Conditional Likelihood Estimators for Hidden Markov Models and Stochastic Volatility Models

VALENTINE GENON-CATALOT

Université de Marne la Vallée

THIERRY JEANTHEAU

Université de Marne-la-Vallée

CATHERINE LAREDO

INRA-Jouy-en-Josas

ABSTRACT. This paper develops a new contrast process for parametric inference of general hidden Markov models, when the hidden chain has a non-compact state space. This contrast is based on the conditional likelihood approach, often used for ARCH-type models. We prove the strong consistency of the conditional likelihood estimators under appropriate conditions. The method is applied to the Kalman filter (for which this contrast and the exact likelihood lead to asymptotically equivalent estimators) and to the discretely observed stochastic volatility models.

Key words: conditional likelihood, diffusion processes, discrete time observations, hidden Markov models, parametric inference, stochastic volatility

1. Introduction

Parametric inference for hidden Markov models (HMMs) has been widely investigated, especially in the last decade. These models are discrete-time stochastic processes including classical time series models and many other non-linear non-Gaussian models. The observed process (Z_n) is modelled via an unobserved Markov chain (U_n) such that, conditionally on (U_n), the (Z_n)s are independent and the distribution of Z_n depends only on U_n. When studying the statistical properties of HMMs, a difficulty arises since the exact likelihood cannot be explicitly calculated. As a consequence, a lot of papers have been concerned with approximations by means of numerical and simulation techniques (see, e.g. Gouriéroux *et al.*, 1993; Durbin & Koopman, 1998; Pitt & Shephard, 1999; Kim *et al.*, 1998; Del Moral & Miclo, 2000; Del Moral *et al.*, 2001).

The theoretical study of the exact maximum likelihood estimator (m.l.e.) is a difficult problem and has only been investigated in the following cases. Leroux (1992) has proved the consistency when the unobserved Markov chain has a finite state space. Asymptotic normality is proved in Bickel *et al.* (1998). The extension of these properties to a compact state space for the hidden chain can be found in Jensen & Petersen (1999) and Douc & Matias (2001).

In previous papers (see Genon-Catalot *et al.*, 1998, 1999, 2000a), we have investigated some statistical properties of discretely observed stochastic volatility models (SV). In particular, when the sampling interval is fixed, the SV models are HMMs, for which the hidden chain has a non-compact state space. Using ergodicity and mixing properties, we have built empirical moment estimators of unknown parameters. Other types of empirical estimators are given in

Gallant *et al.* (1997) using the efficient moment method or by Sørensen (2000) considering the class of prediction-based estimating functions.

In this paper, we propose a new contrast for general HMMs, which is theoretical in nature but seems close to the exact likelihood. The contrast is based on the conditional likelihood method, which is the estimating method used in the field of ARCH-type models (see Jeantheau, 1998). The method is applied to examples. Particular attention is paid throughout this paper to the well-known Kalman filter which lightens our approach. It is a special case of HMM with non-compact state space for the hidden chain. All computations are explicit and detailed herein. The minimum contrast estimator based on the conditional likelihood is asymptotically equivalent to the exact m.l.e. In the case of the SV models, when the unobserved volatility is a positive ergodic diffusion, the method is also applicable. It requires that the state space of the hidden diffusion is open, bounded and bounded away from zero. Contrary to moment methods, only the finiteness of the first moment of the stationary distribution is needed.

In Section 2, we recall definitions, ergodic properties and expressions for the likelihood (Propositions 1 and 2) for HMMs. We define and study in Section 3 the conditional likelihood (Theorem 1 and Definition 2) and build the associated contrasts. Then, we state a general theorem of consistency to study the related minimum contrast estimators (Theorem 2). We apply this method to examples, especially to the Kalman filter and to GARCH(1,1) model. Then, we specialize these results to SV models in Section 4. We consider the conditional likelihood (Proposition 4) and study its properties (Proposition 5). Finally, we consider the case of mean reverting hidden diffusion. The proof of Theorem 2 is given in the appendix.

2. Hidden Markov models

2.1. Definition and ergodic properties

The formal definition of an HMM can be taken from Leroux (1992) or Bickel & Ritov (1996).

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

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

Above, $\mathscr{Z}$ and $\mathscr{U}$ are general Polish spaces equipped with their Borel sigma-fields. Therefore, this definition extends the one given in Leroux (1992) or Bickel & Ritov (1996) since we do not require a finite state space for the hidden chain.

We give now some examples of HMMs. They are all included in the following general framework. We set

$$Z_n = G(U_n, \varepsilon_n), \tag{1}$$

where $(\varepsilon_n)_{n \in \mathbb{N}^*}$ is a sequence of i.i.d. random variables, $(U_n)_{n \in \mathbb{N}^*}$ is a Markov chain with state space $\mathscr{U}$, and these two sequences are independent.

Example 1 (Kalman filter). The above class includes the well-known discrete Kalman filter

$$Z_n = U_n + \varepsilon_n, \tag{2}$$

where (U_n) is a stationary real AR(1) Gaussian process and (ε_n) are i.i.d. $\mathcal{N}(0, \gamma^2)$.

Example 2 (Noisy observations of a discretely observed Markov process). In (1), take $U_n = V_{n\Delta}$, where $(V_t)_{t\in\mathbb{R}_+}$ is any Markov process independent of the sequence $(\varepsilon_n)_{n\in\mathbb{N}^*}$.

Example 3 (Stochastic volatility models). These models are given in continuous time by a two-dimensional process (Y_t, V_t) with

$$\mathrm{d}Y_t = \sigma_t \mathrm{d}B_t,$$

and $V_t = \sigma_t^2$ (the so-called volatility) is an unobserved Markov process independent of the brownian motion (B_t). Then, a discrete observation with sampling interval Δ is taken. We set

$$Z_n = \frac{1}{\sqrt{\Delta}} \int_{(n-1)\Delta}^{n\Delta} \sigma_s \, \mathrm{d}B_s = \frac{1}{\sqrt{\Delta}} (Y_{n\Delta} - Y_{(n-1)\Delta}). \tag{3}$$

Conditionally on $(V_s, s \geq 0)$, the random variables (Z_n) are independent and Z_n has distribution $\mathcal{N}(0, \bar{V}_n)$ with

$$\bar{V}_n = \frac{1}{\Delta} \int_{(n-1)\Delta}^{n\Delta} V_s \, \mathrm{d}s. \tag{4}$$

Noting that $U_n = (\bar{V}_n, V_{n\Delta})$ is Markov, we set, in accordance with (1),

$$Z_n = G(U_n, \varepsilon_n) = \bar{V}_n^{1/2} \varepsilon_n.$$

Such models have been first proposed by Hull & White (1987), with (V_t) a diffusion process. When (V_t) is an ergodic diffusion, it is proved in Genon-Catalot *et al.* (2000a) that (Z_n) is an HMM. In Barndorff-Nielsen & Shephard (2001), the model is analogous, with (V_t) an Ornstein Uhlenbeck Levy process.

An HMM has the following properties.

Proposition 1

(1) *The process* $((U_n, Z_n), n \geqslant 1)$ *is a time-homogeneous Markov chain.*
(2) *If the hidden chain* $(U_n, n \geq 1)$ *is strictly stationary, so is* $((U_n, Z_n), n \geqslant 1)$.
(3) *If, moreover, the hidden chain* $(U_n, n \geqslant 1)$ *is ergodic, so is* $((U_n, Z_n), n \geqslant 1)$.
(4) *If, moreover,* $(U_n, n \geqslant 1)$ *is* α*-mixing, then* $((U_n, Z_n), n \geqslant 1)$ *is also* α*-mixing, and*

$$\alpha_Z(n) \leqslant \alpha_{(U,Z)}(n) \leqslant \alpha_U(n).$$

The first two points are straightforward, the third is proved in Leroux (1992), and the mixing property is proved in Genon-Catalot *et al.* (2000a, b) and Sørensen (1999). This mixing property holds for the previous examples.

Example 1 (continued) (Kalman filter). We set

$$U_n = aU_{n-1} + \eta_n, \tag{5}$$

where (η_n) are i.i.d. $\mathcal{N}(0, \beta^2)$. When $|a| < 1$, and U_n has law $\mathcal{N}(0, \tau^2 = \beta^2/(1 - a^2))$, U_n is α-mixing.

Examples 2 and 3 (continued). In both cases, whenever (V_t) is a diffusion, we have

$$\alpha_Z(n) \leqslant \alpha_V((n-1)\Delta).$$

For (V_t) a strictly stationary diffusion process, it is well known that

$$\alpha_V(t) \longrightarrow 0 \quad \text{as } t \longrightarrow +\infty.$$

For details (such as the rate of convergence of the mixing coefficient), we refer to Genon-Catalot *et al.* (2000a) and the references therein.

The above properties have interesting statistical consequences. When the unobserved chain depends on unknown parameters, the ergodicity and mixing properties of (Z_n) can be used to derive the consistency and asymptotic normality of empirical estimators of the form $1/n \sum_{i=0}^{n-d} \varphi(Z_{i+1}, \ldots, Z_{i+d})$. This leads to various procedures that have been already applied to models [Moment method, GMM (Hansen, 1982), EMM (Tauchen *et al.*, 1996; Gallant *et al.*, 1997) or prediction-based estimating equations (Sørensen, 2000)]. These methods yield estimators with rate $\sqrt{n}$. However, to be accurate, they require high moment conditions that may be not fulfilled, and often lead to biased estimators and/or to cumbersome computations. Another central question lies in the fact that they may be very far from the exact likelihood method.

