



HAL
open science

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

Christophe Chesneau, Salima El Kolei, Fabien Navarro

► To cite this version:

Christophe Chesneau, Salima El Kolei, Fabien Navarro. Parametric estimation of hidden Markov models by least squares type estimation and deconvolution. 2019. hal-01598922v2

HAL Id: hal-01598922

<https://hal.science/hal-01598922v2>

Preprint submitted on 1 Mar 2019 (v2), last revised 28 Jan 2022 (v6)

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

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

Parametric Estimation of Hidden Markov Models by Least Squares type Estimation and Deconvolution

Christophe Chesneau
LMNO, Université de Caen, France

and

Salima El Kolei
CREST, ENSAI, France

and

Fabien Navarro
CREST, ENSAI, France

March 1, 2019

Abstract

This paper develops a computationally efficient parametric approach to the estimation of general hidden Markov models (HMMs). For non-Gaussian HMMs, the calculation of the Maximum Likelihood Estimator (MLE) involves a high-dimensional integral without an explicit solution that is difficult to calculate with precision. We develop a new alternative method based on the theory of estimating functions and deconvolution strategy. Our procedure requires the same assumptions as the MLE and deconvolution estimators. We provide theoretical guarantees on the performance of the resulting estimator; its consistency and asymptotic normality are established. This leads to building confidence intervals in practice. Monte Carlo experiments are investigated and compared with the MLE. Finally, we illustrate our approach on real data for ex-ante interest rate forecasts.

Keywords: Parametric inference; Stochastic volatility; Measurement error; Contrast function; Expected Inflation; Unobserved components model.

1 Introduction

In this paper, a hidden non-linear Markov model (HMM) with heteroscedastic noise is considered; we observe n random variables Y_1, \dots, Y_{n+1} having the following additive structure

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

where $(X_i)_{i \geq 1}$ is a strictly stationary, ergodic unobserved Markov chain that depends on two unknown measurable functions b_{θ_0} and σ_{θ_0} and 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} . 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 . This assumption cannot be easily removed, even if the density of ε_t is completely known up to a scale parameter, the model (1) may be non-identifiable as soon as the invariant density of X_t is smoother than the density of the noise (see, e.g., Butucea and Matias (2005)).

Our objective is to estimate the parameter vector θ_0 for non-linear HMMs with heteroscedastic innovations described by the function σ_{θ_0} in (1) assuming that the model is correctly specified, i.e., θ_0 belongs to the interior of a compact set $\Theta \subset \mathbb{R}^r$, with $r \in \mathbb{N}^*$.

Many articles are devoted to parameter estimation and to the study of asymptotic properties of estimators when $(X_i)_{i \geq 1}$ is an autoregressive moving average (ARMA) process (see Newey and McFadden (1994a), Staudenmayer and Buonaccorsi (2005) and Costa and Alpuim (2010)). However, for more general models, (1) is known as HMM with potentially a non-compact continuous state space. This model constitutes a very famous class of discrete-time stochastic processes, with many applications in various fields such as biology, speech recognition or finance. In Douc et al. (2011), the authors study the consistency of the MLE estimator for general HMMs, but they do not provide a method for calculating it in practice. It is well-known that its computation is extremely expensive due to the non-observability of the Markov chain and the proposed methodologies are essentially based on Expectation-Maximization approach or Monte Carlo-based methods (e.g. Markov chain Monte Carlo, Sequential Monte Carlo or particle Markov chain Monte Carlo, see Andrieu et al. (2010), Chopin et al. (2013) and Olsson and Rydén (2007)). Except in the Gaussian and linear setting, where the MLE can be processed by a Kalman filter and in this particular

case the calculation will be relatively fast but there are few cases where real data satisfy this assumption.

In this paper, we do not consider the Bayesian approach, the model (1) is also known in this case as the so-called convolution model, and our approach is therefore based on Fourier analysis. The restrictions on error distribution and rate of convergence obtained for our estimator are also of the same type. If we focus our attention on (semi-)parametric models, few results exist. To the best of our knowledge, the first study that gives a consistent estimator is Comte and Taupin (2001). The authors propose an estimation procedure based on a least squares minimization. Recently, in Dedecker et al. (2014), 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 El Kolei (2013) and El Kolei and Pelgrin (2017) the authors propose a consistent estimator for parametric models assuming knowledge of the stationary density f_{θ_0} up to the unknown parameters θ_0 for the construction of the estimator. For many processes, this density has no explicit form, and even in some cases where it is known, it may be more complex to apply deconvolution techniques using this density rather than the transition density. For example, the autoregressive conditional heteroscedasticity (ARCH) process is a family of processes for which transition density has an explicit form as opposed to the stationary density. These processes are widely used to model economic or financial variables.

In this work, our goal is to develop a new computationally efficient approach whose construction does not require knowledge of the invariant density. We provide a consistent estimator with a parametric rate of convergence for general models. Our approach is valid for non-linear HMMs with heteroscedastic innovations and our estimation principle is inspired by the one proposed in Lacour (2008). We 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 . The methodology we are developing is purely parametric and we are going further in this direction by proposing an analytical expression of the asymptotic variance matrix $\Sigma(\hat{\theta}_n)$ which thus allows us to consider the construction of confidence intervals (CIs).

Under general assumptions, we prove that our estimator is consistent and give some conditions under which the asymptotic normality can be stated and also provide an analytical expression of the asymptotic variance matrix. We show that this approach is much less

greedy from a computational point of view than the MLE for non-Gaussian HMMs and its implementation is straightforward since it requires to compute only Fourier transforms as in Dedecker et al. (2014). In particular, a numerical illustration is given for three models: a Gaussian AR(1) model for which our approach can be well understood; an AR(1) process with Laplace noises in order to study the influence of the smoothness of observation noise on the estimation of the parameters since it is known in deconvolution that this influences the convergence rate (see, e.g., Fan et al. (1991)); a stochastic volatility model (SV) also referred to as the unobserved components/stochastic volatility model (see, e.g., Stock and Watson (2007), Harvey and Shephard (1996) and Chan (2017)). There is by now a large literature on the fitting of SV models (see, e.g., the reviews in Ghysels et al. (1996), Bos and Shephard (2006) and Omori et al. (2007)). All are based on Bayesian methods and in particular MCMC methods. We therefore propose an alternative estimation method that is simple to implement and quick to calculate for this model, which is widely used in practice. We also illustrate the applicability of our procedure on a real dataset to estimate the ex-ante real interest rate since it is shown in Holston et al. (2017) and recently in Laubach and Williams (2003) that interest rates are subject to considerable real-time measurement error. In particular, we focus on the great inflation period. We show that during this period the Gaussianity hypothesis of observation noise is not verified and that in this study a SV-type model gives better results for the latent variable estimation. In this context, the Kalman filter is no longer optimal and therefore leads to a bias in parameter estimation, since in this case we approach the noise density by a Gaussian density to construct the MLE. This bias on the parameters propagates in the estimation of the latent variable (see El Kolei and Patras (2018)). This cannot be overlooked in models where the latent variable to be predicted is used to make political decisions. It seems important to study estimators other than the MLE that cannot be calculated by the Kalman filter. In this respect, our approach therefore provides better results than the (quasi-)MLE estimate.

The remainder of the paper is organized as follows. In Section 2, we present our assumptions on the Markov chain and give examples of such chains. Section 3 describes our estimator and its statistical properties. The consistency and the asymptotic normality of the estimator are established in Section 4. Simulated examples are provided in Section 5 and the real data application in Section 6. The proofs are gathered in Section 8.

2 Framework

2.1 Notations, Assumptions and Examples

Before presenting in detail the main estimation procedure of our study, let us introduce some preliminary notations and assumptions.

2.1.1 Notations

The Fourier transform of an integrable function u is denoted by $u^*(t) = \int e^{-itx}u(x)dx$. We set $\langle u, v \rangle_f = \int u(x)\bar{v}(x)f_{\theta_0}(x)dx$ with $v\bar{v} = |v|^2$. The norm of the operator T is defined by $\|T\|_f = (\int \int |T(x, y)|^2 f_{\theta_0}(x)dx dy)^{1/2}$. Let us recall that, by the properties of the Fourier transform, we have $(u^*)^*(x) = 2\pi u(-x)$ and $\langle u_1, u_2 \rangle_f = (1/(2\pi)) \langle u_1^*, u_2^* \rangle_f$. 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 ∇_{θ}^2g . 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}_{θ_0} , the expected value \mathbb{E}_{θ_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 $X = (X_i)_i$, $\mathbf{P}_n(X) = (1/n) \sum_{i=1}^n X_i$ (*resp.* $\mathbf{P}(X) = \mathbb{E}[X]$). For the purposes of this study, we work with Π_{θ} on a compact subset $A = A_1 \times A_2$. For more clarity, we write Π_{θ} instead of $\Pi_{\theta}\mathbf{1}_A$.

2.1.2 Assumptions

For the construction of our estimator, we consider three different types of assumptions.

A 1.

