Bernoulli **6**(6), 2000, 1051–1079

Stochastic volatility models as hidden Markov models and statistical applications

VALENTINE GENON-CATALOT[1*], THIERRY JEANTHEAU[1**] and CATHERINE LARÉDO[2]

[1]*Université de Marne-la-Vallée, Equipe d'Analyse et de Mathématiques Appliquées, Cité Descartes, 5 boulevard Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France. E-mail: *genon@univ-mlv.fr; **Thierry.Jeantheau@univ-mlv.fr*
[2]*Institut National de la Recherche Agronomique, Laboratoire de Biométrie, 78350 Jouy-en-Josas, France. E-mail: Catherine.Laredo@jouy.inra.fr*

This paper deals with the fixed sampling interval case for stochastic volatility models. We consider a two-dimensional diffusion process (Y_t, V_t), where only (Y_t) is observed at n discrete times with regular sampling interval Δ. The unobserved coordinate (V_t) is ergodic and rules the diffusion coefficient (volatility) of (Y_t). We study the ergodicity and mixing properties of the observations $(Y_{i\Delta})$. For this purpose, we first present a thorough review of these properties for stationary diffusions. We then prove that our observations can be viewed as a hidden Markov model and inherit the mixing properties of (V_t). When the stochastic differential equation of (V_t) depends on unknown parameters, we derive moment-type estimators of all the parameters, and show almost sure convergence and a central limit theorem at rate $n^{1/2}$. Examples of models coming from finance are fully treated. We focus on the asymptotic variances of the estimators and establish some links with the small sampling interval case studied in previous papers.

Keywords: diffusion processes; discrete-time observations; hidden Markov models; mixing; parametric inference; stochastic volatility

1. Introduction

Continuous-time stochastic volatility (SV) models have recently been the object of growing interest because of their applications in econometry and finance. Moreover, they lead to a series of new statistical issues in the field of inference for stochastic processes. To our knowledge, the first authors to introduce continuous SV models were Hull and White (1987) (for a review of the literature, see Ghysels *et al.* 1996). Continuous-time SV models mainly concern asset-price modelling. The standard model for the evolution of asset prices is the following. Let (Y_t) denote the logarithm of the price process; then it is assumed to be ruled by

$$\mathrm{d}Y_t = \mu(\sigma_t^2)\,\mathrm{d}t + \sigma_t\,\mathrm{d}B_t,$$

where (σ_t^2) is the unobserved instantaneous volatility, (B_t) a Brownian motion and μ is some real function. Among models for (σ_t^2), there are the diffusion processes driven by another

Brownian motion (W_t) – for instance, a GARCH diffusion (Nelson 1990) or a Cox–Ingersoll–Ross process (Heston 1993). Other models including jumps for the volatility have been considered (see, for example, Drost and Werker 1996; Barndorff-Nielsen and Shephard 1998).

In this paper, we study the case where $V_t = \sigma_t^2$ is a diffusion process, and we consider the simplified two-dimensional diffusion process (Y_t, V_t) given by

$$\mathrm{d}Y_t = \sigma_t \, \mathrm{d}B_t, \qquad Y_0 = 0$$

and $V_t = \sigma_t^2$, with

$$\mathrm{d}V_t = b(V_t)\,\mathrm{d}t + a(V_t)\,\mathrm{d}W_t, \qquad V_0 = \eta.$$

Here, $(B_t, W_t)_{t\geqslant 0}$ is a two-dimensional standard Brownian motion, (V_t) is a positive diffusion and η is a positive random variable independent of $(B_t, W_t)_{t\geqslant 0}$. The diffusion (V_t) is unobserved, and the sample path (Y_t) is discretely observed at regularly spaced times $t_i = i\Delta$, $i = 1, \ldots, n$.

For the above model, we have investigated in two previous papers the statistical problem of estimating unknown parameters in the drift and diffusion coefficients of (V_t) (see Genon-Catalot *et al.* 1998; 1999). We have assumed that, while the number of observations n tends to infinity, the sampling interval $\Delta = \Delta_n$ tends to zero and the length of the observation time $n\Delta_n$ tends to infinity. The main assumption on the hidden diffusion (V_t) is its ergodicity. In this framework, we have proved limit theorems for the empirical distribution of the increments $(Y_{t_i} - Y_{t_{i-1}}, i = 1, \ldots, n)$. Furthermore, we have proposed explicit contrast functions to replace the intractable likelihood. This has led to estimators of the unknown parameters present in the stationary distribution of (V_t), which are consistent and asymptotically Gaussian with rate $(n\Delta_n)^{1/2}$.

Complementing this approach is the case – classical in the statistics of diffusion processes – where the sampling interval Δ is fixed (see, for example, Bibby and Sørensen 1995; Kessler 2000). The latter approach, which is the subject of the present paper, will enable us to gain a new insight into the intrinsic properties of SV models. Indeed, we prove here that discretely observed SV models can be viewed as hidden Markov models (HMMs). For a formal definition, we refer to Leroux (1992) or Bickel and Ritov (1996). However, most statistical references in this field assume that the state space of the hidden chain is finite. The SV models provide a concrete example of HMMs with continuous state space for the hidden chain. This more difficult situation has recently been taken into account (see e.g. Jensen and Petersen 1998). In our model we have to deal with a further source of difficulty: the transition probability of the hidden chain that we exhibit is not explicitly known.

The paper is organized as follows. We have to link discrete- and continuous-time processes, so Section 2 revises the ergodicity and mixing properties of strictly stationary Markov processes in both cases. This review is mainly based on Bhattacharya (1982) and Hansen *et al.* (1998). Special attention is given to the study of the ρ-mixing coefficient of an ergodic diffusion. In particular, we give a new proof of a necessary and sufficient condition for ρ-mixing introduced by Hansen *et al.* (1998). This condition can easily be checked on the drift and diffusion coefficients. Moreover, our proof provides explicit upper and lower bounds for the spectral gap of the infinitesimal generator, which in turn gives bounds for the ρ-mixing coefficient.

In Section 3, we define the HMM and prove the key properties for statistical purposes: the observed process inherits the ergodicity and mixing properties of the hidden chain. For a finite state-space hidden chain, such results have been established by Lindgren (1978) and Leroux (1992). Coming back to the SV model, we set, for $i \geqslant 1$,

$$Z_i = \frac{1}{\sqrt{\Delta}} \int_{(i-1)\Delta}^{i\Delta} \sigma_s \, \mathrm{d}B_s,$$

$$U_i = (\overline{V}_i, V_{i\Delta}), \qquad \text{with } \overline{V}_i = \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} V_s \, \mathrm{d}s.$$

We prove that (Z_i) is an HMM with hidden chain (U_i). Applying the results of Section 2 concerning the various mixing coefficients of an ergodic diffusion, we study the α-mixing property of (Z_i). This is achieved using only conditions on the drift and diffusion coefficients of the (V_t) model.

Consider now the problem of estimating unknown parameters in the volatility model. We observe that, conditionally on $(V_t, t \geqslant 0)$, the variables (Z_i) are independent with distribution $\mathcal{N}(0, \overline{V}_i)$. However, neither the joint distribution of $(\overline{V}_i)$ nor the transition probability of (U_i) are explicitly known. Consequently, available results on HMMs cannot be directly applied. Therefore, we study here empirical estimators and prove limit theorems, especially for polynomial functions of $(Z_i, \ldots, Z_{i+d})$. This leads to consistent and asymptotically Gaussian estimators with rate $n^{1/2}$, and all the unknown parameters of the (V_t) model can be estimated. The computation of asymptotic variances enlights the links between fixed and small sampling interval. Section 4 is devoted to examples of widely used parametric models for the volatility.

2. Properties of strictly stationary Markov processes

In view of applications to diffusion processes, we summarize some of the important properties of strictly stationary Markov processes, drawing on several references, among them Bhattacharya (1982), Ethier and Kurtz (1986), Bradley (1986), Doukhan (1994), Hansen and Scheinkman (1995), Hansen *et al.* (1998).

2.1. Continuous time and continuous sample paths

2.1.1. Semigroup and infinitesimal generator

Let us define on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathbf{P})$, $(X_t, t \geqslant 0)$, a time-homogeneous Markov process with state space $(S, \mathscr{B}(S))$, where, for the sake of simplicity, $S = \mathbb{R}^k$, with $k \geqslant 1$. We assume that (X_t) has continuous sample paths, with transition probability $P_t(x, \mathrm{d}y)$, admits a stationary distribution π and that X_0 has distribution π.

Let $\mathscr{F}_t = \sigma(X_s, s \leqslant t)$, $\mathscr{F} = \sigma(X_t, t \geqslant 0)$. Now, for any real π-integrable function f on S, we define

$$P_t f(x) = \int_S f(y) P_t(x, \mathrm{d}y) = \mathrm{E}[f(X_t)/X_0 = x], \qquad \pi \text{ a.s.}, \tag{1}$$

which satisfies, for all $t, s \geqslant 0$, $P_{t+s} f = P_t P_s f$.

Throughout this paper, we will use the following notation:

$$L^2_\pi = L^2(S, \mathscr{B}(S), \pi), \qquad \langle f, g \rangle = \int_S fg \, \mathrm{d}\pi, \qquad \|f\| = \langle f, f \rangle^{1/2}. \tag{2}$$

Whenever $f \in L^2_\pi$, $P_t f \in L^2_\pi$ and $\|P_t f\| \leqslant \|f\|$. Hence, $(P_t, t \geqslant 0)$ is a contraction semigroup on L^2_π.

Property 1. *The semigroup $(P_t, t \geqslant 0)$ is strongly continuous, that is, for all $f \in L^2_\pi$,*

$$\lim_{t \downarrow 0} \|P_t f - f\| = 0.$$

Proof. From the continuity of sample paths, this is clearly satisfied for bounded and continuous functions on S. Now, since this set of functions is dense in L^2_π, the proof is achieved since P_t is a contraction. □

We may now introduce the infinitesimal generator of the semigroup P_t. On the domain

$$\mathscr{D} = \left\{ f \in L^2_\pi : \left\| \frac{P_t f - f}{t} - g \right\| \to 0, \text{ as } t \downarrow 0 \text{ for some } g \in L^2_\pi \right\},$$

the infinitesimal generator $\mathscr{A}$ is defined by

$$f \in \mathscr{D} \to \mathscr{A} f = \lim_{t \downarrow 0} \frac{P_t f - f}{t} \text{ in } L^2_\pi. \tag{3}$$

The following result is a straightforward consequence of the contraction property of P_t.

Property 2. *For any $f \in \mathscr{D}$, $\langle \mathscr{A} f, f \rangle \leqslant 0$.*

The next two properties follow from Property 1 (see, for example, Ethier and Kurtz 1986, pp. 9–10).

Property 3. *For any $f \in \mathscr{D}$ and $t \geqslant 0$, $P_t f \in \mathscr{D}$ and $\mathrm{d}P_t f / \mathrm{d}t = P_t \mathscr{A} f = \mathscr{A} P_t f$.*

Property 4. *$\mathscr{D}$ is (L^2_π)-dense in L^2_π and $\mathscr{A}$ is a closed operator (i.e. the graph of $\mathscr{A}$, $\mathscr{G}(\mathscr{A}) = \{(f, \mathscr{A} f); f \in \mathscr{D}\}$ is a closed subset of $L^2_\pi \times L^2_\pi$).*

2.1.2. Ergodicity

We remark that, for all $t \geqslant 0$, $P_t 1 = 1$, hence $1 \in \mathscr{D}$, and $\mathscr{A} 1 = 0$. So the constant functions are eigenfunctions of P_t with eigenvalue 1, and are also eigenfunctions of $\mathscr{A}$ with eigenvalue

0. The ergodicity of (X_t) is linked with the dimension of the eigenspace of $\mathscr{A}$ associated with the eigenvalue 0.

