



**HAL**  
open science

# PARAMETRIC INFERENCE OF HIDDEN DISCRETE-TIME DIFFUSION PROCESSES BY DECONVOLUTION

Salima El Kolei, Florian Pelgrin

► **To cite this version:**

Salima El Kolei, Florian Pelgrin. PARAMETRIC INFERENCE OF HIDDEN DISCRETE-TIME DIFFUSION PROCESSES BY DECONVOLUTION. 2015. hal-01248486v1

**HAL Id: hal-01248486**

**<https://hal.science/hal-01248486v1>**

Preprint submitted on 26 Dec 2015 (v1), last revised 18 Dec 2016 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PARAMETRIC INFERENCE OF HIDDEN DISCRETE-TIME DIFFUSION PROCESSES BY DECONVOLUTION

SALIMA EL KOLEI AND FLORIAN PELGRIN

ABSTRACT. We study a new parametric approach for hidden discrete-time diffusion models. This method is based on contrast minimization and deconvolution and leads to estimate a large class of stochastic models with nonlinear drift and nonlinear diffusion. It can be applied, for example, for ecological and financial state space models.

After proving consistency and asymptotic normality of the estimation, leading to asymptotic confidence intervals, we provide a thorough numerical study, which compares many classical methods used in practice (Non Linear Least Square estimator, Monte Carlo Expectation Maximization Likelihood estimator and Bayesian estimators) to estimate stochastic volatility model. We prove that our estimator clearly outperforms the Maximum Likelihood Estimator in term of computing time, but also most of the other methods. We also show that this contrast method is the most stable and also does not need any tuning parameter.

Keywords: Deconvolution, Hidden Markov Model, Discrete Stochastic Volatility Models.

## 1. INTRODUCTION

A large class of models encountered in finance fields can be written as:

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = b_{\theta_0}(X_i) + \sigma_{\theta_0}(X_i)\eta_{i+1}, \end{cases} \quad (1)$$

---

*Date:* December 26, 2015.

CREST, Ecole Nationale de la Statistique et de l'Analyse de l'Information (Ensaï), Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. Author email: salima.el.kolei@ensai.fr.

Ecole des Hautes Etudes Commerciales, 24, avenue Gustave Delory CS 50411, 59057 Roubaix Cedex 1-France. Author email: florian.pelgrin@edhec.edu .

where one observes  $Y_1, \dots, Y_n$ , and where the random variables  $\varepsilon_i$ ,  $\eta_i$  and  $X_i$  are unobserved. Notably  $(X_i)_{i \geq 0}$  is a strictly stationary, ergodic process that depends on two measurable functions  $b_{\theta_0}$  and  $\sigma_{\theta_0}$  and its stationary density is  $f_{\theta_0}$ . These two regression functions are known up to a finite dimensional parameter,  $\theta_0$ , and the dependence with respect to  $\theta_0$  is not required to be the same in  $b_{\theta_0}$  and  $\sigma_{\theta_0}$ . Finally, the innovations  $(\eta_i)_{i \geq 0}$  and the errors  $(\varepsilon_i)_{i \geq 0}$  are independent and identically distributed (i.i.d) random variables, the distribution of the innovations being known. This class of parametric models includes, among others, the autoregressive model with measurement errors, the autoregressive stochastic volatility model ([Taylor, 2005]), the discrete time versions of well-known diffusion processes in finance ([Heston, 1993], [Cox et al., 1985]) and some families of stochastic processes: Vasick, CIR, CIR modified and hyperbolic processes (see [Genon-Catalot et al., 1999]).

In this paper, we propose a new parametric estimation method for the two functions  $b_{\theta_0}$  and  $\sigma_{\theta_0}$  driving the dynamics of the hidden variables  $(X_i)_{i \geq 0}$ . The principle of the estimation method goes as follows. Taking that the stationary density,  $f_{\theta_0}$ , is known up to the finite dimensional parameter  $\theta_0$ , our M-estimator consists in optimizing a contrast function that exploits a Fourier deconvolution strategy in a parametric framework. In so doing, we exploit a "Nadaraya-Watson approach" in the sense that we estimate  $b_{\theta_0}$  (respectively,  $b_{\theta_0}^2 + \sigma_{\theta_0}^2$ ) as ratio of an estimator of  $l_{\theta_0} = b_{\theta_0} f_{\theta_0}$  (respectively,  $l_{\theta_0} = (b_{\theta_0}^2 + \sigma_{\theta_0}^2) f_{\theta_0}$ ) and an estimator of  $f_{\theta_0}$ . Notably we provide an analytical expression of the contrast function for a well-known example and characterize their main properties. Moreover we show that this deconvolution-based estimator is consistent and asymptotically normally distributed for  $\alpha$ -mixing processes which leads to obtain confidence intervals in practice for many processes. Finally, our Monte Carlo simulations show that our approach gives good results and is fast computing. All results are illustrated on the famous Heston model which is very used in practice for options prices and our approach is compared with many others methods used in the literature to estimate this model (Non Linear Least Square, Monte Carlo Expectation Maximisation, Sequential Monte Carlo).

We are far from the first ones to take interest in the estimation of stochastic volatility models (see [Genon-Catalot et al., 1999]) but in this paper we propose a general new approach to estimate a large

class of stochastic processes.

Three papers are closely related to our work. On the one hand, Comte et al. (see [Comte et al., 2010]) propose a non-parametric estimation procedure in the case of discrete time stochastic models.<sup>1</sup> Their approach rests on three steps. The first stage leads to define an estimator of  $l_{\theta_0}$  through a penalized contrast function by using an orthogonal projection of  $l_{\theta_0}$  on a (finite) basis of the space of square integrable function (having Fourier transform). The second stage amounts of estimating  $f_{\theta_0}$  in an adaptive way (see [Comte et al., 2006]). By a ratio strategy ("Nadaraya-Watson approach"), estimators of  $b$  and  $\sigma^2$  are then deduced in the third stage. This approach is the cornerstone of the estimator developed in this paper in the sense that it is based on the Fourier deconvolution and the same (non-penalized) contrast function. At the same time, we show that their methodology can be extended in a parametric framework and we go further by obtaining confidence intervals. On the other hand, in [Dedecker et al., 2014] Dedecker et al. propose a new parametric estimation procedure based on a modified least squares criterion. Their assumptions on the process  $X_i$  are less restrictive than those proposed by F. Comte and M. Taupin in [Comte and Taupin, 2001] and they provide consistent estimation of the parameter  $\theta_0$  of the regression function  $b$  with a parametric rate of convergence in a very general framework. Their general estimator is based on the introduction of a kernel deconvolution density and depends on the choice of a weight function.

Finally, in [El Kolei, 2013], the author proposes a new parametric estimation for a class of state space models. This approach is not based on a weighted least squared estimation so that the choice of the weight function is not encountered in her paper. Moreover, it allows to estimate both the parameters of the drift  $b$  and diffusion  $\sigma^2$  functions only in the case where  $\sigma^2$  is a constant function of the hidden variable  $x$ . Our approach is a generalisation of this procedure, that is it permits also to estimate models with nonlinear drift function  $b$  and nonlinear diffusion function  $\sigma^2$ . Although, we focus our approach on the estimation of stochastic volatility models, this method can be applied for many applications.

The paper is organized as follows. Section 2 provides examples of hidden stochastic volatility models

---

<sup>1</sup>See also Comte et Taupin in [Comte and Taupin, 2007].

that can be casted in (1). Section 3 presents the notations and the model assumptions. Section 4 defines the deconvolution-based M-estimator and states all of the theoretical properties. Some Monte Carlo simulations are discussed in Section 5 and some concluding remarks are provided in the last section. The proofs are gathered in Appendix.

## 2. EXAMPLES

In this section, we provide many examples of stochastic processes on which our approach can be easily applied.

- Cox-Ingersoll-Ross process:

$$dX(t) = (\beta_1 - \beta_2 X(t))dt + \beta_3 \sqrt{X(t)}dW(t) \quad (2)$$

The stationary distribution is a Gamma law with shape parameter  $2\beta_1/\beta_3^2$  and scale parameter  $\beta_3^2/2\beta_2$ .

- The modified CIR

$$dX(t) = -\beta_1 X(t)dt + \beta_2 \sqrt{1 + X^2(t)}dW(t)$$

where  $\beta_1 + \beta_2^2 > 0$ . The modified CIR process has a stationary distribution that is proportional to  $1/(1 + x^2)^{1+\beta_1/\beta_2}$ . Thus,

$$X(t) \sim \mathcal{T}(n_v)/\sqrt{(n_v)}$$

where  $\mathcal{T}$  is the Student distribution and the number of degrees of freedom  $n_v$  is given by  $n_v = 1 + \beta_1/\beta_2$ .

- The hyperbolic processes (sometimes used to model log-returns of assets prices in stock markets)

$$dX(t) = \frac{\sigma^2}{2} \left[ \beta - \gamma \frac{X(t)}{\sqrt{\delta^2 + (X(t) - \mu)^2}} \right] dt + \sigma dW(t)$$

Its invariant density is defined by:

$$\pi(x) = \frac{\sqrt{\gamma^2 - \beta^2}}{2\gamma\delta K_1(\cdot)} \exp\{-\gamma\sqrt{\delta^2 + (x - \mu)^2} + \beta(x - \mu)\}$$

where  $K_1(\cdot) = K_1(\delta\sqrt{\gamma^2 - \beta^2})$  is the modified Bessel function of second kind of index one. It is worth noting that  $\delta > 0, 0 \leq |\beta|$ , and  $\gamma > 0$ . As a special case, one gets the hyperbolic process

$$dX(t) = \theta \frac{X(t)}{\sqrt{1 + X^2(t)}} dt + dW(t)$$

- Linear feedback models

- Type 1

TABLE 1. Expression

Drift: $b(x) = r(\theta - x), r > 0$
Diffusion: $\sigma^2(x) = \epsilon$
SDE: $dX(t) = r(\theta - X(t))dt + \sqrt{\epsilon}dW(t)$
Stationary density: $\frac{1}{2\pi\delta} \exp\{-\frac{(x-\theta)^2}{2\delta}\}$
$\delta = \epsilon/r$

- Type 2

TABLE 2. Expression

Drift: $b(x) = r(\theta - x), r > 0$
Diffusion: $\sigma^2(x) = \epsilon x$
SDE: $dX(t) = r(\theta - X(t))dt + \sqrt{\epsilon X(t)}dW(t)$
Stationary density: $(\frac{x}{\delta})^{-1+\frac{\theta}{\delta}} \frac{\exp\{-\frac{x}{\delta}\}}{\Gamma(\theta/\delta)}$
$\delta = \epsilon/r$ and $\Gamma$ is the Gamma function.

- Type 3

TABLE 3. Expression

Drift: $b(x) = r(\theta - x), r > 0$
Diffusion: $\sigma^2(x) = \epsilon x(1 - x)$
SDE: $dX(t) = r(\theta - X(t))dt + \sqrt{\epsilon X(t)(1 - X(t))}dW(t)$
Stationary density: $\frac{\Gamma(1/\delta)}{\Gamma(\theta/\delta)\Gamma(1-\theta/\delta)}x^{-1+\frac{\theta}{\delta}}(1-x)^{-1+\frac{1-\theta}{\delta}}$
$\delta = \epsilon/r$ and $\Gamma$ is the Gamma function.

**Remark 1.** *The state equation in the model (1) is viewed as an approximation of a stochastic differential equation (SDE). One key issue is that the stationary distribution of  $X(t)$  defined by the SDE*

$$dX(t) = a(X(t), \beta_1)dt + b(X(t), \beta_2)dW(t)$$

*can be known but this no longer true when considering the discrete time version of the SDE. Nevertheless, it is not the matter of this paper and can be studied in a next paper. For this paper, we consider exact discrete time nonlinear state space models in which we know exactly the stationary distribution. For example, for Ornstein-Uhlenbeck process or CIR process (2) we know the stationary distribution of the discrete time version of the SDE (see [Cox et al., 1985]).*

### 3. GENERAL SETTING AND ASSUMPTIONS

In this section, we introduce some preliminary main notations and provide the assumptions of model (1).