2.2. The exact likelihood

We recall here general properties of the likelihood in an HMM model. Such likelihoods have already been studied but mainly under the assumption that the state space of the hidden chain is finite (see, e.g. Leroux, 1992; Bickel & Ritov, 1996). Since this is not the case in the examples above, we focus on the properties which hold without this assumption.

Consider a general HMM as in Definition 1 and let us give some more notations and assumptions.

- (H1) The conditional distribution of Z_n given $U_n = u$ is given by a density $f(z/u)$ with respect to a dominating measure $\mu(\mathrm{d}z)$ on $(\mathscr{Z}, \mathscr{B}(\mathscr{Z}))$.
- (H2) The transition operator P_θ of the hidden chain (U_i) depends on an unknown parameter $\theta \in \Theta \subset \mathbb{R}^p, p \geqslant 1$, and the transition probability is specified by a density $p(\theta, u, t)$ with respect to a dominating measure $v(\mathrm{d}t)$ on $(\mathscr{U}, \mathscr{B}(\mathscr{U}))$.
- (H3) For all θ, the transition P_θ admits a stationary distribution having density with respect to the same dominating measure, denoted by $g(\theta, u)$. The initial variable U_1 has this stationary distribution.
- (H4) For all θ, the chain (U_n) with marginal distribution $g(\theta, u)v(\mathrm{d}u)$ is ergodic.

Under these assumptions, the process (U_n, Z_n) is strictly stationary and ergodic. Let us point out that we do not introduce unknown parameters in the density of Z_n given $U_n = u$, since, in our examples, all unknown parameters will come from the hidden chain.

With these notations, the distribution of the initial variable (U_1, Z_1) of the chain (U_n, Z_n) has density $g(\theta, u)f(z/u)$ with respect to $v(\mathrm{d}u) \otimes \mu(\mathrm{d}z)$; the transition density is given by

$$p(\theta, u, t)f(z/t). \tag{6}$$

Now, if we denote by $p_n(\theta, z_1, \ldots, z_n)$ the density of $(Z_1, \ldots, Z_n)$ with respect to $\mathrm{d}\mu(z_1) \cdots \otimes \mathrm{d}\mu(z_n)$, we have

$$p_1(\theta, z_1) = \int_{\mathscr{U}} g(\theta, u) f(z_1/u) \, \mathrm{d}v(u), \tag{7}$$

$$p_n(\theta, z_1, \ldots, z_n) = \int_{\mathcal{U}^n} g(\theta, u_1) f(z_1/u_1) \prod_{i=2}^{n} p(\theta, u_{i-1}, u_i) f(z_i/u_i) \, \mathrm{d}\nu(u_1) \cdots \otimes \mathrm{d}\nu(u_n). \tag{8}$$

In formula (8), the function under the integral sign is the density of $(U_1, \ldots, U_n, Z_1, \ldots, Z_n)$. A more tractable expression for $p_n(\theta, z_1, \ldots, z_n)$ is obtained from the classical formula:

$$p_n(\theta, z_1, \ldots, z_n) = p_1(\theta, z_1) \prod_{i=2}^{n} t_i(\theta, z_i/z_{i-1}, \ldots, z_1), \tag{9}$$

where $t_i(\theta, z_i/z_{i-1}, \ldots, z_1)$ is the conditional density of Z_i given $Z_{i-1} = z_{i-1}, \ldots, Z_1 = z_1$. The interest of this representation appears below.

Proposition 2

(1) *For all* $i \geqslant 2$, *we have (see (9))*

$$t_i(\theta, z_i/z_{i-1}, \ldots, z_1) = \int_{\mathcal{U}} g_i(\theta, u_i/z_{i-1}, \ldots, z_1) f(z_i/u_i) \, \mathrm{d}\nu(u_i), \tag{10}$$

where $g_i(\theta, u_i/z_{i-1}, \ldots, z_1)$ *is the conditional density of* U_i *given* $Z_{i-1} = z_{i-1}, \ldots, Z_1 = z_1$.

(2) *If* $\hat{g}_i(\theta, u_i/z_i, \ldots, z_1)$ *denotes the conditional density of* U_i *given* $Z_i = z_i, \ldots, Z_1 = z_1$, *then it is given by*

$$\hat{g}_i(\theta, u_i/z_i, \ldots, z_1) = \frac{g_i(\theta, u_i/z_{i-1}, \ldots, z_1) f(z_i/u_i)}{t_i(\theta, z_i/z_{i-1}, \ldots, z_1)}.$$

(3) *For all* $i \geqslant 1$,

$$g_{i+1}(\theta, u_{i+1}/z_i, \ldots, z_1) = \int_{\mathcal{U}} \hat{g}_i(\theta, u_i/z_i, \ldots, z_1) p(\theta, u_i, u_{i+1}) \, \mathrm{d}\nu(u_i). \tag{11}$$

The result of Proposition 2 is standard in the field of filtering theory (see, e.g. Liptser & Shiryaev, 1978, and for details, Genon-Catalot *et al*., 2000b). It leads to a recursive expression of the exact likelihood function. Now, we are faced with two kinds of problems, a numerical one and a theoretical one.

The numerical computation of the exact m.l.e. of θ raises difficulties and there is a large amount of literature devoted to approximation algorithms (Markov chain Monte Carlo methods (MCMC), EM algorithm, particle filter method, etc.). In particular, MCMC methods have been recently considered in econometrics and applied to partially observed diffusions (see, e.g. Eberlein *et al*., 2001) .

From a theoretical point of view, for HMMs with a finite state space for the hidden chain, results on the asymptotic behaviour (consistency) of the m.l.e. are given in Leroux (1992), Francq & Roussignol (1997) and Bickel & Ritov (1996). Extensions to the case of a compact state space for the hidden chain have been recently investigated (Jensen & Petersen, 1999; Douc & Matias, 2001). Apart from these cases, the problem is open. Let us stress that, in the SV models, the state space of the hidden chain is not compact.

Nevertheless, there is at least one case where the state space of the hidden chain is non-compact and where the asymptotic behaviour of the m.l.e. is well known: the Kalman filter model defined in (2).

In this paper, we consider and study new estimators that are minimum contrast estimators based on the conditional likelihood. We can justify our approach by the fact that these estimators are asymptotically equivalent to the exact m.l.e. in the Kalman filter.

3. Contrasts based on the conditional likelihood

Our aim is to develop a theoretical tool for obtaining consistent estimators. A specific feature of HMMs is that the conditional law of Z_i given the past observations depends effectively on i and all the observations $Z_1, \dots, Z_{i-1}$. In order to recover some stationarity properties, we need to introduce the infinite past of each observation Z_i. This is the concern of the conditional likelihood. This estimation method derives from the field of discrete time ARCH models. Besides, as a tool, it appears explicitly in Leroux (1992, Section 4).

We first introduce and study the conditional likelihood, and then derive associated contrast functions leading to estimators.

3.1. Conditional likelihood

Since the process (U_n, Z_n) is strictly stationary, we can consider its extension to a process $((U_n, Z_n), n \in \mathbb{Z})$ indexed by $\mathbb{Z}$, with the same finite-dimensional distributions. For all $i \in \mathbb{Z}$, let $\underline{Z}_i = (Z_i, Z_{i-1}, \dots)$ be the vector of $\mathbb{R}^{\mathbb{N}}$ defining the past of the process until i. Below, we prove that the conditional distribution of the sample $(Z_1, \dots, Z_n)$ given the infinite past $\underline{Z}_0$ admits a density with respect to $\mu(\mathrm{d}z_1) \otimes \cdots \otimes \mu(\mathrm{d}z_n)$, which allows to define the conditional likelihood of $(Z_1, \dots, Z_n)$ given $\underline{Z}_0$.

From now on, let us suppose that $\mathscr{Z} = \mathbb{R}$ and $\mathscr{U} = \mathbb{R}^k$, for some $k \geqslant 1$. Denote by $\mathbb{P}_\theta$ the distribution of $((U_n, Z_n), n \in \mathbb{Z})$ on the canonical space, $((U_n, Z_n), n \in \mathbb{Z})$ will be the canonical process, and $E_\theta = E_{\mathbb{P}_\theta}$.

Theorem 1

Under (H1)–(H3), the following holds:

(1) *The conditional distribution, under $\mathbb{P}_\theta$, of U_1 given the infinite past $\underline{Z}_0$ admits a density with respect to $\nu(\mathrm{d}u)$ equal to*

$$\tilde{g}(\theta, u/\underline{Z}_0) = E_\theta(p(\theta, U_0, u)/\underline{Z}_0). \tag{12}$$

The function $\tilde{g}(\theta, u/\underline{Z}_0)$ is well defined under $\mathbb{P}_\theta$ as a measurable function of $(u, \underline{Z}_0)$, since there exists a regular version of the conditional distribution of U_0 given $\underline{Z}_0$.