Denote by $C(\mathbb{R}^+, S)$ the space of continuous functions on $\mathbb{R}^+$ with values in S, endowed with the topology of uniform convergence on compact subsets of $\mathbb{R}^+$. The shift operator θ_t is defined by $\theta_t(x)(\cdot) = x(t + \cdot)$, $x \in C(\mathbb{R}^+, S)$ and $t \geqslant 0$. Then the mapping $(x, t) \to \theta_t(x)$ is jointly continuous on $C(\mathbb{R}^+, S) \times \mathbb{R}^+$. Let $\mathscr{C}$ be the Borel σ-field of $C(\mathbb{R}^+, S)$ and $X = (X_t, t \geqslant 0)$. From strict stationarity, we have, for any $B \in \mathscr{C}$, $\mathbf{P}(X \in B) = \mathbf{P}(\theta_t(X) \in B)$, that is, (θ_t) is measure-preserving. The shift-invariant σ-field is the sub-σ-field of $\mathscr{F}$ given by

$$\mathscr{J} = \{(X \in B);\ B \in \mathscr{C}, \forall t,\ B = \theta_t^{-1}(B)\}.$$

A strictly stationary process is said to be ergodic if $\mathscr{J}$ is $\mathbf{P}$-trivial (i.e. $\forall A \in \mathscr{J}$, $\mathbf{P}(A) = 0$ or 1). In this case, Birkhoff's ergodic theorem (see, for example, Krengel 1985, pp. 9–10) implies that, when $f \in L^1_\pi$,

$$\frac{1}{T}\int_0^T f(X_s)\,\mathrm{d}s \overset{\text{a.s.}}{\to} \mathrm{E}[f(X_0)/\mathscr{J}] = \int f\,\mathrm{d}\pi, \qquad \text{as } T \to \infty.$$

Let us now give some characterizations of ergodicity for Markov process.

Proposition 2.1. *For a strictly stationary Markov process, $\mathscr{J} \subset \sigma(X_0) = \mathscr{F}_0$ (up to null probability sets).*

Proof. Let $A = (X \in B) \in \mathscr{J}$. Then $\forall t$, $B = \theta_t^{-1}(B)$. Set $h(x) = \mathbf{P}(A/X_0 = x)$. We have

$$\mathbf{P}(A/\mathscr{F}_t) = \mathbf{P}(\theta_t(X) \in B/\mathscr{F}_t) = \mathbf{P}(\theta_t(X) \in B/X_t) = h(X_t).$$

From the convergence theorem for martingales, $\mathbf{P}(A/\mathscr{F}_t)$ converges a.s. to $\mathbf{P}(A/\mathscr{F}) = \mathbb{1}_A$ as $t \to \infty$. From strict stationarity, $(\theta_t(X), X_t)$ and (X, X_0) have the same distribution. So, for all $t \geqslant 0$,

$$h(X_t) = \mathbf{P}(\theta_t(X) \in B/X_t) = \mathbf{P}(X \in B/X_0) = h(X_0).$$

We deduce that $\mathbb{1}_A = h(X_0)$ a.s. □

Theorem 2.1. *For a strictly stationary Markov process, the following two statements are equivalent:*

(*i*) *The process (X_t) is ergodic.*
(*ii*) *0 is a simple eigenvalue of $\mathscr{A}$ (this means that the null space $N_{\mathscr{A}} = \{f \in \mathscr{D} : \mathscr{A}f = 0\}$ is the one-dimensional subspace of L^2_π spanned by constants).*

Proof. Assume first that 0 is a simple eigenvalue of $\mathscr{A}$ and consider a square-integrable and $\mathscr{J}$-measurable random variable Z. The previous proof shows that $h(x) = \mathrm{E}(Z/X_0 = x)$ satisfies $h(X_t) = \mathrm{E}(Z/\mathscr{F}_t) = h(X_0) = Z$. Thus, $h(X_t)$ is a square-integrable martingale. We can therefore write, for all $t, s \geqslant 0$,

$$P_t h(X_s) = \mathrm{E}(h(X_{t+s})/\mathscr{F}_s) = h(X_s).$$

Hence $P_t h = h$ (in L^2_π), for all $t \geqslant 0$, and this implies that $h \in \mathscr{D}$ and $\mathscr{A}h = 0$. Using our assumption, h is a constant, so $Z = h(X_0)$ is also a constant. Therefore, $\mathscr{J}$ is **P**-trivial.

We now prove the converse. Let $h \in \mathscr{D}$ and $\mathscr{A}h = 0$. From Property 3, we get $\mathrm{d}P_t h/\mathrm{d}t = P_t \mathscr{A}h = 0$, which implies $P_t h = P_0 h = h$. So, from the Markov property, $h(X_t)$ is a square-integrable martingale, and there exists a square-integrable and $\mathscr{F}$-measurable random variable $Z = \lim h(X_t)$ a.s. (as $t \to \infty$). Moreover, $h(X_t) = \mathrm{E}(Z/\mathscr{F}_t)$. Since, for all $s \geqslant 0$, $Z = \lim h(X_{t+s})$, Z is $\mathscr{J}$-measurable. Since $\mathscr{J}$ is assumed to be **P**-trivial, Z is a.s. constant. Finally, $h(X_t) = \mathrm{E}(Z/\mathscr{F}_t)$ implies that h is a constant in L^2_π. □

Further results are proved in Bhattacharya (1982, Proposition 2.3 and note added in proof, p. 201; Theorem 2.1).

Proposition 2.2. *For a strictly stationary Markov process, the following statements are equivalent:*

(*i*) *0 is a simple eigenvalue of $\mathscr{A}$ (i.e. (X_t) is ergodic).*
(*ii*) *The range of $\mathscr{A}$, $\mathscr{R}_\mathscr{A} = \{\mathscr{A}g;\ g \in \mathscr{D}\}$, is dense in $\{1\}^\perp$.*

Moreover, if this is the case, then $\mathscr{R}_\mathscr{A} = \{1\}^\perp$ if and only if 0 is an isolated point of the spectrum of $\mathscr{A}$.

Theorem 2.2. *Let (X_t) be a strictly stationary and ergodic Markov process and let $f \in L^2_\pi$. Then, if $f \in \mathscr{R}_\mathscr{A}$,*

$$\frac{1}{T^{1/2}} \int_0^T f(X_s)\,\mathrm{d}s \xrightarrow{\mathscr{D}} \mathscr{N}(0, \sigma^2) \qquad \text{as } T \to \infty,$$

where $\sigma^2 = -2\langle f, g\rangle$ and g is any element of $\mathscr{D}$ satisfying $\mathscr{A}g = f$. Moreover,

$$\mathrm{var}\left(\frac{1}{T^{1/2}} \int_0^T f(X_s)\,\mathrm{d}s\right) \to \sigma^2 \qquad \text{as } T \to \infty.$$

Therefore, if 0 is a simple eigenvalue and an isolated point of the spectrum of $\mathscr{A}$, then the above central limit theorem holds for any $f \in L^2_\pi$ such that $\int f\,\mathrm{d}\pi = 0$.

2.2. Discrete time and ergodicity

Now, let $(X_n,\ n \in \mathbb{N})$ be a time-homogeneous Markov process on $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$, with state space $(S, \mathscr{B}(S))$ as above, with one-step transition probability $P(x, \mathrm{d}y)$, admitting a stationary distribution π and such that X_0 has distribution π.

Let $\mathscr{F}_n = \sigma(X_k,\ k \leqslant n)$, $\mathscr{F} = \sigma(X_n,\ n \geqslant 0)$. Now, for any real π-integrable function f on S, we define

$$Pf(x) = \int_S f(y)P(x, \mathrm{d}y) = \mathrm{E}[f(X_1)/X_0 = x], \qquad \pi \text{ a.s.}$$

When $f \in L^2_\pi$, $Pf \in L^2_\pi$, $\|Pf\| \leqslant \|f\|$ and $P1 = 1$ (following the notation of Section 2.1). We study P as a linear operator on L^2_π. The ergodicity of (X_n) may now be linked with the dimension of the eigenspace of P associated with the eigenvalue 1.

Let us define the shift operator $\theta: S^{\mathbb{N}} \to S^{\mathbb{N}}$ by $\theta(x) = (x_1, x_2, \ldots)$, for $x = (x_0, x_1, \ldots)$. The mapping θ is measurable with respect to $\mathscr{B}(S^{\mathbb{N}})$, and setting $X = (X_n, n \geqslant 0)$, the σ-field of shift-invariant sets is the sub-σ-field of $\mathscr{F}$ given by

$$\mathscr{J}_\theta = \{(X \in B); B \in \mathscr{B}(S^{\mathbb{N}}), \theta^{-1}(B) = B\}.$$

The process is said to be ergodic if $\mathscr{J}_\theta$ is $\mathbf{P}$-trivial.

Proposition 2.3. *If* (X_n) *is a strictly stationary Markov,* $\mathscr{J}_\theta \subset \sigma(X_0) = \mathscr{F}_0$ *(up to null probability sets).*

The proof is identical to the proof of Proposition 2.1. We just have to note that, if $\theta_n = \theta \circ \ldots \circ \theta$ is the nth iteration of θ, then for any $B \in \mathscr{B}(S^{\mathbb{N}})$, $B = \theta^{-1}(B)$ implies $B = \theta_n^{-1}(B)$ for all n.

Theorem 2.3. *If* (X_n) *is a strictly stationary Markov, the following two statements are equivalent:*

(*i*) *The process* (X_n) *is ergodic.*
(*ii*) *The value 1 is a simple eigenvalue of* P (*this means that the space* $\{h \in L^2_\pi : Ph = h\}$ *is the one-dimensional subspace of* L^2_π *spanned by constants*).

The proof is similar to the proof of Theorem 2.1. The key tool is even simpler. For $h \in L^2_\pi$, $Ph = h$ holds if and only if $h(X_n)$ is a martingale, which is necessarily such that

$$h(X_n) = h(X_0) = \mathrm{E}(Z/\mathscr{F}_n) = \mathrm{E}(Z/X_0) = Z,$$

for some $\mathscr{J}_\theta$-measurable square-integrable random variable Z.

2.3. Mixing coefficients

In the following, we shall use the so-called α-, β- and ρ-mixing coefficients. We recall their definitions and specific properties for strictly stationary Markov processes (see, for example, Bradley 1986; and Doukhan 1994). Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and $\mathscr{A}$ and $\mathscr{B}$ two σ-fields included in $\mathscr{F}$. The following measures of dependence are classical:

$$\alpha(\mathscr{A}, \mathscr{B}) = \sup_{A \in \mathscr{A}, B \in \mathscr{B}} |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)|, \tag{4}$$

$$\beta(\mathscr{A}, \mathscr{B}) = \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbf{P}(A_i \cap B_j) - \mathbf{P}(A_i)\mathbf{P}(B_j)|, \tag{5}$$

where the latter supremum is taken over all pairs of partitions $\{A_1, \ldots, A_I\}$ and $\{B_1, \ldots, B_J\}$ of Ω such that $A_i \in \mathscr{A}$ for all i and $B_j \in \mathscr{B}$ for all j, and

$$\rho(\mathscr{A}, \mathscr{B}) = \sup\{|\mathrm{corr}(X, Y)|;\ X \in L^2(\mathscr{A}),\ Y \in L^2(\mathscr{B}),\ X, Y \text{ real}\}. \tag{6}$$

The following inequalities hold:

$$2\alpha(\mathscr{A}, \mathscr{B}) \leqslant \beta(\mathscr{A}, \mathscr{B}) \leqslant 1,$$
$$4\alpha(\mathscr{A}, \mathscr{B}) \leqslant \rho(\mathscr{A}, \mathscr{B}) \leqslant 1. \tag{7}$$

The above formulae can be rewritten as:

$$\alpha(\mathscr{A}, \mathscr{B}) = \sup\{|\mathrm{cov}(U, V)|;\ 0 \leqslant U, V \leqslant 1,\ U \text{ is } \mathscr{A}\text{-measurable},\ V \text{ is } \mathscr{B}\text{-measurable}\}, \tag{8}$$

$$\beta(\mathscr{A}, \mathscr{B}) = \mathrm{E}(\mathrm{ess.sup}\{|\mathbf{P}(B/\mathscr{A}) - \mathbf{P}(B)|,\ B \in \mathscr{B}\}), \tag{9}$$

$$\rho(\mathscr{A}, \mathscr{B}) = \sup\left\{\frac{\|\mathrm{E}(X/\mathscr{B}) - \mathrm{E}(X)\|_2}{\|X\|_2};\ X \in L^2(\mathscr{A}),\ X \text{ real}\right\}. \tag{10}$$

Now, for any stochastic process (X_t), with $t \in \mathbb{R}^+$ or $t \in \mathbb{N}$, let $\mathscr{F}_t = \sigma(X_s, s \leqslant t)$, $\mathscr{F}^t = \sigma(X_u, u \geqslant t)$; then $\alpha_X(t)$, $\beta_X(t)$ and $\rho_X(t)$ are defined by

$$c_X(t) = \sup_{s \geqslant 0} c(\mathscr{F}_s, \mathscr{F}^{s+t}), \tag{11}$$

with $c = \alpha$, β or ρ. The process is said to be c-mixing if $c_X(t) \to 0$ when $t \to \infty$. In particular, when (X_t) is a strictly stationary process, it is well known that α-mixing implies ergodicity. For Markov processes, mixing coefficients have simple expressions (see Bradley 1986, Theorem 4.1).