**3.1. Notations.** Subsequently, we denote by  $u^*$  the Fourier transform of the function  $u$ :

$$u^*(t) = \int e^{itx}u(x)dx,$$

and by  $\|u\|_2$ ,  $\|u\|_\infty$ ,  $\langle u, v \rangle$ , and  $u * v$  the quantities:

$$\begin{aligned} \|u\|_2 &= \left( \int |u(x)|^2 dx \right)^{1/2} \\ \|u\|_\infty &= \sup_{x \in \mathbb{R}} |u(x)| \\ \langle u, v \rangle &= \int u(x) \bar{v}(x) dx \quad \text{with} \quad v\bar{v} = |v|^2 \\ u * v &= \int u(t) \bar{v}(x-t) dt. \end{aligned}$$

Moreover, for any integrable and square-integrable functions  $u$ ,  $u_1$ , and  $u_2$ :

$$\begin{aligned} (u^*)^*(x) &= 2\pi u(-x) \\ \langle u_1, u_2 \rangle &= \frac{1}{2\pi} \langle u_1^*, u_2^* \rangle. \end{aligned}$$

Finally,  $\|A\|$  denotes the Euclidean norm of a matrix  $A$ ,  $\mathbf{Y}_i = (Y_i, Y_{i+1})$  and  $\mathbf{y}_i = (y_i, y_{i+1})$ ,  $\mathbf{P}_n$  (respectively,  $\mathbf{P}$ ) the empirical (respectively, theoretical) expectation, that is, for any stochastic variable:  $\mathbf{P}_n(X) = \frac{1}{n} \sum_{i=1}^n X_i$  (respectively,  $\mathbf{P}(X) = \mathbb{E}[X]$ ). Regarding the partial derivatives, for any function  $h_\theta$ ,  $\nabla_\theta h_\theta$  is the vector of the partial derivatives of  $h_\theta$  with respect to (w.r.t)  $\theta$  and  $\nabla_\theta^2 h_\theta$  is the Hessian matrix of  $h_\theta$  w.r.t  $\theta$ .

**3.2. Assumptions.** We consider the hidden discrete-time diffusion model (1). The assumptions are the following.

**A0**  $\theta_0$  belongs to the interior  $\Theta_0$  of a compact set  $\Theta$ ,  $\theta_0 \in \Theta \subset \mathbb{R}^p$ .

**A1** The errors  $(\varepsilon_i)_{i \geq 0}$  are independent and identically distributed centered random variables with finite variance,  $\mathbb{E}[\varepsilon_1^2] = s_\varepsilon^2$ . The density of  $\varepsilon_1$ ,  $f_\varepsilon$ , belongs to  $\mathbb{L}_2(\mathbb{R})$ , and for all  $x \in \mathbb{R}$ ,  $f_\varepsilon^*(x) \neq 0$ .

**A2** The innovations  $(\eta_i)_{i \geq 0}$  are independent and identically distributed centered random variables with unit variance  $\mathbb{E}[\eta_1^2] = 1$  and  $\mathbb{E}[\eta_1^3] = 0$ .

**A3** The  $X_i$ 's are strictly stationary and ergodic with invariant density  $f_{\theta_0}$ .

**A4** The sequences  $(X_i)_{i \geq 0}$  and  $(\varepsilon_i)_{i \geq 0}$  are independent. The sequence  $(\varepsilon_i)_{i \geq 0}$  and  $(\eta_i)_{i \geq 0}$  are independent.



**A5** On  $\Theta_0$ , the functions  $\theta \mapsto b_\theta$  and  $\sigma \mapsto \sigma_\theta$  admit continuous derivatives with respect to  $\theta$  up to order 2.

**A6** The function to estimate  $l_\theta := (b_\theta^2 + \sigma_\theta^2) f_\theta$  belongs to  $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ , is twice continuously differentiable w.r.t  $\theta \in \Theta$  for any  $x$  and measurable w.r.t  $x$  for all  $\theta$  in  $\Theta$ . Each element of  $\nabla_\theta l_\theta$  and  $\nabla_\theta^2 l_\theta$  belongs to  $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ .

The compactness assumption A0 might be relaxed by assuming that  $\theta_0$  is an element of the interior of a convex parameter space  $\Theta \in \mathbb{R}^p$ . In this case, the statistical properties of the M-estimator can be proved in the light of convex optimization arguments. Assumptions **A1-A3** are quite standard when considering estimation in the convolution model. On the other hand, Assumption **A3** implies that if  $(X_i)_{i \geq 0}$  is an ergodic process then  $(Y_i)_{i \geq 0}$  is stationary and ergodic since it is the sum of an ergodic process and an i.i.d. noise process ([Dedecker et al., 2007]). Consequently  $\mathbf{Y}_i = (Y_i, Y_{i+1})$  inherits the ergodicity property. According to Assumption **A4** the unknown density  $g_{\theta_0}$  of the  $Y_i$ 's is defined to be  $f_{\theta_0} * f_\varepsilon$ . It turns out that  $g_{\theta_0}^* = f_{\theta_0}^* f_\varepsilon^*$  and thus  $f_{\theta_0}^* = g_{\theta_0}^* / f_\varepsilon^*$ . Assumption **A5** ensures some smoothness for the functions to estimate. Note that Assumptions **A2** and **A3** imply that the sequence  $(X_i)_{i \geq 0}$  is stationary and ergodic with invariant density  $f_{\theta_0}$ . Assumption **A6** is also quite usual in the literature and serves for the construction and for asymptotic properties of our estimator.

#### 4. PARAMETRIC DECONVOLUTION ESTIMATOR

In this section, we first define the contrast function. Then we state the definition of the parametric deconvolution estimator. Finally the consistency and asymptotic normality of this estimator are established.

**4.1. The contrast function.** Using Comte et al. (see [Comte et al., 2010]), the starting point of our estimation procedure is to consider an estimator of the following contrast function.

**Definition 1** (Contrast function). *Suppose that Assumptions **A1**, **A2**, **A4**, **A5** and **A6** hold true, the quantity  $\varphi(Y_2)u_{l_\theta}^*(Y_1)$  has a finite expectation for some measurable function  $\varphi$ , and  $u_{l_\theta}$  is square integrable (**A7**). Then the contrast function is defined by:*

$$\mathbb{E} [m_\theta(\mathbf{Y}_1)] := \|l_\theta\|^2 - 2\mathbb{E} [\varphi(Y_2)u_{l_\theta}^*(Y_1)],$$

where  $u_{l_\theta}(x) = \frac{1}{2\pi} \frac{v^*(-x)}{f_\varepsilon^*(x)}$ . The function  $\varphi$  depends on the form of the diffusion function  $\sigma_{\theta_0}$  i.e.

- (1)  $\varphi : x \in \mathbb{R} \mapsto x$ , if  $\sigma_{\theta_0}$  is a constant function of the hidden variable
- (2)  $\varphi : x \in \mathbb{R} \mapsto x^2 - s_\varepsilon^2$ , if  $\sigma_{\theta_0}$  is not a constant function of the hidden variable

and it satisfies

$$\mathbb{E} [\varphi(Y_2) u_{l_\theta}^*(Y_1)] = \langle l_\theta, l_{\theta_0} \rangle.$$

Several points are worth commenting. First, the empirical contrast is given by:

$$\mathbf{P}_n m_\theta = \frac{1}{n} \sum_{i=1}^n m_\theta(\mathbf{y}_i), \quad (3)$$

where:

$$m_\theta(\mathbf{y}_i) : (\theta, \mathbf{y}_i) \in (\Theta \times \mathbb{R}^2) \mapsto m_\theta(\mathbf{y}_i) = \|l_\theta\|^2 - 2\varphi(y_{i+1}) u_{l_\theta}^*(y_i).$$

As explained below, using the Ergodicity Theorem, it can be shown that  $\mathbf{P}_n m_\theta$  converges in probability towards  $\mathbf{P} m_\theta = \mathbb{E} [m_\theta(\mathbf{Y}_1)]$  as  $n$  goes to infinity. Second, taking the various applications, the choice of the function  $\varphi$  can be made explicit. Indeed, in the case of a constant diffusion function  $\sigma_\theta$  with respect to  $x$  the function  $\varphi$  is given by the identity function and  $l_\theta$  is given by:

$$l_\theta(x) = b_\theta(x) f_\theta(x)$$

This typical case is studied in [El Kolei, 2013].

On the other hand, when  $\sigma_\theta$  is a nonlinear diffusion function with respect to  $x$  as in model (2), we define the function  $\varphi$  by  $\varphi(x) = x^2 - s_\varepsilon^2$  and  $l_\theta(x) = (b_\theta^2(x) + \sigma_\theta^2(x)) f_\theta(x)$ .

**Remark 2.** As said in the introduction, the case where the diffusion function  $\sigma_{\theta_0}$  is a constant function of the hidden variable has already been studied in [El Kolei, 2013]. Therefore, from now on, we focus on the case 2 in Definition (1) and we refer to the aforementioned paper for the case 1 in Definition (1).

It is straightforward to show that in this case:

$$\mathbb{E} [\varphi(Y_2)u_{l_\theta}^*(Y_1)] = \langle l_\theta, l_{\theta_0} \rangle$$

and

$$\mathbb{E} [m_\theta(\mathbf{Y}_1)] = \|l_\theta\|^2 - 2 \langle l_\theta, l_{\theta_0} \rangle = \|l_\theta - l_{\theta_0}\|^2 - \|l_{\theta_0}\|^2. \quad (4)$$

Third, the information criterion (4) is minimum when  $\theta = \theta_0$ . This requires the following identification assumption:

**A8** The information criterion  $\mathbb{E} [m_\theta(\mathbf{Y}_1)]$  has a unique minimum at  $\theta = \theta_0$ .

In this respect, the associated minimum-contrast estimator  $\hat{\theta}_n$  is defined as follows.

**Definition 2.** *The minimum-contrast estimator  $\hat{\theta}_n$  solves:*

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \mathbf{P}_n m_\theta. \quad (5)$$

**Remark 3.** *In this paper we consider the situation in which the observation noise variance is known. This assumption, which is often not satisfied in practice, is necessary for the identifiability of the model and so is a standard assumption for state-space models given in (1).*

*There is some restrictions on the distribution of the innovations in the Nadaraya-Watson approach. It is known that the rate of convergence for estimating the function  $l_\theta$  is related to the rate of decreasing of  $f_\varepsilon^*$ . In particular, the smoother  $f_\varepsilon$ , the slower the rate of convergence for estimating is (The Gaussian, log-chi squared or Cauchy distributions are super-smooth. The Laplace distribution is ordinary smooth). This rate of convergence can be improved by assuming some additional regularity conditions on  $l_\theta$ . There is a good discussion about this subject in [Comte et al., 2010] and [Comte et al., 2006].*

**The procedure:** Let us explain the choice of the contrast function and how the strategy of deconvolution works. The convergence of  $\mathbf{P}_n m_\theta$  to  $\mathbf{P} m_\theta = \mathbb{E} [m_\theta(\mathbf{Y}_1)]$  as  $n$  tends to the infinity is provided by the Ergodicity Theorem. Moreover, the limit  $\mathbb{E} [m_\theta(\mathbf{Y}_1)]$  of the contrast function can be

explicitly computed. Using (1) and Assumptions **A1-A3**, we obtain:

$$\begin{aligned}
\mathbb{E} [(Y_2^2 - s_\varepsilon^2) u_{l_\theta}^*(Y_1)] &= \mathbb{E} [(X_2^2 + 2X_2\varepsilon_2 + \varepsilon_2^2 - s_\varepsilon^2) u_{l_\theta}^*(Y_1)] \\
&= \mathbb{E} [X_2^2 u_{l_\theta}^*(Y_1)] \text{ by assumption } \mathbf{A1} \\
&= \mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)\eta_2^2 + 2b_{\theta_0}(X_1)\sigma_{\theta_0}(X_1)\eta_2) u_{l_\theta}^*(Y_1)] \\
&= \mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) u_{l_\theta}^*(Y_1)] \text{ by assumption } \mathbf{A2},
\end{aligned}$$

Using Fubini's Theorem and (1), it follows that:

$$\begin{aligned}
\mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) u_{l_\theta}^*(Y_1)] &= \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \int e^{iY_1 y} u_{l_\theta}(z) dz \right] \\
&= \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \int \frac{1}{2\pi} \frac{1}{f_\varepsilon^*(z)} e^{iY_1 z} (l_\theta(-z))^* dy \right] \\
&= \frac{1}{2\pi} \int \mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) e^{i(X_1 + \varepsilon_1)z}] \frac{1}{f_\varepsilon^*(z)} (l_\theta(-z))^* dz \\
&= \frac{1}{2\pi} \int \frac{\mathbb{E} [e^{i\varepsilon_1 z}]}{f_\varepsilon^*(z)} \mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) e^{iX_1 z}] (l_\theta(-z))^* dy \\
&= \frac{1}{2\pi} \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \int e^{iX_1 z} (l_\theta(-z))^* dz \right] \\
&= \frac{1}{2\pi} \mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) ((l_\theta(-X_1))^*)^*] \\
&= \mathbb{E} [(b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) l_\theta(X_1)]. \\
&= \int (b_{\theta_0}^2(x) + \sigma_{\theta_0}^2(x)) f_{\theta_0}(x) (b_{\theta_0}^2(x) + \sigma_{\theta_0}^2(x)) f_{\theta_0}(x) dx \\
&= \langle l_\theta, l_{\theta_0} \rangle.
\end{aligned}$$

Then,

$$\mathbb{E} [m_\theta(\mathbf{Y}_1)] = \|l_\theta\|^2 - 2 \langle l_\theta, l_{\theta_0} \rangle = \|l_\theta - l_{\theta_0}\|^2 - \|l_{\theta_0}\|^2.$$

Using Definition 2, we are now in position to explicit the contrast function and the minimization problem for the examples in Section 2.<sup>2</sup>

---

<sup>2</sup>All the proofs are derived in Appendix 7 and 7.4.

