



**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. 2021. hal-01598922v5

**HAL Id: hal-01598922**

**<https://hal.science/hal-01598922v5>**

Preprint submitted on 20 Oct 2021 (v5), 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\*, Salima El Kolei†, Fabien Navarro‡

October 20, 2021

## Abstract

This paper develops a simple and computationally efficient parametric approach to the estimation of general hidden Markov models (HMMs). For non-Gaussian HMMs, the computation of the maximum likelihood estimator (MLE) involves a high-dimensional integral that has no analytical solution and can be difficult to approach accurately. We develop a new alternative method based on the theory of estimating functions and a deconvolution strategy. Our procedure requires the same assumptions as the MLE and deconvolution estimators. We provide theoretical guarantees about the performance of the resulting estimator; its consistency and asymptotic normality are established. This leads to the construction of confidence intervals. Monte Carlo experiments are investigated and compared with the MLE. Finally, we illustrate our approach using real data for ex-ante interest rate forecasts.

*MSC 2010 subject classifications:* 62F12.

*Keywords:* Contrast function; deconvolution; hidden Markov models; least square estimation; parametric inference.

## 1 Introduction

In this paper, a hidden non-linear Markov model (HMM) with heteroskedastic noise is considered; we observe  $n$  random variables  $Y_1, \dots, Y_n$  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 known measurable functions  $b_{\theta_0}$  and  $\sigma_{\theta_0}$  up to the unknown parameter  $\theta_0$ . In addition to 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  $f_{\theta_0}$ . We assume that

---

\*Christophe Chesneau

Université de Caen - LMNO, France, E-mail: christophe.chesneau@unicaen.fr

†Salima El Kolei

CREST - ENSAI, France, E-mail: salima.el-kolei@ensai.fr

‡Fabien Navarro

Université Paris 1 Panthéon-Sorbonne - SAMM, France, E-mail: fabien.navarro@univ-paris1.fr

the transition distribution admits a density  $\Pi_{\theta_0}$ , defined by  $\Pi_{\theta_0}(x, x')dx' = \mathbb{P}_{\theta_0}(X_{i+1} \in dx' | X_i = x)$ . For the identifiability of (1), we assume that  $\varepsilon_1$  admits a known density with respect to the Lebesgue measure denoted by  $f_\varepsilon$ .

Our objective is to estimate the parameter vector  $\theta_0$  for non-linear HMMs with heteroskedastic innovations described by the function  $\sigma_{\theta_0}$  in (1) assuming that the model is correctly specified, *i.e.*,  $\theta_0$  belongs to the interior of a compact set  $\Theta \subset \mathbb{R}^r$ , with  $r \in \mathbb{N}^*$ .

Many articles have focused on parameter estimation and on the study of asymptotic properties of estimators when  $(X_i)_{i \geq 1}$  is an autoregressive moving average (ARMA) process (see [9], [51] and [12]) or in a regression context with measurement error (see, *e.g.*, [56], [43] or [24]). 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 [16], 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 (EM) approach or Monte Carlo-based methods (*e.g.*, Markov chain Monte Carlo (MCMC), Sequential Monte Carlo (SMC) or particle Markov chain Monte Carlo, see [1], [10] and [48]). 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; we consider the model (1) as a 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 [11]. The authors propose an estimation procedure based on least squares minimization. Recently, in [13], the authors generalize this approach to models defined as  $X_i = b_{\theta_0}(X_{i-1}) + \eta_i$ , where  $b_{\theta_0}$  is the regression function assumed to be known up to  $\theta_0$  and for homoscedastic innovations  $\eta_i$ . Also, in [21] and [23], the authors propose a consistent estimator for parametric models assuming knowledge of the stationary density  $f_{\theta_0}$  up to the unknown parameters  $\theta_0$  for the construction of the estimator. For many processes, this density has no analytic expression, 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 heteroskedasticity (ARCH) process is a family of processes for which transition density has a closed form as opposed to the stationary density. These processes are widely used to model economic or financial variables.

In this work, we aim to develop a new computationally efficient approach whose construction does not require the 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 heteroskedastic innovations, and our estimation principle is based on the contrast function proposed in a nonparametric context by [35, 3]. Thus, we propose to adapt their approach in a parametric context, assuming that the form of the transition density  $\Pi_{\theta_0}$  is known up to some unknown parameter  $\theta_0$ . The proposed methodology is purely parametric, and we go a step further by proposing an analytical expression of the asymptotic variance matrix  $\Sigma(\hat{\theta}_n)$ , which allows us to consider the construction of confidence intervals.

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 only Fourier transforms, as in [13]. In particular, a numerical illustration is given for the three following models: a Gaussian AR(1) model for which our approach can be well understood, an AR(1) process with Laplace's noise in order to study the influence of the smoothness of observation noise on the estimation of the parameters since it is known in deconvolution to affect the convergence rate (see, *e.g.*, [25]), and a stochastic volatility model (SV) also referred to as the unobserved components/stochastic volatility model (see, *e.g.*, [52], [8] and [19]). There is a large literature on the fitting of SV models (see, *e.g.*, the reviews in [27], [2] and [49]). All are based on Bayesian methods, and in particular, Markov chain Monte Carlo (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 provide a simulation study for both linear and nonlinear examples and in a Gaussian and non-Gaussian setting. We then compare the empirical performance of the proposed method with other methods from the literature. 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 [32] and more recently in [39] 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, an 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 in the parameters propagates in the estimation of the latent variable (see [22]). 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 regard, our approach therefore provides better results than the (quasi-)MLE estimate.

The remainder of the paper is organized as follows. We present our assumptions about the Markov chain in Section 2. Section 3 describes our estimator and its statistical properties, and also presents our main results: the consistency and asymptotic normality of the estimator. Simulated examples are provided in Section 4 and the real data application is in Section 5. The proofs are gathered in Section 6.

## 2 Framework

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

### 2.1 Notations

The Fourier transform of an integrable function  $u$  is denoted by  $u^*(t) = \int e^{-itx}u(x)dx$ , and it satisfies the equation  $(u^*)^*(x) = 2\pi u(-x)$ . We denote by  $\nabla_{\theta}g$  the vector of the partial derivatives of  $g$  with respect to (w.r.t.)  $\theta$ . The Hessian matrix of  $g$  w.r.t.  $\theta$  is denoted by  $\nabla_{\theta}^2g$ . For any matrix  $M = (M_{i,j})_{i,j}$ , the Frobenius norm is defined by

$\|M\| = \sqrt{\sum_i \sum_j |M_{i,j}|^2}$ . Finally, 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,  $\mathbb{P}, \mathbb{E}, \text{Var}$  and  $\text{Cov}$  denote respectively the probability  $\mathbb{P}_{\theta_0}$ , the expected value  $\mathbb{E}_{\theta_0}$ , the variance  $\text{Var}_{\theta_0}$  and the covariance  $\text{Cov}_{\theta_0}$  when the true parameter is  $\theta_0$ . 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  $\Pi_\theta$  on a compact subset  $A = A_1 \times A_2$ . For more clarity, we write  $\Pi_\theta$  instead of  $\Pi_\theta \mathbf{1}_A$  and we denote by  $\|\cdot\|_A$  (*resp.*  $\|\cdot\|_A^2$ ) the norm in  $\mathbb{L}_1(A)$  (*resp.*  $\mathbb{L}_2(A)$ ) defined as

$$\|u\|_A = \int \int |u(x, y)| f_{\theta_0}(x) \mathbf{1}_A(x, y) dx dy, \quad \|u\|_A^2 = \int \int u^2(x, y) f_{\theta_0}(x) \mathbf{1}_A(x, y) dx dy.$$

## 2.2 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  $\Pi_\theta$  belongs to  $\mathbb{L}_1(A) \cap \mathbb{L}_2(A)$  and is twice continuously differentiable w.r.t.  $\theta \in \Theta$  for any  $(x, x')$  and measurable w.r.t.  $(x, x')$  for all  $\theta$  in  $\Theta$ . 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  $\alpha$ -mixing with invariant density  $f_{\theta_0}$ .*

#### *Assumptions on the noise $\varepsilon_t$ and innovations $\eta_t$ :*

- (iii) – *The errors  $(\varepsilon_i)_i$  are independent and identically distributed (i.i.d.) centered random variables with finite variance,  $\mathbb{E}[\varepsilon_1^2] = \sigma_\varepsilon^2$ . The random variable  $\varepsilon_1$  admits a known density,  $f_\varepsilon$ , 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 i.i.d. centered random variables.*

#### *Identifiability assumptions:*

- (iv) *The mapping  $\theta \mapsto \mathbf{P}m_\theta = \|\Pi_\theta - \Pi_{\theta_0}\|_A^2 - \|\Pi_{\theta_0}\|_A^2$  with  $m_\theta$  defined in (2) admits a unique minimum at  $\theta_0$  and its Hessian matrix denoted by  $\mathcal{V}_\theta$  is non-singular in  $\theta_0$ .*

An entire section has been devoted to commenting on all the necessary assumptions (see Section 3.3)

## 3 Estimation Procedure and Main Results

### 3.1 Least Squares Contrast Estimation

A key ingredient in the construction of our estimator of the parameter  $\theta_0$  is the choice of a “contrast function” depending on the data. Details about contrast estimators can be found in [55]. For the purpose of this study, we consider the contrast function initially

introduced by [35] in a nonparametric setting, inspired by regression-type contrasts and later used in various works (see, *e.g.*, [3, 36, 37, 38]), that is

$$\mathbf{P}_n m_\theta = \frac{1}{n} \sum_{i=1}^{n-1} m_\theta(\mathbf{y}_i). \quad (2)$$

with  $m_\theta : \mathbf{y} \mapsto Q_{\Pi_\theta^2}(y) - 2V_{\Pi_\theta}(\mathbf{y})$ . The operators  $Q$  and  $V$  are defined for any function  $h \in \mathbb{L}_1(A) \cap \mathbb{L}_2(A)$  as

$$V_h(x, y) = \frac{1}{4\pi^2} \int \int e^{i(xu+yv)} \frac{h^*(u, v)}{f_\varepsilon^*(-u)f_\varepsilon^*(-v)} dudv, \quad Q_h(x) = \frac{1}{2\pi} \int e^{ixu} \frac{h^*(u, 0)}{f_\varepsilon^*(-u)} du,$$

and must meet the following integrability condition:

**A 2.** *The functions  $\Pi_\theta^*/f_\varepsilon^*$ ,  $(\partial\Pi_\theta/\partial\theta_j)^*/f_\varepsilon^*$  and  $(\partial^2\Pi_\theta/\partial\theta_j\partial\theta_k)^*/f_\varepsilon^*$  for  $j, k = 1, \dots, r$  belongs to  $\mathbb{L}_1(A)$ .*

This assumption can be understood as  $\Pi_\theta^*$  and its first two derivatives (resp.  $(\Pi_\theta^2)^*$ ) have to be smooth enough compared to  $f_\varepsilon^*$ .

We are now able to describe in detail the procedure of [35] to understand the choice of this contrast function (see, *e.g.*, [35, 3] for the links between this contrast and regression-type contrasts). A full discussion of the hypothesis is given in Section 3.3.

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 ergodic<sup>1</sup>, 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 analytically computed. To do this, we use the same technique as in the convolution problem (see [36, 37]). 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

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

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 (see [36, 6.1. Proof of Lemma 2] for the proof).