Smoothness and mixing assumptions:

- (i) *The function to estimate Π_{θ} belongs to $\mathbb{L}_1(A) \cap \mathbb{L}_2(A)$ and 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 $\mathbb{L}_1(A) \cap \mathbb{L}_2(A)$.*
- (ii) *The $(X_i)_i$ is strictly stationary, ergodic and α -mixing with invariant density f_{θ_0} .*

Assumptions on the noise ε_t and innovations η_t :

- (iii) – The errors $(\varepsilon_i)_i$ are independent and identically distributed centered random variables with finite variance, $\mathbb{E}[\varepsilon_1^2] = \sigma_\varepsilon^2$. The random variable ε_1 admits a known density, f_ε , belongs to $\mathbb{L}_2(\mathbb{R})$, and for all $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$.
- The innovations $(\eta_i)_i$ are independent and identically distributed centered random variables.

Identifiability assumptions:

- (iv) The mapping $\theta \mapsto \mathbf{P}m_\theta = \|\Pi_\theta - \Pi_{\theta_0}\|_f^2 - \|\Pi_{\theta_0}\|_f^2$ admits a unique minimum at θ_0 and its hessian matrix denoted by \mathcal{V}_θ is non-singular in θ_0 .

Let's make a few remarks on the above assumptions. Assumption **A 1(i)** is not restrictive and is often satisfied for many processes.

Taking the process $(X_i)_i$ defined in (1), we provide conditions on the functions b , σ and η ensuring that assumption **A 1(ii)** is satisfied (see Doukhan (1994)). Subsection 2.1.3 gives some examples of Markov chains satisfying these first two hypotheses.

- (a) The random variables $(\eta_i)_i$ are *i.i.d.* with an everywhere positive and continuous density function independent of $(X_i)_i$.
- (b) The function b_{θ_0} is bounded on every bounded set; that is, for every $K > 0$, $\sup_{|x| \leq K} |b_{\theta_0}(x)| < \infty$.
- (c) The function σ_{θ_0} satisfies, for every $K > 0$ and constant σ_1 , $0 < \sigma_1 \leq \inf_{|x| \leq K} \sigma_{\theta_0}(x)$ and $\sup_{|x| \leq K} \sigma_{\theta_0}(x) < \infty$.
- (d) There exist constants $C_b > 0$ and $C_\sigma > 0$, sufficiently large $M_1 > 0$, $M_2 > 0$, $c_1 \geq 0$ and $c_2 \geq 0$ such that $|b_{\theta_0}(x)| \leq C_b|x| + c_1$, for $|x| \geq M_1$ and $|\sigma_{\theta_0}(x)| \leq C_\sigma|x| + c_2$, for $|x| \geq M_2$ and $C_b + \mathbb{E}[\eta_1]C_\sigma < 1$.

Assumption **A 1(iii)** on f_ε is quite usual when considering deconvolution estimation, in particular, the first part is essential for the identifiability of the model (1). This assumption cannot be easily removed: even if the density of ε_t is completely known up to a scale parameter, the model (1) may be non-identifiable as soon as the invariant density of X_t is smoother than the density of the noise (see Butucea and Matias (2005)). The second part of **A 1(iii)** ensures the existence of the estimation criterion.

2.1.3 Examples

(Non-linear) AR(1) process: let us consider the following process

$$X_i = b_{\theta_0}(X_{i-1}) + \eta_i,$$

where η_i has a density with respect to the Lebesgue measure, which does not depend on $i \geq 1$. We suppose that b_{θ_0} is bounded on any compact set and that there exists $C > 0$ and $\rho < 1$ such that for all $|x| > C$, $|b_{\theta_0}| < \rho|x|$. In Mokkadem (1990) the author shows that if there exists $p > 0$ such that $\sup_x \mathbb{E}[|\eta_1|^p] < \infty$ then the chain is geometrically ergodic.

ARCH process: the ARCH process is given by

$$X_i = b_{\theta_0}(X_{i-1}) + \sigma_{\theta_0}(X_{i-1})\eta_i,$$

where b_{θ_0} and σ_{θ_0} are two continuous functions such that $\sigma(x) \neq 0$ for all x . We suppose that the distribution of η_i has a continuous density with respect to the Lebesgue measure and that there exists $p > 0$ such that $\mathbb{E}[|\eta_1|^p] < \infty$. If

$$\limsup_{|x| \rightarrow \infty} \frac{|b_{\theta_0}(x)| + |\sigma_{\theta_0}(x)|(\mathbb{E}[|\eta_1|^p])^{1/p}}{|x|} < 1,$$

then the chain is geometrically ergodic (see Doukhan (1994)).

ARX(1,1) process: the non-linear process ARX(1,1) is defined by

$$X_i = b_{\theta_0}(X_{i-1}, Z_i) + \eta_i,$$

where b_{θ_0} is bounded and $(\eta_i)_i, (Z_i)_i$ are independent sequences of *i.i.d.* random variables with $\mathbb{E}[|\eta_i|] < \infty$. We assume that the distribution of the variables Z_i admits a density w.r.t. the Lebesgue measure. Suppose that there exist $\rho < 1$ and a locally bounded measure function $g : \mathbb{R} \rightarrow \mathbb{R}^+$ such that $\mathbb{E}[g(Z_i)] < \infty$ and two constants $M > 0$ and $c > 0$ such that, $\forall |(u, v)| > M$, we have

$$|b_{\theta_0}(u, v)| < \rho|u| + h(v) - c \quad \text{and} \quad \sup_{|x| \leq M} |b_{\theta_0}(x)| < \infty,$$

then the chain is geometrically ergodic (see Doukhan (1994) p.105).

3 Estimation Procedure

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 Van der Vaart (1998). For

the purpose of this study, we consider the following contrast function

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

where the operators Q and V are given by

$$V_{\Pi_\theta}(x, y) = \frac{1}{4\pi^2} \int \int e^{i(xu+yv)} \frac{\Pi_\theta^*(u, v)}{f_\varepsilon^*(-u)f_\varepsilon^*(-v)} dudv \quad \text{and} \quad Q_{\Pi_\theta}(x) = \frac{1}{2\pi} \int e^{ixu} \frac{\Pi_\theta^*(u, 0)}{f_\varepsilon^*(-u)} du, \quad (3)$$

We are now able to explain the choice of such contrast function and to describe in detail our procedure.

The procedure: Owing to the definition of the model (1), the \mathbf{Y}_i are not *i.i.d.*. However, by assumption **A1(ii)**, they are stationary ergodic¹, so the convergence of $\mathbf{P}_n m_\theta$ to $\mathbf{P} m_\theta$ as n tends to infinity is provided by the Ergodic Theorem. Moreover, the limit $\mathbf{P} m_\theta$ of the contrast function can be explicitly computed. To do this, we use the same technique as in the convolution problem (see Comte et al. (2006)). Let us denote by F_X the density of \mathbf{X}_i and F_Y the density of 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

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

Similarly, the operator Q is defined to replace the term $\int \Pi_\theta^2(X_i, y) dy$. The operators Q and V are chosen to satisfy the following Lemma.

Lemma 3.1. *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. $\mathbb{E}[V_{\Pi_\theta}(\mathbf{Y}_i)] = \int \int \Pi_\theta(x, y) \Pi_{\theta_0}(x, y) f_{\theta_0}(x) dx dy$.
4. $\mathbb{E}[Q_{\Pi_\theta}(Y_i)] = \int \int \Pi_\theta(x, y) f_{\theta_0}(x) dx dy$.

¹We refer the reader to Doukhan (1994) 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 Genon-Catalot et al. (2000))

The proof of Lemma 3.1 is postponed in Subsection 8.1. It follows from Lemma 3.1 that

$$\begin{aligned} \mathbf{P}m_\theta &= \int \int \Pi_\theta^2(x, y) f_{\theta_0}(x) dx dy - 2 \int \int \Pi_\theta(x, y) \Pi_{\theta_0}(x, y) f_{\theta_0}(x) dx dy \\ &= \|\Pi_\theta\|_f^2 - 2\langle \Pi_\theta, \Pi_{\theta_0} \rangle_f = \|\Pi_\theta - \Pi_{\theta_0}\|_f^2 - \|\Pi_{\theta_0}\|_f^2. \end{aligned} \quad (4)$$

Hence, the contrast defined in (2) is the empirical counterpart of the \mathbb{L}_2 -norm between the true transition density Π_{θ_0} and the transition density Π_θ computed at $\theta \in \Theta$.