The existence of our estimator follows from regularity properties of the function  $l_\theta$  and compactness argument of the parameter space, it is explained in Appendix 7 Section 7.1.

**4.2. Asymptotic properties.** In this section we show that our estimator is weakly consistent and asymptotically normally distributed for mixing processes. In this respect, we further assume that:

$$\mathbf{A9} \text{ (Local dominance): } \mathbb{E} \left[ \sup_{\theta \in \Theta} \left| \varphi(Y_{i+1}) u_{l_\theta}^*(Y_i) \right| \right] < \infty.$$

**A10** The application  $\theta \mapsto \mathbf{P}m_\theta$  admits a unique minimum and its Hessian matrix, denoted by  $V_\theta$ , is non-singular in  $\theta_0$ .

For the CLT, we need to add two assumptions:

$$\mathbf{A11} \text{ (Moment condition): For some } \delta > 0, \mathbb{E} \left[ \left| \varphi(Y_{i+1}) u_{\nabla_\theta^2 l_\theta}^*(Y_i) \right|^{2+\delta} \right] < \infty.$$

**A12** (Hessian Local dominance): For some neighbourhood  $\mathcal{U}$  of  $\theta_0$ :

$$\mathbb{E} \left[ \sup_{\theta \in \mathcal{U}} \left\| \varphi(Y_{i+1}) u_{\nabla_\theta^2 l_\theta}^*(Y_i) \right\| \right] < \infty$$

**4.2.1. Consistency.** The first result regards the (weak) consistency of our estimator.

**Theorem 1.** *Consider the model (1) under the assumptions **A0-A7** and suppose that the conditions **A8-A10** hold true. Then  $\hat{\theta}_n$  defined by (5) is weakly consistent:*

$$\hat{\theta}_n \longrightarrow \theta_0 \quad \text{as } n \rightarrow \infty \text{ in } \mathbb{P}_{\theta_0} - \text{probability. } \square$$

Proof: See Appendix 7 Section 7.2.

The main idea for proving the consistency of a M-estimator comes from the following observation: if  $\mathbf{P}_n m_\theta$  converges to  $\mathbf{P} m_\theta$  in probability, and if the true parameter solves the limit minimization problem, then, the limit of the argminimum  $\widehat{\theta}_n$  is  $\theta_0$ . By using an argument of uniform convergence in probability and by compactness of the parameter space, we show that the argminimum of the limit is the limit of the argminimum. A standard method to prove the uniform convergence is to use *the Uniform Law of Large Numbers* (see Lemma 1 in Appendix 7). Combining these arguments with the dominance argument **(A9)** give the consistency of our estimator, and then, the Theorem 1.

4.2.2. *Asymptotic normality.* The second result states our estimator is  $\sqrt{n}$ -consistent and asymptotically normally distributed. Besides, regarding the variance-covariance matrix, Corollary 1 provides the different terms of the variance-covariance matrix for stochastic processes with nonlinear diffusion.

For the CLT, we need to recall some mixing properties (we refer the reader to [Dedecker et al., 2007] for a complete revue of mixing processes). Let  $K_\theta(x, dy)$  be a Markov transition kernel on a general space  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$  and let  $K_\theta^n(x, dy)$  denotes the  $n$  step Markov transition corresponding to  $K_\theta$ . Then, for  $k \in \mathbb{N}, x \in \mathcal{X}$  and a measurable set  $\mathcal{A}$ :

$$K_\theta^n(x, dy) = \mathbb{P}_\theta(X_{n+k} \in \mathcal{A} | X_k = x)$$

Let  $M(x)$  be a nonnegative function and  $\gamma(n)$  be a nonnegative decreasing function on  $\mathbb{Z}_+$  such that:

$$(C) \quad \|K_\theta^n(x, \cdot) - f_\theta(\cdot)\|_{VT} \leq M(x)\gamma(n)$$

where  $\|\cdot\|_{VT}$  denotes the total variation norm.

**Definition 3** ( $\alpha$ -mixing (strongly mixing process)). *Let  $Y := \{Y_n\}$  denotes a general sequence of random variables on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and let  $\mathcal{F}_k^m = \sigma(Y_k, \dots, Y_m)$ . The sequence  $Y$  is said to be  $\alpha$ -mixing if  $\alpha(n) \rightarrow 0$  as  $n \rightarrow \infty$  where*

$$\alpha(n) := \sup_{k \geq 1} \sup_{\mathcal{A} \in \mathcal{F}_1^k, \mathcal{B} \in \mathcal{F}_{k+n}^\infty} |\mathbb{P}_\theta(\mathcal{A} \cup \mathcal{B}) - \mathbb{P}_\theta(\mathcal{A})\mathbb{P}_\theta(\mathcal{B})|$$

where  $\mathcal{A}$  and  $\mathcal{B}$  are two measurable sets.

**Definition 4** ( $\rho$ -mixing (asymptotically uncorrelated)). *The sequence  $Y$  is said to be  $\rho$ -mixing if  $\rho(n) \rightarrow 0$  as  $n \rightarrow \infty$  where*

$$\rho(n) := \sup \left\{ \text{corr}(U, V), U \in \mathbb{L}_2(\mathcal{F}_1^k), V \in \mathbb{L}_2(\mathcal{F}_{k+n}^\infty), k \geq 1 \right\}.$$

**Remark 4.**  *$X$  is geometrically ergodic if (C) holds with  $\gamma(n) = t^n$  for some  $t < 1$ .  $X$  is uniform ergodic if (C) holds with  $M$  bounded and  $\gamma(n) = t^n$  for some  $t < 1$ .  $X$  is polynomial ergodic of order  $m$  where  $m \geq 0$  if (C) holds with  $\gamma(n) = n^{-m}$ . The strong Markov property implies that  $\rho$ -mixing sequences are also  $\alpha$ -mixing and, in fact,  $4\alpha(n) \leq \rho(n)$ .*

The following theorem states our estimator is asymptotically normally distributed. In this respect, we further assume that:

**A14** The stochastic process  $X_i$  is  $\alpha$ -mixing.

**Theorem 2.** *Consider the model (1) under the assumptions **A0-A8**, and suppose that the conditions **A9-A14** hold true.*

*If (C) holds such that  $\mathbb{E}[M(X_1)] < \infty$  and  $\gamma(n)$  satisfies  $\sum_n \gamma(n)^{\frac{\delta}{2+\delta}} < \infty$  where  $\delta$  is given in assumption **A12**, then  $\hat{\theta}_n$  defined by (5) is a  $\sqrt{n}$ -consistent estimator of  $\theta_0$  which satisfies:*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta_0))$$

where the variance-covariance matrix is given by:

$$\Sigma(\theta_0) = V_{\theta_0}^{-1} \Omega(\theta_0) V_{\theta_0}^{-1'}$$

*Proof.* See Appendix 7 Section 7.3. □

The following corollary gives an expression of the matrix  $\Omega(\theta_0)$  and  $V_{\theta_0}$  of Theorem 2 for the practical implementation:

**Corollary 1.** *Under our assumptions, the matrix  $\Omega(\theta_0)$  is given by:*

$$\Omega(\theta_0) = \Omega_0(\theta_0) + 2 \sum_{j=2}^{+\infty} \Omega_{j-1}(\theta_0),$$

where:

$$\begin{aligned} \Omega(\theta_0) &= \text{Var}(\nabla_{\theta} m_{\theta}(\mathbf{Y}_1)) + 2 \sum_{j=2}^{+\infty} \text{Cov}(\nabla_{\theta} m_{\theta}(\mathbf{Y}_1), \nabla_{\theta} m_{\theta}(\mathbf{Y}_j)), \\ &= \Omega_1(\theta_0) + 2 \sum_{j=2}^{+\infty} \Omega_j(\theta_0) \end{aligned}$$

where

$$\begin{aligned} \Omega_1(\theta_0) &= 4 \left\{ \mathbb{E} \left[ \left( \varphi(Y_2) u_{\nabla_{\theta} l_{\theta_0}}^*(Y_1) \right) \left( \varphi(Y_2) u_{\nabla_{\theta} l_{\theta_0}}^*(Y_1) \right)' \right] \right. \\ &\quad \left. - \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \nabla_{\theta} l_{\theta_0}(X_1) \right] \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \nabla_{\theta} l_{\theta_0}(X_1) \right]' \right\} \\ \Omega_j(\theta_0) &= 4 \left\{ \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \nabla_{\theta} l_{\theta_0}(X_1) \left( (b_{\theta_0}^2(X_j) + \sigma_{\theta_0}^2(X_j)) \nabla_{\theta} l_{\theta_0}(X_j) \right)' \right] \right. \\ &\quad \left. - \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \nabla_{\theta} l_{\theta_0}(X_1) \right] \mathbb{E} \left[ (b_{\theta_0}^2(X_1) + \sigma_{\theta_0}^2(X_1)) \nabla_{\theta} l_{\theta_0}(X_1) \right]' \right\}. \end{aligned}$$

Furthermore, the Hessian matrix  $V_{\theta_0}$  is given by:

$$\left( [V_{\theta_0}]_{j,k} \right)_{1 \leq j,k \leq r} = 2 \left( \left\langle \frac{\partial l_{\theta}}{\partial \theta_k}, \frac{\partial l_{\theta}}{\partial \theta_j} \right\rangle \right)_{j,k} \text{ at point } \theta = \theta_0.$$

Proof: See Appendix 7 Section 7.4.

The asymptotic normality follows essentially from Central Limit Theorem for mixing processes (see [Jones, 2004]). Thanks to the consistency, the proof is based on a moment condition of the Jacobian vector of the function  $m_{\theta}(\mathbf{y})$  and on a local dominance condition of its Hessian matrix. To refer to



likelihood results, one can see these assumptions as a moment condition of the score function and a local dominance condition of the Hessian (see [Van der Vaart, 1998]).

## 5. APPLICATIONS

**5.1. Contrast estimator for the CIR process.** In this part, we will consider the discrete time version of the CIR proces (2):

$$X_{i+1} = X_i + \kappa(\mu - X_i)\Delta + \sigma\sqrt{\Delta X_i}\eta_{i+1} \quad (6)$$

with  $\Delta$  the sampling interval.

The unobserved variance process  $X_i$  is driven by a mean reverting stochastic process which was introduced in [Cox et al., 1985] to model the short term interest rates. The parameter  $\kappa$  is the positive mean reverting parameter,  $\mu$  is the positive long run parameter and  $\sigma$  the positive volatility of the variance process.

We assume that the variance process  $X_i$  is greater than zero. To ensure that this is satisfied, we make the following assumption:

$$\mathbf{F} \quad a = \frac{2\kappa\theta}{\sigma^2} \geq 1 \text{ and } c = \frac{2\kappa}{\sigma^2} > 0,$$

this condition is known as the Feller condition (see [Cox et al., 1985]) and implies that the variance process  $X_i$  is ergodic and  $\rho$ -mixing (see Definition (4)). Furthermore, the stationary distribution  $f_\theta$  writes:

$$f_\theta(x) = \frac{c^a}{\Gamma(a)} x^{a-1} e^{-cx} \quad \forall x > 0.$$

This is the gamma distribution  $\Gamma(a, c)$  (see [Genon-Catalot et al., 1999]).