Proposition 2.4. *Assume that (X_t) is a strictly stationary Markov process. Then, for $c = \alpha, \beta, \rho$,*

$$c_X(t) = c(\sigma(X_0), \sigma(X_t)).$$

Using the above proposition and (9)–(10), explicit expressions for $\beta_X(t)$ and $\rho_X(t)$ are easily deduced (see Doukhan 1994, Section 2.4).

Theorem 2.4. *Assume that (X_t) is a strictly stationary Markov process, with initial distribution π. Then, in the continuous-time case, with the notation of Section 2.1,*

$$\beta_X(t) = \int_S \pi(\mathrm{d}x)\|P_t(x, \mathrm{d}y) - \pi(\mathrm{d}y)\|_{\mathrm{TV}},$$

where $\|\nu\|_{\mathrm{TV}}$ denotes the total variation norm of a signed measure ν, and

$$\rho_X(t) = \sup\left\{\frac{\|P_t(f)\|}{\|f\|};\ f \in L^2_\pi,\ \langle f, 1\rangle = 0\right\}.$$

In the discrete-time case, with the notation of Section 2.2,

$$\beta_X(n) = \int_S \pi(\mathrm{d}x)\|P^n(x, \mathrm{d}y) - \pi(\mathrm{d}y)\|_{\mathrm{TV}}$$

and

$$\rho_X(n) = \sup\left\{\frac{\|P^n(f)\|}{\|f\|};\ f \in L^2_\pi,\ \langle f, 1\rangle = 0\right\},$$

where P^n is the nth iteration of the operator P, and $P^n(x, \mathrm{d}y)$ is the n-step transition probability.

The proofs of central limit theorems rely on the rate of convergence to 0 of the mixing coefficients as t goes to infinity. We say that $c(t)$ tends to 0 'exponentially fast' if there is some $\varepsilon > 0$ such that $c(t) = O(\mathrm{e}^{-\varepsilon t})$. The ρ-mixing coefficient has the following special property (see Bradley 1986, Theorem 4.2):

Proposition 2.5 *Let (X_t) be a strictly stationary Markov process. If $\rho_X(t) \to 0$ as $t \to +\infty$, then $\rho_X(t) \to 0$ 'exponentially fast'.*

It is worth noting that, for α- or β-mixing coefficients, the convergence to 0 may occur, for instance, at a polynomial rate.

2.4. Continuous time and discrete sampling

Consider a continuous-time process $(X_t,\ t \geqslant 0)$, as in Section 2.1, and, for any $\Delta > 0$, the discretely sampled process $(X_{k\Delta},\ k \geqslant 0)$. Hence, the one-step transition probability of $(X_{k\Delta})$ is $P_\Delta(x, \mathrm{d}y)$ and $P^n_\Delta = P_{n\Delta}$. Assume that (X_t) is ergodic. The following question arises: is $(X_{k\Delta})$ ergodic for any Δ? An answer is given below.

Theorem 2.5. *Assume that (X_t) is a strictly stationary Markov process, with transition probability $P_t(x, \mathrm{d}y)$ and marginal distribution π. If $\|P_t(x, \mathrm{d}y) - \pi(\mathrm{d}y)\|_{\mathrm{TV}}\} \to 0$ as $t \to +\infty$ for all $x \in S$, then (X_t) and, for all $\Delta > 0$, $(X_{k\Delta})$ are β-mixing, hence α-mixing and ergodic.*

The proof is just an immediate consequence of the Lebesgue dominated convergence theorem, and of Theorem 2.4. In Orey (1971, Proposition 4.3), a more precise result is proved: the convergence to 0 of $\|P_t(x, \mathrm{d}y) - \pi(\mathrm{d}y)\|_{\mathrm{TV}}$ is equivalent to the property that the tail σ-field $\bigcap_{t\geqslant 0}\sigma(X_s,\ s \geqslant t)$ is trivial for every initial distribution. The ergodicity is obtained since the tail σ-field contains the shift invariant σ-field (see also Bhattacharya 1982, Proposition 2.5).

2.5. Reversible Markov process

A strictly stationary continuous-time Markov process (X_t) is said to be reversible if the joint distributions of (X_0, X_t) and (X_t, X_0) are identical. As a consequence, for all $f, g \in L^2_\pi$, $\langle P_t f, g\rangle = \langle f, P_t g\rangle$, and so the infinitesimal generator $\mathscr{A}$ is also self-adjoint. Some useful properties may be deduced. Since P_t is self-adjoint for all t, we have, for all $f \in L^2_\pi$,

$$\langle P_t f, f\rangle = \mathrm{E}(f(X_0)f(X_t)) = \langle P_{t/2}f, P_{t/2}f\rangle \geqslant 0. \tag{12}$$

This implies the following special feature of the covariance structure:

$$\forall t \geqslant 0, \qquad \mathrm{cov}(f(X_0), f(X_t)) \geqslant 0. \tag{13}$$

Moreover, as noted by Hansen and Scheinkman (1995, pp. 786 and 794), the following result holds:

Proposition 2.6. *Assume that $(X_t, t \geqslant 0)$ is a strictly stationary Markov process, ergodic and reversible. Then, any discretely sampled process $(X_{k\Delta})$ is ergodic.*

Furthermore, the reversibility allows an exact computation of the ρ-mixing coefficient in the ergodic case.

Theorem 2.6. *Assume that (X_t) is a strictly stationary ergodic and reversible Markov process. Then $\rho_X(t) = \mathrm{e}^{-\lambda t}$ with*

$$\lambda = \inf\left\{\frac{\langle f, -\mathscr{A}f\rangle}{\langle f, f\rangle}; f \in \mathscr{D}, \langle f, 1\rangle = 0\right\} \geqslant 0. \tag{14}$$

Therefore, only two cases may occur:

- either $\lambda > 0$, in which case $\rho_X(t)$, and consequently $\alpha_X(t)$ will tend to 0 exponentially fast;
- or $\lambda = 0$, in which case $\rho_X(t) = 1$, for all $t \geqslant 0$ – however, $\alpha_X(t)$ or $\beta_X(t)$ may tend to 0.

Proof of Theorem 2.6. The following direct proof was provided to us by E.M. Ouhabaz. Let $E_0 = \{f \in \mathscr{D}; \mathscr{A}f = 0\}$ be the null space of $\mathscr{A}$ which is, by the ergodicity assumption, the one-dimensional space spanned by constants. Thus $L^2_\pi = E_0^\perp \oplus E_0$, with $E_0^\perp = \{f \in L^2_\pi; \langle f, 1\rangle = 0\}$. Noting that $P_t E_0^\perp \subset E_0^\perp$, $T_t = P_t/E_0^\perp$ defines another strongly continuous contraction semigroup on $E_0^\perp$ with self-adjoint infinitesimal generator $\mathscr{B}$. Its spectrum $\sigma(\mathscr{B})$ is included in $(-\infty, 0]$. By Theorem 2.4,

$$\rho_X(t) = \sup\left\{\frac{\|T_t f\|}{\|f\|}; f \in E_0^\perp\right\} = \|T_t\|.$$

The spectral mapping theorem holds for T_t in the sense that $\|T_t\| = \mathrm{e}^{-\lambda t}$, with

$$-\lambda = \sup\{\sigma(\mathscr{B})\} = \sup\left\{\frac{\langle f, \mathscr{B}f\rangle}{\langle f, f\rangle}; f \in \mathscr{D}(B)\right\}$$

(see, for example, Nagel 1986, Chapter A-III). The domain of $\mathscr{B}$ is $\mathscr{D}(\mathscr{B}) = \{f \in \mathscr{D} \cap E_0^\perp; \mathscr{A}f \in E_0^\perp\}$. But, since, for $f \in \mathscr{D}$, $\langle \mathscr{A}f, 1\rangle = \langle f, \mathscr{A}1\rangle = 0$, $\mathscr{D}(\mathscr{B}) = E_0^\perp \cap \mathscr{D}$. Moreover, $\mathscr{A}$ and $\mathscr{B}$ coincide on $\mathscr{D}(\mathscr{B})$, so the proof is complete. □

It is worth noting that $\lambda > 0$ is equivalent to the fact that the spectrum of $\mathscr{A}$ is included in $(-\infty, \lambda] \cup \{0\}$. Hence, 0 is an isolated point of the spectrum, and λ is the so-called spectral gap.

2.6. One-dimensional diffusion processes

2.6.1. Notation and assumptions

We consider a Markov process which is defined by a stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + a(X_t)\,\mathrm{d}W_t, \qquad X_0 = \eta, \tag{15}$$

where W is a standard Brownian motion in $\mathbb{R}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and η is a real random variable defined on Ω and independent of W. We now make the standard assumptions on functions $b(x)$ and $a(x)$, ensuring that the solution of (15) is a positive recurrent diffusion on an interval (l, r) $(-\infty \leqslant l < r \leqslant +\infty)$ and a strictly stationary ergodic time-reversible process.

(A1) The functions $b(x)$ and $a(x)$ are defined on (l, r), and satisfy

$$b(x) \in C^1(l, r),\ a^2(x) \in C^2(l, r),\ a(x) > 0 \text{ for all } x \in (l, r),$$

and

$$\exists K > 0,\ \forall x \in (l, r),\ |b(x)| \leqslant K(1 + |x|) \text{ and } a^2(x) \leqslant K(1 + x^2).$$

For $x_0 \in (l, r)$, define the scale and speed densities of diffusion (X_t),

$$s(x) = \exp\left(-2\int_{x_0}^{x} \frac{b(u)}{a^2(u)}\,\mathrm{d}u\right), \qquad m(x) = \frac{1}{a^2(x)s(x)}. \tag{16}$$

(A2) $\displaystyle\int_l s(x)\,\mathrm{d}x = +\infty,\ \int^r s(x)\,\mathrm{d}x = +\infty,\ \int_l^r m(x)\,\mathrm{d}x = M < +\infty.$

Let us define the stationary density

$$\pi(x) = \frac{1}{M} m(x) \mathbb{1}_{\{x \in (l,r)\}}. \tag{17}$$

(A3) The initial random variable η has distribution $\pi(\mathrm{d}x) = \pi(x)\,\mathrm{d}x$.

Now consider the two following additional assumptions.

(A4) As $x \downarrow l$ and $x \uparrow r$, $\lim a(x)m(x) = 0$.