(2) *The conditional distribution, under $\mathbb{P}_\theta$, of Z_1 given the infinite past $\underline{Z}_0$ admits a density with respect to $\mathrm{d}\mu(z_1)$ equal to*

$$\tilde{p}(\theta, z_1/\underline{Z}_0) = \int_{\mathscr{U}} \tilde{g}(\theta, u_1/\underline{Z}_0) f(z_1/u_1) \nu(\mathrm{d}u_1). \tag{13}$$

Proof of Theorem 1. We omit θ in all notations for the proof. By the Markov property of (U_n, Z_n), the conditional distribution of U_1 given $(U_0, Z_0), (U_{-1}, Z_{-1}), \dots, (U_{-n}, Z_{-n})$ is identical to the conditional distribution of U_1 given (U_0, Z_0), which is simply the conditional distribution of U_1 given U_0 [see (6)]. Consequently, for $\varphi : \mathscr{U} \to [0, 1]$ measurable,

$$E(\varphi(U_1)/U_0, Z_0, Z_{-1}, \dots, Z_{-n}) = E(\varphi(U_1)/U_0) = P\varphi(U_0), \tag{14}$$

with

$$P\varphi(U_0) = \int \varphi(u) p(U_0, u) \nu(\mathrm{d}u). \tag{15}$$

Hence,

$$E(\varphi(U_1)/Z_0, Z_{-1}, \dots, Z_{-n}) = E(P\varphi(U_0)/Z_0, Z_{-1}, \dots, Z_{-n}). \tag{16}$$

By the martingale convergence theorem, we get

$$E(\varphi(U_1)/\underline{Z}_0) = E(P\varphi(U_0)/\underline{Z}_0). \tag{17}$$

Now, using the fact that there is a regular version of the conditional distribution of U_0 given $\underline{Z}_0$, say $\mathrm{d}P_{U_0}(u_0/\underline{Z}_0)$, we obtain

$$E(\varphi(U_1)/\underline{Z}_0) = \int_{\mathscr{U}} P\varphi(u_0)\,\mathrm{d}P_{U_0}(u_0/\underline{Z}_0). \tag{18}$$

Applying the Fubini theorem yields

$$E(\varphi(U_1)/\underline{Z}_0) = \int_{\mathscr{U}} \varphi(u)E(p(U_0,u)/\underline{Z}_0)\nu(\mathrm{d}u). \tag{19}$$

So, the proof of (1) is complete.

Using again the Markov property of (U_n, Z_n) and (6), we get that the conditional distribution of Z_1 given $(U_1, Z_0, Z_{-1}, \dots, Z_{-n})$ is identical to the conditional distribution of Z_1 given U_1 which is simply $f(z_1/U_1)\mu(\mathrm{d}z_1)$. So taking $\varphi : \mathscr{Z} \to [0,1]$ measurable as above, we get

$$E(\varphi(Z_1)/Z_0, Z_{-1}, \dots, Z_{-n}) = E(E(\varphi(Z_1)/U_1)/Z_0, Z_{-1}, \dots, Z_{-n}). \tag{20}$$

So, using the martingale convergence theorem and Proposition 2, we obtain

$$E(\varphi(Z_1)/\underline{Z}_0) = \int_{\mathbb{R}} \varphi(z_1)\mu(\mathrm{d}z_1) \int_{\mathscr{U}} f(z_1/u_1)\tilde{g}(u_1/\underline{Z}_0)\nu(\mathrm{d}u_1). \tag{21}$$

This achieves the proof. □

Note that, under $\mathbb{P}_\theta$, by the strict stationarity, for all $i \geqslant 1$, the conditional density of Z_i given $\underline{Z}_{i-1}$ is given by

$$\tilde{p}(\theta, z_i/\underline{Z}_{i-1}) = \int_{\mathscr{U}} \tilde{g}(\theta, u_i/\underline{Z}_{i-1}) f(z_i/u_i)\nu(\mathrm{d}u_i). \tag{22}$$

Thus, the conditional distribution of $(Z_1, \dots, Z_n)$ given $\underline{Z}_0 = \underline{z}_0$ has a density given by

$$\tilde{p}_n(\theta, z_1, \dots, z_n/\underline{z}_0) = \prod_{i=1}^{n} \tilde{p}(\theta, z_i/\underline{z}_{i-1}). \tag{23}$$

Hence, we may introduce Definition 2.

Definition 2. Let us assume that, $\mathbb{P}_{\theta_0}$ a.s., the function $\tilde{p}(\theta, Z_n/\underline{Z}_{n-1})$ is well defined for all θ and all n. Then, the conditional likelihood of $(Z_1, \dots, Z_n)$ given $\underline{Z}_0$ is defined by

$$\tilde{p}_n(\theta, \underline{Z}_n) = \prod_{i=1}^{n} \tilde{p}(\theta, Z_i/\underline{Z}_{i-1}) = \tilde{p}_n(\theta, Z_1, \dots, Z_n/\underline{Z}_0). \tag{24}$$

Let us set when defined

$$l(\theta, \underline{z}_1) = \log \tilde{p}(\theta, z_1/\underline{z}_0), \tag{25}$$

so that

$$\log \tilde{p}_n(\theta) = \sum_{i=1}^{n} l(\theta, \underline{Z}_i).$$

We have Proposition 3.

Proposition 3

Under (H1)–(H4), if $E_{\theta_0}|l(\theta, \underline{Z}_1)| < \infty$, *then, we have, almost surely, under* $\mathbb{P}_{\theta_0}$,

$$\frac{1}{n}\log \tilde{p}_n(\theta, \underline{Z}_n) \to E_{\theta_0} l(\theta, \underline{Z}_1).$$

Proof of Proposition 3. Using Proposition 1 (3), (Z_n) is ergodic, and the ergodic theorem may be applied. □

Example 1 (continued) (Kalman filter). For the discrete Kalman filter [see (2)], the successive distributions appearing in Proposition 2 are explicitely known and Gaussian.

With the notations introduced previously, the unknown parameters are a, β^2 while γ^2 is supposed to be known and $|a| < 1$. We assume that (U_n) is in a stationary regime. The following properties are well known and are obtained following the steps of Proposition 2. Under $\mathbb{P}_\theta$, $(\theta = (a, \beta^2))$:

(i) the conditional distribution of U_n given $(Z_{n-1}, \dots, Z_1)$ is the law $\mathbb{N}(x_{n-1}, \sigma^2_{n-1})$;
(ii) the conditional distribution of U_n given $(Z_n, \dots, Z_1)$ is the law $\mathbb{N}(\hat{x}_n, \hat{\sigma}^2_n)$;
(iii) the conditional distribution of Z_n given $(Z_{n-1}, \dots, Z_1)$ is the law $\mathbb{N}(\bar{x}_{n-1}, \bar{\sigma}^2_{n-1})$,

with

$$v_n = \frac{\hat{\sigma}^2_n}{\gamma^2} \quad v_1 = \frac{\tau^2}{\tau^2 + \gamma^2}, \tag{26}$$

$$\hat{x}_{n+1} = a\hat{x}_n + (Z_{n+1} - a\hat{x}_n)v_{n+1} \quad \hat{x}_0 = 0, \tag{27}$$

$$v_{n+1} = f(v_n) \quad \text{with} \quad f(v) = 1 - \left(1 + \frac{\beta^2}{\gamma^2} + a^2 v\right)^{-1}, \tag{28}$$

$$x_{n-1} = a\hat{x}_{n-1} = \bar{x}_{n-1} \quad \sigma^2_{n-1} = \beta^2 + a^2\gamma^2 v_{n-1}, \tag{29}$$

$$\bar{\sigma}^2_{n-1} = \sigma^2_{n-1} + \gamma^2 \quad \bar{\sigma}^2_0 = \tau^2 + \gamma^2. \tag{30}$$

With the above notations, the distribution of Z_1 is the law $\mathbb{N}(\bar{x}_0, \bar{\sigma}^2_0)$ and the exact likelihood of $(Z_1, \dots, Z_n)$ is (up to a constant) equal to

$$p_n(\theta) = p_n(\theta, Z_1, \dots, Z_n) = (\bar{\sigma}_0 \dots \bar{\sigma}_{n-1})^{-1} \exp\left(-\sum_{i=1}^{n} \frac{(Z_i - \bar{x}_{i-1})^2}{2\bar{\sigma}^2_{i-1}}\right), \tag{31}$$

where $\theta = (a, \beta^2)$, $\bar{x}_i = \bar{x}_i(\theta)$ and $\bar{\sigma}_i = \bar{\sigma}_i(\theta)$.

The conditional distribution of Z_1 given $(Z_0, \dots, Z_{-n+2})$ is obtained substituting $(Z_{n-1}, \dots, Z_1)$ by $(Z_0, \dots, Z_{-n+2})$ in the law $\mathbb{N}(\bar{x}_{n-1}, \bar{\sigma}^2_{n-1})$. This sequence of distributions converges weakly to the conditional distribution of Z_1 given $\underline{Z}_0$. Indeed, the deterministic recurrence equation for (v_n) converges to a limit $v(\theta) \in (0, 1)$ with exponential rate [see (28)]. Using the above equations, we find after some computations that the conditional distribution of Z_1 given $\underline{Z}_0$ is the law $\mathbb{N}(\bar{x}(\theta, \underline{Z}_0), \bar{\sigma}^2(\theta))$ with

$$\bar{x}(\theta, \underline{Z}_0) = aE_\theta(U_0/\underline{Z}_0) = av(\theta)\sum_{i=0}^{\infty} a^i(1 - v(\theta))^i Z_{-i} \quad \text{and} \quad \bar{\sigma}^2(\theta) = a\gamma^2 v + \beta^2 + \gamma^2.$$

Therefore, the conditional likelihood is (up to a constant) explicitely given by

$$\tilde{p}_n(\theta, \underline{Z}_n) = \bar{\sigma}(\theta)^{-n} \exp\left(-\sum_{i=1}^{n} \frac{(Z_i - \bar{x}(\theta, \underline{Z}_{i-1}))^2}{2\bar{\sigma}(\theta)^2}\right). \tag{32}$$

Since the series $\bar{x}(\theta, \underline{Z}_0)$ converges in $L^2(\mathbb{P}_{\theta_0})$, this function is well defined for all θ under $\mathbb{P}_{\theta_0}$. Moreover, the assumptions of Proposition 3 are satisfied, and the limit is

$$E_{\theta_0} l(\theta, \underline{Z}_1) = -\frac{1}{2}\left\{\log \bar{\sigma}(\theta)^2 + \frac{1}{\bar{\sigma}(\theta)^2} E_{\theta_0}(Z_1 - \bar{x}(\theta, \underline{Z}_0))^2\right\}. \tag{33}$$

3.2. Associated contrasts

The conditional likelihood cannot be used directly, since we do not observe $\underline{Z}_0$. However, the conditional likelihood suggests a family of appropriate contrasts to build consistent estimators.