Then, the discrete time Heston model is given by the following nonlinear state space model with additive noises:

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = X_i + \kappa(\mu - X_i)\Delta + \sigma\sqrt{\Delta X_i}\eta_{i+1} \end{cases} \quad (7)$$

with  $\varepsilon_i$  follows a log chi-squared distribution and  $\eta_i$  a gaussian distribution.

On the other hand, the functions  $b$ .,  $\sigma$ . and  $l$ . are given by:

$$b_\theta(x) = (1 - \kappa)x + \kappa\theta, \quad \sigma_\theta(x) = \sigma\sqrt{x} \text{ and } l_\theta = (b_\theta^2(x) + \sigma_\theta^2(x)) \Gamma(a, c).$$

where  $\theta = (\kappa, \mu, \sigma)$ .

Using the Fourier transform of the Gamma and the log chi-squared density, we have

$$f_\theta^*(x) = \left(1 - \frac{ix}{c}\right)^{-a} \text{ and } f_\varepsilon^*(x) = \frac{1}{\sqrt{\pi}} 2^{ix} \Gamma\left(\frac{1}{2} + ix\right) \exp(-iCx)$$

with  $C$  the expectation of the logarithm of a chi-squared random variable, *i.e.*  $C = -1.27$  (see [Abramowitz and Stegun, 1992] and Appendix 7.4 for the expression of the Fourier transform).

As well as the one of  $l_\theta$  is given by

$$l_\theta^*(x) = -\alpha_1 \left[ \frac{-a}{c^2} (a+1) \left(1 - \frac{ix}{c}\right)^{-a-2} \right] + i\alpha_2 \frac{a}{c} \left(1 - \frac{ix}{c}\right)^{-a-1} + \alpha_3 \left(1 - \frac{ix}{c}\right)^{-a}.$$

with  $\alpha_1 = (1 - \kappa)^2$ ,  $\alpha_2 = 2(1 - \kappa)\kappa\theta + \sigma^2$ ,  $\alpha_3 = (\kappa\theta)^2$ .

Furthermore, the  $\mathbb{L}_2$ -norm of  $l_\theta$  is given by:

$$\begin{aligned} \|l_\theta\|_2^2 &= \alpha_1^2 2^{-(2a+3)} c^{-3} \frac{\Gamma(2a+3)}{\Gamma^2(a)} + 2\alpha_1\alpha_2 2^{-(2a+2)} c^{-2} \frac{\Gamma(2a+2)}{\Gamma^2(a)} \\ &\quad + (2\alpha_1\alpha_3 + \alpha_2^2) 2^{-(2a+1)} c^{-1} \frac{\Gamma(2a+1)}{\Gamma^2(a)} + \alpha_2\alpha_3 2^{-(2a)} \frac{\Gamma(2a)}{\Gamma^2(a)} \\ &\quad + \alpha_3^2 2^{-2a+1} c \frac{\Gamma(2a-1)}{\Gamma^2(a)}. \end{aligned}$$

where  $\Gamma$  corresponds to the Gamma function given by

$$\Gamma(z) = \int_{\mathbb{R}_+} t^{z-1} \exp(-t) dt$$

*Proof.* see Appendix 7.4. □

**Remark 5.** In our simulation study, we take  $s_\varepsilon^2 = 0.1$  instead of  $\pi^2/2$  which is the variance of a  $\log(\chi^2)$  random variable (see [Abramowitz and Stegun, 1992]). By transformation (see Appendix 7.4) we take

$$f_\varepsilon^*(x) = \frac{1}{\sqrt{\pi}} 2^{i\beta x} \Gamma\left(\frac{1}{2} + i\beta x\right) \exp(-i\tilde{C}x)$$

with  $\beta = \sqrt{2s_\varepsilon^2/\pi^2}$  and  $\tilde{C} = \beta C$  where  $C$  is defined above.

Hence, the M-estimator solves:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \left\{ \|l_\theta\|_2^2 - \frac{2}{n} \sum_{i=1}^n Y_{i+1} u_{l_\theta}^*(Y_i) \right\} \quad (8)$$

where:

$$u_{l_\theta}(y) = \sqrt{\pi} \left( \frac{\alpha_1 \left[ \frac{a}{c^2} (a+1) \left(1 - \frac{iy}{c}\right)^{-a-2} \right] + i\alpha_2 \frac{a}{c} \left(1 - \frac{iy}{c}\right)^{-a-1} + \alpha_3 \left(1 - \frac{iy}{c}\right)^{-a}}{2^{i\beta y} \Gamma\left(\frac{1}{2} + i\beta y\right) \exp(-i\tilde{C}y)} \right)$$

## 5.2. Comparison with others methods.

5.2.1. *NLSE*:. A common and popular solution to find the vector of parameters is to find the set of parameters which gives the correct market prices of options. This is called an inverse problem since we

find indirectly the vector of parameters. The approach for solving this inverse problem is to minimize the error between the model prices and the market prices. This turns out to be a nonlinear least square optimisation problem and we can use one of the numerous algorithms of optimisation. More precisely, for each time  $t$ , we have to minimize the following function:

$$F_t(\tilde{\theta}) = \sum_{i=1}^{N_{opt}} w_i \left( C_t^{i,market} - C_t^{i,Heston}(\tilde{\theta}, Y_t) \right)^2, \quad (9)$$

with respect to  $\tilde{\theta} = (X_t, \theta)$ . The variable  $N$  is the number of available options used for the calibration,  $C_t^{i,market}$  and  $C_t^{i,Heston}(\tilde{\theta}, Y_t)$  are the  $i^{th}$  option prices given by the market and the model respectively. The variables  $w_i$  are the weights corresponding to the contribution of the  $i^{th}$  option.

The choice of the weights  $w_i$  is very important. For example, by choosing  $w_i = 1$  for all  $i$ , we assign more weight to the expensive options and less weight for the cheap options (see [Chen, 2007] for different definitions of weights).

5.2.2. *Particle filters estimators: Bootstrap, APF and KSAPF.* For the particle filters, the vector of parameters  $\theta$  is supposed random obeying the prior distribution assumed to be known. We propose to use the Kitagawa and al.'s approach (see [Doucet et al., 2001] chapter 10 p.189) in which the parameters are supposed time-varying:  $\theta_{i+1} = \theta_i + \mathcal{G}_{i+1}$  where  $\mathcal{G}_{i+1}$  is a centered Gaussian random with a variance matrix  $Q$  supposed to be known. Now, we consider the augmented state vector  $\tilde{X}_{i+1} = (X_{i+1}, \theta_{i+1})'$  where  $X_{i+1}$  is the hidden state variable and  $\theta_{i+1}$  the unknown vector of parameters. For initialisation the distribution of  $X_1$ <sup>3</sup> conditionally to  $\theta_1$  is given by the stationary density  $f_{\theta_1}$ , that is the Gamma density in our case.

For the comparison with our contrast estimator given in (8), we use the three methods: the Bootstrap filter, the Auxiliary Particle filter (APF) and the Kernel Smoothing Auxiliary Particle filter (KSAPF).

---

<sup>3</sup>To avoid confusions between the true value  $\theta_0$  and the initial value  $\theta_1$  in the Bayesian algorithms, we start the algorithms with  $i = 1$ .

We refer the reader to [Doucet et al., 2001], [Pitt and Shephard, 1999] and [Liu and West, 2001] for a complete revue of these methods.

## 6. A SIMULATION STUDY

In this section we present some Monte Carlo simulations using the model (7). For the analysis we consider  $\theta_0 = (\kappa_0, \mu_0, \sigma_0^2) = (4, 0.03, 0.4)$ . This choice is consistent with empirical applications of daily data (see [Do, 2005]). Thus, we have sampled the trajectory of the  $X_i$ , and conditionally to the trajectory, we have sampled the variables  $Y_i$  with a variance noise  $s_\varepsilon^2 = 0.1$ .

For particles methods, we take a number of particles  $M$  equal to 5000. Note that for the Bayesian procedure (Bootstrap, APF and KSAPF) we need a prior on  $\theta$ , and this only at the first step. The prior for  $\theta_1$  is taken to be the Uniform law and conditionally to  $\theta_1$  the distribution of  $X_1$  is the stationary law:

$$\begin{cases} p(\theta_1) = \mathcal{U}(3, 5) \times \mathcal{U}(0.02, 0.04) \times \mathcal{U}(0.3, 0.5) \\ f_{\theta_1}(X_1) = \Gamma(a, b) \end{cases}$$

with  $a = \frac{2\kappa\mu}{\sigma^2}$  and  $b = \frac{\sigma^2}{2\kappa}$ .

For the KSAPF, we take a bandwidth  $h = 0.1$  and  $Q = \begin{pmatrix} 1.10^{-3} & 0 & 0 \\ 0 & 0.1.10^{-4} & 0 \\ 0 & 0 & 1.10^{-4} \end{pmatrix}$  for the APF

and Bootstrap filter.

For the NLSE, we take for initial values of the parameters  $\theta_1 = (3, 0.02, 0.3)$  and we compute  $N_{opt}$  equal to 183 options prices with a maturity  $T$  between one month and one year and with a strike  $K$  between 65 and 106 (see [Heston, 1993] for the closed formula of option prices). Furthermore, in our case, we take  $w_i = 1$  for all  $i = 1, \dots, N_{opt}$ .

The minimisation is not trivial since in general the function  $F_t(\tilde{\theta})$  is neither convex nor it has any particular structure. Therefore, finding a global minimum is difficult and depends on the algorithm

used. Furthermore, there is not an unique solution for Eq.(9), in which case only local minima can be found (we refer the reader to [Moodley, 2005] in which Adaptive Simulated Annealing methods are used to find a global minima). Another drawback is the initial condition for the vector  $\theta_1$ .

**6.1. Numerical Results.** In the numerical section we compare the different estimations: the NLS estimator defined in Section 5.2.1, the Bayesian estimators defined in Section 5.2.2 and our contrast estimator defined in Eq (8). For the comparison of the computing time, we also compare our contrast estimator with the Monte Carlo Expectation Maximisation (MCEM) estimator (see [Lindstrom, 2012]).

*6.1.1. Computing time.* From a theoretical point of view, the MLE is asymptotically efficient. However, in practice since the states  $(X_1 \dots, X_n)$  are unobservable and the Heston model is non Gaussian, the likelihood is untractable. We have to use numerical methods to approximate it. In this section, we illustrate the MCEM estimator which consists in approximating the likelihood and applying the Expectation-Maximisation algorithm introduced by Dempster [Dempster et al., 1977] to find the parameter  $\theta$ .

To illustrate the MCEM for the Heston model, we run an estimator with a number of observations  $n$  equal to 1000. Although the estimation is good the computing time is very long compared with the others methods (see Table [4]). This result illustrates the numerical complexity of the MCEM. Therefore, in the following, we only compare our contrast estimator with the NLS estimator and particles estimators.

TABLE 4. MCEM estimation for Heston model.

$\hat{\kappa}_n$	$\hat{\mu}_n$	$\hat{\sigma}_n^2$	CPU (sec)
4.07	0.02892	0.3878	217430

*6.1.2. Parameter estimates.* For the Heston model, we run  $N = 100$  estimates for each method (NLS, APF, KSAPF and Bootsrap filter). The number of observations  $n$  is equal to 1000.

In order to compare the performance of our estimator with others methods, we compute for each method the Mean Square Error (MSE) defined by:

$$MSE = \frac{1}{N} \sum_{i=1}^p \sum_{j=1}^N (\hat{\theta}_j^i - \theta_0^i)^2 \quad (10)$$

where  $p$  corresponds to the dimension of the vector of parameters.

We illustrate by boxplots the different estimates (see Figures [1] up to [3]). We also illustrate in Table [5] the MSE for each estimator computed by equation (10) and the CPU for a number of observations  $n = 1000$ .

We note that for all parameters, the EKF estimator is very bad since the Heston model is strongly nonlinear, and its corresponding boxplots have the largest dispersion meaning that this filter is not stable and not appropriated to estimate this model. Among particle filters, the KSAPF and the APF are the best estimators although the dispersion is huge for the mean reversion parameter  $\kappa$  and the volatility parameter  $\sigma$ .

Besides the Bootstrap filter is less efficient than the others particle filters. Our estimator and the NLSE are stable, and if we compare the MSE, it is smallest for the contrast estimator. From a computational point of view, all particles filters have a CPU equivalent. Besides, we can see that the NLSE is faster than our contrast estimator since for the Heston model the function  $u_{l_\theta}$  has not an explicit form, so the function  $u_{l_\theta}^*$  is approximated numerically in our approach.<sup>4</sup> In spite of this approximation, our contrast estimator is fast and its implementation is straightforward and the MSE is smaller (see Table 5).

