* $ARMA(M,M)$ *equation of the form*

$$\mathbf{X}_t^{(M)}-\mu_{\mathbf{X}^{(M)}}-\sum_{i=1}^{M}\Phi_i^{(M)}(\mathbf{X}_{t-i}^{(M)}-\mu_{\mathbf{X}^{(M)}})=\mathbf{u}_t^{(M)}-\sum_{i=1}^{M}\Psi_i^{(M)}\mathbf{u}_{t-i}^{(M)}, \tag{9}$$

where $(\mathbf{u}_t^{(M)})$ *is a white noise,* $\mu_{\mathbf{X}^{(M)}}$ *is a vector and* $(\Phi_i^{(M)})$, $(\Psi_i^{(M)})$ *are sequences of* $M\times M$ *matrices.*

(ii) For any $m\in\{1,\ldots,M\}$, (X_t^m) *is solution of an* $ARMA(m,m)$ *equation of the form*

$$\prod_{l=1}^{m}(1-\beta^l L)(X_t^m-\mu_{X^m})=u_t^{(m)}-\sum_{i=1}^{m}\alpha_i^{(m)}u_{t-i}^{(m)}, \tag{10}$$

where L *is the lag operator and* $(u_t^{(m)})$ *is a white noise.*

Remark 4. It should be noted that if (η_t) and (v_t) have the $\mathcal{N}(0,1)$ distribution, which is a standard assumption, the moment conditions are met for all m. Therefore, the ARMA representations hold for any power M.

Remark 5. The coefficients μ_{X^m} and $\alpha_i^{(m)}$ in (10) can be straightforwardly, albeit tediously obtained. We have

$$\mu_{X^m} = \sum_{i=0}^{m} \binom{m}{i} E(Y_t^i) E(Z_t^{m-i})$$

and the moments of (Y_t) can be derived recursively by

$$E(Y_t^i) = \sum_{l=0}^{i} \binom{i}{l} \beta^l E(Y_t^l) E(\omega + \sigma v_t)^{i-l}.$$

Similarly, the $\alpha_i^{(m)}$s are functions of the first m autocorrelations of (X_t^m), which can be derived from (22) to (24).

Example 2. (Example 1 continued): with the same model as in section 3.1, the squared process, (X_t^2) satisfies the following ARMA(2,2) representation

$$(X_t^2 - 132.80) - 1.71(X_{t-1}^2 - 132.80) + 0.73(X_{t-2}^2 - 132.80) = u_t^{(2)} - 1.63u_{t-1}^{(2)} + 0.67u_{t-2}^{(2)},$$

where $(u_t^{(2)})$ is a white noise with variance 3744.79.

4. Estimation

This section introduces an estimation method based on the ARMA representations. We start by defining estimators using the univariate ARMA representations of (X_t^m) given in part (ii) of theorem 1, because they are easier to handle numerically. Then we define estimators based on the vector representation of part (i) of this theorem.

4.1. Linear representation and Vectorial linear representation

We denote by $\Theta := \mathbb{R} \times]-1, 1[\times]0, +\infty[$ the space of the parameters $\theta = (\omega, \beta, \sigma)$. We have shown that, under suitable moment conditions, for all positive integer m, (X_t^m) satisfies an ARMA (m, m) representation given by (10). For ease of exposition, we shall first assume that the moments of $\ln \eta_t^2$ and v_t are known up to the order $2m$. Let $u_t^{(m)} = u_t^{(m)}(\theta)$ and $u_t = u_t^{(1)}$. Provided the MA part in (10) has no unit root, which is the case for $m = 1$ by remark 3, we have a unique AR(∞) representation of the form

$$u_t^{(m)}(\theta) = c^{(m)}(\theta) + \sum_{i=0}^{\infty} b_i^{(m)}(\theta) X_{t-i}^m, \tag{11}$$

where $b_0^{(m)}(\theta) = 1$ and $\sum_{i=0}^{\infty} |b_i^{(m)}(\theta)| < \infty$. For instance, when $m = 1$ we have

$$c^{(1)}(\theta) = -\frac{\omega + (1-\beta)\mu}{1-\alpha} \quad \text{and,} \quad \text{for } i \geq 0, \quad b_{i+1}^{(1)}(\theta) = (\alpha - \beta)\alpha^i, \tag{12}$$

where α is given by (8). Let $X_1, \ldots, X_n$ be a realization of length n of (X_t) and let $\theta_0 = (\omega_0, \beta_0, \sigma_0)$ be the true value of the parameter. We assume that A1 and A2 are satisfied when $\beta = \beta_0$ and $\sigma = \sigma_0$. For $0 < t \leq n$, $u_t^{(m)}(\theta)$ is approximated by $\breve{u}_t^{(m)}(\theta)$, obtained by replacing the unknown starting values by zero, i.e.

$$\breve{u}_t^{(m)}(\theta) = c^{(m)}(\theta) + \sum_{i=0}^{t-1} b_i^{(m)}(\theta) X_{t-i}^m, \quad t = 1, \ldots, n. \tag{13}$$

This choice of the initial values is made for simplicity, but in finite samples a more sensible choice may be $E_\theta X_t^m$. Let Θ_K be a compact set of the form

$$\Theta_K = [-K, +K] \times [-1 + K^{-1}, 1 - K^{-1}] \times [0, K],$$

where $K > 1$. For some strictly positive integer M, we define a *linear representation* (LR) based estimator of θ as any solution of

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta_K} Q_n(\theta), \quad \text{where} \quad Q_n(\theta) = \sum_{m=1}^{M} \tau_n^{(m)} \left\{ \frac{1}{n} \sum_{t=1}^{n} [\breve{u}_t^{(m)}(\theta)]^2 \right\} \tag{14}$$

and $(\tau_n^{(m)})$ is a sequence of random weights, independent of θ, and almost surely converging to $\tau^{(m)} \geq 0$ as $n \to \infty$. The limiting weights $\tau^{(m)}$ are assumed to be non-random, and such that $\tau^{(1)} = 1$. We also assume that $\tau^{(m)} = 0$ whenever the MA part in (10) has a unit root, for some $\theta \in \Theta$. Recall that, by remark 3, this does not occur for $m = 1$. This assumption provides against numerical troubles in the computation of the $\breve{u}_t^{(m)}(\theta)$.

Remark 6. This definition involves conditional least squares criteria, generally used to estimate ARMA coefficients. However, it is important to note that the minimization is performed over the parameters (ω, β, σ) of model (1), and not over the ARMA parameters as is usually carried out.

Remark 7. The proposed estimator belongs to the general class of M-estimators. It presents some common features with GMM estimators, in the sense that several criteria are averaged by means of weights. It is necessary to introduce such weights, because the magnitudes of the $\sum_{t=1}^{n} [\breve{u}_t^{(m)}(\theta)]^2$ can be very different when m varies: without the correction brought by the weights, the estimator would be mainly determined by the sums of larger module.

Remark 8. The condition $\tau^{(1)} = 1$ allows to obtain a consistent estimator even if $M = 1$. We should like to emphasize that the LR estimator for $M = 1$ is also obtained for $M > 1$, by choosing $\tau_n^{(m)} = 0$ for $m = 2, \ldots, M$. However, a better asymptotic accuracy may be obtained for other choices of the weights. See remark 10 below and section 4.3.1 for further discussion on the choice of the weights.

Now consider the LR estimators based on the vectorial ARMA representations. We have shown that, under suitable moment conditions, $\mathbf{X}_t^{(M)} = (X_t, \ldots, X_t^M)'$ admits a weak ARMA (M, M) representation given by (9). For all $\theta \in \Theta$, write $\mathbf{u}_t^{(M)}(\theta) = \mathbf{u}_t^{(M)}$ for all $t \in \mathbb{Z}$, and

$$\breve{\mathbf{u}}_t^{(M)}(\theta) = \mathbf{X}_t^{(M)} - \mu_{\mathbf{X}^{(m)}} - \sum_{i=1}^{M} \Phi_i^{(M)} (\mathbf{X}_{t-i}^{(M)} - \mu_{\mathbf{X}^{(M)}}) + \sum_{i=1}^{M} \Psi_i^{(M)} \breve{\mathbf{u}}_{t-i}^{(m)}(\theta)$$

for all $t = 1, \ldots, n$, where $\breve{u}_t^{(m)}(\theta) = \mathbf{X}_t^{(M)} - \mu_{\mathbf{X}^{(m)}}(\theta) = 0$ for all $t \leq 0$. We define the vectorial linear representation (VLR) based estimator of θ as any solution of

$$\hat{\theta}_n^V = \arg\min_{\theta \in \Theta_K} Q_n^V(\theta), \quad \text{where} \quad Q_n^V(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left(\breve{\mathbf{u}}_t^{(M)}(\theta) \right)' \mathbf{S}_n \left(\breve{\mathbf{u}}_t^{(M)}(\theta) \right) \tag{15}$$

and $(\mathbf{S}_n)$ is a sequence of random matrices independent of θ, which converges, almost surely, to a symmetric positive definite constant matrix $\mathbf{S}$ as $n \to \infty$.