**Lemma 3.1.** *For all  $i \in \{1, \dots, n\}$ , we have*

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

<sup>1</sup>We refer the reader to [18] 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 [26])

It follows from Lemma 3.1 that

$$\begin{aligned} \mathbf{P}m_\theta &= \int \int \Pi_\theta^2(x, y) f_{\theta_0}(x) dx dy - 2 \int \int \Pi_\theta(x, y) \Pi_{\theta_0}(x, y) f_{\theta_0}(x) dx dy \\ &= \|\Pi_\theta\|_A^2 - 2\langle \Pi_\theta, \Pi_{\theta_0} \rangle_A = \|\Pi_\theta - \Pi_{\theta_0}\|_A^2 - \|\Pi_{\theta_0}\|_A^2. \end{aligned} \quad (3)$$

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

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \mathbf{P}_n m_\theta. \quad (4)$$

### 3.2 Asymptotic Properties of the Estimator

The following result shows the consistency of our estimator and the central limit theorem (CLT) for  $\alpha$ -mixing processes. To achieve 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  $\theta_0$  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

$$\Sigma(\theta) = \mathcal{V}_\theta^{-1} \Omega(\theta) \mathcal{V}_\theta^{-1'}, \quad \Omega(\theta) = \Omega_0(\theta) + 2 \sum_{j=2}^{+\infty} \Omega_{j-1}(\theta), \quad (5)$$

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 3.1.** *Under Assumptions **A1–A3** let  $\hat{\theta}_n$  be the least square estimator defined in (4), 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 3.1 is provided in Subsection 6.1.

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 **A1–A3**, the matrix  $\Omega(\theta_0)$  is given by*

$$\Omega(\theta_0) = \Omega_1(\theta_0) + \Omega_2(\theta_0) + 2 \sum_{j=3}^{+\infty} \Omega_j(\theta_0),$$

where

$$\begin{aligned} \Omega_1(\theta_0) = & \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}^2(Y_1)] + 4\mathbb{E}[V_{\nabla_\theta \Pi_\theta}^2(\mathbf{Y}_1)] - 4\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)] \\ & - \left\{ \mathbb{E} \left( \int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right)^2 + 4\mathbb{E} \left( \nabla_\theta \Pi_\theta(\mathbf{X}_1) \right)^2 \right. \\ & \left. - 4\mathbb{E} \left( \int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right) \mathbb{E} \left( \nabla_\theta \Pi_\theta(\mathbf{X}_1) \right) \right\} \end{aligned}$$

and,

$$\begin{aligned} \Omega_2(\theta) = & \mathbb{E} \left[ \int \nabla_\theta \Pi_\theta^2(X_1, y) dy \int \nabla_\theta \Pi_\theta^2(X_2, y) dy \right] - \mathbb{E} \left[ \int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right]^2 \\ & - 2 \left( \mathbb{E} \left[ \left( \int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right) \nabla_\theta \Pi_\theta(\mathbf{X}_2) \right] + \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_2)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)] \right) \\ & - 4 \left( \mathbb{E} \left[ \int \nabla_\theta \Pi_\theta^2(X_1, y) dy \right] \mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_2)] + \mathbb{E}[\nabla_\theta \Pi_\theta(\mathbf{X}_1)]^2 - \mathbb{E}[V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_2)] \right) \end{aligned}$$

and, the covariance terms are given for  $j > 2$  as

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

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

Furthermore, the Hessian matrix  $\mathcal{V}_{\theta_0}$  is given by

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

The proof of Corollary 3.1 is given in Subsection 6.2.

**Sketch of proof:** Let us now state the strategy of the proof. The full proof is given in Section 6. Clearly, the proof of Theorem 3.1 relies on M-estimator properties and on the deconvolution strategy. The following observation explains the consistency of our estimator: 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  $\theta_0$ . 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)**, we prove the consistency of our estimator, and then, the first part of Theorem 3.1.

Asymptotic normality follows essentially from CLT for mixing processes (see [33]). Thanks to the consistency, the proof is based on a moment condition of the Jacobian vector of the function  $m_\theta(\mathbf{y}) = Q_{\Pi_\theta^2}(y) - 2V_{\Pi_\theta}(\mathbf{y})$  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.



### 3.3 Comments on the Assumptions

In the following, we provide a discussion of the hypotheses.

- Assumption **A 1(i)** is not restrictive; it is satisfied for many processes. Taking the process  $(X_i)_i$  defined in (1), we provide conditions on the functions  $b$ ,  $\sigma$  and  $\eta$  ensuring that Assumption **A 1(ii)** is satisfied (see [18] for more details).
  - (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  $b_{\theta_0}$  is bounded on every bounded set; that is, for every  $K > 0$ ,  $\sup_{|x| \leq K} |b_{\theta_0}(x)| < \infty$ .
  - (c) The function  $\sigma_{\theta_0}$  satisfies, for every  $K > 0$  and constant  $\sigma_1$ ,  $0 < \sigma_1 \leq \inf_{|x| \leq K} \sigma_{\theta_0}(x)$  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  $f_\varepsilon$  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  $\varepsilon_i$  is completely known up to a scale parameter, the model (1) may be non-identifiable as soon as the invariant density of  $X_i$  is smoother than the density of the noise (see [6]). The second part of **A 1(iii)** is a classical assumption ensuring the existence of the estimation criterion.

- For some models, the integrability assumption **A 2** is not satisfied. In particular, for models where this integrability assumption is not valid, we propose inserting a weight function  $\varphi$  or a truncation kernel as in [13, p. 285] to circumvent the issue of integrability. More precisely, we define the operators as follows:

$$Q_{h \star K_{B_n}}(x) = \frac{1}{2\pi} \int e^{ixu} \frac{(h \star K_{B_n})^*(u, 0)}{f_\varepsilon^*(-u)} du, \quad V_{h \star K_{B_n}}^* = \frac{(h \star K_{B_n})^*}{f_\varepsilon^* \otimes f_\varepsilon^*},$$

where  $K_{B_n}^*$  denotes the Fourier transform of a density deconvolution kernel with compact support and satisfies  $|1 - K_{B_n}^*(t)| \leq \mathbf{1}_{|t| > 1}$  and  $B_n$  is a sequence which tends to infinity with  $n$ . The contrast is then defined as

$$\mathbf{P}_n m_\theta = \frac{1}{n} \sum_{i=1}^{n-1} Q_{\Pi_\theta^* \star K_{B_n}}(Y_i) - 2V_{\Pi_\theta^* \star K_{B_n}}(\mathbf{Y}_i). \quad (6)$$

This contrast is still valid under Assumptions **A 1–3** by taking  $K_{B_n}(t)^* = \mathbf{1}_{|t| \leq B_n}$  with  $B_n = +\infty$ .

- It should be noted that the construction of the contrasts (2)-(6) does not need to know the stationary density  $f_{\theta_0}$ . The second part of Assumption **A1(iv)** is an exception because the first part concerns the uniqueness of  $\mathbf{P}m_\theta$ , which is strictly convex w.r.t.  $\theta$ . Nevertheless, the second part requires computing the Hessian matrix of  $\mathbf{P}m_\theta$  to ensure that it is invertible and secondly to calculate confidence

intervals. For the latter, we propose to use in practice the following consistent estimator for the confidence bounds:

$$\mathcal{V}_{\hat{\theta}_n} = \frac{1}{n} \sum_{i=1}^{n-1} Q_{\frac{\partial^2 \Pi_{\theta}^*}{\partial \theta^2}}(Y_i) - 2V_{\frac{\partial^2 \Pi_{\theta}}{\partial \theta^2}}(\mathbf{Y}_i), \quad (7)$$

because, 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  $\mathcal{V}_{\theta_0}$ . This matrix and its inverse can be computed in practice for a large class of models.

- The local dominance Assumptions **A 3(i)** and **(iii)** are not more restrictive than Assumption **A 2** and are satisfied if for  $j, k = 1, \dots, r$  the functions  $\sup_{\theta \in \Theta} \Pi_{\theta}^*/f_{\varepsilon}^*$  and  $\sup_{\theta \in \Theta} (\partial^2 \Pi_{\theta} / \partial \theta_j \partial \theta_k)^*/f_{\varepsilon}^*$  are integrable.
- In most applications, we do not know the bounds of the true parameter. The compactness assumption can be replaced with:  $\theta_0$  is an element of the interior of a convex parameter space  $\Theta \subset \mathbb{R}^r$ . Then, under our assumptions except for 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 [31] for this discussion.

## 4 Simulations

### 4.1 Linear Autoregressive Processes

We start from this following HMM

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \quad (8)$$

where the noises  $\varepsilon_i$  and the innovations  $\eta_i$  are supposed to be *i.i.d.* centered random variables with variance respectively  $\sigma_{\varepsilon}^2$  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  $\phi_0$  satisfies  $|\phi_0| < 1$  (see [18]). In this setting, the innovations  $\eta_i$  are assumed Gaussian with zero mean and variance  $\sigma_{0,\eta}^2$ . Hence, the transition function  $\Pi_{\theta_0}(x, y)$  is also Gaussian with mean  $\phi_0 x$  and variance  $\sigma_{0,\eta}^2$ . To analyse the effect of the regularity of the density of observation noises on the estimation of parameters, we consider three types of noises:

**Case 1: ARMA model with Gaussian noise (super smooth).** The density of  $\varepsilon_1$  is given by

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

We have  $f_{\varepsilon}^*(x) = \exp(-\sigma_{\varepsilon}^2 x^2/2)$ . The vector of parameters  $\theta_0$  belongs to the compact subset  $\Theta$  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  $\sigma_{\max}^2$  are positive real constants. We consider this condition ( $\sigma_{0,\eta}^2 > \sigma_{\varepsilon}^2$ ) for integrability assumption but one can relax this assumption (see Subsection 6.3 for the discussion on Assumptions **A 1-A 3**).

**Case 2: ARMA model with Laplace's noise (ordinary smooth).** The density of  $\varepsilon_1$  is given by

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

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

**Case 3: SV model: log- $\mathcal{X}^2$  noise (super smooth).** The density of  $\varepsilon_1$  is given by

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

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 by  $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ . For the cases 2 and 3, the vector of parameters  $\theta = (\phi, \sigma^2)$  belongs to the compact subset  $\Theta$  given by  $[-1 + r; 1 - r] \times [\sigma_{min}^2; \sigma_{max}^2]$  with  $r, \sigma_{min}^2$  and  $\sigma_{max}^2$  positive real constants. Furthermore, this latter case corresponds to the SV model introduced by Taylor in [53]:

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

where  $\beta$  denotes a positive constant. The noises  $\xi_{i+1}$  and  $\eta_{i+1}$  are two centered Gaussian random variables with standard variance  $\sigma_\varepsilon^2$  and  $\sigma_0^2$ .

In the original paper [53], the constant  $\beta$  is equal to 1<sup>2</sup>. 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 defined model (8).

For all these models, we postponed in Subsection 6.3 the verification of Theorem 3.1 assumptions.

#### 4.1.1 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 4.1.** *For the HMM model (8) the theoretical contrast defined in (3) is given by*

$$\mathbf{P}m_\theta = -1 + \frac{\sigma_\eta + \sigma_{0,\eta}}{2\sqrt{\pi}\sigma_\eta\sigma_{0,\eta}} - \sqrt{\frac{2(\phi_0^2 - 1)}{\pi\sigma_\eta^2(-1 - \phi_0^2) - \pi\sigma_{0,\eta}^2(1 + \phi^2 - 2\phi\phi_0)}}, \quad (9)$$

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

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

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

<sup>2</sup>We argue that our approach can be applied when we introduce a long mean parameter  $\mu$  in the volatility process.