As mentioned earlier, our approach is motivated by the estimation methods used for discrete time ARCH models (see Engle, 1982). In these models, the exact likelihood is untractable, whereas the conditional likelihood is explicit and simple. Moreover, the common feature of these series is that they usually do not have even second-order moments. This rules out any standard estimation method based on moments. Elie & Jeantheau (1995) and Jeantheau (1998) have proved an appropriate theorem to deal with this difficulty. We also use this theorem later and recall it. For the sake of clarity, its proof is given in the appendix.

3.2.1. A general result of consistency

For a given known sequence $\underline{z}_0$ of $\mathbb{R}^{\mathbb{N}}$, we define the random vector of $\mathbb{R}^{\mathbb{N}}$

$$\underline{Z}_n(\underline{z}_0) = (Z_n, \ldots, Z_1, \underline{z}_0).$$

Let f be a real function defined on $\Theta \times \mathbb{R}^{\mathbb{N}}$ and set

$$F_n(\theta, \underline{z}_n) = n^{-1} \sum_{i=1}^{n} f(\theta, \underline{z}_i). \tag{34}$$

Let us introduce the random variable

$$\theta_n^* = \arg\inf_{\theta} F_n(\theta, \underline{Z}_n) = \theta_n^*(\underline{Z}_n). \tag{35}$$

Now, we introduce the estimator defined by the equation

$$\tilde{\theta}_n(\underline{z}_0) = \arg\inf_{\theta} F_n(\theta, \underline{Z}_n(\underline{z}_0)) = \theta_n^*(\underline{Z}_n(\underline{z}_0)). \tag{36}$$

The estimator $\tilde{\theta}_n(\underline{z}_0)$ is a function of the observations $(Z_1, \ldots, Z_n)$, but also depends on f and $\underline{z}_0$. Theorem 2 gives conditions on f and $\underline{z}_0$ to obtain strong consistency for this type of estimators.

We have in mind the case of $f = -\log l$ [see (25)] so that $F_n(\theta, \underline{Z}_n) = -\log \tilde{p}_n(\theta)$. Nevertheless, other functions of f could be used.

As usual, let θ_0 be the true value of the parameter and consider the following conditions:

- **C0** Θ is compact.
- **C1** The function f is such that
  (i) For all n, $f(\theta, \underline{Z}_n)$ is measurable on $\Theta \times \Omega$ and continuous in θ, $\mathbb{P}_{\theta_0}$ a.s.
  (ii) Let $B(\theta, \rho)$ be the open ball of centre θ and radius ρ, and set for $i \in \mathbb{Z}$, $f_*(\theta, \rho, \underline{Z}_i) = \inf\{f(\theta', \underline{Z}_i), \theta' \in B(\theta, \rho) \cap \Theta\}$. Then, $\forall \theta \in \Theta$, $E_{\theta_0}(f_*^-(\theta, \rho, \underline{Z}_1)) > -\infty$ [with the notation $a^- = \inf(a, 0)$] .
- **C2** The function $\theta \to F(\theta_0, \theta) = E_{\theta_0}(f(\theta, \underline{Z}_1))$ has a unique (finite) minimum at θ_0.
- **C3** The function f and $\underline{z}_0$ are such that
  (i) For all n, $f(\theta, \underline{Z}_n(\underline{z}_0))$ is measurable on $\Theta \times \Omega$ and continuous in θ, $\mathbb{P}_{\theta_0}$ a.s.
  (ii) $f(\theta, \underline{Z}_n) - f(\theta, \underline{Z}_n(\underline{z}_0)) \longrightarrow 0$ as $n \to \infty$, $\mathbb{P}_{\theta_0}$ a.s., uniformly in θ.

Theorem 2

Assume (H1)–(H4). Then, under (C0)–(C2), the random variable θ_n^* *defined in (35) converges* $\mathbb{P}_{\theta_0}$ *a.s. to* θ_0 *when* $n \to \infty$. *Under (C0)–(C3),* $\tilde{\theta}_n(\underline{z}_0)$ *converges* $\mathbb{P}_{\theta_0}$ *a.s. to* θ_0.

Let us make some comments on Theorem 2. It holds not only for HMMs, but for any strictly stationary and ergodic process under condition (C0)–(C3). It is an extension of a result of Pfanzagl (1969) proved for i.i.d. data and for the exact likelihood. Its main interest is to obtain strongly consistent estimators under weaker assumptions than the classical ones. In particular, (C1) and (C2) are weak moment and regularity conditions. Condition (C3) appears as the most difficult one. We give below examples where these conditions may be checked. Let us stress that in econometric literature, (C3) is generally not checked and only θ_n^* is considered and treated as the standard estimator. Note also that (C3) may be weakened into

$$\frac{1}{n}\sum_{i=1}^{n}(f_*(\theta,\rho,\underline{Z}_i) - f_*(\theta,\rho,\underline{Z}_i(\underline{z}_0))) \longrightarrow 0 \quad \text{as } n \to \infty,$$

uniformly in θ, in $\mathbb{P}_{\theta_0}$-probability. This leads to a convergence in $\mathbb{P}_{\theta_0}$-probability of $\tilde{\theta}_n(\underline{z}_0)$ to θ_0.

Example 4 (ARCH-type models). Consider observations Z_n given by

$$Z_n = U_n^{1/2}\varepsilon_n \quad \text{and} \quad U_n = \varphi(\theta, \underline{Z}_{n-1}),$$

where (ε_n) is a sequence of i.i.d. random variables with mean 0 and variance 1, and with $\mathscr{I}_n = \sigma(Z_k, k \leqslant n)$, ε_n is $\mathscr{I}_n$-measurable and independent of $\mathscr{I}_{n-1}$. Although U_n is a Markov chain, Z_n is not an HMM in the sense of Definition 1. Still, under appropriate assumptions, Z_n is strictly stationary and ergodic. If the ε_ns are Gaussian, we choose

$$f(\theta, \underline{Z}_1) = \frac{1}{2}\left(\log\varphi(\theta,\underline{Z}_0) + \frac{Z_1^2}{\varphi(\theta,\underline{Z}_0)}\right),$$

which corresponds to the conditional loglikelihood. If the ε_ns are not Gaussian, we may still consider the same function.

To clarify, consider the special case of a GARCH(1,1) model, given by

$$U_n = \omega + \alpha Z_{n-1}^2 + \beta U_{n-1} = \omega + (\alpha\varepsilon_{n-1}^2 + \beta)U_{n-1},$$

where ω, α and β are positive real numbers. It is well known that, if $E(\log(\alpha\varepsilon_1^2 + \beta)) < 0$, there exists a unique stationary and ergodic solution, and, almost surely,

$$U_n = \frac{\omega}{1-\beta} + \alpha\sum_{i\geqslant 1}\beta^{i-1}Z_{n-i}^2.$$

Let us remark that this solution does not even have necessarily a first-order moment. However, it is enough to add the assumption $\omega \geqslant c > 0$ to prove that assumptions (C1)–(C2) hold. Fixing $\underline{Z}_0 = \underline{z}_0$ is equivalent to fixing $U_1 = u_1$, and (C3) can be proved (see Elie & Jeantheau, 1995).

3.2.2. Identifiability assumption

Condition (C2) is an identifiability assumption. The most interesting case is when f directly comes from the conditional likelihood, that is to say

$$f(\theta, \underline{z}_1) = -l(\theta, \underline{z}_1) = -\log\tilde{p}(\theta, z_1/\underline{z}_0).$$

In this case, the limit appearing in (C2) admits a representation in terms of a Kullback–Leibler information. Consider the random probability measure

$$\tilde{P}_\theta(\mathrm{d}z) = \tilde{p}(\theta, z/\underline{Z}_0)\mu(\mathrm{d}z) \tag{37}$$

and the assumptions

- (H5) $\forall \theta_0, \theta$, $E_{\theta_0}|l(\theta, \underline{Z}_1)| < \infty$
- (H6) $\tilde{P}_\theta = \tilde{P}_{\theta_0}, \mathbb{P}_{\theta_0}$ a.s. $\Longrightarrow \theta = \theta_0$.

Set

$$K(\theta_0, \theta) = E_{\theta_0} K(\tilde{P}_{\theta_0}, \tilde{P}_\theta), \tag{38}$$

where $K(\tilde{P}_{\theta_0}, \tilde{P}_\theta)$ denotes the Kullback information of $\tilde{P}_{\theta_0}$ with respect to $\tilde{P}_\theta$.