(A5) Set $\gamma(x) = a'(x) - 2b(x)/a(x)$. As $x \downarrow l$ and $x \uparrow r$, the limits of $1/\gamma(x)$ exist.

Assumptions (A1) and (A2) ensure the existence and uniqueness of the solution of (15) together with the positive recurrence on (l, r), and (A3) provides strict stationarity. Assumptions (A4) and (A5) are needed to study the ρ-mixing property. Note that, in view of (A2), (A4) is not a strong assumption.

In the further central limit theorems of the paper, a precise rate of convergence for $\alpha_X(t)$

is required. A direct computation being difficult, we can use the two inequalities (7) since, for diffusion processes, $\beta_X(t)$ and $\rho_X(t)$ are easier to obtain, as we will see below.

2.6.2. The β-mixing coefficient

First, we have the following property.

Proposition 2.7. *Under (A1)–(A3), (X_t) is time reversible, and (X_t) as well as $(X_{k\Delta})$, for all Δ, are ergodic and β-mixing.*

Proof. According to Kent (1978, pp. 830–831), under (A1)–(A3), (X_t) is time-reversible. If $P_t(x, \mathrm{d}y)$ denotes the transition probability, then (see Rogers and Williams 1987, pp. 302–303)

$$\|P_t(x, \mathrm{d}y) - \pi(\mathrm{d}y)\|_{\mathrm{TV}} \to 0 \qquad \text{as } t \to +\infty.$$

Theorem 2.5 gives the result. □

Second, the rate of convergence to 0 of $\beta_X(t)$ has been investigated in several papers. In particular, Veretennikov (1988) gives sufficient conditions for an exponential rate of convergence to 0 and Veretennikov (1997) for a polynomial rate (see also Lindvall 1983).

2.6.3. The ρ-mixing coefficient

Let us denote by $(P_t, t \geqslant 0)$ the transition semigroup of (X_t) on L^2_π and by $\mathscr{A}$ its infinitesimal generator with domain $\mathscr{D}$. In order to investigate the ρ-mixing property of (X_t), we rely on results stated in Hansen *et al.* (1998, Section 3). First, we have to find a subset D of $\mathscr{D}$ on which $\langle \mathscr{A}f, f\rangle$ has a simple expression in terms of the functions b and a. This subset D has to be large enough to obtain

$$\lambda = \inf\left\{\frac{\langle f, -\mathscr{A}f\rangle}{\langle f, f\rangle}; f \in D, \langle f, 1\rangle = 0\right\}. \tag{18}$$

This will be achieved if D is a core for $\mathscr{A}$, that is, if D is dense in $\mathscr{D}$ with respect to the graph norm $\|f\| + \|\mathscr{A}f\|$. For instance, D can be taken as the space $C^2_c(l, r)$ of twice continuously differentiable real functions with compact support included in (l, r) and, for $f \in C^2_c(l, r)$, we have $\mathscr{A}f = Lf$, with (see (A2) and (A3))

$$Lf = \frac{a^2}{2} f'' + bf' = \frac{1}{2m}\left(\frac{f'}{s}\right)'. \tag{19}$$

Now

$$\langle \mathscr{A}f, f\rangle = M\int_l^r \frac{1}{2m}\left(\frac{f'}{s}\right)' fm\,\mathrm{d}x$$

Integrating by parts yields

$$\langle \mathscr{A}f, f\rangle = \frac{M}{2}\left[\frac{f'}{s}f\right]_l^r - \frac{M}{2}\int_l^r \frac{f'}{s}f'\,\mathrm{d}x$$
$$= -\frac{1}{2}\int_l^r (f'a)^2\,\mathrm{d}\pi.$$

Therefore,

$$\lambda = \frac{1}{2}\inf\left\{\frac{\int_l^r (f'a)^2\,\mathrm{d}\pi}{\int_l^r f^2\,\mathrm{d}\pi};\ f\in C_c^2(l, r),\ \int_l^r f\,\mathrm{d}\pi = 0\right\}. \tag{20}$$

Let us now return to Theorem 2.2 to throw some light on the computation of the asymptotic variance for diffusion processes. First, recall the exact description of $(\mathscr{D}, \mathscr{A})$ under (A1)–(A3). We have

$$\mathscr{D} = \{g\in L_\pi^2;\ g' \text{ absolutely continuous, } Lg\in L_\pi^2,\ \lim g'(x)/s(x) = 0 \text{ as } x\downarrow l \text{ and } x\uparrow r\}. \tag{21}$$

On $\mathscr{D}$, $\mathscr{A}g = Lg$ and the following formula holds:

$$-2\langle \mathscr{A}g, g\rangle = \int_l^r (g'a)^2\,\mathrm{d}\pi. \tag{22}$$

Starting with f such that $\langle f, 1\rangle = 0$ and for g solving $Lg = -f$, we obtain, by (19), $g'(x) = -2s(x)\int_l^x f(u)m(u)\,\mathrm{d}u$. Computation yields

$$-2\langle \mathscr{A}g, g\rangle = -2\langle f, g\rangle = 4M\int_l^r s(x)\,\mathrm{d}x\left(\int_l^x f(u)\,\mathrm{d}\pi(u)\right)^2. \tag{23}$$

The property that $f\in\mathscr{R}_{\mathscr{A}}$ is equivalent to the fact that the above integral is finite.

We can now state the following equivalence.

Proposition 2.8. *Under (A1)–(A5), (X_t) is ρ-mixing if and only if the limits in (A5) are finite.*

The proof of the proposition, somewhat technical, is postponed to the Appendix. One or two comments need to be made. Introducing (A4)–(A5), Hansen and Scheinkman (1995) obtained the sufficient condition for ρ-mixing. The equivalence is then proved in Hansen *et al.* (1998) using the spectral theory of ordinary differential equations. In the latter paper another result is obtained. Under (A5), if

$$\tau = \frac{1}{8}\inf\{\gamma^2(l), \gamma^2(r)\},$$

then the discrete spectrum of $\mathscr{A}$ is included in $(-\tau, 0]$. This implies that $\tau = +\infty$ is equivalent to the fact that the spectrum is entirely discrete.

We give another proof based on close study of λ as given in (14). The interest of our proof relies in the obtainment of lower and upper bounds for the spectral gap λ which are new and explicitly computable from the diffusion coefficients $b(.)$ and $a(.)$.

The following corollary is useful.

Corollary 2.1. *Assume* (*A1*)–(*A5*) *and that the limits in* (*A5*) *are finite. Then there exists a positive* λ *such that* $\alpha_X(t) \leqslant e^{-\lambda t}/4$.

Proof. This is a consequence of (7), Theorem 2.6 and Proposition 2.8. □

Turning to examples, let us consider diffusions with mean reverting drift given by

$$dX_t = \alpha(\beta - X_t)\,dt + cX_t^\nu\,dW_t,$$

where $\frac{1}{2} \leqslant \nu \leqslant 1$. The state space is $(l, r) = (0, +\infty)$. Assumption (A1) holds. To check (A2), we must distinguish three cases. Thus assumption (A2) holds:

(i) for $\nu = \frac{1}{2}$, if $\alpha > 0$ and $\alpha\beta \geqslant c^2/2$;
(ii) for $\frac{1}{2} < \nu < 1$, if $\alpha > 0$ and $\beta > 0$;
(iii) for $\nu = 1$, if $\alpha\beta > 0$ and $\alpha > -c^2/2$.

Under these conditions on the parameters, (A4)–(A5) also hold. For the first two models, both limits of $1/\gamma(x)$ are equal to 0. For $\nu = 1$, we have $\lim 1/\gamma(x)$ equal to 0 as $x \to 0$ but being strictly positive (and equal to $c/(2\alpha + c^2)$) as $x \to +\infty$. Under the stationarity condition (A3), we obtain, by Proposition 2.8, that the process (X_t) is ρ-mixing. Therefore, 0 is an isolated point of the infinitesimal generator spectrum and $\mathscr{R}_{\mathscr{A}} = \{1\}^\perp$. The case $\nu = \frac{1}{2}$ is well known since the spectrum is discrete and equal to $\{-n\alpha, n \geqslant 0\}$.

3. Stochastic volatility model as a hidden Markov model

In this section, we show that a stochastic volatility model can be viewed as a hidden Markov model.

3.1. Definition and properties of a hidden Markov model

We follow Leroux (1992) and Bickel and Ritov (1996) for a formal definition.

Definition 3.1. *A stochastic process* $(Z_n, n \geqslant 1)$, *with state space* $(\mathscr{Z}, \mathscr{B}(\mathscr{Z}))$, *is a hidden Markov model if the following hold:*

(*i*) *(Hidden chain.) We are given (but do not observe) a strictly stationary Markov chain* $U_1, U_2, \ldots, U_n, \ldots$ *with state space* $(\mathscr{U}, \mathscr{B}(\mathscr{U}))$.
(*ii*) *For all* n, *given* $(U_1, U_2, \ldots, U_n)$, *the* Z_i, $i = 1, \ldots, n$, *are conditionally independent, and the conditional distribution of* Z_i *depends only on* U_i.
(*iii*) *The conditional distribution of* Z_i *given* $U_i = u$ *does not depend on* i.

This is the definition given by Leroux. $\mathscr{Z}$ and $\mathscr{U}$ are Polish spaces equipped with their Borel σ-field. Note that condition (ii) is replaced in Bickel and Ritov's paper by:

(ii′) For all n, given $(U_1, U_2, \ldots, U_n)$, the Z_i, $i = 1, \ldots, n$, are conditionally independent, and, given U_i, Z_i is independent of $(U_j, j \neq i, j = 1, \ldots, n)$.

It is simple to check that conditions (i)–(ii) are equivalent to (i)–(ii′). It is worth noting that, in general, authors only consider the case where the hidden chain has a finite state-space. In view of our applications, we do not make this assumption. A hidden Markov model has the following property.

Proposition 3.1. *The process $(Z_i, i \geqslant 1)$ is strictly stationary. If the hidden Markov chain $(U_i, i \geqslant 1)$ is ergodic, then $(Z_i, i \geqslant 1)$ is also ergodic. Moreover, if $(U_i, i \geqslant 1)$ is α-mixing, then $(Z_i, i \geqslant 1)$ is also α-mixing, and*

$$\alpha_Z(k) \leqslant \alpha_U(k).$$

Proof. Let $\varphi : \mathscr{Z}^n \to \mathbb{R}$ be a positive measurable function. If we denote by $P(.; u)$ a regular version of the conditional distribution of Z_i given $U_i = u$, we can define, using (ii) and (iii) of Definition 3.1,

$$\begin{aligned} h_\varphi(u_1, \ldots, u_n) &= \mathrm{E}[\varphi(Z_1, \ldots, Z_n)/U_1 = u_1, \ldots, U_n = u_n] \\ &= \int_{\mathscr{Z}^n} \varphi(z_1, \ldots, z_n) P(dz_1; u_1) \otimes \ldots \otimes P(\mathrm{d}z_n; u_n). \end{aligned}$$

Using (i), we have

$$\mathrm{E}\varphi(Z_{k+1}, \ldots, Z_{k+n}) = \mathrm{E}h_\varphi(U_{k+1}, \ldots, U_{k+n}) = \mathrm{E}h_\varphi(U_1, \ldots, U_n),$$

This implies that $(Z_1, \ldots, Z_n)$ and $(Z_{k+1}, \ldots, Z_{k+n})$ have the same distribution. Hence, Z_i is strictly stationary.

The ergodicity is proved in Lemma 1 of Leroux (1992) for a hidden chain with finite space state. Actually, Leroux's proof does not require this assumption, but only relies on the ergodicity of the hidden chain. Thus, it applies here.

