Série des Documents de Travail

n° 2017-66
Parametric estimation of hidden Markov models by least squares type estimation and deconvolution

C. CHESNEAU[1]
S. EL KOLEI[2]
F. NAVARRO[3]

[1] Université de Caen - LMNO, France, E-mail: christophe.chesneau@unicaen.fr
[2] ENSAI, Université Bretagne Loire, France, E-mail: salima.el-kolei@ensai.fr
[3] CREST-ENSAI, E-mail : fabien.navarro@ensai.fr

Parametric estimation of hidden Markov models by least squares type estimation and deconvolution

Christophe Chesneau*, Salima El Kolei†, Fabien Navarro‡

September 30, 2017

Abstract

In this paper, we study a specific hidden Markov chain defined by the equation: $Y_i = X_i + \varepsilon_i$, $i = 1, \ldots, n+1$, where $(X_i)_{i\geq 1}$ is a real-valued stationary Markov chain and $(\varepsilon_i)_{i\geq 1}$ is a noise independent of $(X_i)_{i\geq 1}$. We develop a new parametric approach obtained by minimization of a particular contrast taking advantage of the regressive problem and based on deconvolution strategy. We provide theoretical guarantees on the performance of the resulting estimator; its consistency and its asymptotic normality are established.

Keywords: Contrast function; deconvolution; least square estimation; parametric inference; stochastic volatility.

1 Introduction

In this paper, a particular additive hidden Markov model (HMM) is considered; we observe n random variables $Y_1, \ldots, Y_{n+1}$ having the following additive structure:

$$Y_i = X_i + \varepsilon_i, \tag{1}$$

where $(X_i)_{i\geq 1}$ is an unobserved real-valued Markov chain, $(\varepsilon_i)_{i\geq 1}$ is a sequence of independent and identically distributed (*i.i.d.*) random variables and independent of $(X_i)_{i\geq 1}$. Besides its initial distribution, the chain $(X_i)_{i\geq 1}$ is characterized by its transition, i.e. the distribution of X_{i+1} given X_i and by its stationary density f_{θ_0} which we assume unknown. We assume that the transition distribution admits a density Π_{θ_0}, defined by $\Pi_{\theta_0}(x,y)dy = \mathbb{P}_{\theta_0}(X_{i+1} \in dy | X_i = x)$. For the identifiability of (1), we assume that ε_1 admits a known density with respect to the Lebesgue measure denoted by f_ε. Furthermore, we assume that $(X_i)_{i\geq 1}$ is strictly stationary which means that the initial distribution of X_1 is an invariant distribution for the transition kernel Π_{θ_0} of the homogeneous Markov chain $(X_i)_{i\geq 1}$.

*Christophe Chesneau
Université de Caen - LMNO, France, E-mail: christophe.chesneau@unicaen.fr
†Salima El Kolei
ENSAI, Université Bretagne Loire, France, E-mail: salima.el-kolei@ensai.fr
‡Fabien Navarro
CREST, ENSAI, Université Bretagne Loire, France, E-mail: fabien.navarro@ensai.fr

We aim to estimate the vector of parameters θ_0 and suppose that the model is correctly specified: that is, θ_0 belongs to the parameter space $\Theta \subset \mathbb{R}^r$, with $r \in \mathbb{N}^*$.

Many papers are devoted to the case where $(X_i)_{i\geq 1}$ is an autoregressive moving average (ARMA) process (see [2], [18] and [6]). Therefore, all existing known results can be applied for ARMA models. Nevertheless, for more general models, (1) is known as HMM with potentially a non-compact continuous state space. In a Bayesian setting, various results are already stated and most of them are based on Monte Carlo inference (see [8], [1],[3] and [17]). In this paper, we do not consider the Bayesian approach, the model (1) is known in this case as the so-called convolution model. If we focus our attention on (semi)-parametric models, few results exist. The first study which gives a consistent estimator is [5]. The authors propose an estimation procedure based on least squares minimization. Recently, in [7], the authors extend this approach in a general context for models defined as $X_i = b_{\theta_0}(X_{i-1}) + \eta_i$, where b_{θ_0} is the regression function assumed to be known up to θ_0 and for homoscedastic innovations η_i. Also, in [10] the author proposes a consistent estimator for parametric model by assuming the knowledge of the stationary density f_{θ_0} up to θ_0. Nevertheless, for many processes as the class of autoregressive conditional heteroscedastic (ARCH) processes and their extensions, the transition density has an explicit form contrary to the stationary density. So, in this context, it is more appropriate to use the transition density Π_{θ_0} instead of f_{θ_0} in the construction of the estimator.

In this paper, we propose a new estimation approach which provides a consistent estimator with a parametric rate of convergence for more general models. Our approach holds for nonlinear HMMs (1) with heteroscedastic innovations, that is when $X_i = b_{\theta_0}(X_{i-1}) + \sigma_{\theta_0}(X_{i-1})\eta_i$, where σ_{θ_0} corresponds to the heteroscedastic function. Our principle of estimation relies on the procedure proposed by [15] in a non-parametric case to estimate the function Π_{θ_0}. We propose to adapt their approach in a parametric context, assuming that the form of the transition density Π_{θ_0} is known up to some unknown parameter θ_0. Our work is purely parametric but we go further in this direction by proposing an analytical expression of the asymptotic variance matrix $\Sigma(\hat{\theta}_n)$ which allows to construct confidence intervals. The procedure of estimation requires to compute only Fourier transforms of some functions as in [7]. Under general assumptions, we prove that our estimator is consistent. Moreover, we give some conditions under which the asymptotic normality can be stated and provide an analytical expression of the asymptotic variance matrix. These results hold under α-mixing dependency.