Under the identifiability assumption **A1**(*iv*), this quantity is minimal when $\theta = \theta_0$. So, the associated minimum-contrast estimator $\hat{\theta}_n$ is defined as any solution of

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

For this contrast to be well defined, operators Q and V must satisfy:

A 2. *The functions V_{Π_θ} and $Q_{\Pi_\theta^2}$ defined in (3) must be integrable 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.*

This assumption can be understood as “ Π_θ^* and its first two derivatives (resp. $(\Pi_\theta^2)^*$) have to be smooth enough compared to f_ε^* ”. For some models, this integrability assumption is not satisfied. In particular, for models where the functions Π_θ^* and $(\Pi_\theta^2)^*$ are less smooth than f_ε^* . In this case, we propose to insert a weight function or a truncation Kernel as in Dedecker et al. (2014) p. 285 to circumvent the issue of integrability. More precisely, define

$$Q_{\Pi_\theta, K}(x) = \frac{1}{2\pi} \int e^{ixu} \frac{(\Pi_\theta \star K)^*(u, 0)}{f_\varepsilon^*(-u)} du \quad \text{and} \quad V_{\Pi_\theta, K}^* = \frac{(\Pi_\theta \star K)^*}{\overline{f_\varepsilon^*} \otimes \overline{f_\varepsilon^*}},$$

where K^* is the Fourier transform of a density deconvolution kernel with compact support and satisfies $|1 - K^*(t)| \leq \mathbf{1}_{|t| > 1}$. This assumption is for instance a problem for Gaussian noise, and the weight function is then mandatory.

4 Asymptotic Properties of the Estimator

The following result 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 3.

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

(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 $j, k \in \{1, \dots, r\}$

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

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

$$\Sigma(\theta) = \mathcal{V}_\theta^{-1} \Omega(\theta) \mathcal{V}_\theta^{-1'} \quad \text{with} \quad \Omega(\theta) = \Omega_0(\theta) + 2 \sum_{j=2}^{+\infty} \Omega_{j-1}(\theta), \quad (6)$$

where $\Omega_0(\theta) = \text{Var}(\nabla_\theta m_\theta(\mathbf{Y}_1))$ and $\Omega_{j-1}(\theta) = \text{Cov}(\nabla_\theta m_\theta(\mathbf{Y}_1), \nabla_\theta m_\theta(\mathbf{Y}_j))$.

Theorem 4.1. Under Assumptions **A1–A3** let $\hat{\theta}_n$ be the least square estimator defined in (5), we have

$$\hat{\theta}_n \longrightarrow \theta_0 \quad \text{in probability as } n \rightarrow \infty.$$

Moreover,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}(0, \Sigma(\theta_0)) \quad \text{in law as } n \rightarrow \infty.$$

The proof of Theorem 4.1 is provided in Subsection 8.2.

The following corollary gives an expression of the matrices $\Omega(\theta_0)$ and \mathcal{V}_{θ_0} defined in $\Sigma(\theta)$ of Theorem 4.1.

Corollary 4.1. Under Assumptions **A1–A3**, 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

$$\begin{aligned} \Omega_0(\theta_0) = & \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}^2(Y_1)] + 4\mathbb{E}[V_{\nabla_\theta \Pi_\theta}^2(\mathbf{Y}_1)] \\ & - \left\{ \mathbb{E} \left(\int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right)^2 + 4\mathbb{E} \left(\nabla_\theta \Pi_\theta(\mathbf{X}_1) \right)^2 \right. \\ & \left. - 4\mathbb{E} \left(\int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right) \mathbb{E} \left(\nabla_\theta \Pi_\theta(\mathbf{X}_1) \right) \right\}, \end{aligned}$$

and, the covariance terms are given by

$$\begin{aligned} \Omega_{j-1}(\theta_0) = & \text{Cov} \left(\int \nabla_{\theta} \Pi_{\theta}^2(X_1, y) dy, \int \nabla_{\theta} \Pi_{\theta}^2(X_j, y) dy \right) \\ & + 4 \left(\text{Cov}(\nabla_{\theta} \Pi_{\theta}(\mathbf{X}_1), \nabla_{\theta} \Pi_{\theta}(\mathbf{X}_j)) - \text{Cov} \left(\int \nabla_{\theta} \Pi_{\theta}^2(X_1, y) dy, \nabla_{\theta} \Pi_{\theta}(\mathbf{X}_j) \right) \right), \end{aligned}$$

where the differential $\nabla_{\theta} \Pi_{\theta}$ is taken at point $\theta = \theta_0$.

Furthermore, the Hessian matrix \mathcal{V}_{θ_0} is given by

$$\left([\mathcal{V}_{\theta_0}]_{j,k} \right)_{j,k} = 2 \left(\left\langle \frac{\partial \Pi_{\theta}}{\partial \theta_k}, \frac{\partial \Pi_{\theta}}{\partial \theta_j} \right\rangle_f \right)_{j,k} \quad \text{at the point } \theta = \theta_0.$$

The proof of Corollary 4.1 is given in Subsection 8.3.

Sketch of proof: Let us now state the strategy of the proof, the full proof is given in Appendix 8. Clearly, the proof of Theorem 4.1 relies on M-estimators properties and on the deconvolution strategy. The existence of our estimator follows from regularity properties of the function Π_{θ} and compactness argument of the parameter space, it is explained in Appendix 8.2.1. The key ingredient of the proof consists in proving the asymptotic properties of our estimator. This is done by splitting the proof into two parts: we first give the consistency result in Appendix 8.2.2 and then give the asymptotic normality in Appendix 8.2.3. Let us introduce the principal arguments:

The main idea for proving the consistency of a M-estimator comes from the following observation: if $\mathbf{P}_n m_{\theta}$ converges to $\mathbf{P} m_{\theta}$ in probability, and if the true parameter solves the limit minimization problem, then, the limit of the argument of the minimum $\hat{\theta}_n$ is θ_0 . By using the uniform convergence in probability and the compactness of the parameter space, we show that the argmin of the limit is the limit of the argmin. Combining these arguments with the dominance argument **A3(i)** give the consistency of our estimator, and then, the first part of Theorem 4.1.

The asymptotic normality follows essentially from CLT for mixing processes (see Jones (2004)). Thanks to the consistency, the proof is based on a moment condition of the Jacobian vector of the function $m_{\theta}(\mathbf{y}) = Q_{\Pi_{\theta}^2}(y) - V_{\Pi_{\theta}}(\mathbf{y})$ and on a local dominance condition of its Hessian matrix. These conditions are given in **A3(ii)** and **A3(iii)**. To refer to likelihood results, one can see these assumptions as a moment condition of the score function and a local dominance condition of the Hessian.

Remark 4.1. *It should be noted that the construction of the estimator defined in (5) does not need to know the stationary density and assumptions **A3** need only its regularity. The exception is the identifiability assumption **A1(iv)** which appears to be non-trivial and in particular the second part of this assumption since the first part concerns the uniqueness of $\mathbf{P}m_\theta$ which is strictly convex w.r.t. θ . Nevertheless, the second part implies computing the Hessian matrix of $\mathbf{P}m_\theta$ to obtain the asymptotic variance matrix and this computation needs to know the stationary density by our definition of the norm. Therefore, we propose for the computation of the confidence bounds a consistent estimator of this variance matrix in Appendix 8.4.*

5 Simulations

5.1 The Models

We start from this following HMM

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \quad (7)$$

where the noises ε_i and the innovations η_i are supposed to be *i.i.d.* centered random variables with variance respectively σ_ε^2 and $\sigma_{0,\eta}^2$.

Here, the unknown vector of parameters is $\theta_0 = (\phi_0, \sigma_{0,\eta}^2)$ and for stationary and ergodic properties of the process X_i , we assume that the parameter ϕ_0 satisfies $|\phi_0| < 1$ (see Doukhan (1994)). In this setting, the transition function $\Pi_{\theta_0}(x, y)$ is also Gaussian with mean equal to $\phi_0 x$ and variance $\sigma_{0,\eta}^2$. And, to analyse the influence of the regularity of the density of observation noises on parameter estimation, we consider three types of noises:

Case 1: Gaussian noise (super-smooth). The density of ε_1 is given by

$$f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_\varepsilon^2}\right).$$

We have $f_\varepsilon^*(x) = \exp(-\sigma_\varepsilon^2 x^2/2)$. The vector of parameters θ_0 belongs to the compact subset Θ given by $\Theta = [-1 + r; 1 - r] \times [\sigma_{\min}^2; \sigma_{\max}^2]$ with $\sigma_{\min}^2 \geq \sigma_\varepsilon^2 + \bar{r}$ where $r, \bar{r}, \sigma_{\min}^2$ and σ_{\max}^2 are positive real constants. We consider this subset since by construction the operators $Q_{\Pi_\theta^2}$ and V_{Π_θ} are well defined for $\sigma_{0,\eta}^2 > \sigma_\varepsilon^2$.

Case 2: Laplace noise (ordinary smooth). The density of ε_1 is given by

$$f_\varepsilon(x) = \frac{1}{\sqrt{2}\sigma_\varepsilon} \exp\left(-\frac{\sqrt{2}}{\sigma_\varepsilon}|x|\right).$$

It satisfies $f_\varepsilon^*(x) = 1/(1 + \sigma_\varepsilon^2 x^2/2)$.

