Quasi Maximum Likelihood Inference for Stochastic Volatility Models

Maddalena Cavicchioli *

Abstract

In the present paper we consider the Quasi Maximum Likelihood (QML) procedure for the estimation of stationary Stochastic Volatility models. We prove the consistency of the QML estimators and compute explicitly their asymptotic variances. This allows us to obtain also consistent estimators of the asymptotic variances in explicit forms. The knowledge of the asymptotic variance-covariance matrix of the QML estimators gives a concrete possibility for the use of the classical testing procedures. Our results are related to those obtained in Ruiz (1994) and Bartolucci and De Luca (2001) (2003).

Keywords: Stochastic volatility models, asymptotically stationary process, consistency, asymptotic normality, asymptotic variance matrix.

JEL Classification: C01, C13, C58.

---

* Department of Economics "Marco Biagi", University of Modena and Reggio Emilia, Viale Berengario 51, 41121, Modena (Italy). E-mail: maddalena.cavicchioli@unimore.it.
I would like to thank Professor E. Ruiz for sending me, as personal communication, some explicit computations concerning with results stated in Ruiz (1994). These were most valuable in writing the present paper. Finally, I thank the Editor-in-Chief Professor Ephraim Clark and the anonymous referee for helpful comments and remarks. The authors are solely responsible for any errors.

1 – Introduction

Stochastic Volatility (SV) models have been received a growing interest in time series analysis since they find many financial applications as, for example, option pricing, asset allocation and risk management. For a comprehensive discussion on SV models see, for example, Taylor (1994). Estimation of SV models is difficult since it is not easy to derive their exact likelihood function. In Monfardini (1998) the estimation of SV models is based on the Indirect Inference procedure proposed by Gallant and Tauchen (1992, and published on 1996) and Gouriéroux, Monfort and Renault (1993). Another approach is represented by the Quasi Maximum Likelihood (QML) procedure which is based on the maximization of the approximated Likelihood function. The asymptotic and finite sample properties of the QML estimators were analyzed in Ruiz (1994). For parameter values corresponding to very-high-frequency financial time series, this author showed that such estimators are more efficient than some estimators based on the generalized method of moments. In the literature there has been employed a linearized filtering method (Extended Kalman Filter) to obtain QML estimation. See, for example, Harvey et al. (1994) and Harvey and Shephard (1996). Fridman and Harris (1998) proposed a method for approximating the likelihood of the basic SV model. Then they suggested to estimate the parameters of such a model maximizing the approximate likelihood by an algorithm which makes use of numerical derivatives. An extension of this method was employed by Bartolucci and De Luca (2001) (2003) in order to determine explicitly the first and second analytical derivatives of the approximate likelihood. In particular, the latter derivatives may be used to compute the standard errors of the estimators and confidence intervals for the parameters. Finally, Bayesian analysis of SV models can be found, for example, in Jacquier, Polson and Rossi (1994). In the present paper we consider the QML procedure for the estimation of stationary SV models. We prove the consistency of the QML estimators and compute explicitly their asymptotic variances. This allows us to obtain also consistent estimators of the asymptotic variances in explicit forms. The knowledge of the asymptotic variance-covariance matrix of the QML estimators gives a concrete possibility for the use of the classical testing procedures. Our results are related to those obtained in Ruiz (1994) and Bartolucci and De Luca (2001) (2003).

Let us consider the basic stochastic volatility model given by

$$y_t = exp\left\{\frac{1}{2}h_t(\theta)\right\}u_t \qquad (1.1)$$
$$h_t(\theta) = \mu + \rho h_{t-1}(\theta) + v_t$$

with $\theta = (\mu \quad \rho \quad \sigma_v^2)'$ in a compact parameter set $\Theta \subset (0, +\infty)^3$. The model does not contain exogenous variables. Here the two error terms $u_t \sim IIN(0,1)$ and $v_t \sim IIN(0, \sigma_v^2)$ are assumed to be independent of one other. To ensure stationarity, we always set $|\rho| < 1$. This model specifies the variance $\sigma_t^2(\theta) = exp\{h_t(\theta)\}$ of the observable variable y_t to be a function of the unobservable $h_t(\theta)$ which follows a first order autoregressive process. The true parameter is denoted by $\theta_0 = (\mu_0 \quad \rho_0 \quad \sigma_0^2)'$, and it belongs to the interior of Θ. For any asymptotically stationary process $(X_t)_{t\geq 0}$, let

$$E_\infty(X_t) = \lim_{T\to+\infty} E_T(X_t) = \lim_{T\to+\infty} \frac{1}{T}\sum_{t=1}^{T} X_t$$

provided this limit exists. In the stationary case ($|\rho| < 1$), we get

$$h_t(\theta) = (1-\rho L)^{-1}\mu + (1-\rho L)^{-1}v_t = \mu(1-\rho)^{-1} + \sum_{i=0}^{\infty}\rho^i\, v_{t-i} < +\infty \qquad (1.2)$$

hence

$$var(y_t) = \sigma_t^2(\theta) = exp\{h_t(\theta)\} = exp(\mu(1-\rho)^{-1})\ \prod_{i=0}^{\infty}\exp(\rho^i v_{t-i}).$$

Here L denotes the lag operator as usual. Define $\boldsymbol{y} = (y_1 \ldots y_T)'$ and $\boldsymbol{h} = (h_1 \ldots h_T)'$. Then we have $\boldsymbol{y}|\boldsymbol{h} \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = diag(exp(h_1) \ldots exp(h_T))$. So we get

$$P(\boldsymbol{y}|\boldsymbol{h};\ \theta) = (2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}exp\left(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}\right)$$

where

$$\boldsymbol{y}'\boldsymbol{\Omega}^{-1}\boldsymbol{y} = \sum_{t=1}^{T} y_t^2\exp(-h_t).$$

Then we have

$$logP(\boldsymbol{y}|\boldsymbol{h};\ \theta) = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} h_t(\theta) - \frac{1}{2}\sum_{t=1}^{T} y_t^2 \exp\bigl(-h_t(\theta)\bigr)$$

hence

$$\frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta} = -\frac{1}{2}\sum_{t=1}^{T}\frac{\partial h_t}{\partial\theta} + \frac{1}{2}\sum_{t=1}^{T} y_t^2 \exp(-h_t)\frac{\partial h_t}{\partial\theta}\ . \tag{1.3}$$

Here log denotes the natural logarithm as usual. The Quasi Maximum Likelihood estimator (in short, QMLE) $\widehat{\theta_T} = \left(\widehat{\mu_T} \quad \widehat{\rho_T} \quad \widehat{\sigma_{\nu T}^2}\right)'$ is any measurable solution of

$$\widehat{\theta_T} = \underset{\theta\in\Theta}{\operatorname{argmax}} - \frac{1}{2T}\sum_{t=1}^{T} h_t(\theta) - \frac{1}{2T}\sum_{t=1}^{T} y_t^2 \exp\bigl(-h_t(\theta)\bigr).$$

The paper is organized as follows. In Section 2 we describe the asymptotic properties of the QML estimators for Model (1.1). In Section 3 we review some results on the approximation of the likelihood function from Fridman and Harris (1998) and Bartolucci and De Luca (2001) (2003), and use them to describe the asymptotic variance of the QML estimators. Section 4 is devoted to give an explicit computation of the asymptotic variance matrix of the QML estimators by using the frequency-domain approach. Section 5 concludes.