The remainder of the paper is organized as follows. Section 2 describes our estimator and its statistical properties. The consistency and the asymptotic normality of the estimator are established in Section 3. The proofs are gathered in Section 4.

2 Procedure: Least squares estimator

Before presenting the main procedure of the study, let us introduce some notations and assumptions which will be useful.

Notations: The Fourier transform of an integrable function u is denoted by $u^*(t) = \int e^{-itx}u(x)dx$. We set $\langle u, v\rangle_f = \int u(x)\overline{v}(x)f_{\theta_0}(x)dx$ with $v\overline{v} = |v|^2$. The norm of the operator T is defined by $||T||_f = \left(\int\int |T(x,y)|^2 f_{\theta_0}(x)dxdy\right)^{1/2}$. Let us recall that, by the properties of the Fourier transform, we have $(u^*)^*(x) = 2\pi u(-x)$ and $\langle u_1, u_2\rangle_f = \frac{1}{2\pi}\langle u_1^*, u_2^*\rangle_f$. We denote by $\nabla_\theta g$ the vector of the partial derivatives of g with respect to (w.r.t) θ. The Hessian matrix of g w.r.t θ is denoted by $\nabla^2_\theta g$. For any matrix $A = (A_{i,j})_{i,j}$, the Frobenius norm is defined by $\|A\| = \sqrt{\sum_i \sum_j |A_{i,j}|^2}$. We set $\mathbf{Y}_i = (Y_i, Y_{i+1})$ and $\mathbf{y}_i = (y_i, y_{i+1})$ is a given realization of $\mathbf{Y}_i$. We set $(t \otimes s)(x, y) = t(x)s(y)$.

In the following, for the sake of conciseness, $\mathbb{P}, \mathbb{E}, \mathbb{V}ar$ and $\mathbb{C}ov$ denote respectively the probability $\mathbb{P}_{\theta_0}$, the expected value $\mathbb{E}_{\theta_0}$, the variance $\mathbb{V}ar_{\theta_0}$ and the covariance $\mathbb{C}ov_{\theta_0}$ when the true parameter is θ_0. Additionally, we write $\mathbf{P}_n$ (*resp.* $\mathbf{P}$) the empirical expectation (*resp.* theoretical), that is, for any stochastic variable $X = (X_i)_i$, $\mathbf{P}_n(X) = \frac{1}{n}\sum_{i=1}^n X_i$ (*resp.* $\mathbf{P}(X) = \mathbb{E}[X]$).

Assumptions:

A 1.

(i) *θ_0 belongs to the interior Θ_0 of a compact set Θ, $\theta_0 \in \Theta \subset \mathbb{R}^r$;*

(ii) *the errors $(\varepsilon_i)_{i\geq 0}$ are i.i.d. centered random variables with finite variance, $\mathbb{E}[\varepsilon_1^2] = s_\epsilon^2$. The density of ε_1, f_ε, belongs to $\mathbb{L}_2(\mathbb{R})$, and for all $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$;*

(iii) *the innovations $(\eta_i)_{i\geq 0}$ are i.i.d. centered random variables with unit variance $\mathbb{E}[\eta_1^2] = 1$ and $\mathbb{E}[\eta_1^3] = 0$;*

(iv) *the X_i's are strictly stationary and ergodic with invariant density f_{θ_0};*

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

(vi) *the function to estimate Π_θ belongs to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ and the functions V_{Π_θ} and $Q_{\Pi_\theta^2}$ defined in (2) and (3) respectively must be integrable.*

The assumption **A** 1(*ii*) on f_ε is quite usual when considering deconvolution estimation. Furthermore, the first part of **A** 1(*vi*) is not restrictive and the second part can be understood as "Π_θ^* (resp. $(\Pi_\theta^2)^*$) has to be smooth enough compared to f_ε^*".

A key ingredient in the construction of our estimator of the parameter θ_0 is the choice of a "contrast function". Details about this notion can be found in [19]. For the purpose of this study, we consider the contrast function proposed by [15], that is

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

where

$$m_\theta(\mathbf{y}_i) = Q_{\Pi_\theta^2}(y_i) - 2V_{\Pi_\theta}(\mathbf{y}_i),$$

and the functions Q and V are two operators such that

$$\begin{cases} \mathbb{E}[Q_{\Pi_\theta^2}(Y_1)] = \mathbb{E}[\Pi_\theta^2(X_1)] \\ \mathbb{E}[V_{\Pi_\theta}(\mathbf{Y}_1)] = \mathbb{E}[\Pi_\theta(\mathbf{X}_1)]. \end{cases}$$

We are now able to describe in detail our procedure.

The procedure: Let us explain the choice of the contrast function and how the strategy of deconvolution works under assumptions **A** 1(*i*) up to (*vi*). Obviously, owing to the definition of the model (1), the $\mathbf{Y}_i$ are not *i.i.d.*. However, by assumption **A** 1(*iv*), they are stationary ergodic [1], so the convergence of $\mathbf{P}_n m_\theta$ to $\mathbf{P}m_\theta = \mathbb{E}\left[m_\theta(\mathbf{Y}_1)\right]$ as n tends to infinity is provided by the Ergodic Theorem. Moreover, the limit $\mathbb{E}\left[m_\theta(\mathbf{Y}_1)\right]$ of the contrast function can be explicitly computed. To do this, we use the same technique as in the convolution problem (see [4]). Let us denote by F_X the density of $\mathbf{X}_i$ and F_Y the density of $\mathbf{Y}_i$. We remark that $F_Y = F_X \star (f_\varepsilon \otimes f_\varepsilon)$ and $F_Y^* = F_X^*(f_\varepsilon^* \otimes f_\varepsilon^*)$, where $\star$ stands for the convolution product, and then by Parseval equality we have

$$\mathbb{E}[\Pi_\theta(\mathbf{X}_i)] = \int\int \Pi_\theta F_X = \frac{1}{2\pi}\int\int \Pi_\theta^* \overline{F_X^*} = \int\int \frac{\Pi_\theta^*}{\overline{f_\varepsilon^*}\otimes\overline{f_\varepsilon^*}}\overline{F_Y^*}.$$