---

<sup>4</sup>We use a quadrature method implemented in Matlab to approximate the Fourier transform of  $u_{l_\theta}(y)$ . One can also use the FFT method since the computation of the contrast estimator will be faster in this case.

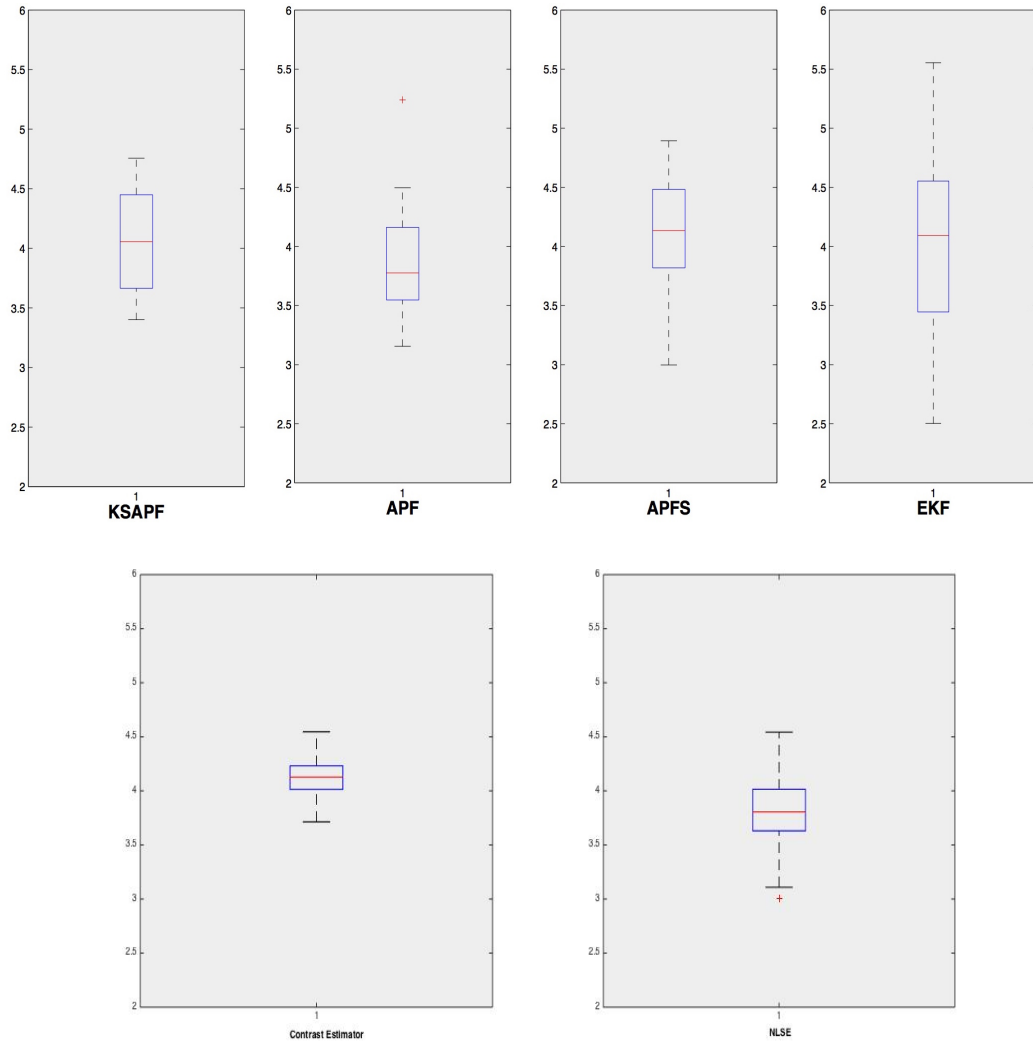


FIGURE 1. Boxplot of the parameter  $\kappa$ . True value equal to 4.



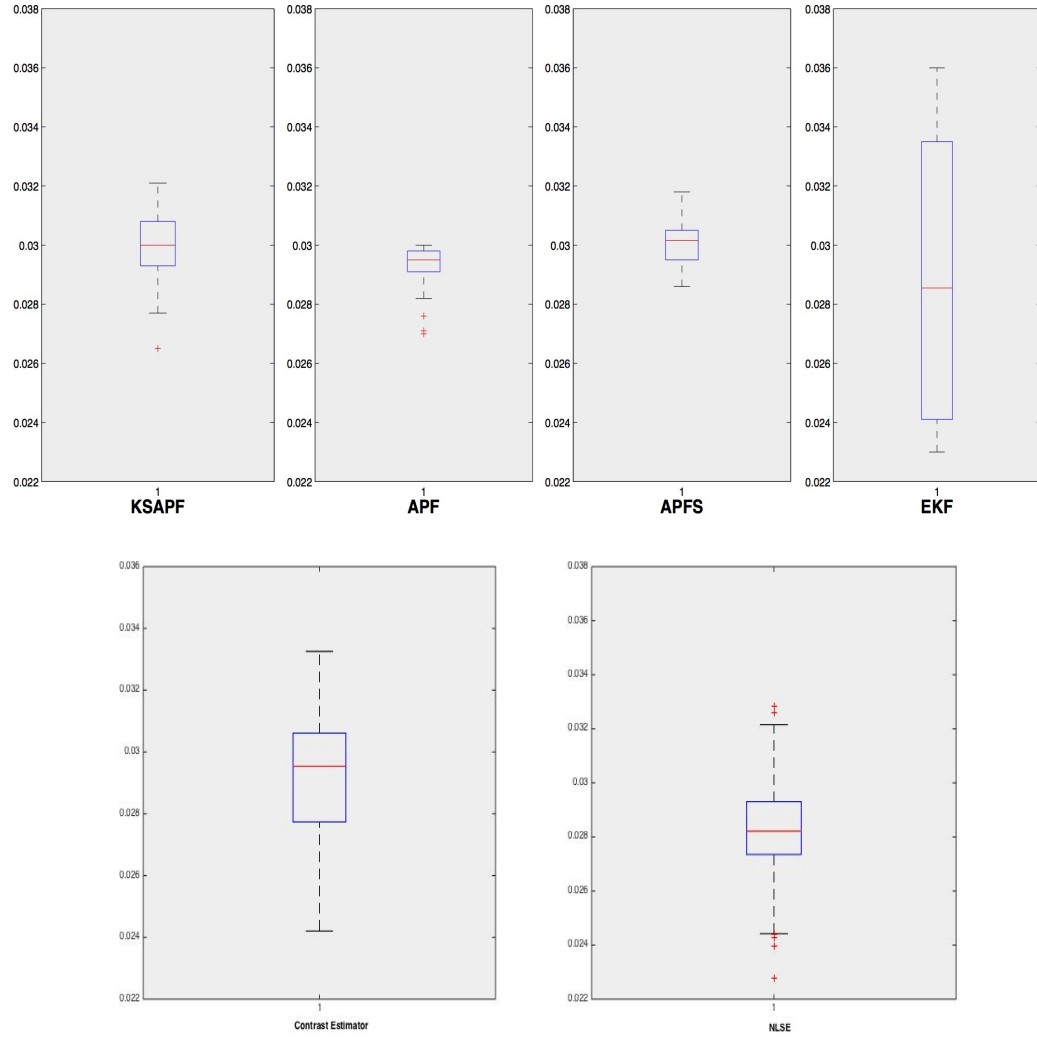


FIGURE 2. Boxplot of the parameter  $\mu$ . True value equal to 0.03.

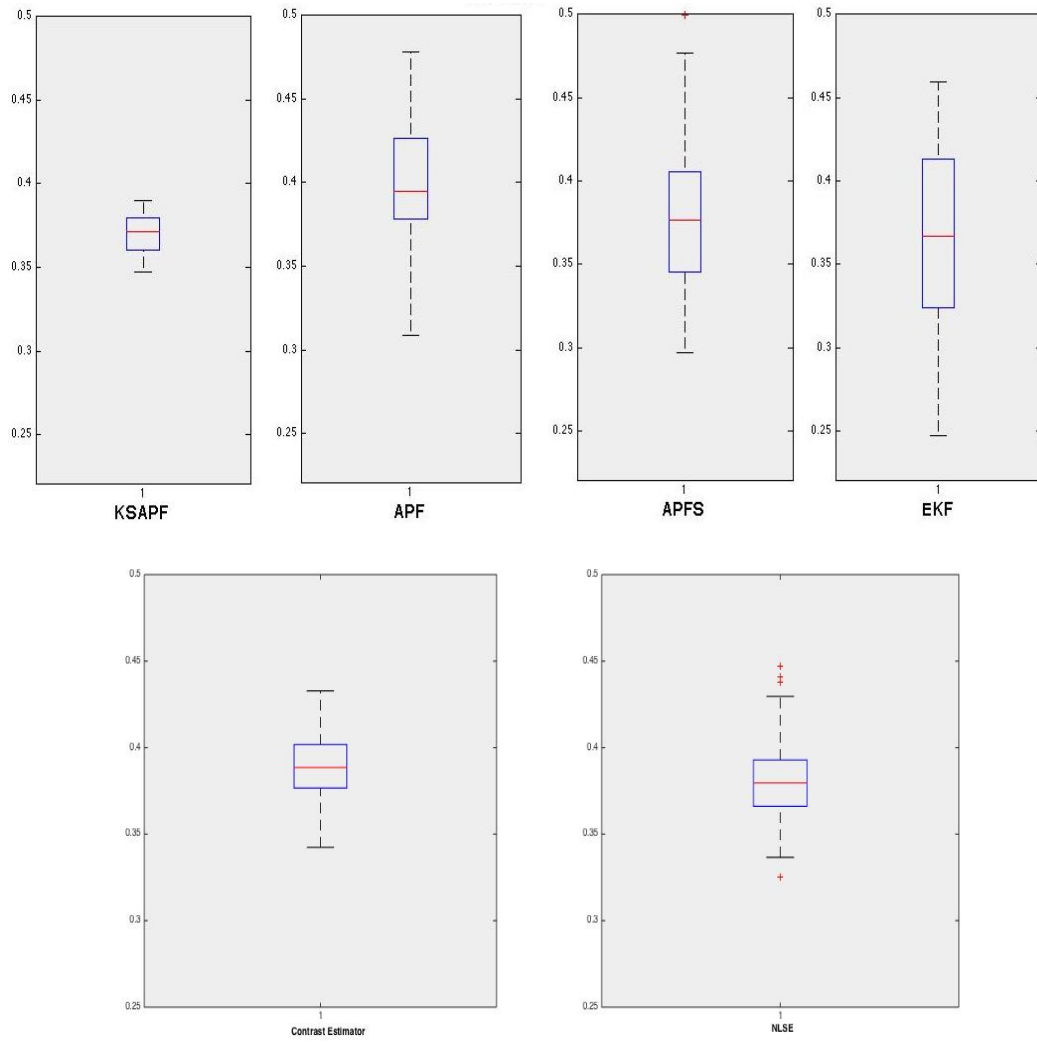


FIGURE 3. Boxplot of the parameter  $\sigma^2$ . True value equal to 0.4.

	CPU	MSE
Contrast	20.4074	0.124
Bootstrap	192.2166	0.205
EKF	0.2	0.43
APF	105.1695	0.189
KSAPF	93.8846	0.169
NLSE	0.0702	0.198

TABLE 5. Comparison of the computing time (CPU in seconds) and the MSE for the number of observations  $n = 1000$ . The number of estimators is  $N = 100$  for the MSE (see Eq.(10)).

6.1.3. *Confidence Interval of the contrast estimator.* To illustrate the statistical properties of our contrast estimator, we compute the confidence intervals computed with the confidence level  $1 - \alpha$  equal to 0.95 for  $N = 1$  estimator. The coverage corresponds to the number of times for which the true parameter  $\theta_0^i, i = 1, \dots, p$  belongs to the confidence interval. The results are illustrated in Figure [4]. We note that the coverage converges to 95% for a small number of observations and as expected, the confidence interval decreases with the number of observations,. Note that of course a MLE confidence interval would be smaller since the MLE is efficient but the corresponding computing time would be huge (see Table 4).