Finally, we remark that, conditioning by $(U_i, i \geqslant 1)$, we obtain, for functions $\varphi : \mathscr{Z}^i \to [0, 1]$ and $\psi : \mathscr{Z}^j \to [0, 1]$,

$$\mathrm{cov}(\varphi(Z_1, \ldots, Z_i), \psi(Z_{i+k+1}, \ldots, Z_{i+k+j})) = \mathrm{cov}(h_\varphi(U_1, \ldots, U_i), h_\psi(U_{i+k+1}, \ldots, U_{i+k+j})).$$

Since h_φ is positive and bounded by 1, (8) gives the result. □

3.2 The stochastic volatility model

We now consider the model studied in two previous papers (see Genon-Catalot *et al.* 1998; 1999). Let $(Y_t, V_t)_{t \geqslant 0}$ be a two-dimensional diffusion process defined by

$$\mathrm{d}Y_t = \sigma_t \, \mathrm{d}B_t, \qquad Y_0 = 0, \tag{24}$$

$$V_t = \sigma_t^2, \qquad \mathrm{d}V_t = b(V_t)\,\mathrm{d}t + a(V_t)\,\mathrm{d}W_t, \qquad V_0 = \eta. \tag{25}$$

We assume that

(A0) (B, W) is a standard Brownian motion in $\mathbb{R}^2$ defined on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$, and η is a random variable defined on Ω, independent of (B, W).

We denote by P_t the transition semigroup of V_t on L^2_π and by $\mathscr{A}$ its infinitesimal generator with domain $\mathscr{D}$. Using the notation of Section 2.6, we assume (A1)–(A3) hold, with $(l, r) \subset (0, +\infty)$. Then, the diffusion (V_t) is strictly stationary, ergodic and time-reversible, and all the conclusions of Theorem 2.5 hold.

For Δ positive, we define, for $i \geqslant 1$,

$$Z_i = \frac{1}{\sqrt{\Delta}} \int_{(i-1)\Delta}^{i\Delta} \sigma_s \, \mathrm{d}B_s, \tag{26}$$

and

$$U_i = (\overline{V}_i, V_{i\Delta}), \qquad \overline{V}_i = \frac{1}{\Delta} \int_{(i-1)\Delta}^{i\Delta} V_s \, \mathrm{d}s. \tag{27}$$

We remark that, conditionally on $(V_s, s \geqslant 0)$, the random variables (Z_i) are independent with distribution $\mathscr{N}(0, \overline{V}_i)$. However, $(\overline{V}_i)$ is not Markov.

Theorem 3.1. *Under (A0)–(A3), we have that*

(*i*) $(U_i, i \geqslant 1)$ *is a strictly stationary Markov chain with state-space* $(l, r)^2$;
(*ii*) $(Z_i, i \geqslant 1)$ *is a hidden Markov model with hidden chain* $(U_i, i \geqslant 1)$.

Proof. Let $\mathscr{G}_t = \sigma(V_s, s \in [0, t])$, $E = C([0, \Delta], (l, r))$ the space of continuous functions defined on $[0, \Delta]$ with values in (l, r) and $\mathscr{B}$ the Borel σ-algebra associated with the uniform topology. Let $\varphi : (l, r)^2 \to \mathbb{R}$ be a bounded Borel function. From the Markov property of (V_t), we have

$$\mathrm{E}[\varphi(\overline{V}_i, V_{i\Delta})/\mathscr{G}_{(i-1)\Delta}] = \mathrm{E}[\varphi(\overline{V}_i, V_{(i\Delta)})/V_{(i-1)\Delta}] = \psi(V_{(i-1)\Delta}),$$

where

$$\psi(v) = \mathrm{E}[\varphi(\overline{V}_1, V_\Delta)/V_0 = v].$$

This comes from the fact that

$$\overline{V}_i = \frac{1}{\Delta} \int_0^\Delta V_{(i-1)\Delta+s} \, \mathrm{d}s \quad \text{and} \quad V_{i\Delta} = V_{(i-1)\Delta+\Delta}.$$

Thus, $(U_i, i \geqslant 1)$ is a Markov chain with transition probability given by

$$\mathscr{P}\varphi(\overline{v}, v) = \mathrm{E}[\varphi(\overline{V}_i, V_{i\Delta})/\overline{V}_{i-1} = \overline{v}, V_{(i-1)\Delta} = v] = \psi(v),$$

which only depends on v.

Set, for $i \geqslant 1$, $X_i = (V_{(i-1)\Delta+s}, s \in [0, \Delta])$. The process (X_i) has state-space $(E, \mathscr{B})$. Since the process (V_t) is strictly stationary, the same is true for (X_i). Now, note that $U_i = T(X_i)$ with

$$T(x) = \left(\frac{1}{\Delta} \int_0^\Delta x(s)\,\mathrm{d}s,\, x(\Delta) \right).$$

Since T is continuous on E, (U_i) is strictly stationary.

Turning to (ii), we remark that, conditionally on $\mathscr{G}_{n\Delta}$, the random variables $Z_1, \ldots, Z_n$ are independent and Z_i has distribution $\mathscr{N}(0, \overline{V}_i)$. Thus, for real numbers $\lambda_1, \ldots, \lambda_n$,

$$\mathrm{E}\left[\exp \sum_{j=1}^n \mathrm{i}\lambda_j Z_j / \mathscr{G}_{n\Delta} \right] = \exp -\frac{1}{2} \sum_{j=1}^n \lambda_j^2 \overline{V}_j.$$

Because the right-hand side above is $\sigma(U_1, \ldots, U_n)$-measurable and $\sigma(U_1, \ldots, U_n) \subset \mathscr{G}_{n\Delta}$,

$$\mathrm{E}\left[\exp \sum_{j=1}^n \mathrm{i}\lambda_j Z_j | U_1, \ldots, U_n \right] = \exp -\frac{1}{2} \sum_{j=1}^n \lambda_j^2 \overline{V}_j.$$

This gives properties (ii) and (iii) of Definition 3.1. □

Remark. The proof above shows that the process (X_i) is itself a strictly stationary Markov chain and that (Z_i) is also a hidden Markov model with (X_i) as hidden Markov chain. This allows us to extend our results to the case where (Y_t) has a drift term depending only on (V_t).

Since $(V_{i\Delta})$ is ergodic, it can be proved, using Theorem 2.3, that this implies the ergodicity of (U_i). But we have also the following stronger result.

Proposition 3.2. *Under (A0)–(A3), the process (Z_i) is α-mixing, with $\alpha_Z(k) \leqslant \alpha_V((k-1)\Delta)$.*

Proof. In view of Proposition 3.1, it is enough to prove that (U_i) is α-mixing. Since (U_i) is strictly stationary Markov, $c_U(k) = c(\sigma(U_1), \sigma(U_{k+1}))$, with $c = \alpha$, β or ρ. Hence

$$c_U(k) \leqslant c_V((k-1)\Delta).$$

By Proposition 2.7, (V_t) is β-mixing, so α-mixing. Proposition 3.1 gives the result. □

3.3. Limit theorems

In Genon-Catalot *et al.* (1998; 1999), where $\Delta = \Delta_n \to 0$, we have considered empirical estimators of the form $\sum \varphi(Z_i)/n$. For fixed Δ, we can study functions of successive observations to keep the information contained in the covariance structure. Let d be a positive integer and $\varphi : \mathbb{R}^d \to \mathbb{R}$ a Borel function, and consider, when defined, the function $h_\varphi : (\mathbb{R}^+)^d \to \mathbb{R}$ given by

$$h_\varphi(v_1, \ldots, v_d) = \mathrm{E}(\varphi(\varepsilon_1 v_1^{1/2}, \ldots, \varepsilon_d v_d^{1/2})), \tag{28}$$

where $(\varepsilon_1, \ldots, \varepsilon_d)$ are independent and identically distributed standard Gaussian random variables. In particular, when $d = 1$ and $\varphi(z) = z^{2p}$, we have $h_\varphi(v) = C_{2p} v^p$, where

$$C_k = \mathrm{E}|\varepsilon_1|^k. \tag{29}$$

When $d = 2$ and $\varphi(z_1, z_2) = z_1^{2p} z_2^{2q}$, we have $h_\varphi(v_1, v_2) = C_{2p} C_{2q} v_1^p v_2^q$.

Theorem 3.2. *Assume (A0)–(A3). Then, if* φ is such that $\mathrm{E}|h_\varphi(\overline{V}_1, \ldots, \overline{V}_d)| < \infty$,

$$\frac{1}{n}\sum_{i=0}^{n-d} \varphi(Z_{i+1}, \ldots, Z_{i+d}) \overset{\text{a.s.}}{\to} \mathrm{E}h_\varphi(\overline{V}_1, \ldots, \overline{V}_d) \qquad \text{as } n \to \infty.$$

Proof. From Proposition 3.2, the process (Z_i) is ergodic. Therefore, it is enough to check that $\mathrm{E}|\varphi(Z_1, \ldots, Z_d)|$ is finite and

$$\mathrm{E}(\varphi(Z_1, \ldots, Z_d)) = \mathrm{E}h_\varphi(\overline{V}_1, \ldots, \overline{V}_d).$$

This is obtained by conditioning on $\mathscr{G}_{d\Delta} = \sigma(V_s, 0 \leqslant s \leqslant d\Delta)$. □

Theorem 3.3. *Assume (A0)–(A3). Set* $\Phi_i = \varphi(Z_{i+1}, \ldots, Z_{i+d})$*. If, for some positive* δ*,* $\mathrm{E}|\Phi_0|^{2+\delta} < \infty$ *and* $\sum_{k\geqslant 1}\alpha_V^{2/(2+\delta)}(k\Delta) < \infty$*, the quantity*

$$\Sigma_\Delta(\varphi, d) = \mathrm{var}(\Phi_0) + 2\sum_{i=1}^{\infty} \mathrm{cov}(\Phi_0, \Phi_i) \tag{30}$$

is well defined and non-negative. If $\Sigma_\Delta(\varphi, d) > 0$*, then*

$$\frac{1}{n^{1/2}}\sum_{i=0}^{n-d}(\Phi_i - \mathrm{E}h_\varphi(\overline{V}_1, \ldots, \overline{V}_d)) \overset{\mathscr{D}}{\to} \mathscr{N}(0, \Sigma_\Delta(\varphi, d)) \qquad \text{as } n \to \infty.$$

Proof. We apply the Ibragimov central limit theorem for strictly stationary α-mixing sequences (see, for example, Hall and Heyde 1980, Corollary 5.1, p. 132). The α-mixing coefficient of the sequence (Φ_i) satisfies (see Proposition 3.2)

$$\alpha_\Phi(n) \leqslant \alpha_Z((n+1-d)\Delta) \leqslant \alpha_V((n-d)\Delta).$$

Therefore, by our assumptions $\Sigma_\Delta(\varphi, d) = \lim \mathrm{var}(\Phi_0 + \ldots + \Phi_{n-d})/n$ exists and is non-negative. Now, if it is positive, the central limit theorem holds. □

Theorem 3.3 has an immediate multidimensional version. Considering p functions φ_l, let us set, as above, $\Phi_i^l = \varphi_l(Z_{i+1}, \ldots, Z_{i+d})$ and

$$\Sigma_\Delta(\varphi_l, \varphi_k; d) = \mathrm{cov}(\Phi_0^l, \Phi_0^k) + \sum_{i=1}^{\infty}(\mathrm{cov}(\Phi_0^l, \Phi_i{}^k) + \mathrm{cov}(\Phi_0^k, \Phi_i^l)),$$

so that $\Sigma_\Delta(\varphi, d) = \Sigma_\Delta(\varphi, \varphi; d)$.