The idea is then to define

$$V_{\Pi_\theta}^* = \frac{\Pi_\theta^*}{\overline{f_\varepsilon^*}\otimes\overline{f_\varepsilon^*}} \tag{2}$$

so that

$$\mathbb{E}[\Pi_\theta(\mathbf{X}_i)] = \frac{1}{2\pi}\int\int V_{\Pi_\theta}^* \overline{F_Y^*} = \int\int V_{\Pi_\theta} F_Y = \mathbb{E}[V_{\Pi_\theta}(\mathbf{Y}_i)].$$

In the same way, we find an operator Q to replace the term $\int \Pi_\theta^2(X_i, y)dy$. More precisely, for all function Π_θ, let Q_{Π_θ} be the inverse Fourier transform of $\frac{\Pi_\theta^*(x,0)}{f_\varepsilon^*(-x)}$, that is

$$Q_{\Pi_\theta}(x) = \frac{1}{2\pi}\int e^{ixu}\frac{\Pi_\theta^*(u,0)}{f_\varepsilon^*(-u)}du. \tag{3}$$

The operators Q and V are chosen to satisfy the following Lemma.

Lemma 2.1. *For all $i \in \{1,\dots,n+1\}$, we have*

1. $\mathbb{E}[V_{\Pi_\theta}(\mathbf{Y}_i)|X_1,\dots,X_{n+1}] = \Pi_\theta(\mathbf{X}_i)$.
2. $\mathbb{E}[Q_{\Pi_\theta}(Y_i)|X_1,\dots,X_{n+1}] = \int \Pi_\theta(X_i, y)dy$.
3. $\mathbb{E}[V_{\Pi_\theta}(\mathbf{Y}_i)] = \int\int \Pi_\theta(x,y)\Pi_{\theta_0}(x,y)f_{\theta_0}(x)dxdy$.
4. $\mathbb{E}[Q_{\Pi_\theta}(Y_i)] = \int\int \Pi_\theta(x,y)f_{\theta_0}(x)dxdy$.

The proof of Lemma 2.1 is postponed in Subsection 4.1.

By using the operators Q and V, the contrast is defined as

$$\mathbf{P}_n m_\theta = \frac{1}{n}\sum_{i=1}^{n} Q_{\Pi_\theta^2}(Y_i) - 2V_{\Pi_\theta}(\mathbf{Y}_i).$$

It follows from Lemma 2.1 that

$$\begin{aligned}\mathbf{P}m_\theta &= \mathbb{E}[m_\theta(\mathbf{Y}_1)] = \mathbb{E}[\Pi_\theta^2(X_1)] - 2\mathbb{E}[\Pi_\theta(\mathbf{X}_1)]\\ &= \int\int \Pi_\theta^2(x,y)f_{\theta_0}(x)dxdy - 2\int\int \Pi_\theta(x,y)\Pi_{\theta_0}(x,y)f_{\theta_0}(x)dxdy\\ &= ||\Pi_{\theta_0}||_f^2 - 2\langle \Pi_\theta, \Pi_{\theta_0}\rangle_f = \|\Pi_\theta - \Pi_{\theta_0}\|_f^2 - \|\Pi_{\theta_0}\|_f^2.\end{aligned} \tag{4}$$

Under the uniqueness assumption **A** 2 presented in the next section, this quantity is minimal when $\theta=\theta_0$. Hence, the associated minimum-contrast estimator $\widehat{\theta}_n$ is defined as any solution of

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \mathbf{P}_n m_\theta. \tag{5}$$

[1]We refer the reader to [9] for the proof that if $(X_i)_i$ is an ergodic process then the process $(Y_i)_i$, which is the sum of an ergodic process with an *i.i.d.* noise, is again stationary ergodic. Moreover, by the definition of an ergodic process, if $(Y_i)_i$ is an ergodic process then the couple $\mathbf{Y}_i = (Y_i, Y_{i+1})$ inherits the property (see [11])

3 Asymptotic properties of the least squares estimator

The following Theorem states the consistency of our estimator and the Central Limit Theorem (CLT) for α-mixing processes. To this aim, we further assume that the following assumptions hold true:

A 2. *The application $\theta \mapsto \mathbf{P}m_\theta$ admits a unique minimum and its Hessian matrix denoted by $\mathcal{V}_\theta$ is non-singular in θ_0.*

A 3. *(Local dominance):* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|Q_{\Pi_\theta^2}(Y_1)\right|\right] < \infty.$

A 4.