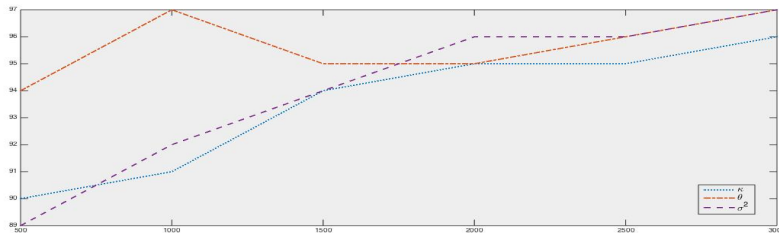


FIGURE 4. Coverage with respect to the number of observations  $n = 500$  up to 1000 for  $N = 100$  estimators.

6.1.4. *Ratio signal-noise of the contrast estimator.* We denote by  $r = \frac{\theta_0^2}{\sigma^2}$  the ratio signal-noise and in Table (6) we compare the MSE for different  $r$  and different number of observations  $n$  for the contrast

estimator. We note that the MSE decreases with the number of observations and is smaller for small ratio-signal-noise. As we explained in Section 4.1, see [Comte et al., 2010] for more details, the rate of convergence of our approach depends on the regularity of the noise density  $f_\varepsilon$ . And, in particular, the smoother are the noises, the slower the rate of convergence is. For the Heston model, the density of the noises and the function  $l_\theta$  are ordinary smooth, so we are in a favourable case.

TABLE 6. Ratio Signal-noise for the estimation of the Heston model

	Mean( $\hat{\mu}_n$ )	Mean( $\hat{\kappa}_n$ )	Mean( $\hat{\sigma}_n^2$ )	MSE
$n = 500$ and $r = 0.1$	0.0315	3.88	0.401	0.14
$n = 500$ and $r = 1$	0.0303	3.89	0.405	0.16
$n = 1000$ and $r = 0.1$	0.0312	3.76	0.401	0.11
$n = 1000$ and $r = 1$	0.0308	3.83	0.41	0.18

**6.2. Summary and Conclusions.** In this paper we have proposed a new method to estimate hidden nonlinear diffusion process. This method is based on a deconvolution strategy and leads to consistent and asymptotically normal estimator. We have empirically studied the performance of our estimator in the Heston model and we were able to construct confidence interval (see Figure [4]). As the boxplots [1] up to [3] show, only Contrast, NLS, APF, and KSAPF estimators are comparable. Indeed EKF and Bootstrap Filter estimators are biased and their MSE are bad, especially for the EKF method since the Heston model is nonlinear. Furthermore, if one compares the MSE of the particle filters, the KSAPF estimator is the best method. Among particles filters, it is clearly known that the Bootstrap filter is less efficient than the APF filter since the propagation step of the particles is made according to the transition density which doesn't take care the observations.

Then, the Contrast, NLS APF, and KSAPF methods lead to unbiased and not so much varying estimator. We emphasize that our estimator performs the others in a MSE aspect (see Table 5). Most

importantly, our estimator can be constructed without any arbitrary parameters choice, is straightforward to implement, fast and allows to construct confidence interval.

## REFERENCES

- [Abramowitz and Stegun, 1992] Abramowitz, M. and Stegun, I. A., editors (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications Inc., New York. Reprint of the 1972 edition.
- [Chen, 2007] Chen, B. (2007). *Calibration of the Heston model with Application in Derivative Pricing and Hedging*. PhD thesis.
- [Comte et al., 2010] Comte, F., Lacour, C., and Rozenholc, Y. (2010). Adaptive estimation of the dynamics of a discrete time stochastic volatility model. *J. Econometrics*, 154(1):59–73.
- [Comte et al., 2006] Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34(3):431–452.
- [Comte and Taupin, 2001] Comte, F. and Taupin, M.-L. (2001). Semiparametric estimation in the (auto)-regressive  $\beta$ -mixing model with errors-in-variables. *Math. Methods Statist.*, 10(2):121–160.
- [Comte and Taupin, 2007] Comte, F. and Taupin, M.-L. (2007). Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statist. Sinica*, 17(3):1065–1090.
- [Cox et al., 1985] Cox, J. C., Ingersoll, Jr., J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.
- [Dedecker et al., 2007] Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S., and Prieur, C. (2007). *Weak dependence: with examples and applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York.
- [Dedecker et al., 2014] Dedecker, J., Samson, A., and Taupin, M.-L. (2014). Estimation in autoregressive model with measurement error. *ESAIM Probab. Stat.*, 18:277–307.
- [Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38. With discussion.
- [Do, 2005] Do, B. (2005). *Estimating the Heston stochastic volatility model from option prices by filtering. A simulation exercise*.
- [Doucet et al., 2001] Doucet, A., de Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York.
- [El Kolei, 2013] El Kolei, S. (2013). Parametric estimation of hidden stochastic model by contrast minimization and deconvolution. *Metrika*, 76(8):1031–1081.
- [Fan et al., 1990] Fan, J., Truong, Y. K., and Wang, Y. (1990). Nonparametric function estimation involving errors-in-variables.
- [Genon-Catalot et al., 1999] Genon-Catalot, V., Jeantheau, T., and Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5(5):855–872.

- [Hansen and Horowitz, 1997] Hansen, B. E. and Horowitz, J. L. (1997). Handbook of econometrics, vol. 4 robert f. engle and daniel l. mcfadden, editors elsevier science b. v., 1994. *Econometric Theory*, 13(01):119–132.
- [Hayashi, 2000] Hayashi, F., editor (2000). *Econometrics*. Princeton University Press, Princeton NJ.
- [Heston, 1993] Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6(2):327–43.
- [Jones, 2004] Jones, G. L. (2004). On the Markov chain central limit theorem. *Probab. Surv.*, 1:299–320.
- [Lindstrom, 2012] Lindstrom, E. (2012). A monte carlo em algorithm for discretely observed diffusions, jump-diffusions and lvy-driven stochastic differential equations. *n International journal of mathematical models and methods in applied sciences 6(5)*, pages 643–651.
- [Liu and West, 2001] Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. in arnaud doucet, nando de freitas, and neil gordon, editors. *Sequential Monte Carlo Methods in Practice*.
- [Moodley, 2005] Moodley, N. (2005). The heston model: A pratical approach with matlab code. Technical report.
- [Newey and McFadden, 1994] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam.
- [Pitt and Shephard, 1999] Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.*, 94(446):590–599.
- [Taylor, 2005] Taylor, S. (2005). *Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices.*, volume 1, pages 203–226. Oxford University Press.
- [Van der Vaart, 1998] Van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

## 7. APPENDIX: PROOFS

For the reader convenience we split the proof of Theorems 1 and 2 into three parts: in Subsection 7.1, we give the proof of the existence of our contrast estimator defined in (5). In Subsection 7.2, we prove the consistency, that is, the Theorem 1. Then, we prove the asymptotic normality of our estimator in Subsection 7.3, that is, the Theorem 2. The Subsection 7.4 is devoted to Corollary 1.

Recall from Remark 2 that we only made the proof for the function  $\varphi$  defined by (2) in Definition 1 and we refer to [El Kolei, 2013] for the proof in the case (1) of Definition 1.

**7.1. Proof of the existence and measurability of the M-Estimator.** By assumption, the function  $\theta \mapsto \|l_\theta\|_2^2$  is continuous. Moreover,  $l_\theta^*$  and then  $u_{l_\theta}^*(x) = \frac{1}{2\pi} \int e^{ixy} \frac{l_\theta^*(-y)}{f_\varepsilon^*(y)} dy$  are continuous w.r.t  $\theta$ . In particular, the function  $m_\theta(\mathbf{y}_i) = \|l_\theta\|_2^2 - 2\varphi(y_{i+1})u_{l_\theta}^*(y_i)$  is continuous w.r.t  $\theta$ , for  $\varphi : x \mapsto x^2 - s_\varepsilon^2$ . Hence, the function  $\mathbf{P}_n m_\theta = \frac{1}{n} \sum_{i=1}^n m_\theta(\mathbf{Y}_i)$  is continuous w.r.t  $\theta$  belonging to the compact subset  $\Theta$ . So, there exists  $\tilde{\theta}$  that belongs to  $\Theta$  such that:

$$\inf_{\theta \in \Theta} \mathbf{P}_n m_\theta = \mathbf{P}_n m_{\tilde{\theta}}. \quad \square$$

**7.2. Proof of the Consistency.** For the consistency of our estimator, we need to use the uniform convergence given in the following Lemma. Let us consider the following quantities:

$$\mathbf{P}_n h_\theta = \frac{1}{n} \sum_{i=1}^n h_\theta(Y_i); \quad \mathbf{P}_n S_\theta = \frac{1}{n} \sum_{i=1}^n \nabla_\theta h_\theta(Y_i) \quad \text{and} \quad \mathbf{P}_n H_\theta = \frac{1}{n} \sum_{i=1}^n \nabla_\theta^2 h_\theta(Y_i)$$

where  $h_\theta(y)$  is real function from  $\Theta \times \mathcal{Y}$  with value in  $\mathbb{R}$ .

**Lemma 1.** *Uniform Law of Large Numbers (ULLN)(see [Newey and McFadden, 1994] for the proof.)*

Let  $(Y_i)$  be an ergodic stationary process and suppose that:

- (1)  $h_\theta(y)$  is continuous in  $\theta$  for all  $y$  and measurable in  $y$  for all  $\theta$  in the compact subset  $\Theta$ .
- (2) There exists a function  $s(y)$  (called the dominating function) such that  $|h_\theta(y)| \leq s(y)$  for all  $\theta \in \Theta$  and  $\mathbb{E}[s(Y_1)] < \infty$ . Then:

$$\sup_{\theta \in \Theta} |\mathbf{P}_n h_\theta - \mathbf{P} h_\theta| \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty.$$

Moreover,  $\mathbf{P}h_\theta$  is a continuous function of  $\theta$ .

By assumption  $l_\theta$  is continuous w.r.t  $\theta$  for any  $x$  and measurable w.r.t  $x$  for all  $\theta$  which implies the continuity and the measurability of the function  $\mathbf{P}_n m_\theta$  on the compact subset  $\Theta$ . Furthermore, the local dominance assumption **(A9)** implies that  $\mathbb{E}[\sup_{\theta \in \Theta} |m_\theta(\mathbf{Y}_i)|]$  is finite. Indeed,

$$\begin{aligned} |m_\theta(\mathbf{y}_i)| &= \left| \|l_\theta\|_2^2 - 2\varphi(y_{i+1})u_{l_\theta}^*(y_i) \right| \\ &\leq \|l_\theta\|_2^2 + 2|\varphi(y_{i+1})u_{l_\theta}^*(y_i)|. \end{aligned}$$

with  $\varphi$  the function  $x \mapsto x^2 - s_\varepsilon^2$ .

As  $\|l_\theta\|_2^2$  is continuous on the compact subset  $\Theta$ ,  $\sup_{\theta \in \Theta} \|l_\theta\|_2^2$  is finite. Therefore,  $\mathbb{E}[\sup_{\theta \in \Theta} |m_\theta(\mathbf{Y}_i)|]$  is finite if  $\mathbb{E}[\sup_{\theta \in \Theta} |\varphi(Y_{i+1})u_{l_\theta}^*(Y_i)|]$  is finite. Lemma *ULLN 1* gives us the uniform convergence in probability of the contrast function: for any  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \sup_{\theta \in \Theta} |\mathbf{P}_n m_\theta - \mathbf{P} m_\theta| \leq \varepsilon \right) = 1.$$

Combining the uniform convergence with Theorem 2.1 p. 2121 chapter 36 in [[Hansen and Horowitz, 1997](#)] yields the weak (convergence in probability) consistency of the estimator.  $\square$

**7.3. Proof of the asymptotic normality.** Consider the model (1) under the assumptions **A0-A8**. The proof of the asymptotic normality results from assumptions **A9-A14** and is a straightforward application of Hayashi (2000, proposition 7.8. p. 472) and Galin (2004). In this respect, we need to check that

- (1) (Moment condition):  $\mathbb{E}[|m_\theta(\mathbf{Y}_i)|^{2+\delta}]$  for some  $\delta > 0$ .
- (2) (Hessian Local condition): For some neighbourhood  $\mathcal{U}$  of  $\theta_0$ :  $\mathbb{E}[\sup_{\theta \in \mathcal{U}} \|\nabla_\theta^2 m_\theta(\mathbf{Y}_i)\|] < \infty$ .