Corollary 3.1. *Assume (A0)-(A3) and that, for some positive* δ*,* $\sum_{k\geqslant 1}\alpha_V^{2/(2+\delta)}(k\Delta) < \infty$*. If, for all* $1 \leqslant l \leqslant p$*,* $\mathrm{E}|\Phi_0^l|^{2+\delta} < \infty$*, then the matrix* $\Sigma_\Delta = (\Sigma_\Delta(\varphi_l, \varphi_k))_{1\leqslant l,k\leqslant p}$ *is well defined. Moreover, if it is positive definite, we have*

$$\frac{1}{n^{1/2}}\sum_{i=0}^{n-d}\begin{pmatrix} \Phi_0^1 - \mathrm{E}h_{\varphi_1}(\overline{V}_1, \ldots, \overline{V}_d) \\ \vdots \\ \Phi_0^p - \mathrm{E}h_{\varphi_p}(\overline{V}_1, \ldots, \overline{V}_d) \end{pmatrix} \overset{\mathscr{D}}{\to} \mathscr{N}_p(0, \Sigma_\Delta) \qquad \text{as } n \to \infty.$$

The results of Section 2 enable us to check the condition on the α-mixing coefficient of (V_t) (see Proposition 2.8 and Corollary 2.1). In particular, if (V_t) is ρ-mixing, this condition holds automatically for any δ.

To check the moment assumption of Theorem 3.3 for functions with polynomial growth, the following proposition is useful:

Proposition 3.3. *Assume that, for some constant* K, $|\varphi(z_1, \ldots, z_d)| \leqslant K(1 + \sum_{i=1}^d |z_i|^q)$. *Then, if* $\mathrm{E}V_0^{q(1+\delta/2)} < \infty$, *we have* $\mathrm{E}|\Phi_0|^{2+\delta} < \infty$.

Proof. The assumption implies, for another constant K',

$$|\Phi_0|^{2+\delta} \leqslant K'\left(1 + \sum_{i=1}^{d} |Z_i|^{q(2+\delta)}\right).$$

Using notation (29) and the Hölder inequality, we obtain

$$\mathrm{E}|Z_i|^{q(2+\delta)} = C_{q(2+\delta)}\mathrm{E}\overline{V}_i^{q(1+\delta/2)} \leqslant C_{q(2+\delta)}\mathrm{E}V_0^{q(1+\delta/2)}.$$

This gives the result. □

3.4 Asymptotic variances for empirical moments

The idea is to use the previous limit theorems for statistical applications. In particular, for polynomial functions φ, we shall obtain empirical estimators of the parameters of the hidden diffusion. This justifies the computation of the above asymptotic variances. By conditional independence, we have, for $i \geqslant d$,

$$\mathrm{cov}(\Phi_0, \Phi_i) = \mathrm{cov}(h_\varphi(\overline{V}_1, \ldots, \overline{V}_d), h_\varphi(\overline{V}_{i+1}, \ldots, \overline{V}_{i+d})). \tag{31}$$

Proposition 3.4. *Assume (A0)–(A3) and that, for some positive* δ, $\sum_{k\geqslant 1}\alpha_V^{2/(2+\delta)}(k\Delta) < \infty$.

(i) For $\varphi(z) = z^{2p}$, *we have, if* $\mathrm{E}V_0^{2p(1+\delta/2)} < \infty$,

$$\Sigma_\Delta(z^{2p}, 1) = C_{2p}^2\left(\mathrm{var}(\overline{V}_1^p) + 2\sum_{i=1}^{\infty}\mathrm{cov}(\overline{V}_1^p, \overline{V}_{i+1}^p)\right) + (C_{4p} - C_{2p}^2)(\mathrm{E}\overline{V}_1^{2p}).$$

(ii) For $\varphi(z_1, z_2) = z_1^{2p}z_2^{2q}$, *we have, if* $q \leqslant p$ *and* $\mathrm{E}V_0^{4p(1+\sigma/2)} < \infty$,

$$\begin{aligned}\Sigma_\Delta(z_1^{2p}z_2^{2q}, 2) = {} & C_{2p}^2C_{2q}^2\left(\mathrm{var}(\overline{V}_1^p\overline{V}_2^q) + 2\sum_{i=1}^{\infty}\mathrm{cov}(\overline{V}_1^p\overline{V}_2^q, \overline{V}_{i+1}^p\overline{V}_{i+2}^q)\right)\\ & + (C_{4p}C_{4q} - C_{2p}^2C_{2q}^2)\mathrm{E}(\overline{V}_1^{2p}\overline{V}_2^{2q})\\ & + 2C_{2p}C_{2q}(C_{2(p+q)} - C_{2p}C_{2q})\mathrm{E}(\overline{V}_1^p\overline{V}_2^{p+q}\overline{V}_3^q).\end{aligned}$$

Proof. For the two parts, the moment condition comes from Proposition 3.3. We have to compute the terms appearing in (30). For (i), using (28)–(29), we obtain

$$\operatorname{var}\Phi_0 = \operatorname{var} Z_1^{2p} = C_{4p}\mathrm{E}\overline{V}_1^{2p} - C_{2p}^2(\mathrm{E}\overline{V}_1^p)^2.$$

Now, using (31), for $i \geqslant 1$,

$$\operatorname{cov}(\Phi_0, \Phi_i) = C_{2p}^2 \operatorname{cov}(\overline{V}_1^p, \overline{V}_{i+1}^p)$$

This gives the first expression. For (ii),

$$\operatorname{var}\Phi_0 = \operatorname{var} Z_1^{2p} Z_2^{2q} = C_{4p}C_{4q}\mathrm{E}\overline{V}_1^{2p}\overline{V}_2^{2q} - C_{2p}^2 C_{2q}^2(\mathrm{E}\overline{V}_1^p\overline{V}_2^q)^2$$

and

$$\operatorname{cov}(\Phi_0, \Phi_1) = \operatorname{cov}(Z_1^{2p} Z_2^{2q},\ Z_2^{2p} Z_3^{2q} = C_{2p}C_{2(p+q)}C_{2q}\mathrm{E}\overline{V}_1^p\overline{V}_2^{p+q}\overline{V}_3^q - C_{2p}^2 C_{2q}^2(\mathrm{E}\overline{V}_1^p\overline{V}_2^q)^2.$$

Using (31) again, we have, for $i \geqslant 2$,

$$\operatorname{cov}(\Phi_0, \Phi_i) = C_{2p}^2 C_{2q}^2 \operatorname{cov}(\overline{V}_1^p\overline{V}_2^q, \overline{V}_{i+1}^p\overline{V}_{i+2}^q).$$

Rearranging the terms together leads to the result. □

Let us remark that the first terms in the above formulae correspond (up to a multiplicative constant) respectively to the asymptotic variance of

$$n^{-1/2}\sum_{i=1}^{n}(\overline{V}_{i+1}^p - \mathrm{E}\overline{V}_1^p) \quad \text{and} \quad n^{-1/2}\sum_{i=1}^{n-1}(\overline{V}_{i+1}^p\overline{V}_{i+1}^q - \mathrm{E}\overline{V}_1^p\overline{V}_2^q).$$

The case $\varphi(z) = z^2$ is of special interest.

Corollary 3.2. *Assume (A0)–(A3) and that, for some positive δ, $\sum_{k\geqslant 1}\alpha_V^{2/(2+\delta)}(k\Delta) < \infty$. If $\mathrm{E}V_0^{2+\delta} < \infty$, and $\sigma^2 = 4M\int_l^r s(x)\,\mathrm{d}x(\int_l^x(u-\beta)\,\mathrm{d}\pi(u))^2 < \infty$, with $\beta = \mathrm{E}V_0$ and M given in (A2), then*

$$\Sigma_\Delta(z^2, 1) = 2\mathrm{E}(\overline{V}_1^2) + \sigma^2/\Delta.$$

Proof. From Proposition 3.4, it is enough to prove that $\sigma^2/\Delta = \Sigma$, with

$$\Sigma = \operatorname{var}(\overline{V}_1) + 2\sum_{i=1}^{\infty}\operatorname{cov}(\overline{V}_1, \overline{V}_{i+1}).$$

First, note that $\beta = \mathrm{E}V_0 = \mathrm{E}(\overline{V}_1)$ and that

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n}(\overline{V}_i - \beta) = \frac{1}{\Delta^{1/2}}\frac{1}{(n\Delta)^{1/2}}\int_0^{n\Delta}(V_s - \beta)\,\mathrm{d}s. \tag{32}$$

We have

$$\mathrm{cov}(\overline{V}_1, \overline{V}_{i+1}) = \frac{1}{\Delta^2}\int_{0\leqslant s\leqslant\Delta, i\Delta\leqslant s'\leqslant(i+1)\Delta} \mathrm{cov}(V_0, V_{s'-s})\,\mathrm{d}s\,\mathrm{d}s'.$$

Then, using (12), we see that the above covariance is non-negative for $i \geqslant 1$, so that $\Sigma \geqslant \mathrm{var}(\overline{V_1}) > 0$. Thus, the central limit theorem holds, and the left-hand side of (32) converges to $\mathcal{N}(0, \Sigma)$. By Theorem 2.2 and (23), the right-hand side term of (32) also converges in distribution to $\mathcal{N}(0, \sigma^2/\Delta)$. So the asymptotic variances are equal. □

When (V_t) is ρ-mixing, and $\mathrm{E}V_0^2 < \infty$, the condition on the α-mixing coefficient is satisfied, and, by Proposition 2.2 and the comment given afterwards, the condition $\sigma^2 < +\infty$ is automatic, and the equality of Corollary 3.2 holds without any further assumption.

Let us point out the consistency of the above result with the one obtained for small sampling intervals. In Genon-Catalot *et al.* (1998) it was proved, when $\Delta = \Delta_n \to 0$, that

$$(n\Delta_n)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \beta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2) \qquad \text{as } n \to \infty.$$

Here,

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \beta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2/\Delta + 2\mathrm{E}(\overline{V}_1^2)) \qquad \text{as } n \to \infty.$$

Thus, for the same estimator, the fixed sampling framework provides an extra term in the asymptotic variance.

4. Application to some classical models in finance

4.1. Mean reverting hidden diffusion

We focus our attention on models of the form

$$\mathrm{d}Y_t = \sigma_t\,\mathrm{d}B_t, \qquad Y_0 = 0,$$
$$\mathrm{d}V_t = \alpha(\beta - V_t)\,\mathrm{d}t + a(V_t)\,\mathrm{d}W_t, \qquad V_0 = \eta, \quad V_t = \sigma_t^2,$$

where $\alpha > 0$, $\beta > 0$ and $a(V_t)$ may also depend on unknown parameters. Due to the mean reverting drift of (V_t), these models possess some special features.

Proposition 4.1. *Assume that the above hidden diffusion* (V_t) *satisfies* (*A1*)–(*A3*) *and that* $\mathrm{E}V_0^2$ *is finite. Then,* $\mathrm{E}\overline{V}_1 = \mathrm{E}V_0 = \beta$, *and*

$$\mathrm{E}\overline{V}_1^2 = \beta^2 + \mathrm{var}(V_0)\frac{2(\alpha\Delta - 1 + \mathrm{e}^{-\alpha\Delta})}{\alpha^2\Delta^2}, \qquad \mathrm{E}\overline{V}_1\overline{V}_2 = \beta^2 + \mathrm{var}(V_0)\frac{(1 - \mathrm{e}^{-\alpha\Delta})^2}{\alpha^2\Delta^2}. \qquad (33)$$

Moreover, if (V_t) *is* ρ*-mixing, the term* σ^2 *in Corollary 3.2 is equal to* $2\,\mathrm{var}(V_0)/\alpha$.