4.2. Consistency and asymptotic distribution

Let $\overline{X}_n$ be the sample mean of $X_1, \ldots, X_n$ and let $\hat{\gamma}_X(k)$ be the sample autocovariance at lag k. We have the following consistency result.

Theorem 2

Assume that the moments of $Z_t = \ln \eta_t^2$ and v_t exist and are known up to order $2M$. Let $\hat{\theta}_n$ be a sequence of LR estimators of θ satisfying (14). Then for any sufficiently large K such that $\theta_0 \in \Theta_K$,

(i) if $\beta_0 \neq 0$, $\quad \hat{\theta}_n \to \theta_0$ $\quad$ almost surely as $\quad n \to \infty$;
(ii) if $\beta_0 = 0$, $\quad (\hat{\omega}_n, \hat{\beta}_n) \to (\omega_0, \beta_0)$ $\quad$ almost surely as $\quad n \to \infty$.

In both cases, a strongly consistent estimator of σ_0^2 is $\tilde{\sigma}_n^2 = (1 - \hat{\beta}_n^2)(\hat{\gamma}_X(0) - \sigma_Z^2)$.

Remark 9. The assumption that certain moments of the processes $(\ln \eta_t^2)$ and (v_t) are known may be found restrictive. However, it is not a particularity of the method of this paper. Indeed, papers dealing with the estimation of SV models generally make stronger assumptions, most of them assuming the standard normality of the processes (η_t) and (v_t) (e.g. see the references cited in the introduction). In section 5.2 below we will obtain results in cases where the moments of $(\ln \eta_t^2)$ and (v_t) are unknown (except those fixed for identifiability).

We now turn to the asymptotic distribution of the LR estimator $\hat{\theta}_n$ defined by (14).

Theorem 3

Assume that $\beta_0 \neq 0$, that the assumptions of theorem 2 hold, and that $\sigma^2_{Z^{2M}} < \infty$ and $\sigma^2_{v^{2M}} < \infty$. Then for any sufficiently large K such that θ_0 belong to the interior of Θ_K, the limiting distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ is a centred normal with covariance matrix

$$\Sigma(\tau) = \left(\sum_{m=1}^{M} \tau^{(m)} J^{(m)} \right)^{-1} \left(\sum_{m=1}^{M} \sum_{\ell=1}^{M} \tau^{(m)} \tau^{(\ell)} I^{(m,\ell)} \right) \left(\sum_{m=1}^{M} \tau^{(m)} J^{(m)} \right)^{-1},$$

where $\tau = (\tau^{(1)}, \ldots, \tau^{(m)})$ and

$$J^{(m)} = E_{\theta_0} \left(\frac{\partial u_t^{(m)}}{\partial \theta} \frac{\partial u_t^{(m)}}{\partial \theta'} (\theta_0) \right), \qquad I^{(m,\ell)} = \sum_{h=-\infty}^{+\infty} E_{\theta_0} \left(u_t^{(m)} u_{t+h}^{(\ell)} \frac{\partial u_t^{(m)}}{\partial \theta} \frac{\partial u_{t+h}^{(\ell)}}{\partial \theta'} (\theta_0) \right).$$

The asymptotic properties of the VLR estimator defined by (15) are given in the following result.

Theorem 4

Let the assumptions of theorem 2 hold, and assume that for all $\theta \in \Theta_K$ the MA part in (9) is invertible. Then, almost surely, $\hat{\theta}_n^V \to \theta_0$, as $n \to \infty$. Moreover, if $E(\ln \eta_t^2)^{4M} < \infty$ and $E(v_t^{4M}) < \infty$ then $\sqrt{n}\left(\hat{\theta}_n^V - \theta_0\right)$ has a limiting centred normal distribution with covariance matrix $\Sigma^V = \left(J^V\right)^{-1} I^V \left(J^V\right)^{-1}$, where

$$J^V = E_{\theta_0} \left(\frac{\partial \mathbf{u}_t^{(M)}(\theta_0)'}{\partial \theta} \mathbf{S} \frac{\partial \mathbf{u}_t^{(M)}(\theta_0)}{\partial \theta'} \right)$$

and

$$I^V = \sum_{h=-\infty}^{+\infty} E_{\theta_0}\left(\frac{\partial \mathbf{u}_t^{(M)}(\theta_0)'}{\partial\theta}\mathbf{S}\mathbf{u}_t^{(M)}(\theta_0)\mathbf{u}_{t-h}^{(M)}(\theta_0)'\mathbf{S}\frac{\partial \mathbf{u}_{t-h}^{(M)}(\theta_0)}{\partial\theta'}\right).$$

Remark 10. It is worth noticing that the matrices $J^{(m)}$ and $I^{(m,\ell)}$ in theorem 3 do not depend on the asymptotic weights $\tau^{(m)}$. Suppose that we are interested in estimating $\lambda'\theta$, where λ is a given three-dimensional vector. An optimal vector of asymptotic weights is any solution of $\tau^* = \arg\min_\tau \lambda'\Sigma(\tau)\lambda$. This solution, which may not be unique, and may not exist unless the minimization is made over a compact set, is a function of the sole matrices $J^{(m)}$ and $I^{(m,\ell)}$. Practical implementation of this solution requires, of course, estimating these matrices, which will be briefly discussed in the next section. By contrast, the matrices J^V and I^V of theorem 4 depend on S, and the search of an optimal matrix S is a more difficult task.

Remark 11. An interesting feature of the LR estimator when $M=1$, is that its asymptotic covariance matrix $\Sigma = \left(J^{(1)}\right)^{-1} I^{(1,1)} \left(J^{(1)}\right)^{-1}$ can be numerically computed for each value of θ_0. Matrix $J^{(1)}$ can be derived from AR(∞) expansions of the derivatives of $u_t^{(1)}$. For the sake of brevity, we do not present here the method to evaluate $I^{(1,1)}$, but it is available on request. A consistent estimator of Σ is then obtained by replacing θ_0 by $\hat\theta_n$ in the formulas of $I^{(1,1)}$ and $J^{(1)}$.

4.3. Interest of using several representations

One could question the usefulness, for statistical inference, of considering ARMA representations for X_t^M with $M>1$. Two main reasons can be advanced.

4.3.1. Efficiency gains

By contrast with the classical method of moments, the GMM uses over-identifying restrictions (i.e. more moments than required for identifiability) to obtain some efficiency gains. Similarly, using more ARMA representations than required for identifiability may reduce the mean squared error of the estimator.

In view of remark 10, estimating optimal weights amounts to estimating $J^{(m)}$ and $I^{(m,\ell)}$, for $\ell, m = 1,\ldots,M$. This can be performed on the basis of a first-step consistent estimator $\hat\theta_n$, for instance the LR estimator with $M=1$ (and $\tau_n^{(1)} = \tau^{(1)} = 1$). At least numerically, we are able to compute the derivatives $\partial u_t^{(m)}/\partial\theta$ at $(\hat\theta_n)$. A natural estimator of $J^{(m)}$ is the empirical moment

$$\hat J^{(m)} = n^{-1}\sum_t \left\{\frac{\partial u_t^{(m)}}{\partial\theta}\right\}\left\{\frac{\partial u_t^{(m)}}{\partial\theta'}\right\}.$$

The estimation of $I^{(m,\ell)}$ requires more sophisticated methods, such as the so-called HAC procedure (see e.g. Andrews, 1991).

4.3.2. Testing for the adequacy of the SV model

There are many types of parametric volatility models following the seminal work by Engle (1982) and these models are prone to mis-specification. An interesting particularity of the ARMA(m,m) representation (10) for X_t^{2m}, $m>1$, is that the roots of the AR polynomial are β^{-i} for $i=1,\ldots,m$. This property can be exploited to assess the validity of the SV model for a given series. In particular, a very simple test can be built from the ARMA(2,2) representation of X_t^2. The AR polynomial of this representation is $(1-\beta L)((1-\beta^2 L)=1-aL+bL^2$, with $H_0: a=b^{1/3}+b^{2/3}$. Let $(\hat{a},\hat{b})$ be the least squares estimator of (a,b) [i.e. the AR coefficients in the least squares estimator of the ARMA(2,2) parameters]. It can be shown that $\sqrt{n}\{(\hat{a},\hat{b})'-(a,b)\}$ converges in distribution to a centred Gaussian law with covariance matrix $\underline{\Sigma}$. Let $\hat{\underline{\Sigma}}$ be a consistent estimate of this invertible matrix and let the Jacobian vector

$$J(a,b)=\left(1,\ -\frac{1}{3}b^{-2/3}-\frac{2}{3}b^{-1/3}\right)'.$$

A straightforward application of the delta method shows that, under H_0 the statistic

$$S_n:=n(\hat{a}-\hat{b}^{1/3}-\hat{b}^{2/3})^2\left\{J'(\hat{a},\hat{b})\hat{\underline{\Sigma}}J(\hat{a},\hat{b})\right\}^{-1}$$

has an asymptotic chi-square distribution with 1 d.f., denoted χ_1^2. Therefore, the SV model can be rejected at the asymptotic level α if $S_n>\chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the χ_1^2 distribution.