(i) *(Regularity): We assume that the function Π_θ is twice continuously differentiable w.r.t $\theta \in \Theta$ for any x and measurable w.r.t x for all θ in Θ. Additionally, each coordinate of $\nabla_\theta \Pi_\theta$ and each coordinate of $\nabla_\theta^2 \Pi_\theta$ belongs to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ and each coordinate of $Q_{\nabla_\theta \Pi_\theta^2}$ and $Q_{\nabla_\theta^2 \Pi_\theta^2}$ have to be integrable as well. In the same way, each coordinate of $V_{\nabla_\theta \Pi_\theta}$ and $V_{\nabla_\theta^2 \Pi_\theta}$ have to be integrable.*

(ii) *(Moment condition): For some $\delta > 0$ and for $j \in \{1, \dots, r\}$:*

$$\mathbb{E}\left[\left|Q_{\frac{\partial \Pi_\theta^2}{\partial \theta_j}}(Y_1)\right|^{2+\delta}\right] < \infty.$$

(iii) *(Hessian Local dominance): For some neighborhood $\mathcal{U}$ of θ_0 and for $j, k \in \{1, \dots, r\}$:*

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{U}}\left|Q_{\frac{\partial^2 \Pi_\theta^2}{\partial \theta_j \partial \theta_k}}(Y_1)\right|\right] < \infty.$$

A 5. *(Statistical assumptions):*

- *The stochastic process $(X_i)_{i\geq 1}$ is α-mixing (see Subsection 4.2.3 for a definition and [9] for a complete details on mixing processes).*
- *Let $g(x)$ be a nonnegative function and $\beta(q)$ be a nonnegative decreasing function on $\mathbb{Z}_+$ such that*

$$\|\Pi^q(x, .) - f_\theta\|_{\mathrm{TV}} \leq g(x)\beta(q), \qquad \mathbf{(M)}$$

  where $\Pi^q(x, .)$ the distribution of X_{i+q} given $X_i = x$ and $\|.\|_{\mathrm{TV}}$ the total variation distance.

Remark 3.1. *The stochastic process $(X_i)_{i\geq 1}$ is said to be*

- *geometrically ergodic if (**M**) holds with $\beta(q) = t^q$ for some $t < 1$.*
- *uniform ergodic if (**M**) holds with g bounded and $\beta(q) = t^q$ for some $t < 1$.*
- *polynomial ergodic of order m where $m \geq 0$ if (**M**) holds with $\beta(q) = q^{-m}$.*

A lot of processes as AR, ARCH, GARCH processes satisfy these mixing assumptions.

The regularity conditions **A** 4(*i*) are not restrictive and are similar to **A** 1(*vi*) for the first and second derivatives of Π_θ (resp. Π^2_θ).

Let us now introduce the matrix $\Sigma(\theta)$ given by

$$\Sigma(\theta) = \mathcal{V}_\theta^{-1}\Omega(\theta)\mathcal{V}_\theta^{-1'} \quad \text{with} \quad \Omega(\theta) = \Omega_0(\theta) + 2\sum_{j=2}^{+\infty}\Omega_{j-1}(\theta),$$

where $\Omega_0(\theta) = \mathbb{V}ar\left(\nabla_\theta m_\theta(\mathbf{Y_1})\right)$ and $\Omega_{j-1}(\theta) = \mathbb{C}ov\left(\nabla_\theta m_\theta(\mathbf{Y_1}), \nabla_\theta m_\theta(\mathbf{Y_j})\right)$.

Theorem 3.1. *Under Assumptions **A** 1–**A** 5 and if* (**M**) *holds such that* $\mathbb{E}[g(X_1)] < \infty$ *and* $\beta(q)$ *satisfies* $\sum_q \beta(q)^{\frac{\delta}{2+\delta}} < \infty$*, where* δ *is given in the moment assumption **A** 4(ii), let* $\widehat{\theta}_n$ *be the least square estimator defined in* (5)*. Then we have*

$$\widehat{\theta}_n \longrightarrow \theta_0 \qquad \textit{in probability} \quad \textit{as } n \to \infty.$$

Moreover,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to \mathcal{N}\left(0, \Sigma(\theta_0)\right) \qquad \textit{in law as } n \to \infty.$$

The proof of Theorem 3.1 is provided in Subsection 4.2.

The following corollary gives an expression of the matrices $\Omega(\theta_0)$ and $\mathcal{V}_{\theta_0}$ defined in $\Sigma(\theta)$ of Theorem 3.1.

Corollary 3.1. *Under assumptions **A** 1,**A** 2 and **A** 4(i), the matrix* $\Omega(\theta_0)$ *is given by*

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

where:

$$\begin{aligned}\Omega_0(\theta_0) =& \mathbb{E}[Q^2_{\nabla_\theta \Pi^2_\theta}(Y_1)] + 4\mathbb{E}[V^2_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)] \\ &- \left(\mathbb{E}[\nabla_\theta \Pi^2_\theta(X_1)]^2 + 4\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]^2 - 4\mathbb{E}[\nabla_\theta \Pi^2_\theta(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]\right)\end{aligned}$$

and, the covariance terms are given by

$$\begin{aligned}\Omega_{j-1}(\theta_0) = \mathbb{C}ov\left(\nabla_\theta \Pi^2_\theta(X_1), \nabla_\theta \Pi^2_\theta(X_j)\right) + 4\Big(&\mathbb{C}ov\left(\nabla_\theta \Pi_\theta(\mathbf{X}_1), \nabla_\theta \Pi_\theta(\mathbf{X}_j)\right) \\ &- \mathbb{C}ov\left(\nabla_\theta \Pi^2_\theta(X_1), \nabla_\theta \Pi_\theta(\mathbf{X}_j)\right)\Big),\end{aligned}$$

where the differential $\nabla_\theta \Pi_\theta$ *is taken at point* $\theta = \theta_0$*.*
Furthermore, the Hessian matrix $\mathcal{V}_{\theta_0}$ *is given by*

$$\left([\mathcal{V}_{\theta_0}]_{j,k}\right)_{1\le j,k\le r} = 2\left(\left\langle \frac{\partial \Pi_\theta}{\partial \theta_k}, \frac{\partial \Pi_\theta}{\partial \theta_j}\right\rangle\right)_{j,k} \quad \textit{at point } \theta = \theta_0.$$

The proof of Corollary 3.1 is given in Subsection 4.3.