Proof. Using (A1), $\mathrm{E}a^2(V_0) < \infty$, so $\int_0^t a(V_s)\,\mathrm{d}W_s$ is a martingale. We easily deduce that

$\beta = \mathrm{E} V_0 = \mathrm{E}\overline{V}_1$. Setting $U_t = \mathrm{e}^{\alpha t} V_t$, we obtain

$$U_t = U_0 + \alpha\beta \int_0^t \mathrm{e}^{\alpha s}\, \mathrm{d}s + \int_0^t \mathrm{e}^{\alpha s} a(V_s)\, \mathrm{d}W_s,$$

where the last term is also a martingale. Using this property, we obtain, for $s \leqslant s'$,

$$\mathrm{E} U_s U_{s'} = \alpha\beta\, \mathrm{e}^{\alpha s} \int_s^{s'} \mathrm{e}^{\alpha u}\, \mathrm{d}u \mathrm{E} V_0 + \mathrm{e}^{2\alpha s} \mathrm{E} V_0^2.$$

Hence

$$\mathrm{E}(V_s V_{s'}) = \mathrm{e}^{-\alpha(s'-s)} \operatorname{var}(V_0) + \beta^2.$$

We obtain (33), noting that

$$\mathrm{E}\overline{V}_1^2 = \frac{2}{\Delta^2} \int_{0 \leqslant s \leqslant s' \leqslant \Delta} \mathrm{E}(V_s V_{s'})\, \mathrm{d}s\, \mathrm{d}s', \qquad \mathrm{E}\overline{V}_1 \overline{V}_2 = \frac{1}{\Delta^2} \int_0^\Delta \mathrm{d}s \int_\Delta^{2\Delta} \mathrm{d}s' \mathrm{E}(V_s V_{s'}).$$

Now, if (V_t) is ρ-mixing, the function $f(v) = v - \beta$ satisfies the assumptions of Theorem 2.2, and $\mathscr{A} f = Lf = -\alpha f$ (see (19)). So, using (23) $g = (-1/\alpha) f$, we obtain $\sigma^2 = (2/\alpha)\langle f, f\rangle$, which is the result. □

The last part of the above result relies on the fact that $v - \beta$ is an eigenfunction of the infinitesimal generator with eigenvalue $-\alpha$, as pointed out by Hansen *et al.* (1998, paragraph 5.2).

4.2. Two examples

Let us first consider the model

$$\mathrm{d}Y_t = \sigma_t\, \mathrm{d}B_t, \qquad Y_0 = 0,$$
$$\mathrm{d}V_t = \alpha(\beta - V_t)\, \mathrm{d}t + c V_t\, \mathrm{d}W_t, \qquad V_0 = \eta, \quad V_t = \sigma_t^2,$$

where α, β, c are real numbers. This appears as the diffusion approximation of a GARCH(1,1)-M model (see Nelson 1990). First, we must check that assumptions (A1) and (A2) hold. Assumption (A1) is clearly verified, with $(l, r) = (0, +\infty)$. Let us set

$$a = 1 + 2\alpha/c^2, \qquad \mu = 2\beta\alpha/c^2.$$

Then the function $s(v)$ and the constant M are given by

$$s(v) = K v^{a-1} \exp\frac{\mu}{v}, \qquad M = \frac{1}{kc^2} \frac{\Gamma(a)}{\mu^a},$$

where $\Gamma(a)$ is the usual gamma function. Assumption (A2) holds if and only if $\mu > 0$ and $a > 0$, and the stationary distribution π has density

$$\pi(v) = \frac{\mu^a}{\Gamma(a)} v^{-a-1} \exp\left(-\frac{\mu}{v}\right) \mathbb{1}_{\{v>0\}}.$$

This is the inverse gamma distribution with parameters (a, μ). Thus $V_0 = \eta$ has distribution π if and only if $T = \eta^{-1}$ is distributed according to a gamma distribution $\Gamma(a, \mu)$. From Section 2.6, the conditions for ρ-mixing are $\alpha\beta > 0$, $\alpha > -c^2/2$, that is to say, the conditions required for (A2).

Now we compute the moments of π, which are given by

$$m(p) = \mathrm{E}(\eta^p) = \mu^p \frac{\Gamma(a-p)}{\Gamma(a)} \qquad \text{if } p < a,$$

and $+\infty$ if $p \geqslant a$. In particular, if $a > 2$, $\mathrm{E}(\eta) = \mu/(a-1) = \beta$ and $\mathrm{var}(\eta) = \mu^2/(a-1)^2(a-2) = \beta^2/((2\alpha/c^2) - 1)$. Note that the conditions $a > 2$, $\mu > 0$ are equivalent to $\beta > 0$, $\alpha > c^2/2$.

The following functions of the observations

$$\widehat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} Z_i^2, \qquad \widehat{m}_2 = \frac{1}{3n}\sum_{i=1}^{n} Z_i^4, \qquad \widehat{m}_{12} = \frac{1}{n}\sum_{i=1}^{n-1} Z_i^2 Z_{i+1}^2, \tag{34}$$

are respectively consistent estimators of β, $\mathrm{E}\overline{V}_1^2$ and $\mathrm{E}\overline{V}_1\overline{V}_2$. The last two quantities can be computed using (33). Inverting the formulae leads to consistent estimators of α, β and c^2. Furthermore, Proposition 3.4, Corollary 3.2 and Proposition 4.1 enable us to compute the asymptotic variances.

Our second example was proposed by Heston (1993). We consider for (V_t) the classical square-root process used by Cox *et al.* (1985) for interest rates:

$$\mathrm{d}V_t = \alpha(\beta - V_t)\,\mathrm{d}t + cV_t^{1/2}\,\mathrm{d}W_t, \qquad V_0 = \eta, \quad V_t = \sigma_t^2,$$

where α, β, c are real numbers. Now we set

$$a = 2\beta\alpha/c^2, \qquad \mu = 2\alpha/c^2.$$

Then the function $s(v)$ and the constant M are given by

$$s(v) = Kv^{-a}\,\mathrm{e}^{\mu v}, \qquad M = \frac{1}{Kc^2}\frac{\Gamma(a)}{\mu^a}.$$

Assumption (A2) holds if and only if $\mu > 0$ and $a \geqslant 1$ (i.e. $\alpha > 0$, $\alpha\beta \geqslant c^2/2$). The stationary distribution π has density

$$\pi(v) = \frac{\mu^a}{\Gamma(a)} v^{a-1}\,\mathrm{e}^{-\mu v} 1_{\{v>0\}}.$$

This is the gamma distribution with parameters (a, μ). The ρ-mixing condition holds (this is well known, since, for this model, the spectrum of the infinitesimal generator is discrete). The gamma distribution has moments of order p for all positive p which are

$$m(p) = \mathrm{E}(\eta^p) = \frac{\Gamma(a+p)}{\Gamma(a)\mu^p}.$$

In particular, $\mathrm{E}(\eta) = a/\mu = \beta$ and $\mathrm{var}(\eta) = a/\mu^2 = (c^2/2\alpha)\beta$. Again, $\widehat{m}_1$, $\widehat{m}_2$ and $\widehat{m}_{12}$ (see (34)) yield consistent estimators of α, β and c^2, whose asymptotic variances may be computed as above.

Appendix

This appendix is devoted to the proof of Proposition 2.8. Let us fix x_0 in (l, r), and consider the scale function $S(x) = \int_{x_0}^{x} s(t)\,\mathrm{d}t$ of (X_t) (see (16)). Let us define, for $x \in (l, r)$,

$$\varphi(S(x)) = a(x)s(x)\int_x^r m(v)\,\mathrm{d}v, \qquad \psi(S(x)) = a(x)s(x)\int_l^x m(v)\,\mathrm{d}v, \tag{35}$$

$$C_1(z) = \sup\{\varphi^2(y);\ y \geqslant z\}, \quad C_0(z) = \sup\{\psi^2(y);\ y \leqslant z\}, \quad C = \inf_{z\in\mathbb{R}} \max\{C_1(z), C_0(z)\}, \tag{36}$$

$$C' = M \sup_{\epsilon>0} \frac{\epsilon}{(\epsilon+1)^2} \int_l^r s(v)\,\mathrm{d}v \left(\inf\left\{ \int_l^x \pi(v)\,\mathrm{d}v,\ \int_x^r \pi(v)\,\mathrm{d}v \right\}\right)^{\epsilon+1} \mathrm{d}y. \tag{37}$$

The sketch of the proof is the following. If the limits in (A5) are finite, then $C < \infty$ and

$$\lambda \geqslant \frac{1}{8C}. \tag{38}$$

On the other hand, if $\lambda > 0$, $\lambda^{-1} \geqslant 8C'$ implies that $C' < \infty$, which in turn implies that the limits in (A5) are finite.

By (A2), the scale function S is increasing and one-to-one from (l, r) to $\mathbb{R}$. Therefore, $Y_t = S(X_t)$ has the same ρ-mixing coefficient as (X_t). So, we compute the ρ-mixing coefficient of (Y_t). By Itô's formula, (Y_t) satisfies $\mathrm{d}Y_t = V(Y_t)\,\mathrm{d}W_t$ with

$$v(y) = a(S^{-1}(y))s(S^{-1}(y)). \tag{39}$$

The stationary distribution of (Y_t) is equal to

$$\mu(\mathrm{d}y) = h(y)\,\mathrm{d}y, \qquad h(y) = \frac{1}{M}\frac{1}{v^2(y)} \tag{40}$$

with M given in (A2). Let $\mathscr{A}_0$ be the infinitesimal generator of (Y_t) on $L^2(\mathbb{R}, \mu)$, with domain $\mathscr{D}_0$. Applying (14), (18), (21) and (22) to (Y_t) yields

$$\lambda = \frac{1}{2} \inf\left\{ \frac{\int_{\mathbb{R}} (f'(y))^2\,\mathrm{d}y}{\int_{\mathbb{R}} f^2(y)/v^2(y)\,\mathrm{d}y};\ f \in D,\ \int_{\mathbb{R}} \frac{f(y)}{v^2(y)}\,\mathrm{d}y = 0 \right\}, \tag{41}$$

where D is a core of $\mathscr{A}_0$. Let us note that, for $y \in \mathbb{R}$ (see (35)),

$$\varphi(y) = v(y)\int_y^{+\infty} \frac{\mathrm{d}u}{v^2(u)}, \qquad \psi(y) = v(y)\int_{-\infty}^{y} \frac{\mathrm{d}u}{v^2(u)} \tag{42}$$

The proof of Proposition 2.8 is obtained using the three lemmas stated below.

Lemma A.1. *Under (A4) and (A5),*

$$\lim_{y\to+\infty} \varphi(y) = \lim_{x\to r} 1/\gamma(x) = \delta_1 \in [0, +\infty], \qquad \lim_{y\to-\infty} \psi(y) = -\lim_{x\to l} 1/\gamma(x) = \delta_0 \in [0, +\infty]$$

Proof. Using (39) and noting that $v'(y) = \gamma(S^{-1}(y))$, we remark that, applying (A4) and (A5),

$$\lim_{y\to+\infty} \frac{1}{v(y)} = 0 \qquad \text{and} \qquad \lim_{y\to+\infty} \frac{1}{v'(y)} = \delta_1$$

exist in $\overline{\mathbb{R}}$. By L'Hôpital's rule,

$$\lim_{y\to+\infty} \varphi(y) = \lim_{y\to+\infty} \frac{1/v^2(y)}{v'(y)/v^2(y)} = \delta_1 \in [0, +\infty].$$

Analogously, (A4) and (A5) imply that

$$\lim_{y\to-\infty} \psi(y) = -\lim_{y\to+\infty} \frac{1/v^2(y)}{v'(y)/v^2(y)} = \delta_0 \in [0, +\infty].$$

□

Lemma A.2. (Sufficient condition). *Under (A1)–(A5), if* $\delta_0 < +\infty$ *and* $\delta_1 < +\infty$*, then* $C < +\infty$ and $\lambda \geqslant 1/8C > 0$*, where C is defined in (36).*