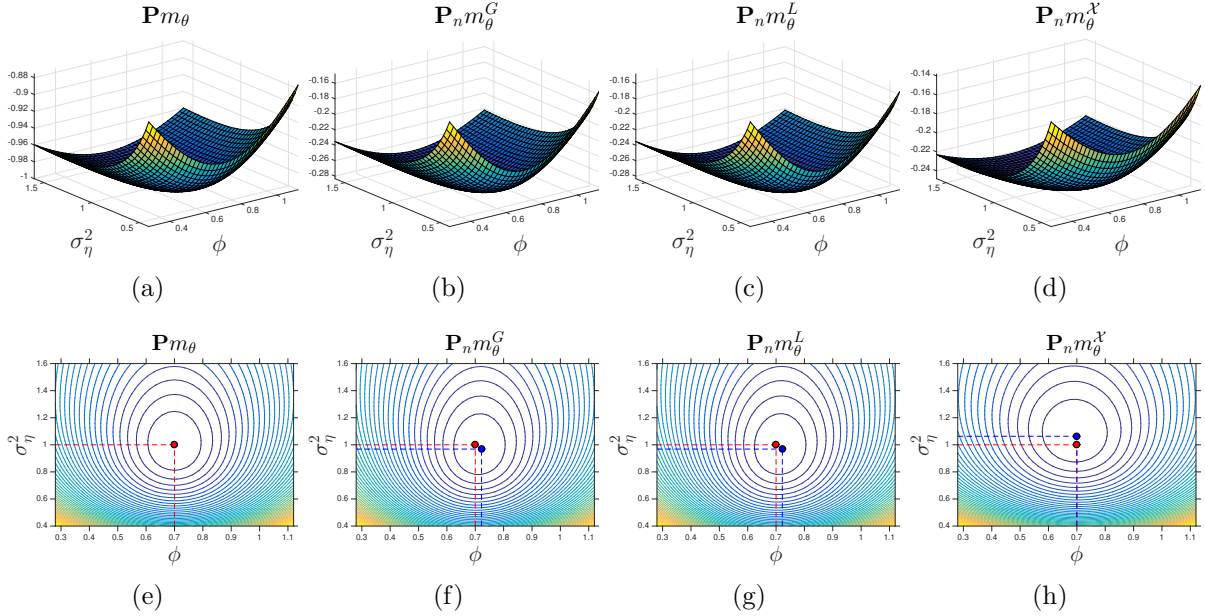


Figure 1: Contrast functions. (a):  $\mathbf{P}m_\theta$  as a function of the parameters  $\phi$  and  $\sigma_\eta^2$  for one realization of (8), with  $n = 2000$ . (b):  $\mathbf{P}_n m_\theta^G$ . (c):  $\mathbf{P}_n m_\theta^L$ . (d):  $\mathbf{P}_n m_\theta^X$ . (e)-(h): Corresponding contour lines. The red circle represents the global minimizer  $\theta_0$  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.

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

The notations  $\mathbf{P}_n m_\theta^G$  and  $\mathbf{P}_n m_\theta^L$  corresponds to the case 1 and case 2, respectively, (see Figure 1 for a visual representation of these contrasts as a function of the parameters  $\phi$  and  $\sigma_\eta^2$ ).

For the SV model (case 3), the inverse Fourier transform of  $\Pi_\theta^*/f_\varepsilon^*$  does not have an explicit expression in particular because of the gamma function  $\Gamma$  in  $f_\varepsilon^*$ . Nevertheless, the contrast can be approached numerically using a fast Fourier transform (FFT) and is also represented. Figure 1 depicts the three empirical and true contrast curves as a function of the parameters  $\phi$  and  $\sigma_\eta^2$  for one realization of (8), 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 considered in this simulation study.

#### 4.1.2 Monte Carlo Simulation and Comparison with the MLE

Let us present the results of our simulation experiments. First, we perform a Monte Carlo (MC) study for the first two cases to study 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 non-Gaussian noises such as Laplace noises. For noises other than Gaussian noises, the calculation of the MLE is generally more complicated in practice and requires the use of algorithms such as MCMC, EM, Stochastic Approximation of EM (SAEM), SMC, or alternative estimation strategies (see [1], [10] and [48]), which require a longer computation time. In the case

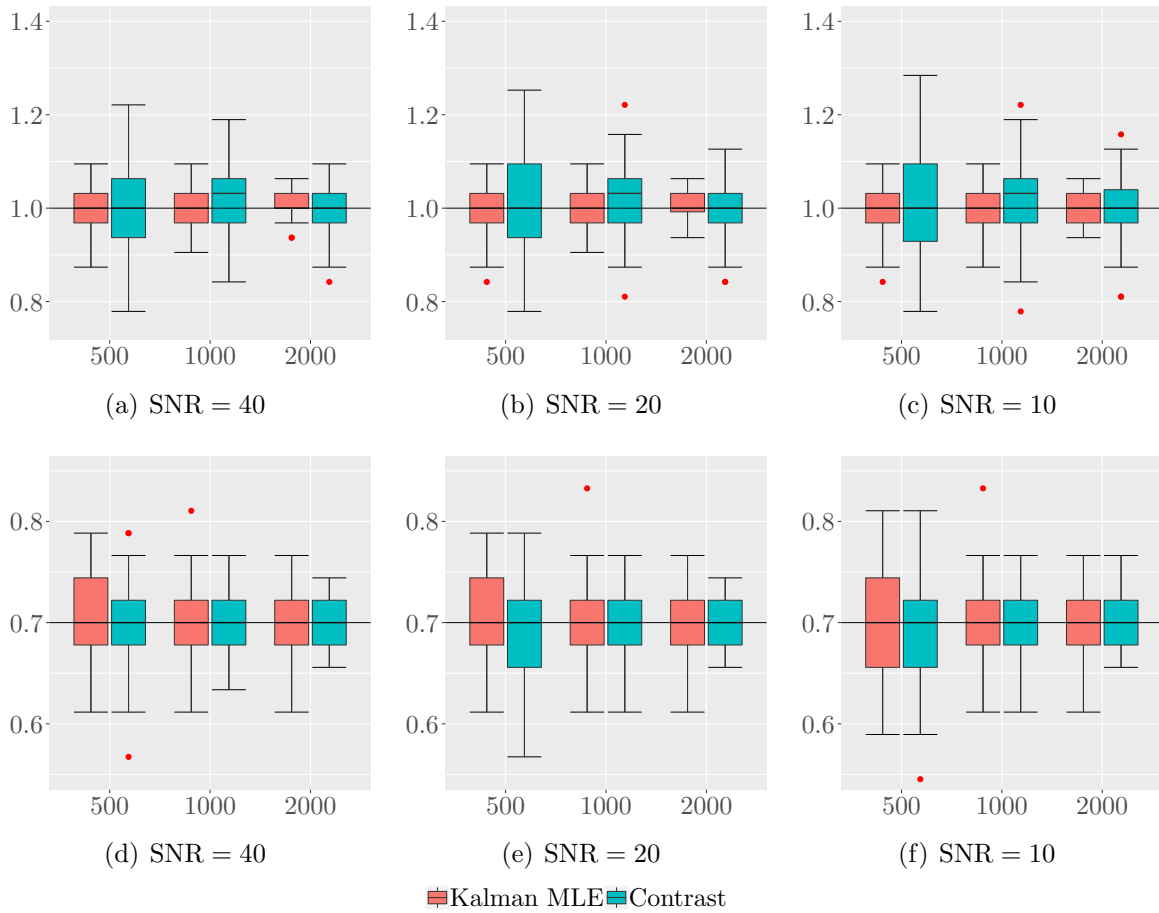


Figure 2: Boxplot of parameter estimators for different sample sizes in case 1. The horizontal lines represent the values of true parameters  $\sigma_\eta^2$  and  $\phi$ .

of Laplace’s noise, we use the R package **tseries** [54] to fit an ARMA(1,1) model to the  $Y_i$  observations by a conditional least squares method [30]. 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 as opposed to the MLE where the Kalman filter is used to construct the likelihood of the model to be maximized.

For each simulation, we consider three different signal-to-noise ratios denoted by SNR (*i.e.*,  $\text{SNR} = \sigma_\eta^2/\sigma_\varepsilon^2 = 1/\sigma_\varepsilon^2$ , with  $\sigma_\varepsilon^2$  equals to  $1/40$ ,  $1/20$  and  $1/10$  corresponding to low, medium, and high noise levels). For all experiments, we set  $\phi = 0.7$  and generate samples of different sizes (*i.e.*,  $n = 500$  up to  $2000$ ) while keeping the financial and economic context in mind. We represent the results obtained, on boxplots for each parameter and for 100 repetitions in Figure 2 and Figure 3 (corresponding to cases 1 and 2, respectively). We can see that, as already noticed in the deconvolution setting, there is little difference between Laplace and Gaussian  $\varepsilon_i$ ’s. The convergence is slightly better for Laplace noise. Moreover, increasing the sample size leads to noticeable improvements in the results. In the Gaussian case (Figure 2), the MLE provides better overall results, which is not surprising since in case 1, it is the best unbiased linear estimator. However, we can see that for the  $\phi$  parameter, our estimator is slightly better. The estimation of the variance  $\sigma_\eta^2$  is more difficult, especially as the SNR decreases. For Laplace’s noise (Figure 3, our contrast estimator provides better results, whatever the number of observations and the SNR. The

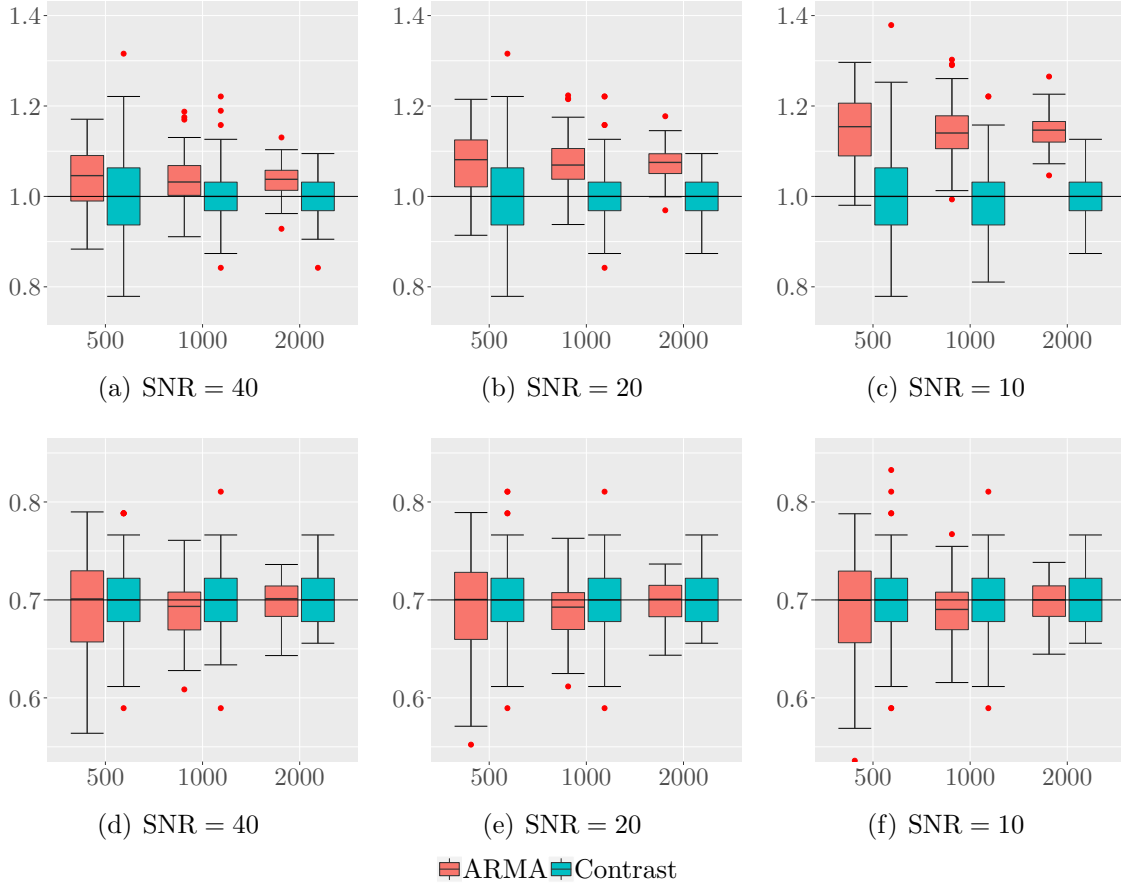


Figure 3: Boxplot of parameter estimators for different sample sizes in case 2. The horizontal lines represent the values of true parameters  $\sigma_\eta^2$  and  $\phi$ .

differences with the MLE are even more significant when estimating the variance  $\sigma_\eta^2$ . The contrast approach is an interesting alternative to the MLE in the non-Gaussian case.

