

Parametric estimation of hidden stochastic model by contrast minimization and deconvolution

Application to the stochastic volatility model

Salima El Kolei

Received: 7 March 2012 / Published online: 6 March 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract We study a new parametric approach for particular hidden stochastic models. This method is based on contrast minimization and deconvolution and can be applied, for example, for ecological and financial state space models. After proving consistency and asymptotic normality of the estimation leading to asymptotic confidence intervals, we provide a thorough numerical study, which compares most of the classical methods that are used in practice (Quasi-Maximum Likelihood estimator, Simulated Expectation Maximization Likelihood estimator and Bayesian estimators) to estimate the Stochastic Volatility model. We prove that our estimator clearly outperforms the Maximum Likelihood Estimator in term of computing time, but also most of the other methods. We also show that this contrast method is the most robust with respect to non Gaussianity of the error and also does not need any tuning parameter.

Keywords Contrast function · Deconvolution · Parametric inference · Stochastic volatility

1 Introduction

This paper is concerned with the particular *hidden stochastic model*:

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

S. El Kolei (✉)
Laboratoire de Mathématiques J.A. Dieudonné, UMR n 7351 CNRS UNSA,
Université de Nice, Sophia Antipolis, 06108 Nice Cedex 02, France
e-mail: salima@unice.fr

where $(\varepsilon_i)_{i \geq 1}$ and $(\eta_i)_{i \geq 1}$ are two independent sequences of independent and identically distributed (i.i.d) centered random variables with variance σ_ε^2 and σ_η^2 . It is assumed that the variance σ_ε^2 is known. The terminology *hidden* comes from the unobservable character of the process $(X_i)_{i \geq 1}$ since the only available observations are Y_1, \dots, Y_n .

The dynamics of the process X_i is described by a measurable function b_{ϕ_0} which depends on an unknown parameter ϕ_0 and by a sequence of i.i.d centered random variables with unknown variance σ_0^2 . We denote by θ_0 the vector of parameters governing the process X_i and suppose that the model is correctly specified: that is, θ_0 belongs to the parameter space $\Theta \subset \mathbb{R}^r$, with $r \in \mathbb{N}^*$.

Inference in hidden Markov models is a real challenge and has been studied by many authors (see Cappé et al. 2005a; Doucet et al. 2001; Robert et al. 2000). Chanda provided in (1995) an asymptotically normal estimator for the vector of parameters θ_0 by using modified Yule Walker equation but it assumes that the function b_{ϕ_0} is linear in ϕ_0 and X_i , so the model (1) is reduced to an autoregressive model with measurement error.

Recently, in Douc et al. (2011), the authors propose an efficient estimator of the vector of parameters θ_0 for nonlinear function b_{ϕ_0} . They prove that their *Maximisation Likelihood Estimator* (MLE) is consistent and asymptotically normal. The main difficulty with their approach comes from the unobservable character of the process X_i making the calculus of the exact likelihood intractable in practice: the likelihood is only available in the form of a multiple integral, so exact likelihood methods require simulations and have therefore an intensive computational cost. In many case, the MLE has to be approximated. A popular approach to approximate the MLE consists in using Monte Carlo Markov Chain (MCMC) simulation techniques. Thanks to the development of these methods, the MLE has known a huge progress and Bayesian estimations have received more attention (see Smith and Roberts 1993). Another method for performing the MLE consists in using the Expectation-Maximization (EM) algorithm proposed by Dempster et al. (1977). Nevertheless, since X_i is unobservable, this method requires to introduce a MCMC in the Expectation step. Although these methods are used in practice, they are expensive from a computational point of view.

Some authors have proposed Sequential Monte Carlo algorithms (SMC) known as Particles Filtering methods which allow to approximate the likelihood. The computational cost is reduced by a recursive construction. We refer to the book of Doucet et al. (2001) and Cappé et al. (2005a) for a complete review of these methods.

Particle Markov Chain Monte Carlo (PMCMC) is another method for estimating the model (1). This method combines Particles filtering methods and MCMC methods to estimate the vector of parameters θ_0 . From a computational point of view, this approach is expensive and we refer the reader to Andrieu et al. (2010) for more details. In Peters et al. (2010), they propose an adaptive PMCMC method to estimate ecological hidden stochastic models.

We propose here an approach based on M-estimation: It consists in the optimisation of a well-chosen contrast function (see Van der Vaart 1998, chapter p. 41 for a partial review) and deconvolution strategy. The deconvolution problem is encountered in many statistical situations where the observations are collected with random errors. In this approach, a method for estimating the parameter ϕ_0 has been proposed by

Comte and Taupin (2001). Their procedure of estimation is based on a modified least squared minimization. In the same perspective, Dedecker et al. in (2011) propose also the same procedure of estimation based on a weighted least squared estimation: Their assumptions on the process X_i are less restrictive than those proposed by F. Comte and M. Taupin and they provide consistent estimation of the parameter ϕ_0 with a parametric rate of convergence in a very general framework. Their general estimator is based on the introduction of a kernel deconvolution density and depends on the choice of a weight function.

The approach proposed here is different: it is not based on a weighted least squared estimation so that the choice of the weight function is not encountered in this paper. Moreover, it allows to estimate both the parameters ϕ_0 and σ_0^2 . Our principle of estimation relies on the Nadaraya–Watson strategy and is proposed by Comte et al. in (2011) in a non parametric case to estimate the function b_ϕ as a ratio of an estimate of $l_\theta = b_\phi f_\theta$ and an estimate of f_θ , where f_θ represents the invariant density of the X_i . We propose to adapt their approach in a parametric context and suppose that the form of the stationary density f_{θ_0} is known up to some unknown parameter θ_0 . Our work is purely parametric but we go further in this direction by proposing an analytical expression of the asymptotic variance matrix $\Sigma(\hat{\theta}_n)$ which allows to construct confidence interval. Furthermore, this approach is much less greedy from a computational point of view than the MLE and its implementation is straightforward.

Applications: Applications include epidemiology, meteorology, neuroscience, ecology (see Ionides et al. 2011) and finance (see Johannes et al. 2009). For example, our approach can be applied to the five ecological state space models described in Peters et al. (2010). Although the scope of our method is general, we have chosen to focus on the so-called discrete time Stochastic Volatility model (SV) introduced by Taylor (2005). We also investigate the behavior of our method on the simpler autoregressive process AR(1) with measurement noise which has been widely studied and on which our method can be more easily understood and compared with other ones. Our procedure allows to estimate the parameters of a large class of discrete Stochastic Volatility models (ARCH-E model, Vasicek model, Merton model..), which is a real challenge in financial application.

(i) *Gaussian Autoregressive AR(1) with measurement noise:* It has the following form:

$$\begin{cases} Y_{i+1} = X_{i+1} + \varepsilon_{i+1} \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \tag{2}$$

where ε_{i+1} and η_{i+1} are two centered Gaussian random variables with variance σ_ε^2 assumed to be known and σ_0^2 assumed to be unknown. Additionally, we assume that $|\phi_0| < 1$ which implies the stationary and ergodic property of the process X_i (see Dedecker et al. 2007).

(ii) *SV model:* It is directly connected to the type of diffusion process used in asset-pricing theory (see Melino and Turnbull 1990):

$$\begin{cases} R_{i+1} = \exp\left(\frac{X_{i+1}}{2}\right) \xi_{i+1}, \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \tag{3}$$

where ξ_{i+1} and η_{i+1} are two centered Gaussian random variables with variance σ_ξ^2 assumed to be known and equal to one and σ_0^2 assumed to be unknown. The variables R_{i+1} represent the returns and X_{i+1} is the log-volatility process.

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 SV model is a particular version of (1). We assume that $|\phi_0| < 1$ and we refer the reader to [Genon-Catalot et al. \(2000\)](#) for the mixing properties of stochastic volatility models.

Most of the computational problems stem from the assumptions that the innovation of the returns are Gaussian which translates into a logarithmic chi-square distribution when the model (12) is transformed in a linear state space model. Many authors have ignored it in their implementation and many authors use some mixture of Gaussian to approximate the log-chi-square density. For example, in the Quasi-Maximum Likelihood (QML) method implemented by [Jacquier et al. \(2002\)](#) and in the Simulated Expectation-Maximization Likelihood estimator proposed (SIEMLE) by [Kim et al. \(1994\)](#) they used a mixture of Gaussian distribution to approximate the log-chi-square distribution. [Harvey \(1994\)](#) used the Kalman filter to estimate the likelihood of the transform state space model, hence the model was also assumed to be Gaussian.

Organization of the paper: The first purpose of the paper is to present our estimator and its statistical properties in Sect. 1.1: Under weak assumptions, we show that it is a consistent and asymptotically normal estimator (Table 1).

The second purpose of this paper consists in comparing our contrast estimator with different estimations: the QML, the SIEMLE and Bayesian estimators. Section 2 contains the numerical study: In Sect. 2.4 we give the parameter estimates and the comparison with others ones for simulation data and Sect. 2.6 contains the study on real data. We compare our contrast estimator with other ones on the SP&500 and FTSE index. From a computational point of view, we show that the implementation of our estimator is straightforward and it is faster than the SIEMLE (see Table 2 in Sect. 2.5.1). Besides, we show that our estimator outperforms the QML and Bayesian estimators.

Notations: We denote by: $u^*(t) = \int e^{itx} u(x) dx$ the Fourier transform of the function $u(x)$ and $\langle u, v \rangle = \int u(x) \bar{v}(x) dx$ with $v\bar{v} = |v|^2$. We write $\|u\|_2 = (\int |u(x)|^2 dx)^{1/2}$ the norm of $u(x)$ on the space of functions $\mathbb{L}^2(\mathbb{R})$. By property of the Fourier transform, we have $(u^*)^*(x) = 2\pi u(-x)$ and $\langle u_1, u_2 \rangle = \frac{1}{2\pi} \langle u_1^*, u_2^* \rangle$. The vector of the partial derivatives of f with respect to (w.r.t) θ is denoted by $\nabla_\theta f$ and the Hessian matrix of f w.r.t θ is denoted by $\nabla_\theta^2 f$. The Euclidean norm matrix, that is, the square root of the sum of the squares of all its elements will be written by $\|A\|$. We denote by \mathbf{Y}_i the pair (Y_i, Y_{i+1}) and $\mathbf{y}_i = (y_i, y_{i+1})$ is a given realisation of \mathbf{Y}_i .

In the following, $\mathbb{P}, \mathbb{E}, \text{Var}$ and Cov denote respectively the probability \mathbb{P}_{θ_0} , the expected value \mathbb{E}_{θ_0} , the variance Var_{θ_0} and the covariance Cov_{θ_0} when the true parameter is θ_0 . Additionally, we write \mathbf{P}_n (resp. \mathbf{P}) the empirical expectation (resp. theoretical), that is: for any stochastic variable X : $\mathbf{P}_n(X) = \frac{1}{n} \sum_{i=1}^n X_i$ (resp. $\mathbf{P}(X) = \mathbb{E}[X]$).

1.1 Procedure: contrast estimator

Hereafter, we propose explicit estimators of the parameter θ_0 . This estimator called the contrast estimator is based on minimization of suitable functions of the observations usually called “contrasts functions”. We refer the reader to [Van der Vaart \(1998\)](#) for a general account on this notion. Furthermore, in this part, we use the contrast function proposed by [Comte et al. \(2010\)](#), that is:

$$\mathbf{P}_n m_\theta = \frac{1}{n} \sum_{i=1}^n m_\theta(\mathbf{Y}_i), \tag{4}$$

with n the number of observations and:

$$m_\theta(\mathbf{y}_i) : (\theta, \mathbf{y}_i) \in (\Theta \times \mathbb{R}^2) \mapsto m_\theta(\mathbf{y}_i) = \|l_\theta\|_2^2 - 2y_{i+1}u_{l_\theta}^*(y_i),$$

where the function l_θ and u_v are given by:

$$l_\theta(x) = b_\phi(x) f_\theta(x) \quad \text{and} \quad u_v(x) = \frac{1}{2\pi} \frac{v^*(-x)}{f_\varepsilon^*(x)} \tag{5}$$

with f_θ the invariant density of X_i .

Some assumptions. As our procedure relies on the Fourier deconvolution strategy, in order to construct our estimator, we assume that the density of the noise ε_i , denoted by f_ε , is fully known and belongs to $\mathbb{L}_2(\mathbb{R})$, and for all $x \in \mathbb{R}$ $f_\varepsilon^*(x) \neq 0$. Furthermore, we assume that the function l_θ belongs to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$. The function u_{l_θ} must be integrable.

For the statistical study, the key assumption is that the process $(X_i)_{i \geq 1}$ is stationary and ergodic (see [Genon-Catalot et al. 2000](#) for a definition).

Remark 1 In this paper we consider the situation in which the observation noise variance is known. This assumption which is not in general the case in practice is necessary for the identifiability of the model and so is a standard assumption for state-space models given in (1).

There is some restrictions on the distribution of the observation and process errors in the Nadaraya-Watson approach. It is known that the rate of convergence for estimating the function l_θ is related to the rate of decrease of f_ε^* . In particular, the smoother f_ε , the slower the rate of convergence for estimating is (The Gaussian, log-chi squared or Cauchy distributions are super-smooth. The Laplace distribution is ordinary smooth). This rate of convergence can be improved by assuming some additional regularity conditions on l_θ . There is a good discussion about this subject in [Comte et al. \(2010\)](#) and [Dedecker et al. \(2011\)](#).

The procedure Let us explain the choice of the contrast function and how the strategy of deconvolution works. Obviously, as the model (1) shows, the \mathbf{Y}_i are not i.i.d. However, by assumption, they are stationary ergodic, so the convergence of $\mathbf{P}_n m_\theta$

to $\mathbf{P}m_\theta = \mathbb{E}[m_\theta(\mathbf{Y}_1)]$ as n tends to the infinity is provided by the Ergodic Theorem. Moreover, the limit $\mathbb{E}[m_\theta(\mathbf{Y}_1)]$ of the contrast function can be explicitly computed:

$$\mathbb{E}[m_\theta(\mathbf{Y}_1)] = \|l_\theta\|_2^2 - 2\mathbb{E}[Y_2 u_{l_\theta}^*(Y_1)].$$

By Eq. (1) and under the independence assumptions of the noise (ε_2) and (η_2) , we have:

$$\mathbb{E}[Y_2 u_{l_\theta}^*(Y_1)] = \mathbb{E}[b_{\phi_0}(X_1) u_{l_\theta}^*(Y_1)].$$

Using Fubini's Theorem and Eq. (1), we obtain:

$$\begin{aligned} \mathbb{E}[b_{\phi_0}(X_1) u_{l_\theta}^*(Y_1)] &= \mathbb{E}\left[b_{\phi_0}(X_1) \int e^{iY_1 z} u_{l_\theta}(z) dz\right] \\ &= \mathbb{E}\left[b_{\phi_0}(X_1) \int \frac{1}{2\pi} \frac{1}{f_\varepsilon^*(z)} e^{iY_1 z} (l_\theta(-z))^* dz\right] \\ &= \frac{1}{2\pi} \int \mathbb{E}\left[b_{\phi_0}(X_1) e^{i(X_1 + \varepsilon_1)z}\right] \frac{1}{f_\varepsilon^*(z)} (l_\theta(-z))^* dz \\ &= \frac{1}{2\pi} \int \frac{\mathbb{E}[e^{i\varepsilon_1 z}]}{f_\varepsilon^*(z)} \mathbb{E}\left[b_{\phi_0}(X_1) e^{iX_1 z}\right] (l_\theta(-z))^* dz \\ &= \frac{1}{2\pi} \mathbb{E}\left[b_{\phi_0}(X_1) \int e^{iX_1 z} (l_\theta(-z))^* dz\right] \\ &= \frac{1}{2\pi} \mathbb{E}\left[b_{\phi_0}(X_1) ((l_\theta(-X_1))^*)^*\right] \\ &= \mathbb{E}[b_{\phi_0}(X_1) l_\theta(X_1)]. \\ &= \int b_{\phi_0}(x) f_{\theta_0}(x) b_\phi(x) f_\theta(x) dx \\ &= \langle l_\theta, l_{\theta_0} \rangle. \end{aligned} \tag{6}$$

Then,

$$\mathbb{E}[m_\theta(\mathbf{Y}_1)] = \|l_\theta\|_2^2 - 2\langle l_\theta, l_{\theta_0} \rangle, \tag{7}$$

$$= \|l_\theta - l_{\theta_0}\|_2^2 - \|l_{\theta_0}\|_2^2. \tag{8}$$

Under the uniqueness assumption (CT) given just later this quantity is minimal when $\theta = \theta_0$. Hence, the associated minimum-contrast estimators $\hat{\theta}_n$ is defined as any solution of:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \mathbf{P}_n m_\theta. \tag{9}$$

Remark 2 One can see in the deconvolution strategy described in Eq. (6) that temporal correlation in the observation or latent process errors can be authorized. The procedure still be applicable but the covariance matrix $\Omega_{j-1}(\theta_0)$ for the CLT has not an analytic

expression in this case since the use of the Fourier deconvolution approach does not work.

We refer the reader to [Dedecker et al. \(2007\)](#) for the proof that if X_i is an ergodic process then the process Y_i , which is the sum of an ergodic process with an i.i.d. noise, is again stationary ergodic. Furthermore, by the definition of an ergodic process, if Y_i is an ergodic process then the couple $\mathbf{Y}_i = (Y_i, Y_{i+1})$ inherits the property (see [Genon-Catalot et al. 2000](#)).

1.2 Asymptotic properties of the contrast estimator

Our proof holds under the following assumptions. For the reader convenience, we denote by **(E)** [*resp.* **(C)** and **(T)**] the assumptions which serve us for the existence [*resp.* Consistency and Central Limit Theorem]. If the same assumption is needed for two results, for example for the existence and the consistency, it is denoted by **(EC)**.

(ECT): The parameter space Θ is a compact subset of \mathbb{R}^r and θ_0 is an element of the interior of Θ .

(C): (Local dominance): $\mathbb{E} \left[\sup_{\theta \in \Theta} \left| Y_2 u_{l_\theta}^* (Y_1) \right| \right] < \infty$.

(CT): The application $\theta \mapsto \mathbf{P}m_\theta$ admits an unique minimum and its Hessian matrix denoted by V_θ is non-singular in θ_0 .

(T): (Regularity): We assume that the function l_θ is twice continuously differentiable w.r.t $\theta \in \Theta$ for any x and measurable w.r.t x for all θ in Θ . Additionally, each coordinate of $\nabla_\theta l_\theta$ and each coordinate of $\nabla_\theta^2 l_\theta$ belong to $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ and each coordinate of $u_{\nabla_\theta l_\theta}$ and $u_{\nabla_\theta^2 l_\theta}$ have to be integrable as well.