5. Extensions

In this section we show that the assumptions of theorem 2 can be weakened.

5.1. Non-compact parameter space

As the objective function $Q_n(\theta)$ is differentiable over Θ, theorem 2 entails that, for $\beta_0\neq 0$ and K large enough, the LR estimator over the compact set Θ_K is asymptotically a solution of the so-called estimating equations

$$\frac{\partial}{\partial\theta}Q_n(\theta)=0,\quad \theta\in\Theta. \tag{16}$$

It should be noted that in (16) the parameter θ is no longer supposed to be in a compact subset of Θ. In view of the previous remarks we have the following result.

Corollary 1

When $\beta_0\neq 0$, and under the assumptions of theorem 2, there exists a sequence of roots $\hat{\theta}_n$ of (16) such that $\hat{\theta}_n\to\theta_0$ a.s. as $n\to\infty$.

Theorem 2 and corollary 1 do not guarantee the consistency of the LR estimator when it is computed over the whole parameter space Θ. Relaxing the compactness assumption of the parameter space is generally a difficult task (see Potscher & Prucha, 1997). This can be carried out in our framework, at least when $M=1$, by introducing a suitable compactification of the parameter set.

By remark 3 $|\alpha|<1$ for all $(\beta,\sigma)\in[-1,1]\times]0,\infty[$ and for all $(\beta,\sigma)\in]-1,1[\times[0,\infty[$. Therefore, using (6), (11) and (12), the process $\{u_t(\theta)\}_t$ is well defined by

$$\begin{aligned}u_t(\theta)=u_t^{(1)}(\theta)&=(1-\alpha L)^{-1}\{X_t-\beta X_{t-1}-\omega-(1-\beta)\mu\}\\&=-\frac{\omega+(1-\beta)\mu}{1-\alpha}+X_t+\sum_{i=1}^{\infty}\alpha^i(\alpha-\beta)X_{t-i}\end{aligned}$$

for any

$$\theta=(\omega,\beta,\sigma)\in\{\mathbb{R}\times[-1,1]\times]0,+\infty[\}\cup\{\mathbb{R}\times]-1,1[\times[0,+\infty[\}.$$

From (8), we have $\alpha=\beta$ when $\beta=\pm 1$ and $\sigma=0$. Thus we set

$$u_t(\theta)=X_t-\omega/2-\mu \quad \text{for} \quad (\omega,\beta,\sigma)=(\omega,-1,0) \quad \text{with} \quad \omega\in\mathbb{R},$$
$$u_t(\theta)=X_t-\mu \quad \text{for} \quad (\omega,\beta,\sigma)=(0,1,0),$$
$$u_t(\theta)=+\infty \quad \text{for} \quad (\omega,\beta,\sigma)=(\omega,1,0) \quad \text{with} \quad \omega<0,$$
$$u_t(\theta)=-\infty \quad \text{for} \quad (\omega,\beta,\sigma)=(\omega,1,0) \quad \text{with} \quad \omega>0.$$

Therefore, in view of Billingsley (1995, Theorem 36.4), the process $\{u_t(\theta)\}_t$, valued in the extended real line $\overline{\mathbb{R}}=[-\infty,+\infty]$, is ergodic and stationary for any

$$\theta=(\omega,\beta,\sigma)\in\mathbb{R}\times[-1,1]\times[0,+\infty[.$$

Recall however that, for the existence of the stationary solution (3) it is essential that $|\beta_0|<1$, and that $\sigma_0>0$ is required for identifiability reasons. Note also that $u_t(\theta)$ is not continuous at the point $\theta=(0,1,0)$, but is continuous elsewhere.

Let $\underline{\omega}=\arctan\omega$, $\underline{\sigma}=\arctan\sigma$, $\underline{\theta}=(\underline{\omega},\beta,\underline{\sigma})$ and the one-to-one map $T:\mathbb{R}\times[-1,1]\times[0,+\infty[\to]-\pi/2,\pi/2[\times[-1,1]\times[0,\pi/2[$ such that $T(\theta)=\underline{\theta}$. We can write

$$\underline{u}_t(\underline{\theta})=u_t\{T^{-1}(\underline{\theta})\}, \quad \breve{\underline{u}}_t(\underline{\theta})=\breve{u}_t^{(1)}\{T^{-1}(\underline{\theta})\}.$$

We will now extend the definition domain of $\underline{u}_t(\underline{\theta})$. In view of (7), we have $\alpha\to 0$ when $\sigma\to\infty$. Therefore, we set

$$\underline{u}_t(\underline{\omega},\beta,\pi/2)=X_t-\beta X_{t-1}-\tan\underline{\omega}-(1-\beta)\mu$$

for all $(\underline{\omega},\beta)\in]-\pi/2,\pi/2[\times[-1,1]$. For all the already defined expression of $u_t(\theta)$, we have $u_t(\theta)\to\pm\infty$ as $\omega\to\mp\infty$, we set $\underline{u}_t(+\pi/2,\beta,\underline{\sigma})\equiv-\infty$ and $\underline{u}_t(-\pi/2,\beta,\underline{\sigma})\equiv+\infty$ for all $(\beta,\underline{\sigma})\in[-1,1]\times[0,\pi/2]$. With these conventions, $\{\underline{u}_t(\underline{\theta})\}_t$ is well defined and constitutes an ergodic stationary sequence in $\overline{\mathbb{R}}$ for all $\underline{\theta}$ in the compact set $\underline{\Theta}:=[-\pi/2,\pi/2]\times[-1,1]\times[0,\pi/2]$. We also extend the definition domain of T, by considering T as a function from $\overline{\Theta}:=[-\infty,\infty]\times[-1,1]\times[0,\infty]$ to $\underline{\Theta}$. It suffices to set $T(\pm\infty,\beta,\sigma)=(\pm\pi/2,\beta,\arctan\sigma)$ for all $(\beta,\sigma)\in[-1,1]\times[0,\infty[$, $T(\pm\infty,\beta,+\infty)=(\pm\pi/2,\beta,+\pi/2)$ for all $\beta\in[-1,1]$, and $T(\omega,\beta,+\infty)=(\arctan\omega,\beta,+\pi/2)$ for all $(\omega,\beta)\in]-\infty,+\infty[\times[-1,1]$. Note that T remains a one-to-one map.

It can be seen that $\underline{u}_t(\underline{\theta})$ is not continuous at the point $\underline{\theta}=(0,1,0)$, but is continuous at any other $\underline{\theta}$ [when $\underline{\Theta}$ is equipped with any metric and $\overline{\mathbb{R}}$ is equipped with a metric of the form $\mathrm{d}(x,y)=|F(x)-F(y)|$, where F denotes any bounded and strictly increasing continuous function]. We similarly extend the definition domain of $\breve{\underline{u}}_t(\underline{\theta})$ by setting $\breve{\underline{u}}_t(\pm\pi/2,\beta,\underline{\sigma})\equiv\mp\infty$, $\breve{\underline{u}}_t(\underline{\omega},\beta,\pi/2)=\underline{u}_t(\underline{\omega},\beta,\pi/2)$ for $t\geq 2$, and $\breve{\underline{u}}_1(\underline{\omega},\beta,\pi/2)=X_1-\tan\underline{\omega}-(1-\beta)\mu$.

We are now in a position to state the following theorem.

Theorem 5

Assume that $\beta_0\neq 0$ and that the assumptions of theorem 2 are satisfied for $M=1$.

Let $(\hat{\underline{\theta}}_n)$ be a sequence of estimators satisfying

$$\hat{\underline{\theta}}_n=\arg\min_{\underline{\theta}\in\underline{\Theta}}\underline{Q}_n(\underline{\theta}), \quad \textit{where} \quad \underline{Q}_n(\underline{\theta})=\frac{1}{n}\sum_{t=1}^{n}\{\breve{\underline{u}}_t(\underline{\theta})\}^2.$$

Then, almost surely, $T^{-1}\left(\hat{\underline{\theta}}_n\right)\to\theta_0$ as $n\to\infty$.

If the assumptions of theorem 3 are satisfied for $M=1$, then the limiting distribution of $\sqrt{n}\{T^{-1}(\hat{\underline{\theta}}_n)-\theta_0\}$ is a centred normal with covariance matrix $J^{-1}IJ^{-1}$, where $J=J^{(1)}$ and $I=I^{(1,1)}$ are defined in theorem 3.

5.2. Estimation when moments of the latent processes are unknown

In this section, we maintain the assumptions that $\mathrm{var}(v_t)=1$ and A2, in particular $E(Z_t)=\mu$ for an arbitrary constant μ. These assumptions, or equivalent ones, cannot be avoided for already discussed identifiability reasons. However, the assumption that higher moments of the unobserved processes (v_t) and (Z_t) are known can be seen as a restriction to the applicability of our estimation procedure. This criticism applies to most of the estimation procedures mentioned in the introduction as well. Theorem 2 requires the knowledge of σ_Z^2. Fortunately, as can be seen from (5), the mean and AR coefficient of the ARMA(1,1) representation do not depend on σ_Z^2. This is illustrated in the following example.