Our estimation procedure allows us to deepen our analysis since Theorem 3.1 applies and as we already mentioned, Corollary 3.1 allows to compute confidence intervals (CIs) in practice, *i.e.*, for  $i = 1, 2$ :

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

as  $n \rightarrow \infty$  where  $z_{1-\alpha/2}$  is the  $1 - \alpha/2$  quantile of the Gaussian distribution,  $\theta_{0,i}$  is the  $i^{\text{th}}$  coordinate of  $\theta_0$  and  $\mathbf{e}_i$  is the  $i^{\text{th}}$  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 (7).

We investigate the coverage probabilities for sample sizes ranging from 100 to 2000 with a replication number of 1000. We compute the confidence interval for each sample and plot the proportion of samples for which the true parameters are contained in the confidence interval for the two cases 1 and 2. The results are shown in Figure 4. It can be seen that the computed proportions provide good estimators for the empirical coverage probability for the confidence intervals 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.

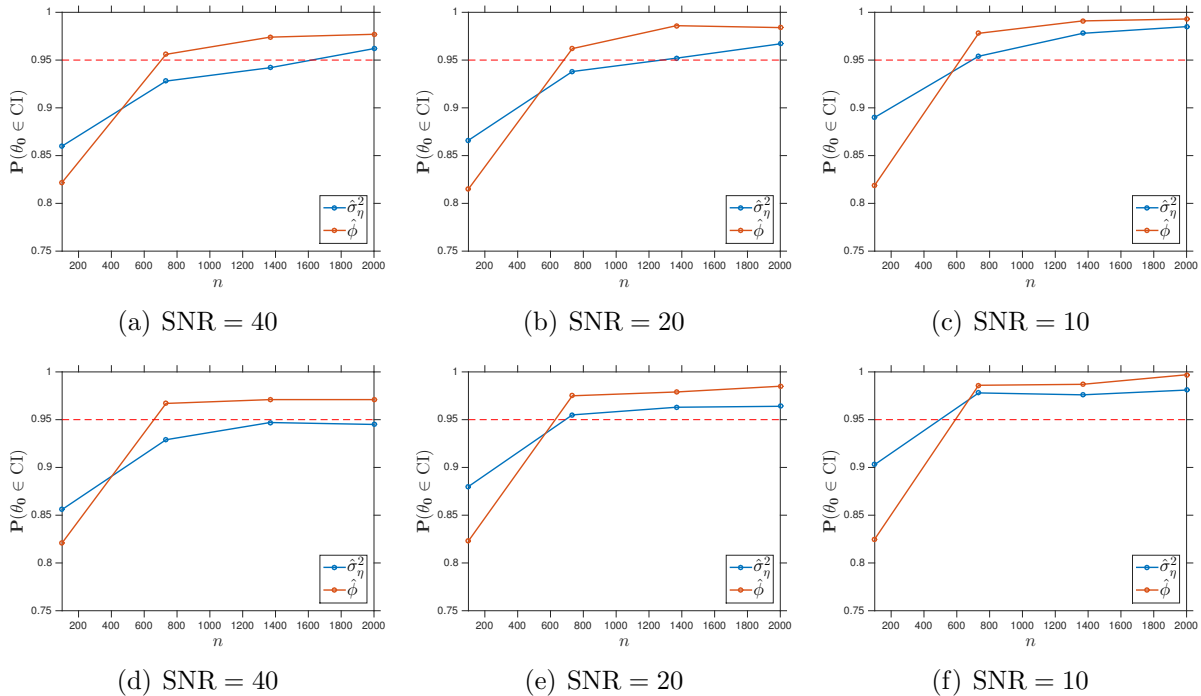


Figure 4: Coverage probability for 95% CIs versus sample size  $n$  (Case 1 top Case 2 bottom) based on 1000 simulations.

#### 4.1.3 Sensibility of the Contrast w.r.t. $\sigma_\varepsilon^2$

In this section, we focus on the sensitivity of our contrast estimator when we relax the assumption **A1(iii)**, that is, we assume that  $f_\varepsilon$  is known up to the unknown  $\sigma_\varepsilon^2$ . Many authors have addressed this issue from both a theoretical and practical perspective. We can highlight the following works, although this list is not exhaustive: in the context of [20] and [45], where the error density is unknown and estimated from additional direct observations which come from the error density. [42] considered a testing procedure for two possible densities competing to be the error density. [7] introduce a uniformly consistent estimation procedure when the error variance  $\sigma_\varepsilon^2$  is unknown but restricted to a known compact interval. In [34], the authors consider two different models: a first one where an additional sample of pure noise is available, as well as the model of repeated measurements, where the contaminated random variables of interest can be observed repeatedly, with independent errors. In [14] the authors propose a completely new approach to this problem, which does not require any extra data of any kind and does not require smoothness assumptions on the signal distribution.

To estimate  $\sigma_\varepsilon^2$  here, we use the approach proposed in [42], which is very simple to implement and yields good results. Let us present the methodology. We consider the absolute empirical Fourier transform defined by

$$\widehat{\varphi}_n(t) = \left| \frac{1}{n} \sum_{j=1}^n e^{itY_j} \right|.$$

In the sequel we denote  $(k_n)_{n \in \mathbb{N}}$ ,  $(\omega_n)_{n \in \mathbb{N}}$  and  $(\sigma_n^2)_{n \in \mathbb{N}}$  three sequences of positive numbers described below. We set

$$\tilde{\sigma}_n^2 = -2k_n^{-2} \log \left( \frac{\widehat{\varphi}_n(k_n)}{Ck_n^{-\beta}} \right),$$



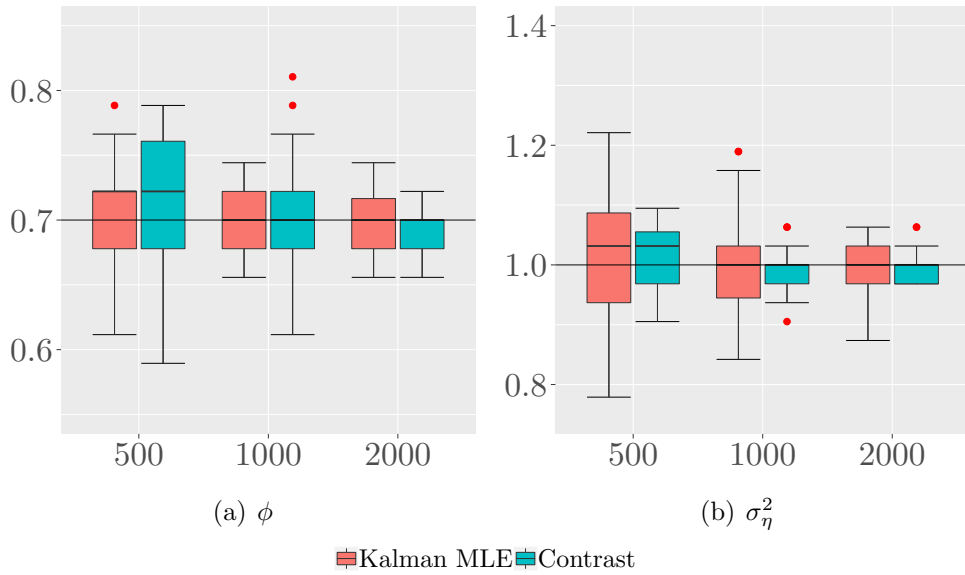


Figure 5: Left. Estimation of  $\phi$  with unknown  $\sigma_\varepsilon^2$ . (Horizontal line: true value). Right. Estimation of  $\sigma_\eta^2$  with unknown  $\sigma_\varepsilon^2$ . (Horizontal line: true value).

with  $C$  and  $\beta > 0$  arbitrary constants not stipulated to be known, so they might be misspecified in practice. If we know the regularity of the density of  $X_i$ , that is, if  $f_\theta \in \mathcal{F}$ , where  $\mathcal{F}$  is the ordinary smooth or super smooth function class, one can choose these parameters according to it. For all examples considered in this section, the stationary density  $f_\theta$  is super smooth. Otherwise, we fix these parameters arbitrarily and set  $\omega_n = k_n/\log(k_n)$  with  $k_n \rightarrow \infty$ . Thus, the estimator of  $\sigma_\varepsilon^2$  is defined as

$$\hat{\sigma}_{n,\varepsilon}^2 = \begin{cases} 0 & \text{if } \tilde{\sigma}_n^2 < 0 \\ \tilde{\sigma}_n^2 & \text{if } \tilde{\sigma}_n^2 \in [0, \sigma_n^2] \\ \sigma_n^2 & \text{if } \tilde{\sigma}_n^2 > \sigma_n^2 \end{cases} \quad (10)$$

To analyse the sensitivity of our approach w.r.t. the noise variance  $\sigma_\varepsilon^2$ , we applied our contrast estimator defined in Proposition 4.1 for the AR Gaussian model where we plugged the estimator given in (10). We then compared our results with the MLE estimation where the same variance estimator is used. The results are shown in Figure 5 for each parameter and for a different number of observations. For the construction of the estimator (10), we took the following parameters  $C = 1/4$ ,  $\beta = 4$ ,  $\sigma_n^2 = 1$  and  $k_n = \sqrt{n}$  and a number of MC trials equal to 50. We tested different sets of parameters to see the influence of these parameters. As the influence was negligible, we did not report all the tests.

The results obtained are similar to those of the Section 4.1.2. We note that whatever the number of observations, the MLE is more robust in this favourable case (Gaussian measurement noise). Nevertheless, our approach gives as good results as when the variance is assumed to be known. The given estimator (10) is a good alternative in practice when we relax the hypothesis on the knowledge of the observation noises and is readily implementable.



Table 1: Asymptotic variance  $\Sigma(\widehat{\theta}_n)$  w.r.t. the truncation  $J$  for  $n = 500, 1000$  and  $5000$ .

$J$	2	5	10	20
<b>n = 500</b>				
$\phi_0 = 0.7$	0.2367	0.2364	0.2338	0.2338
$\sigma_\eta^2 = 1$	0.4299	0.4287	0.4211	0.4211
<b>n = 1000</b>				
$\phi_0 = 0.7$	0.1671	0.1664	0.1653	0.1653
$\sigma_\eta^2 = 1$	0.3032	0.3008	0.2977	0.2977
<b>n = 5000</b>				
$\phi_0 = 0.7$	0.0747	0.0744	0.0739	0.0739
$\sigma_\eta^2 = 1$	0.1356	0.1345	0.1332	0.1332

#### 4.1.4 Sensibility of the Contrast w.r.t. the truncature of the Asymptotic Variance Matrix

The asymptotic variance of our estimator  $\Sigma(\theta)$  defined in (5) suggests computing the covariance matrix  $\Omega(\theta)$ , which requires the computation of the infinite sum of covariance terms. In practice, we use the following empirical covariance with lag  $j$  defined as

$$\widehat{\Omega}_j(\widehat{\theta}_n) = \frac{1}{n} \sum_{i=|j|+1}^n \left( \nabla_{\theta} m_{\theta}(\mathbf{Y}_{i-|j|}) \right) \left( \nabla_{\theta} m_{\theta}(\mathbf{Y}_i) \right),$$

and the sum in (5) has been truncated for a large value of  $j$ . In this part we wish to analyze the influence of this truncation on the asymptotic variance of our estimator. The illustration is made for the Gaussian AR model. In Table 1, we report the asymptotic variance w.r.t. the truncation denoted  $J$  and the number of observations  $n$ . We note that the asymptotic variance decreases w.r.t.  $J$  and no longer varies for a given moderate lag, whatever the number of observations. This phenomenon is explained by the fact that at a certain rank  $j_0$ , we have that for  $j \geq j_0$  the covariance terms  $\Omega_j$  are infinitely small and negligible in front of  $\Omega_1$ , which is in agreement with the mixing assumption **A1(ii)**, satisfied for the AR process. Furthermore, as expected, the asymptotic variance also decreases with the number of observations.