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

$$\mathbb{E} \left[\left| Y_2 u_{\frac{\partial l_\theta}{\partial \theta_j}}^* (Y_1) \right|^{2+\delta} \right] < \infty$$

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

$$\mathbb{E} \left[\sup_{\theta \in \mathcal{U}} \left| Y_2 u_{\frac{\partial^2 l_\theta}{\partial \theta_j \partial \theta_k}}^* (Y_1) \right| \right] < \infty$$

Let us introduce the matrix:

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

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 1 Under our assumptions, let $\widehat{\theta}_n$ be the minimum-contrast estimator defined by (9). Then:

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

Moreover, if \mathbf{Y}_i is geometrically ergodic (see Definition 1 in “Appendix A”), then:

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

The following corollary gives an expression of the matrix $\Omega(\theta_0)$ and V_{θ_0} of Theorem 1 for the practical implementation:

Corollary 1 Under our assumptions, the matrix $\Omega(\theta_0)$ is given by:

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

where:

$$\Omega_0(\theta_0) = 4\mathbb{E} \left[Y_2^2 (u_{\nabla_{\theta} l_{\theta}}^*(Y_1))^2 \right] - 4\mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1))] \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1))]',$$

and, the covariance terms are given by:

$$\Omega_{j-1}(\theta_0) = 4 \left[\tilde{C}_{j-1} - \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1))] \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1))]' \right],$$

where $\tilde{C}_{j-1} = \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1)) (b_{\phi_0}(X_j) \nabla_{\theta} l_{\theta}(X_j))']$ and the differential $\nabla_{\theta} l_{\theta}$ is taken at point $\theta = \theta_0$.

Furthermore, the Hessian matrix V_{θ_0} is given by:

$$([V_{\theta_0}]_{j,k})_{1 \leq j,k \leq r} = 2 \left(\left\langle \frac{\partial l_{\theta}}{\partial \theta_k}, \frac{\partial l_{\theta}}{\partial \theta_j} \right\rangle \right)_{j,k} \text{ at point } \theta = \theta_0.$$

Let us now state the strategy of the proof, the full proof is given in “Appendix B”. Clearly, the proof of Theorem 1 relies on M-estimators properties and on the deconvolution strategy. The existence of our estimator follows from regularity properties of the function l_{θ} and compactness argument of the parameter space, it is explained in “Appendix B.1”. The key of the proof consists in proving the asymptotic properties of our estimator. This is done by splitting the proof into two parts: we first give the consistency result in “Appendix B.2” and then give the asymptotic normality in “Appendix B.3”. Let us introduce the principal arguments:

The main idea for proving the consistency of a M-estimator comes from the following observation: if $\mathbf{P}_n m_{\theta}$ converges to $\mathbf{P} m_{\theta}$ in probability, and if the true parameter solves the limit minimization problem, then, the limit of the argminimum $\widehat{\theta}_n$ is θ_0 . By

using an argument of uniform convergence in probability and by compactness of the parameter space, we show that the argminimum of the limit is the limit of the argminimum. A standard method to prove the uniform convergence is to use *the Uniform Law of Large Numbers* (see Lemma 1 in “Appendix A”). Combining these arguments with the dominance argument (C) give the consistency of our estimator, and then, the first part of Theorem 1.

The asymptotic normality follows essentially from Central Limit Theorem for a mixing process (see Jones 2004). Thanks to the consistency, the proof is based on a moment condition of the Jacobian vector of the function $m_\theta(\mathbf{y})$ and on a local dominance condition of its Hessian matrix. 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.

2 Applications

2.1 Contrast estimator for the Gaussian AR(1) model with measurement noise

Consider the following autoregressive process AR(1) with measurement noise:

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \tag{10}$$

The noises ε_i and η_i are supposed to be centered Gaussian randoms with variance respectively σ_ε^2 and σ_0^2 . We assume that σ_ε^2 is known. Here, the unknown vector of parameters is $\theta_0 = (\phi_0, \sigma_0^2)$ and for stationary and ergodic properties of the process X_i , we assume that the parameter ϕ_0 satisfies $|\phi_0| < 1$ (see Dedecker et al. 2007). The functions b_ϕ and l_θ are defined by:

$$\begin{aligned} b_\phi(x) &: (x, \theta) \in (\mathbb{R} \times \Theta) \mapsto b_\phi(x) = \phi x, \\ l_\theta(x) &: (x, \theta) \in (\mathbb{R} \times \Theta) \mapsto l_\theta(x) = b_\phi(x) f_\theta(x) = \frac{\phi}{\sqrt{2\pi\gamma^2}} x \exp\left(-\frac{1}{2\gamma^2} x^2\right), \end{aligned}$$

where $\gamma^2 = \frac{\sigma^2}{1-\phi^2}$. The vector of parameter θ belongs to the compact subset Θ given by $\Theta = [-1+r; 1-r] \times [\sigma_{min}^2; \sigma_{max}^2]$ with $\sigma_{min}^2 \geq \sigma_\varepsilon^2 + \bar{r}$ where $r, \bar{r}, \sigma_{min}^2$ and σ_{max}^2 are positive real constants. We consider this subset since by stationary of X_i , the parameter $|\phi| < 1$ and by construction the function $u_{l_\theta}^*$ is well defined for $\sigma^2 > \sigma_\varepsilon^2(1-\phi^2)$ with $\phi \in [-1+r; 1-r]$ which is implied by $\sigma^2 > \sigma_\varepsilon^2$. The contrast estimator defined in (1.1) has the following form:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \left\{ \frac{\phi^2 \gamma}{4\sqrt{\pi}} - \sqrt{\frac{2}{\pi}} \frac{\phi \gamma^2}{n(\gamma^2 - \sigma_\varepsilon^2)^{3/2}} \sum_{j=1}^n Y_{j+1} Y_j \exp\left(-\frac{1}{2} \frac{Y_j^2}{(\gamma^2 - \sigma_\varepsilon^2)}\right) \right\} \tag{11}$$

with n the number of observations. Theorem 1 applies for $\theta_0 = (0.7, 0.3)$ and the corresponding result for the Gaussian AR(1) model is given in ‘‘Appendix C.1’’. As we already mentioned, Corollary 1 allows to compute confidence intervals: For all $i = 1, 2$:

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

as $n \rightarrow \infty$ where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the Gaussian law, $\theta_{0,i}$ is the i th coordinate of θ_0 and \mathbf{e}_i is the i th coordinate of the vector of the canonical basis of \mathbb{R}^2 . The covariance matrix $\Sigma(\hat{\theta}_n)$ is computed in Lemma 3 in ‘‘Appendix C.1.3’’.

2.2 Contrast estimator for the SV model

We consider the following SV model:

$$\begin{cases} R_{i+1} = \exp\left(\frac{X_{i+1}}{2}\right) \xi_{i+1}, \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \tag{12}$$

The noises ξ_{i+1} and η_{i+1} are two centered Gaussian random variables with standard variance σ_ξ^2 assumed to be known and σ_0^2 . We assume that $|\phi_0| < 1$ and we refer the reader to Genon-Catalot et al. (2000) for the mixing properties of this model.

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 given by:

$$\begin{cases} Y_{i+1} = X_{i+1} + \varepsilon_{i+1} \\ X_{i+1} = \phi_0 X_i + \eta_{i+1}, \end{cases} \tag{13}$$

The Fourier transform of the noise ε_{i+1} is given by:

$$f_\varepsilon^*(x) = \frac{1}{\sqrt{\pi}} 2^{ix} \Gamma\left(\frac{1}{2} + ix\right) e^{-i\mathcal{E}x}$$

where $\mathcal{E} = \mathbb{E}[\log(\xi_{i+1}^2)] = -1.27$ and $\text{Var}[\log(\xi_{i+1}^2)] = \sigma_\varepsilon^2 = \frac{\pi^2}{2}$. Here, Γ represents the gamma function given by:

$$\Gamma : u \rightarrow \int_0^{+\infty} t^{u-1} e^{-t} dt \quad \forall u \in \mathbb{C} \text{ such that } \Re_e(u) > 0.$$

The vector of parameters $\theta = (\phi, \sigma^2)$ belongs to the compact subset Θ given by $[-1 + r; 1 - r] \times [\sigma_{min}^2; \sigma_{max}^2]$ with r, σ_{min}^2 and σ_{max}^2 positive real constants.

Our contrast estimator (1.1) is given by:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \left\{ \frac{\phi^2 \gamma}{4\sqrt{\pi}} - \frac{2}{n} \sum_{i=1}^n Y_{i+1} u_{l_\theta}^*(Y_i) \right\}, \tag{14}$$

with $u_{l_\theta}(y) = \frac{1}{2\sqrt{\pi}} \left(\frac{-i\phi y \gamma^2 \exp\left(-\frac{y^2}{2} \gamma^2\right)}{\exp(-i\mathcal{E}y) 2^{iy} \Gamma\left(\frac{1}{2} + iy\right)} \right)$.

Theorem 1 applies for $\theta_0 = (0.7, 0.3)$ and by Slutsky’s Lemma we also obtain confidence intervals. We refer the reader to “Appendix C.2” for the proof.

2.3 Comparison with the others methods

2.3.1 QML estimator

For the SV model, the QML estimator, proposed independently by Harvey et al. 1994 is based on the log-transform model given in (13). Making as if the ε_i were Gaussian in the log-transform of the SV model, the Kalman filter Kalman (1960) can be applied in order to obtain the Quasi-Maximum Likelihood function of $Y_{1:n} = (Y_1, \dots, Y_n)$ where n is the sample data length. For the AR(1) and the log-transform of the SV model, the log-likelihood $l(\theta)$ is given by:

$$l(\theta) = \log f_\theta(Y_{1:n}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log F_i - \frac{1}{2} \sum_{i=1}^n \frac{v_i^2}{F_i},$$

where v_i is the one-step ahead prediction error for Y_i , and F_i is the corresponding mean square error. More precisely, the two quantities are given by:

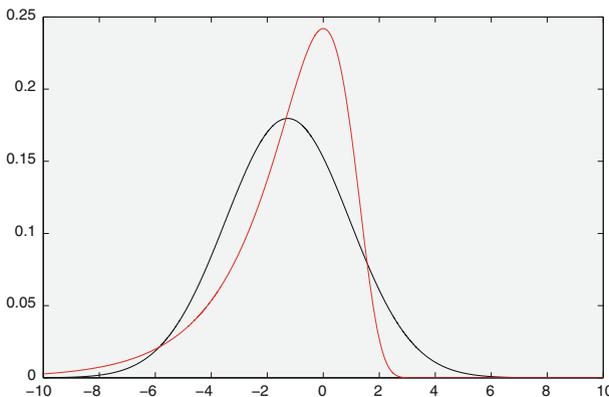


Fig. 1 Approximation of the log-chi-square density (red) by a Gaussian density with mean $\mathcal{E} = -1.27$ and variance $\sigma_{\mathcal{E}}^2 = \frac{\pi^2}{2}$ (black) (color figure online)

$$v_i = (Y_i - \hat{Y}_i^-) \quad \text{and} \quad F_i = \text{Var}_\theta[v_i] = P_i^- + \sigma_\varepsilon^2,$$

where $\hat{Y}_i^- = \mathbb{E}_\theta[Y_i|Y_{1:i-1}]$ is the one-step ahead prediction for Y_i and $P_i^- = \text{Var}_\theta[(X_i - \hat{X}_i^-)^2]$ is the one-step ahead error variance for X_i .

Hence, the associated estimator of θ_0 is defined as a solution of:

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} l(\theta).$$

Note that this procedure can be inefficient: the method does not rely on the exact likelihood of the $Z_{1:n}$ and approximating the true log-chi-square density by a normal density can be rather inappropriate (see Fig. 1 below).

2.3.2 Particle filters estimators: bootstrap, APF and KSAPF

For the particle filters, the vector of parameters $\theta = (\phi, \sigma^2)$ is supposed random obeying the prior distribution assumed to be known. We propose to use the Kitagawa and al. approach (2001 chapter 10 p. 189) in which the parameters are supposed time-varying: $\theta_{i+1} = \theta_i + \mathcal{G}_{i+1}$ where \mathcal{G}_{i+1} is a centered Gaussian random with a variance matrix Q supposed to be known. Now, we consider the augmented state vector $\tilde{X}_{i+1} = (X_{i+1}, \theta_{i+1})'$ where X_{i+1} is the hidden state variable and θ_{i+1} the unknown vector of parameters. In this paragraph, we use the terminology of the particle filtering method, that is: we call particle a random variable. The sequential particle estimation of the vector \tilde{X}_{i+1} consists in a combined estimation of X_{i+1} and θ_{i+1} . For initialisation the distribution of X_1 ¹ conditionally to θ_1 is given by the stationary density f_{θ_1} .

For the comparison with our contrast estimator (1.1), we use the three methods: the Bootstrap filter, the Auxiliary Particle filter (APF) and the Kernel Smoothing Auxiliary Particle filter (KSAPF). We refer the reader to Doucet et al. (2001), Pitt and Shephard (1999) and Liu and West (2001) for a complete revue of these methods.

Remark 3 Let us underline some particularity of the combined state and parameters estimation: For the Bootstrap and APF estimator, an important issue concerns the choice of the parameter variance Q since the parameter is itself unobservable. If one can choose an optimal variance Q the APF estimator could be a very good estimator since with arbitrary variance the results are acceptable (see Table 4). In practice, Q is chosen by an empirical optimization. The KSAPF is an enhanced version of the APF and depends on a smooth factor $0 < h < 1$ (see Liu and West 2001). Therefore, the choice of h is another problem in practice.

A common approach to estimate the vector of parameters is to maximize the likelihood. Nevertheless, for state space models, the main difficulty with the Maximum Likelihood Estimator (MLE) comes from the unobservable character of the state x_t making the calculus of the likelihood untractable in practice: the likelihood is only available in the form of a multiple integral, so exact likelihood methods require simulations and have therefore an intensive computational cost. In many cases, the MLE has

¹ To avoid confusions between the true value θ_0 and the initial value θ_1 in the Bayesian algorithms, we start the algorithms with $i = 1$.

to be approximated. A popular approach to approximate it consists in using MCMC simulation techniques (see [Smith and Roberts 1993](#); [Cappé et al. 2005b](#)). Another approach to approximate the likelihood consists in using particles filtering algorithms. Recently, in [Rue et al. \(2009\)](#) the authors propose an approach of Integrated Nested Laplace Approximations to obtain approximations of the likelihood.

In [Chopin et al. \(2011\)](#) the authors propose a sequential SMC^2 algorithm which allows an efficient approximation of the complete distribution $p(x_{0:t}, \theta|y_{1:t})$. Their approach is an extension of the Iterated Batch Importance Sampling (IBIS) proposed in [Chopin \(2002\)](#). In [Andrieu et al. \(2010\)](#) the authors develop a general algorithm which is a MCMC algorithm that uses the particles filter to approximate the intractable density $p_\theta(y_{1:n})$ combined with a MCMC step that samples from $p(\theta|y_{1:n})$. They show that their PMCMC algorithm admits as stationary density the distribution of interest $p(x_{0:t}, \theta|y_{1:t})$. There exist others methods and we refer the reader to [Johansen et al. \(2008\)](#); [Poyiadjis et al. \(2011\)](#) for more details.

2.4 A simulation study

For the AR(1) and SV model, we sample the trajectory of the X_i with the parameters $\phi_0 = 0.7$ and $\sigma_0^2 = 0.3$. Conditionally to the trajectory, we sample the variables Y_i for $i = 1 \dots n$ where n represents the number of observations. We take $n = 1000$ and $\sigma_\varepsilon^2 = 0.1$ for the two models. This means that we consider the following model:

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

with $\beta = \frac{1}{\sqrt{5\pi}}$. In this case, the Fourier transform of ε_{i+1} is given by: $f_\varepsilon^*(y) = \exp(-i\tilde{\varepsilon}y) \frac{2^{i\beta y}}{\sqrt{\pi}} \Gamma\left(\frac{1}{2} + i\beta y\right)$ with $\tilde{\varepsilon} = \beta\varepsilon$ (see ‘‘Appendix C.2’’).

For the three methods, we take a number of particles M equal to 5000. Note that for the Bayesian procedure (Bootstrap, APF and KSAPF), we need a prior on θ , and this only at the first step. The prior for θ_1 is taken to be the Uniform law and conditionally to θ_1 the distribution of X_1 is the stationary law:

$$\begin{cases} p(\theta_1) = \mathcal{U}(0.5, 0.9) \times \mathcal{U}(0.1, 0.4) \\ f_{\theta_1}(X_1) = \mathcal{N}\left(0, \frac{\sigma_1^2}{1-\phi_1^2}\right) \end{cases}$$

We take $h = 0.1$ for the KSAPF and $Q = \begin{pmatrix} 0.6 \cdot 10^{-6} & 0 \\ 0 & 0.1 \cdot 10^{-6} \end{pmatrix}$ for the APF and Bootstrap filter.

Remark 4 Note that, in practice, there is no constraint on the parameters for the contrast function contrary to the particle filters where we take the stationary law for $p_\theta(X_0)$ and the Uniform law around the true parameters. Hence, we bias favourably the particle filters.

2.5 Numerical results

In the numerical section we compare the different estimations: the QML estimator defined in Sect. 2.3.1, the Bayesian estimators defined in Sect. 2.3.2 and our contrast estimator defined in Sect. 1.1. For the comparison of the computing time, we also compare our contrast estimator with the SIEMLE proposed by Kim et al. [see “Appendix D.1” and Kim and Shephard (1994) for more general details].

2.5.1 Computing time

From a theoretical point of view, the MLE is asymptotically efficient. However, in practice since the states $(X_1 \dots, X_n)$ are unobservable and the SV model is non Gaussian, the likelihood is untractable. We have to use numerical methods to approximate it. In this section, we illustrate the SIEMLE which consists in approximating the likelihood and applying the Expectation-Maximisation algorithm introduced by Dempster (1977) to find the parameter θ .