Example 3. (Examples 1 and 2 continued): let us consider the model defined in example 1, but assume that η_t is $\mathcal{U}[-1.44025, 1.44025]$, instead of $\mathcal{N}(0,1)$, distributed. Then $\mu_Z=-1.270$, as in the case where $\eta_t\sim\mathcal{N}(0,1)$, but $\sigma_Z^2=4$ instead of 4.935. The ARMA(1,1) representation becomes $(X_t-11.27)-0.9(X_{t-1}-11.27)=u_t-0.79u_{t-1}$, where (u_t) is a white noise with variance $\sigma_u^2=5.78$.

Hence, when σ_Z^2 changes, only σ_u^2 and the MA coefficient change in the ARMA(1,1) representation. It follows that, based on the sole ARMA(1,1) representation, ω_0 and β_0 are consistently estimated even if σ_Z^2 is mis-specified. In the following theorem we also obtain a consistent estimator for the remaining parameter σ_0.

Theorem 6

Assume that $\beta_0\neq 0$, and that σ_Z^2 exists but is unknown. Let σ_Z^* be an arbitrary strictly positive number, and let $\hat{\theta}_n(\sigma_Z^*)$ be a sequence of LR estimators of θ obtained for $M=1$ where $\breve{u}_t(\theta)$ is computed by taking σ_Z^* instead of σ_Z in (8). For any K such that θ_0 and $\theta_0^*:=(\omega_0,\beta_0,(\sigma_Z^*/\sigma_Z)\sigma_0)$ belong to the interior of Θ_K, we have

$$\hat{\theta}_n(\sigma_Z^*)\to\theta_0^* \quad \text{almost surely as} \quad n\to\infty.$$

Moreover, a strongly consistent estimator of σ_0^2 is

$$\hat{\hat{\sigma}}_n^2=(1-\hat{\beta}_n^2)\hat{\beta}_n^{-1}\hat{\gamma}_X(1),$$

where $\hat{\beta}_n=\hat{\beta}_n(\sigma_Z^*)$ is the second component of $\hat{\theta}_n(\sigma_Z^*)$.

A straightforward consequence of theorem 6, and of the second equality in (5), is that σ_Z^2 can in turn be consistently estimated by

$$\hat{\sigma}_Z^2=\hat{\gamma}_X(0)-\hat{\gamma}_X(1)\hat{\beta}_n^{-1}. \tag{17}$$

The higher-order representations can similarly be used without the knowledge of higher-order moments for (v_t) and (Z_t). The following result shows that β_0 is consistently estimated, whatever the values attributed to those moments in the criterion.

Theorem 7

Assume that $\beta_0\neq 0$, and that the moments of Z_t and v_t exist up to order $2M$ but are unknown. Let $\hat{\theta}_n^*=(\hat{\omega}_n^*,\hat{\beta}_n^*,\hat{\sigma}_n^*)$ be a sequence of LR estimators satisfying (14), the unknown moments Z_t and v_t being replaced by any (compatible) arbitrary values. For any K such that $\theta_0\in\Theta_K$, we have $\hat{\beta}_n^*\to\beta_0$ almost surely as $n\to\infty$.

To obtain consistent estimators of ω_0 and σ_0 based on the ARMA(m,m) representations for X_t^m $(m=1,\ldots,M)$, some consistent estimates of the moments of Z_t and v_t up to order $2M$ are required. To this aim, the following estimation scheme can be used. For ease of presentation we consider the case $M=2$. Let $\Gamma_X^{(m,m')}(k)=\mathrm{cov}(X_t^m, X_{t-k}^{m'})$. Using section 3.2, one can derive the following relations:

$$\Gamma_X^{(1,2)}(1)=\frac{\beta\sigma^3}{1-\beta^3}E(v_t^3)+\frac{2\omega\sigma^2\beta}{(1-\beta^2)(1-\beta)}+\frac{2\mu_Z\beta\sigma^2}{1-\beta^2} \tag{18}$$

$$\Gamma_X^{(1,2)}(0)=\frac{\sigma^3E(v_t^3)}{1-\beta^3}+\mathrm{cov}(Z_t^2,Z_t)+\frac{2\omega\sigma^2}{(1-\beta^2)(1-\beta)}+\frac{2\mu_Z\sigma^2}{1-\beta^2}+\frac{2\omega\sigma_Z^2}{1-\beta} \tag{19}$$

$$\Gamma_X^{(2,2)}(1)=\frac{\beta^2\sigma^4}{1-\beta^4}\mathrm{var}(v_t^2)+E(v_t^3)f_1(\omega,\beta,\sigma,\mu_Z)+f_2(\omega,\beta,\sigma,\mu_Z) \tag{20}$$

$$\begin{aligned}\Gamma_X^{(2,2)}(0)=&\frac{\sigma^4}{1-\beta^4}\mathrm{var}(v_t^2)+\mathrm{var}(Z_t^2)+\frac{4\omega}{1-\beta}\mathrm{cov}(Z_t,Z_t^2)\\&+4\frac{\omega+(1-\beta)\mu_Z}{(1-\beta^3)(1-\beta)}\sigma^3E(v_t^3)+f_3(\omega,\beta,\sigma,\mu_Z,\sigma_Z)\end{aligned} \tag{21}$$

for some functions f_1, f_2 and f_3 of the parameters and the first two moments of Z_t. The details are not reported here but are available from the authors. By theorem 6 and (17), some first-step consistent estimators of θ_0 and σ_Z are available. Recalling that μ_Z is known, the previous equations allow to consistently estimate the unknown moments required to compute the LR estimator. Replacing in (18)–(21) the autocovariances of the powers of X_t by their empirical counterparts, the method of moments can be used. Equation (18) provides a consistent estimator of $E(v_t^3)$. Substituting that into (19) we get a consistent estimator of $\mathrm{cov}(Z_t^2, Z_t)$. By (20) and (21) we then obtain estimators for $\mathrm{var}(v_t^2)$ and $\mathrm{var}(Z_t^2)$. Next, these estimators can be used to produce a second-step estimator of θ_0 based on the ARMA representations for X_t and X_t^2.

6. Finite-sample properties

In this section, we carry out several Monte Carlo experiments to analyse the finite-sample properties of the estimators introduced above.

As in section 4, the model parameters, namely ω, β and σ, can be estimated from the sole ARMA(1,1) representation for X_t. Following Jacquier *et al.* (1994), Andersen (1994), Andersen & Sørensen (1996), we use $\theta=(\omega,\beta,\sigma)=(-0.736, 0.90, 0.363)$ as our leading parameter setting and we simulate model (1) with independent standard Gaussian distributions for (η_t) and (v_t). However, it is important to note that the estimation procedure does not hinge on any assumption on the distribution of the latent processes. It involves three successive steps: (i) derivation of the ARMA representation for all values of θ: the first values of the autocovariance function are computed recursively; (ii) inversion of the MA part to obtain the sequences $\{u_t(\theta)\}$; (iii) minimization of the sum of squares with respect to θ. Note that the method does not require estimating the ARMA representation.

The results of several simulation experiments are given in Table 1. We consider sample sizes of $n=500, 1000$ and 2000. The table was computed using 1000 replications, NAG random number generation and a numerical optimization routine which does not require analytic derivatives. In a few cases and for the lower sample sizes, our optimization routine was unable to locate a meaningful value for the criterion function. It seems that the likelihood is

Table 1. *Simulated mean and root mean squared error (RMSE), asymptotic standard errors (ASE)*

Size	Non-convergence	Estimator	Mean	RMSE	ASE
Design 1: $\omega = -0.736, \beta = 0.900, \sigma = 0.363$					
$n = 500$	19	$\hat{\omega}$	−1.075	0.843	0.458
		$\hat{\beta}$	0.854	0.114	0.062
		$\hat{\sigma}$	0.432	0.222	0.157
$n = 1000$	2	$\hat{\omega}$	−0.892	0.520	0.324
		$\hat{\beta}$	0.879	0.071	0.044
		$\hat{\sigma}$	0.398	0.152	0.111
$n = 2000$	0	$\hat{\omega}$	−0.797	0.267	0.229
		$\hat{\beta}$	0.892	0.036	0.031
		$\hat{\sigma}$	0.376	0.084	0.079
Design 2: $\omega = 0.000, \beta = 0.990, \sigma = 0.200$					
$n = 500$	2	$\hat{\omega}$	−0.001	0.026	0.009
		$\hat{\beta}$	0.976	0.032	0.008
		$\hat{\sigma}$	0.225	0.080	0.053
$n = 1000$	0	$\hat{\omega}$	−0.000	0.011	0.006
		$\hat{\beta}$	0.985	0.012	0.005
		$\hat{\sigma}$	0.209	0.047	0.038
$n = 2000$	0	$\hat{\omega}$	0.000	0.006	0.004
		$\hat{\beta}$	0.988	0.006	0.004
		$\hat{\sigma}$	0.205	0.033	0.027

The statistics are based on 1000 simulated samples of size n, using starting values $\beta = 0$ and $\beta = 0.5$; the starting values for ω and σ were obtained by replacing μ_X and $\gamma_X(0)$ by their empirical counterparts, for the two starting values of β. The last column shows the finite sample approximation to the ASE obtained as the square roots of the diagonal terms of Σ/n.