2 – Consistency and Asymptotic Normality

The following are the main results in this section:

Theorem 1. *For* $\Theta \subset (0,+\infty)^3$ *such that for every* $\theta \in \Theta$, $|\rho| < 1$, *then*

$$\lim_{T\to+\infty}\widehat{\theta_T} = \theta_0 \quad (a.s.)$$

and

$$\sqrt{T}\left(\widehat{\theta_T} - \theta_0\right) \to N(0, 2\mathcal{J}^{-1}) \qquad as \qquad T \to +\infty$$

where

$$\mathcal{J} = E_\infty\left(\frac{\partial h_t(\theta)}{\partial \theta}|_{\theta=\theta_0}\frac{\partial h_t(\theta)}{\partial \theta'}|_{\theta=\theta_0}\right)$$

is a positive constant symmetric matrix.

Theorem 2. *Under the assumptions of Theorem 1, the asymptotic score matrix is given by*

$$E_\infty\left(\frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta_0}\right) = -\frac{1}{2}\mathcal{J},$$

where $\mathcal{L}(\theta) = P(\boldsymbol{y}|\theta)$ *is the likelihood function. Then* $\widehat{\theta_T}$ *is the maximizer of the objective function for T sufficiently large.*

The Likelihood function is given by

$$\begin{aligned}\mathcal{L}(\theta) &= P(\boldsymbol{y}|\theta)\\ &= \int_{\boldsymbol{h}} P(\boldsymbol{y},\boldsymbol{h}|\theta)\, d\boldsymbol{h} = \int_{\boldsymbol{h}} P(\boldsymbol{y}|\boldsymbol{h};\ \theta)\, P(\boldsymbol{h}|\ \theta)d\boldsymbol{h}\\ &= E_H[P(\boldsymbol{y}|\boldsymbol{h};\ \theta)].\end{aligned}$$

Then we have

$$\begin{aligned}\frac{\partial log\mathcal{L}(\theta)}{\partial\theta} &= \frac{1}{\mathcal{L}(\theta)}\frac{\partial\mathcal{L}(\theta)}{\partial\theta}\\ &= \frac{1}{\mathcal{L}(\theta)}\int_{\boldsymbol{h}} \frac{\partial P(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}\, P(\boldsymbol{h}|\ \theta)d\boldsymbol{h}\\ &+ \frac{1}{\mathcal{L}(\theta)}\int_{\boldsymbol{h}} P(\boldsymbol{y}|\boldsymbol{h};\ \theta)\frac{\partial P(\boldsymbol{h}|\ \theta)}{\partial\theta}\, d\boldsymbol{h}\\ &= E_H\left[\frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}|\boldsymbol{y},\theta\right]\\ &+ E_H\left[\frac{\partial logP(\boldsymbol{h}|\ \theta)}{\partial\theta}|\boldsymbol{y},\theta\right] \qquad (2.1)\end{aligned}$$

See Appendix I. The First Order Condition (FOC) is given by

$$\frac{\partial log\mathcal{L}(\theta)}{\partial\theta}|_{\theta=\widehat{\theta_T}} = 0. \qquad (2.2)$$

The asymptotic score is given by

$$\begin{aligned}\lim_{T\to+\infty}\frac{1}{T}\frac{\partial log\mathcal{L}(\theta)}{\partial\theta} \\ &= \lim_{T\to+\infty}\frac{1}{T}E_H\left[\sum_{t=1}^{T}\frac{\partial logP(y_t|h_t;\,\theta)}{\partial\theta}|\boldsymbol{y},\theta\right] \\ &+ \lim_{T\to+\infty}\frac{1}{T}E_H\left[\sum_{t=1}^{T}\frac{\partial logP(h_t|\,h_{t-1},\theta)}{\partial\theta}|\boldsymbol{y},\theta\right] \\ = E_YE_H\left[\frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\,\theta)}{\partial\theta}|\boldsymbol{y},\theta\right] + E_YE_H\left[\frac{\partial logP(\boldsymbol{h}|\,\theta)}{\partial\theta}|\boldsymbol{y},\theta\right] \\ &= E_H\left[\frac{\partial logP(\boldsymbol{y}|\boldsymbol{h})}{\partial\theta}\right] + E_H\left[\frac{\partial logP(\boldsymbol{h}|\theta)}{\partial\theta}\right]\end{aligned}$$

by using the double expectation. This relation is based on the Law of Large Numbers from which it follows that $\frac{1}{T}\sum_{t=1}^{T}\nu(y_t)$, with $\boldsymbol{y}\sim P(\boldsymbol{y}|\theta)$, converges almost surely to

$$E_Y[\nu(\boldsymbol{y})] = \int_{\boldsymbol{h}} \nu(\boldsymbol{y})P(\boldsymbol{y}|\,\theta)\,d\boldsymbol{y}$$

as T goes to infinity. Then we can write

$$E_\infty[\nu(y_t)] = \lim_{T\to+\infty}\frac{1}{T}\sum_{t=1}^{T}\nu(y_t) = E_Y[\nu(\boldsymbol{y})].$$

Thus we get

$$\begin{aligned}\lim_{T\to+\infty}\frac{1}{T}\frac{\partial log\mathcal{L}(\theta)}{\partial\theta} &= E_H\left[\frac{\partial logP(\boldsymbol{y}|\boldsymbol{h})}{\partial\theta}\right] \\ &= \lim_{T\to+\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial logP(y_t|h_t;\,\theta)}{\partial\theta} \qquad (2.3)\end{aligned}$$

as

$$E_H\left[\frac{\partial logP(\boldsymbol{h}|\theta)}{\partial\theta}\right]=0.$$