To illustrate the SIEMLE for the SV model, we run an estimator with a number of observations n equal to 1000. Although the estimation is good the computing time is very long compared with the others methods (see Tables 1 and 2). This result illustrates the numerical complexity of the SIEMLE (see “Appendix D.1”). Therefore, in the following, we only compare our contrast estimator with the QML and Bayesian estimators. The results are illustrated by Fig. 1. We can see that our contrast estimator is the fastest for the Gaussian AR(1) model. The QML is the most rapid for the SV

Table 1 Comparison of the computing time (CPU in seconds) and the MSE with respect to the number of observations $n = 200$ up to 1,500 for the Gaussian AR(1) and the SV models

n	SV		AR(1)	
	CPU	MSE	CPU	MSE
Contrast				
200	4.2695	0.0425	0.032146	0.0411
300	5.1015	0.0453	0.022588	0.0398
400	7.0502	0.0239	0.028062	0.0374
500	6.9109	0.0175	0.026517	0.0306
750	11.8555	0.0117	0.031353	0.0218
1,000	20.4074	0.0078	0.056931	0.0133
1,500	29.3910	0.0061	0.08432	0.0091
Bootstrap filter				
200	41.4780	0.0275	85.65	0.0225
300	57.5201	0.0261	103.7212	0.0211
400	67.9421	0.0248	155.0456	0.0199
500	107.9450	0.0228	169.5578	0.0187
750	138.0307	0.0186	241.1891	0.0154
1,000	192.2166	0.0174	318.5656	0.0133

Table 1 continued

n	SV		AR(1)	
	CPU	MSE	CPU	MSE
1,500	158.3680	0.0166	388.7098	0.0122
APF				
200	19.4471	0.0209	49.6784	0.0138
300	39.2457	0.0182	69.3421	0.0125
400	46.9590	0.0123	86.9111	0.0118
500	54.5811	0.0189	108.9087	0.0112
750	91.5288	0.0171	166.3432	0.0100
1,000	105.1695	0.0163	189.5432	0.0087
1,500	122.1278	0.0159	326.7654	0.0074
KSAPF				
200	32.8328	0.0131	55.039200	0.0121
300	47.4919	0.0129	90.691115	0.0116
400	58.3216	0.0118	107.767974	0.110
500	66.3554	0.0114	127.565273	0.102
750	76.4818	0.0103	173.311428	0.0086
1,000	93.8846	0.0093	246.09729	0.0073
1,500	151.7971	0.0084	376.8976	0.0068
QML				
200	0.0268	0.172	0.0283	0.0444
300	0.0201	0.164	0.0312	0.0331
400	0.0532	0.153	0.0386	0.0336
500	0.0675	0.146	0.0476	0.0327
750	0.1046	0.132	0.0631	0.0311
1,000	0.0702	0.118	0.0712	0.0278
1,500	0.2148	0.110	0.0854	0.0253

The number of particles in Bayesian estimations is $M = 5,000$ particles and the number of estimators is $N = 100$ for the MSE [see Eq. (15)]

Table 2 SIEMLE estimation for the SV model

ϕ_0	σ_0^2	$\hat{\phi}_n$	$\hat{\sigma}_n^2$	CPU (s)
0.7	0.3	0.667	0.2892	74,300

The number of observations is $n = 1,000$ and the number of sweeps for the Gibbs sampler is $\tilde{M} = 100$ (see ‘‘Appendix D.1’’)

model since it assumes that the measurement errors are Gaussian but we show in Figs. 2, 3 and 4 that it is a biased estimator with large mean square error. For our algorithm, for the Gaussian AR(1) model, the function $u_{l_\theta}^*$ has an explicit expression but for the SV model, the function $u_{l_\theta}^*$ is approximated numerically since the Fourier transform of the function u_{l_θ} has not an explicit form. This explains why our algorithm is slower

on the SV model than on the Gaussian AR(1) model.² In spite of this approximation, our contrast estimator is fast and its implementation is straightforward.

2.5.2 Parameter estimates

For the AR(1) Gaussian model, we run $N = 1,000$ estimates for each method (QML, APF, KSAPF and Bootstrap filter) and $N = 500$ for the SV model. The number of observations n is equal to 1,000 for the two models.

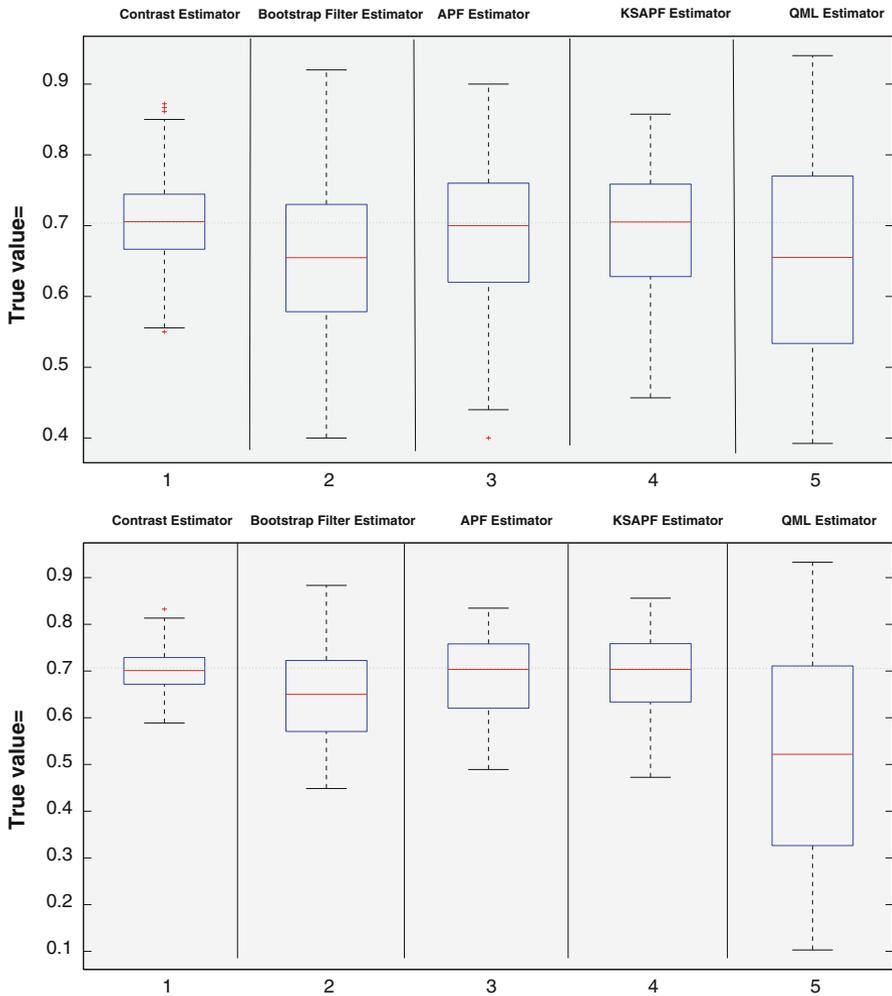


Fig. 2 Boxplot of ϕ . True value: $\phi_0 = 0.7$. *Top panel* Gaussian AR(1) model. *Bottom panel* SV model

² We use a quadrature method implemented in Matlab to approximate the Fourier transform of $u_{I_\theta}(y)$. One can also use the FFT method and we expect that the contrast estimator will be more rapid in this case.

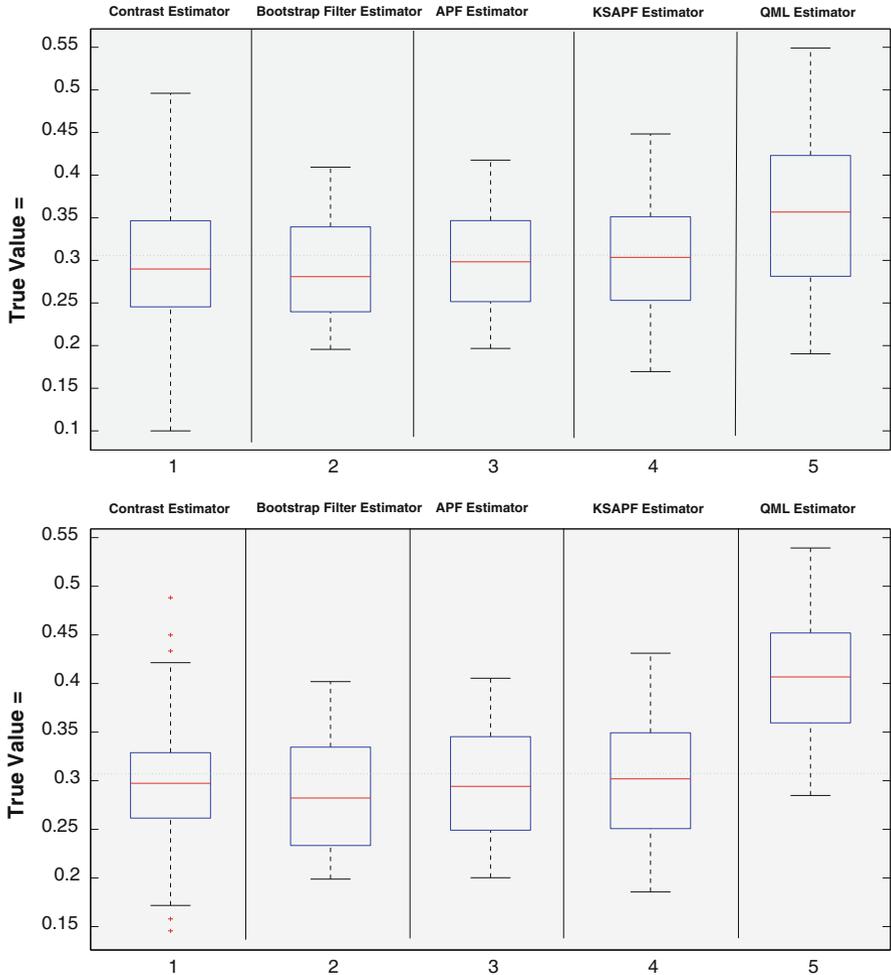


Fig. 3 Boxplot of σ^2 . True value: $\sigma_0^2 = 0.3$. *Left* Gaussian AR(1) model. *Right* SV model

In order to compare with others the performance of our estimator, we compute for each method the Mean Square Error (MSE) defined by:

$$MSE = \frac{1}{N} \left(\sum_{j=1}^N (\hat{\phi}_j - \phi_0)^2 + (\hat{\sigma}_j^2 - \sigma_0^2)^2 \right), \tag{15}$$

We illustrate by boxplots the different estimates (see Figs. 2, 3). We also illustrate in Fig. 4 the MSE for each estimator computed by Eq. (15). We can see that, for the parameter ϕ_0 , the QML estimator is better for the Gaussian AR(1) model than for the SV model (see Fig. 2). Indeed, the Gaussianity assumption is wrong for the SV model. Moreover, the estimate of σ_0^2 by QML is very bad for the two models (see Fig. 3) and its corresponding boxplots have the largest dispersion meaning that the

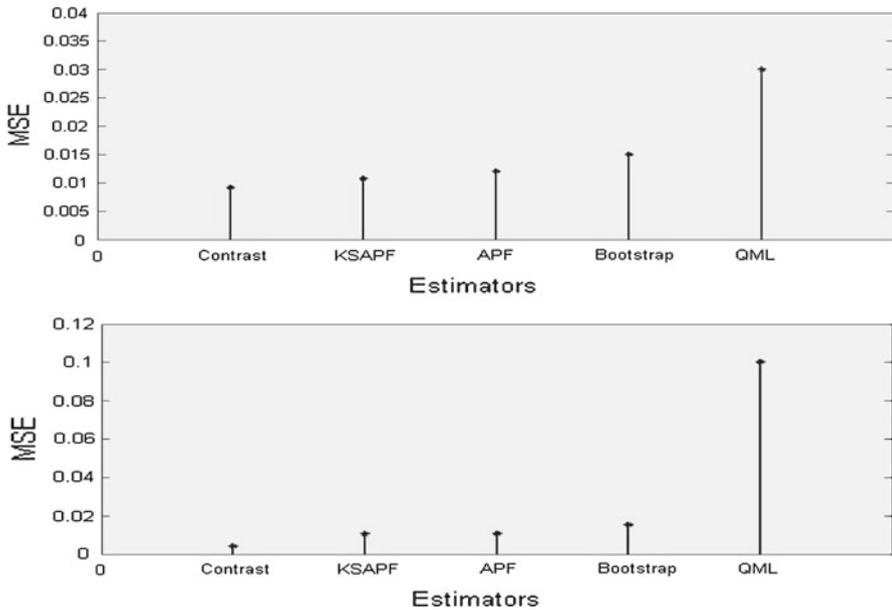


Fig. 4 MSE computed by Eq. (15). *Top panel* Gaussian AR(1) model. *Bottom panel* SV model

QML method is not very stable. The Bootstrap, APF and KSAPF have also a large dispersion of their boxplots, in particular for the parameter ϕ_0 (see Fig. 2). Besides, the Bootstrap filter is less efficient than the APF and KSAPF. For the Gaussian and SV model, the boxplots of our contrast estimator show that our estimator is the most stable with respect to ϕ_0 and we obtain similar results for σ_0^2 . The MSE is better for the SV model and the smallest for our contrast estimator.

2.5.3 Confidence interval of the contrast estimator

To illustrate the statistical properties of our contrast estimator, we compute for each model the confidence intervals computed with the confidence level $1 - \alpha$ equal to 0.95 for $N = 1$ estimator and the coverages for $N = 1,000$ with respect to the number of observations. The coverage corresponds to the number of times for which the true parameter $\theta_{0,i}$, $i = 1, 2$ belongs to the confidence interval. The results are illustrated by the Figs. 5, 6 and 7: for the Gaussian and SV models, the coverage converges to 95 % for a small number of observations. As expected, the confidence interval decreases with the number of observations. Note that of course a MLE confidence interval would be smaller since the MLE is efficient but the corresponding computing time would be huge.

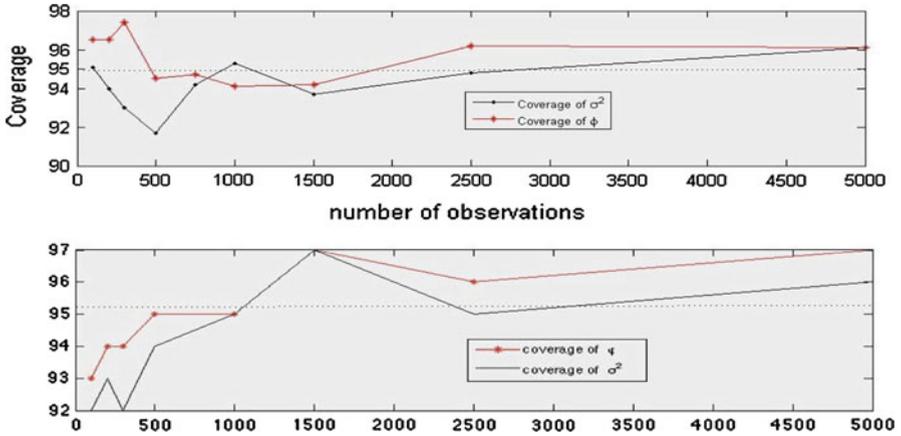


Fig. 5 Coverage with respect to the number of observations $n = 100$ up to 5,000 for $N = 1,000$ estimators. *Top panel* Gaussian AR(1) model. *Bottom panel* SV model (color figure online)

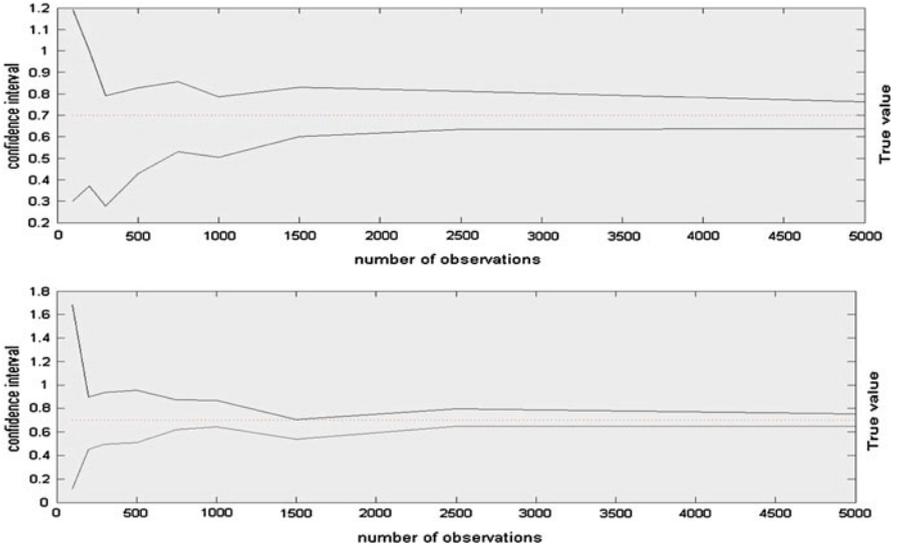


Fig. 6 Confidence interval for the parameter ϕ_0 with respect to the number of observations $n = 100$ up to 5,000 for $N = 1$ estimator. *Top panel* Gaussian AR(1) model. *Bottom panel* SV model

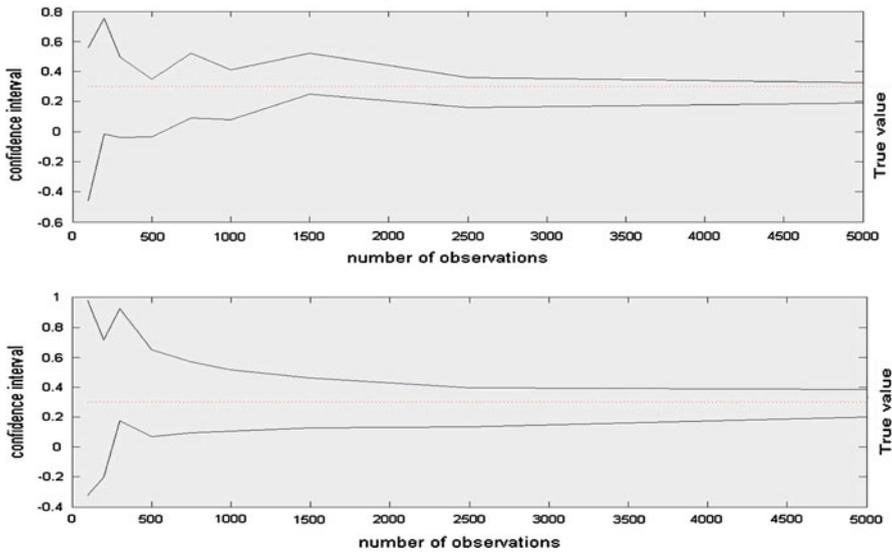


Fig. 7 Confidence interval for the parameter σ_0^2 with respect to the number of observations $n = 100$ up to 5,000 for $N = 1$ estimator. *Top panel* Gaussian AR(1) model. *Bottom panel* SV model

2.6 Application to real data

The data consist of daily observations on FTSE stock price index and S&P500 stock price index. The series taken in boursorama.com are closing prices from January, 3, 2004 to January, 2,2007 for the FTSE and S&P500 leaving a sample of 759 observations for the two series.