The proof is based on the following Lemma:

**Lemma 2.** *Suppose that the conditions of the consistency hold. Suppose further that:*

- (1)  $\mathbf{Y}_i$  is  $\alpha$ -mixing.
- (2) (Moment condition): for some  $\delta > 0$  and for each  $j \in \{1, \dots, r\}$ :

$$\mathbb{E} \left[ \left| \frac{\partial m_\theta(\mathbf{Y}_1)}{\partial \theta_j} \right|^{2+\delta} \right] < \infty$$



(3) Assumption (C) given in Section 4.2.2 holds such that  $\mathbb{E}[M(X_1)] < \infty$  and  $\gamma(n)$  satisfies  $\sum_n \gamma(n)^{\frac{\delta}{2+\delta}} < \infty$  where  $\delta$  is given in condition 2.

(4) (Hessian Local condition): for some neighbourhood  $\mathcal{U}$  of  $\theta_0$  and for  $j, k \in \{1, \dots, r\}$

$$\mathbb{E} \left[ \sup_{\theta \in \mathcal{U}} \left| \frac{\partial^2 m_\theta(\mathbf{Y}_1)}{\partial \theta_j \partial \theta_k} \right| \right] < \infty.$$

Then,  $\widehat{\theta}_n$  defined in Eq.(5) is asymptotically normal with asymptotic covariance matrix given by:

$$\Sigma(\theta_0) = V_{\theta_0}^{-1} \Omega(\theta_0) V_{\theta_0}^{-1}$$

where  $V_{\theta_0}$  is the Hessian of the application  $\mathbf{P}m_\theta$  given in Eq.(4).

*Proof.* The proof follows from Fumio's [Hayashi, 2000] and Galin Proposition [Jones, 2004].  $\square$

It just remains to check that the conditions (2) and (4) of Lemma 2 hold under our assumptions.

Moment condition. As the function  $l_\theta$  is twice continuously differentiable w.r.t  $\theta$ , for all  $\mathbf{y}_i \in \mathbb{R}^2$ , the application  $m_\theta(\mathbf{y}_i) : \theta \in \Theta \mapsto m_\theta(\mathbf{y}_i) = \|l_\theta\|_2^2 - 2\varphi(y_{i+1})u_{l_\theta}^*(y_i)$  is twice continuously differentiable for all  $\theta \in \Theta$  and its first derivatives are given by:

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \nabla_\theta \|l_\theta\|_2^2 - 2\varphi(y_{i+1})\nabla_\theta u_{l_\theta}^*(y_i).$$

By assumption, for each  $j \in \{1, \dots, r\}$ ,  $\frac{\partial l_\theta}{\partial \theta_j} \in \mathbb{L}_1(\mathbb{R})$ , therefore one can apply the Lebesgue differentiation Theorem and Fubini's Theorem to obtain :

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \left[ \nabla_\theta \|l_\theta\|_2^2 - 2\varphi(y_{i+1})u_{\nabla_\theta l_\theta}^*(y_i) \right]. \quad (11)$$

Then, for some  $\delta > 0$ :

$$\begin{aligned} |\nabla_\theta m_\theta(\mathbf{y}_i)|^{2+\delta} &= \left| \nabla_\theta \|l_\theta\|_2^2 - 2\varphi(y_{i+1})u_{\nabla_\theta l_\theta}^*(y_i) \right|^{2+\delta} \\ &\leq C_1 \left| \nabla_\theta \|l_\theta\|_2^2 \right|^{2+\delta} + C_2 \left| \varphi(y_{i+1})u_{\nabla_\theta l_\theta}^*(y_i) \right|^{2+\delta}, \end{aligned} \quad (12)$$

where  $C_1$  and  $C_2$  are two positive constants. By assumption, the function  $\|l_\theta\|_2^2$  is twice continuously differentiable w.r.t  $\theta$ . Hence,  $\nabla_\theta \|l_\theta\|_2^2$  is continuous on the compact subset  $\Theta$  and the first term of equation (12) is

finite. The second term is finite by the moment assumption **(A12)**.

Hessian Local dominance. For  $j, k \in \{1, \dots, r\}$ ,  $\frac{\partial^2 l_\theta}{\partial \theta_j \partial \theta_k} \in \mathbb{L}_1(\mathbb{R})$ , the Lebesgue differentiation Theorem gives:

$$\nabla_\theta^2 m_\theta(\mathbf{y}_i) = \nabla_\theta^2 \|l_\theta\|_2^2 - 2\varphi(y_{i+1})u_{\nabla_\theta^2 l_\theta}^*(y_i),$$

and, for some neighbourhood  $\mathcal{U}$  of  $\theta_0$ :

$$\mathbb{E} \left[ \sup_{\theta \in \mathcal{U}} \|\nabla_\theta^2 m_\theta(\mathbf{Y}_i)\| \right] \leq \sup_{\theta \in \mathcal{U}} \|\nabla_\theta^2 \|l_\theta\|_2^2\| + 2\mathbb{E} \left[ \sup_{\theta \in \mathcal{U}} \|\varphi(Y_{i+1})u_{\nabla_\theta^2 l_\theta}^*(Y_i)\| \right].$$

The first term of the above equation is finite by continuity and compactness argument. And, the second term is finite by the Hessian local dominance assumption **(A13)**.  $\square$

**7.4. Proof of Corollary 1.** By replacing  $\nabla_\theta m_\theta(\mathbf{Y}_1)$  by its expression (11), we have:

$$\begin{aligned} \Omega_0(\theta) &= \text{Var} \left[ \nabla_\theta \|l_\theta\|_2^2 - 2\varphi(Y_2)u_{\nabla_\theta l_\theta}^*(Y_1) \right] \\ &= 4\text{Var} \left[ \varphi(Y_2)u_{\nabla_\theta l_\theta}^*(Y_1) \right] \\ &= 4 \left[ \mathbb{E} \left[ \varphi(Y_2)^2 (u_{\nabla_\theta l_\theta}^*(Y_1)) (u_{\nabla_\theta l_\theta}^*(Y_1))' \right] - \mathbb{E} \left[ \varphi(Y_2)u_{\nabla_\theta l_\theta}^*(Y_1) \right] \mathbb{E} \left[ \varphi(Y_2)u_{\nabla_\theta l_\theta}^*(Y_1) \right]' \right]. \end{aligned}$$

Furthermore, by Eq.(1) and by independence of the centered noise  $(\varepsilon_2)$  and  $(\eta_2)$ , we have:

$$\mathbb{E} \left[ \varphi(Y_2)u_{\nabla_\theta l_\theta}^*(Y_1) \right] = \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)u_{\nabla_\theta l_\theta}^*(Y_1) \right].$$

Using Fubini's Theorem and Eq.(1) we obtain:

$$\begin{aligned} \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)u_{\nabla_\theta l_\theta}^*(Y_1) \right] &= \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1) \int e^{iY_1 z} u_{\nabla_\theta l_\theta}(z) dz \right] \\ &= \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1) \int \frac{1}{2\pi} \frac{1}{f_\varepsilon^*(z)} e^{iY_1 z} (\nabla_\theta l_\theta)^*(-z) dz \right] \\ &= \frac{1}{2\pi} \int \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1) e^{i(X_1 + \varepsilon_1)z} \right] \frac{1}{f_\varepsilon^*(z)} (\nabla_\theta l_\theta)^*(-z) dz \\ &= \frac{1}{2\pi} \int \frac{\mathbb{E} [e^{i\varepsilon_1 z}]}{f_\varepsilon^*(z)} \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1) e^{iX_1 z} \right] (\nabla_\theta l_\theta)^*(-z) dz, \end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_1)] &= \frac{1}{2\pi} \int \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)e^{i(X_1+\varepsilon_1)z}] \frac{1}{f_{\varepsilon}^*(z)} (\nabla_{\theta}l_{\theta})^*(-z) dz \\
&= \frac{1}{2\pi} \int \frac{\mathbb{E} [e^{i\varepsilon_1 z}]}{f_{\varepsilon}^*(z)} \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)e^{iX_1 z}] (\nabla_{\theta}l_{\theta})^*(-z) dz \\
&= \frac{1}{2\pi} \mathbb{E} \left[ (b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1) \int e^{iX_1 z} (\nabla_{\theta}l_{\theta})^*(-z) dz \right] \\
&= \frac{1}{2\pi} \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1) ((\nabla_{\theta}l_{\theta})^*(-X_1))^*] \\
&= \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)\nabla_{\theta}l_{\theta}(X_1)]. \tag{13}
\end{aligned}$$

Hence,

$$\Omega_0(\theta) = 4(P_2 - P_1),$$

where

$$\begin{aligned}
P_1 &= \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)\nabla_{\theta}l_{\theta}(X_1)] \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)\nabla_{\theta}l_{\theta}(X_1)]', \\
P_2 &= \mathbb{E} [\varphi(Y_2)^2 (u_{\nabla_{\theta}l_{\theta}}^*(Y_1)) (u_{\nabla_{\theta}l_{\theta}}^*(Y_1))'].
\end{aligned}$$

*Calculus of the covariance matrix of Corollary (1):* By replacing  $(\nabla_{\theta}m_{\theta}(Y_1))$  by its expression (11) we have:

$$\begin{aligned}
\Omega_{j-1}(\theta) &= \text{Cov} (\nabla_{\theta}\|l_{\theta}\|_2^2 - 2\varphi(Y_2)u_{\nabla_{\theta}l_{\theta}}^*(Y_1), \nabla_{\theta}\|l_{\theta}\|_2^2 - 2\varphi(Y_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_j)), \\
&= 4\text{Cov} (\varphi(Y_2)u_{\nabla_{\theta}l_{\theta}}^*(Y_1), \varphi(Y_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_j)), \\
&= 4 \left[ \mathbb{E} (\varphi(Y_2)u_{\nabla_{\theta}l_{\theta}}^*(Y_1)\varphi(Y_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_j)) - \mathbb{E} (\varphi(Y_2)u_{\nabla_{\theta}l_{\theta}}^*(Y_1)) \mathbb{E} (\varphi(Y_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_j))' \right].
\end{aligned}$$

By using Eq.(13) and the stationary property of the  $Y_i$ , one can replace the second term of the above equation by:

$$\mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)\nabla_{\theta}l_{\theta}(X_1)] \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)\nabla_{\theta}l_{\theta}(X_1)]'.$$

Furthermore, by using Eq.(1) we obtain:

$$\begin{aligned} \mathbb{E} [\varphi(Y_2)\varphi(Y_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_j)] &= \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_j)u_{\nabla_{\theta}l_{\theta}}^*(Y_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_j)] \\ &+ \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(\eta_{j+1} + \varepsilon_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_j)] \end{aligned} \quad (14)$$

$$+ \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_j)(\eta_2 + \varepsilon_2)u_{\nabla_{\theta}l_{\theta}}^*(Y_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_j)] \quad (15)$$

$$+ \mathbb{E} [(\eta_2 + \varepsilon_2)(\eta_{j+1} + \varepsilon_{j+1})u_{\nabla_{\theta}l_{\theta}}^*(Y_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_j)]. \quad (16)$$

By independence of the centered noise, the term (14), (15) and (16) are equal to zero. Now, if we use Fubini's Theorem we have:

$$\mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_j)u_{\nabla_{\theta}l_{\theta}}^*(Y_1)u_{\nabla_{\theta}l_{\theta}}^*(Y_j)] = \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_j)\nabla_{\theta}l_{\theta}(X_1)\nabla_{\theta}l_{\theta}(X_j)]. \quad (17)$$

Hence, the covariance matrix is given by:

$$\begin{aligned} \Omega_{j-1}(\theta) &= 4 \left( \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_j)(\nabla_{\theta}l_{\theta}(X_1))(\nabla_{\theta}l_{\theta}(X_j))'] \right. \\ &\quad \left. - \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(\nabla_{\theta}l_{\theta}(X_1))] \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(\nabla_{\theta}l_{\theta}(X_1))]' \right) \\ &= 4 \left( \tilde{C}_{j-1} - \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(\nabla_{\theta}l_{\theta}(X_1))] \mathbb{E} [(b_{\theta_0}^2 + \sigma_{\theta_0}^2)(X_1)(\nabla_{\theta}l_{\theta}(X_1))]' \right) \\ &= 4 \left( \tilde{C}_{j-1} - P_1 \right). \end{aligned}$$

Finally, we obtain:  $\Omega(\theta) = \Omega_0(\theta) + 2 \sum_{j>1}^{\infty} \Omega_{j-1}(\theta)$  with  $\Omega_0(\theta) = 4(P_2 - P_1)$  and  $\Omega_{j-1}(\theta) = 4(\tilde{C}_{j-1} - P_1)$ .