Note that those results do not require the knowledge of f_{θ_0} as it is the case in [10]. Furthermore, they apply to a large class of HMMs with homoscedastic or heteroscedastic innovations. Under **A** 1–**A** 5, our estimation procedure allows to achieve the parametric rate and an analytical expression of the asymptotic variance matrix is obtained to construct confidence intervals.

4 Proofs

4.1 Proof of Lemma 2.1

Proof. We prove only points 1. and 2. since the other assertions are immediate consequences.

1. Let use set

$$V_{\Pi_\theta}(Y_i, Y_{i+1}) = \frac{1}{4\pi^2}\iint e^{iY_iu+iY_{i+1}v}\frac{\Pi_\theta^*(u,v)}{f_\varepsilon^*(-u)f_\varepsilon^*(-v)}dudv.$$

By denoting $X_{1:n+1} = (X_1, \ldots, X_n)$, we have

$$\mathbb{E}[V_{\Pi_\theta}(Y_i, Y_{i+1})|X_{1:n+1}] = \frac{1}{4\pi^2}\iint \mathbb{E}[e^{iY_iu+iY_{i+1}v}|X_{1:n+1}]\frac{\Pi_\theta^*(u,v)}{f_\varepsilon^*(-u)f_\varepsilon^*(-v)}dudv.$$

By using the independence between (X_i) and (ε_i), we have

$$\begin{aligned}\mathbb{E}[e^{iY_iu+iY_{i+1}v}|X_{1:n+1}] &= \mathbb{E}[e^{iX_iu+iX_{i+1}v} + e^{i\varepsilon_iu} + e^{i\varepsilon_{i+1}v}|X_{1:n+1}]\\ &= e^{iX_iu+iX_{i+1}v}\mathbb{E}[e^{i\varepsilon_iu}]\mathbb{E}[e^{i\varepsilon_{i+1}v}]\\ &= e^{iX_iu+iX_{i+1}v}f_\varepsilon^*(-u)f_\varepsilon^*(-v).\end{aligned}$$

Hence

$$\mathbb{E}[V_{\Pi_\theta}(\mathbf{Y}_i)|X_{1:n+1}] = \frac{1}{4\pi^2}\iint e^{iX_iu+iX_{i+1}v}\Pi_\theta^*(u,v)dudv = \Pi_\theta(X_i, X_{i+1}) = \Pi_\theta(\mathbf{X}_i).$$

The point 1. is proved.

2. For the operator Q, we proceed in a similar manner. We have

$$Q_{\Pi_\theta}(Y_i) = \frac{1}{2\pi}\int e^{iY_iu}\frac{\Pi_\theta^*(u,0)}{f_\varepsilon^*(-u)}du.$$

Hence

$$\mathbb{E}[Q_{\Pi_\theta}(Y_i)|X_{1:n}] = \frac{1}{2\pi}\int \mathbb{E}[e^{iY_iu}|X_{1:n}]\frac{\Pi_\theta^*(u,0)}{f_\varepsilon^*(-u)}du.$$

By using the independence between X_i and ε_i, we have

$$\mathbb{E}[e^{iY_iu}|X_{1:n+1}] = \mathbb{E}[e^{iX_iu} + e^{i\varepsilon_iu}|X_{1:n+1}] = e^{iX_iu}\mathbb{E}[e^{i\varepsilon_iu}] = e^{iX_iu}f_\varepsilon^*(-u).$$

Thus

$$\mathbb{E}[Q_{\Pi_\theta}(Y_i)|X_{1:n}] = \frac{1}{2\pi}\int e^{iX_iu}\Pi_\theta^*(u,0)du$$

By denoting by $\Pi_{\theta,y}$ the function $x \mapsto \Pi_{\theta,y}(x) = \Pi_\theta(x,y)$, we obtain

$$\Pi_\theta^*(u,0) = \int\int e^{-ixu}\Pi_{\theta,y}(x)dxdy = \int \Pi_{\theta,y}^*(u)dy.$$

So

$$\frac{1}{2\pi}\iint e^{iX_iu}\Pi_{\theta,y}^*(-u)dudy = \int \Pi_\theta(x,y)dy.$$

The point 2. is proved.

□

4.2 Proofs of Theorem 3.1

For the reader convenience we split the proof of Theorem 3.1 into three parts: in Subsection 4.2.1, we give the proof of the existence of our contrast estimator defined in (2). In Subsection 4.2.2, we prove the consistency, that is, the first part of Theorem 3.1. Then, we prove the asymptotic normality of our estimator in Subsection 4.2.3, that is, the second part of Theorem 3.1. The Section 4.3 is devoted to Corollary 3.1.

4.2.1 Proof of the existence and measurability of the M-Estimator

By assumption, the function $m_\theta(\mathbf{y}_i) = Q_{\Pi_\theta^2}(y_i) - 2V_{\Pi_\theta}(\mathbf{y}_i)$ is continuous w.r.t θ. Hence, the function $\mathbf{P}_n m_\theta = \frac{1}{n}\sum_{i=1}^n m_\theta(\mathbf{Y}_i)$ is continuous w.r.t θ belonging to the compact subset Θ. So, there exists $\tilde{\theta}$ belongs to Θ such that $\inf_{\theta\in\Theta} \mathbf{P}_n m_\theta = \mathbf{P}_n m_{\tilde{\theta}}$. □

4.2.2 Proof of the Consistency

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

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

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

Lemma 4.1. *Uniform Law of Large Numbers (see [16] for the proof).*
Let $(Y_i)_{i\geq 1}$ be an ergodic stationary process and suppose that:

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

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

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

By assumption Π_θ is continuous w.r.t θ for any x and measurable w.r.t x for all θ which implies the continuity and the measurability of the function $\mathbf{P}_n m_\theta$ on the compact subset Θ. Furthermore, the local dominance assumption **A** 3 implies that $\mathbb{E}\left[\sup_{\theta\in\Theta} |m_\theta(\mathbf{Y}_i)|\right]$ is finite. Indeed, by assumption **A** 3, we have

$$|m_\theta(\mathbf{y}_i)| = \left|Q_{\Pi_\theta^2}(y_i) - 2V_{\Pi_\theta}(\mathbf{y}_i)\right| \leq \left|Q_{\Pi_\theta^2}(y_i)\right| + 2\left|V_{\Pi_\theta}(\mathbf{y}_i)\right| < \infty.$$

Lemma 4.1 gives us the uniform convergence in probability of the contrast function: for any $\varepsilon > 0$:

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

Combining the uniform convergence with [12, Theorem 2.1 p. 2121 chapter 36] yields the weak (convergence in probability) consistency of the estimator. □

Remark 4.1. *In most applications, we do not know the bounds for the true parameter. So the compactness assumption is sometimes restrictive, one can replace the compactness assumption by: θ_0 is an element of the interior of a convex parameter space $\Theta \subset \mathbb{R}^r$. Then, under our assumptions except the compactness, the estimator is also consistent. The proof is the same and the existence is proved by using convex optimization arguments. One can refer to [13] for this discussion.*

4.2.3 Proof of the asymptotic normality

For the CLT, we need to define the α-mixing property of a process (we refer the reader to [9] for a complete review of mixing processes).

Definition 4.1 (α-mixing (strongly mixing process))**.** *Let $Y := (Y_i)_i$ denotes a general sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ and let $\mathcal{F}_k^m = \sigma(Y_k, \ldots, Y_m)$. The sequence Y is said to be α-mixing if $\alpha(n) \to 0$ as $n \to \infty$, where*

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

The proof of the CLT is based on the following Lemma.

Lemma 4.2. *Suppose that the conditions of the consistency hold. Suppose further that:*

(i) $(\mathbf{Y}_i)_i$ *is α-mixing.*

(ii) *(Moment condition): for some $\delta > 0$ and for each $j \in \{1, \ldots, r\}$:*

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

(iii) *(Hessian Local condition): for some neighborhood $\mathcal{U}$ of θ_0 and for $j, k \in \{1, \ldots, r\}$:*

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

(iv) *Assumption* (**M**) *given in Section 3 holds such that $\mathbb{E}[g(X_1)] < \infty$ and $\beta(q)$ satisfies $\sum_q \beta(q)^{\frac{\delta}{2+\delta}} < \infty$, where δ is given in the moment condition (ii).*

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

$$\Sigma(\theta_0) = \mathcal{V}_{\theta_0}^{-1}\Omega(\theta_0)\mathcal{V}_{\theta_0}^{-1},$$

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

Proof. The proof follows from Proposition 7.8 p.472 of [13] and [14], and by using the fact that, by regularity assumptions **A** 4(*i*) and the Lebesgue Differentiation Theorem, we have $\mathbb{E}[\nabla_\theta^2 m_\theta(\mathbf{Y}_1)] = \nabla_\theta^2\mathbb{E}[m_\theta(\mathbf{Y}_1)]$. □

It just remains to check that the conditions (ii) and (iii) of Lemma 4.2 hold under our assumptions **A** 4(*ii*) and (*iii*).

(ii): As the function Π_θ is twice continuously differentiable w.r.t θ, for all $\mathbf{y}_i \in \mathbb{R}^2$ and so also Π_θ^2, the application $m_\theta(\mathbf{y}_i) : \theta \in \Theta \mapsto m_\theta(\mathbf{y}_i) = Q_{\Pi_\theta^2}(y_i) - 2V_{\Pi_\theta}(\mathbf{y}_i)$ is twice continuously differentiable for all $\theta \in \Theta$ and its first derivatives are given by

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \nabla_\theta Q_{\Pi_\theta^2}(y_i) - 2\nabla_\theta V_{\Pi_\theta}(\mathbf{y}_i).$$

By assumption, for each $j \in \{1, \ldots, r\}$, $\frac{\partial\Pi_\theta}{\partial\theta_j}$ and $\frac{\partial\Pi_\theta^2}{\partial\theta_j}$ belong to $\mathbb{L}_1(\mathbb{R})$, therefore one can apply the Lebesgue Differentiation Theorem and Fubini Theorem to obtain

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \left[Q_{\nabla_\theta\Pi_\theta^2}(y_i) - 2V_{\nabla_\theta\Pi_\theta}(\mathbf{y}_i)\right]. \tag{6}$$

Then, for some $\delta > 0$, by the moment assumption **A** 4(*ii*), we have

$$|\nabla_\theta m_\theta(\mathbf{y}_i)|^{2+\delta} = \left|Q_{\nabla_\theta \Pi_\theta^2}(y_i) - 2V_{\nabla_\theta \Pi_\theta}(\mathbf{y}_i)\right|^{2+\delta} \le C_1 \left|Q_{\nabla_\theta \Pi_\theta^2}(y_i)\right|^{2+\delta} + C_2 \left|V_{\nabla_\theta \Pi_\theta}(\mathbf{y}_i)\right|^{2+\delta} < \infty,$$

where C_1 and C_2 and C_3 denote three positive constants.