Lemma 1

Assume (H1)–(H6), we have, almost surely, under $\mathbb{P}_{\theta_0}$,

$$\frac{1}{n}(\log \tilde{p}_n(\theta_0, \underline{Z}_n) - \log \tilde{p}_n(\theta, \underline{Z}_n)) \to K(\theta_0, \theta),$$

and (C2) holds for $f = -l$.

Proof of Lemma 1. Under (H5), the convergence is obtained by the ergodic theorem. Conditioning on $\underline{Z}_0$, we get

$$E_{\theta_0} \log \tilde{p}(\theta, Z_1/\underline{Z}_0) = E_{\theta_0} \int_{\mathbb{R}} \log \tilde{p}(\theta, z/\underline{Z}_0)\, \mathrm{d}\tilde{P}_{\theta_0}(z).$$

Therefore,

$$E_{\theta_0}[l(\theta, \underline{Z}_1) - l(\theta_0, \underline{Z}_1)] = K(\theta_0, \theta).$$

This quantity is non negative [see (38)] and, by (H6), equal to 0 if and only if $\tilde{P}_\theta = \tilde{P}_{\theta_0}$ $\mathbb{P}_{\theta_0}$ a.s. □

Example 1 (continued) (Kalman filter). Recall that [see (33)] the conditional likelihood is based on the function

$$f(\theta, \underline{Z}_1) = -l(\theta, \underline{Z}_1) = \frac{1}{2}\left\{\log \bar{\sigma}(\theta)^2 + \frac{1}{\bar{\sigma}(\theta)^2}(Z_1 - \bar{x}(\theta, \underline{Z}_0))^2\right\}. \tag{39}$$

Condition (C1) is immediate. To check (C2), we use Lemma 1. Assumption (H5) is also immediate. Let us check (H6). We have seen that

$$\tilde{P}_\theta = \mathbb{N}(\bar{x}(\theta, \underline{Z}_0), \bar{\sigma}^2(\theta)).$$

Thus,

$$\tilde{P}_\theta = \tilde{P}_{\theta_0} \quad \mathbb{P}_{\theta_0} \text{ a.s.}$$

is equivalent to

$$\bar{x}(\theta, \underline{Z}_0) = \bar{x}(\theta_0, \underline{Z}_0)\ \mathbb{P}_{\theta_0} \text{ a.s.} \quad \text{and} \quad \bar{\sigma}(\theta)^2 = \bar{\sigma}(\theta_0)^2. \tag{40}$$

The first equality in (40) writes $\sum_{i \geqslant 0} \lambda_i Z_i = 0$ $\mathbb{P}_{\theta_0}$ a.s. with

$$\lambda_i = av(\theta)a^i(1 - v(\theta))^i - a_0 v(\theta_0)a_0^i(1 - v(\theta_0))^i.$$

Suppose that the λ_is are not all equal to 0. Denote by i_0 the smallest integer such that $\lambda_{i_0} \neq 0$. Then, Z_{i_0} becomes a deterministic function of its infinite past. This is impossible. Hence, for all

i, $\lambda_i = 0$. Thus, we get $av(\theta) = a_0 v(\theta_0)$ and $a(1 - v(\theta)) = a_0(1 - v(\theta_0))$. Since $0 < v(\theta), v(\theta_0) < 1$, we get $a = a_0$ and $v(\theta) = v(\theta_0)$. The second equality in (40) yields $\beta = \beta_0$.

3.2.3. Stability assumption

Condition (C3) can be viewed as a stability assumption, since it states an asymptotic forgetting of the past. But, here, the stability condition has only to be checked on the specific function f and point $\underline{z}_0$ chosen to build the estimator. This holds for Example 4 (ARCH-type models). We can also check it for the Kalman filter.

Example 1 (continued) (Kalman filter). Let us prove that (C3) holds with $\underline{z}_0 = (0, 0, \ldots) = \underline{0}$ and f given by (39). Note that for abirtrary $\underline{z}_0$ in $\mathbb{R}^{\mathbb{N}}$, $\bar{x}(\theta, \underline{z}_0)$ may be undefined. For $\underline{z}_0 = \underline{0}$, we have

$$f(\theta, \underline{Z}_n) - f(\theta, \underline{Z}_n(\underline{0})) = \frac{1}{2\bar{\sigma}(\theta)^2}(\bar{x}(\theta, \underline{Z}_{n-1}) - \bar{x}(\theta, \underline{Z}_{n-1}(\underline{0})))(2Z_n - \bar{x}(\theta, \underline{Z}_{n-1}) - \bar{x}(\theta, \underline{Z}_{n-1}(\underline{0}))).$$

Now,

$$\bar{x}(\theta, \underline{Z}_{n-1}) - \bar{x}(\theta, \underline{Z}_{n-1}(\underline{0})) = a^{n-1}(1 - v(\theta))^{n-1} x(\theta, \underline{Z}_0).$$

Using that Θ is compact, we easily deduce (C3) for this example.

Remark

To conclude, let us stress that it is possible to compare the exact m.l.e. and the minimum contrast estimator $\tilde{\theta}_n(\underline{0})$ in the Kalman filter example. Indeed, (Z_n) is a stationary ARMA(1,1) Gaussian process. The exact likelihood requires the knowledge of $Z^{\mathrm{t}}\Sigma_{1,n}^{-1}Z$ where $\Sigma_{1,n}$ is the covariance matrix of $(Z_1, \ldots, Z_n)$. To avoid the difficult computation of $\Sigma_{1,n}^{-1}$, two approximations are classical. The first one is the Whittle approximation which consists in computing $\tilde{Z}^{\mathrm{t}}\Sigma_{-\infty,\infty}^{-1}\tilde{Z}$, where $\Sigma_{-\infty,\infty}$ is the covariance matrix of the infinite vector $(Z_n, n \in \mathbb{Z})$ and $\tilde{Z} = (\ldots, 0, 0, Z_1, \ldots, Z_n, 0, 0, \ldots)$. The second one is the case described here. It corresponds to computing $\tilde{Z}^{\mathrm{t}}\Sigma_{-\infty,0}^{-1}\tilde{Z}$ with $\tilde{Z} = \underline{Z}_n(\underline{0}) = (\ldots, 0, 0, Z_1, \ldots, Z_n)$. It is well known that the three estimators are asymptotically equivalent. It is also classical to use the previous estimators even for non-Gaussian stationary processes (for details, see Beran, 1995).

4. Stochastic volatility models

In this section, we give more details on Example 3, in the case where the volatility V_t is a strictly stationary diffusion process.

4.1. Model and assumptions

We consider for $t \in \mathbb{R}$, (Y_t, V_t) defined by, for $s \leqslant t$

$$Y_t - Y_s = \int_s^t \sigma_u \, \mathrm{d}B_u, \tag{41}$$

$$V_t = \sigma_t^2 \quad \text{and} \quad V_t - V_s = \int_s^t b(\theta, V_u) \, \mathrm{d}u + \int_s^t a(\theta, V_u) \, \mathrm{d}W_u. \tag{42}$$

For positive Δ, we observe a discrete sampling $(Y_{i\Delta}, i = 1, \ldots, n)$ of (41) and the problem is to estimate the unknown $\theta \in \Theta \subset \mathbb{R}^d$ of (42) from this observation.

We assume that

- (A0) $(B_t, W_t)_{t\in\mathbb{R}}$ is a standard Brownian motion of $\mathbb{R}^2$ defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$.

Equation (42) defines a one-dimensional diffusion process indexed by $t \in \mathbb{R}$. We make now the standard assumptions on functions $b(\theta, u)$ and $a(\theta, u)$ ensuring that (42) admits a unique strictly stationary and ergodic solution with state space (l, r) included is $(0, \infty)$.

- (A1) For all $\theta \in \Theta, b(\theta, v)$ and $a(\theta, v)$ are continuous (in v) real functions on $\mathbb{R}$, and C^1 functions on (l, r) such that

$$\exists k > 0, \quad \forall v \in (l, r), \quad b^2(\theta, v) + a^2(\theta, v) \leqslant k(1 + v^2) \quad \text{and} \quad \forall v \in (l, r), \quad a(\theta, v) > 0.$$

For $v_0 \in (l, r)$, define the derivative of the scale function of diffusion (V_t),

$$s(\theta, v) = \exp\left(-2\int_{v_0}^{v} \frac{b(\theta, u)}{a^2(\theta, u)} \mathrm{d}u\right). \tag{43}$$

- (A2) For all $\theta \in \Theta$,

$$\int_{l+} s(\theta, v)\,\mathrm{d}v = +\infty, \qquad \int^{r-} s(\theta, v)\,\mathrm{d}v = +\infty, \qquad \int_{l}^{r} \frac{\mathrm{d}v}{a^2(\theta, v)s(\theta, v)} = M_\theta < +\infty.$$

Under (A0)–(A2), the marginal distribution of (V_t) is $\pi_\theta(\mathrm{d}v) = \pi(\theta, v)\mathrm{d}v$, with

$$\pi(\theta, v) = \frac{1}{M_\theta} \frac{1}{a^2(\theta, v)s(\theta, v)} 1_{(v\in(l,r))}. \tag{44}$$

In order to study the conditional likelihood, we consider the additional assumptions

- (A3) For all $\theta \in \Theta$, $\int_l^r v\pi_\theta(dv) < \infty$.
- (A4) $0 < l < r < \infty$.

Let us stress that (A3) is a weak moment condition. The condition $l > 0$ is crucial. Intuitively, it is natural to consider volatilities bounded away from 0 in order to estimate their parameters from the observation of (Y_t).