In fact, from

$$\int_{\boldsymbol{h}} P(\boldsymbol{h}|\theta)\,d\boldsymbol{h}=1$$

we get

$$\int_{\boldsymbol{h}} \frac{\partial P(\boldsymbol{h}|\theta)}{\partial\theta}\,d\boldsymbol{h}=0$$

hence

$$E_H\left[\frac{\partial logP(\boldsymbol{h}|\theta)}{\partial\theta}\right]=\int_{\boldsymbol{h}} \frac{\partial logP(\boldsymbol{h}|\theta)}{\partial\theta}P(\boldsymbol{h}|\theta)\,d\boldsymbol{h}=\int_{\boldsymbol{h}} \frac{\partial P(\boldsymbol{h}|\theta)}{\partial\theta}d\boldsymbol{h}=0\,.$$

From (2.2) and (2.3) we obtain

$$\lim_{T\to+\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial logP(y_t|h_t;\ \theta)}{\partial\theta}\ |_{\theta=\widehat{\theta_T}}=0. \qquad (2.4)$$

Using (1.3) we have

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial logP(y_t|h_t;\ \theta)}{\partial\theta}\ |_{\theta=\widehat{\theta_T}}$$
$$=-\frac{1}{2T}\sum_{t=1}^{T}[1-y_t^2\exp\{-h_t(\widehat{\theta_T})\}]\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\widehat{\theta_T}}.$$

Taking the 1st Taylor expansions around θ_0 of $1-y_t^2\exp\{-h_t(\widehat{\theta_T})\}$ and $\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\widehat{\theta_T}}$ and using the relation $y_t^2=\exp\{h_t(\theta_0)\}u_t^2$, we get

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial logP(y_t|h_t;\ \theta)}{\partial\theta}\ |_{\theta=\widehat{\theta_T}}$$
$$= -\frac{1}{2}\mathcal{I}_T\left(\widehat{\theta_T}-\theta_0\right)+O_p(1)$$
$$-\frac{1}{2T}\sum_{t=1}^{T}(1-u_t^2)\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\theta_0} \qquad (2.5)$$

where

$$\mathcal{I}_T=\frac{1}{T}\sum_{t=1}^{T}\left\{(1-u_t^2)\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\ |_{\theta=\theta_0}\ +u_t^2\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\theta_0}\ \frac{\partial h_t(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}\right\}.$$

Then we have

$$\mathcal{I}=\lim_{T\to+\infty}\mathcal{I}_T=E_\infty\left(\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\theta_0}\ \frac{\partial h_t(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}\right)<+\infty$$

as $E_\infty(u_t^2)=1$ and u_t is independent of $h_t(\theta)$ (and its derivatives). In fact, the latter is a function of v_t which is independent of u_t by hypothesis. Since $|\rho|<1$, the matrix $\mathcal{I}<+\infty$. Thus taking the limit for $T\to+\infty$ in (2.5) and using (2.4) gives the consistency of $\widehat{\theta_T}$, that is,

$$\lim_{T\to+\infty}\left(\widehat{\theta_T}-\theta_0\right)=-\mathcal{I}^{-1}\lim_{T\to+\infty}\frac{1}{T}\sum_{t=1}^{T}(1-u_t^2)\frac{\partial h_t(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}$$
$$=-\mathcal{I}^{-1}\,E_\infty(1-u_t^2)E_\infty\left(\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\theta_0}\right)=0$$

as $E_\infty(1-u_t^2)=0$. For the asymptotic variance of $\widehat{\theta_T}$, we have

$$\left(\widehat{\theta_T}-\theta_0\right)\left(\widehat{\theta_T}-\theta_0\right)'=\mathcal{I}_T^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}(1-u_t^2)\frac{\partial h_t(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}\right]\times$$

$$\left[\frac{1}{T}\sum_{t=1}^{T}(1-u_\tau^2)\frac{\partial h_\tau(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}\ \right]\mathcal{J}_T^{-1}$$

for T sufficiently large. Thus

$$var_\infty\left(\widehat{\theta_T}\right) = \lim_{T\to+\infty} E_T\left(T\left(\widehat{\theta_T}-\theta_0\right)\left(\widehat{\theta_T}-\theta_0\right)'\right) =$$

$$= \mathcal{J}^{-1}E_\infty((1-u_t^2)^2)\ E_\infty\left(\frac{\partial h_t(\theta)}{\partial\theta}\ |_{\theta=\theta_0}\ \frac{\partial h_t(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}\ \right)\mathcal{J}^{-1}$$
$$= 2\mathcal{J}^{-1}\mathcal{J}\mathcal{J}^{-1} = 2\mathcal{J}^{-1}$$

as $E_\infty((1-u_t^2)^2) = 1 + E_\infty(u_t^4) - 2E_\infty(u_t^2) = 2$ which is the statement of Theorem 1. We remark that, for T sufficiently large, the matrix $\mathcal{J}_T$ is invertible, so we can write

$$\widehat{\theta_T}\sim\theta_0 + \mathcal{J}_T^{-1}\left[-\frac{1}{T}\sum_{t=1}^{T}(1-u_t^2)\frac{\partial h_t(\theta)}{\partial\theta'}\ |_{\theta=\theta_0}\ \right]$$

From (1.3) we get

$$\frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'} = -\frac{1}{2}\ \sum_{t=1}^{T}\frac{\partial^2 h_t}{\partial\theta\partial\theta'} - \frac{1}{2}\sum_{t=1}^{T} y_t^2\exp(-h_t)\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta'}$$
$$+\frac{1}{2}\sum_{t=1}^{T} y_t^2\exp(-h_t)\frac{\partial^2 h_t}{\partial\theta\partial\theta'}\ .$$

Using the relation $y_t^2 = \exp\{h_t(\theta_0)\}u_t^2$, we obtain

$$\frac{1}{T}\frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'}|_{\theta=\widehat{\theta}_T}$$
$$= -\frac{1}{2T}\sum_{t=1}^{T}[1 - u_t^2 \exp\{h_t\,(\theta_0) - h_t(\widehat{\theta_T})\}]\frac{\partial^2 h_t}{\partial\theta\partial\theta'}|_{\theta=\widehat{\theta}_T}$$
$$-\frac{1}{2T}\sum_{t=1}^{T} u_t^2 \exp\{h_t(\theta_0) - h_t(\widehat{\theta_T})\}\frac{\partial h_t}{\partial\theta}|_{\theta=\widehat{\theta}_T}\frac{\partial h_t}{\partial\theta'}|_{\theta=\widehat{\theta}_T}$$