Case 3: $\log - \mathcal{X}^2$ noise (super smooth). The density of ε_1 is given by

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{x}{2} - \frac{1}{2} \exp(x)\right).$$

We have $f_\varepsilon^*(x) = (1/\sqrt{\pi})2^{ix}\Gamma(1/2 + ix)e^{-i\mathcal{E}x}$, where $\mathcal{E} = \mathbb{E}[\log(\xi_{i+1}^2)]$ and $\text{Var}[\log(\xi_{i+1}^2)] = \sigma_\varepsilon^2 = \pi^2/2$, and $\Gamma(x)$ denotes the gamma function defined $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. For the cases 2 and 3 the vector of parameters $\theta = (\phi, \sigma^2)$ belongs to the compact subset Θ given by $[-1+r; 1-r] \times [\sigma_{min}^2; \sigma_{max}^2]$ with r, σ_{min}^2 and σ_{max}^2 positive real constants. Furthermore, this latter case corresponds to the SV model introduced by Taylor (2005)

$$\begin{cases} R_{i+1} = \exp\left(\frac{X_{i+1}}{2}\right) \xi_{i+1}^\beta, \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases}$$

where β denotes a positive constant. The noises ξ_{i+1} and η_{i+1} are two centered Gaussian random variables with standard variance σ_ε^2 and σ_0^2 .

In the original paper of Taylor (1994), the constant β is equal to 1². In this case, by applying a log transformation $Y_{i+1} = \log(R_{i+1}^2) - \mathbb{E}[\log(\xi_{i+1}^2)]$ and $\varepsilon_{i+1} = \log(\xi_{i+1}^2) - \mathbb{E}[\log(\xi_{i+1}^2)]$, the log-transform SV model is a special case of the model (7). We postponed in Appendix 8.4 the verification of Theorem 4.1 assumptions for these three models.

5.2 Expression of the Contrasts

For all the models described above, we can express the theoretical and empirical contrasts regardless of the type of observation noise used. These expressions are given in the following proposition

Proposition 5.1. *For the HMM model (7) the theoretical contrast defined in (4) is given by*

$$\mathbf{P}m_\theta = -1 + \frac{\sigma_\eta + \sigma_{0,\eta}}{2\sqrt{\pi}\sigma_\eta\sigma_{0,\eta}} - \sqrt{\frac{2(\phi_0^2 - 1)}{\pi\sigma_\eta^2(-1 - \phi_0^2) - \pi\sigma_{0,\eta}^2(1 + \phi^2 - 2\phi\phi_0)}}, \quad (8)$$

and the empirical contrasts used in our simulations are obtained as follows

$$\mathbf{P}_n m_\theta^G = \frac{1}{2\sqrt{\pi}\sigma_\eta} - \frac{1}{n} \sqrt{\frac{2}{\pi(\sigma_\eta^2 - (\phi^2 + 1)\sigma_\varepsilon^2)}} \sum_{i=1}^{n-1} \exp\left(-\frac{\mathcal{A}_i}{2(\sigma_\eta^2 - (\phi^2 + 1)\sigma_\varepsilon^2)}\right),$$

²We argue that our approach can be applied when we introduce a long mean parameter μ into the volatility process.

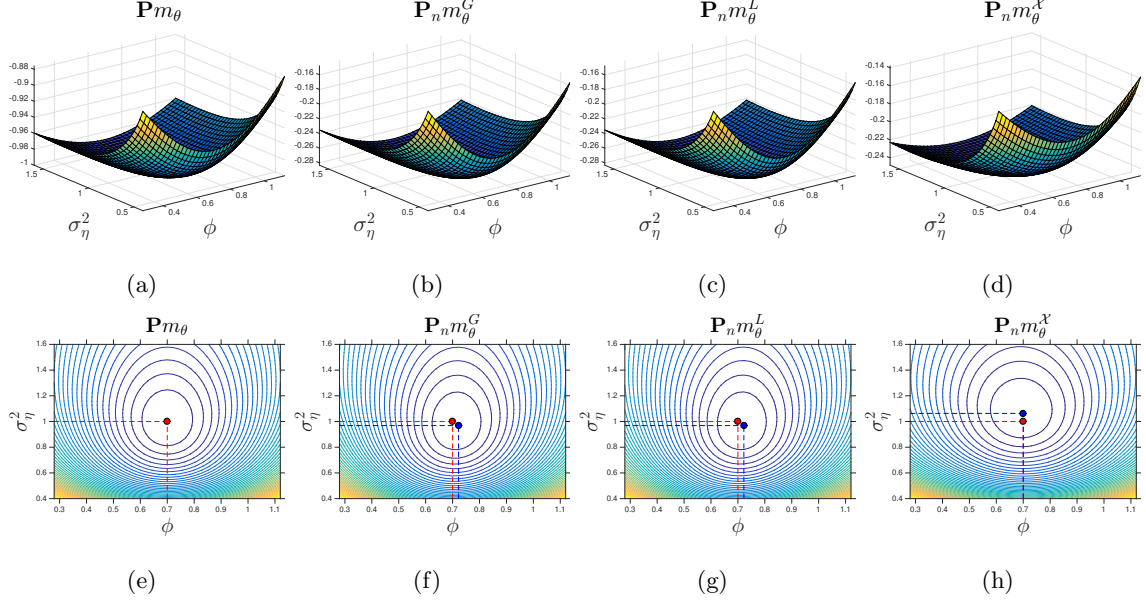


Figure 1: Contrast functions. (a): $\mathbf{P}m_\theta$ as a function of the parameters ϕ and σ_η^2 for one realization of (7), with $n = 2000$. (b): $\mathbf{P}_n m_\theta^G$. (c): $\mathbf{P}_n m_\theta^L$. (d): $\mathbf{P}_n m_\theta^X$. (e)-(h): Corresponding contour lines. The red circle represents the global minimizer θ_0 of $\mathbf{P}m_\theta$ and the blue circle, the one of $\mathbf{P}_n m_\theta^G$, $\mathbf{P}_n m_\theta^L$ and $\mathbf{P}_n m_\theta^X$ respectively.

$$\mathbf{P}_n m_\theta^L = \frac{1}{2\sqrt{\pi}\sigma_\eta} - \frac{1}{2n\sqrt{2\pi}\sigma_\eta^9} \sum_{i=1}^{n-1} \exp\left(-\frac{\mathcal{A}_i}{2\sigma_\eta^2}\right) \left(4\sigma_\eta^8 - 2(1 + \phi^2)\sigma_\eta^4\sigma_\varepsilon^2(\mathcal{A}_i - \sigma_\eta^2) + \phi^2\sigma_\varepsilon^4(\mathcal{A}_i^2 - 6\mathcal{A}_i\sigma_\eta^2 + 3\sigma_\eta^4)\right),$$

where $\mathcal{A}_i = (Y_{i+1} - \phi Y_i)^2$.

$$\mathbf{P}_n^X m_\theta = \frac{1}{2\sqrt{\pi}\sigma_\eta} - \frac{2}{n} \sum_{i=1}^{n-1} V_{\Pi_\theta}^*(Y_i, Y_{i+1})$$

with $V_{\Pi_\theta}(x, y) = \frac{1}{4\pi^2} \frac{e^{-\sigma_\eta^2 y^2} \mathbf{1}_{(x+\phi y)}}{\exp(i\mathcal{E}x)2^{-ix}\Gamma(\frac{1}{2}-ix) \exp(i\mathcal{E}y)2^{-iy}\Gamma(\frac{1}{2}-iy)}$ where $\mathbf{1}^*(x)$ denotes the Fourier transform of the indicator function and computed in practice over a large truncated interval such that

$$\mathbf{1}^*(x) = \begin{cases} T \operatorname{sinc}\left(\frac{Tx}{2}\right) & \text{for } T \text{ large and } x \neq 0 \\ T & \text{for } x = 0 \end{cases}$$

The notation $\mathbf{P}_n m_\theta^G$, $\mathbf{P}_n m_\theta^L$ and $\mathbf{P}_n m_\theta^X$ corresponds to the case 1, case 2 and case 3 respectively (see Figure 1 for a visual representation of these contrasts).

Figure 1 depicts the three empirical and true contrast curves as a function of the parameters ϕ and σ_η^2 for one realization of (7), with $n = 2000$. It can be observed that these three empirical contrasts give reliable estimates for the theoretical contrast, and in turn, also a high-quality estimate of the optimal parameter for the three types of noise distributions.

5.3 Monte Carlo Simulation and Comparison with the MLE

Let us present the results of our simulation experiments. First, we perform a MC study for the first two cases to analyse the influence of the noise regularity on the performance of the estimate (case 3 has the same regularity as case 1). For the first case, we compare our approach to the MLE. This case is favorable for the MLE since its calculation is fast via the Kalman filter. Indeed, the linearity of the model and the gaussianity of the observation noises make the use of the Kalman filter suitable for the computation of the MLE. However, this is no longer the case for Laplace's noises. For noises other than Gaussian noises, the calculation of the MLE is more complicated in practice and requires the use of algorithms such as MCMC, EM, SAEM, SMC or alternative estimation strategies (see Andrieu et al. (2010), Chopin et al. (2013) and Olsson and Rydén (2007)), which require a longer computation time. Moreover, it is important to note that even in the most favourable case for calculating the MLE our approach is faster than the Kalman filter as it only requires the minimization of an explicitly known contrast function.