(iii): For $j,k \in \{1,\dots,r\}$, $\frac{\partial^2 \Pi_\theta}{\partial\theta_j \partial\theta_k}$ and $\frac{\partial^2 \Pi_\theta^2}{\partial\theta_j \partial\theta_k}$ belong to $\mathbb{L}_1(\mathbb{R})$, the Lebesgue Differentiation Theorem gives

$$\nabla_\theta^2 m_\theta(\mathbf{y}_i) = \left[Q_{\nabla_\theta^2 \Pi_\theta^2}(y_i) - 2V_{\nabla_\theta^2 \Pi_\theta}(\mathbf{y}_i)\right],$$

and, for some neighborhood $\mathcal{U}$ of θ_0, by the local dominance assumption **A** 4(*iii*),

$$\mathbb{E}\left[\sup_{\theta\in\mathcal{U}} \left\|\nabla_\theta^2 m_\theta(\mathbf{Y}_i)\right\|\right] \le \mathbb{E}\left[\sup_{\theta\in\mathcal{U}} \left\|Q_{\nabla_\theta^2 \Pi_\theta^2}(Y_i)\right\|\right] + 2\mathbb{E}\left[\sup_{\theta\in\mathcal{U}} \left\|V_{\nabla_\theta^2 \Pi_\theta}(\mathbf{Y}_i)\right\|\right] < \infty.$$

This ends the proof of Theorem 3.1. □

4.3 Proof of Corollary 3.1

By replacing $\nabla_\theta m_\theta(\mathbf{Y}_1)$ by its expression (6), we have

$$\begin{aligned}
\Omega_0(\theta) &= \mathbb{V}ar\left[Q_{\nabla_\theta \Pi_\theta^2}(Y_1) - 2V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)\right] \\
&= \mathbb{V}ar\left(Q_{\nabla_\theta \Pi_\theta^2}(Y_1)\right) + 4\mathbb{V}ar\left(V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)\right) - 4\mathbb{C}ov\left(Q_{\nabla_\theta \Pi_\theta^2}(Y_1), V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)\right).
\end{aligned}$$

Owing to Lemma 2.1, we obtain

$$\begin{aligned}
\mathbb{V}ar\left(Q_{\nabla_\theta \Pi_\theta^2}(Y_1)\right) &= \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)^2] - \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)]^2 \\
&= \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)^2] - \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]^2.
\end{aligned}$$

In a similar manner, using again Lemma 2.1, we have

$$\begin{aligned}
\mathbb{V}ar\left(V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)\right) &= \mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)^2] - \mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)]^2 \\
&= \mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)^2] - \mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]^2
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}ov\left(Q_{\nabla_\theta \Pi_\theta^2}(Y_1), V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)\right) &= \mathbb{E}\left[\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1) V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1) | X_{1:n+1}]\right] \\
&\quad - \mathbb{E}\left[\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1) | X_{1:n+1}]\right]\mathbb{E}\left[\mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1) | X_{1:n+1}]\right] \\
&= \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1) V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)] - \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)].
\end{aligned}$$

Hence

$$\begin{aligned}
\Omega_0(\theta) &= \mathbb{V}ar\left(\nabla_\theta m_\theta(\mathbf{Y}_1)\right) \\
&= \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)^2] + 4\mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)^2] - 4\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1) V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)] \\
&\quad - \left(\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]^2 + 4\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]^2 - 4\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]\right).
\end{aligned}$$

Calculus of the covariance matrix of Corollary 3.1: By replacing $\nabla_\theta m_\theta(Y_1)$ by its expression (6), we have

$$\begin{aligned}
\Omega_{j-1}(\theta) &= \mathbb{C}ov\left(\nabla_\theta m_\theta(\mathbf{Y}_1), \nabla_\theta m_\theta(\mathbf{Y}_j)\right) \\
&= \mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_1)\nabla_\theta m_\theta(\mathbf{Y}_j)] - \mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_1)]\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_j)].
\end{aligned}$$

It follows from Lemma 2.1 and the stationarity assumption **A** 1(*iv*) of $(Y_i)_{i\geq 1}$ that

$$\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_1)] = \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)] - 2\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)].$$

Moreover

$$\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_j)] = \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_j)] - 2\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_j)].$$

Hence

$$\begin{aligned}\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_1)]\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_j)] = \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_j)] - 2\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_j)]\\ - 2\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_j)] + 4\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_j)].\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_1)\nabla_\theta m_\theta(\mathbf{Y}_j)] &= \mathbb{E}\Big[(Q_{\nabla_\theta \Pi_\theta^2}(Y_1) - 2V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1))(Q_{\nabla_\theta \Pi_\theta^2}(Y_j) - 2V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_j))\Big]\\ &= \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)Q_{\nabla_\theta \Pi_\theta^2}(Y_j)] - 2\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)V_{\nabla_\theta \Pi_\theta}(Y_j)]\\ &\quad - 2\mathbb{E}[V_{\nabla_\theta \Pi_\theta}(Y_1)Q_{\nabla_\theta \Pi_\theta^2}(Y_j)] + 4\mathbb{E}[V_{\nabla_\theta \Pi_\theta}(Y_1)V_{\nabla_\theta \Pi_\theta}(Y_j)].\end{aligned}$$

Furthermore, the Fubini Theorem yields

$$\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)Q_{\nabla_\theta \Pi_\theta^2}(Y_j)] = \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)\nabla_\theta \Pi_\theta^2(X_j)].$$

Similarly, we have

$$\mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_j)] = \mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)\nabla_\theta \Pi_\theta(\mathbf{X}_j)].$$

Hence, by the stationarity of $(Y_i)_{i\geq 1}$,

$$\begin{aligned}\mathbb{E}[\nabla_\theta m_\theta(\mathbf{Y}_1)\nabla_\theta m_\theta(\mathbf{Y}_j)] &= \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)\nabla_\theta \Pi_\theta^2(X_j)]\\ &\quad + 4\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)\nabla_\theta \Pi_\theta(\mathbf{X}_j)] - 4\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_j)].\end{aligned}$$

By using Lemma 2.1, the last term is equal to

$$\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_j)] = \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)\nabla_\theta \Pi_\theta(\mathbf{X}_j)].$$