Using Theorem 1 and taking the limit for $T \to +\infty$ gives

$$E_\infty\left(\frac{\partial^2 log\mathcal{L}(\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta_0}\right)$$
$$= -\frac{1}{2}E_\infty(1-u_t^2)E_\infty\left(\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}|_{\theta=\theta_0}\right)$$
$$-\frac{1}{2}E_\infty(u_t^2)E_\infty\left(\frac{\partial h_t}{\partial\theta}|_{\theta=\theta_0}\frac{\partial h_t}{\partial\theta'}|_{\theta=\theta_0}\right)$$
$$= -\frac{1}{2}E_\infty\left(\frac{\partial h_t}{\partial\theta}|_{\theta=\theta_0}\frac{\partial h_t}{\partial\theta'}|_{\theta=\theta_0}\right) = -\frac{1}{2}\mathcal{J}$$

which is negative definite. Thus $\widehat{\theta}_T$ is a maximizer of the objective function for T sufficiently large. This proves Theorem 2.

3 – Approximated Likelihood

ML estimation of dynamic latent variable models requires the integration of the latent volatility out of the joint density of $\boldsymbol{y} = (y_1 \dots y_T)'$ and $\boldsymbol{h} = (h_1 \dots h_T)'$**:**

$$\mathcal{L}(\theta) = \int_{\boldsymbol{h}} P(\boldsymbol{y},\boldsymbol{h}|\theta)\,d\boldsymbol{h} = \int_{\boldsymbol{h}}\left[\prod_{t=1}^{T}P(y_t, h_t|h_{t-1};\,\theta)\right]d\boldsymbol{h}$$

where $P(y_t, h_t|h_{t-1};\,\theta)$ is the joint density of y_t and h_t given h_{t-1} and θ. The T-dimensional integral $\mathcal{L}(\theta)$ cannot be factorized into a product of T one-dimensional integrals because the dependence of h_t on the past. However, it can be evaluated using an iterated numeric integration procedure. See Fridman and Harris (1998) and Bartolucci and De Luca (2001) (2003). According to these authors, one can approximate $\mathcal{L}(\theta)$ with the required precision through the function

$$\tilde{\mathcal{L}}(\theta) = a_T \sum_{i_1=1}^{n} \left[\sum_{i_2=1}^{n} \dots \left[\sum_{i_T=1}^{n} P\left(y_T, x_{i_T} \middle| x_{i_{T-1}}, \theta\right) \right] \dots P\left(y_2, x_{i_2} \middle| x_{i_1}, \theta\right) \right] P\left(y_1, x_{i_1} \middle| h_0, \theta\right)$$

where $\{x_k\}, k = 1, \dots, n,$ is a set of quadrature points chosen as $x_k = x_1 + (k-1)a_T$ and $a_T = (x_n - x_1)/(n-1)$. Obviously, the higher n and the wider the interval $[x_1, x_n]$, the better the accuracy of the approximation. In order to facilitate the computation of $\tilde{\mathcal{L}}(\theta)$, it is convenient to use a more compact form. Let

$$\tilde{\mathcal{L}}(\theta) = a_T \mathbf{1}' \boldsymbol{F}_T \boldsymbol{F}_{T-1} \cdots \boldsymbol{F}_2 \boldsymbol{f}_1$$

where $\boldsymbol{f}_1$ is the $n \times 1$ vector with entries $P(y_1, x_i | h_0), i = 1, \dots, n,$ $\boldsymbol{F}_T$ is the $n \times n$ matrix with entries $P\left(y_t, x_i \middle| x_j\right), i, j = 1, \dots, n,$ and $\mathbf{1}$ is the $n \times 1$ vector of ones. The approximated likelihood $\tilde{\mathcal{L}}(\theta)$ may be expressed as

$$\tilde{\mathcal{L}}(\theta) = a_T \mathbf{1}' \boldsymbol{f}_T$$

where $\boldsymbol{f}_t, t = 2, \dots, T,$ is defined through the recursion

$$\boldsymbol{f}_t = \boldsymbol{F}_t \boldsymbol{f}_{t-1}$$

that may be easily evaluated. See the quoted papers. Using the last recursive representation, it is possible to compute the first and second derivatives of $\tilde{\mathcal{L}}(\theta)$ with respect to θ, that is,

$$\frac{\partial \tilde{\mathcal{L}}(\theta)}{\partial \theta_i} = a_T \mathbf{1}' \frac{\partial \boldsymbol{f}_T}{\partial \theta_i} \qquad \qquad \frac{\partial^2 \tilde{\mathcal{L}}(\theta)}{\partial \theta_i \partial \theta_j} = a_T \mathbf{1}' \frac{\partial^2 \boldsymbol{f}_T}{\partial \theta_i \partial \theta_j}$$

for $i, j = 1,2,3$. The derivatives of $\boldsymbol{f}_t$ may be also computed in a recursive way

$$\frac{\partial \boldsymbol{f}_t}{\partial \theta_i} = \frac{\partial \boldsymbol{F}_t}{\partial \theta_i} \boldsymbol{f}_{t-1} + \boldsymbol{F}_t \frac{\partial \boldsymbol{f}_{t-1}}{\partial \theta_i}$$

$$\frac{\partial^2 \boldsymbol{f}_t}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 \boldsymbol{F}_t}{\partial \theta_i \partial \theta_j} \boldsymbol{f}_{t-1} + \frac{\partial \boldsymbol{F}_t}{\partial \theta_i} \frac{\partial \boldsymbol{f}_{t-1}}{\partial \theta_j} + \frac{\partial \boldsymbol{F}_t}{\partial \theta_j} \frac{\partial \boldsymbol{f}_{t-1}}{\partial \theta_i} + \boldsymbol{F}_t \frac{\partial^2 \boldsymbol{f}_{t-1}}{\partial \theta_i \partial \theta_j}$$

for $t = 2, \ldots, T$ and $i, j = 1,2,3$. The parameter vector that maximizes $\tilde{\mathcal{L}}(\theta)$ is obtained after running the Newton-Raphson algorithm implemented on the basis of such derivatives. It is an approximated ML estimate of θ and will be denoted as $\widetilde{\theta_T}$. The following result was given in Bartolucci and De Luca (2001) (2003).