For each simulation, we consider three different signal-to-noise ratios denoted SNR (i.e. $\text{SNR} = \sigma_\eta^2/\sigma_\varepsilon^2 = 1/\sigma_\varepsilon^2$, with σ_ε^2 equals to $1/40, 1/20$ and $1/10$ corresponding to low, medium and high noise level). For all experiments, we set $\phi = 0.7$ and having in mind financial and economic setting we generate samples of different sizes (i.e. $n = 500$ up to 2000). We represent the results obtained, on Boxplots for each parameter and for 100 repetitions in Figure 2. We can see that, as already noticed in the deconvolution setting there is little difference between Laplace and Gaussian ε_i 's. The convergence is slightly better for Laplace noise. Moreover, increasing the sample size leads to noticeable improvements of the results and in all cases and for all the noise distributions the estimation is good. The MLE provides better overall results which is not surprising since in the case 1 it is the best unbiased linear estimator but we can see that for the parameter ϕ our estimator is slightly better. The estimation of the variance σ_η^2 is more difficult and this is all the more true as the SNR decreases.

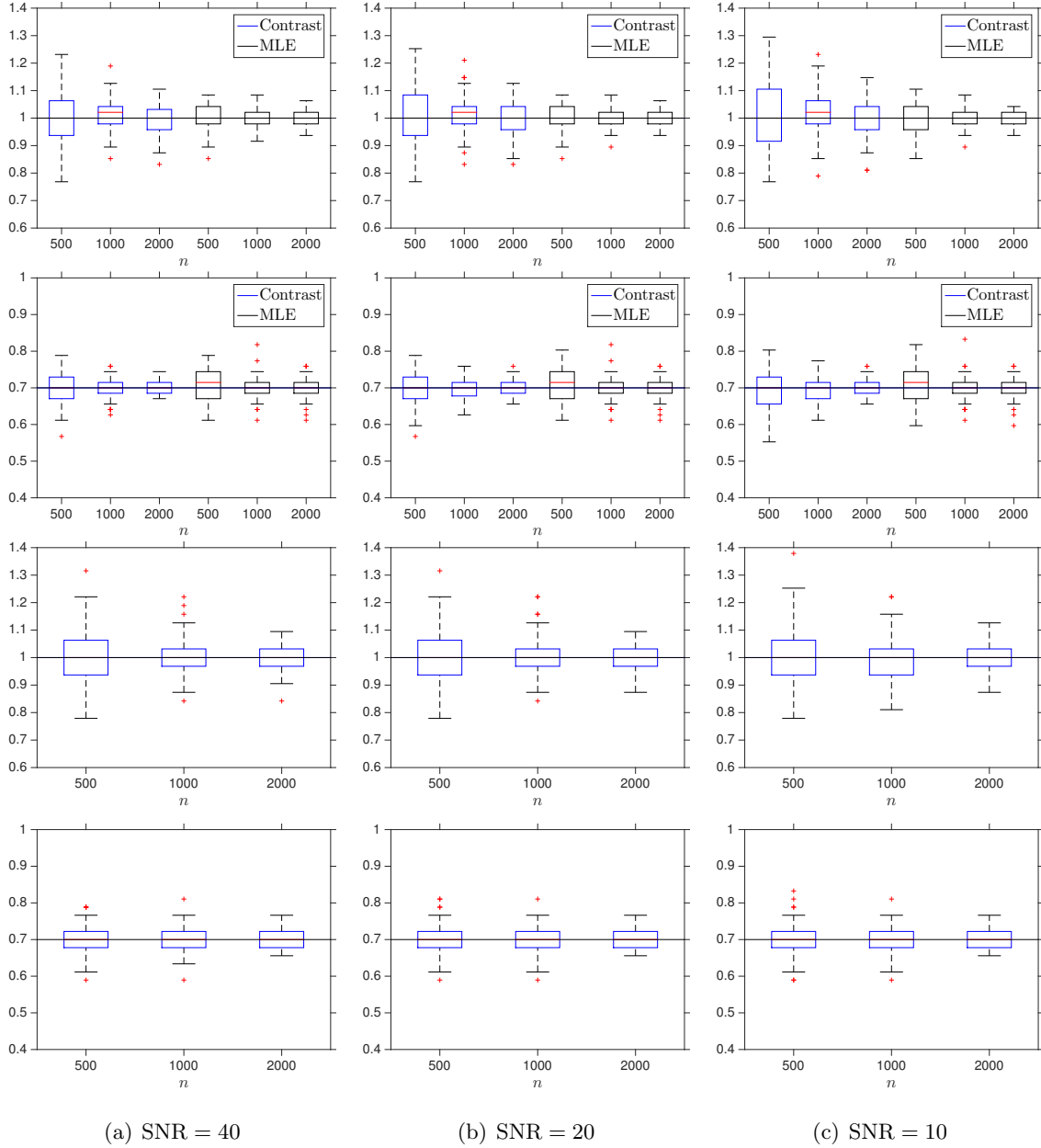


Figure 2: Boxplot of parameter estimators for different sample sizes (first two rows: Case 1, last two rows: Case 2). The horizontal lines represent the values of true parameters σ_η^2 and ϕ . Our proposed contrast estimator is represented by the blue boxes, and the Kalman MLE by the black ones.

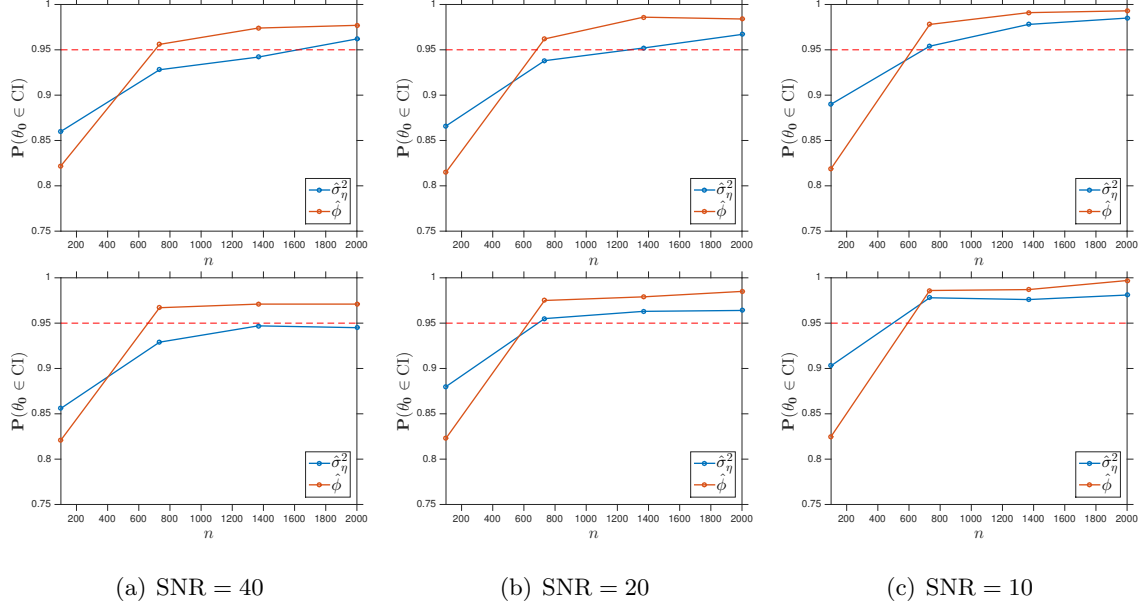


Figure 3: Coverage probability for 95% CIs versus sample size n (Case 1 top Case 2 bottom) based on 1000 simulations.

Our estimation procedure allows us to deepen our analysis since Theorem 4.1 applies and as we already mentioned, Corollary 4.1 allows to compute CIs in practice: For all $i = 1, 2$:

$$\mathbb{P} \left(\hat{\theta}_{n,i} - z_{1-\alpha/2} \sqrt{\frac{\mathbf{e}_i' \Sigma(\hat{\theta}_n) \mathbf{e}_i}{n}} \leq \theta_{0,i} \leq \hat{\theta}_{n,i} + z_{1-\alpha/2} \sqrt{\frac{\mathbf{e}_i' \Sigma(\hat{\theta}_n) \mathbf{e}_i}{n}} \right) \rightarrow 1 - \alpha,$$

as $n \rightarrow \infty$ where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the Gaussian law, $\theta_{0,i}$ is the i^{th} coordinate of θ_0 and \mathbf{e}_i is the i^{th} coordinate of the vector of the canonical basis of \mathbb{R}^2 . The covariance matrix $\Sigma(\hat{\theta}_n)$ is computed by plug-in the variance matrix defined in (11) (see Appendix 8.4) in (6). We investigate the coverage probabilities for sample size ranging from 100 to 2000 with a replication number of 1000. We compute the CI for each sample and plot the proportion of samples for which the true parameters are contained in the CI for the two cases 1 and 2. The results are shown in the Figure 3. It can be seen that the computed proportions provide good estimators for the empirical coverage probability for the CIs of both parameters whatever the type of noise. We can also see that these proportions deviate slightly from the theoretical value as the noise level increases. Indeed, the size of the ICs tends to increase as the noise level increases.