Proof. We proceed as in Ledoux (1999, Section 4, pp. 173–176). From Lemma A.1, if δ_0 and δ_1 are finite, it is clear that $C_1(z)$, $C_0(z)$ and C are finite. To bound λ, we use (41) with $D = C_c^2(\mathbb{R})$ as a core for $\mathscr{A}_0$. For $z \in \mathbb{R}$, if $f \in C_c^2(\mathbb{R})$ and $\int_{\mathbb{R}} f(y)/v^2(y)\,\mathrm{d}y = 0$,

$$\int_{\mathbb{R}} \frac{f^2(y)}{v^2(y)}\,\mathrm{d}y = \inf_{z\in\mathbb{R}} I(z), \tag{43}$$

with

$$I(z) = \int_z^{+\infty} \frac{(f(y)-f(z))^2}{v^2(y)}\,\mathrm{d}y + \int_{-\infty}^{z} \frac{(f(y)-f(z))^2}{v^2(y)}\,\mathrm{d}y = I_1(z) + I_0(z). \tag{44}$$

Integrating by parts yields

$$I_1(z) = \left[-(f(y)-f(z))^2 \int_y^{+\infty} \frac{\mathrm{d}u}{v^2(u)}\right]_z^{+\infty} + \int_z^{+\infty} 2f'(y)(f(y)-f(z)) \int_y^{+\infty} \frac{\mathrm{d}u}{v^2(u)}\,\mathrm{d}y$$
$$= 2\int_z^{+\infty} f'(y)\frac{f(y)-f(z)}{v(y)} v(y) \int_y^{+\infty} \frac{\mathrm{d}u}{v^2(u)}\,\mathrm{d}y.$$

Using the Cauchy–Schwarz inequality and (42),

$$I_1^2(z) \leqslant 4\int_z^{+\infty} f'^2(y)\,\mathrm{d}y \int_z^{+\infty} \left(\frac{f(y)-f(z)}{v(y)}\varphi(y)\right)^2 \mathrm{d}y.$$

We obtain

$$I_1^2(z) \leqslant 4C_1(z) I_1(z) \int_z^{+\infty} f'^2(y)\,\mathrm{d}y.$$

Analogously,

$$I_0^2(z) \leqslant 4C_0(z)I_0(z)\int_{-\infty}^{z} f'^2(y)\,\mathrm{d}y.$$

Therefore,

$$I(z) \leqslant 4\max\{C_1(z),\, C_0(z)\}\int_{\mathbb{R}} f'^2(y)\,\mathrm{d}y.$$

By (36) and (44),

$$\int_{\mathbb{R}} \frac{f^2(y)}{v^2(y)}\,\mathrm{d}y \leqslant 4C\int_{\mathbb{R}} f'^2(y)\,\mathrm{d}y.$$

Hence, using (41), $\lambda \geqslant 1/(8C)$. □

Lemma A.3 (Necessary condition). *Under* (*A1*)–(*A5*)*, assume that* $\lambda > 0$*. Then,* $1/\lambda \geqslant 8C'$*, where* C' *is given in* (*37*)*. Hence,* $C' < +\infty$*. On the other hand,* $C' < +\infty$ *implies that* δ_0 *and* δ_1 *are finite.*

Proof. We use (41) with another core D of $\mathscr{A}_0$ equal to

$$D = \{f \in L^2(\mathbb{R}, \mu),\ f' \text{ absolutely continuous with compact support}\}.$$

Set

$$D^+ = \left\{ f \in D,\ f' \geqslant 0 \quad \text{and} \quad \int_{\mathbb{R}} f(y)/v^2(y)\,\mathrm{d}y = 0 \right\}.$$

Then

$$\frac{1}{\lambda} \geqslant 2\sup\left\{ \frac{\int_{\mathbb{R}} f^2(y)/v^2(y)\,\mathrm{d}y}{\int_{\mathbb{R}} f'^2(y)\,\mathrm{d}y};\ f \in D^+ \right\}. \tag{45}$$

From now on, we follow closely Klaassen (1985, Theorem 2.2). For $f \in D^+$, using (40), (43) and (44), we obtain

$$\int_{\mathbb{R}} f^2(y)h(y)\,\mathrm{d}y = \inf_{z\in\mathbb{R}}\left\{ \int_z^{+\infty} \left(\int_z^y f'(t)\,\mathrm{d}t \right)^2 h(y)\,\mathrm{d}y + \int_{-\infty}^{z} \left(\int_y^z f'(t)\,\mathrm{d}t \right)^2 h(y)\,\mathrm{d}y \right\}. \tag{46}$$

Let $\theta\colon \mathbb{R} \to \mathbb{R}_+$ be an arbitrary measurable function and set

$$\Phi(t) = \min\left\{ \int_{-\infty}^{t} \theta(y)h(y)\,\mathrm{d}y,\ \int_t^{+\infty} \theta(y)h(y)\,\mathrm{d}y \right\}.$$

Since f' is non-negative, Klaassen's proof can be applied and leads to

$$\int f^2(y)h(y)\,\mathrm{d}y \geqslant \frac{(\int f'(y)\Phi(y)\,\mathrm{d}y)^2}{\int \theta^2(y)h(y)\,\mathrm{d}y}. \tag{47}$$

Using the notation of (40) and (37), set $F(y) = \mu((-\infty, y])$ and $G(y) = \min\{F(y), 1 - F(y)\}$. First, we choose $\theta = G^{(\varepsilon-1)/2}$. Computation yields

$$\int \theta^2(y)h(y)\,\mathrm{d}y \leqslant \frac{1}{\varepsilon}, \qquad \Phi \geqslant \frac{2}{\varepsilon+1} G^{(\varepsilon+1)/2}. \tag{48}$$

Second, for $K > 0$, it is possible to choose f_K such that $\int f_K h = 0$ and $f'_K = G^{(\varepsilon+1)/2}\varphi_K$ with $\varphi_K \in C_c^1(\mathbb{R})$, $\varphi_K \geqslant 0$ and $\varphi_K \uparrow 1$ as $K \uparrow +\infty$. Thus, $f_K \in D^+$. Therefore, by (45), (47) and (48), for all K, noting that $\varphi_K^2 \leqslant \varphi_K$,

$$\lambda^{-1} \geqslant \frac{8\varepsilon}{(\varepsilon+1)^2} \frac{(\int \varphi_K(y)G^{\varepsilon+1}(y)\,\mathrm{d}y)^2}{\int \varphi_K^2(y)G^{\varepsilon+1}(y)\,\mathrm{d}y} \geqslant \frac{8\varepsilon}{(\varepsilon+1)^2} \int \varphi_K(y)G^{\varepsilon+1}(y)\,\mathrm{d}y.$$

Now taking the limit as $K \to \infty$ yields

$$\lambda^{-1} \geqslant \frac{8\varepsilon}{(\varepsilon+1)^2} \int G^{\varepsilon+1}(y)\,\mathrm{d}y.$$

Since, setting $y = S(x)$ in (37),

$$C' = \sup_{\varepsilon>0} \frac{\varepsilon}{(\varepsilon+1)^2} \int G^{\varepsilon+1}(y)\,\mathrm{d}y,$$

we obtain $\lambda^{-1} \geqslant 8C'$, which implies $C' < +\infty$. Thus, we must have

$$C'' = \limsup_{\varepsilon\downarrow 0} \varepsilon \int G^{\varepsilon+1}(y)\,\mathrm{d}y < +\infty.$$

Now we must link this condition with δ_0 and δ_1. For this, note that

$$\delta_0 = \lim_{y\downarrow-\infty} Mv(y)F(y) = \lim_{t\downarrow 0} Mtv(F^{-1}(t))$$

and, analogously

$$\delta_1 = \lim_{t\uparrow 1} M(1-t)v(F^{-1}(t)).$$

Following the end of Klaassen's proof, if $\delta_0 = +\infty$ or $\delta_1 = +\infty$, then C'' is infinite. We conclude that δ_0 and δ_1 have to be finite. □

Acknowledgement

The authors would like to thank E.M. Ouhabaz for several helpful discussions.

References

Barndorff-Nielsen, O.E. and Shephard, N. (1998) Aggregation and model construction for volatility models. Centre for Analytical Finance, University of Aarhus, Working Paper Series No. 10.

Bhattacharya, R.N. (1982) On the functional limit theorem and the law of iterated logarithm for

Markov processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **60**, 185–201.
Bibby, B. and Sørensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.
Bickel, P.J. and Ritov, Y. (1996) Inference in hidden Markov models I: Local asymptotic normality in the stationary case. *Bernoulli*, **2**, 199–228.
Bradley, R.C. (1986) Basic properties of strong mixing conditions. In E. Eberlein and M.S. Taqqu (eds), *Dependence in Probability and Statistics, A Survey of Recent Results*, pp. 165–192, Boston: Birkhäuser.
Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) A theory of term structure of interest rates. *Econometrica*, **53**, 385–407.
Doukhan, P. (1994) *Mixing: Properties and Examples,* Lecture Notes in Statist. 85. Berlin: Springer-Verlag.
Drost, F.C. and Werker, B. (1996) Closing the GARCH gap: Continuous time GARCH modelling. *J. Econometrics*, **74**, 31–57.
Ethier, S.N. and Kurtz, T.G. (1986) *Markov Processes*. New York: Wiley.
Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998) Limit theorems for discretely observed stochastic volatility models. *Bernoulli*, **4**, 283–303.
Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999) Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, **5**, 855–872.
Ghysels, E., Harvey, A. and Renault, E. (1996) Stochastic volatility. In G.S. Maddala and C.R. Rao (eds), *Handbook of Statistics 14: Statistical Methods in Finance*, pp. 119–192. Amsterdam: North-Holland.
Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and its Application*. New York: Academic Press.
Hansen, L.P. and Scheinkman, J.A. (1995) Back to the future: generating moment implication for continuous-time Markov processes. *Econometrica*, **63**, 767–804.
Hansen, L.P., Scheinkman, J.A. and Touzi, N. (1998) Spectral methods for identifying scalar diffusions. *J. Econometrics*, **86**(1), 1–32.
Heston, S.L. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Stud.*, **6**, 327–343.
Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities. *J. Finance*, **42**, 281–300.
Jensen, J.L. and Petersen, N.V. (1998) Asymptotic normality of the maximum likelihood estimator in state space models. Research Reports of University of Aarhus, No. 392.
Kent, J. (1978) Time-reversible diffusions. *Adv. Appl. Probab.*, **10**, 819–835.
Kessler, M. (1996) Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.*, **27**(1), 65–82.
Klaassen, C.A.J. (1985) On a inequality of Chernoff. *Ann. Probab.*, **13**, 966–974.
Krengel, U. (1985) *Ergodic Theorems*. Berlin: de Gruyter.
Ledoux, M. (1999) Concentration of measure and logarithmic Sobolev inequalities. In J. Azéma, M. Emery, M. Ledoux and M. Yor (eds), *Séminaire de Probabilités XXXIII*, Lecture Notes in Math. 1709, pp. 120–216. Berlin: Springer-Verlag.
Leroux, B.G. (1992) Maximum likelihood estimation for hidden Markov models. *Stochastic Process. Appl.*, **40**, 127–143.
Lindgren, G. (1978) Markov regime models for mixed distributions and switching regressions. *Scand. J. Statist.*, **5**, 81–91.
Lindvall, T. (1983) On coupling of diffusion processes. *J. Appl. Probab.*, **20**, 82–93.

Nagel, P. (ed.) (1986) *One-Parameter Semigroups of Positive Operators*, *Lecture Notes in Math. 1184*. Berlin: Springer-Verlag.

Nelson, D.B. (1990) ARCH models as diffusion approximations. *J. Econometrics*, **45**, 7–38.

Orey, S. (1971) *Limit Theorems for Markov Chain Transition Probabilities*. London: Van Nostrand.

Rogers, L.C.G. and Williams, D. (1987) *Diffusions, Markov processes and Martingales. Vol. 2: Ito calculus*. New York: Wiley.

Veretennikov, A.Y. (1988) Bounds for the mixing rate in the theory of stochastic differential equations. *Theory Probab. Appl.*, **32**, 273–281.

Veretennikov, A.Y. (1997) On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.*, **70**, 115–127.

Received October 1998 and revised November 1999.