Therefore, the covariance matrix is given by

$$\begin{aligned}\Omega_{j-1}(\theta) &= \mathbb{C}ov\left(\nabla_\theta m_\theta(\mathbf{Y}_1), \nabla_\theta m_\theta(\mathbf{Y}_j)\right)\\ &= \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)\nabla_\theta \Pi_\theta^2(X_j)] - \mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_j)]\\ &\quad + 4\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)\nabla_\theta \Pi_\theta(\mathbf{X}_j)] - 4\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_j)]\\ &\quad - 4\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)\nabla_\theta \Pi_\theta(\mathbf{X}_j)] + 4\mathbb{E}[\nabla_\theta \Pi_\theta^2(X_1)]\mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_j)].\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{C}ov\left(\nabla_\theta m_\theta(\mathbf{Y}_1), \nabla_\theta m_\theta(\mathbf{Y}_j)\right) &= \mathbb{C}ov\left(\nabla_\theta \Pi_\theta^2(X_1), \nabla_\theta \Pi_\theta^2(X_j)\right)\\ &\quad + 4\left(\mathbb{C}ov\left(\nabla_\theta \Pi_\theta(\mathbf{X}_1), \nabla_\theta \Pi_\theta(\mathbf{X}_j)\right) - \mathbb{C}ov\left(\nabla_\theta \Pi_\theta^2(X_1), \nabla_\theta \Pi_\theta(\mathbf{X}_j)\right)\right).\end{aligned}$$

Expression of the Hessian matrix $\mathcal{V}_\theta$: We have

$$\mathbf{P}m_\theta = \|\Pi_\theta\|_2^2 - 2\left\langle \Pi_\theta, \Pi_{\theta_0}\right\rangle.$$

Under **A** 4(*i*) ,for all θ in Θ, the application $\theta \mapsto \mathbf{P}m_\theta$ is twice differentiable w.r.t θ on the compact subset Θ. For $j \in \{1, \dots, r\}$, at the point $\theta = \theta_0$, we have

$$\frac{\partial \mathbf{P}m}{\partial \theta_j}(\theta) = 2\left\langle \frac{\partial \Pi_\theta}{\partial \theta_j}, \Pi_\theta \right\rangle - 2\left\langle \frac{\partial \Pi_\theta}{\partial \theta_j}, \Pi_{\theta_0} \right\rangle = 2\left\langle \frac{\partial \Pi_\theta}{\partial \theta_j}, \Pi_\theta - \Pi_{\theta_0} \right\rangle = 0$$

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

$$\frac{\partial^2 \mathbf{P}m}{\partial \theta_j \partial \theta_k}(\theta) = 2\left(\left\langle \frac{\partial^2 \Pi_\theta}{\partial \theta_j \theta_k}, \Pi_\theta - l_{\theta_0} \right\rangle + \left\langle \frac{\partial \Pi_\theta}{\partial \theta_k}, \frac{\partial \Pi_\theta}{\partial \theta_j} \right\rangle\right)_{j,k} = 2\left(\left\langle \frac{\partial \Pi_\theta}{\partial \theta_k}, \frac{\partial \Pi_\theta}{\partial \theta_j} \right\rangle\right)_{j,k}.$$

The proof of Corollary 3.1 is completed. □

References

[1] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342, 2010.

[2] K. C. Chanda. Large sample analysis of autoregressive moving-average models with errors in variables. *J. Time Ser. Anal.*, 16(1):1–15, 1995.

[3] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC^2: an efficient algorithm for sequential analysis of state-space models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(3):397–426, 2013.

[4] F. Comte, Y. Rozenholc, and M.-L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3):431–452, 2006.

[5] F. Comte and M.-L. Taupin. Semiparametric estimation in the (auto)-regressive β-mixing model with errors-in-variables. *Math. Methods Statist.*, 10(2):121–160, 2001.

[6] M. Costa and T. Alpuim. Parameter estimation of state space models for univariate observations. *J. Statist. Plann. Inference*, 140(7):1889–1902, 2010.

[7] J. Dedecker, A. Samson, and M.-L. Taupin. Estimation in autoregressive model with measurement error. *ESAIM Prob. Stat.*, 18:277–307, 2014.

[8] R. Douc, E. Moulines, J. Olsson, and R. van Handel. Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, 39(1):474–513, 2011.

[9] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

[10] S. El Kolei. Parametric estimation of hidden stochastic model by contrast minimization and deconvolution. *Metrika*, 76(8):1031–1081, 2013.

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

[12] B. E. Hansen and J. L. Horowitz. Handbook of econometrics, vol. 4 robert f. engle and daniel l. mcfadden, editors elsevier science b. v., 1994. *Econometric Theory*, 13(01):119–132, February 1997.

[13] F. Hayashi. *Econometrics*. Princeton University Press, Princeton, NJ, 2000.

[14] G. L. Jones. On the Markov chain central limit theorem. *Probab. Surv.*, 1:299–320, 2004.

[15] C. Lacour. Least squares type estimation of the transition density of a particular hidden Markov chain. *Electron. J. Stat.*, 2:1–39, 2008.

[16] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam, 1994.

[17] J. Olsson and T. Rydén. Particle filter-based approximate maximum likelihood inference asymptotics in state-space models. In *Conference Oxford sur les méthodes de Monte Carlo séquentielles*, volume 19 of *ESAIM Proc.*, pages 115–120. EDP Sci., Les Ulis, 2007.

[18] J. Staudenmayer and J. P. Buonaccorsi. Measurement error in linear autoregressive models. *J. Amer. Statist. Assoc.*, 100(471):841–852, 2005.

[19] A. W. Van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 1998.