6 Application on the Ex-Ante Real Interest Rate

Ex-ante real interest rate is important in finance and economics because it provides a measure of the real return on an asset between now and the future. It is a fruitful indicator of the monetary policy direction of central banks. Nevertheless, it is important to make a distinction between the ex-post real rate, which is the observed series, and the ex-ante real rate which is unobserved. While the ex-post real rate is simply the difference between the observed nominal interest rate and the observed actual inflation, the ex-ante real rate is defined as the difference between the nominal rate and the unobserved expected inflation rate. Since monetary policy makers cannot observe inflation within the period, they must establish their interest rate decisions on the expected inflation rate. Hence, the ex-ante real rate is probably a better indicator of the monetary policy orientation.

There are two different strategies for the estimation of the ex-ante real rate. The first consists in using a proxy variable for the ex-ante real rate (see, e.g., Holston et al. (2017) for the US region and recently in Laubach and Williams (2003) for Canada, the Euro Area and the UK). The second strategy consists in treating the ex-ante real rate as an unknown variable using a latent factor model (see, e.g., Burmeister et al. (1986), Hamilton (1994b) Hamilton (1994a)). Our procedure is in line with this second strategy in which the factor model is specified as follows

$$\begin{aligned} Y_t &= \alpha + X_t + \varepsilon_t \\ X_t &= \phi X_{t-1} + \eta_t \end{aligned} \tag{9}$$

where Y_t is the observed ex-post real interest rate, X_t is the latent ex-ante real interest rate adjusted by a parameter α , ϕ a parameter of persistence and ε_t (resp. η_t) centered Gaussian random variables with variance σ_ε^2 (resp. σ_η^2).

This modelization comes from the fact that if we denote by Y_t^e the ex-ante real interest rate, we have that $Y_t^e = \mathcal{R}_t - \mathcal{I}_t^e$ with \mathcal{R}_t the observed nominal interest rate and \mathcal{I}_t^e the expected inflation rate. So, the unobserved part of Y_t^e comes from the expected inflation rate. Furthermore, the ex-post real interest rate Y_t is obtained from $Y_t = \mathcal{R}_t - \mathcal{I}_t$ with \mathcal{I}_t the observed actual inflation rate. Hence, expanding these expressions to allow for expected inflation rate \mathcal{I}_t^e gives

$$Y_t = \mathcal{R}_t - \mathcal{I}_t^e + \mathcal{I}_t^e - \mathcal{I}_t = Y_t^e + \varepsilon_t$$

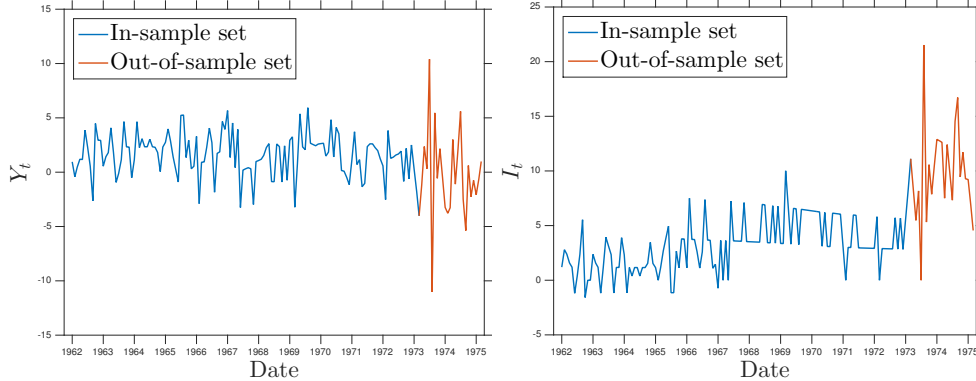


Figure 4: Observed ex-post real rate (left) and observed actual inflation rate (right).

where $\varepsilon_t = \mathcal{I}_t^e - \mathcal{I}_t$ is the inflation expectations random variables. If people do not make systematic errors in forecasting inflation, then ε_t , might reasonably be assumed to be a Gaussian white noise with variance denotes by σ_ε^2 . This assumption is known as the Rational Expectation Hypothesis (REH). Thus, the REH is very naturally suitable for a state-space representation (see, e.g., Burmeister and Wall (1982)) and defining $X_t = Y_t^e - \alpha$ as the ex-ante real interest rate adjusted by its population mean α gives the modelization (9).

The dataset was split into in and out-of-sample monthly data sets. The in-sample contained 75% (it ranges from 1 January 1962 to 1 March 1973) of the total dataset and the out-of-sample the remaining 25% (from 1 April 1973 to 1 March 1975, i.e. for a horizon of two years). This places us in the period of great inflation for the US region. More precisely, the ex-post real interest rate Y_t (represented in Figure 4(a)) is then computed as the difference between the annualized nominal funds rate and the annualized percentage inflation rate I_t (depicted in Figure 4(b)) obtained from the consumer price index for all urban consumers. The data are available from the Federal Reserve Economic Data. We consider two of the models previously studied, the Gaussian AR model and the SV model.

Let us denote $\theta = (\phi, \sigma_\eta^2)$ the vector of unknown parameters to be estimated. We make a two-stage estimation: in the first step we estimate the unknown parameter vector θ on the in-sample by minimizing the contrast introduced in this paper. The second step is devoted to the ex-ante real rate forecasts and the expected inflation rate forecasts on the out-of-sample by plugging $\hat{\theta}$ obtained from the in-sample set in the first stage by running a Kalman filter. So that all forecasts are computed using the pseudo out-of-sample.

The value of the estimation obtained in the first step is as follows: $\hat{\alpha}_{\text{AR}} = 1.5699$,

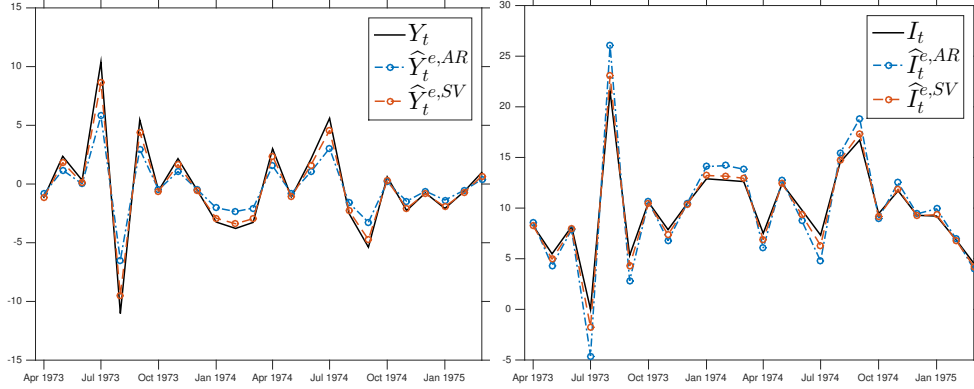


Figure 5: (Ex-post real rate forecasts (left) and expected inflation rate forecasts (right).

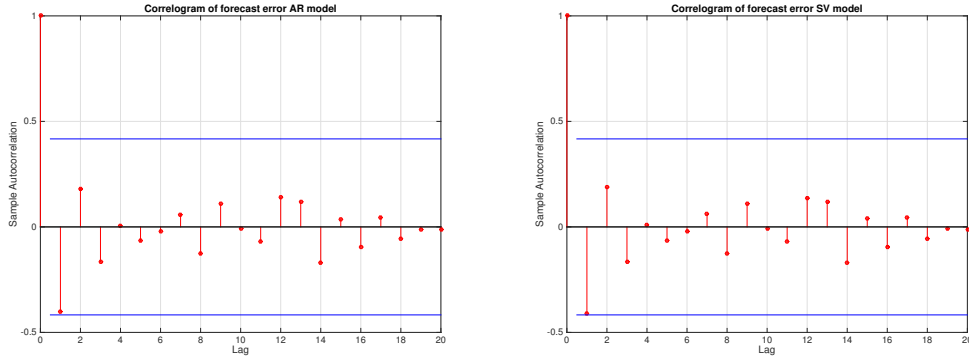


Figure 6: Correlogram of \hat{e}_t for AR model (left) and of \hat{e}_t for SV model (right).

$\hat{\alpha}_{SV} = 2.1895$, $\hat{\phi}_{AR} = 0.5750$, $\hat{\phi}_{SV} = 0.8500$, $\hat{\sigma}_{AR}^2 = 4.4048$ and $\hat{\sigma}_{SV}^2 = 3.1756$ and the forecasts obtained from the out-of-sample are shown in Figure 5. A first examination of our results reveals that our forecasts expected inflation series is plausible for the two models. Nevertheless, the results are better for the SV model: the Mean Squared Forecast Error is divided by a factor 10 for the SV model. The Kolmogorov-Smirnov test suggests that during the great inflation period the data are no longer Gaussian and exhibit more like a distribution with heavy tails. The null hypothesis is rejected at level $\beta = 0.05$ but is accepted at level $\beta = 0.1$. This result may explain why the SV model is significantly better than the AR model. The mean of the forecast error $\hat{e}_t = \mathcal{I}_t - \hat{\mathcal{I}}_t^e$ is sufficiently close to zero and the Ljung box test accepts the null hypothesis, meaning that the forecast errors are not correlated for the two models. The correlograms for the two models are given in Figure 6. These results are consistent with the rational expectation hypothesis. If one

compares the results of the parameters estimation for the two models one can see that the persistence parameter ϕ is higher for the SV model than the AR model, on the other hand, the variance is lower. Therefore, the variance of $\widehat{\mathcal{I}}_t^e$ is smaller than that of \mathcal{I}_t for the SV model. These results are consistent with the economically intuitive notion that expectations are smoother than realizations. Most importantly, these results corroborate those of the thorough analysis in Stock and Watson (2007) and Pivetta and Reis (2007) whose find that the persistence parameter is high and close to one for this period of study.