The daily prices S_i are transformed into compounded rates returns centered around their sample mean c for self-normalization (see [Mathieu and Schotman 1998](#); [Ghysels](#)

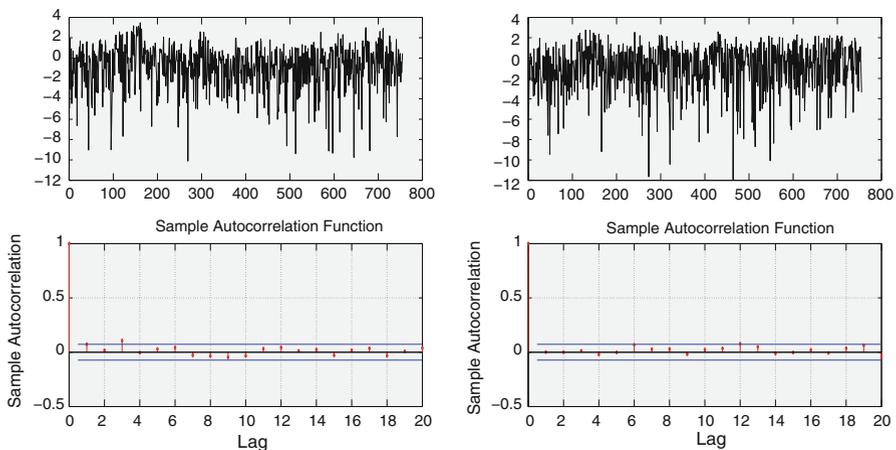


Fig. 8 *Top left panel* Graph of $Y_i = FTSE$. *Top right panel* Graph of $Y_i = SP500$. *Bottom left panel* Autocorrelation of $Y_i = FTSE$. *Bottom right panel* Autocorrelation of $Y_i = SP500$

et al. 1996) $R_i = 100 \times \log \left(\frac{S_i}{S_{i-1}} \right) - c$. We want to model those data by the SV model defined in (13) leading to:

$$\begin{aligned} Y_i &= \log(R_i^2) - \mathbb{E}[\log(\xi_i^2)] \\ &= \log(R_i^2) + 1.27 \end{aligned}$$

Those data are represented on Fig. [8].

2.6.1 Parameter estimates

In the empirical analysis, we compare the QML, the Bootstrap filter, the APF and the KSAPF estimators. The last one is our contrast estimator. The variance of the measurement noise is $\sigma_\varepsilon^2 = \frac{\pi^2}{2}$, that is β is equal to 1 (see Sect. 2.4). Table 3 summarises the parameter estimates and the computing time for the five methods. For initialization of the Bayesian procedure, we take the Uniform law for the parameters $p(\theta_1) = \mathcal{U}(0.4, 0.95) \times \mathcal{U}(0.1, 0.5)$ and the stationary law for the log-volatility process X_1 , i.e. $f_{\theta_1}(X_1) = \mathcal{N}\left(0, \frac{\sigma_1^2}{1-\phi_1^2}\right)$.

The estimates of ϕ are in full accordance with results reported in previous studies of SV models. This parameter is in general close to 1 which implies persistent logarithmic volatility data. We compute the corresponding confidence intervals at level 5 % (see Table 4). For the SP500 and the FTSE, note that the Bootstrap filter and the QML are not in the confidence interval for the two parameters ϕ and σ^2 . These results are consistent with the simulations where we showed that both methods were biased for the SV model (see Sect. 2.5.2). Note also that as expected the computing time for the QML is the shortest because it assumes Gaussianity which is probably not the case

Table 3 Parameter estimates: $n = 1,000$ and the number of particles $M = 5,000$ for the particle filters

Index	FTSE			SP500		
	$\hat{\phi}_n$	$\hat{\sigma}_n^2$	CPU	$\hat{\phi}_n$	$\hat{\sigma}_n^2$	CPU
Contrast	0.69	0.27	26	0.78	0.13	38
Bootstrap filter	0.91	0.15	204	0.830	0.247	214
APF	0.693	0.29	169	0.734	0.108	182
KSAPF	0.697	0.29	152	0.80	0.12	175
QML	0.649	0.08	0.07	0.895	0.257	0.1

The CPU is in seconds

Table 4 Confidence interval at level 5 %

Index	Confidence interval	
	ϕ	σ^2
FTSE	[0.6627; 0.7173]	[0.1771; 0.3629]
SP500	[0.7086; 0.8514]	[0.0278; 0.2322]

here. Except of QML, the contrast is the fastest method. The results are presented in Table 3 below.

2.7 Summary and conclusions

In this paper we propose a new method to estimate an hidden stochastic model on the form (1). This method is based on the deconvolution strategy and leads to a consistent and asymptotically normal estimator. We empirically study the performance of our estimator for the Gaussian AR(1) model and SV model and we are able to construct a confidence interval (see Figs. 6, 7). As the boxplots [2] and [3] show, only the Contrast, the APF, and the KSAPF estimators are comparable. Indeed the QML and the Bootstrap Filter estimators are biased and their MSE are bad, and in particular, the QML method is the worst estimator (see Fig. 4). One can see that the QML estimator proposed by Harvey et al. is not suitable for the SV model because the approximation of the log-chi-square density by the Gaussian density is not robust (see Fig. 1). Furthermore, if we compare the MSE of the three Sequential Bayesian estimation, the KSAPF estimator is the best method. From a Bayesian point of view, it is known that the Bootstrap filter is less efficient than the APF and KSAPF filter since by using the density transition as the importance density, the propagation step of the particles will be made without taking care the observations (see Doucet et al. 2001).

Among the three estimators (Contrast, APF, and KSAPF) which give good results our estimator outperforms the others in a MSE aspect (see Fig. 4). Moreover, as we already mentioned, in the combined state and parameters estimation the difficulties are the choice of Q , h and the prior law since the results depend on these choices. In the numerical section, we have used the stationary law for the variable X_1 and this choice yields good results but we expect that the behavior of the Bayesian estimation will be worse for another prior. The implementation of the contrast estimator is the easiest and it leads to confidence intervals with a larger variance than the SIEMLE but at a smaller computing cost, in particular for the AR(1) Gaussian model (see Table 1). Furthermore, the contrast estimator does not require an arbitrary choice of parameter in practice.

Acknowledgments I thank my co-director Patricia Reynaud-Bouret for her idea about this paper and for her help and generosity. I thank also my director Frédéric Patras for his supervisory throughout this paper. I thank him for his careful reading of the paper and his large comments. I would like to thank F. Comte, A. Samson, N. Chopin, M. Miniconi and F. Pelgrin for their suggestions and their interest about this framework.

Appendix A: M-estimator

Definition 1 Geometrical ergodic process

Denote by $Q^n(x, \cdot)$ the transition kernel at step n of a (discrete-time) stationary Markov chain $(X_n)_n$ which started at x at time 0. That is, $Q^n(x, F) = \mathbb{P}(X_n \in F | X_0 = x)$. Let π denote the stationary law of X_n and let f be any measurable function. We call mixing coefficients $(\beta_n)_n$ the coefficients defined by, for each n :

$$\beta_n = \int \left[\sup_{\|f\|_\infty \leq 1} |Q^n(x, f) - \pi(f)| \right] \pi(dx),$$

where $\pi(f) = \int f(y)\pi(dy)$. We say that a process is geometrically ergodic if the decreasing of the sequence of the mixing coefficients $(\beta_n)_n$ is geometrical, that is:

$$\exists 0 < \eta < 1, \text{ such that } \beta_n \leq \eta^n.$$

The following results are the main tools for the proof of Theorem 1.

Consider the following quantities:

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

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

Lemma 1 *Uniform Law of Large Numbers (ULLN) (see Newey and McFadden 1994 for the proof).*

Let (Y_i) be an ergodic stationary process and suppose that:

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

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

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

Proposition 1 (Proposition 7.8 p. 472 in Hayashi (2000). The proof is in Newey (1987) Theorem 4.1.5.)

Suppose that:

1. θ_0 is in the interior of Θ .
2. $h_\theta(y)$ is twice continuously differentiable in θ for any y .
3. The Hessian matrix of the application $\theta \mapsto \mathbf{P} h_\theta$ is non-singular.
4. $\sqrt{n} \mathbf{P}_n S_\theta \rightarrow \mathcal{N}(0, \Omega(\theta_0))$ in law as $n \rightarrow \infty$, with $\Omega(\theta_0)$ a positive definite matrix.
5. Local dominance on the Hessian: for some neighbourhood \mathcal{U} of θ_0 :

$$\mathbb{E} \left[\sup_{\theta \in \mathcal{U}} \left\| \nabla_\theta^2 h_\theta(Y_1) \right\| \right] < \infty,$$

so that, for any consistent estimator $\hat{\theta}$ of θ_0 we have: $\mathbf{P}_n H_{\hat{\theta}} \rightarrow \mathbb{E}[\nabla_\theta^2 h_\theta(Y_1)]$ in probability as $n \rightarrow \infty$.

Then, $\hat{\theta}$ is asymptotically normal with asymptotic covariance matrix given by:

$$\Sigma(\theta_0) = \mathbb{E}[\nabla_{\theta}^2 h_{\theta}(Y_1)]^{-1} \Omega(\theta_0) \mathbb{E}[\nabla_{\theta}^2 h_{\theta}(Y_1)]^{-1}$$

where the differential $\nabla_{\theta}^2 h_{\theta}(Y_1)$ is taken at point $\theta = \theta_0$.

Proposition 2 (The proof is in Jones 2004)

Let Y_i be an ergodic stationary Markov chain and let $g: \mathcal{Y} \rightarrow \mathbb{R}$ a borelian function. Suppose that Y_i is geometrically ergodic and $\mathbb{E}[|g(Y_1)|^{2+\delta}] < \infty$ for some $\delta > 0$. Then, when $n \rightarrow \infty$,

$$\sqrt{n}(\mathbf{P}_n g - \mathbf{P}g) \rightarrow \mathcal{N}(0, \sigma_g^2) \text{ in law,}$$

where $\sigma_g^2 := \text{Var} [(g(Y_1))] + 2 \sum_{j=1}^{\infty} \text{Cov} (g(Y_1), g(Y_j)) < \infty$.

Appendix B: Proofs of theorem 1

For the reader convenience we split the proof of Theorem 1 into three parts: in Sect. 2.7, we give the proof of the existence of our contrast estimator defined in (1.1). In Sect. 2.7, we prove the consistency, that is, the first part of Theorem 1. Then, we prove the asymptotic normality of our estimator in Sect. 2.7, that is, the second part of Theorem 1. The Sect. 2.7 is devoted to Corollary 1. Finally, in Sect. 2.7 we prove that Theorem 1 applies for the AR(1) and SV models.

B.1 Proof of the existence and measurability of the M-estimator

By assumption, the function $\theta \mapsto \|l_{\theta}\|_2^2$ is continuous. Moreover, l_{θ}^* and then $u_{l_{\theta}^*}^*(x) = \frac{1}{2\pi} \int e^{ixy} \frac{l_{\theta}^*(-y)}{f_{\theta}^*(y)} dy$ are continuous w.r.t θ . In particular, the function $m_{\theta}(\mathbf{y}_i) = \|l_{\theta}\|_2^2 - 2y_{i+1}u_{l_{\theta}^*}^*(y_i)$ is continuous w.r.t θ . Hence, the function $\mathbf{P}_n m_{\theta} = \frac{1}{n} \sum_{i=1}^n m_{\theta}(\mathbf{Y}_i)$ is continuous w.r.t θ belonging to the compact subset Θ . So, there exists $\tilde{\theta}$ belongs to Θ such that:

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

B.2 Proof of the consistency

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

$$\begin{aligned} |m_{\theta}(\mathbf{y}_i)| &= \left| \|l_{\theta}\|_2^2 - 2y_{i+1}u_{l_{\theta}^*}^*(y_i) \right| \\ &\leq \|l_{\theta}\|_2^2 + 2|y_{i+1}u_{l_{\theta}^*}^*(y_i)|. \end{aligned}$$

As $\|l_\theta\|_2^2$ is continuous on the compact subset Θ , $\sup_{\theta \in \Theta} \|l_\theta\|_2^2$ is finite. Therefore, $\mathbb{E}[\sup_{\theta \in \Theta} |m_\theta(\mathbf{Y}_i)|]$ is finite if $\mathbb{E}\left[\sup_{\theta \in \Theta} |Y_{i+1}u_{l_\theta}^*(Y_i)|\right]$ is finite. Lemma *ULLN 1* gives us the uniform convergence in probability of the contrast function: for any $\varepsilon > 0$:

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

Combining the uniform convergence with Theorem 2.1 p. 2121 chapter 36 in [Hansen and Horowitz \(1997\)](#) yields the weak (convergence in probability) consistency of the estimator. □

Remark 5 In most applications, we do not know the bounds for the true parameter. So the compactness assumption is sometimes restrictive, one can replace the compactness assumption by: θ_0 is an element of the interior of a convex parameter space $\Theta \subset \mathbb{R}^r$. Then, under our assumptions except the compactness, the estimator is also consistent. The proof is the same and the existence is proved by using convex optimization arguments. One can refer to [Hayashi \(2000\)](#) for this discussion.

B.3 Proof of the asymptotic normality

The proof is based on the following Lemma:

Lemma 2 *Suppose that the conditions of the consistency hold. Suppose further that:*

1. \mathbf{Y}_i geometrically ergodic.
2. (Moment condition): for some $\delta > 0$ and for each $j \in \{1, \dots, r\}$:

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

3. (Hessian Local condition): For some neighbourhood \mathcal{U} of θ_0 and for $j, k \in \{1, \dots, r\}$:

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

Then, $\widehat{\theta}_n$ defined in Eq. (9) is asymptotically normal with asymptotic covariance matrix given by:

$$\Sigma(\theta_0) = V_{\theta_0}^{-1} \Omega(\theta_0) V_{\theta_0}^{-1}$$

where V_{θ_0} is the Hessian of the application $\mathbf{P}m_\theta$ given in Eq. (7).

Proof The proof follows from Proposition 1 and Proposition 2 and by using the fact that by assumption we have $\mathbb{E}[\nabla_\theta^2 m_\theta(\mathbf{Y}_1)] = \nabla_\theta^2 \mathbb{E}[m_\theta(\mathbf{Y}_1)]$. □

It just remains to check that the conditions (2) and (3) of Lemma 2 hold under our assumptions **(T)**.

Moment condition: As the function l_θ is twice continuously differentiable w.r.t θ , for all $\mathbf{y}_i \in \mathbb{R}^2$, the application $m_\theta(\mathbf{y}_i) : \theta \in \Theta \mapsto m_\theta(\mathbf{y}_i) = \|l_\theta\|_2^2 - 2y_{i+1}u_{l_\theta}^*(y_i)$ is twice continuously differentiable for all $\theta \in \Theta$ and its first derivatives are given by:

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \nabla_\theta \|l_\theta\|_2^2 - 2y_{i+1}\nabla_\theta u_{l_\theta}^*(y_i).$$

By assumption, for each $j \in \{1, \dots, r\}$, $\frac{\partial l_\theta}{\partial \theta_j} \in \mathbb{L}_1(\mathbb{R})$, therefore one can apply the Lebesgue Derivation Theorem and Fubini's Theorem to obtain:

$$\nabla_\theta m_\theta(\mathbf{y}_i) = \left[\nabla_\theta \|l_\theta\|_2^2 - 2y_{i+1}u_{\nabla_\theta l_\theta}^*(y_i) \right]. \tag{16}$$

Then, for some $\delta > 0$:

$$\begin{aligned} |\nabla_\theta m_\theta(\mathbf{y}_i)|^{2+\delta} &= \left| \nabla_\theta \|l_\theta\|_2^2 - 2y_{i+1}u_{\nabla_\theta l_\theta}^*(y_i) \right|^{2+\delta} \\ &\leq C_1 \left| \nabla_\theta \|l_\theta\|_2^2 \right|^{2+\delta} + C_2 |y_{i+1}u_{\nabla_\theta l_\theta}^*(y_i)|^{2+\delta}, \end{aligned} \tag{17}$$

where C_1 and C_2 are two positive constants. By assumption, the function $\|l_\theta\|_2^2$ is twice continuously differentiable w.r.t θ . Hence, $\nabla_\theta \|l_\theta\|_2^2$ is continuous on the compact subset Θ and the first term of Eq. (17) is finite. The second term is finite by the moment assumption **(T)**.

Hessian Local dominance: For $j, k \in \{1, \dots, r\}$, $\frac{\partial^2 l_\theta}{\partial \theta_j \partial \theta_k} \in \mathbb{L}_1(\mathbb{R})$, the Lebesgue Derivation Theorem gives:

$$\nabla_\theta^2 m_\theta(\mathbf{y}_i) = \nabla_\theta^2 \|l_\theta\|_2^2 - 2y_{i+1}u_{\nabla_\theta^2 l_\theta}^*(y_i),$$

and, for some neighbourhood \mathcal{U} of θ_0 :

$$\mathbb{E} \left[\sup_{\theta \in \mathcal{U}} \left\| \nabla_\theta^2 m_\theta(\mathbf{Y}_i) \right\| \right] \leq \sup_{\theta \in \mathcal{U}} \left\| \nabla_\theta^2 \|l_\theta\|_2^2 \right\| + 2\mathbb{E} \left[\sup_{\theta \in \mathcal{U}} \left\| Y_{i+1}u_{\nabla_\theta^2 l_\theta}^*(Y_i) \right\| \right].$$

The first term of the above equation is finite by continuity and by compactness argument. And, the second term is finite by the Hessian local dominance assumption **(T)**. □

B.4 Proof of corollary 1

By replacing $\nabla_{\theta} m_{\theta}(\mathbf{Y}_1)$ by its expression (16), we have:

$$\begin{aligned} \Omega_0(\theta) &= \text{Var} \left[\nabla_{\theta} \|l_{\theta}\|_2^2 - 2Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right] \\ &= 4 \text{Var} \left[Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right] \\ &= 4 \left[\mathbb{E} \left[Y_2^2 \left(u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right) \left(u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right)' \right] - \mathbb{E} \left[Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right] \mathbb{E} \left[Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right]' \right]. \end{aligned}$$

Furthermore, by Eq. (1) and by independence of the centered noise (ε_2) and (η_2) , we have:

$$\mathbb{E} \left[Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right] = \mathbb{E} \left[b_{\phi_0}(X_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right].$$

Using Fubini’s Theorem and Eq. (1) we obtain:

$$\begin{aligned} \mathbb{E} \left[b_{\phi_0}(X_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right] &= \mathbb{E} \left[b_{\phi_0}(X_1) \int e^{iY_1 z} u_{\nabla_{\theta} l_{\theta}}(z) dz \right] \\ &= \mathbb{E} \left[b_{\phi_0}(X_1) \int \frac{1}{2\pi} \frac{1}{f_{\varepsilon}^*(z)} e^{iY_1 z} (\nabla_{\theta} l_{\theta})^*(-z) dz \right] \\ &= \frac{1}{2\pi} \int \mathbb{E} \left[b_{\phi_0}(X_1) e^{i(X_1 + \varepsilon_1)z} \right] \frac{1}{f_{\varepsilon}^*(z)} (\nabla_{\theta} l_{\theta})^*(-z) dz \\ &= \frac{1}{2\pi} \int \frac{\mathbb{E} \left[e^{i\varepsilon_1 z} \right]}{f_{\varepsilon}^*(z)} \mathbb{E} \left[b_{\phi_0}(X_1) e^{iX_1 z} \right] (\nabla_{\theta} l_{\theta})^*(-z) dz \\ &= \frac{1}{2\pi} \mathbb{E} \left[b_{\phi_0}(X_1) \int e^{iX_1 z} (\nabla_{\theta} l_{\theta})^*(-z) dz \right] \\ &= \frac{1}{2\pi} \mathbb{E} \left[b_{\phi_0}(X_1) \left((\nabla_{\theta} l_{\theta})^*(-X_1) \right)^* \right] \\ &= \mathbb{E} \left[b_{\phi_0}(X_1) \nabla_{\theta} l_{\theta}(X_1) \right]. \end{aligned} \tag{18}$$