Theorem 3. *With the above notation, an estimator of the covariance matrix* $V_{\widetilde{\theta_T}}$ *of* $\widetilde{\theta_T}$ *may be obtained as*

$$\hat{V}_{\widetilde{\theta_T}} = \left[- \frac{\partial^2 log \tilde{\mathcal{L}}(\theta)}{\partial \theta \partial \theta'} |_{\theta = \widetilde{\theta_T}} \right]^{-1} .$$

Using Theorem 2 and Theorem 3 we can give an approximated description of the asymptotic score matrix $\mathcal{J}$:

Corollary 4. *An asymptotic estimator of the matrix* $\mathcal{J}$ *may be obtained as*

$$E_\infty \left[- \frac{\partial^2 log \tilde{\mathcal{L}}(\theta)}{\partial \theta \partial \theta'} |_{\theta = \theta_0} \right] = \frac{1}{2} \hat{\mathcal{J}}$$

or, equivalently,

$$Var_\infty \left(\widetilde{\theta_T} \right) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \hat{V}_{\widetilde{\theta_T}} = E_\infty \left[- \frac{\partial^2 log \tilde{\mathcal{L}}(\theta)}{\partial \theta \partial \theta'} |_{\theta = \theta_0} \right]^{-1} = 2 \mathcal{J}^{-1}$$

$$= Var_\infty \left(\hat{\theta}_T \right)$$

4 – Explicit Computation of the Asymptotic Variance Matrix

Following Ruiz (1994) and Breidt and Carriquiry (1996) an estimation method for the parameter vector $\theta = (\mu \quad \rho \quad \sigma_v^2)'$ is given by linearizing the model in (1.1) and transforming the linearized process into an ARMA(1,1) with one-to-one local mappings between the parameter vectors. Squaring (1.1) and taking logs, we get the state space model

$$\begin{aligned} x_t &= \alpha + h_t + e_t \\ h_t &= \mu + \rho h_{t-1} + v_t \end{aligned} \tag{4.1}$$

where $x_t = log y_t^2,\ \alpha = E[log\, u_t^2]$ is a real constant ($\alpha \cong -1.27$) and $e_t = log u_t^2 - \alpha$ is a non-Gaussian zero mean white noise with variance $\sigma_e^2 = \pi^2/2.$ The higher moments of (e_t) are known. See, for example, the quoted paper. Model (4.1) has an ARMA(1,1) reduced form

$$(1 - \rho L)x_t = \xi + (1 + \beta L)w_t \tag{4.2}$$

where

$$\xi = (1 - \rho)\alpha + \mu \qquad and \qquad v_t = -e_t + \rho e_{t-1} + w_t + \beta w_{t-1}.$$

Then we have $w_t \sim IID(0, \sigma_w^2)$. From Breidt and Carriquiry (1996), this implies the following locally one-to-one mappings

$$\sigma_v^2 = (1 + \rho^2)\sigma_e^2 + (1 + \beta^2)\sigma_w^2 \tag{4.3}$$

and

$$\sigma_w^2 = -\rho\sigma_e^2\beta^{-1} \tag{4.4}$$

hence $\beta < 0$. Substituting (4.4) into (4.3) and then multiplying by β, we get

$$\rho\sigma_e^2\beta^2 + [\sigma_v^2 - (1 + \rho^2)\sigma_e^2]\beta + \rho\sigma_e^2 = 0$$

which gives

$$\beta = \frac{-\sigma_v^2 + (1 + \rho^2)\sigma_e^2 + \sqrt{\Delta}}{2\rho\sigma_e^2} \tag{4.5}$$

where

$$\Delta = [\sigma_v^2 - (1+\rho^2)\sigma_e^2]^2 - 4\rho^2\sigma_e^4 \tag{4.6}$$

Here we assume $\delta \geq 0$. Note that $-\sqrt{\Delta}$ in (4.5) is not acceptable since we set $-1 < \beta < 0$ to ensure invertibility of the process. An important reason for imposing the invertibility condition is to get identifiability. From (4.4) and (4.5) we obtain

$$\sigma_w^2 = \frac{\sigma_v^2 - (1+\rho^2)\sigma_e^2 + \sqrt{\Delta}}{2}.$$

Model (4.1) can be written as

$$\begin{aligned} x_t^* &= h_t^* + e_t \\ h_t^* &= \rho h_{t-1}^* + v_t \end{aligned} \tag{4.7}$$

where $x_t^* = x_t - E(x_t)$, $h_t^* = h_t - E(h_t)$, $E(x_t) = \alpha + E(h_t)$ and $E(h_t) = \mu(1-\rho)^{-1}$. As remarked in Ruiz (1994), p.292, $\mu^* = E(x_t)$ can be consistenly estimated by the sample mean of $x_t = log y_t^2$, that is,

$$\hat{\mu}^* = \frac{1}{T}\sum_{t=1}^{T} \log \quad y_t^2 .$$

Thus, given a QML estimation $\hat{\rho}_{ML}$ of ρ, we get a QML estimation of μ by the formula

$$\hat{\mu}_{ML} = (\hat{\mu}^* + 1.27)(1 - \hat{\rho}_{ML}) .$$

Moreover, $\hat{\mu}^*$ is a QML estimator of μ^* which is uncorrelated with the QML estimator of the stochastic part of the model. See Harvey (1989) and Ruiz (1994). Thus, in what follows, we treat μ^* (or, equivalently, μ) as known and concentrate on the estimation of the parameter vector $(\rho \quad \sigma_v^2)'$, also denoted by θ. Using (4.7), the ARMA(1,1) process in (4.2) becomes

$$(1-\rho L)x_t^* = (1+\beta L)w_t \tag{4.8}$$

with unknown parameter vector $\theta_* = (\rho \quad \beta)'$. Note that σ_w^2 is given by (4.4). Let $\hat{\theta}_*$ be a QML estimator of θ_*. The asymptotic covariance matrix V_θ of $\hat{\theta}$ can be found by calculating the asymptotic covariance matrix V_{θ_*} of $\hat{\theta}_*$ and applying the delta method to the map $\theta = \theta(\theta_*)$. We can express σ_v^2 in terms of ρ and β by using (4.3) and (4.4). Then we have

$$\sigma_v^2 = (1+\rho^2)\sigma_e^2 - \rho\sigma_e^2\beta^{-1} - \rho\sigma_e^2\beta \,. \tag{4.9}$$