7 Conclusion

In this paper we propose a new estimation strategy for non-linear and non-Gaussian HMM models. This methodology makes it possible to bypass the MLE estimate known to be difficult to calculate for these models. Our approach is not based on MC methods (MCMC or particle filtering methods), which avoids the instability problems of most of the proposed methods when minimizing the criterion following MC errors (see Doucet et al. (2001)). The parameter estimation step in HMM models is very important since it is shown in El Kolei and Patras (2018) that the bias on the parameters propagates in the estimation of the latent variable. This cannot be overlooked in models where the latent variable to be predicted is used to make political decisions. For example, in this paper, we have looked at the prediction of the ex-ante interest rates and show that during the period of high inflation the annualized inflation rate exhibits distribution with heavy tails. Thus, in this context the SV model seems more appropriate and gives better results. Nevertheless, since this model is no longer Gaussian it seems important to study estimators other than the MLE that cannot be calculated by the Kalman filter. In this context, we provide a new easily way to estimate the parameters in a Gaussian and non-Gaussian setting. This provides an alternative estimation method to those proposed in the literature that are largely based on MC methods.

8 Proofs

8.1 Proof of Lemma 3.1

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

1. Let us set

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

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

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

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

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

Hence

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

The point 1. is proved.

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

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

Hence

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

By using the independence between X_i and ε_i , we have

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

Thus

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

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

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

So

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

The point 2. is proved. \square

8.2 Proofs of Theorem 4.1

For the reader convenience we split the proof of Theorem 4.1 into three parts: in Subsection 8.2.1, we give the proof of the existence of our contrast estimator defined in (3). In Subsection 8.2.2, we prove the consistency, that is, the first part of Theorem 4.1. Then, we prove the asymptotic normality of our estimator in Subsection 8.2.3, that is, the second part of Theorem 4.1. The Section 8.3 is devoted to Corollary 4.1.

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

8.2.2 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 8.1. *Uniform Law of Large Numbers (see Newey and McFadden (1994a) 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

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

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 **A3**(i) implies that $\mathbb{E} [\sup_{\theta \in \Theta} |m_\theta(\mathbf{Y}_i)|]$ is finite. Indeed, by assumption **A3**(i), we have

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

Lemma 8.1 gives us the uniform convergence in probability of the contrast function: for any $\varepsilon > 0$:

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

Combining the uniform convergence with (Newey and McFadden, 1994b, Theorem 2.1 p. 2121 chapter 36) yields the weak (convergence in probability) consistency of the estimator. \square

Remark 8.1. *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 Hayashi (2000) for this discussion.*

8.2.3 Proof of the asymptotic normality

For the CLT, we need to define the α -mixing property of a process (we refer the reader to Doukhan (1994) for a complete review of mixing processes).

Definition 8.1 (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 $\mathcal{F}_k^m = \sigma(Y_k, \dots, Y_m)$. The sequence Y is said to be α -mixing if $\alpha(n) \rightarrow 0$ as $n \rightarrow \infty$, where*

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

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

Lemma 8.2. *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 $j \in \{1, \dots, r\}$

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

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

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

Then, $\widehat{\theta}_n$ defined in (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}_{θ_0} is the Hessian of the mapping $\mathbf{P}m_\theta$ given in (4).

Proof. The proof follows from Proposition 7.8 p. 472 of Hayashi (2000) and Jones (2004), and by using the fact that, by regularity assumptions **A1**(i) and the Lebesgue Differentiation Theorem, we have $\mathbb{E}[\nabla_\theta^2 m_\theta(\mathbf{Y}_1)] = \nabla_\theta^2 \mathbb{E}[m_\theta(\mathbf{Y}_1)]$. \square

It just remains to check that the conditions (ii) and (iii) of Lemma 8.2 hold under our assumptions **A3**(ii) and **A**(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 mapping $m_\theta(\mathbf{y}_i) : \theta \in \Theta \mapsto m_\theta(\mathbf{y}_i) = Q_{\Pi_\theta^2}(\mathbf{y}_i) - 2V_{\Pi_\theta}(\mathbf{y}_i)$ is twice continuously differentiable for all $\theta \in \Theta$ and its first derivatives are given by

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

By assumption, for each $j \in \{1, \dots, r\}$, $\frac{\partial \Pi_\theta}{\partial \theta_j}$ 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

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \left[Q_{\nabla_\theta \Pi_\theta^2}(\mathbf{y}_i) - 2V_{\nabla_\theta \Pi_\theta}(\mathbf{y}_i) \right]. \quad (10)$$

Then, for some $\delta > 0$, by the moment assumption **A3**(ii), we have

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

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 $L_1(\mathbb{R})$, 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 **A3**(iii),

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

This ends the proof of Theorem 4.1. \square

8.3 Proof of Corollary 4.1

By replacing $\nabla_\theta m_\theta(\mathbf{Y}_1)$ by its expression (10), we have

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

Owing to Lemma 3.1, we obtain

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

In a similar manner, using again Lemma 3.1, we have

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

and

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

Hence

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

Calculus of the covariance matrix of Corollary 4.1: By replacing $\nabla_{\theta}m_{\theta}(Y_1)$ by its expression (10), we have

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

It follows from Lemma 3.1 and the stationarity assumption $\mathbf{A1}(iv)$ of $(Y_i)_{i \geq 1}$ that

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

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

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

On the other hand, we have

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

Furthermore, the Tower property yields

$$\mathbb{E}[Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1)Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_j)] = \mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_1, y)dy \int \nabla_{\theta}\Pi_{\theta}^2(X_j, y)dy\right].$$

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 $(Y_i)_{i \geq 1}$,

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

By using Lemma 3.1, the last term is equal to

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

Therefore, the covariance matrix is given by

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

Thus

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

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

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

Under **A1(i)**, for all θ in Θ , the mapping $\theta \mapsto \mathbf{P}m_{\theta}$ 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 $j, k \in \{1, \dots, r\}$:

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

The proof of Corollary 4.1 is completed. \square

8.4 Contrast and Checking Assumptions for the Simulations

Contrasts for the simulations: To compute the several contrasts defined in Proposition 8 the following quantities are essentially required: $(\Pi_{\theta}^2(x, 0))^*$, $\Pi_{\theta}^*(x, y)$ and $f_{\varepsilon}^*(x)$. For

the model defined in (7), the square of the transition density is also Gaussian up to the parameter $1/(2\sqrt{\pi\sigma_\eta^2})$ with mean ϕx and variance $\sigma_\eta^2/2$. Hence, if we are interested in computing the following Fourier transform

$$(\Pi_\theta^2(x, 0))^* = \int e^{-ixu} \left(\int \Pi_\theta^2(u, v) dv \right) du = \int e^{-ixu} \tilde{\Pi}_\theta(u) du = (\tilde{\Pi}_\theta(x))^*.$$

And we have $\tilde{\Pi}_\theta(x) = 1/(2\sqrt{\pi\sigma_\eta^2})$ by integration of the Gaussian density. Hence,

$$(\Pi_\theta^2(x, 0))^* = 1/(2\sqrt{\pi\sigma_\eta^2}) \int e^{-ixu} = 1/(2\sqrt{\pi\sigma_\eta^2}) \mathbf{1}^*(x).$$

For $\Pi_\theta^*(x, y)$ the computation is the same and given by

$$\begin{aligned} (\Pi_\theta(x, y))^* &= \int e^{-ixu} \left(\int e^{-iyv} \int \Pi_\theta(u, v) dv \right) du \\ &= \int e^{-ixu} \left(\int e^{-iyv} \mathcal{N}(v, \phi u, \sigma_\eta) du dv \right) \\ &= \int e^{-ixu} \left(\mathcal{N}(y, \phi u, \sigma_\eta) \right)^* du = e^{-\frac{\sigma_\eta^2}{2}y^2} \int e^{-iu(x+\phi y)} du \\ &= e^{-\frac{\sigma_\eta^2}{2}y^2} \mathbf{1}^*(x + \phi y), \end{aligned}$$

where $\mathcal{N}(v, a, b)$ stands for the Gaussian density computed at the point v with mean a and standard deviation b . Furthermore, for each simulation case the Fourier transform of the noise density are given in Section 5.