very flat for estimated values of β close to zero (recall that the model is not identifiable from one ARMA representation when $\beta = 0$). We decided to discard the simulations when either the estimated values for $|\beta|$ or $|\alpha|$ were <0.01, or the empirical variance of the noise in (6) was too far from the variance computed from the model, i.e.

$$\frac{|1/n\sum u_t^2(\hat{\theta}_n) - \sigma_u^2(\hat{\theta}_n)|}{\{1/n\sum u_t^2(\hat{\theta}_n)\}} > 0.5.$$

The numbers of failures have been reported. Of course, these failures could have been avoided by constraining the parameter estimates to belong to a neighbourhood of the true value (e.g. β between 0.5 and 1.0 as in Harvey & Shephard, 1996), but we found it preferable to work with unconstrained parameters. When the sample size increases, the occurrences of such non-converging samples tend to vanish rapidly. The estimation results obtained for $n = 500$ display an important bias, mainly for the intercept ω. The asymptotic distribution seems to provide a good approximation for the standard deviation of the estimator. At least for $n \geq 2000$, the finite-sample approximations to the asymptotic standard errors (ASE) are very close to the root mean squared errors (RMSE) of the parameter estimates computed over the 1000 simulated samples. Note that the ASEs are given by the square roots of the diagonal terms of Σ/n. When the sample size increases, the quality of the inference improves rapidly. Another set of experiments have been conducted with a high degree of persistence in the volatility equation, namely $\theta = (0.000, 0.990, 0.200)$. The results displayed in Table 1 show that the estimator works quite well in this situation. Even in small samples, the number of non-convergences is close to zero and the biases are small. The adequacy between the ASEs and the Monte Carlo standard errors is even more pronounced than in the former set of experiments.

Similar experiments, not reported here, have been conducted by using two ARMA representations, i.e. the ARMA(1,1) for (X_t) and the ARMA(2,2) for (X_t^2), with weights chosen equal to the inverses of the empirical standard deviations of X_t and X_t^2. The most noticeable output of these experiments is that the non-convergences vanish (for $n=500$, the number of non-convergences out of 1000 simulations is 0, instead of 19 for $M=1$). Discarding the non-convergences, the bias and accuracy turn out to be roughly the same for $M=1$ and $M=2$.

It seems natural to want to compare the novel estimation method of this paper with alternative procedures. Apart from the GMM and QML procedures, all the methods mentioned in the introduction are computationally demanding and often difficult to implement. However, the GMM and QML methods are known to yield a lesser degree of accuracy in finite samples than more sophisticated approaches such as the Bayesian MCMC procedure of Jacquier *et al.* (1994) (henceforth denoted as JPR). To retain a benchmark, we therefore compare our approach with the GMM, QML and JPR procedures. It can be seen from the results displayed in Table 2 that, for the sample size of 2000 and under correct specification of the noise distributions, the Bayes estimator clearly dominates the other methods. The LR estimator performs better than the QML estimator, and it appears to be highly competitive relative to the GMM estimators. In particular, the RMSEs are always smaller with the LR estimator than with the GMM estimator. It should also be noticed that the performances of GMM procedures are very sensitive to the choice of the number of moments (the results worsen when five or 24 moments are used, see Andersen, 1994) and to the estimation of the optimal weighting matrix. Finally, the number of discarded simulations can become very significant when a small number of moments are used or for smaller sample sizes (see Andersen, 1994).

In Table 3 we have reported ASEs of GMM, QML and LR estimators, taking the same parameter values as in Ruiz (1994) and Andersen & Sørensen (1997). It should be noted that ω is fixed, so the unknown parameters are β and σ. Andersen & Sørensen (1997, p. 399) and Ruiz (1994) seem to have fixed $\omega=0$, arguing that the estimation of ω should not have much of an effect on the estimation of the other two parameters. But it appears, from both

Table 2. *Comparison of method of moments, QML, Bayes and LR estimators for sample size of 2000*

Method	Number of moments (GMM) or representations (LR)	Non-convergence	ω −0.736	β 0.900	σ 0.363
GMM (JPR)	24	0	−0.86	0.88	0.31
			(0.42)	(0.06)	(0.10)
QML (JPR)		0	−0.853	0.88	0.383
			(0.46)	(0.06)	(0.11)
Bayes (JPR)		0	−0.762	0.896	0.359
			(0.15)	(0.02)	(0.034)
GMM1 (A)	14	11	−0.75	0.90	0.30
			(0.39)	(0.05)	(0.11)
GMM2 (A)	14	3	−0.73	0.90	0.30
			(0.35)	(0.05)	(0.10)
GMM3 (A)	14	10	−0.59	0.92	0.28
			(0.31)	(0.04)	(0.12)
LR	1	0	−0.797	0.892	0.376
			(0.267)	(0.036)	(0.084)

QML, quasi-maximum likelihood; GMM, generalized method of moments; LR, linear representation. The first two sets of results are reproduced from Jacquier *et al.* (1994, Table 9) (JPR) and Andersen (1994, Tables 2, 4 and 6). This table reports the mean parameter estimates with the associated root mean squared error in parentheses below. The JPR results are based on a set of 500 simulated samples, the Andersen results are based on 1000 converged parameter estimates and our results are based on 1000 simulated samples. GMM1, GMM2 and GMM3 correspond to different GMM procedures for estimating the weighting matrix (see Andersen, 1994).

Table 3. *ASEs of the GMM, QML and LR estimators of β and σ (ω known)*

	GMM	QML	LR ($\omega=0$)	LR ($\omega=1$)	LR ($\omega=-2$)
$\beta=0.90$	0.28	0.68	0.69 (0.64)	0.10 [0.09]	0.05 (0.05)
$\sigma=1$	2.75	6.46	4.03 (3.84)	3.00 [2.87]	3.02 [3.07]
$\beta=0.70$	2.03	2.31	2.64 [2.79]	0.36 [0.32]	0.18 [0.19]
$\sigma=1$	6.44	10.74	7.78 [8.48]	3.90 [3.67]	3.97 [3.89]
$\beta=0.90$	1.60	1.77	1.71 [1.86]	0.04 [0.04]	0.02 [0.02]
$\sigma=0.3$	1.57	2.24	3.70 [4.18]	1.74 [1.65]	1.74 [2.05]
$\beta=0.70$	6.37	15.12	16.04 [16.20]	0.22 [0.20]	0.11 [0.12]
$\sigma=0.3$	2.23	7.62	14.07 [17.35]	5.03 [5.35]	5.04 [5.89]
$\beta=0.95$	1.27	0.92	0.81 [0.91]	0.01 [0.01]	0.006 [0.006]
$\sigma=0.2$	1.12	0.95	2.16 [2.49]	1.23 [1.19]	1.23 [1.32]
$\beta=0.97$	0.74	0.47	0.38 [0.51]	0.01 [0.01]	0.003 [0.003]
$\sigma=0.2$	1.02	0.69	1.48 [1.95]	1.12 [1.14]	1.12 [1.07]
$\beta=0.99$	0.36	0.18	0.13 [0.18]	0.002 [0.002]	0.001 [0.001]
$\sigma=0.2$	1.46	0.47	1.12 [1.39]	1.25 [0.95]	1.25 [0.92]

GMM, generalized method of moments; QML, quasi-maximum likelihood; LR, linear representation; RMSE, root mean squared error. The first two columns are reproduced from Andersen & Sørensen (1997) and Ruiz (1994). The numbers in parenthesis are 100 times the RMSEs of the LR estimator, obtained from 100 simulated samples of size 10,000 (100 times RMSE should be close to the theoretical ASE).

the ASE and RMSE of Table 3 that the value of ω is of importance for the accuracy of the other parameters. A huge difference is observed, for instance, when $\beta=0.7$, whether $\omega=0$, 1 or -2. A general comment on this table is that our estimator of β dominates the GMM and QML estimators, in every instance, when β is >0.95. This is worth emphasizing, as β usually takes values close to unity for real daily financial time series. Finally, we found it important to compute RMSEs from simulation experiments, in order to validate the ASEs. In this respect, the agreement of the quantities, in all designs reported in the table, is very satisfactory. Unfortunately we are not able to similarly validate the ASEs obtained from the two other methods.

7. Conclusion

An estimation procedure exploiting the ARMA representations for powers of the log-squared SV process is introduced in this paper. Compared with other estimation techniques recently suggested in the literature, the method is easy to implement and computationally attractive. Simulation experiments show that the method is competitive with more complicated estimation procedures. On the theoretical front, the relative simplicity of the proposed estimator allowed us to derive its limiting distribution. Moreover, for the simplest version of the method, the asymptotic accuracy of the estimator is computable.

The estimation procedure developed in this paper is not limited to the canonical SV model. It is well known that numerous non-linear processes (or transformations of such processes) admit ARMA representations. The strategy, consisting in selecting the parameter of a non-linear model that leads to optimal ARMA predictions (of the process itself or of some relevant transformations), could be extended to other non-linear processes. Of course, the task requires a case-by-case study and such extensions are left for future research.