Let $C = C(\mathbb{R}, \mathbb{R}^2)$ be the space of continuous functions on $\mathbb{R}$ and $\mathbb{R}^2$-valued, equipped with the Borel σ-field $\mathscr{C}$ associated with the uniform topology on each compact subset of $\mathbb{R}$. We shall assume that (Y_t, V_t) is the canonical diffusion solution of (41) and (42) on $(C, \mathscr{C})$, and we keep the notation P_θ for its distribution. For given positive Δ, we observe $(Y_{i\Delta} - Y_{(i-1)\Delta}, i = 1, \ldots, n)$ and we define (Z_i) as in (3). As recalled in Example 3, (Z_i) is an HMM with hidden chain $U_n = (\overline{V_n}, V_{n\Delta})$.

Setting $t = (\overline{v}, v) \in (l, r)^2$, the conditional distribution of Z_i given $U_i = t$ is the Gaussian law $\mathbb{N}(0, \overline{v})$, so that $\mu(\mathrm{d}z) = \mathrm{d}z$ is the Lebesgue measure on $\mathbb{R}$ and

$$f(z/t) = f(z/\overline{v}) = \frac{1}{(2\pi\overline{v})^{1/2}} \exp\left(-\frac{z^2}{2\overline{v}}\right). \tag{45}$$

For the transition density of the hidden chain $U_i = (\overline{V}_i, V_{i\Delta})$, it is natural to have, as dominating measure, the Lebesgue measure

$$\nu(\mathrm{d}t) = 1_{(l,r)^2}(\overline{v}, v)\mathrm{d}\overline{v}\,\mathrm{d}v. \tag{46}$$

Actually, it amounts to proving that the two-dimensional diffusion $(\int_0^t V_s\,\mathrm{d}s, V_t)$ admits a transition density. Two points of view are possible. In some models, a direct proof may be feasible. Otherwise, this will be ensured under additional regularity assumptions on functions

$a(\theta, .)$ and $b(\theta, .)$ (see, e.g. Gloter, 2000). For sake of clarity, we introduce the following assumption:

- (A5) The transition probability distribution of the chain (U_i) is given by

$$p(\theta, u, t)\nu(\mathrm{d}t), \tag{47}$$

where $(u, t) \in (l, r)^2 \times (l, r)^2$ and $\nu(\mathrm{d}t)$ is the Lebesgue measure (46).

This has several interesting consequences. First, note that this transition density has a special form. Setting $u = (\bar{a}, a) \in (l, r)^2$,

$$p(\theta, u, t) = p(\theta, a, t) \tag{48}$$

only depends on a and is equal to the conditional density of $U_1 = (\overline{V}_1, V_\Delta)$ given $V_0 = a$. Therefore, the (unconditional) density of U_1 is (with $t = (\bar{v}, v)$)

$$g(\theta, t) = \int_l^r p(\theta, a, t)\pi(\theta, a)\,\mathrm{d}a, \tag{49}$$

where $\pi(\theta, a)\,\mathrm{d}a$, defined in (44), is the stationary distribution of the hidden diffusion (V_t). Of course, $g(\theta, t)$ is the stationary density of the chain (U_i). The densities of $\overline{V}_1$ and Z_1 are, therefore [see (45)],

$$\bar{\pi}(\theta, \bar{v}) = \int_l^r g(\theta, (\bar{v}, v))\,\mathrm{d}v, \tag{50}$$

$$p_1(\theta, z_1) = \int_l^r f(z_1/\bar{v})\bar{\pi}(\theta, \bar{v})\,\mathrm{d}\bar{v}. \tag{51}$$

Second, the conditional distribution of $\overline{V}_i$ given $Z_{i-1} = z_{i-1}, \ldots, Z_1 = z_1$ has a density with respect to the Lebesgue measure on (l, r), say

$$\bar{\pi}_i(\theta, \bar{v}_i/z_{i-1}, \ldots, z_1). \tag{52}$$

So, applying Proposition 2, we can integrate with respect to the second coordinate $V_{i\Delta}$ of $U_i = (\overline{V}_i, V_{i\Delta})$ to obtain that the conditional density of Z_i given $Z_{i-1} = z_{i-1}, \ldots, Z_1 = z_1$, is equal to

$$t_i(\theta, z_i/z_{i-1}, \ldots, z_1) = \int_l^r \bar{\pi}_i(\theta, \bar{v}_i/z_{i-1}, \ldots, z_1)f(z_i/\bar{v}_i)\,\mathrm{d}\bar{v}_i \tag{53}$$

for all $i \geqslant 2$ [see (9)]. Therefore, (53) is a variance mixture of Gaussian distributions, the mixing distribution being $\bar{\pi}_i(\theta, \bar{v}_i/z_{i-1}, \ldots, z_1)\,\mathrm{d}\bar{v}_i$.

Let us establish some links between the likelihood and a contrast previously used in the case of the small sampling interval (see Genon-Catalot *et al.*, 1999). The contrast method is based on the property that the random variables (Z_i) behave asymptotically (as $\Delta = \Delta_n$ goes to zero) as a sample of the distribution

$$q(\theta, z) = \int_l^r \pi(\theta, v)f(z/v)\,\mathrm{d}v.$$

The same property of variance mixture of Gaussian distributions appears in (53), with a change of mixing distribution [see also (54)].

4.2. Conditional likelihood

Applying Theorem 1 and integrating with respect to the second coordinate V_Δ of $U_1 = (\overline{V}_1, V_\Delta)$, we obtain the following proposition:

Proposition 4
Assume (A0)–(A2) and (A5). Then, under $\mathbb{P}_\theta$:

(1) *The conditional distribution of $\overline{V}_1$ given $\underline{Z}_0 = \underline{z}_0$ admits a density with respect to the Lebesgue measure on (l, r), which we denote by $\tilde{\pi}_\theta(\overline{v}/\underline{z}_0)$.*
(2) *The conditional distribution of Z_1 given $\underline{Z}_0 = \underline{z}_0$ has the density*

$$\tilde{p}(\theta, z_1/\underline{z}_0) = \int_l^r \tilde{\pi}_\theta(\overline{v}/\underline{z}_0) f(z_1/\overline{v})\, \mathrm{d}\overline{v}. \tag{54}$$

Hence, the conditional likelihood of $(Z_1, \ldots, Z_n)$ given $\underline{Z}_0$ is given by

$$\tilde{p}_n(\theta, \underline{Z}_n) = \prod_{i=1}^n \tilde{p}(\theta, Z_i/\underline{Z}_{i-1}).$$

Therefore, the distribution given by (54) is a variance mixture of Gaussian distributions, the mixing distribution being now the conditional distribution $\tilde{\pi}_\theta(\overline{v}/\underline{Z}_0)\,\mathrm{d}\overline{v}$ of $\overline{V}_1$ given $\underline{Z}_0$ [compare with (53)].

In accordance with Definition 2, let us assume that, $\mathbb{P}_{\theta_0}$ a.s., the function $\tilde{\pi}_\theta(\overline{v}/\underline{Z}_0)$ is well defined for all θ and is a probability density on (l, r). We keep the following notations:

$$f(\theta, \underline{Z}_n) = -\log \tilde{p}(\theta, Z_n/\underline{Z}_{n-1})) \quad \text{and} \quad \tilde{P}_\theta(\mathrm{d}z) = \tilde{p}(\theta, z/\underline{Z}_0)\,\mathrm{d}z.$$

Then, we have

Proposition 5
Under (A0)–(A5):

(1) $\forall \theta_0, \theta$, $E_{\theta_0}|f(\theta, \underline{Z}_1)| < \infty$.
(2) *We have, almost surely, under $\mathbb{P}_{\theta_0}$,*

$$\frac{1}{n}(\log \tilde{p}_n(\theta_0, \underline{Z}_n) - \log \tilde{p}_n(\theta, \underline{Z}_n)) \to K(\theta_0, \theta) = E_{\theta_0} K(\tilde{P}_{\theta_0}, \tilde{P}_\theta),$$

see (38).

Proof of Proposition 5. Using (A4) and the fact that $\tilde{\pi}_\theta(\overline{v}/\underline{Z}_0)$ is a probability density over (l, r), we get

$$\frac{1}{\sqrt{2\pi r}} \exp -\frac{z_1^2}{2l} \leqslant \tilde{p}(\theta, z_1/\underline{Z}_0)) \leqslant \frac{1}{\sqrt{2\pi l}}. \tag{55}$$

So, for some constant C (independent of θ, involving only the boundaries l, r), we have

$$|f(\theta, \underline{Z}_1)| \leqslant C(1 + Z_1^2). \tag{56}$$

By (A3), $\mathbb{E}_{\theta_0} Z_1^2 = \mathbb{E}_{\theta_0} V_0 < \infty$. Therefore, we get the first part. The second follows from the ergodic theorem. □

So, we have checked (H5). We do not know how to check the identifiability assumption (H6). However, in statistical problems for which the identifiability assumption contains randomness, this assumption can rarely be verified. Hence, if we know that regularity condition (C1) holds, we get that θ_n^* converges a.s. to θ_0. Condition (C3) remains to be checked (see, in this respect, our comments after Theorem 2).

To conclude, the above results on the SV models are of theoretical nature but clarify the difficulties of the statistical inference in this model and enlight the set of minimal assumptions.