Then we get

$$J = \frac{\partial\theta}{\partial\theta_*'} = \begin{pmatrix} 1 & 0 \\ \sigma_e^2(2\rho - \beta - \beta^{-1}) & \rho\sigma_e^2(\beta^{-2} - 1) \end{pmatrix}$$

where β is given in (4.5) and (4.6). Then V_θ is obtained as $JV_{\theta_*}J'$. From Dunsmuir (1979) and Ruiz (1994), the asymptotic distribution of the QML estimator $\hat{\theta}_*$ of θ_* from Model (4.8) is given by

$$\sqrt{T}\left(\hat{\theta}_* - \theta_*\right) \xrightarrow{d} N(\mathbf{0}, V_{\theta_*})$$

where T is the sample size and the analytic expression of V_{θ_*} is derived in the Appendix II. So we have

Theorem 5. *The asymptotic variance matrix V_{θ_*} of the QML estimator $\hat{\theta}_*$ of $\theta_* = (\rho \quad \beta)'$ from Model (4.8) is given by*

$$V_{\theta_*} = 2A^{-1} + A^{-1}BA^{-1}$$

where

$$A = \begin{pmatrix} \dfrac{1+\rho^2}{\rho^2(1-\rho^2)} & \dfrac{\rho\beta - 1}{\rho\beta(\rho\beta+1)} \\ \dfrac{\rho\beta - 1}{\rho\beta(\rho\beta+1)} & \dfrac{1+\beta^2}{\beta^2(1-\beta^2)} \end{pmatrix}$$

and

$$B = \boldsymbol{k}\begin{pmatrix} \frac{1}{\rho^2} & -\frac{1}{\rho\beta} \\ -\frac{1}{\rho\beta} & \frac{1}{\beta^2} \end{pmatrix}.$$

Here $\boldsymbol{k}$ is the measure of excess kurtosis of the reduced form disturbance term, that is,

$$\boldsymbol{k} = \frac{E(w_t^4)}{\sigma_w^4} - 3 = \frac{4(1+\rho^4)(1+\beta^2)^2\sigma_e^4}{[\sigma_v^2 + (1+\rho^2)\sigma_e^2]^2(1+\beta^4)}$$

where σ_v^2 is given by (4.9).

5 – Conclusions

Stochastic Volatility (SV) models are highly applied to many financial applications but their estimation is notoriously difficult. Here we consider a Quasi Maximum Likelihood (QML) approach and we prove the asymptotic theory of QML estimators. In particular, we are able to compute the variance-covariance matrix of the QML estimators in an explicit form, opening up the possibility to use classical testing procedure in future empirical applications.

Appendix I

For the first summand in (2.1), we have

$$\frac{1}{\mathcal{L}(\theta)}\int_{\boldsymbol{h}} \frac{\partial P(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}\ P(\boldsymbol{h}|\ \theta)d\boldsymbol{h}$$
$$= \frac{1}{P(\boldsymbol{y}|\ \theta)}\int_{\boldsymbol{h}} \frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}\ P(\boldsymbol{y}|\ \boldsymbol{h},\theta)P(\boldsymbol{h}|\ \theta)d\boldsymbol{h}$$
$$= \frac{1}{P(\boldsymbol{y}|\ \theta)}\int_{\boldsymbol{h}} \frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}\ P(\boldsymbol{y},\boldsymbol{h}|\theta)d\boldsymbol{h}$$
$$= \int_{\boldsymbol{h}} \frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}\ \frac{P(\boldsymbol{y},\boldsymbol{h}|\theta)}{P(\boldsymbol{y}|\ \theta)}d\boldsymbol{h}$$
$$= \int_{\boldsymbol{h}} \frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}P(\boldsymbol{h}|\boldsymbol{y};\ \theta)\ d\boldsymbol{h} = E_H\left[\frac{\partial logP(\boldsymbol{y}|\boldsymbol{h};\ \theta)}{\partial\theta}|\boldsymbol{y},\theta\right]\ .$$

For the second summand in (2.1), we have

$$\frac{1}{\mathcal{L}(\theta)}\int_{\boldsymbol{h}}\ P(\boldsymbol{y}|\boldsymbol{h};\ \theta)\frac{\partial P(\boldsymbol{h}|\ \theta)}{\partial\theta}\ d\boldsymbol{h}$$
$$= \frac{1}{P(\boldsymbol{y}|\ \theta)}\int_{\boldsymbol{h}}\ P(\boldsymbol{y}|\boldsymbol{h};\ \theta)P(\boldsymbol{h}|\ \theta)\frac{\partial logP(\boldsymbol{h}|\ \theta)}{\partial\theta}\ d\boldsymbol{h}$$
$$= \frac{1}{P(\boldsymbol{y}|\ \theta)}\int_{\boldsymbol{h}}\ P(\boldsymbol{y},\boldsymbol{h}|\theta)\frac{\partial logP(\boldsymbol{h}|\ \theta)}{\partial\theta}\ d\boldsymbol{h} = \int_{\boldsymbol{h}}\ \frac{P(\boldsymbol{y},\boldsymbol{h}|\theta)}{P(\boldsymbol{y}|\ \theta)}\frac{\partial logP(\boldsymbol{h}|\ \theta)}{\partial\theta}\ d\boldsymbol{h}$$
$$= \int_{\boldsymbol{h}}\ P(\boldsymbol{h}|\boldsymbol{y},\theta)\frac{\partial logP(\boldsymbol{h}|\ \theta)}{\partial\theta}\ d\boldsymbol{h}$$
$$= E_H\left[\frac{\partial logP(\boldsymbol{h}|\ \theta)}{\partial\theta}\ |\boldsymbol{y},\theta\right]\ .$$

Appendix II

The asymptotic variance matrix of the QML estimator θ_* is given by

$$V_{\theta_*} = 2A^{-1} + A^{-1}BA^{-1}$$

where

$$A = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \theta_*} \frac{\partial \log \quad g(\lambda)}{\partial \theta'_*} d\lambda$$

and

$$B = \boldsymbol{k}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \theta_*} d\lambda\right]\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \theta'_*} d\lambda\right]$$

where $\boldsymbol{k}$ is the measure of excess kurtosis of the reduced form disturbance term, $g(\lambda)$ is the spectral generating function from Model (4.8). See Harvey (1989), Section 4.5.5, for the conditions under which this result works. Here we use a method from Ruiz (1994) received as personal communication and adapted to our case. Then we have

$$g(\lambda) = \sigma_w^2 \frac{1 + \beta^2 + 2\beta\cos(\lambda)}{1 + \rho^2 - 2\rho\cos(\lambda)} = \sigma_e^2 \frac{-\beta^{-1} - \beta - 2\cos(\lambda)}{\rho^{-1} + \rho - 2\cos(\lambda)}$$

by using (4.4). So we get

$$\log \mathrm{g}(\lambda) = log\sigma_e^2 + log\left[\frac{-\beta^{-1} - \beta - 2\cos(\lambda)}{\rho^{-1} + \rho - 2\cos(\lambda)}\right]$$

hence

$$\frac{\partial \log \quad g(\lambda)}{\partial \rho} = \frac{1 - \rho^2}{\rho + \rho^3 - 2\rho^2\cos(\lambda)}$$

and

$$\frac{\partial \log \quad g(\lambda)}{\partial \beta} = \frac{1-\beta^2}{-\beta-\beta^3-2\beta^2\cos(\lambda)} \ .$$

(I). *Computation of the matrix* A.