## 4.2 NonLinear Autoregressive Processes

We conclude this simulation study with an example of a nonlinear process of the form

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = \phi_0 \sin(X_i) + \eta_{i+1}, \end{cases} \quad (11)$$

where the noises  $\varepsilon_i$  and the innovations  $\eta_i$  are supposed to be *i.i.d.* centered random variables with variance respectively  $\sigma_\varepsilon^2$  and  $\sigma_{0,\eta}^2$ . The transition density is given by  $\Pi_{\theta_0}(x, y) = f_\eta(y - \phi_0 \sin(x))$  (see [44]). In spite of the explicit expressions of the Fourier transforms of  $\Pi_{\theta_0}$  and  $f_\varepsilon$  the contrast does not admit an explicit expression but can easily be calculated by FFT or numerical integration. More precisely, in this typical example, the operator  $V_{\Pi_\theta}$  in the contrast is not explicit, but we give in Section 6.3 some details

Table 2: Nonlinear AR(1): MSE for  $n = 500$  up to 2000 and  $MC = 50$ . The number of particles for the SAEM is 50.

$n$	500	1000	2000
<b>Contrast</b>			
	0.0049	0.0029	0.0014
<b>EKF</b>			
	0.0074	0.0049	0.0042
<b>SAEM</b>			
nbIter=100	0.0073	0.0061	0.0051
nbIter=400	0.0062	0.0032	0.0026

of its computation. For this example, we compare our contrast to the MLE computed with the Extended Kalman filter (EKF) adapted to nonlinear models. The latter, after a linearization of the model, can be used to compute the likelihood. Nevertheless, the properties of EKF are known to be good when the model is not strongly nonlinear, which is not the case for the drift function given by the sinus. We therefore propose to also compare our approach with the more general SAEM (see [15] for more details). Given the unobservable character of the variables  $X_i$ , the latter uses a conditional particle filter (see [41]) in the expectation step. For this experiment we take  $SNR = 1/40$ ,  $\phi_0 = 0.7$ ,  $\sigma_{0,\eta}^2 = 0.5$ , a number of replications  $MC = 50$  and we generate samples of different sizes  $n = 500, 1000$  and  $2000$ . For the SAEM, we took a number of particles equal to 50 in the conditional particle filter and a number of iterations (denoted nbIter) for the convergence of the EM equal to 400 (below this threshold, there was no convergence). The results are represented by boxplots for each parameter (see Figures 6) and synthesized in the Table 2, where the mean square error (MSE) is given and computed as

$$\text{MSE}(\theta_0, \hat{\theta}_0) = \frac{1}{MC} \sum_{j=1}^{MC} (\hat{\phi}_j - \phi_0)^2 + (\hat{\sigma}_{\eta,j}^2 - \sigma_{\eta,0}^2)^2$$

We observe from Table 2 that the EKF gives worse results than the two other approaches, in terms of MSE. This result was expected and in accordance with the properties of the EKF when the model is not linear. The contrast gives better results than the SAEM method, whatever the number of observations and the number of iterations considered in the EM step. From Table 2, we can see that the results for SAEM tend to be better and closer to the contrast when we increase the number of iterations (nbIter=100 to nbIter=400). Indeed, the results of SAEM depend strongly on the initial condition for  $\theta$  and the number of iterations that control the convergence of the algorithm. We voluntarily tested different thresholds for the number of iterations since, for nbIter=100, the convergence did not take place for the drift parameter  $\phi$ . On the other hand, the SAEM results do not depend on the number of particles chosen since a conditional filter has been used with an ancestral resampling step, which allows the use of a reasonable number of particles in terms of computation time (see [41] for more details). Contrary to the SAEM, our approach does not require parameter calibration and gives good results, even in a nonlinear framework, whatever the initial condition.

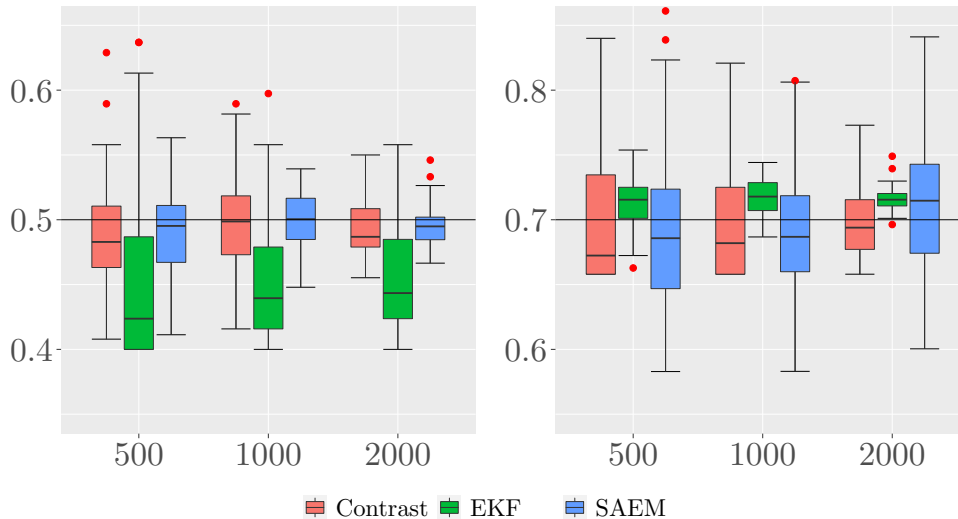


Figure 6: Boxplot of parameter estimators for different sample sizes and  $MC = 50$ ,  $\text{nbIter}=400$  and  $\text{NbParticles}=50$  for the SAEM. The horizontal lines represent the values of true parameters  $\sigma_\eta^2$  and  $\phi$ . Left: parameter  $\sigma_\eta^2$ . Right: parameter  $\phi$ .

## 5 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 of using a proxy variable for the ex-ante real rate (see, *e.g.*, [32] for the US region and more recently in [39] for Canada, the Euro Area and the UK). The second strategy consists of treating the ex-ante real rate as an unknown variable using a latent factor model (see, *e.g.*, [5, 29, 28]). Our procedure is in line with this second strategy in which the factor model is specified as follows:

$$\begin{cases} Y_t &= \alpha + X_t + \varepsilon_t \\ X_t &= \phi X_{t-1} + \eta_t, \end{cases} \quad (12)$$

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  $\alpha$ ,  $\phi$  a parameter of persistence and  $\varepsilon_t$  (resp.  $\eta_t$ ) centered Gaussian random variables with variance  $\sigma_\varepsilon^2$  (resp.  $\sigma_\eta^2$ ).

This model is derived 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. As a result, expanding these expressions to allow for

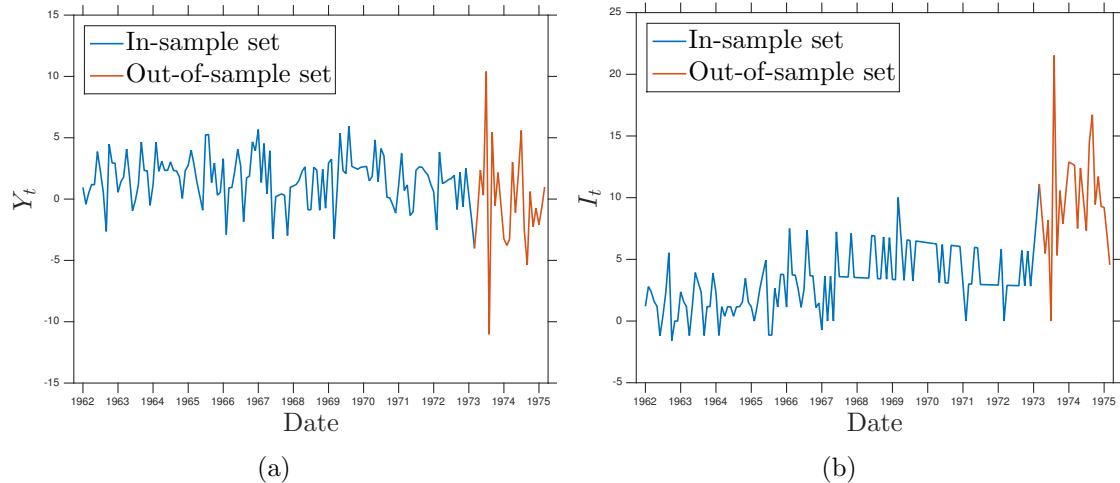


Figure 7: (a): Observed ex-post real rate. (b): Observed actual inflation rate.

expected inflation rate  $\mathcal{I}_t^e$  gives

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

where  $\varepsilon_t = \mathcal{I}_t^e - \mathcal{I}_t$  is the random variable for inflation expectation. If people do not make systematic errors in forecasting inflation,  $\varepsilon_t$  can be assumed to be a Gaussian white noise with variance denoted by  $\sigma_\varepsilon^2$ . This assumption is known as the Rational Expectation Hypothesis (REH). Thus, the REH lends itself very naturally to a state-space representation (see, *e.g.*, [4]), and defining  $X_t = Y_t^e - \alpha$  as the ex-ante real interest rate adjusted by its population mean,  $\alpha$ , yields the modelization (12).

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 contained 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 rates,  $Y_t$  (depicted in Figure 7(a)), is calculated as the difference between the logarithm annualized nominal funds rate and the logarithm annualized percentage inflation rate,  $I_t$  (depicted in Figure 7(b)), obtained from the consumer price index for all urban consumers. The data are available from the Federal Reserve Economic Data and for this data, we consider two of the models previously studied, the Gaussian AR model and the SV model.

Let us denote  $\theta = (\alpha, \phi, \sigma_\eta^2)$  the vector of unknown parameters to be estimated. We estimate the unknown parameter vector  $\theta$  on the in-sample in two stages: first, we minimize the contrast introduced in this paper to estimate the unknown parameter vector  $\theta$  on the in-sample. 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 all forecasts are computed using the pseudo out-of-sample method.