4.3. Mean reverting hidden diffusion

In the SV models, we cannot have explicit expressions for the conditional densities. Therefore, we must use other functions $f(\theta, \underline{Z}_n)$ to build estimators. To illustrate this, let us consider mean-reverting volatilities, that is, models of the form

$$dV_t = \alpha(\beta - V_t)\,dt + a(V_t)\,dW_t, \tag{57}$$

where $\alpha > 0$, $\beta > 0$ and $a(V_t)$ may also depend on unknown parameters. Due to the mean reverting drift, these models present some special features. In particular, many authors have remarked that the covariance structure of the process $(\overline{V_i})$ is simple (see, e.g. Genon-Catalot *et al.*, 2000a; Sørensen, 2000).

Assume that the above hidden diffusion (V_t) satisfies (A1) and (A2), and that $\mathbb{E}V_0^2$ is finite. Then, $\mathbb{E}\overline{V_1} = \mathbb{E}V_0 = \beta$,

$$\mathbb{E}\overline{V_1}^2 = \beta^2 + \mathrm{Var}(V_0)\frac{2(\alpha\Delta - 1 + e^{-\alpha\Delta})}{\alpha^2\Delta^2} \tag{58}$$

and for $k \geqslant 1$,

$$\mathbb{E}\overline{V_1}\,\overline{V_{k+1}} = \beta^2 + \mathrm{Var}(V_0)\frac{(1 - e^{-\alpha\Delta})^2}{\alpha^2\Delta^2}e^{-\alpha(k-1)\Delta}. \tag{59}$$

The previous formulae allow to compute the covariance function of $(Z_i^2, i \geqslant 1)$.

Proposition 6

Assume (A0)–(A2) and that $\mathbb{E}V_0^2$ *is finite. Then, the process defined for* $i \geqslant 1$ *by*

$$X_i = Z_{i+1}^2 - \beta - e^{-\alpha\Delta}(Z_i^2 - \beta) \tag{60}$$

satisfies, for $j \geqslant 2$, $\mathrm{Cov}(X_i, X_{i+j}) = 0$. *Hence,* $((Z_i^2 - \beta), i \geqslant 1)$ *is centred and* ARMA(1,1).

Proof of proposition 6. The process $(Z_i^2, i \geqslant 1)$ is strictly stationary and ergodic. Straightforward computations lead to $\mathbb{E}Z_1^2 = \beta$,

$$\mathrm{Var}(Z_1^2) = 2\mathbb{E}\overline{V_1}^2 + \mathrm{Var}(\overline{V_1}) \tag{61}$$

and for $j \geqslant 1$,

$$\mathrm{Cov}(Z_1^2, Z_{1+j}^2) = \mathrm{Cov}(\overline{V_1}, \overline{V_{1+j}}). \tag{62}$$

Then, the computation of $\mathrm{Cov}(X_i, X_{i+j})$ easily follows from (58)–(60). □

Estimation by the Whittle approximation of the likelihood is, therefore, feasible as suggested by Barndorff-Nielsen & Shephard (2001). To apply our method, as in the Kalman filter, we can use the linear projection of $Z_1^2 - \beta$ on its infinite past to build a Gaussian conditional likelihood. To be more specific, let us set $\theta = (\alpha, \beta, c^2)$ with $c^2 = \mathrm{Var}V_0$ and define, under $\mathbb{P}_\theta$,

$$\gamma_\theta(0) = \mathrm{Var}_\theta X_i, \quad \gamma_\theta(1) = \mathrm{Cov}_\theta(X_i, X_{i+1}).$$

Straightforward computations show that $\gamma_\theta(1) < 0$. The $L^2(\mathbb{P}_\theta)$-projection of $Z_1^2 - \beta$ on the linear space spanned by $(Z_{-i}^2 - \beta, i \geqslant 0)$ has the following form:

$$Z_1^2 - \beta = \sum_{i \geqslant 0} \alpha_i(\theta)(Z_{-i}^2 - \beta) + U(\theta, \underline{Z}_1),$$

where, for all $i \geqslant 0$,

$$\mathbb{E}_\theta((Z_{-i}^2 - \beta)U(\theta, \underline{Z}_1)) = 0 \tag{63}$$

and the one-step prediction error is

$$\sigma^2(\theta) = \mathbb{E}_\theta(U^2(\theta, \underline{Z}_1)). \tag{64}$$

The coefficients $(\alpha_i(\theta), i \geqslant 0)$ can be computed using $\gamma_\theta(0)$ and $\gamma_\theta(1)$ as follows. By the canonical representation of (X_i) as an MA(1)-process, we have,

$$X_i = W_{i+1} - b(\theta)W_i,$$

where (W_i) is a centred white noise (in the wide sense), the so-called innovation process. It is such that $|b(\theta)| < 1$ and

$$\sigma^2(\theta) = \mathrm{Var}_\theta W_i.$$

Therefore, the spectral density $f_\theta(\lambda)$ satisfies

$$f_\theta(\lambda) = \gamma_\theta(0) + 2\gamma_\theta(1)\cos(\lambda) = \sigma^2(\theta)(1 + b^2(\theta) - 2b(\theta)\cos(\lambda)).$$

Since, for all λ, $f_\theta(\lambda) > 0$, we get that $\gamma_\theta^2(0) - 4\gamma_\theta^2(1) > 0$. Now, using that $\gamma_\theta(1) < 0$, the following holds

$$b(\theta) = \frac{\gamma_\theta(0) - \left(\gamma_\theta^2(0) - 4\gamma_\theta^2(1)\right)^{1/2}}{-2\gamma_\theta(1)}, \quad \sigma^2(\theta) = \frac{-\gamma_\theta(1)}{b(\theta)}.$$

Then, setting $a(\theta) = \exp(-\alpha\Delta)$, $v(\theta) = 1 - [b(\theta)/a(\theta)]$,

$$\alpha_i(\theta) = a(\theta)v(\theta)(a(\theta)(1 - v(\theta)))^i. \tag{65}$$

Now, we can define

$$f(\theta, \underline{Z}_1) = \log\sigma^2(\theta) + \frac{1}{\sigma^2(\theta)}\left(Z_1^2 - \beta - \sum_{i\geqslant 0}\alpha_i(\theta)(Z_{-i}^2 - \beta)\right)^2.$$

Easy computations using (63)–(65) yield that

$$\begin{aligned}\mathbb{E}_{\theta_0}(f(\theta, \underline{Z}_1) - f(\theta_0, \underline{Z}_1)) = {} & \frac{\sigma^2(\theta_0)}{\sigma^2(\theta)} - 1 - \log\frac{\sigma^2(\theta_0)}{\sigma^2(\theta)} + \frac{(\beta_0 - \beta)^2}{\sigma^2(\theta)}\left(\frac{1 - a(\theta)}{1 - b(\theta)}\right)^2 \\ & + \frac{1}{\sigma^2(\theta)}\mathbb{E}_{\theta_0}\left(\sum_{i\geqslant 0}(\alpha_i(\theta_0) - \alpha_i(\theta))(Z_{-i}^2 - \beta_0)\right)^2.\end{aligned}$$

Hence, all the conditions of Theorem 2 may be checked and θ can be identified by this method.

5. Concluding remarks

The conditional likelihood method is classical in the field of ARCH-type models. In this paper, we have shown that it can be used for HMMs, and in particular for SV models. The approach is theoretical but enlightens the minimal assumptions needed for statistical inference. From this point of view, these assumptions do not require the existence of high-order moments for the hidden Markov process. This is consistent with financial data that usually exhibit fat tailed marginals.

In order to illustrate on an explicit example the conditional likelihood method, we revisit in full detail the Kalman filter. SV models with mean-reverting volatility provide another example where the method can be used.

This method may be applied to other classes of models for financial data: models including leverage effects (see, e.g. Tauchen *et al.* 1996); complete models with SV (see, e.g. Hobson & Rogers, 1998; Jeantheau, 2002).

6. Acknowledgements

The authors were partially supported by the research training network 'Statistical Methods for Dynamical Stochastics Models' under the Improving Human Potential programme, financed by the Fifth Framework Programme of the European Commission. The authors would like to thank the referees and the associate editor for helpful comments.

References

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. Roy. Statist. Soc. Ser. B* **63**, 167–241.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. Ser. B* **57**, 659–672.

Bickel, P. J. & Ritov, Y. (1996). Inference in hidden Markov models I: local asymptotic normality in the stationary case. *Bernoulli* **2**, 199–228.

Bickel, P. J., Ritov, Y. & Ryden, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614–1635.

Del Moral, P., Jacod, J. & Protter, Ph. (2001). The Monte-Carlo method for filtering with discrete time observations. *Probab. Theory Related Fields* **120**, 346–368.

Del Moral, P. & Miclo, L. (2000). Branching and interacting particle systems approximation of Feynman-Kac formulae with applications to non linear filtering. In *Séminaire de probabilités XXXIV* (eds J. Azéma, M. Emery, M. Ledoux & M. Yor). Lecture Notes in Mathematics, vol. **1729**, Springer-Verlag, Berlin, pp. 1–145.

Douc, P. & Matias, L. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7**, 381–420.

Durbin, J. & Koopman, S. J. (1998). Monte-Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**, 669–684.

Eberlein, O., Chib, S. & Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica* **69**, 959–993.

Elie, L. & Jeantheau, T. (1995). Consistance dans les modèles hétéroscédastiques. *C.R. Acad. Sci. Paris, Sér. I* **320**, 1255–1258.

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

Francq, C. & Roussignol, M. (1997). On white noises driven by hidden Markov chains. *J. Time. Ser. Anal.* **18**, 553–578.