Acknowledgement

We are grateful to the associate editor for his very helpful comments.

References

Abramowitz, M. & Stegun, I. A. (1970). *Handbook of mathematical functions*. Dover Publications, New York, NY.

Andersen, T. G. (1994). Comment on 'Bayesian Analysis of Stochastic Volatility Models' by Jacquier, Polson & Rossi, *J. Bus. Econom. Statist.* **12**, 389–392.

Andersen, T. G. & Sørensen, B. E. (1996). GMM estimation of a stochastic volatility model: a Monte Carlo study. *J. Bus. Econom. Statist.* **14**, 328–352.

Andersen, T. G. & Sørensen, B. E. (1997). GMM and QML asymptotic standard deviations in stochastic volatility models: comments on Ruiz (1994). *J. Econometrics* **76**, 397–403.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Berkes, I., Horváth, L. & Kokoszka, P. S. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201–227.

Billingsley, P. (1995). *Probability and measure*. John Wiley, New York, NY.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

Breidt, F. J. & Carriquiry, A. L. (1996). Improved quasi-maximum likelihood estimation for stochastic volatility models. In *Modelling and prediction: honoring Seymour Geisser* (eds C. Lee, W. O. Johnson & A. Zellner), 288–247. Springer-Verlag, New York, NY.

Brockwell, P. J. & Davis, R. A. (1991). *Time series: theory and methods*. Springer-Verlag, New York, NY.

Chib, S., Nardari, F. & Shephard, N. (2002). Markov Chain Monte Carlo methods for stochastic volatility models. *J. Econometrics* **108**, 281–316.

Danielsson, J. (1994). Stochastic volatility in asset prices, estimation with simulated maximum likelihood. *J. Econometrics* **64**, 375–400.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987–1007.

Francq, C. & Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605–637.

Fridman, M. & Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *J. Bus. Econom. Statist.* **16**, 284–291.

Ghysels, E., Harvey, A. C. & Renault, E. (1996). Stochastic volatility. In *Statistical Methods in Finance* (eds C. R. Rao & G. S. Maddala), 119–191. North-Holland, Amsterdam.

Harvey, A. C. & Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *J. Bus. Econom. Statist.* **14**, 429–434.

Harvey, A. C., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models. *Rev. Econom. Stud.* **61**, 247–264.

Hoffmann, M. (2002). Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Process. Appl.* **97**, 147–170.

Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42**, 281–300.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *J. Bus. Econom. Statist.* **12**, 371–389.

Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.* **65**, 361–393.

Lee, S. W. & Hansen, B. E. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* **10**, 29–58.

Linton, O. (1997). An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory* **13**, 558–581.

Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **64**, 575–596.

Mahieu, R. J. & Schotman, P. C. (1998). An empirical application of stochastic volatility models. *J. Appl. Econometrics* **13**, 333–360.

Melino, A. & Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45**, 7–39.

Nelson, D. B. (1988). Time series behavior of stock market volatility and returns. Unpublished PhD dissertation, Economics Department, Massachusetts Institute of Technology. Cambridge, MA.

Potscher, B. M. & Prucha, I. R. (1997). *Dynamic nonlinear econometric models: asymptotic theory*. Springer-Verlag, New York, NY.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *J. Econometrics* **63**, 289–306.
Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time series models in econometrics, finance and other fields*. (eds D. R. Cox, O. E. Barndorff-Nielson & D. V. Hinkley), 1–67. Chapman and Hall, London.
Taylor, S. J. (1986). *Modelling financial time series*. Wiley, New York, NY.
Taylor, S. J. (1994). Modelling stochastic volatility. *Math. Finance* **4**, 183–204.

Received February 2004, in final form December 2005

Jean-Michel Zakoïan, CREST, 3 Avenue Pierre Larousse, 92245 Malakoff Cedex, France.
E-mail: zakoian@ensae.fr

Appendix: proofs

Proof of theorem 1. We have

$$X_t^M = \sum_{i=0}^{M} \binom{M}{i} Y_t^i Z_t^{M-i}$$

and (X_t) is clearly $2M$th order stationary. To show that $(\mathbf{X}_t^{(M)})$ admits a vector ARMA representation, we compute the following covariances, for all $k \geq 0$ and all $m, m' \in \{1, \ldots, M\}$

$$\Gamma_X^{(m,m')}(k) = \mathrm{cov}(X_t^m, X_{t-k}^{m'}) = \sum_{i=0}^{m} \sum_{j=0}^{m'} \binom{m}{i} \binom{m'}{j} \mathrm{cov}(Y_t^i Z_t^{m-i}, Y_{t-k}^j Z_{t-k}^{m'-j}). \tag{22}$$

Hence, for all $k > 0$,

$$\Gamma_X^{(m,m')}(k) = \sum_{i=1}^{m} \sum_{j=1}^{m'} \binom{m}{i} \binom{m'}{j} E(Z_t^{m-i}) E(Z_t^{m'-j}) \mathrm{cov}(Y_t^i, Y_{t-k}^j). \tag{23}$$

Let, for $i, j > 0$, $\Gamma_Y^{(i,j)}(k) = \mathrm{cov}(Y_t^i, Y_{t-k}^j)$ and let $v_i = E(\omega + \sigma v_t)^i$. We have from (4), for $i, j, k > 0$,

$$\Gamma_Y^{(i,j)}(k) = \mathrm{cov}\left(\sum_{l=0}^{i} \binom{i}{l} \beta^l Y_{t-1}^l (\omega + \sigma v_t)^{i-l}, Y_{t-k}^j \right) = \sum_{l=1}^{i} \binom{i}{l} v_{i-l} \beta^l \Gamma_Y^{(l,j)}(k-1). \tag{24}$$

In particular we have for $k > 0$,

$$\Gamma_Y^{(1,j)}(k) = \beta \Gamma_Y^{(1,j)}(k-1), \quad \Gamma_Y^{(2,j)}(k) = \beta^2 \Gamma_Y^{(2,j)}(k-1) + 2\beta v_1 \Gamma_Y^{(1,j)}(k-1).$$

Hence, for $k > 1, (1 - \beta^2 L)(1 - \beta L)\Gamma_Y^{(2,j)}(k) = 0$ [where $L\Gamma_Y^{(i,j)}(k) = \Gamma_Y^{(i,j)}(k-1)$]. More generally, it is straightforward to show that for $i, j > 0$, $\forall k > i-1$, $\prod_{l=1}^{i}(1 - \beta^l L)\Gamma_Y^{(i,j)}(k) = 0$. In view of (23) we then have $\forall k > m$, $\prod_{l=1}^{m}(1 - \beta^l L)\Gamma_X^{(m,m')}(k) = 0$. In particular, when $m' = m$ this equation shows, by lemma 1 that (X_t^m) is an ARMA(m, m). Moreover, we have $\forall k > M$, $\prod_{l=1}^{M}(1 - \beta^l L)\Gamma_{\mathbf{X}^{(M)}}(k) = 0$ where $\Gamma_{\mathbf{X}^{(M)}}(\cdot)$ denotes the autocovariance function of $(\mathbf{X}_t^{(M)})$, which proves that $(\mathbf{X}_t^{(M)})$ is an ARMA(M, M).

Proof of theorem 2. It will be convenient to consider the theoretical criteria:$\forall \theta \in \Theta$,

$$O_n(\theta) = \sum_{m=1}^{M} \tau_n^{(m)} \left\{ \frac{1}{n} \sum_{t=1}^{n} [u_t^{(m)}(\theta)]^2 \right\}, \quad O_\infty(\theta) = \sum_{m=1}^{M} \tau^{(m)} E_{\theta_0}[u_t^{(m)}(\theta)]^2.$$

As $u_t^{(m)}(\theta_0)$ is the linear innovation of X_t^m, and as, by (11), $u_t^{(m)}(\theta) - u_t^{(m)}(\theta_0)$ is a linear combination of the X_{t-i}^m for $i > 0$, we have $E_{\theta_0} u_t^{(m)}(\theta_0)\{u_t^{(m)}(\theta) - u_t^{(m)}(\theta_0)\} = 0$. Hence

$$O_\infty(\theta) = O_\infty(\theta_0) + \sum_{m=1}^{M} \tau^{(m)} E_{\theta_0}[u_t^{(m)}(\theta) - u_t^{(m)}(\theta_0)]^2 \geq O_\infty(\theta_0),$$

with equality if and only if, for all m such that $\tau^{(m)} > 0$, $u_t^{(m)}(\theta) = u_t^{(m)}(\theta_0)$ with probability one. In particular, if $O_\infty(\theta) = O_\infty(\theta_0)$ then $u_t^{(1)}(\theta) = u_t^{(1)}(\theta_0)$ with probability one. Hence $c^{(1)}(\theta) = c^{(1)}(\theta_0)$ and, for $i \geq 0, b_{i+1}^{(1)}(\theta) = b_{i+1}^{(1)}(\theta_0)$. In view of (12) we easily find that $(\omega, \beta, \alpha) = (\omega_0, \beta_0, \alpha_0)$. Here we have to distinguish two different cases.