*Expression of the Hessian matrix  $V_{\theta}$  :* We have:

$$\mathbf{P}m_{\theta} = \|l_{\theta}\|_2^2 - 2 \langle l_{\theta}, l_{\theta_0} \rangle. \quad (18)$$

For all  $\theta$  in  $\Theta$ , the application  $\theta \mapsto \mathbf{P}m_{\theta}$  is twice differentiable w.r.t  $\theta$  on the compact subset  $\Theta$ . And for  $j \in \{1, \dots, r\}$ :

$$\begin{aligned}
\frac{\partial \mathbf{P}m}{\partial \theta_j}(\theta) &= 2 \left\langle \frac{\partial l_\theta}{\partial \theta_j}, l_\theta \right\rangle - 2 \left\langle \frac{\partial l_\theta}{\partial \theta_j}, l_{\theta_0} \right\rangle \\
&= 2 \left\langle \frac{\partial l_\theta}{\partial \theta_j}, l_\theta - l_{\theta_0} \right\rangle, \\
&= 0 \text{ at the point } \theta_0,
\end{aligned}$$

and for  $j, k \in \{1, \dots, r\}$ :

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}m}{\partial \theta_j \partial \theta_k}(\theta) &= 2 \left( \left\langle \frac{\partial^2 l_\theta}{\partial \theta_j \partial \theta_k}, l_\theta - l_{\theta_0} \right\rangle + \left\langle \frac{\partial l_\theta}{\partial \theta_k}, \frac{\partial l_\theta}{\partial \theta_j} \right\rangle \right)_{j,k} \\
&= 2 \left( \left\langle \frac{\partial l_\theta}{\partial \theta_k}, \frac{\partial l_\theta}{\partial \theta_j} \right\rangle \right)_{j,k} \text{ at the point } \theta_0. \square
\end{aligned}$$

### APPENDIX 3: M-ESTIMATOR USING THE EXAMPLE IN SECTION 5

Expression of  $f_\varepsilon^*$ . Consider the random variable  $\bar{\varepsilon} = \frac{\varepsilon - C}{\sqrt{V}}$  with  $\varepsilon = \log(X^2)$  where  $X$  is standard Gaussian random variable,  $C = \mathbb{E}[\log(X^2)]$  and  $V = \mathbb{V}[\log(X^2)]$ . The Fourier transform of  $\bar{\varepsilon}$  is given by:

$$\begin{aligned}
\mathbb{E}[\exp(i\bar{\varepsilon}y)] &= \exp\left(-\frac{iC}{\sqrt{V}}y\right) \mathbb{E}[\exp(i\varepsilon y)] \\
&= \exp\left(-\frac{iC}{\sqrt{V}}y\right) \mathbb{E}\left[X^{\frac{2iy}{\sqrt{V}}}\right] \\
&= \exp\left(-\frac{iC}{\sqrt{V}}y\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{\frac{2iy}{\sqrt{V}}} \exp\left(-\frac{x^2}{2}\right) dx
\end{aligned}$$

Using a change of variable  $z = \frac{x^2}{2}$ , we get:

$$\begin{aligned}
\mathbb{E}[\exp(i\bar{\varepsilon}y)] &= \exp\left(-\frac{iC}{\sqrt{V}}y\right) \frac{2^{\frac{iy}{\sqrt{V}}}}{\sqrt{\pi}} \int_0^{+\infty} z^{\frac{iy}{\sqrt{V}} - \frac{1}{2}} e^{-z} dz \\
&\equiv \exp\left(-\frac{iC}{\sqrt{V}}y\right) \frac{2^{\frac{iy}{\sqrt{V}}}}{\sqrt{\pi}} \Gamma\left(\frac{1}{2} + \frac{iy}{\sqrt{V}}\right).
\end{aligned}$$

Then

$$\begin{aligned}
f_\varepsilon^* &= \exp\left(-\frac{iC}{\sqrt{V}}y\right) \frac{2^{\frac{iy}{\sqrt{V}}}}{\sqrt{\pi}} \int_0^{+\infty} z^{\frac{iy}{\sqrt{V}} - \frac{1}{2}} e^{-z} dz \\
&\equiv \exp\left(-\frac{iC}{\sqrt{V}}y\right) \frac{2^{\frac{iy}{\sqrt{V}}}}{\sqrt{\pi}} \Gamma\left(\frac{1}{2} + \frac{iy}{\sqrt{V}}\right).
\end{aligned}$$

Discrete time stochastic volatility model: Heston model. Taking that  $\eta_{i+1} \sim \mathcal{N}(0, 1)$  and  $\varepsilon_i$  has a (log-) Chi-squared probability density function, if the Feller's condition holds true ( $a = \frac{2\kappa\theta}{\sigma^2} \geq 1$ ) and  $c = \frac{2\kappa}{\sigma^2} > 0$ , then the volatility process  $X_i$  is stationary and ergodic and  $\rho$ -mixing. The stationary distribution  $f_\theta$  writes:

$$f_\theta(x) = \frac{c^a}{\Gamma(a)} x^{a-1} e^{-cx} \quad \forall x > 0.$$

This is the gamma distribution  $\Gamma(a, c)$  (Genon-Catalot et al., 2000). On the other hand, the functions  $b_\theta(x)$ ,  $\sigma_\theta(x)$  and  $l_\theta(x)$  are given by:

$$b_\theta(x) = (1 - \kappa)x + \kappa\theta \text{ and } \sigma_\theta(x) = \sigma\sqrt{x}$$

$$\begin{aligned} l_\theta &= (b_\theta^2(x) + \sigma_\theta^2(x)) \Gamma(a, c) \\ &= \left( (1 - \kappa)^2 x^2 + 2x(1 - \kappa)\kappa\theta + \sigma^2 + (\kappa\theta)^2 \right) f_\theta(x), \\ &= (\alpha_1 x^2 + \alpha_2 x + \alpha_3) f_\theta(x). \end{aligned}$$

where  $\theta = (\kappa, \mu, \sigma)$  and  $\alpha_1 = (1 - \kappa)^2$ ,  $\alpha_2 = 2(1 - \kappa)\kappa\theta + \sigma^2$ ,  $\alpha_3 = (\kappa\theta)^2$ .

Therefore

$$\begin{aligned} l_\theta^*(t) &= \mathbb{E}[e^{itX} l_\theta(X)] \\ &= \alpha_1 \mathbb{E}[X^2 e^{itX}] + \alpha_2 \mathbb{E}[X e^{itX}] + \alpha_3 \mathbb{E}[e^{itX}] \text{ with } X \sim \Gamma(a, c) \\ &= -\alpha_1 \frac{\partial^2 f_\theta^*}{\partial t^2}(t) - i\alpha_2 \frac{\partial f_\theta^*}{\partial t}(t) + \alpha_3 f_\theta^*(t). \end{aligned}$$

After replacing  $f_\theta^*(t)$  by  $(1 - \frac{it}{c})^{-a}$ , we obtain:

$$l_\theta^*(x) = -\alpha_1 \left[ \frac{-a}{c^2} (a+1) \left(1 - \frac{ix}{c}\right)^{-a-2} \right] + i\alpha_2 \frac{a}{c} \left(1 - \frac{ix}{c}\right)^{-a-1} + \alpha_3 \left(1 - \frac{ix}{c}\right)^{-a}.$$

It follows that the squared norm of  $l_\theta(x)$  is given by:

$$\begin{aligned} \|l_\theta\|^2 &= \int (b_\theta^2 + \sigma_\theta^2(x))^2 \Gamma^2(a, c) dx \\ &= \int (\beta_1 x^4 + \beta_2 x^3 + \beta_3 x^2 + \beta_4 x + \beta_5) \Gamma^2(a, c) dx, \end{aligned}$$

where  $\beta_1 = \alpha_1^2$ ,  $\beta_2 = 2\alpha_1\alpha_2$ ,  $\beta_3 = 2\alpha_1\alpha_3 + \alpha_2^2$ ,  $\beta_4 = 2\alpha_2\alpha_3$ ,  $\beta_5 = \alpha_3^2$ . Finally, using the non-centered moments of a Gamma-distributed random variable,  $\mathbb{E}[X^r] = \frac{\Gamma(a+r)}{\Gamma(a)c^r}$ , we get:

$$\begin{aligned} \beta_1 \int x^4 \Gamma^2(a, c) dx &= \beta_1 \int x^4 \frac{c^{2a}}{\Gamma^2(a)} e^{-2cx} x^{2a-2} \\ &= \beta_1 2^{-(2a+3)} c^{-3} \frac{\Gamma(2a+3)}{\Gamma^2(a)} \int \frac{(2c)^{2a+3}}{\Gamma(2a+3)} e^{-(2c)x} x^{(2a+3)-1} dx \\ &= \beta_1 2^{-(2a+3)} c^{-3} \frac{\Gamma(2a+3)}{\Gamma^2(a)} \int \Gamma(2a+3, 2c) dx \\ &= \beta_1 2^{-(2a+3)} c^{-3} \frac{\Gamma(2a+3)}{\Gamma^2(a)} \end{aligned}$$

and

$$\begin{aligned} \beta_2 \int x^3 \Gamma^2(a, c) dx &= \beta_2 2^{-(2a+2)} c^{-2} \frac{\Gamma(2a+2)}{\Gamma^2(a)} \int \Gamma(2a+2, 2c) dx = \beta_2 2^{-(2a+2)} c^{-2} \frac{\Gamma(2a+2)}{\Gamma^2(a)}. \\ \beta_3 \int x^2 \Gamma^2(a, c) dx &= \beta_3 2^{-(2a+1)} c^{-1} \frac{\Gamma(2a+1)}{\Gamma^2(a)} \int \Gamma(2a+1, 2c) dx = \beta_3 2^{-(2a+1)} c^{-1} \frac{\Gamma(2a+1)}{\Gamma^2(a)}. \\ \beta_4 \int x \Gamma^2(a, c) dx &= \beta_4 2^{-(2a)} \frac{\Gamma(2a)}{\Gamma^2(a)}. \\ \beta_5 \int \Gamma^2(a, c) dx &= \beta_5 2^{-2a+1} c \frac{\Gamma(2a-1)}{\Gamma^2(a)}. \end{aligned}$$

and the expression of the contrast function (8) is obtained. It is worth noting that the function  $u_{l_\theta}^*(y)$  must be approximated numerically by using standard quadrature methods.

Checking assumptions for the Heston model.

**Mixing:** Under the Feller's condition, the volatility process  $X_t$  is  $\rho$ -mixing and so  $\alpha$ -mixing by using the strong Markov property.

**Regularity conditions:** For the Heston model, the function  $l_\theta$  is given by the following polynomial function  $(\alpha_1 x^2 + \alpha_2 x + \alpha_3) f_\theta(x)$  with  $\alpha_1 = (1 - \kappa)^2$ ,  $\alpha_2 = 2(1 - \kappa)\kappa\theta + \sigma^2$ ,  $\alpha_3 = (\kappa\theta)^2$  and is regular w.r.t  $\theta \in \Theta$ . Hence, it remains to prove the moment condition and the local dominance to apply Theorem 2.

Since the function  $l_\theta$  is polynomial w.r.t  $\theta$  belonging to the compact subset  $\Theta$ , all the derivatives exist and in particular  $\sup_{\theta \in \Theta} l_\theta$  and  $\sup_{\theta \in \Theta} \nabla_\theta^2 l_\theta$  are finite. Furthermore, by combining the compactness argument and as the Fourier transform  $f_\epsilon^*$  satisfies (see [Fan et al., 1990]):

$$|f_\varepsilon^*(x)| = \sqrt{2} \exp\left(-\frac{\pi}{2}|x|\right) \left(1 + O\left(\frac{1}{|x|}\right)\right), \quad |x| \rightarrow \infty,$$

which means that  $f_\varepsilon$  is ordinary-smooth in its terminology, we obtain:

$$\begin{cases} \mathbb{E} \left( \sup_{\theta \in \Theta} \|\varphi(Y_2) u_{l_\theta}^*(Y_1)\| \right) < \infty \\ \mathbb{E} \left( |\varphi(Y_2) u_{\nabla_{\theta} l_\theta}^*(Y_1)|^{2+\delta} \right) < \infty & \text{for some } \delta > 0, \\ \mathbb{E} \left( \sup_{\theta \in \mathcal{U}} \|\varphi(Y_2) u_{\nabla_{\theta}^2 l_\theta}^*(Y_1)\| \right) < \infty & \text{for some neighbourhood } \mathcal{U} \text{ of } \theta_0. \end{cases}$$

for  $\varphi : x \mapsto x^2 - s_\varepsilon^2$ .

CREST (ENSAI),

EDHEC BUSINESS SCHOOL