Gallant, A. R., Hsieh, D. & Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *J. Econometrics* **81**, 159–192.

Genon-Catalot, V., Jeantheau, T. & Larédo, C. (1998). Limit theorems for discretely observed stochastic volatility models. *Bernoulli* **4**, 283–303.

Genon-Catalot, V., Jeantheau, T. & Larédo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5**, 855–872.

Genon-Catalot, V., Jeantheau, T. & Larédo, C. (2000a). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051–1079.

Genon-Catalot, V., Jeantheau, T. & Larédo, C. (2000b). Consistency of conditional likelihood estimators for stochastic volatility models. Prépublication Marne la Vallée, 03/2000

Gloter, A. (2000). Estimation des paramètres d'une diffusion cachée. PhD Thesis, Université de Marne-la-Vallée.

Gourieroux, C., Monfort, A. & Renault, E. (1993). Indirect inference. *J. Appl. Econometrics.* **8**, S85–S118.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.

Hobson, D. G. & Rogers, L. C. G. (1998). Complete models with stochastic volatility. *Math Finance* **8**, 27–48.

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

Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**, 70–86.

Jeantheau, T. (2002). A link between complete models with stochastic volatility and ARCH models. *Finance Stoch.*, to appear.

Jensen, J. L. & Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27**, 514–535.
Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH-models. *Rev. Econom. Stud.* **65**, 361–394.
Leroux, B. G. (1992). Maximum likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* **40**, 127–143.
Lipster, R. S. & Shiryaev, A. N. (1978). *Statistics of random processes II*. Springer Verlag, Berlin.
Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14**, 249–272.
Pitt, M. K. & Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.* **94**, 590–599.
Sørensen, M. (2000). Prediction-based estimating functions. *Econom. J.* **3**, 121–147.
Tauchen, G., Zhang, H. & Ming, L. (1996). Volume, volatility and leverage. *J. Econometrics* **74**, 177–208.

Received March 2000, in final form July 2002

Catherine Larédo, I.N.R.A., Labratoire de Biométrie, 78350, Jouy-en-Josas, France.
E-mail: Catherine.Laredo@jouy.inra.fr

Appendix

This appendix is devoted to prove Theorem 2 due to Elie & Jeantheau (1995). Recall that we have set $a^+ = \sup(a, 0)$ and $a^- = \inf(a, 0)$, so that $a = a^+ + a^-$. The following proof holds for any strictly stationary and ergodic process under (C0)–(C3).

Proof. First, let us introduce the random variable defined by the equation

$$\theta_n^* = \arg\inf_\theta F_n(\theta, \underline{Z}_n).$$

Note that θ_n^* is not an estimator, since it is a function of the infinite past $(\underline{Z}_n)$. The first part of the proof is devoted to show that, under (H0)–(H2), θ_n^* converges to θ_0 a.s.

By the continuity assumption on f, $f_*(\theta, \rho, \underline{Z}_i)$ is measurable. Moreover, under (C1), $E_{\theta_0}(f^-(\theta), \underline{Z}_1) > -\infty$, therefore $F(\theta_0, \theta)$ is well defined, but may be equal to $+\infty$.

For all $\theta \in \Theta$ and $\theta' \in B(\theta, \rho) \cap \Theta$, the function $\theta' \longrightarrow f(\theta', \underline{Z}_1) - f_*^-(\theta, \rho, \underline{Z}_1)$ is non-negative. Using (C1) and the continuity of f with respect to θ, Fatou's Lemma implies $\liminf_{\theta' \to \theta} F(\theta_0, \theta') \geqslant F(\theta_0, \theta)$. Therefore, F is lower semicontinuous in θ.

Applying the monotone convergence theorem to $f_*^+(\theta, \rho, \underline{Z}_1)$ and the Lebesgue dominated convergence theorem to $f_*^-(\theta, \rho, \underline{Z}_1)$, we get

$$\lim_{\rho \to 0} E_{\theta_0}(f_*(\theta, \rho, \underline{Z}_1)) = E_{\theta_0}(f(\theta, \underline{Z_1})) = F(\theta_0, \theta). \tag{66}$$

Let $\varepsilon > 0$ and consider the compact set $K_\varepsilon = \Theta \cap (B(\theta_0, \varepsilon))^c$. By (C2) and the lower semi-continuity of F, there exists a real $\eta > 0$ such that:

$$\forall \theta \in K_\varepsilon, \quad F(\theta_0, \theta) - F(\theta_0, \theta_0) > \eta. \tag{67}$$

Consider $\theta \in K_\varepsilon$. If $F(\theta_0, \theta) < +\infty$, using (66), there exists $\rho(\theta) > 0$ such that

$$0 \leqslant F(\theta_0, \theta) - E_{\theta_0}(f_*(\theta, \rho(\theta), \underline{Z}_1)) < \eta/2.$$

Combining the above inequality with (67), we obtain

$$E_{\theta_0}(f_*(\theta, \rho(\theta), \underline{Z}_1)) - F(\theta_0, \theta_0) > \eta/2. \tag{68}$$

If $F(\theta_0, \theta) = +\infty$, since $F(\theta_0, \theta_0)$ is finite, using (66), we can also associate $\rho(\theta) > 0$ such that (68) holds. So we cover the compact set K_ε with a finite number L of balls, say $\{B(\theta_k, \rho(\theta_k)), k = 1, \ldots, L\}$.

The ergodic theorem implies that there exists a measurable $N_\varepsilon \subset \Omega$ such that $P_{\theta_0}(N_\varepsilon) = 0$, and for all $\omega \in N_\varepsilon^c$, and for $k = 1, \dots, L$,

$$n^{-1} \sum_{i=1}^{n} f_*(\theta_k, \rho(\theta_k), \underline{Z}_i(\omega)) \longrightarrow E_{\theta_0}(f_*(\theta_k, \rho(\theta_k), \underline{Z}_1)).$$

This holds even if $E_{\theta_0} f_*^+(\theta_k, \rho(\theta_k), \underline{Z}_1) = +\infty$. Since the sequence $(\theta_n^*(\omega))_{n \geqslant 1}$ is in the compact Θ, we can extract, for all $\omega \in N_\varepsilon^c$, a converging subsequence $(\theta_{n_j}^*(\omega))$.

Let us assume that $\theta_{n_j}^*(\omega)$ converges in K_ε. Therefore, it converges in one of the balls, say $B(\theta_1, \rho(\theta_1))$ and we have, for n_j large enough:

$$\frac{1}{n_j} \sum_{i=1}^{n_j} f_*(\theta_1, \rho(\theta_1), \underline{Z}_i(\omega)) \leqslant F_{n_j}(\theta_{n_j}^*(\omega), \underline{Z}_{n_j}(\omega)).$$

But, $F_{n_j}(\theta_{n_j}^*(\omega), \underline{Z}_{n_j}(\omega)) - F_{n_j}(\theta_0, \underline{Z}_{n_j}(\omega)) \leqslant 0$, which implies

$$\frac{1}{n_j} \sum_{i=1}^{n_j} f_*(\theta_1, \rho(\theta_1), \underline{Z}_i(\omega))(\omega) - F_{n_j}(\theta_0, \underline{Z}_{n_j}(\omega)) \leqslant 0.$$

The above term converges as $n_j \to \infty$ to

$$E_{\theta_0}(f_*(\theta_1, \rho(\theta_1), \underline{Z}_1)) - F(\theta_0, \theta_0) \leqslant 0,$$

which is in contradiction with (68).

For $\omega \in \Omega$, denote by $\lambda(\omega)$ the set of limit points of $(\theta_n^*(\omega))$. We have proved that, for all $\omega \in N_\varepsilon^c$, $\lambda(\omega) \subset B(\theta_0, \varepsilon) \cap \Theta$. Now, choose $\varepsilon = 1/n$, and $N = \cup_{n \geqslant 1} N_{1/n}$. Then, N is P_{θ_0}-negligible and

$$\forall \omega \in N^c, \quad \forall n \geqslant 1, \quad \lambda(\omega) \subset B(\theta_0, 1/n) \cap \Theta.$$

Therefore, for all $\omega \in N^c, \lambda(\omega) = \{\theta_0\}$.

Now, we prove the consistency of our estimators using the additional assumption (C3). It is enough to show

$$\frac{1}{n} \sum_{i=1}^{n} (f_*(\theta, \rho, \underline{Z}_i) - f_*(\theta, \rho, \underline{Z}_i(\underline{z}_0))) \overset{P_{\theta_0}\text{a.s.}}{\longrightarrow} 0 \quad \text{when } n \to \infty. \tag{69}$$

Set $D_i = \sup_{\alpha \in \Theta} |f(\alpha, \underline{Z}_i) - f(\alpha, \underline{Z}_i(\underline{z}_0))|$. We have, for all $\theta' \in \Theta$, $f(\theta', \underline{Z}_i) \leqslant f(\theta', \underline{Z}_i(\underline{z}_0)) + D_i$. Thus, for $\rho > 0$,

$$f_*(\theta, \rho, \underline{Z}_i) \leqslant f_*(\theta, \rho, \underline{Z}_i(\underline{z}_0)) + D_i.$$

We easily deduce

$$|f_*(\theta, \rho, \underline{Z}_i) - f_*(\theta, \rho, \underline{Z}_i(\underline{z}_0))| \leqslant D_i.$$

By (C3), D_i converges a.s. to 0, and, using the Cesaro theorem, we get (69). □