The value of the estimation obtained in the first step is as follows:  $\hat{\alpha}_{\text{AR}} = 1.5699$ ,  $\hat{\alpha}_{\text{SV}} = 2.1895$ ,  $\hat{\phi}_{\text{AR}} = 0.5750$ ,  $\hat{\phi}_{\text{SV}} = 0.8500$ ,  $\hat{\sigma}_{\text{AR}}^2 = 4.4048$  and  $\hat{\sigma}_{\text{SV}}^2 = 3.1756$  and the forecasts obtained from the out-of-sample are shown in Figure 8. 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 Lilliefors test [40], a variant of the

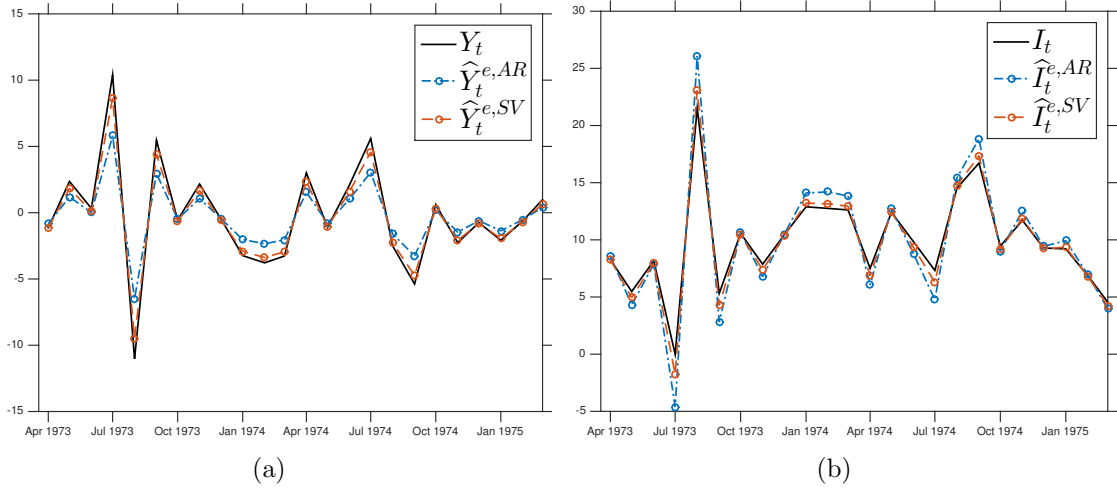


Figure 8: (a): Ex-post real rate forecasts. (b): Expected inflation rate forecasts.

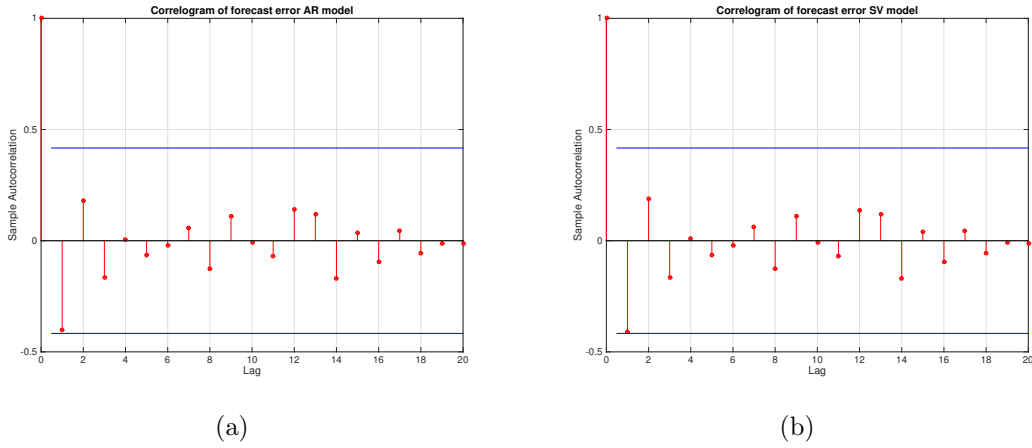


Figure 9: (a): Correlogram of  $\hat{e}_t$  for AR model. (b): Correlogram of  $\hat{e}_t$  for SV model.

Kolmogorov-Smirnov test when certain parameters of the distributions must be estimated which is the case here, 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  $\alpha = 0.05$ . 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 9. 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  $\phi$  is higher for the SV model than the AR model. On the other hand, the variance is lower. Therefore, the variance of  $\hat{\mathcal{I}}_t^e$  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 [52] and [50], whose findings show that the persistence parameter is high and close to one for this period of study.

## Conclusion

In this paper, we propose a new parametric estimation strategy for non-linear and non-Gaussian HMM models inspired by [37]. We also provide an analytical expression of the asymptotic variance matrix  $\Sigma(\hat{\theta}_n)$  which allows us to consider the construction of confidence intervals. 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 [17]). The parameter estimation step in HMM models is very important since it is shown in [22] that the bias in 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. In this paper, for example, we looked at the prediction of ex-ante interest rates and found that during periods of high inflation, the annualized inflation rate has a 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 and simple 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.

## 6 Proofs

### 6.1 Proofs of Theorem 3.1

For the reader's convenience, we split the proof of Theorem 3.1 into three parts: in Subsection 6.1.1, we give the proof of the existence of our contrast estimator defined in (3). In Subsection 6.1.2, we prove consistency, that is, the first part of Theorem 3.1. Then, we prove the asymptotic normality of our estimator in Subsection 6.1.3, that is, the second part of Theorem 3.1. Section 6.2 is devoted to Corollary 3.1.

#### 6.1.1 Existence 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.  $\theta$ . Hence, the function  $\mathbf{P}_n m_\theta = \frac{1}{n} \sum_{i=1}^n m_\theta(\mathbf{Y}_i)$  is continuous w.r.t.  $\theta$  belonging to the compact subset  $\Theta$ . So, there exists  $\hat{\theta}$  belongs to  $\Theta$  such that  $\inf_{\theta \in \Theta} \mathbf{P}_n m_\theta = \mathbf{P}_n m_{\hat{\theta}}$ .  $\square$

#### 6.1.2 Consistency

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

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

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

**Lemma 6.1.** *Uniform Law of Large Numbers (see [46] for the proof). Let  $(Y_i)_{i \geq 1}$  be an ergodic stationary process and suppose that:*

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

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

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

By assumption  $\Pi_\theta$  is continuous w.r.t.  $\theta$  for any  $x$  and measurable w.r.t.  $x$  for all  $\theta$  which implies the continuity and the measurability of the function  $\mathbf{P}_n m_\theta$  on the compact subset  $\Theta$ . Furthermore, the local dominance assumption **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 6.1 gives 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 [47, Theorem 2.1 p. 2121 chapter 36] yields the weak (convergence in probability) consistency of the estimator.  $\square$

### 6.1.3 Asymptotic Normality

For the CLT, we need to define the  $\alpha$ -mixing property of a process (we refer the reader to [18] for a complete review of mixing processes).

**Definition 6.1** ( $\alpha$ -mixing (strongly mixing process)). *Let  $Y := (Y_i)_i$  denote 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  $\alpha$ -mixing if  $\alpha(n) \rightarrow 0$  as  $n \rightarrow \infty$ , where*

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

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

**Lemma 6.2.** *Suppose that the conditions of the consistency hold. Suppose further that:*

- (i)  $(\mathbf{Y}_i)_i$  is  $\alpha$ -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  $\theta_0$  and for  $j, k \in \{1, \dots, r\}$ :

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



Then,  $\widehat{\theta}_n$  defined in (4) 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 mapping  $\mathbf{P}m_{\theta}$  given in (3).

*Proof.* The proof follows from [31, Proposition 7.8 p. 472] and [33], 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 6.2 hold under our assumptions **A3(ii)** and **A(iii)**.

(ii): As the function  $\Pi_{\theta}$  is twice continuously differentiable w.r.t.  $\theta$ ,  $\forall \mathbf{y}_i \in \mathbb{R}^2$  and so also  $\Pi_{\theta}^2$ , the mapping  $m_{\theta}(\mathbf{y}_i) : \theta \in \Theta \mapsto m_{\theta}(\mathbf{y}_i) = Q_{\Pi_{\theta}^2}(y_i) - 2V_{\Pi_{\theta}}(\mathbf{y}_i)$  is twice continuously differentiable  $\forall \theta \in \Theta$  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  $j \in \{1, \dots, r\}$ ,  $\frac{\partial \Pi_{\theta}}{\partial \theta_j}$  and  $\frac{\partial \Pi_{\theta}^2}{\partial \theta_j}$  belong to  $\mathbb{L}_1(A)$ , therefore one can apply the Lebesgue Differentiation Theorem and Fubini Theorem to obtain

$$\nabla_{\theta} m_{\theta}(\mathbf{y}_i) = \left[ Q_{\nabla_{\theta} \Pi_{\theta}^2}(y_i) - 2V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{y}_i) \right]. \quad (13)$$

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

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

where  $C_1$  and  $C_2$  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  $\mathbb{L}_1(A)$ , 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  $\theta_0$ , by the local dominance assumption **A3(iii)**,

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

This ends the proof of Theorem 3.1.  $\square$

## 6.2 Proof of Corollary 3.1

By replacing  $\nabla_{\theta} m_{\theta}(\mathbf{Y}_1)$  by its expression (13), we have for  $j = 1$

$$\begin{aligned} \Omega_1(\theta) &= \text{Var} \left[ Q_{\nabla_{\theta} \Pi_{\theta}^2}(Y_1) - 2V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{Y}_1) \right] \\ &= \text{Var} \left( Q_{\nabla_{\theta} \Pi_{\theta}^2}(Y_1) \right) + 4\text{Var} \left( V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{Y}_1) \right) - 4\text{Cov} \left( Q_{\nabla_{\theta} \Pi_{\theta}^2}(Y_1), V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{Y}_1) \right). \end{aligned}$$



Owing to Lemma 3.1, we obtain

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

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

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

and

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

Hence

$$\begin{aligned}\Omega_1(\theta) &= \mathbb{V}ar(\nabla_{\theta}m_{\theta}(\mathbf{Y}_1)) \\ &= \mathbb{E}[Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1)^2] + 4\mathbb{E}[V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1)^2] - 4\mathbb{E}[Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1)V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1)] \\ &\quad - \left(4\mathbb{E}[\nabla_{\theta}\Pi_{\theta}(\mathbf{X}_1)]^2 + \mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(\mathbf{X}_1, y)dy\right]^2 - 4\mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_1, y)dy\right]\mathbb{E}[\nabla_{\theta}\Pi_{\theta}(\mathbf{X}_1)]\right).\end{aligned}$$

For  $j = 2$ , we have

$$\begin{aligned}\Omega_2(\theta) &= \mathbb{C}ov(\nabla_{\theta}m_{\theta}(\mathbf{Y}_1), \nabla_{\theta}m_{\theta}(\mathbf{Y}_2)) \\ &= \mathbb{C}ov\left(Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1), Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_2)\right) - 2\mathbb{C}ov\left(Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1), V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_2)\right) \\ &\quad - 2\mathbb{C}ov\left(Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_2), V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1)\right) + 4\mathbb{C}ov(V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1), V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_2))\end{aligned}$$

where the different terms are obtained from Lemma 3.1

$$\begin{aligned}\mathbb{C}ov\left(Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1), Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_2)\right) &= \mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_1, y)dy \int \nabla_{\theta}\Pi_{\theta}^2(X_2, y)dy\right] \\ &\quad - \mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_1, y)dy\right]\mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_2, y)dy\right] \\ \mathbb{C}ov\left(Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_1), V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_2)\right) &= \mathbb{E}\left[\left(\int \nabla_{\theta}\Pi_{\theta}^2(X_1, y)dy\right) \nabla_{\theta}\Pi_{\theta}(\mathbf{X}_2)\right] \\ &\quad - \mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_1, y)dy\right]\mathbb{E}[\nabla_{\theta}\Pi(\mathbf{X}_2)] \\ \mathbb{C}ov\left(Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_2), V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1)\right) &= \mathbb{E}[Q_{\nabla_{\theta}\Pi_{\theta}^2}(Y_2)V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1)] - \mathbb{E}\left[\int \nabla_{\theta}\Pi_{\theta}^2(X_2, y)dy\right]\mathbb{E}[\nabla_{\theta}\Pi(\mathbf{X}_1)] \\ \mathbb{C}ov(V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1), V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_2)) &= \mathbb{E}[V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_1)V_{\nabla_{\theta}\Pi_{\theta}}(\mathbf{Y}_2)] - \mathbb{E}[\nabla_{\theta}\Pi(\mathbf{X}_1)]\mathbb{E}[\nabla_{\theta}\Pi(\mathbf{X}_2)]\end{aligned}$$