To compute the elements of the matrix $A = (a_{ij})_{i,j=1,2}$, we need the following integrals

$$\int_{-\pi}^{\pi}\left[\frac{\partial \log \quad g(\lambda)}{\partial \rho}\right]^2 d\lambda = \int_{-\pi}^{\pi}\frac{(1-\rho^2)^2}{[a-bcos(\lambda)]^2}d\lambda \qquad (I.1)$$

$$\int_{-\pi}^{\pi}\frac{\partial \log \quad g(\lambda)}{\partial \rho}\frac{\partial \log \quad g(\lambda)}{\partial \beta}d\lambda = \int_{-\pi}^{\pi}\frac{(1-\rho^2)(1-\beta^2)}{[a-bcos(\lambda)][c-dcos(\lambda)]}d\lambda \qquad (I.2)$$

and

$$\int_{-\pi}^{\pi}\left[\frac{\partial \log \quad g(\lambda)}{\partial \beta}\right]^2 d\lambda = \int_{-\pi}^{\pi}\frac{(1-\beta^2)^2}{[c-dcos(\lambda)]^2}d\lambda \qquad (I.3)$$

where $a=\rho+\rho^3, b=2\rho^2, c=-\beta-\beta^3$ and $d=2\beta^2$. For (I.1) we have

$$\int_{-\pi}^{\pi}\left[\frac{\partial \log \quad g(\lambda)}{\partial \rho}\right]^2 d\lambda = 2(1-\rho^2)^2\int_0^{\pi}\frac{1}{[a-bcos(\lambda)]^2}d\lambda$$
$$= 2(1-\rho^2)^2\frac{\rho(1+\rho^2) \quad \pi}{\rho^3(1-\rho^2)^3} = \frac{2\pi(1+\rho^2)}{\rho^2(1-\rho^2)}$$

hence

$$a_{11} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\partial \log \quad g(\lambda)}{\partial \rho}\right]^2 d\lambda = \frac{1+\rho^2}{\rho^2(1-\rho^2)} \quad .$$

For (I.3) we can repeat the above arguments. So we get

$$a_{22} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\partial \log \quad g(\lambda)}{\partial \beta}\right]^2 d\lambda = \frac{1+\beta^2}{\beta^2(1-\beta^2)} \quad .$$

For (I.2) we have

$$\int_{-\pi}^{\pi}\frac{\partial \log \quad g(\lambda)}{\partial \rho}\frac{\partial \log \quad g(\lambda)}{\partial \beta}d\lambda$$
$$= (1-\rho^2)(1-\beta^2)\int_{-\pi}^{\pi}\frac{1}{[a-bcos(\lambda)][c-dcos(\lambda)]}d\lambda$$
$$= (1-\rho^2)(1-\beta^2)\left[M\int_{-\pi}^{\pi}\frac{1}{a-bcos(\lambda)}d\lambda + N\int_{-\pi}^{\pi}\frac{1}{c-dcos(\lambda)}d\lambda\right]$$
$$= (1-\rho^2)(1-\beta^2)\left[\frac{2\pi M}{\rho(1-\rho^2)} - \frac{2\pi N}{\beta(1-\beta^2)}\right]$$
$$= \frac{2\pi[M\beta(1-\beta^2) - N\rho(1-\rho^2)]}{\rho\beta}$$

where

$$M = \frac{-\rho}{\beta(1+\beta\rho)(\beta+\rho)} \qquad N = \frac{\beta}{\rho(1+\beta\rho)(\beta+\rho)} \quad .$$

Then we get

$$a_{12} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \rho} \frac{\partial \log \quad g(\lambda)}{\partial \beta} d\lambda = \frac{\rho\beta - 1}{\rho\beta(1+\rho\beta)} \quad .$$

(II). *Computation of the matrix* B.

To compute the elements of the matrix $B = \boldsymbol{k}(b_{ij})_{i,j=1,2}$, we need the following integrals

$$\begin{aligned}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \rho} d\lambda &= \int_{-\pi}^{\pi} \frac{(1-\rho^2)}{\rho(1+\rho^2) - 2\rho^2 cos(\lambda)} d\lambda \\ &= \frac{2(1-\rho^2)}{\rho}\int_0^{\pi} \frac{1}{1+\rho^2 - 2\rho \quad cos(\lambda)} d\lambda \\ &= \frac{2(1-\rho^2)}{\rho}\frac{\pi}{1-\rho^2} \\ &= \frac{2\pi}{\rho} \qquad (II.1)\end{aligned}$$

and

$$\begin{aligned}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \beta} d\lambda &= \int_{-\pi}^{\pi} \frac{(1-\beta^2)}{-\beta(1+\beta^2) - 2\beta^2 cos(\lambda)} d\lambda \\ &= -\frac{2\pi}{\beta} \quad . \qquad (II.2)\end{aligned}$$

Then we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial \log \quad g(\lambda)}{\partial \theta_*} d\lambda = \begin{pmatrix} \frac{1}{\rho} \\ -\frac{1}{\beta} \end{pmatrix}$$

hence

$$B = \boldsymbol{k}\begin{pmatrix} \frac{1}{\rho} \\ -\frac{1}{\beta} \end{pmatrix}\begin{pmatrix} \frac{1}{\rho} & -\frac{1}{\beta} \end{pmatrix} = \boldsymbol{k}\begin{pmatrix} \frac{1}{\rho^2} & -\frac{1}{\rho\beta} \\ -\frac{1}{\rho\beta} & \frac{1}{\beta^2} \end{pmatrix} \ .$$