Hence,

$$\Omega_0(\theta) = 4 (P_2 - P_1),$$

where

$$\begin{aligned} P_1 &= \mathbb{E} \left[b_{\phi_0}(X_1) \nabla_{\theta} l_{\theta}(X_1) \right] \mathbb{E} \left[b_{\phi_0}(X_1) \nabla_{\theta} l_{\theta}(X_1) \right]', \\ P_2 &= \mathbb{E} \left[Y_2^2 \left(u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right) \left(u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right)' \right]. \end{aligned}$$

Calculus of the covariance matrix of Corollary (1): By replacing $(\nabla_{\theta} m_{\theta}(Y_1))$ by its expression (16) we have:

$$\begin{aligned} \Omega_{j-1}(\theta) &= \text{Cov} \left(\nabla_{\theta} \|l_{\theta}\|_2^2 - 2Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1), \nabla_{\theta} \|l_{\theta}\|_2^2 - 2Y_{j+1} u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right), \\ &= 4 \text{Cov} \left(Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1), Y_{j+1} u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right), \\ &= 4 \left[\mathbb{E} \left(Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) Y_{j+1} u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right) - \mathbb{E} \left(Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right) \mathbb{E} \left(Y_{j+1} u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right)' \right]. \end{aligned}$$

By using Eq. (18) and the stationary property of the Y_i , one can replace the second term of the above equation by:

$$\mathbb{E} \left[b_{\phi_0}(X_1) \nabla_{\theta} l_{\theta}(X_1) \right] \mathbb{E} \left[b_{\phi_0}(X_1) \nabla_{\theta} l_{\theta}(X_1) \right]'$$

Furthermore, by using Eq. (1) we obtain:

$$\begin{aligned} \mathbb{E} \left[Y_2 Y_{j+1} u_{\nabla_{\theta} l_{\theta}}^*(Y_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right] &= \mathbb{E} \left[b_{\phi_0}(X_1) b_{\phi_0}(X_j) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right] \\ &\quad + \mathbb{E} \left[b_{\phi_0}(X_1) (\eta_{j+1} + \varepsilon_{j+1}) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right] \end{aligned} \tag{19}$$

$$+ \mathbb{E} \left[b_{\phi_0}(X_j) (\eta_2 + \varepsilon_2) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right] \tag{20}$$

$$+ \mathbb{E} \left[(\eta_2 + \varepsilon_2) (\eta_{j+1} + \varepsilon_{j+1}) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right]. \tag{21}$$

By independence of the centered noise, the term (19), (20) and (21) are equal to zero. Now, if we use Fubini’s Theorem we have:

$$\mathbb{E} \left[b_{\phi_0}(X_1) b_{\phi_0}(X_j) u_{\nabla_{\theta} l_{\theta}}^*(Y_1) u_{\nabla_{\theta} l_{\theta}}^*(Y_j) \right] = \mathbb{E} \left[b_{\phi_0}(X_1) b_{\phi_0}(X_j) \nabla_{\theta} l_{\theta}(X_1) \nabla_{\theta} l_{\theta}(X_j) \right]. \tag{22}$$

Hence, the covariance matrix is given by:

$$\begin{aligned} \Omega_{j-1}(\theta) &= 4 \left(\mathbb{E} \left[b_{\phi_0}(X_1) b_{\phi_0}(X_j) (\nabla_{\theta} l_{\theta}(X_1)) (\nabla_{\theta} l_{\theta}(X_j))' \right] \right. \\ &\quad \left. - \mathbb{E} \left[b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1)) \right] \mathbb{E} \left[b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1)) \right]' \right) \\ &= 4 \left(\tilde{C}_{j-1} - \mathbb{E} \left[b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1)) \right] \mathbb{E} \left[b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1)) \right]' \right) \\ &= 4 \left(\tilde{C}_{j-1} - P_1 \right). \end{aligned}$$

Finally, we obtain: $\Omega(\theta) = \Omega_0(\theta) + 2 \sum_{j>1}^{\infty} \Omega_{j-1}(\theta)$ with $\Omega_0(\theta) = 4(P_2 - P_1)$ and $\Omega_{j-1}(\theta) = 4(\tilde{C}_{j-1} - P_1)$.

Expression of the Hessian matrix V_{θ} : We have:

$$Pm_{\theta} = \|l_{\theta}\|_2^2 - 2 \langle l_{\theta}, l_{\theta_0} \rangle. \tag{23}$$

For all θ in Θ , the application $\theta \mapsto \mathbf{P}m_\theta$ is twice differentiable w.r.t θ on the compact subset Θ . And for $j \in \{1, \dots, r\}$:

$$\begin{aligned} \frac{\partial \mathbf{P}m}{\partial \theta_j}(\theta) &= 2 \left\langle \frac{\partial l_\theta}{\partial \theta_j}, l_\theta \right\rangle - 2 \left\langle \frac{\partial l_\theta}{\partial \theta_j}, l_{\theta_0} \right\rangle \\ &= 2 \left\langle \frac{\partial l_\theta}{\partial \theta_j}, l_\theta - l_{\theta_0} \right\rangle, \\ &= 0 \text{ at the point } \theta_0, \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2 \mathbf{P}m}{\partial \theta_j \partial \theta_k}(\theta) &= 2 \left(\left\langle \frac{\partial^2 l_\theta}{\partial \theta_j \partial \theta_k}, l_\theta - l_{\theta_0} \right\rangle + \left\langle \frac{\partial l_\theta}{\partial \theta_k}, \frac{\partial l_\theta}{\partial \theta_j} \right\rangle \right)_{j,k} \\ &= 2 \left(\left\langle \frac{\partial l_\theta}{\partial \theta_k}, \frac{\partial l_\theta}{\partial \theta_j} \right\rangle \right)_{j,k} \text{ at the point } \theta_0. \end{aligned}$$

Appendix C: Proof of the applications

C.1 The Gaussian AR(1) model with measurement noise

C.1.1 Contrast function

We have:

$$l_\theta(x) = \frac{1}{\sqrt{2\pi\gamma^2}} \phi x \exp\left(-\frac{1}{2\gamma^2}x^2\right).$$

So that:

$$\|l_\theta\|_2^2 = \int |l_\theta(x)|^2 dx = \frac{\phi^2\gamma}{4\sqrt{\pi}},$$

and the Fourier Transform of l_θ is given by:

$$\begin{aligned} l_\theta^*(y) &= \int e^{iyx} l_\theta(x) dx = \int e^{iyx} \frac{1}{\sqrt{2\pi\gamma^2}} \phi x \exp\left(-\frac{1}{2\gamma^2}x^2\right) dx \\ &= -i\phi \mathbb{E}\left[iG e^{iyG}\right] = -i\phi \frac{\partial}{\partial y} \mathbb{E}\left[e^{iyG}\right] \quad \text{where } G \sim \mathcal{N}(0, \gamma^2) \\ &= -i\phi \frac{\partial}{\partial y} \left[e^{-\frac{y^2}{2}\gamma^2} \right] \\ &= i\phi y \gamma^2 e^{-\frac{y^2}{2}\gamma^2}. \end{aligned}$$

As ε_i is a centered Gaussian noise with variance σ_ε^2 , we have:

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

Define:

$$u_{l_\theta}(y) = \frac{1}{2\pi} \frac{l_\theta^*(-y)}{f_\varepsilon^*(y)}.$$

Then:

$$\begin{aligned} u_{l_\theta}^*(y) &= \frac{1}{2\pi} \int \frac{l_\theta^*(-x)}{f_\varepsilon^*(x)} e^{iyx} dx = \frac{-i}{2\pi} \phi \gamma^2 \int x e^{iyx} \exp\left(\frac{x^2}{2}\sigma_\varepsilon^2\right) \exp\left(-\frac{x^2}{2}\gamma^2\right) dx \\ &= \frac{-i}{2\pi} \phi \gamma^2 \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)^{1/2}} \int x e^{iyx} (\gamma^2 - \sigma_\varepsilon^2)^{1/2} \exp\left(-\frac{1}{2}x^2(\gamma^2 - \sigma_\varepsilon^2)\right) dx \\ &= -\frac{1}{\sqrt{2\pi}} \phi \gamma^2 \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)^{1/2}} \mathbb{E} \left[i G e^{iyG} \right] = -\frac{1}{\sqrt{2\pi}} \phi \gamma^2 \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)^{1/2}} \frac{\partial}{\partial y} \mathbb{E} \left[e^{iyG} \right] \\ &= -\frac{1}{\sqrt{2\pi}} \phi \gamma^2 \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)^{1/2}} \frac{\partial}{\partial y} \left[e^{-\frac{y^2}{2(\gamma^2 - \sigma_\varepsilon^2)}} \right] \\ &= \frac{1}{\sqrt{2\pi}} \phi \gamma^2 \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)^{3/2}} y e^{-\frac{y^2}{2(\gamma^2 - \sigma_\varepsilon^2)}}, \end{aligned}$$

where $G \sim \mathcal{N}\left(0, \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)}\right)$. We deduce that the function $m_\theta(\mathbf{y}_i)$ is given by:

$$\begin{aligned} m_\theta(\mathbf{y}_i) &= \|l_\theta\|_2^2 - 2y_{i+1} u_{l_\theta}^*(y_i) \\ &= \frac{\phi^2 \gamma}{4\sqrt{\pi}} - 2y_i y_{i+1} \frac{1}{\sqrt{2\pi}} \phi \gamma^2 \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)^{3/2}} \exp\left(-\frac{y_i^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right). \end{aligned}$$

Then, the contrast estimator defined in (1.1) is given by:

$$\begin{aligned} \hat{\theta}_n &= \arg \min_{\theta \in \Theta} \mathbf{P}_n m_\theta \\ &= \arg \min_{\theta \in \Theta} \left\{ \frac{\phi^2 \gamma}{4\sqrt{\pi}} - \sqrt{\frac{2}{\pi}} \frac{\phi \gamma^2}{n(\gamma^2 - \sigma_\varepsilon^2)^{3/2}} \sum_{j=1}^n Y_{j+1} Y_j \exp\left(-\frac{1}{2} \frac{Y_j^2}{(\gamma^2 - \sigma_\varepsilon^2)}\right) \right\}. \quad \square \end{aligned}$$

C.1.2 Checking assumptions of Theorem 1

Mixing properties. If $|\phi| < 1$, the process \mathbf{Y}_i is geometrically ergodic. For further details, we refer to [Dedecker et al. \(2007\)](#).

Regularity conditions: It remains to prove that the assumptions of Theorem 1 hold. It is easy to see that the only difficulty is to check the moment condition and the local

dominance **(C)-(T)** and the uniqueness assumption **(CT)**. The others assumptions are easily to verify since the function $l_\theta(x)$ is regular in θ belonging to Θ .

(CT): The limit contrast function $\mathbf{P}m_\theta : \theta \in \Theta \mapsto \mathbf{P}m_\theta$ given by:

$$\begin{aligned} \theta \mapsto \mathbf{P}m_\theta &= \|l_\theta\|_2^2 - 2\langle l_\theta, l_{\theta_0} \rangle \\ &= \frac{\phi^2\gamma}{4\sqrt{\pi}} - \sqrt{\frac{2}{\pi}} \frac{\phi\phi_0\gamma^2\gamma_0^2}{(\gamma^2 + \gamma_0^2)^{\frac{3}{2}}}, \end{aligned}$$

is differentiable for all θ in Θ and $\nabla_\theta \mathbf{P}m_\theta = 0_{\mathbb{R}^2}$ if and only if θ is equal to θ_0 . More precisely its first derivatives are given by:

$$\begin{aligned} \frac{\partial \mathbf{P}m_\theta}{\partial \phi} &= \frac{1}{4\sqrt{\pi}} \frac{\phi\gamma(2 - \phi^2)}{(1 - \phi^2)} - \sqrt{\frac{2}{\pi}} \phi_0\gamma_0^2(\gamma^2 + \gamma_0^2)^{-3/2} \\ &\quad \left(\frac{\gamma^2 + \gamma^2\phi^2}{(1 - \phi^2)} - \frac{3\phi^2\gamma^4}{(1 - \phi^2)(\gamma^2 + \gamma_0^2)} \right), \\ \frac{\partial \mathbf{P}m_\theta}{\partial \sigma^2} &= \frac{\phi^2}{8\sqrt{\pi}\sigma(1 - \phi^2)^{1/2}} - \sqrt{\frac{2}{\pi}} \frac{\phi_0\gamma_0^2}{(1 - \phi^2)(\gamma^2 + \gamma_0^2)^{3/2}} \left(\phi - \frac{3\phi\gamma^2}{(\gamma^2 + \gamma_0^2)} \right), \end{aligned}$$

and

$$\nabla_\theta \mathbf{P}m_\theta = 0_{\mathbb{R}^2} \Leftrightarrow \theta = \theta_0$$

The partial derivatives of l_θ w.r.t θ are given by:

$$\begin{aligned} \frac{\partial l_\theta}{\partial \phi}(x) &= \left(\left(\frac{-\phi^2}{1 - \phi^2} + 1 \right) x + \frac{\phi^2}{(1 - \phi^2)\gamma^2} x^3 \right) \frac{1}{\sqrt{2\pi\gamma^2}} e^{-\frac{x^2}{2\gamma^2}}, \\ \frac{\partial l_\theta}{\partial \sigma^2}(x) &= \left(-\frac{\phi}{2(1 - \phi^2)\gamma^2} x + \frac{\phi}{2(1 - \phi^2)\gamma^4} x^3 \right) \frac{1}{\sqrt{2\pi\gamma^2}} e^{-\frac{x^2}{2\gamma^2}}. \end{aligned}$$

For the reader convenience let us introduce the following notations:

$$a_1 = \frac{-\phi^2}{(1 - \phi^2)} + 1 = \frac{1 - 2\phi^2}{(1 - \phi^2)} \quad \text{and} \quad a_2 = \frac{\phi^2}{(1 - \phi^2)\gamma^2}, \tag{24}$$

$$b_1 = \frac{-\phi}{2(1 - \phi^2)\gamma^2} \quad \text{and} \quad b_2 = \frac{\phi}{2(1 - \phi^2)\gamma^4}. \tag{25}$$

We rewrite:

$$\begin{aligned} \nabla_\theta l_\theta(x) &= \left(\frac{\partial l_\theta}{\partial \phi}(x), \frac{\partial l_\theta}{\partial \sigma^2}(x) \right)' \\ &= \left((a_1x + a_2x^3) \times g_{0,\gamma^2}(x), (b_1x + b_2x^3) \times g_{0,\gamma^2}(x) \right)', \end{aligned}$$

where the function g_{0,γ^2} defines the normal probability density of a centered random variable with variance γ^2 . Now, we can use Corollary 1 to compute the Hessian matrix V_{θ_0} :

$$\begin{aligned}
 V_{\theta_0} &= 2 \left(\begin{array}{c} \left\| \frac{\partial l_{\theta}}{\partial \phi} \right\|_2^2 \\ \frac{\partial l_{\theta}}{\partial \sigma^2}, \frac{\partial l_{\theta}}{\partial \phi} \end{array} \left\| \frac{\partial l_{\theta}}{\partial \phi}, \frac{\partial l_{\theta}}{\partial \sigma^2} \right\|_2^2 \right) \\
 &= \frac{1}{\gamma_0 \sqrt{\pi}} \\
 &\quad \left(\begin{array}{cc} a_1^2 \mathbb{E}[X^2] + 2a_1 a_2 \mathbb{E}[X^4] + a_2^2 \mathbb{E}[X^6] & a_1 b_1 \mathbb{E}[X^2] + a_1 b_2 \mathbb{E}[X^4] + a_2 b_1 \mathbb{E}[X^4] + a_2 b_2 \mathbb{E}[X^6] \\ a_1 b_1 \mathbb{E}[X^2] + a_1 b_2 \mathbb{E}[X^4] + a_2 b_1 \mathbb{E}[X^4] + a_2 b_2 \mathbb{E}[X^6] & b_1^2 \mathbb{E}[X^2] + 2b_1 b_2 \mathbb{E}[X^4] + b_2^2 \mathbb{E}[X^6] \end{array} \right), \tag{26}
 \end{aligned}$$

with $X \sim \mathcal{N}(0, \frac{\gamma_0^2}{2})$. By replacing the terms a_1, a_2, b_1 and b_2 at the point θ_0 we obtain:

$$V_{\theta_0} = \frac{1}{8\sqrt{\pi}(1 - \phi_0^2)^2} \begin{pmatrix} \gamma_0(7\phi_0^4 - 4\phi_0^2 + 4) & \frac{-5\phi_0^5 + 3\phi_0^3 + 2\phi_0}{2\gamma_0(1 - \phi_0^2)} \\ \frac{-5\phi_0^5 + 3\phi_0^3 + 2\phi_0}{2\gamma_0(1 - \phi_0^2)} & \frac{7\phi_0^2}{4\gamma_0^3} \end{pmatrix}, \tag{27}$$

which has a positive determinant equal to 0.0956 at the true value $\theta_0 = (0.7, 0.3)$. Hence, V_{θ_0} is non-singular. Furthermore, the strict convexity of the function $\mathbf{P}m_{\theta}$ gives that θ_0 is a minimum.