Now, by using the stationarity assumption **A1(iv)** of  $(X_i)_{i \geq 1}$  we obtain that

$$\begin{aligned}\Omega_2(\theta) = & \mathbb{E} \left[ \int \nabla_{\theta} \Pi_{\theta}^2(X_1, y) dy \int \nabla_{\theta} \Pi_{\theta}^2(X_2, y) dy \right] - \mathbb{E} \left[ \int \nabla_{\theta} \Pi_{\theta}^2(X_1, y) dy \right]^2 \\ & - 2 \left( \mathbb{E} \left[ \left( \int \nabla_{\theta} \Pi_{\theta}^2(X_1, y) dy \right) \nabla_{\theta} \Pi_{\theta}(\mathbf{X}_2) \right] + \mathbb{E}[Q_{\nabla_{\theta} \Pi_{\theta}^2}(Y_2) V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{Y}_1)] \right) \\ & - 4 \left( \mathbb{E} \left[ \int \nabla_{\theta} \Pi_{\theta}^2(X_1, y) dy \right] \mathbb{E}[\nabla_{\theta} \Pi(\mathbf{X}_2)] + \mathbb{E}[\nabla_{\theta} \Pi(\mathbf{X}_1)]^2 - \mathbb{E}[V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{Y}_1) V_{\nabla_{\theta} \Pi_{\theta}}(\mathbf{Y}_2)] \right)\end{aligned}$$

*Calculus of the covariance matrix of Corollary 3.1 for  $j > 2$ :* By replacing  $\nabla_{\theta} m_{\theta}(\mathbf{Y}_1)$  by its expression (13), we have

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

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

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

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

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

Furthermore, conditioning by  $X_{1:n}$  and using 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

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

Noting that for  $j > 2$  the stationarity of  $(X_i)_{i \geq 1}$  implies that  $\mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_1)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_j)] = \mathbb{E}[Q_{\nabla_\theta \Pi_\theta^2}(Y_j)V_{\nabla_\theta \Pi_\theta}(\mathbf{Y}_1)]$ . Hence,

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

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

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

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  $\mathcal{V}_\theta$ :* We have

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

Under **A1(i)**,  $\forall \theta$  in  $\Theta$ , the mapping  $\theta \mapsto \mathbf{P}m_\theta$  is twice differentiable w.r.t.  $\theta$  on the compact subset  $\Theta$ . 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 3.1 is completed.  $\square$

### 6.3 Contrast and Checking Assumptions for the Simulations

**Contrasts for the linear AR simulations:** To compute the several contrasts defined in Proposition 4.1, the following quantities are essentially required:  $(\Pi_\theta^2(x, 0))^*$ ,  $\Pi_\theta^*(x, y)$  and  $f_\varepsilon^*(x)$ . For the model defined in (8), the square of the transition density is also Gaussian up to the parameter  $1/(2\sqrt{\pi\sigma_\eta^2})$  with mean  $\phi x$  and variance  $\sigma_\eta^2/2$ . Hence, we are interested in computing the following Fourier transform:

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

By integration of the Gaussian density, we have that  $\tilde{\Pi}_\theta(x) = 1/(2\sqrt{\pi\sigma_\eta^2}) \forall x$ , which is integrable on  $\mathbb{L}_1(A)$ . Nevertheless, for the cases 1 and 3 (super smooth noises), Assumptions A 2 and A 3(i)-(iii) are not satisfied since  $x \mapsto (\tilde{\Pi}_\theta(x))^*/f_\varepsilon^*(x)$  is not integrable despite the fact that the numerator and denominator taken separately can be integrated. In this case, we introduce a weight function  $\varphi$  belongs to  $\mathcal{S}(\mathbb{R})$ , where  $\mathcal{S}(\mathbb{R})$  is the Schwartz space of functions defined by  $\mathcal{S}(\mathbb{R}) = \{f \in \mathcal{C}^\infty(\mathbb{R}), \forall \alpha, N \text{ there exists } C_{N,\alpha} \text{ s.t. } |\nabla_x^\alpha f(x)| \leq C_{N,\alpha}(1 + |x|)^{-N}\}$ .

Hence,  $\forall \varphi \in \mathcal{S}(\mathbb{R})$ , we have

$$\left\langle \varphi, \tilde{\Pi}_\theta^* \right\rangle = \int \varphi(x) dx \int \tilde{\Pi}_\theta(u) e^{-ixu} du = \frac{1}{2\sqrt{\pi\sigma_\eta^2}} \int \varphi(x) dx \int e^{-ixu} du = \frac{1}{2\sqrt{\pi\sigma_\eta^2}} \left\langle \varphi, \delta_0 \right\rangle,$$

where  $\delta_x$  is the Dirac distribution at point  $x$ .

Hence, by taking  $\varphi : u \mapsto \tilde{\varphi}(u) e^{ixu} / f_\varepsilon^*(-u) \in \mathcal{S}(\mathbb{R})$  with  $\tilde{\varphi} : u \mapsto 2\pi e^{-\sigma_\varepsilon^2 u^2}$ , we obtain the operator  $Q$  as follows

$$Q_{\Pi_\theta^2}(x) = \frac{1}{2\pi} \int e^{ixu} \frac{\tilde{\varphi}(u) \tilde{\Pi}_\theta^*(u)}{f_\varepsilon^*(-u)} du = \left\langle \varphi, \tilde{\Pi}_\theta^* \right\rangle = \frac{1}{2\sqrt{\pi\sigma_\eta^2}} \left\langle \varphi, \delta_0 \right\rangle = \frac{1}{2\sqrt{\pi\sigma_\eta^2}} \varphi(0),$$

where  $\varphi(0) = 1$  for all cases in Section 4. Here, we take  $\tilde{\varphi}$  dependent of  $\sigma_\varepsilon^2$  since we assume that this variance is known but one can take any function  $\tilde{\varphi}$  such that  $\tilde{\varphi}/f_\varepsilon^*$  is in  $\mathcal{S}$ .

For  $\Pi_\theta^*(x, y)$  we make the same analogy, that is let  $\Pi_{u,\theta}(v)$  the function  $v \mapsto \Pi_\theta(u, v) \forall u$ . For the Gaussian transition density  $\Pi_\theta$  we have  $\forall u$ ,

$$(\Pi_{u,\theta}(y))^* = \int e^{-iyv} \Pi_\theta(u, v) dv = \int e^{-iyv} \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{(v-\phi u)^2}{2\sigma_\eta^2}} dv = e^{(-i\phi u y - \frac{\sigma_\eta^2}{2} y^2)}.$$

Let  $\Pi_{y,\theta}(u)$  be the function  $u \mapsto (\Pi_{u,\theta}(y))^* \forall y$ . Then, we have  $\forall \varphi \in \mathcal{S}$  and  $\forall y$

$$\begin{aligned} \left\langle \varphi, \Pi_{y,\theta} \right\rangle &= \int \varphi(x) dx \int e^{-ixu} (\Pi_{u,\theta}(y))^* du \\ &= \int \varphi(x) dx \int e^{-ixu} e^{-i\phi u y - \frac{\sigma_\eta^2}{2} y^2} du \\ &= e^{-\frac{\sigma_\eta^2}{2} y^2} \int \varphi(z - \phi y) dx \int e^{-iuz} du \\ &= e^{-\frac{\sigma_\eta^2}{2} y^2} \left\langle \varphi(\cdot - \phi y), \mathbf{1}^* \right\rangle \\ &= e^{-\frac{\sigma_\eta^2}{2} y^2} \left\langle \varphi(\cdot - \phi y), \delta_0 \right\rangle = e^{-\frac{\sigma_\eta^2}{2} y^2} \varphi(-\phi y). \end{aligned}$$

Hence, the operator  $V_{\Pi_\theta}$  is obtained as follows for the case 1, *i.e.*,

$$\begin{aligned}
V_{\Pi_\theta}(x, y) &= \frac{1}{4\pi^2} \int \int \tilde{\varphi}_1(u) \tilde{\varphi}_2(v) e^{i(xu+yv)} \frac{\Pi_\theta^*(u, v)}{f_\varepsilon^*(-u) f_\varepsilon^*(-v)} du dv \\
&= \frac{1}{4\pi^2} \int \int \frac{\tilde{\varphi}_1(u) \tilde{\varphi}_2(v)}{f_\varepsilon^*(-u) f_\varepsilon^*(-v)} e^{i(xu+yv)} \left( \int e^{-i w u} (\Pi_{w, \theta}(v))^* dw \right) du dv \\
&= \frac{1}{4\pi^2} \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v} \left( \int \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i x u} \left( \int e^{-i w u - i \phi w v - \frac{\sigma_\eta^2}{2} v^2} dw \right) du \right) dv \\
&= \frac{1}{4\pi^2} \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v - \frac{\sigma_\eta^2}{2} v^2} \left( \int \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i x u} \left( \int e^{-i w (u + \phi v)} dw \right) du \right) dv \\
&= \frac{1}{4\pi^2} \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v - \frac{\sigma_\eta^2}{2} v^2} \left( \int \frac{\tilde{\varphi}_1(z - \phi v)}{f_\varepsilon^*(\phi v - z)} e^{i x (z - \phi v)} \left( \int e^{-i w z} dw \right) dz \right) dv \\
&= \frac{1}{2\pi} \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v - \frac{\sigma_\eta^2}{2} v^2} \left( \left\langle \varphi(\cdot - \phi v), \mathbf{1}^* \right\rangle \right) dv \\
&= \frac{1}{2\pi} \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v - \frac{\sigma_\eta^2}{2} v^2} \varphi(-\phi v) dv \\
&= \frac{1}{2\pi} \int e^{-i v y} \left( e^{i v (y - \phi x) - (\sigma_\eta^2 - \sigma_\varepsilon^2 (1 + \phi^2)) \frac{v^2}{2}} \right) dv \\
&= \frac{1}{2\pi} \int e^{i v y} \left( e^{-i v (y - \phi x) - (\sigma_\eta^2 - \sigma_\varepsilon^2 (1 + \phi^2)) \frac{v^2}{2}} \right) dv \\
&= \frac{1}{\sqrt{2\pi(\sigma_\eta^2 - \sigma_\varepsilon^2(1 + \phi^2))}} \exp \left( - \frac{(y - \phi x)^2}{2(\sigma_\eta^2 - \sigma_\varepsilon^2(1 + \phi^2))} \right),
\end{aligned}$$

where  $\varphi : u \mapsto e^{i x u} \tilde{\varphi}_1(u) / f_\varepsilon^*(-u)$  with  $\tilde{\varphi}_1 : u \mapsto 2\pi e^{-\sigma_\varepsilon^2 u^2}$  and  $\tilde{\varphi}_2 : v \mapsto e^{-i v y - \sigma_\varepsilon^2 v^2}$  and such that  $\varphi, \varphi_1$  and  $\varphi_2 \in \mathcal{S}$ . For the cases 2 and 3, one can make the same computations by replacing  $f_\varepsilon^*$  by its expression given in Section 4.

**Checking assumptions A1–A3:** By inspecting the function  $b_{\theta_0} : x \mapsto \phi_0 x$  one can easily see that regularity assumptions are well satisfied and, if  $\phi_0$  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  $\mathbf{P}m_\theta$  implies that  $\theta_0$  is a minimum and Assumption **A1(iv)** also requires to compute the Hessian matrix belonging to  $\mathcal{S}_{2 \times 2}^{ym}$  (where  $\mathcal{S}^{ym}$  represents the space of symmetric matrix). The stationary density  $f_{\theta_0}$  is here a centered Gaussian density with zero mean and variance  $\sigma_{0,\eta}^2 / (1 - \phi_0^2)$ , so the Hessian matrix  $\mathcal{V}_{\theta_0}$  is given by

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

(see Corollary 3.1). Nevertheless, we assume here that  $f_{\theta_0}$  is unknown, so the Hessian matrix is consistently estimated by

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

The computation of this matrix can be easily done for Gaussian AR processes whatever the noises since all derivatives of the Gaussian densities are explicit.