(III). *Computation of excess kurtosis $\boldsymbol{k}$.*

Let $\boldsymbol{k_1}$ denote the kurtosis of $(1-\rho L)x_t^*$ and $\boldsymbol{k}$ the excess kurtosis of the reduced form disturbances, that is,

$$\boldsymbol{k} = \frac{E(w_t^4)}{\sigma_w^4} - 3 \ .$$

From (4.8) we get

$$\boldsymbol{k_1} = \frac{E(x_t^* - \rho x_{t-1}^*)^4}{[E(x_t^* - \rho x_{t-1}^*)^2]^2} = \frac{E(w_t + \beta w_{t-1})^4}{[E(w_t + \beta w_{t-1})^2]^2} = \frac{(1+\beta^4)E(w_t^4) + 6\beta^2\sigma_w^4}{(1+\beta^2)^2\sigma_w^4} = \frac{1+\beta^4}{(1+\beta^2)^2}\frac{E(w_t^4)}{\sigma_w^4} + \frac{6\beta^2}{(1+\beta^2)^2} \qquad (III.1)$$

hence

$$\boldsymbol{k} = \frac{E(w_t^4)}{\sigma_w^4} - 3 = \boldsymbol{k_1}\frac{(1+\beta^2)^2}{1+\beta^4} - \frac{6\beta^2}{1+\beta^4} - 3 \ . \qquad (III.2)$$

Now it remains to compute $\boldsymbol{k_1}$. For this, we need the moments up to order four of the process x_t^*. From

$$E\left(x_t^{*2}\right) = \sigma_v^2(1-\rho^2)^{-1} + \sigma_e^2 \qquad E(x_t^* x_{t-1}^*) = \sigma_v^2\rho(1-\rho^2)^{-1}$$

we get

$$E(x_t^* - \rho x_{t-1}^*)^2 = (1+\rho^2)\, E\left(x_t^{*2}\right) - 2\rho\, E(x_t^* x_{t-1}^*) = (1+\rho^2)[\sigma_v^2(1-\rho^2)^{-1} + \sigma_e^2] - 2\sigma_v^2\rho^2(1-\rho^2)^{-1} = \sigma_v^2 + (1+\rho^2)\sigma_e^2 \ . \qquad (III.3)$$

From

$$E\left(x_t^{*4}\right) = 3\sigma_v^4(1-\rho^2)^{-2} + 6\sigma_v^2\sigma_e^2(1-\rho^2)^{-1} + 7\sigma_e^4$$

$$E\left(x_t^{*3}x_{t-1}^*\right) = 3\sigma_v^4\rho^3(1-\rho^2)^{-2} + 3\sigma_v^2(\sigma_v^2+\sigma_e^2)\rho(1-\rho^2)^{-1}$$

$$E\left(x_t^{*2}{x_{t-1}^*}^2\right) = 3\sigma_v^4\rho^2(1-\rho^2)^{-2} + \sigma_v^2(\sigma_v^2+\sigma_e^2)(1-\rho^2)^{-1} + \sigma_v^2\sigma_e^2\rho^2(1-\rho^2)^{-1} + \sigma_v^2\sigma_e^2 + \sigma_e^4$$

$$E\left(x_t^* \quad {x_{t-1}^*}^3\right) = 3\sigma_v^4\rho(1-\rho^2)^{-2} + 3\sigma_v^2\sigma_e^2\rho(1-\rho^2)^{-1}$$

we obtain

$$\begin{aligned} E(x_t^* - \rho x_{t-1}^*)^4 &= (1+\rho^4)E\left(x_t^{*4}\right) - 4\rho E\left(x_t^{*3}x_{t-1}^*\right) + 6\rho^2 E\left(x_t^{*2}{x_{t-1}^*}^2\right) - 4\rho^3 E\left(x_t^* \quad {x_{t-1}^*}^3\right) \\ &= [\sigma_v^2 + (1+\rho^2)\sigma_e^2][3\sigma_v^2 + 3\sigma_e^2(1+\rho^2)] + 4(1+\rho^4)\sigma_e^4 \end{aligned} \qquad (III.4)$$

Substituting (III.3) and (III.4) into (III.1) we get

$$\boldsymbol{k_1} = \frac{4(1+\rho^4)\sigma_e^4}{[\sigma_v^2 + (1+\rho^2)\sigma_e^2]^2} + 3. \qquad (III.5)$$

Substituting (III.5) into (III.2) we have the expression of $\boldsymbol{k}$ as in the statement of Theorem 5.

References

Bartolucci, F. and De Luca, G. 2001. Maximum likelihood estimation of a latent variable time-series model. *Applied Stochastic Models in Business and Industry* 17, 5-17.

Bartolucci, F. and De Luca, G. 2003. Likelihood-based inference for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis* 42 (3), 445-449.

Breidt, F.J. and Carriquiry, A.L. 1996. Improved quasi-maximum likelihood estimation for stochastic volatility models. in: *Modelling and Prediction, Honoring Seymour Geisel* (Zellner, A., and Lee, J.S., eds), Springer-Verlag, Heidelberg-Berlin-New York.

Dunsmuir, W. 1979. A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *The Annals of Statistics* 7 (3), 490-506.

Fridman, M. and Harris, L. 1998. A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal of Business and Economic Statistics* 16, 284-291.

Gallant, A.R. and Tauchen, G. 1996. Which moments to match? *Econometric Theory* 12, 657-681.

Gouriéroux, C., Monfort, A. and Renault, E. 1993. Indirect Inference. *Journal of Applied Econometrics* 8, S85-S118.

Harvey, A.C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press, Cambridge-New York-Sidney.

Harvey, A.C., Ruiz, E. and Shephard, N. 1994. Multivariate stochastic variance models. *Review of Economic Studies* 61, 247-264.

Harvey, A.C. and Shephard, N. 1996. Estimation of an asymmetric model of asset prices. *Journal of Business & Economic Statistics* 14 (4), 429-434.

Jacquier, E., Polson, N.G. and Rossi, P.E. 1994. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics* 12 (4), 69-87.

Monfardini, C. 1998. Estimating Stochastic Volatility Models through Indirect Inference. *Econometric Journal* 1, 113-128.

Ruiz, E. 1994. Quasi-maximum likelihood estimation of Stochastic Volatility Models. *Journal of Econometric* 63, 289-306.

Ruiz, E. 2013. Personal communication.

Taylor, S.J. 1994. Modelling Stochastic Volatility. *Mathematical Finance* 4, 183-204.