Checking assumptions A1–A3: By inspecting the functions $b_{\theta_0} : x \mapsto \phi_0 x$ one can easily see that regularity assumptions are well-satisfied and if ϕ_0 satisfies $|\phi_0| < 1$ the process is strictly stationary. It remains to check assumptions A1(iv) and A2–A3. The strict convexity of the function $\mathbf{P}m_\theta$ gives that θ_0 is a minimum and Assumption A1(iv) also requires to compute the Hessian matrix belonging to $\mathcal{S}_{2 \times 2}$ (where \mathcal{S} represents the space of symmetric matrix) and given by

$$\mathcal{V}_{\theta_0} = \begin{pmatrix} \left\| \frac{\partial \Pi_\theta}{\partial \phi} \right\|_f^2 & \left\langle \frac{\partial \Pi_\theta}{\partial \phi}, \frac{\partial \Pi_\theta}{\partial \sigma^2} \right\rangle_f \\ \left\langle \frac{\partial \Pi_\theta}{\partial \phi}, \frac{\partial \Pi_\theta}{\partial \sigma^2} \right\rangle_f & \left\| \frac{\partial \Pi_\theta}{\partial \sigma^2} \right\|_f^2 \end{pmatrix}$$

(see Corollary 4.1). The computation of this matrix can be easily done for AR(1) processes since all densities are explicit, and we have

$$\mathcal{V}_{\theta_0} = \frac{1}{4\sqrt{\pi\sigma_\eta^3}} \begin{pmatrix} -\frac{\sigma_{0,\eta}^2}{-1+\phi_0^2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

For processes with not an explicit expression one can use the following estimator

$$\mathcal{V}_{\hat{\theta}_n} = \frac{1}{n} \sum_{i=1}^{n-1} Q_{\frac{\partial^2 \Pi_{\hat{\theta}}^2}{\partial \theta^2}}(Y_i) - 2V_{\frac{\partial^2 \Pi_{\hat{\theta}}}{\partial \theta^2}}(\mathbf{Y}_i) \quad (11)$$

since under the integrability assumption **A2** and the Hessian local dominance assumption **A3(iii)** the matrix $\mathcal{V}_{\hat{\theta}_n}$ is a consistent estimator of \mathcal{V}_{θ_0} . For Gaussian noise the integrability assumption **A2** is not satisfied and as we explained in Section 2 the use of a kernel truncation or a weight function is then mandatory. The existence of Assumptions **A3** depends on the regularity of the following functions: Π_{θ}^* , $(\frac{\partial \Pi_{\theta}}{\partial \theta_j})^*$ and $(\frac{\partial^2 \Pi_{\theta}}{\partial \theta_j \partial \theta_l})^*$ (resp. $(\Pi_{\theta}^2)^*$, $(\frac{\partial \Pi_{\theta}^2}{\partial \theta_j})^*$ and $(\frac{\partial^2 \Pi_{\theta}^2}{\partial \theta_j \partial \theta_l})^*$) for $j, l \in \{1, \dots, r\}^2$ and especially on the type of error distribution. As explained in Comte et al. (2010) and Fan (1991), the convergence rate is determined by the tail behavior of the modulus of the characteristic function of the errors distribution: the faster the modulus function goes to zero in the tail, the slower the convergence. More specifically, one has:

Laplace noise: the convergence towards zero of the modulus of the characteristic function is polynomial. In addition, by the compactness of the parameter space Θ and since the functions $(\Pi_{\theta}^*/f_{\varepsilon}^*)$, $(\frac{\partial \Pi_{\theta}}{\partial \theta_j})^*/f_{\varepsilon}^*$ and $(\frac{\partial^2 \Pi_{\theta}}{\partial \theta_j \partial \theta_l})^*/f_{\varepsilon}^*$ have the following form $C_1(\theta)P(x) \exp(-C_2(\theta)x^2)$ (meaning that they are super-smooth) where $C_1(\theta)$ and $C_2(\theta)$ are two constants well-defined in the compact parameter set Θ and $P(x)$ a polynomial function independent of θ , we obtain that the moment condition and locale dominance assumptions **A3** are satisfied.

Gaussian noise/ log - χ^2 noise: convergence to zero of the modulus of the characteristic function of f_{ε} is exponential. In this case, the rate of convergence is slower than the one of the ordinary smooth case. This result fully agrees with recent advances in deconvolution theory: the smoother are the noises the slower the convergence (see Comte et al. (2010)).

References

- Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72(3), 269–342.
- Bos, C. S. and N. Shephard (2006). Inference for adaptive time series models: Stochastic volatility and conditionally gaussian state space form. *Economet. Rev.* 25(2-3), 219–244.
- Burmeister, E. and K. D. Wall (1982). Kalman filtering estimation of unobserved rational

- expectations with an application to the german hyperinflation. *J. Econometrics* 20(2), 255–284.
- Burmeister, E., K. D. Wall, and J. D. Hamilton (1986). Estimation of unobserved expected monthly inflation using kalman filtering. *J. Bus. Econ. Stat.* 4(2), 147–160.
- Butucea, C. and C. Matias (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* 11(2), 309–340.
- Chan, J. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. 35(1), 17–28.
- Chopin, N., P. E. Jacob, and O. Papaspiliopoulos (2013). Smc2: an efficient algorithm for sequential analysis of state space models. *J. R. Stat. Soc. Ser. B. Stat. Methodol* 75(3), 397–426.
- Comte, F., C. Lacour, and Y. Rozenholc (2010). Adaptive estimation of the dynamics of a discrete time stochastic volatility model. *J. Econometrics* 154(1), 59–73.
- Comte, F., Y. Rozenholc, and M.-L. Taupin (2006). Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics* 34(3), 431–452.
- Comte, F. and M.-L. Taupin (2001). Semiparametric estimation in the (auto)-regressive β -mixing model with errors-in-variables. *Math. Methods Statist.* 10(2), 121–160.
- Costa, M. and T. Alpuim (2010). Parameter estimation of state space models for univariate observations. *J. Statist. Plann. Inference* 140(7), 1889–1902.
- Dedecker, J., A. Samson, and M.-L. Taupin (2014). Estimation in autoregressive model with measurement error. *ESAIM Prob. Stat.* 18, 277–307.
- Douc, R., E. Moulines, J. Olsson, and R. van Handel (2011). Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.* 39(1), 474–513.
- Doucet, A., N. De Freitas, and N. Gordon (2001). An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pp. 3–14. Springer.
- Doukhan, P. (1994). *Mixing*, Volume 85 of *Lecture Notes in Statistics*. New York: Springer-Verlag. Properties and examples.

- El Kolei, S. (2013). Parametric estimation of hidden stochastic model by contrast minimization and deconvolution. *Metrika* 76(8), 1031–1081.
- El Kolei, S. and F. Patras (2018). Analysis, detection and correction of misspecified discrete time state space models. *J. Comput. Appl. Math.* 333, 200–214.
- El Kolei, S. and F. Pelgrin (2017). Parametric inference of autoregressive heteroscedastic models with errors in variables. *Statist. Probab. Lett.* 130, 63–70.
- Fan, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* 53(1), 97–110.
- Fan, J., Y. K. Truong, and Y. Wang (1991). Nonparametric function estimation involving errors-in-variables. In *Nonparametric Functional Estimation and Related Topics*, pp. 613–627. Springer.
- Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* 6(6), 1051–1079.
- Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. *Handbook of statistics* 14, 119–191.
- Hamilton, J. D. (1994a). State-space models. *Handbook of econometrics* 4, 3039–3080.
- Hamilton, J. D. (1994b). *Time series analysis*, Volume 2. Princeton university press Princeton, NJ.
- Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *J. Bus. Econ. Stat* 14(4), 429–434.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton, NJ.
- Holston, K., T. Laubach, and J. C. Williams (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics* 108, S59 – S75. 39th Annual NBER International Seminar on Macroeconomics.
- Jones, G. L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* 1, 299–320.
- Lacour, C. (2008). Least squares type estimation of the transition density of a particular hidden Markov chain. *Electron. J. Stat.* 2, 1–39.

- Laubach, T. and J. C. Williams (2003). Measuring the natural rate of interest. *The Review of Economics and Statistics* 85(4), 1063–1070.
- Mokkadem, A. (1990). Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.* 26(2), 219–260.
- Newey, W. K. and D. McFadden (1994a). Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, Volume 2 of *Handbooks in Econom.*, pp. 2111–2245. Amsterdam: North-Holland.
- Newey, W. K. and D. McFadden (1994b). Large sample estimation and hypothesis testing. *Handbook of econometrics* 4, 2111–2245.
- Olsson, J. and T. Rydén (2007). 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.*, pp. 115–120. EDP Sci., Les Ulis.
- Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: fast and efficient likelihood inference. *J. Econometrics* 140(2), 425–449.
- Pivetta, F. and R. Reis (2007). The persistence of inflation in the united states. *Journal of Economic dynamics and control* 31(4), 1326–1358.
- Staudenmayer, J. and J. P. Buonaccorsi (2005). Measurement error in linear autoregressive models. *J. Amer. Statist. Assoc.* 100(471), 841–852.
- Stock, J. H. and M. W. Watson (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking* 39, 3–33.
- Taylor, S. (2005). *Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices.*, Volume 1, pp. 203–226. Oxford University Press.
- Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4(2), 183–204.
- Van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.