(C): (Local dominance): We have:

$$\mathbb{E} \left[\sup_{\theta \in \Theta} |Y_2 u_{l_{\theta}}^*(Y_1)| \right] = \frac{1}{\sqrt{2\pi}} \mathbb{E} \left[\sup_{\theta \in \Theta} \left| \frac{\phi \gamma^2}{(\gamma^2 - \sigma_{\epsilon}^2)^{(3/2)}} Y_2 Y_1 \exp \left(-\frac{Y_1^2}{2(\gamma^2 - \sigma_{\epsilon}^2)} \right) \right| \right].$$

The multivariate normal density of the pair $\mathbf{Y}_1 = (Y_1, Y_2)$ denoted $g_{(0, \mathcal{J}_{\theta_0})}$ is given by:

$$\frac{1}{2\pi} \det(\mathcal{J}_{\theta_0})^{-1/2} \exp \left(-\frac{1}{2} \mathbf{y}'_1 \mathcal{J}_{\theta_0}^{-1} \mathbf{y}_1 \right),$$

with:

$$\mathcal{J}_{\theta_0} = \begin{pmatrix} \sigma_{\epsilon}^2 + \gamma_0^2 & \phi_0 \gamma_0^2 \\ \phi_0 \gamma_0^2 & \sigma_{\epsilon}^2 + \gamma_0^2 \end{pmatrix} \quad \text{and} \quad \mathcal{J}_{\theta_0}^{-1} = \frac{1}{(\sigma_{\epsilon}^2 + \gamma_0^2)^2 - \phi_0^2 \gamma_0^4} \begin{pmatrix} \sigma_{\epsilon}^2 + \gamma_0^2 & -\phi_0 \gamma_0^2 \\ -\phi_0 \gamma_0^2 & \sigma_{\epsilon}^2 + \gamma_0^2 \end{pmatrix}.$$

By definition of the parameter space Θ and as all moments of the pair \mathbf{Y}_1 exist, the quantity $\mathbb{E} \left[\sup_{\theta \in \Theta} |Y_2 u_{l_{\theta}}^*(Y_1)| \right]$ is finite.

Moment condition (T): We recall that:

$$\begin{aligned} \nabla_{\theta} l_{\theta}(x) &= \left(\frac{\partial l_{\theta}}{\partial \phi}(x), \frac{\partial l_{\theta}}{\partial \sigma^2}(x) \right)' \\ &= \left((a_1x + a_2x^3) \times g_{0,\gamma^2}(x), (b_1x + b_2x^3) \times g_{0,\gamma^2}(x) \right)' . \end{aligned}$$

The Fourier transforms of the first derivatives are:

$$\begin{aligned} \left(\frac{\partial l_{\theta}}{\partial \phi}(x) \right)^* &= \int \exp(ixy) (a_1y + a_2y^3) \times g_{0,\gamma^2}(y) dy \\ &= -ia_1 \mathbb{E} [iG \exp(ixG)] + ia_2 \mathbb{E} [-iG^3 \exp(ixG)] \text{ where } G \sim \mathcal{N}(0, \gamma^2) \\ &= -ia_1 \frac{\partial}{\partial x} \mathbb{E} [\exp(ixG)] + ia_2 \frac{\partial^3}{\partial x^3} \mathbb{E} [\exp(ixG)] \\ &= -ia_1 \frac{\partial}{\partial x} \exp\left(-\frac{x^2}{2}\gamma^2\right) + ia_2 \frac{\partial^3}{\partial x^3} \exp\left(-\frac{x^2}{2}\gamma^2\right) \\ &= (ia_1\gamma^2x + 3ia_2\gamma^4x - ia_2\gamma^6x^3) \exp\left(-\frac{x^2}{2}\gamma^2\right), \end{aligned}$$

and

$$\left(\frac{\partial l_{\theta}}{\partial \sigma^2}(x) \right)^* = (ib_1\gamma^2x + 3ib_2\gamma^4x - ib_2\gamma^6x^3) \exp\left(-\frac{x^2}{2}\gamma^2\right).$$

We can compute the function $u_{\nabla_{\theta} l_{\theta}}(x)$:

$$\begin{aligned} u_{\frac{\partial l_{\theta}}{\partial \phi}}(x) &= \frac{1}{2\pi} \frac{\left(\frac{\partial l_{\theta}}{\partial \phi}(-x) \right)^*}{f_{\varepsilon}^*(x)} \\ &= \frac{1}{\sqrt{2\pi}} (\gamma^2 - \sigma_{\varepsilon}^2)^{1/2} \exp\left(-\frac{x^2}{2}(\gamma^2 - \sigma_{\varepsilon}^2)\right) \\ &\quad \times \left\{ \frac{1}{\sqrt{2\pi}} \frac{1}{(\gamma^2 - \sigma_{\varepsilon}^2)^{1/2}} \left((-ia_1\gamma^2 - 3ia_2\gamma^4)x + ia_2\gamma^6x^3 \right) \right\} \\ &= -i\bar{C} \left(A_1x - A_2x^3 \right) g_{0, \frac{1}{(\gamma^2 - \sigma_{\varepsilon}^2)}}(x), \end{aligned}$$

with $\bar{C} = \frac{1}{\sqrt{2\pi}} \frac{1}{(\gamma^2 - \sigma_{\varepsilon}^2)^{1/2}}$ and $A_1 = a_1\gamma^2 + 3a_2\gamma^4 = \gamma^2 \frac{(1+\phi^2)}{(1-\phi^2)}$ and $A_2 = a_2\gamma^6 = \gamma^4 \frac{\phi^2}{(1-\phi^2)}$. The Fourier transform of the function $u_{\frac{\partial l_{\theta}}{\partial \phi}}(x)$ is given by:

$$\begin{aligned} u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(x) &= -i\bar{C} \int \exp(iyx) (A_1y - A_2y^3) g_{\left(0, \frac{1}{(\gamma^2 - \sigma_{\varepsilon}^2)}\right)}(y) dy \\ &= -\bar{C}A_1 \frac{\partial}{\partial x} \mathbb{E} [\exp(ixG)] - \bar{C}A_2 \frac{\partial^3}{\partial x^3} \mathbb{E} [\exp(ixG)] \end{aligned}$$

$$\begin{aligned}
 &\text{where } G \sim \mathcal{N}\left(0, \frac{1}{(\gamma^2 - \sigma_\varepsilon^2)}\right) \\
 &= -\bar{C}A_1 \frac{\partial}{\partial x} \left(\exp\left(-\frac{x^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right) \right) - \bar{C}A_2 \frac{\partial^3}{\partial x^3} \left(\exp\left(-\frac{x^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right) \right) \\
 &= \left(\Psi_1^{\phi_0} x + \Psi_2^{\phi_0} x^3\right) \exp\left(-\frac{x^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right), \tag{28}
 \end{aligned}$$

with $\Psi_1^{\phi_0} = \bar{C} \left(\frac{A_1}{(\gamma^2 - \sigma_\varepsilon^2)} - \frac{3A_2}{(\gamma^2 - \sigma_\varepsilon^2)^2} \right)$ and $\Psi_2^{\phi_0} = \bar{C} \left(\frac{A_2}{(\gamma^2 - \sigma_\varepsilon^2)^3} \right)$. By the same arguments, we obtain:

$$u_{\frac{\partial l_\theta}{\partial \sigma^2}}^*(x) = \left(\Psi_1^{\sigma_0^2} x + \Psi_2^{\sigma_0^2} x^3\right) \exp\left(-\frac{x^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right), \tag{29}$$

with $\Psi_1^{\sigma_0^2} = \bar{C} \left(\frac{B_1}{(\gamma^2 - \sigma_\varepsilon^2)} - \frac{3B_2}{(\gamma^2 - \sigma_\varepsilon^2)^2} \right)$, $\Psi_2^{\sigma_0^2} = \bar{C} \left(\frac{B_2}{(\gamma^2 - \sigma_\varepsilon^2)^3} \right)$, $B_1 = b_1\gamma^2 + 3b_2\gamma^4 = \frac{\phi}{(1-\phi^2)}$ and $B_2 = b_2\gamma^6 = \gamma^2 \frac{\phi}{2(1-\phi^2)}$.

Hence, for some $\delta > 0$, $\mathbb{E} \left[\left| Y_2 u_{\frac{\partial l_\theta}{\partial \sigma^2}}^*(Y_1) \right|^{2+\delta} \right]$ is finite if:

$$\begin{aligned}
 &\mathbb{E} \left[\left| \left(\Psi_1^{\phi_0} Y_1 Y_2 + \Psi_2^{\phi_0} Y_1^3 Y_2 \right) \exp\left(-\frac{Y_1^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right) \right|^{2+\delta} \right] < \infty, \\
 &\mathbb{E} \left[\left| \left(\Psi_1^{\sigma_0^2} Y_1 Y_2 + \Psi_2^{\sigma_0^2} Y_1^3 Y_2 \right) \exp\left(-\frac{Y_1^2}{2(\gamma^2 - \sigma_\varepsilon^2)}\right) \right|^{2+\delta} \right] < \infty,
 \end{aligned}$$

which is satisfied by the existence of all moments of the pair \mathbf{Y}_1 . One can check that the Hessian local assumption **(T)** is also satisfied by the same arguments.

C.1.3 Explicit form of the covariance matrix

Lemma 3 *The matrix $\Sigma(\theta_0)$ in the Gaussian AR(1) model is given by:*

$$\Sigma(\theta_0) = V_{\theta_0}^{-1} \Omega(\theta_0) V_{\theta_0}^{-1}$$

with

$$V_{\theta_0} = \frac{1}{8\sqrt{\pi}(1 - \phi_0^2)^2} \begin{pmatrix} \gamma_0(7\phi_0^4 - 4\phi_0^2 + 4) & \frac{-5\phi_0^5 + 3\phi_0^3 + 2\phi_0}{2\gamma_0(1 - \phi_0^2)} \\ \frac{-5\phi_0^5 + 3\phi_0^3 + 2\phi_0}{2\gamma_0(1 - \phi_0^2)} & \frac{7\phi_0^2}{4\gamma_0^3} \end{pmatrix},$$

and

$$\Omega(\theta_0) = \Omega_0(\theta_0) + 2 \sum_{j>1}^{\infty} \Omega_{j-1}(\theta_0) = 4[P_2 - P_1] + 8 \sum_{j>1}^{\infty} (\tilde{C}_{j-1} - P_1)$$

where:

$$P_1 = \begin{pmatrix} \frac{\phi_0^2 \gamma_0^2 (2 - \phi_0^2)^2}{64\pi(1 - \phi_0^2)^2} & \frac{\phi_0^3 (2 - \phi_0^2)}{128\pi(1 - \phi_0^2)^2} \\ \frac{\phi_0^3 (2 - \phi_0^2)}{128\pi(1 - \phi_0^2)^2} & \frac{\phi_0^4}{256\pi(1 - \phi_0^2)^2 \gamma_0^2} \end{pmatrix},$$

and P_2 is the 2×2 symmetric matrix multiplied by a factor $\frac{1}{\sqrt{\pi(\gamma_0^2 - \sigma_\varepsilon^2)}}$ and its coefficients $(P_{lm}^2)_{1 \leq l, m \leq 2}$ are given by:

$$\begin{aligned} P_{11}^2 &= (\psi_1^{\phi_0})^2 \mathcal{F} \tilde{V}_1 \left(\tilde{V}_2 + 3 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 15 (\psi_2^{\phi_0})^2 \mathcal{F} \tilde{V}_1^3 \left(\tilde{V}_2 + 7 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 6 \psi_1^{\phi_0} \psi_2^{\phi_0} \mathcal{F} \tilde{V}_1^2 \left(\tilde{V}_2 + 5 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right). \\ P_{22}^2 &= (\psi_1^{\sigma_0^2})^2 \mathcal{F} \tilde{V}_1 \left(\tilde{V}_2 + 3 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 15 (\psi_2^{\sigma_0^2})^2 \mathcal{F} \tilde{V}_1^3 \left(\tilde{V}_2 + 7 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 6 \psi_1^{\sigma_0^2} \psi_2^{\sigma_0^2} \mathcal{F} \tilde{V}_1^2 \left(\tilde{V}_2 + 5 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right). \\ P_{12}^2 &= \psi_1^{\phi_0} \psi_1^{\sigma_0^2} \mathcal{F} \tilde{V}_1 \left(\tilde{V}_2 + 3 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 15 \psi_2^{\phi_0} \psi_2^{\sigma_0^2} \mathcal{F} \tilde{V}_1^3 \left(\tilde{V}_2 + 7 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 3 \psi_1^{\phi_0} \psi_2^{\sigma_0^2} \mathcal{F} \tilde{V}_1^2 \left(\tilde{V}_2 + 5 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right) \\ &\quad + 3 \psi_1^{\sigma_0^2} \psi_2^{\phi_0} \mathcal{F} \tilde{V}_1^2 \left(\tilde{V}_2 + 5 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right), \end{aligned}$$

with $\mathcal{F} = \frac{1}{(\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2} \tilde{V}_1^{1/2} \tilde{V}_2^{1/2}$, $\tilde{V}_1^{-1} = \frac{2}{(\gamma_0^2 - \sigma_\varepsilon^2)} + \left(\frac{\gamma_0^2 + \sigma_\varepsilon^2}{(\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2} \right) \left(1 - \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \right)$, $\tilde{V}_2 = \frac{(\gamma_0^2 + \sigma_\varepsilon^2)^2 - \phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)}$, and:

$$\begin{aligned} \psi_1^{\phi_0} &= \frac{1}{\sqrt{2\pi}} \frac{1}{(\gamma_0^2 - \sigma_\varepsilon^2)^{3/2}} \left(\frac{(1 + \phi_0^2)\gamma_0^2}{(1 - \phi_0^2)} - \frac{3\phi_0^2\gamma_0^4}{(1 - \phi_0^2)(\gamma_0^2 - \sigma_\varepsilon^2)} \right). \\ \psi_1^{\sigma_0^2} &= \frac{1}{\sqrt{2\pi}} \frac{1}{(\gamma_0^2 - \sigma_\varepsilon^2)^{3/2}} \left(\frac{\phi_0}{(1 - \phi_0^2)} - \frac{3\phi_0\gamma_0^2}{2(1 - \phi_0^2)(\gamma_0^2 - \sigma_\varepsilon^2)} \right). \\ \psi_2^{\phi_0} &= \frac{1}{\sqrt{2\pi}} \frac{1}{(\gamma_0^2 - \sigma_\varepsilon^2)^{7/2}} \frac{\gamma_0^4\phi_0^2}{(1 - \phi_0^2)}. \\ \psi_2^{\sigma_0^2} &= \frac{1}{\sqrt{2\pi}} \frac{1}{(\gamma_0^2 - \sigma_\varepsilon^2)^{7/2}} \frac{\gamma_0^2\phi_0}{2(1 - \phi_0^2)} \end{aligned}$$

The covariance terms are given by:

$$\begin{aligned} \tilde{c}_{j-1} &= \frac{\phi_0^2}{2\pi\gamma_0^2} \\ &\left(\begin{array}{cc} (4\phi_0^4 - 4\phi_0^2 + 1)\tilde{c}_1(j) + \frac{2\phi_0^2(1-2\phi_0^2)}{\gamma_0^2}\tilde{c}_2(j) + \frac{\phi_0^4}{\gamma_0^4}\tilde{c}_3(j) & \frac{\phi_0(2\phi_0^2-1)}{2\gamma_0^2}\tilde{c}_1(j) + \frac{\phi_0(1-3\phi_0^2)}{2\gamma_0^4}\tilde{c}_2(j) + \frac{\phi_0^3}{2\gamma_0^6}\tilde{c}_3(j) \\ \frac{\phi_0(2\phi_0^2-1)}{2\gamma_0^2}\tilde{c}_1(j) + \frac{\phi_0(1-3\phi_0^2)}{2\gamma_0^4}\tilde{c}_2(j) + \frac{\phi_0^3}{2\gamma_0^6}\tilde{c}_3(j) & \frac{\phi_0^2}{4\gamma_0^4}\tilde{c}_1(j) - \frac{\phi_0^2}{2\gamma_0^6}\tilde{c}_2(j) + \frac{\phi_0^2}{4\gamma_0^8}\tilde{c}_3(j) \end{array} \right), \end{aligned}$$

with:

$$\begin{aligned} \tilde{c}_1(j) &= \frac{1}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} V_j^{3/2} \left(\gamma + \frac{3\phi_0^{2j} V_j}{(2 - \phi_0^{2j})^2} \right), \\ \tilde{c}_2(j) &= \frac{3}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} V_j^{5/2} \left(\gamma + 5 \frac{\phi_0^{2j} V_j}{(2 - \phi_0^{2j})^2} \right), \\ \tilde{c}_3(j) &= \frac{3(2 - \phi_0^{2j})^{-1/2}}{\gamma_0} V_j^{5/2} \left[3\gamma^2 + 5V_j(4\gamma + 2) \frac{\phi_0^{2j}}{(2 - \phi_0^{2j})^2} + 35V_j^2 \frac{\phi_0^{4j}}{(2 - \phi_0^{2j})^4} \right], \end{aligned}$$

where:

$$V_j = \frac{\gamma_0^2(1 - \phi_0^{2j})(2 - \phi_0^{2j})}{(2 - \phi_0^{2j})^2 - \phi_0^{2j}} \quad \text{and} \quad \gamma = \frac{\gamma_0^2(1 - \phi_0^{2j})}{2 - \phi_0^{2j}}$$

Moreover $\lim_{j \rightarrow \infty} \Omega_{j-1}(\theta_0) = 0_{\mathcal{M}_{2 \times 2}}$.

Remark 6 In practice, for the computing of the covariance matrix $\Omega_{j-1}(\theta)$ that appears in Corollary 1, we have truncated the infinite sum ($q_{trunc} = 100$).

Proof Calculus of ∇m

For all $x \in \mathbb{R}$, the function $l_\theta(x)$ is two times differentiable w.r.t θ on the compact subset Θ . More precisely, note that since $\gamma^2 = \sigma^2/(1 - \phi^2)$, it follows from the definition of the subset Θ that $(\gamma^2 - \sigma_\varepsilon^2) > 0$. So that for all \mathbf{y}_i in \mathbb{R}^2 the function $m_\theta(\mathbf{y}_i) : \theta \in \Theta \mapsto m_\theta(\mathbf{y}_i)$ is differentiable and:

$$\begin{aligned} \nabla_{\theta}(m_{\theta}(\mathbf{y}_i)) &= \left(\frac{\partial m_{\theta}(\mathbf{y}_i)}{\partial \phi}, \frac{\partial m_{\theta}(\mathbf{y}_i)}{\partial \sigma^2} \right)' \\ &= \left(\frac{\partial \|l_{\theta}\|_2^2}{\partial \phi} - 2y_{i+1}u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(y_i), \frac{\partial \|l_{\theta}\|_2^2}{\partial \sigma^2} - 2y_{i+1}u_{\frac{\partial l_{\theta}}{\partial \sigma^2}}^*(y_i) \right)', \end{aligned}$$

with:

$$\begin{aligned} \frac{\partial}{\partial \phi} \|l_{\theta}\|_2^2 &= \frac{\phi\gamma(2 - \phi^2)}{4\sqrt{\pi}(1 - \phi^2)}, \\ \frac{\partial}{\partial \sigma^2} \|l_{\theta}\|_2^2 &= \frac{\phi^2}{8\sqrt{\pi}(1 - \phi^2)}. \end{aligned}$$

And, the function $u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(x)$ and $u_{\frac{\partial l_{\theta}}{\partial \sigma^2}}^*(x)$ are given in Eqs. (28)–(29). Therefore,

$$\begin{aligned} \nabla_{\theta}m_{\theta}(\mathbf{y}_i) &= \left(\left(\frac{\phi_0\gamma_0(2 - \phi_0^2)}{4\sqrt{\pi}(1 - \phi_0^2)} - 2y_{i+1} \left(\Psi_1^{\phi_0} y_i + \Psi_2^{\phi_0} y_i^3 \right) \exp \left(-\frac{y_i^2}{2(\gamma_0^2 - \sigma_{\varepsilon}^2)} \right) \right) \right) \\ &\quad \left(\left(\frac{\phi_0^2}{8\sqrt{\pi}(1 - \phi_0^2)} - 2y_{i+1} \left(\Psi_1^{\sigma_0^2} y_i + \Psi_2^{\sigma_0^2} y_i^3 \right) \exp \left(-\frac{y_i^2}{2(\gamma_0^2 - \sigma_{\varepsilon}^2)} \right) \right) \right) \end{aligned} \tag{30}$$

at the point θ_0 .