As we have pointed out, the integrability Assumptions **A2** and **A3(i)** and **(iii)** are not satisfied for Gaussian AR processes with super smooth noises (cases 1 and 3), hence the introduction in practice of a weight function  $\varphi$  belonging to the Swchartz space  $\mathcal{S}$  is then mandatory. On the other hand, for Laplace noises the convergence towards zero of the modulus of the Fourier transform is polynomial and the functions  $(\Pi_\theta^*/f_\varepsilon^*)$ ,  $(\partial\Pi_\theta/\partial\theta_j)^*/f_\varepsilon^*$  and  $(\partial^2\Pi_\theta/\partial\theta_j\partial\theta_l)^*/f_\varepsilon^*$  have the following form  $C_1(\theta)P(x)\exp(-C_2(\theta)x^2)$  (meaning that they are super smooth and so integrable) where  $C_1(\theta)$  and  $C_2(\theta)$  are two constants well-defined in the compact parameter set  $\Theta$  and  $P(x)$  a polynomial function independent of  $\theta$ . Hence, moment conditions and local dominance are satisfied.

**Contrasts for the nonlinear AR simulations:** Consider the nonlinear process in (11). For this model the transition density is Gaussian with mean  $b_{\theta_0}(x) = \phi_0 \sin(x)$  and variance  $\sigma_{0,\eta}^2$ . In the same manner the square of the transition density is also Gaussian with the same mean  $b_{\theta_0}(x)$  and variance  $\sigma_{0,\eta}^2/2$  up to the constant of normalization  $1/(2\sqrt{\pi}\sigma_{0,\eta})$ . Hence, the computation of the operator  $Q_{\Pi_\theta^2}$  remains unchanged, and we have to compute the operator  $V_{\Pi_\theta}$ . Because of the nonlinearity of the drift function, this operator does not admit an explicit form and is given by

$$\begin{aligned}
V_{\Pi_\theta}(x, y) &= \frac{1}{4\pi^2} \int \int \tilde{\varphi}_1(u) \tilde{\varphi}_2(v) e^{i(xu+yv)} \frac{\Pi_\theta^*(u, v)}{f_\varepsilon^*(-u) f_\varepsilon^*(-v)} dudv \\
&= \frac{1}{4\pi^2} \int \int \frac{\tilde{\varphi}_1(u) \tilde{\varphi}_2(v)}{f_\varepsilon^*(-u) f_\varepsilon^*(-v)} e^{i(xu+yv)} \left( \int e^{-i w u} (\Pi_{w, \theta}(v))^* dw \right) dudv \\
&= \frac{1}{4\pi^2} \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v} \left( \int \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i x u} \left( \int e^{-i w u - i v b_\theta(w) - \frac{\sigma_\eta^2}{2} v^2} dw \right) du \right) dv \\
&= \frac{1}{4\pi^2} \iint \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i u(x-w)} \left( \int \frac{\tilde{\varphi}_2(v)}{f_\varepsilon^*(-v)} e^{i y v - i b_\theta(w) v - \frac{\sigma_\eta^2}{2} v^2} dv \right) dw du \\
&= \frac{1}{4\pi^2} \iint \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i u(x-w)} \left( \int e^{i y v} \left( e^{-i v(y - b_\theta(w)) - \frac{(\sigma_\eta^2 + \sigma_\varepsilon^2)}{2} v^2} \right) dv \right) dw du \\
&= \frac{1}{4\pi^2} \iint \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i u(x-w)} \left( 2\pi \frac{1}{\sqrt{2\pi(\sigma_\eta^2 + \sigma_\varepsilon^2)}} e^{-\frac{(y - b_\theta(w))^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}} \right) dw du \\
&= \frac{1}{2\pi} \frac{1}{\sqrt{2\pi(\sigma_\eta^2 + \sigma_\varepsilon^2)}} \int \frac{\tilde{\varphi}_1(u)}{f_\varepsilon^*(-u)} e^{i u x} \left( \int e^{-i u w} e^{-\frac{(y - b_\theta(w))^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}} dw \right) du \\
&= \int \varphi(u) \left( \int e^{-i u w} \frac{1}{\sqrt{2\pi(\sigma_\eta^2 + \sigma_\varepsilon^2)}} e^{-\frac{(y - b_\theta(w))^2}{2(\sigma_\eta^2 + \sigma_\varepsilon^2)}} dw \right) du
\end{aligned}$$

where  $\varphi : u \mapsto 2\pi e^{i x u} e^{-\sigma_\varepsilon^2 u^2} / f_\varepsilon^*(-u)$  and  $\tilde{\varphi}_2 : v \mapsto e^{-i y v - \sigma_\varepsilon^2 v^2}$ .

## 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] C. S. Bos and N. Shephard. Inference for adaptive time series models: Stochastic volatility and conditionally gaussian state space form. *Economet. Rev.*, 25(2-3):219–244, 2006.

- 
- [3] E. Brunel, F. Comte, and C. Lacour. Adaptive estimation of the conditional density in the presence of censoring. *Sankhyā: The Indian Journal of Statistics*, pages 734–763, 2007.
- [4] E. Burmeister and K. D. Wall. Kalman filtering estimation of unobserved rational expectations with an application to the german hyperinflation. *J. Econometrics*, 20(2):255–284, 1982.
- [5] E. Burmeister, K. D. Wall, and J. D. Hamilton. Estimation of unobserved expected monthly inflation using kalman filtering. *J. Bus. Econ. Stat.*, 4(2):147–160, 1986.
- [6] C. Butucea and C. Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11(2):309–340, 2005.
- [7] C. Butucea and C. Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11(2):309–340, 2005.
- [8] J. C. Chan. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. 35(1):17–28, 2017.
- [9] K. C. Chanda. Large sample analysis of autoregressive moving-average models with errors in variables. *J. Time Ser. Anal.*, 16(1):1–15, 1995.
- [10] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. Smc2: an efficient algorithm for sequential analysis of state space models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(3):397–426, 2013.
- [11] F. Comte and M.-L. Taupin. Semiparametric estimation in the (auto)-regressive  $\beta$ -mixing model with errors-in-variables. *Math. Methods Statist.*, 10(2):121–160, 2001.
- [12] M. Costa and T. Alpuim. Parameter estimation of state space models for univariate observations. *J. Statist. Plann. Inference*, 140(7):1889–1902, 2010.
- [13] J. Dedecker, A. Samson, and M.-L. Taupin. Estimation in autoregressive model with measurement error. *ESAIM Prob. Stat.*, 18:277–307, 2014.
- [14] A. Delaigle and P. Hall. Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 231–252, 2016.
- [15] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128, 1999.
- [16] 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.
- [17] A. Doucet, N. De Freitas, and N. Gordon. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.
- [18] P. Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. Properties and examples.

- 
- [19] B. Ebner, B. Klar, and S. G. Meintanis. Fourier inference for stochastic volatility models with heavy-tailed innovations. *Statistical Papers*, 59(3):1043–1060, 2018.
- [20] S. Efromovich. Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92(438):526–535, 1997.
- [21] S. El Kolei. Parametric estimation of hidden stochastic model by contrast minimization and deconvolution. *Metrika*, 76(8):1031–1081, 2013.
- [22] S. El Kolei and F. Patras. Analysis, detection and correction of misspecified discrete time state space models. *J. Comput. Appl. Math.*, 333:200–214, 2018.
- [23] S. El Kolei and F. Pelgrin. Parametric inference of autoregressive heteroscedastic models with errors in variables. *Statist. Probab. Lett.*, 130:63–70, 2017.
- [24] G.-L. Fan, H.-Y. Liang, and J.-F. Wang. Empirical likelihood for heteroscedastic partially linear errors-in-variables model with  $\alpha$ -mixing errors. *Statistical Papers*, 54(1):85–112, 2013.
- [25] J. Fan, Y. K. Truong, and Y. Wang. Nonparametric function estimation involving errors-in-variables. In *Nonparametric Functional Estimation and Related Topics*, pages 613–627. Springer, 1991.
- [26] V. Genon-Catalot, T. Jeantheau, and C. Larédo. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*, 6(6):1051–1079, 2000.
- [27] E. Ghysels, A. C. Harvey, and E. Renault. Stochastic volatility. *Handbook of statistics*, 14:119–191, 1996.
- [28] J. D. Hamilton. State-space models. *Handbook of econometrics*, 4:3039–3080, 1994.
- [29] J. D. Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, NJ, 1994.
- [30] E. J. Hannan and J. Rissanen. Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69(1):81–94, 1982.
- [31] F. Hayashi. *Econometrics*. Princeton University Press, Princeton, NJ, 2000.
- [32] K. Holston, T. Laubach, and J. C. Williams. Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108:S59 – S75, 2017. 39th Annual NBER International Seminar on Macroeconomics.
- [33] G. L. Jones. On the Markov chain central limit theorem. *Probab. Surv.*, 1:299–320, 2004.
- [34] J. Kappus and G. Mabon. Adaptive density estimation in deconvolution problems with unknown error distribution. *Electronic journal of statistics*, 8(2):2879–2904, 2014.
- [35] C. Lacour. Adaptive estimation of the transition density of a markov chain. In *Annales de l’IHP Probabilités et statistiques*, volume 43, pages 571–597, 2007.



- [36] C. Lacour. Adaptive estimation of the transition density of a particular hidden markov chain. *Journal of Multivariate Analysis*, 99(5):787–814, 2008.
- [37] C. Lacour. Least squares type estimation of the transition density of a particular hidden Markov chain. *Electron. J. Stat.*, 2:1–39, 2008.
- [38] C. Lacour. Nonparametric estimation of the stationary density and the transition density of a markov chain. *Stochastic Processes and their Applications*, 118(2):232–260, 2008.
- [39] T. Laubach and J. C. Williams. Measuring the natural rate of interest. *The Review of Economics and Statistics*, 85(4):1063–1070, 2003.
- [40] H. W. Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402, 1967.
- [41] F. Lindsten. An efficient stochastic approximation em algorithm using conditional particle filters. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6274–6278. IEEE, 2013.
- [42] A. Meister. On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, 32(4):439–449, 2004.
- [43] Y. Miao, F. Zhao, K. Wang, and Y. Chen. Asymptotic normality and strong consistency of ls estimators in the ev regression model with na errors. *Statistical Papers*, 54(1):193–206, 2013.
- [44] A. Mokkadem et al. Sur un modèle autorégressif non linéaire, ergodicité et ergodicité géométrique. *Journal of time series analysis*, 8(2):195–204, 1987.
- [45] M. H. Neumann and O. Hössjer. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997.
- [46] 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.
- [47] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.
- [48] 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.
- [49] Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage: fast and efficient likelihood inference. *J. Econometrics*, 140(2):425–449, 2007.
- [50] F. Pivetta and R. Reis. The persistence of inflation in the united states. *Journal of Economic dynamics and control*, 31(4):1326–1358, 2007.

- 
- [51] J. Staudenmayer and J. P. Buonaccorsi. Measurement error in linear autoregressive models. *J. Amer. Statist. Assoc.*, 100(471):841–852, 2005.
- [52] J. H. Stock and M. W. Watson. Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39:3–33, 2007.
- [53] S. Taylor. *Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices.*, volume 1, pages 203–226. Oxford University Press, 2005.
- [54] A. Trapletti and K. Hornik. *tseries: Time Series Analysis and Computational Finance*, 2019. R package version 0.10-47.
- [55] A. W. Van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [56] H. Zhou, J. You, and B. Zhou. Statistical inference for fixed-effects partially linear regression models with errors in variables. *Statistical Papers*, 51(3):629–650, 2010.