First suppose that $\beta_0 \neq 0$. By (7), $\alpha = \alpha_0$ and $\beta = \beta_0$ imply $\sigma = \sigma_0$. Therefore, we have shown that

$$\text{if } \theta \neq \theta_0 \quad \text{then} \quad O_\infty(\theta) > O_\infty(\theta_0). \tag{25}$$

As $\max_{1 \leq m \leq M} \sup_{\theta \in \Theta_K} |b_i^{(m)}(\theta)| = O(\rho^i)$, for some $0 \leq \rho < 1$, it is easy to show that, almost surely, for $m = 1, \ldots, M$,

$$\lim_{t \to \infty} \sup_{\theta \in \Theta_K} |u_t^{(m)}(\theta) - \breve{u}_t^{(m)}(\theta)| = 0$$

and

$$\lim_{n \to \infty} \sup_{\theta \in \Theta_K} |O_n(\theta) - Q_n(\theta)| = 0. \tag{26}$$

Let, for any positive real d, $V_d(\theta)$ be the open sphere with centre θ and radius $1/d$. For any θ_1 belonging to the interior of Θ_K, and any m, the process $(\inf_{\theta \in V_d(\theta_1) \cap \Theta_K} \{u_t^{(m)}(\theta)\}^2)_t$ is stationary and ergodic, as a measurable function of the stationary and ergodic process (ϵ_t). It follows, by (26) and the ergodic theorem that

$$\begin{aligned} \liminf_{n \to \infty} \inf_{\theta \in V_d(\theta_1) \cap \Theta_K} Q_n(\theta) &\geq \sum_{m=1}^{M} \tau^{(m)} \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} \inf_{\theta \in V_d(\theta_1) \cap \Theta_K} \left\{u_t^{(m)}(\theta)\right\}^2 \\ &= \sum_{m=1}^{M} \tau^{(m)} E_{\theta_0} \inf_{\theta \in V_d(\theta_1) \cap \Theta_K} \left\{u_t^{(m)}(\theta)\right\}^2 \qquad a.s. \end{aligned}$$

By the monotone convergence theorem and the continuity of $u_t^{(m)}(\cdot)$, we have

$$\lim_{d \uparrow \infty} \sum_{m=1}^{M} \tau^{(m)} E_{\theta_0} \inf_{\theta \in V_d(\theta_1) \cap \Theta_K} \left\{u_t^{(m)}(\theta)\right\}^2 = O_\infty(\theta_1).$$

Using (25), we deduce that for any $\theta_1 \neq \theta_0$ there exists a neighbourhood $V(\theta_1)$ of θ_1 such that

$$\liminf_{n \to \infty} \inf_{\theta \in V(\theta_1)} Q_n(\theta) > \lim_{n \to \infty} Q_n(\theta_0). \tag{27}$$

By a standard compactness argument we conclude that the sequence $(\hat{\theta}_n)$ almost surely converges to θ_0.

Now we consider the case $\beta_0 = 0$. We have already seen that $O_\infty(\theta) = O_\infty(\theta_0)$ implies $(\omega, \beta) = (\omega_0, \beta_0)$ but it does not entail $\sigma = \sigma_0$ as $\alpha = 0$ for any value of σ. Adapting the previous arguments to the sequence $(\hat{\omega}_n, \hat{\beta}_n)$ we similarly conclude that this sequence converges almost surely to (ω_0, β_0). The strong consistency of $\tilde{\sigma}^2$ is a straightforward consequence of the second equality in (5) and the ergodic theorem.

Proof of theorem 3. It will be convenient to consider truncated variables. For a positive integer r, write

$$
{}_rY_t = \frac{\omega}{1-\beta} + \sigma \sum_{i=0}^{r} \beta^i \sigma v_{t-i}, \quad {}_rX_t = {}_rY_t + Z_t
$$

$$
{}_ru_t^{(m)}(\theta_0) = c^{(m)}(\theta_0) + \sum_{i=0}^{r} b_i^{(m)}(\theta_0)\,{}_rX_{t-i}^m, \quad {}^ru_t^{(m)}(\theta_0) = u_t^{(m)}(\theta_0) - {}_ru_t^{(m)}(\theta_0)
$$

$$
\boldsymbol{D}_t = 2\sum_{m=1}^{M} \tau^{(m)} \left\{ u_t^{(m)}(\theta_0) \frac{\partial u_t^{(m)}}{\partial \theta}(\theta_0) \right\},
$$

$$
{}_r\boldsymbol{D}_t = 2\sum_{m=1}^{M} \tau^{(m)} \left\{ {}_ru_t^{(m)}(\theta_0) \frac{\partial_r u_t^{(m)}}{\partial \theta}(\theta_0) \right\}, \quad {}^r\boldsymbol{D}_t = \boldsymbol{D}_t - {}_r\boldsymbol{D}_t.
$$

We have, with notations introduced in the proof of theorem 2,

$$
\sqrt{n}\frac{\partial}{\partial \theta} O_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{D}_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} ({}_r\boldsymbol{D}_t - E_{\theta_0}\,{}_r\boldsymbol{D}_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} ({}^r\boldsymbol{D}_t - E_{\theta_0}\,{}^r\boldsymbol{D}_t).
$$

Observe that ${}_r\boldsymbol{D}_t$ is a function of $(\eta_t, \eta_{t-1}, \ldots, \eta_{t-r}, v_t, v_{t-1}, \ldots, v_{t-2r})$. Therefore, the central limit theorem (e.g. see Brockwell & Davis, 1991, Proposition 6.4.2) applied to this $2r$-dependent process entails that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}({}_r\boldsymbol{D}_t - E_{\theta_0}\,{}_r\boldsymbol{D}_t)$ has a limiting $\mathcal{N}(0, 4\,{}_r\boldsymbol{I})$ distribution, where

$$
{}_r\boldsymbol{I} = \sum_{m=1}^{M}\sum_{\ell=1}^{M} \tau^{(m)}\tau^{(\ell)} \sum_{h=-2r}^{+2r} \operatorname{cov}\left({}_ru_t^{(m)}(\theta_0)\frac{\partial_r u_t^{(m)}}{\partial\theta}(\theta_0),\, {}_ru_{t-h}^{(m)}(\theta_0)\frac{\partial_r u_{t-h}^{(m)}}{\partial\theta}(\theta_0) \right).
$$

We can show that

$$
\lim_{r\to\infty} \|{}_r\boldsymbol{I} - \boldsymbol{I}\| \to 0, \quad \left\| \operatorname{var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} {}^r\boldsymbol{D}_t \right) \right\| \leq C\rho^r,
$$

where

$$
\boldsymbol{I} = \sum_{m=1}^{M}\sum_{\ell=1}^{M} \tau^{(m)}\tau^{(\ell)} \boldsymbol{I}^{(m,\ell)},
$$

for some constants $C > 0$ and $0 < \rho < 1$. Details on this derivation can be obtained from the authors. Using a standard argument (e.g. see Brockwell & Davis, 1991, Proposition 6.3.9) we can conclude that

$$
\sqrt{n}\frac{\partial}{\partial\theta} O_n(\theta_0) \text{ has a limiting } \mathcal{N}(0, 4\boldsymbol{I}) \text{ distribution.} \tag{28}
$$

Now, with arguments given in the proof of theorem 2, it is easy to show that $\sqrt{n}\frac{\partial}{\partial\theta} O_n(\theta_0)$ and $\sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0)$ have the same asymptotic distribution (when existing). When $\hat{\theta}_n$ belongs to the interior of Θ_K, a Taylor expansion yields

$$
0 = \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\hat{\theta}_n) = \sqrt{n}\frac{\partial}{\partial\theta} Q_n(\theta_0) + \frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta^*)\sqrt{n}\left(\hat{\theta}_n - \theta_0\right), \tag{29}
$$

where θ^* is a random vector which lies between $\hat{\theta}_n$ and θ_0. As $\hat{\theta}_n$ almost surely converges to θ_0, which belongs to the interior of Θ_K, almost surely the equalities (29) hold for n large enough. From the ergodic theorem and the consistency of $\hat{\theta}_n$, it is easy to see that

$$
\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta^*) \to 2\sum_{m=1}^{M} \tau^{(m)} \boldsymbol{J}^{(m)}, \quad a.s.
$$

The conclusion follows from (28) and Slutsky's lemma.