Calculus of P_1 : Recall that we have:

$$\begin{aligned} P_1 &= \mathbb{E} [b_{\theta_0}(X_1) (\nabla_{\theta}l_{\theta}(X_1))] \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta}l_{\theta}(X_1))]' \\ P_2 &= \mathbb{E} [Y_2^2 (u_{\nabla_{\theta}l_{\theta}}^*(Y_1))^2]. \end{aligned}$$

And the moments $(\mu_{2k})_{k \in \mathbb{N}}$ of a centered Gaussian random variable with variance σ^2 are given by:

$$\mu_{2k} = \binom{(2k)!}{2^k k!} \sigma^{2k}.$$

We define by $P(x)$ a polynomial function of ordinary degree. We are interested in the calculus of $\mathbb{E} [P(X)g_{0,\gamma^2}(X)]$, where $X \sim \mathcal{N}(0, \gamma^2)$. We have:

$$\begin{aligned} \mathbb{E} [P(X)g_{0,\gamma^2}(X)] &= \int P(x) \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{x^2}{2\gamma^2}} \frac{1}{\sqrt{2\pi}\gamma} e^{-\frac{x^2}{2\gamma^2}} dx \\ &= \frac{1}{2\pi\gamma^2} \int P(x) e^{-\frac{x^2}{\gamma^2}} dx \\ &= \frac{1}{2\sqrt{\pi}\gamma} \mathbb{E} [P(\bar{X})], \end{aligned}$$

where $\bar{X} \sim \mathcal{N}\left(0, \frac{\gamma^2}{2}\right)$.

Denote by B_1 the constant $\frac{1}{2\sqrt{\pi}\gamma_0^2}$. We obtain:

$$P_1 = \begin{pmatrix} \mathbb{E} \left[b_{\phi_0}(X_1) \frac{\partial l_{\theta}}{\partial \phi}(\theta, X_1) \right]^2 & \mathbb{E} \left[b_{\phi_0}(X_1) \frac{\partial l_{\theta}}{\partial \phi}(\theta, X_1) \right] \mathbb{E} \left[b_{\phi_0}(X_1) \frac{\partial l_{\theta}}{\partial \sigma^2}(\theta, X_1) \right] \\ \mathbb{E} \left[b_{\phi_0}(X_1) \frac{\partial l_{\theta}}{\partial \phi}(\theta, X_1) \right] \mathbb{E} \left[b_{\phi_0}(X_1) \frac{\partial l_{\theta}}{\partial \sigma^2}(\theta, X_1) \right] & \mathbb{E} \left[b_{\phi_0}(X_1) \frac{\partial l_{\theta}}{\partial \sigma^2}(\theta, X_1) \right]^2 \end{pmatrix}$$

$$= B_1^2 \phi_0^2 \begin{pmatrix} \mathbb{E} [H_{11}(\bar{X})]^2 & \mathbb{E} [H_{12}(\bar{X})] \mathbb{E} [H_{21}(\bar{X})] \\ \mathbb{E} [H_{21}(\bar{X})] \mathbb{E} [H_{12}(\bar{X})] & \mathbb{E} [H_{22}(\bar{X})]^2 \end{pmatrix},$$

where $\bar{X} \sim \mathcal{N}(0, \frac{\gamma_0^2}{2})$. The polynomials $(H_{ij}(x))_{1 \leq i, j \leq 2}$ are given by:

$$H_{11}(x) = (a_1 x^2 + a_2 x^4),$$

$$H_{12}(x) = (b_1 x^2 + b_2 x^4),$$

$$H_{21}(x) = (a_1 x^2 + a_2 x^4),$$

$$H_{22}(x) = (b_1 x^2 + b_2 x^4).$$

Lastly, by replacing the terms B_1 , a_1 , and a_2 by their expressions given in Eq. (24) at the point θ_0 , we obtain:

$$P_1 = \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1))] \mathbb{E} [b_{\phi_0}(X_1) (\nabla_{\theta} l_{\theta}(X_1))]'$$

$$= \begin{pmatrix} \frac{\phi_0^2 \gamma_0^2 (2 - \phi_0^2)^2}{64\pi (1 - \phi_0^2)^2} & \frac{\phi_0^3 (2 - \phi_0^2)}{128\pi (1 - \phi_0^2)^2} \\ \frac{\phi_0^3 (2 - \phi_0^2)}{128\pi (1 - \phi_0^2)^2} & \frac{\phi_0^4}{256\pi (1 - \phi_0^2)^2 \gamma_0^2} \end{pmatrix}.$$

Calculus of P_2 :

$$\mathbb{E} \left[(Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1)) (Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1))' \right] = \begin{pmatrix} \mathbb{E} \left[Y_2^2 \left(u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(Y_1) \right)^2 \right] & \mathbb{E} \left[Y_2^2 \left(u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(Y_1) \right) \left(u_{\frac{\partial l_{\theta}}{\partial \sigma^2}}^*(Y_1) \right) \right] \\ \mathbb{E} \left[Y_2^2 \left(u_{\frac{\partial l_{\theta}}{\partial \sigma^2}}^*(Y_1) \right) \left(u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(Y_1) \right) \right] & \mathbb{E} \left[Y_2^2 \left(u_{\frac{\partial l_{\theta}}{\partial \sigma^2}}^*(Y_1) \right)^2 \right] \end{pmatrix}.$$

We have:

$$\left(2\pi \frac{(\gamma_0^2 - \sigma_{\varepsilon}^2)}{2} \right)^{-1/2} \mathbb{E} \left[Y_2^2 \left(u_{\frac{\partial l_{\theta}}{\partial \phi}}^*(Y_1) \right)^2 \right]$$

$$= \mathbb{E} \left[Y_2^2 \left(\Psi_1^{\phi_0} Y_1 + \Psi_2^{\phi_0} Y_1^3 \right)^2 \times g \left(0, \frac{(\gamma_0^2 - \sigma_{\varepsilon}^2)}{2} \right) \right]$$

$$= \left(\Psi_1^{\phi_0} \right)^2 \mathbb{E} \left[Y_2^2 Y_1^2 \times g \left(0, \frac{(\gamma_0^2 - \sigma_{\varepsilon}^2)}{2} \right) \right]$$

$$\begin{aligned}
 &+ (\Psi_2^{\phi_0})^2 \mathbb{E} \left[Y_2^2 Y_1^6 \times g \left(0, \frac{(\gamma_0^2 - \sigma_\varepsilon^2)}{2} \right) \right] \\
 &+ 2\Psi_1^{\phi_0} \Psi_2^{\phi_0} \mathbb{E} \left[Y_2^2 Y_1^4 \times g \left(0, \frac{(\gamma_0^2 - \sigma_\varepsilon^2)}{2} \right) \right].
 \end{aligned} \tag{31}$$

The density of Y_1 is $g(0, \mathcal{J}_{\theta_0})$. Then, $g(0, \mathcal{J}_{\theta_0}) \times \exp \left(-\frac{y_1^2}{(\gamma_0^2 - \sigma_\varepsilon^2)} \right)$ is equal to:

$$\begin{aligned}
 &\frac{1}{2\pi} \frac{1}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)^{1/2}} \exp \left(-\frac{1}{2} \frac{1}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)} \right. \\
 &\quad \left. ((\sigma_\varepsilon^2 + \gamma_0^2)(y_1^2 + y_2^2) - 2\phi_0 \gamma_0^2 y_1 y_2) \right) \times \exp \left(-\frac{1}{2} \frac{2}{(\gamma_0^2 - \sigma_\varepsilon^2)} y_1^2 \right) \\
 &= \frac{1}{2\pi} \frac{1}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)^{1/2}} \\
 &\quad \times \exp \left(-\frac{1}{2} y_1^2 \left(\frac{2}{(\gamma_0^2 - \sigma_\varepsilon^2)} - \frac{(\gamma_0^2 + \sigma_\varepsilon^2)}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)} \right) \right) \\
 &\quad \times \exp \left(-\frac{1}{2} y_2^2 \left(\frac{(\gamma_0^2 + \sigma_\varepsilon^2)}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)} \right) \right) \\
 &\quad \times \exp \left(-\frac{1}{2} y_1 y_2 \left(\frac{2\phi_0 \gamma_0^2}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)} \right) \right) \\
 &= \frac{1}{2\pi} \frac{1}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)^{1/2}} \exp \left(-\frac{1}{2} \left(\tilde{V}_2^{-1} \left(y_2 - \frac{\phi_0 \gamma_0^2}{\gamma_0^2 + \sigma_\varepsilon^2} y_1 \right)^2 \right) \right) \\
 &\quad \times \exp \left(-\frac{1}{2} y_1^2 \tilde{V}_1^{-1} \right),
 \end{aligned}$$

with $\tilde{V}_1^{-1} = \frac{2}{(\gamma_0^2 - \sigma_\varepsilon^2)} + \left(\frac{\gamma_0^2 + \sigma_\varepsilon^2}{(\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2} \right) \left(1 - \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \right)$ and $\tilde{V}_2 = \frac{(\gamma_0^2 + \sigma_\varepsilon^2)^2 - \phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)}$.

Then, we obtain:

$$\begin{aligned}
 g(0, \mathcal{J}_{\theta_0}) \times \exp \left(-\frac{y_1^2}{(\gamma_0^2 - \sigma_\varepsilon^2)} \right) &= \frac{1}{((\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2)^{1/2}} \tilde{V}_1^{1/2} \tilde{V}_2^{1/2} \\
 &\quad g(\phi_0 \gamma_0^2 y_1 / (\gamma_0^2 + \sigma_\varepsilon^2), \tilde{V}_2)(y_2) g(0, \tilde{V}_1)(y_1).
 \end{aligned}$$

In the following, we set $\mathcal{F} = \frac{1}{(\sigma_\varepsilon^2 + \gamma_0^2)^2 - \gamma_0^4 \phi_0^2} \tilde{V}_1^{1/2} \tilde{V}_2^{1/2}$. Now, we can compute the moments:

$$\begin{aligned} & (\Psi_1^{\phi_0})^2 \mathbb{E} \left[Y_2^2 Y_1^2 \exp \left(-\frac{Y_1^2}{(\gamma_0^2 - \sigma_\varepsilon^2)} \right) \right] \\ &= (\Psi_1^{\phi_0})^2 \mathcal{F} \int y_1^2 g_{(0, \tilde{v}_1)}(y_1) dy_1 \int y_2^2 g_{(\phi_0 \gamma_0^2 y_1 / (\gamma_0^2 + \sigma_\varepsilon^2), \tilde{v}_2)}(y_2) dy_2 \\ &= (\Psi_1^{\phi_0})^2 \mathcal{F} \int y_1^2 g_{(0, v_1)}(y_1) dy_1 \mathbb{E} [G^2] \text{ where } G \sim \mathcal{N}(\phi_0 \gamma_0^2 y_1 / (\gamma_0^2 + \sigma_\varepsilon^2), \tilde{V}_2) \\ &= (\Psi_1^{\phi_0})^2 \mathcal{F} \int \left(\tilde{V}_2 y_1^2 + \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} y_1^4 \right) g_{(0, \tilde{v}_1)}(y_1) dy_1 \\ &= (\Psi_1^{\phi_0})^2 \mathcal{F} \tilde{V}_2 \tilde{V}_1 + 3 (\Psi_1^{\phi_0})^2 \mathcal{F} \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1^2 \\ &= (\Psi_1^{\phi_0})^2 \mathcal{F} \tilde{V}_1 \left(\tilde{V}_2 + 3 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right). \end{aligned}$$

In a similar manner, we have:

$$\begin{aligned} & (\Psi_2^{\phi_0})^2 \mathbb{E} \left[Y_2^2 Y_1^6 \exp \left(-\frac{Y_1^2}{(\gamma_0^2 - \sigma_\varepsilon^2)} \right) \right] \\ &= (\Psi_2^{\phi_0})^2 \mathcal{F} \int y_1^6 g_{(0, v_1)}(y_1) dy_1 \mathbb{E} [G^2] \text{ where } G \sim \mathcal{N}(\phi_0 \gamma_0^2 y_1 / (\gamma_0^2 + \sigma_\varepsilon^2), \tilde{V}_2) \\ &= (\Psi_2^{\phi_0})^2 \mathcal{F} \int \left(\tilde{V}_2 y_1^6 + \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} y_1^8 \right) g_{(0, \tilde{v}_1)}(y_1) dy_1 \\ &= 15 (\Psi_2^{\phi_0})^2 \mathcal{F} \tilde{V}_2 \tilde{V}_1^3 + 105 (\Psi_2^{\phi_0})^2 \mathcal{F} \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1^4 \\ &= 15 (\Psi_2^{\phi_0})^2 \mathcal{F} \tilde{V}_1^3 \left(\tilde{V}_2 + 7 \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} \tilde{V}_1 \right), \end{aligned}$$

and

$$\begin{aligned} & 2\Psi_1^{\phi_0} \Psi_2^{\phi_0} \mathbb{E} \left[Y_2^2 Y_1^4 \exp \left(-\frac{Y_1^2}{(\gamma_0^2 - \sigma_\varepsilon^2)} \right) \right] \\ &= 2\Psi_1^{\phi_0} \Psi_2^{\phi_0} \mathcal{F} \int y_1^4 g_{(0, v_1)}(y_1) dy_1 \mathbb{E} [G^2] \text{ where } G \sim \mathcal{N}(\phi_0 \gamma_0^2 y_1 / (\gamma_0^2 + \sigma_\varepsilon^2), \tilde{V}_2) \\ &= 2\Psi_1^{\phi_0} \Psi_2^{\phi_0} \mathcal{F} \int \left(\tilde{V}_2 y_1^4 + \frac{\phi_0^2 \gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2} y_1^6 \right) g_{(0, \tilde{v}_1)}(y_1) dy_1 \end{aligned}$$

$$\begin{aligned}
 &= 6\psi_1^{\phi_0}\psi_2^{\phi_0}\mathcal{F}\tilde{V}_2\tilde{V}_1^2 + 30\psi_1^{\phi_0}\psi_2^{\phi_0}\mathcal{F}\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1^3 \\
 &= 6\psi_1^{\phi_0}\psi_2^{\phi_0}\mathcal{F}\tilde{V}_1^2\left(\tilde{V}_2 + 5\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right).
 \end{aligned}$$

By replacing all the terms of Eq. (31) we obtain:

$$\begin{aligned}
 \mathbb{E}\left[Y_2^2\left(u_{\frac{\partial l_\theta}{\partial \phi}}^*(Y_1)\right)^2\right] &= (\psi_1^{\phi_0})^2\mathcal{F}\tilde{V}_1\left(\tilde{V}_2 + 3\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right) \\
 &\quad + 15(\psi_2^{\phi_0})^2\mathcal{F}\tilde{V}_1^3\left(\tilde{V}_2 + 7\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right) \\
 &\quad + 6\psi_1^{\phi_0}\psi_2^{\phi_0}\mathcal{F}\tilde{V}_1^2\left(\tilde{V}_2 + 5\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right), \tag{32}
 \end{aligned}$$

and

$$\begin{aligned}
 \mathbb{E}\left[Y_2^2\left(u_{\frac{\partial l_\theta}{\partial \sigma^2}}^*(Y_1)\right)^2\right] &= (\psi_1^{\sigma_0^2})^2\mathcal{F}\tilde{V}_1\left(\tilde{V}_2 + 3\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right) \\
 &\quad + 15(\psi_2^{\sigma_0^2})^2\mathcal{F}\tilde{V}_1^3\left(\tilde{V}_2 + 7\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right) \\
 &\quad + 6\psi_1^{\sigma_0^2}\psi_2^{\sigma_0^2}\mathcal{F}\tilde{V}_1^2\left(\tilde{V}_2 + 5\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right), \tag{33}
 \end{aligned}$$

and

$$\begin{aligned}
 &\mathbb{E}\left[Y_2^2\left(u_{\frac{\partial l_\theta}{\partial \phi}}^*(Y_1)\right)\left(u_{\frac{\partial l_\theta}{\partial \sigma^2}}^*(Y_1)\right)\right] \\
 &= \psi_1^{\phi_0}\psi_1^{\sigma_0^2}\mathbb{E}\left[Y_2^2Y_1^2 \times g\left(0, \frac{(\gamma_0^2 - \sigma_\varepsilon^2)}{2}\right)\right] + \psi_2^{\phi_0}\psi_2^{\sigma_0^2}\mathbb{E}\left[Y_2^2Y_1^6 \times g\left(0, \frac{(\gamma_0^2 - \sigma_\varepsilon^2)}{2}\right)\right] \\
 &\quad + \psi_1^{\phi_0}\psi_2^{\sigma_0^2}\mathbb{E}\left[Y_2^2Y_1^4 \times g\left(0, \frac{(\gamma_0^2 - \sigma_\varepsilon^2)}{2}\right)\right] + \psi_2^{\phi_0}\psi_1^{\sigma_0^2}\mathbb{E}\left[Y_2^2Y_1^4 \times g\left(0, \frac{(\gamma_0^2 - \sigma_\varepsilon^2)}{2}\right)\right] \\
 &= \psi_1^{\phi_0}\psi_1^{\sigma_0^2}\mathcal{F}\tilde{V}_1\left(\tilde{V}_2 + 3\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right) + 15\psi_2^{\phi_0}\psi_2^{\sigma_0^2}\mathcal{F}\tilde{V}_1^3 \\
 &\quad \times \left(\tilde{V}_2 + 7\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2 + \sigma_\varepsilon^2)^2}\tilde{V}_1\right)
 \end{aligned}$$

$$\begin{aligned}
 &+3\Psi_1^{\phi_0}\Psi_2^{\sigma_0^2}\mathcal{F}\tilde{V}_1^2\left(\tilde{V}_2+5\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2+\sigma_\varepsilon^2)^2}\tilde{V}_1\right)+3\Psi_2^{\phi_0}\Psi_1^{\sigma_0^2}\mathcal{F}\tilde{V}_1^2 \\
 &\times\left(\tilde{V}_2+5\frac{\phi_0^2\gamma_0^4}{(\gamma_0^2+\sigma_\varepsilon^2)^2}\tilde{V}_1\right). \tag{34}
 \end{aligned}$$