Proof of theorem 4. As the proof is very similar to those of theorems 2 and 3, we do not give all the details. We have $\mathbf{S}=\mathbf{P}'\mathbf{P}$, where $\mathbf{P}=\big(\mathbf{P}(i,j)\big)$ is a non-singular matrix. The limit criterion satisfies, denoting by $u_{t,j}^{(M)}(\theta)$ the components of $\mathbf{u}_t^{(M)}(\theta)=\mathbf{u}_t^{(M)}$,

$$\begin{aligned} O_\infty^V(\theta) &:= E_{\theta_0}\left\{\mathbf{u}_t^{(M)}(\theta)\right\}'\mathbf{S}\left(\mathbf{u}_t^{(M)}(\theta)\right)=\sum_{i=1}^M E_{\theta_0}\left\{\sum_{j=1}^M \mathbf{P}(i,j)u_{t,j}^{(M)}(\theta)\right\}^2 \\ &=\sum_{i=1}^M\left[E_{\theta_0}\left\{\sum_{j=1}^M \mathbf{P}(i,j)u_{t,j}^{(M)}(\theta_0)\right\}^2+E_{\theta_0}\left\{\sum_{j=1}^M \mathbf{P}(i,j)\left(u_{t,j}^{(M)}(\theta)-u_{t,j}^{(M)}(\theta_0)\right)\right\}^2\right. \\ &\quad\left.+\sum_{j=1}^M\sum_{k=1}^M \mathbf{P}(i,j)\mathbf{P}(i,k)E_{\theta_0}\left\{u_{t,j}^{(M)}(\theta_0)\left(u_{t,k}^{(M)}(\theta)-u_{t,k}^{(M)}(\theta_0)\right)\right\}\right] \\ &=O_\infty^V(\theta_0)+\sum_{i=1}^M E_{\theta_0}\left\{\sum_{j=1}^M \mathbf{P}(i,j)\left(u_{t,j}^{(M)}(\theta)-u_{t,j}^{(M)}(\theta_0)\right)\right\}^2\geq O_\infty^V(\theta_0), \end{aligned}$$

as the components of the vectorial innovation $\mathbf{u}_t^{(M)}(\theta_0)$ are non-correlated with the components of $\mathbf{u}_t^{(M)}(\theta)-\mathbf{u}_t^{(M)}(\theta_0)$, which belongs to the linear past of $\mathbf{X}_t^{(M)}$. The previous inequality is strict if and only if $\mathbf{P}(\mathbf{u}_t^{(M)}(\theta)-\mathbf{u}_t^{(M)}(\theta_0))$ is not almost surely equal to zero. As $\mathbf{P}$ is non-singular, the inequality is strict if and only if $\mathbf{u}_t^{(M)}(\theta)$ is not almost surely equal to $\mathbf{u}_t^{(M)}(\theta_0)$, which means $\theta\neq\theta_0$. The rest of the consistency proof is the same as the proof of theorem 2.

By the arguments used to show (28),

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n^V(\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\frac{\partial\breve{\mathbf{u}}_t^{(M)}(\theta_0)'}{\partial\theta}\mathbf{S}_n\breve{\mathbf{u}}_t^{(M)}(\theta_0)+\frac{\partial\breve{\mathbf{u}}_t^{(M)}(\theta_0)'}{\partial\theta}\mathbf{S}_n'\breve{\mathbf{u}}_t^{(M)}(\theta_0)\right\}$$

has a limiting centred normal distribution with variance $4I^V$. As the components of $\breve{\mathbf{u}}_t^{(M)}(\theta_0)$ and $\left(\partial^2/\partial\theta_k\partial\theta_\ell\right)\breve{\mathbf{u}}_t^{(M)}(\theta_0)$ are non-correlated, the ergodic theorem yields,

$$\frac{\partial^2}{\partial\theta\partial\theta'}Q_n^V(\theta_0)=\frac{1}{n}\sum_{t=1}^n\frac{\partial\breve{\mathbf{u}}_t^{(M)}(\theta_0)'}{\partial\theta}\left(\mathbf{S}_n+\mathbf{S}_n'\right)\frac{\partial\breve{\mathbf{u}}_t^{(M)}(\theta_0)}{\partial\theta'}+o_{a.s.}(1)\to 2J^V$$

almost surely. Following the line of the proof of theorem 3, the conclusion follows.

Proof of theorem 5. With obvious notations, the analogous of (25) and (26) hold:

$$\text{if}\quad \underline{\theta}\in\underline{\Theta},\quad \underline{\theta}\neq\underline{\theta}_0\quad\text{then}\quad \underline{O}_\infty(\underline{\theta})>\underline{O}_\infty(\underline{\theta}_0)$$

and

$$\lim_{n\to\infty}\sup_{\underline{\theta}\in\underline{\Theta}}|\underline{O}_n(\underline{\theta})-\underline{Q}_n(\underline{\theta})|=0\quad a.s.$$

Note that $\hat{\underline{\theta}}_n$ exists because the objective function $\underline{Q}_n(\underline{\theta})$ is continuous over the compact set $\underline{\Theta}$, except at the single point $\underline{\theta}=(1,0,1)$. We begin to show that $\hat{\underline{\theta}}_n$ does not converge to $(0,1,0)$. Let $\underline{\theta}_n=(\underline{\omega}_n,\beta_n,\underline{\sigma}_n)$ be a sequence of elements of $\underline{\Theta}$ which converges to $(0,1,0)$. Let $\omega_n=\tan\underline{\omega}_n$ and $\sigma_n=\tan\underline{\sigma}_n$. As $n\to\infty$, $\alpha_n:=\alpha(\beta_n,\sigma_n;\sigma_Z)\to 1$ and $\alpha_n\neq 1$ when $(\beta_n,\sigma_n)\neq(1,0)$, we have $\alpha_n-\beta_n\to 0$ and

$$\liminf_{n\to\infty} \underline{u}_t^2(\underline{\theta}_n) \geq \liminf_{n\to\infty} \left[\left\{ X_t - \frac{\omega_n + (1-\beta_n)\mu}{1-\alpha_n} \right\}^2 1_{\{(\beta_n,\sigma_n)\neq(1,0)\}} + \{X_t - \mu\}^2 1_{\{(\beta_n,\sigma_n)=(1,0)\}} \right]$$

and

$$E \liminf_{n\to\infty} \underline{u}_t^2(\underline{\theta}_n) \geq \min_{c\in\mathbb{R}} E(X_t - c)^2 = \operatorname{var} X_t.$$

Let V_d be the open sphere of $\mathbb{R}^3$ with centre $(0, 1, 0)$ and radius $1/d$. With the arguments given in the proof of theorem 2, we obtain

$$\lim_{d\uparrow\infty} \liminf_{n\to\infty} \inf_{\underline{\theta}\in V_d\cap\underline{\Theta}} \underline{Q}_n(\underline{\theta}) \geq \operatorname{var} X_t > \underline{O}_\infty(\underline{\theta}_0)$$

because, in view of remark 3, X_t is not a noise. It follows that for some sufficiently large d_0, the estimator $\hat{\underline{\theta}}_n$ belongs to $\underline{\Theta} \setminus V_{d_0}$. Using the continuity of the functions $\breve{\underline{u}}_t(\cdot)$ [and therefore of $\underline{Q}_n(\cdot)$] on the compact $\underline{\Theta} \setminus V_{d_0}$, and the arguments used to show (27), we obtain that for any $\underline{\theta}_1 \neq \underline{\theta}_0$ there exists a neighborhood $V(\underline{\theta}_1)$ of $\underline{\theta}_1$ such that

$$\liminf_{n\to\infty} \inf_{\underline{\theta}\in V(\underline{\theta}_1)} \underline{Q}_n(\underline{\theta}) > \lim_{n\to\infty} \underline{Q}_n(\underline{\theta}_0).$$

By the usual compactness argument we conclude that $\hat{\underline{\theta}}_n \to \underline{\theta}_0$ *a.s.* The proof of the consistency follows because T is a one-to-one map.

When θ_0 belongs to the interior of Θ_K, for some sufficiently large K, the consistency result shows that, almost surely, the unconstrained estimator $T^{-1}(\hat{\underline{\theta}}_n)$ coincides with the constrained estimator $\hat{\theta}_n$ for some sufficiently large n. Therefore, the two estimators have the same asymptotic distribution, which completes the proof.

Proof of theorem 6. For any $\theta = (\omega, \beta, \sigma) \in \Theta$ and for an arbitrary value σ_Z^*, let $\alpha^*(\theta) = \alpha(\mathcal{S}(\theta); \sigma_Z^*)$, where $\mathcal{S}(\theta) = (\beta, \sigma)$ and the function $\alpha(\cdot, \cdot\,; \cdot)$ is defined in (8). Note that $\alpha^*(\theta)$ only depends on the parameter β and on the ratio σ/σ_Z^* and that $\alpha(\theta) = \alpha^*(\mathcal{S}(\theta^*); \sigma_Z)$ with $\theta^* = \mathcal{T}(\theta) = (\omega, \beta, \sigma\sigma_Z^*/\sigma_Z)$, where σ_Z^2 is the (unknown) true value of $\operatorname{var}(Z_t)$. Therefore, the objective function based on σ_Z takes the same value at θ than the objective function based on σ_Z^* at $\theta^* = \mathcal{T}(\theta) = (\omega, \beta, \sigma\sigma_Z^*/\sigma_Z)$. It is then easy to see that

$$\hat{\theta}_n(\sigma_Z^*) = \mathcal{T}(\hat{\theta}_n) = \left(\hat{\omega}_n, \hat{\beta}_n, \hat{\sigma}_n \frac{\sigma_Z^*}{\sigma_Z} \right) \to \theta_0^* \qquad a.s.$$

The strong consistency of $\hat{\hat{\sigma}}_n^2$ follows from the third equality in (5) and the ergodic theorem.

Proof of theorem 7. The theorem is proved by arguments similar to those of theorem 6.