Calculus of $\text{Cov}(\nabla_\theta m_\theta(Y_1), \nabla_\theta m_\theta(Y_j))$: We want to compute:

$$\text{Cov}(\nabla_\theta m_\theta(Y_1), \nabla_\theta m_\theta(Y_j)) = 4\left[\tilde{C}_{j-1} - P_1\right].$$

Since we have already computed the terms of the matrix P_1 , it remains to compute the terms of the covariance matrix \tilde{C}_{j-1} given by:

$$\tilde{C}_{j-1} = \mathbb{E}\left[b_{\phi_0}(X_1)b_{\phi_0}(X_j)(\nabla_\theta l_\theta(X_1))(\nabla_\theta l_\theta(X_j))'\right].$$

For all $j > 1$, the pair (X_1, X_j) has a multivariate normal density $g_{(0,\mathcal{W})}$ where \mathcal{W} is given by:

$$\mathcal{W} = \gamma_0^2 \begin{pmatrix} 1 & \phi_0^j \\ \phi_0^j & 1 \end{pmatrix} \quad \text{and} \quad \mathcal{W}^{-1} = \frac{1}{\gamma_0^2(1-\phi_0^{2j})} \begin{pmatrix} 1 & -\phi_0^j \\ -\phi_0^j & 1 \end{pmatrix}.$$

The density of the couple (X_1, X_j) is:

$$g_{(0,\mathcal{W})}(x_1, x_j) = \frac{1}{2\pi} \det(\mathcal{W})^{-1/2} \exp\left(-\frac{1}{2}(x_1, x_j)'\mathcal{W}^{-1}(x_1, x_j)\right).$$

We start by computing:

$$g_{(0,\mathcal{W})}(x_1, x_j) \times \exp\left(-\frac{1}{2\gamma_0^2}(x_1^2 + x_j^2)\right).$$

We have:

$$\begin{aligned}
 &g_{(0,\mathcal{W})}(x_1, x_j) \times \exp\left(-\frac{1}{2\gamma_0^2}(x_1^2 + x_j^2)\right) \\
 &= \frac{1}{2\pi} \det(\mathcal{W})^{-1/2} \exp\left(-\frac{1}{2(1-\phi_0^{2j})\gamma_0^2}\left(x_1^2(1-\phi_0^{2j}) + x_j^2(1-\phi_0^{2j}) + x_1^2 - 2\phi_0^j x_1 x_j + x_j^2\right)\right), \\
 &= \frac{1}{2\pi} \det(\mathcal{W})^{-1/2} \exp\left(-\frac{1}{2(1-\phi_0^{2j})\gamma_0^2}\left[\left(x_1^2(1-\phi_0^{2j}) + x_1^2 - 2\phi_0^j x_1 x_j\right)\right.\right. \\
 &\quad \left.\left.+ \left(x_j^2(1-\phi_0^{2j}) + x_j^2\right)\right]\right), \\
 &= \frac{1}{2\pi} \det(\mathcal{W})^{-1/2} \exp\left(-\frac{1}{2(1-\phi_0^{2j})\gamma_0^2}\left[(2-\phi_0^{2j})\left(x_1^2 - 2\frac{\phi_0^j}{(2-\phi_0^{2j})}x_1 x_j\right)\right.\right.
 \end{aligned}$$

$$\begin{aligned}
 & + \left(x_j^2(1 - \phi_0^{2j}) + x_j^2 \right) \Big], \\
 & = \frac{1}{2\pi} \det(\mathcal{W})^{-1/2} \exp \left(-\frac{(2 - \phi_0^{2j})}{2(1 - \phi_0^{2j})\gamma_0^2} \left(x_1 - \frac{\phi_0^j x_j}{(2 - \phi_0^{2j})} \right)^2 \right) \\
 & \quad \times \exp \left(-\frac{(2 - \phi_0^{2j})}{2(1 - \phi_0^{2j})\gamma_0^2} x_j^2 \left(1 - \frac{\phi_0^{2j}}{(2 - \phi_0^{2j})} \right) \right).
 \end{aligned}$$

For all $j > 1$, we define:

$$V_j = \frac{\gamma_0^2(1 - \phi_0^{2j})(2 - \phi_0^{2j})}{(2 - \phi_0^{2j})^2 - \phi_0^{2j}} \quad \text{and} \quad \mathcal{V} = \frac{\gamma_0^2(1 - \phi_0^{2j})}{(2 - \phi_0^{2j})}.$$

We can rewrite:

$$\begin{aligned}
 & g_{(0, \mathcal{W})}(x_1, x_j) \times \exp \left(-\frac{1}{2\gamma_0^2} (x_1^2 + x_j^2) \right) \\
 & = \frac{V_j^{1/2}}{\gamma_0} \frac{1}{\sqrt{2\pi} V_j^{1/2}} \exp \left(-\frac{1}{2V_j} x_j^2 \right) \times \frac{1}{(2 - \phi_0^{2j})^{1/2} \sqrt{2\pi} \mathcal{V}^{1/2}} \exp \left(-\frac{1}{2\mathcal{V}} \left(x_1 - \frac{\phi_0^j x_j}{(2 - \phi_0^{2j})} \right)^2 \right).
 \end{aligned}$$

So, by Fubini's Theorem, we obtain:

$$\begin{aligned}
 & \mathbb{E} \left[X_1^2 X_j^2 \exp \left(-\frac{1}{2\gamma_0^2} (X_1^2 + X_j^2) \right) \right] \\
 & = \frac{1}{\gamma_0} V_j^{1/2} \int x_j^2 \frac{1}{\sqrt{2\pi} V_j^{1/2}} \exp \left(-\frac{1}{2V_j} x_j^2 \right) \int x_1^2 \frac{(2 - \phi_0^{2j})^{-1/2}}{\sqrt{2\pi} \mathcal{V}^{1/2}} \exp \left(-\frac{1}{2\mathcal{V}} \left(x_1 - \frac{\phi_0^j x_j}{(2 - \phi_0^{2j})} \right)^2 \right) dx_1 dx_j, \\
 & = \frac{1}{\gamma_0} V_j^{1/2} \int x_j^2 \frac{1}{\sqrt{2\pi} V_j^{1/2}} \exp \left(-\frac{1}{2V_j} x_j^2 \right) (2 - \phi_0^{2j})^{-1/2} \mathbb{E}[G^2] dx_j,
 \end{aligned}$$

where $G \sim \mathcal{N} \left(\frac{\phi_0^j x_j}{(2 - \phi_0^{2j})}, \mathcal{V} \right)$. Thus, $\mathbb{E}[G^2] = \mathcal{V} + \left(\frac{\phi_0^j x_j}{(2 - \phi_0^{2j})} \right)^2$. We obtain:

$$\begin{aligned}
 & \mathbb{E} \left[X_1^2 X_j^2 \exp \left(-\frac{1}{2\gamma_0^2} (X_1^2 + X_j^2) \right) \right] \\
 & = (2 - \phi_0^{2j})^{-1/2} \frac{V_j^{1/2}}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} \int x_j^2 \left(\mathcal{V} + \left(\frac{\phi_0^j x_j}{(2 - \phi_0^{2j})} \right)^2 \right) \frac{1}{\sqrt{2\pi} V_j^{1/2}} \exp \left(-\frac{1}{2V_j} x_j^2 \right) dx_j \\
 & = \frac{V_j^{1/2}}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} \left(\mathcal{V} \mathbb{E}[G_j^2] + \frac{\phi_0^{2j}}{(2 - \phi_0^{2j})^2 \gamma_0} \mathbb{E}[G_j^4] \right) \\
 & = \frac{V_j^{3/2}}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} \left(\mathcal{V} + \frac{3\phi_0^{2j} V_j}{(2 - \phi_0^{2j})^2} \right) \\
 & = \tilde{c}_1(j),
 \end{aligned} \tag{35}$$

where $G_j \sim \mathcal{N}(0, V_j)$. Additionally, we have:

$$\begin{aligned} & \mathbb{E} \left[X_1^2 X_j^4 \exp \left(-\frac{1}{2\gamma_0^2} (X_1^2 + X_j^2) \right) \right] \\ &= \frac{V_j^{1/2}}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} \mathcal{V} \mathbb{E}[G_j^4] + V_j^{1/2} \frac{(2 - \phi_0^{2j})^{-1/2} \phi_0^{2j}}{(2 - \phi_0^{2j})^2} \mathbb{E}[G_j^6], \\ &= \frac{3V_j^{5/2}}{\gamma_0} (2 - \phi_0^{2j})^{-1/2} \left(\mathcal{V} + 5 \frac{\phi_0^{2j} V_j}{(2 - \phi_0^{2j})^2} \right), \\ &= \tilde{c}_2(j). \end{aligned} \tag{36}$$

Now, we are interested in $\mathbb{E} \left[X_1^4 X_j^4 \exp \left(-\frac{1}{2\gamma_0^2} (X_1^2 + X_j^2) \right) \right]$. In a similar manner, we obtain:

$$\begin{aligned} & \mathbb{E} \left[X_1^4 X_j^4 \exp \left(-\frac{1}{2\gamma_0^2} (X_1^2 + X_j^2) \right) \right] \\ &= \frac{V_j^{1/2}}{\gamma_0} \int x_j^4 \frac{1}{\sqrt{2\pi} V_j^{1/2}} \exp \left(-\frac{1}{2V_j} x_j^2 \right) (2 - \phi_0^{2j})^{-1/2} \mathbb{E}[G^4] dx_j, \end{aligned} \tag{37}$$

where $G \sim \mathcal{N} \left(\frac{\phi_0^j x_j}{(2 - \phi_0^{2j})}, \mathcal{V} \right)$. We use the fact that the moments of a random variable $X \sim \mathcal{N}(\mu, v)$ are:

$$\begin{aligned} \mathbb{E}[X^n] &= (n - 1)v\mathbb{E}[X^{n-2}] + \mu\mathbb{E}[X^{n-1}] \\ \mathbb{E}[G^4] &= 3\mathcal{V}\mathbb{E}[G^2] + \left(\frac{\phi_0^j x_j}{(2 - \phi_0^{2j})} \mathbb{E}[G^3] \right) \\ &= 3\mathcal{V}^2 + (4\mathcal{V} + 2) \frac{\phi_0^{2j} x_j^2}{(2 - \phi_0^{2j})^2} + \frac{\phi_0^{4j} x_j^4}{(2 - \phi_0^{2j})^4}. \end{aligned}$$

By replacing $\mathbb{E}[G^4]$ in Eq. (37), we have:

$$\begin{aligned} & \mathbb{E} \left[X_1^4 X_j^4 \exp \left(-\frac{1}{2\gamma_0^2} (X_1^2 + X_j^2) \right) \right] \\ &= \frac{3(2 - \phi_0^{2j})^{-1/2}}{\gamma_0} V_j^{5/2} \left[3\mathcal{V}^2 + 5V_j(4\mathcal{V} + 2) \frac{\phi_0^{2j}}{(2 - \phi_0^{2j})^2} + 35V_j^2 \frac{\phi_0^{4j}}{(2 - \phi_0^{2j})^4} \right], \\ &= \tilde{c}_3(j). \end{aligned} \tag{38}$$

For all $j > 1$, the matrix \tilde{C}_{j-1} is given by:

$$\tilde{C}_{j-1} = \frac{\phi_0^2}{2\pi\gamma_0^2} \begin{pmatrix} a_1^2\tilde{c}_1(j) + 2a_1a_2\tilde{c}_2(j) + a_2^2\tilde{c}_3(j) & a_1b_1\tilde{c}_1(j) + a_1b_2 + a_2b_1\tilde{c}_2(j) + a_2b_2\tilde{c}_3(j) \\ a_1b_1\tilde{c}_1(j) + a_1b_2 + a_2b_1\tilde{c}_2(j) + a_2b_2\tilde{c}_3(j) & b_1^2\tilde{c}_1(j) + 2b_1b_2\tilde{c}_2(j) + b_2^2\tilde{c}_3(j) \end{pmatrix},$$

where the coefficients $\tilde{c}_1(j)$, $\tilde{c}_2(j)$, and $\tilde{c}_3(j)$ are given by (35), (36) and (38).

Finally, by replacing the terms a_1 , a_2 , b_1 and b_2 , the matrix \tilde{C}_{j-1} is equal to:

$$\tilde{C}_{j-1} = A \begin{pmatrix} (4\phi_0^4 - 4\phi_0^2 + 1)\tilde{c}_1(j) + \frac{2\phi_0^2(1-2\phi_0^2)}{\gamma_0^2}\tilde{c}_2(j) + \frac{\phi_0^4}{\gamma_0^4}\tilde{c}_3(j) & \frac{\phi_0(2\phi_0^2-1)}{2\gamma_0^2}\tilde{c}_1(j) + \frac{\phi_0(1-3\phi_0^2)}{2\gamma_0^4}\tilde{c}_2(j) + \frac{\phi_0^3}{2\gamma_0^6}\tilde{c}_3(j) \\ \frac{\phi_0(2\phi_0^2-1)}{2\gamma_0^2}\tilde{c}_1(j) + \frac{\phi_0(1-3\phi_0^2)}{2\gamma_0^4}\tilde{c}_2(j) + \frac{\phi_0^3}{2\gamma_0^6}\tilde{c}_3(j) & \frac{\phi_0^2}{4\gamma_0^4}\tilde{c}_1(j) + \frac{-\phi_0^2}{2\gamma_0^6}\tilde{c}_2(j) + \frac{\phi_0^2}{4\gamma_0^8}\tilde{c}_3(j) \end{pmatrix},$$

where $A = \frac{\phi_0^2}{2\pi\gamma_0^2(1-\phi_0^2)^2}$.

Asymptotic behaviour of the covariance matrix $\Omega_{j-1}(\theta_0)$: By the stationary assumption $|\phi_0| < 1$, the limits of the following terms are:

$$\lim_{j \rightarrow \infty} V_j = \frac{\gamma_0^2}{2} \quad \text{and} \quad \lim_{j \rightarrow \infty} \mathcal{V} = \frac{\gamma_0^2}{2},$$

and

$$\lim_{j \rightarrow \infty} \tilde{c}_1(j) = \frac{\gamma_0^4}{8}, \quad \lim_{j \rightarrow \infty} \tilde{c}_2(j) = \frac{3\gamma_0^6}{16}, \quad \lim_{j \rightarrow \infty} \tilde{c}_3(j) = \frac{9\gamma_0^8}{32}.$$

Therefore,

$$\lim_{j \rightarrow \infty} \tilde{C}_{j-1} = \begin{pmatrix} \frac{\phi_0^2\gamma_0^2(2-\phi_0^2)^2}{64\pi(1-\phi_0^2)^2} & \frac{\phi_0^3(2-\phi_0^2)}{128\pi(1-\phi_0^2)^2} \\ \frac{\phi_0^3(2-\phi_0^2)}{128\pi(1-\phi_0^2)^2} & \frac{\phi_0^4}{256\pi(1-\phi_0^2)^2\gamma_0^2} \end{pmatrix} = P_1.$$

We obtain:

$$\begin{aligned} \lim_{j \rightarrow \infty} \text{Cov}(\nabla_{\theta} m_{\theta_0}(Y_1), \nabla_{\theta} m_{\theta_0}(Y_j)) &= 4 \lim_{j \rightarrow \infty} (\tilde{C}_{j-1} - P_1) \\ &= 0_{\mathcal{M}_{2 \times 2}}. \end{aligned}$$

We conclude that the covariance between the two vectors $\nabla_{\theta} m_{\theta_0}(Y_1)$, $\nabla_{\theta} m_{\theta_0}(Y_j)$ vanishes when the lag between the two observations Y_1 and Y_j goes to the infinity.

Calculus of V_{θ_0} : The Hessian matrix V_{θ_0} is given in Eq. (27).

C.2 The SV model

C.2.1 Contrast function

The \mathbb{L}_2 -norm and the Fourier transform of the function l_θ are the same as the Gaussian AR(1) model. The only difference is the law of the measurement noise which is a log-chi-square for the log-transform SV model.

Consider the random variable $\varepsilon = \beta \log(X^2) - \tilde{\varepsilon}$ where $\tilde{\varepsilon} = \beta \mathbb{E}[\log(X^2)]$ such that ε is centered. The random variable X is a standard Gaussian random. The Fourier transform of ε is given by:

$$\begin{aligned} \mathbb{E} [\exp (i \varepsilon y)] &= \exp \left(-i \tilde{\varepsilon} y\right) \mathbb{E}\left[X^{2 i \beta y}\right] \\ &= \exp \left(-i \tilde{\varepsilon} y\right) \frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{+\infty} x^{2 i \beta y} \exp \left(-\frac{x^2}{2}\right) dx \end{aligned}$$

By a change of variable $z = \frac{x^2}{2}$, one has:

$$\begin{aligned} \mathbb{E} [\exp (i \varepsilon y)] &= \exp \left(-i \tilde{\varepsilon} y\right) \frac{2^{i \beta y}}{\sqrt{\pi}} \underbrace{\int_0^{+\infty} z^{i \beta y-\frac{1}{2}} e^{-z} dz}_{\Gamma\left(\frac{1}{2}+i \beta y\right)} \\ &= \exp \left(-i \tilde{\varepsilon} y\right) \frac{2^{i \beta y}}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}+i \beta y\right), \end{aligned}$$

and the expression (14) of the contrast function follows with $u_{l_\theta}(y) = \frac{1}{2\sqrt{\pi}}$

$$\left(\frac{-i \phi y \gamma^2 \exp \left(-\frac{y^2}{2} \gamma^2\right)}{\exp \left(-i \tilde{\varepsilon} y\right) 2^{i \beta y} \Gamma\left(\frac{1}{2}+i \beta y\right)}\right).$$

C.2.2 Checking assumption of Theorem 1

Regularity conditions: The proof is essentially the same as for the Gaussian case since the functions $l_\theta(x)$ and $\mathbf{P}m_\theta$ are the same. We need only to check the assumptions (C) and (T). These assumptions are satisfied since Fan (1991) showed that the noises ε_i have a Fourier transform f_ε^* which satisfies:

$$\left|f_\varepsilon^*(x)\right|=\sqrt{2} \exp \left(-\frac{\pi}{2}|x|\right)\left(1+O\left(\frac{1}{|x|}\right)\right), \quad|x| \rightarrow \infty,$$

which means that f_ε is super-smooth in its terminology. Furthermore, by the compactness of the parameter space Θ and as the functions l_θ^* , and for $j, k \in\{1,2\}$, the functions $\left(\frac{\partial l_\theta}{\partial \theta_j}\right)^*\left(\frac{\partial^2 l_\theta}{\partial \theta_j \partial \theta_k}\right)^*$, have the following form $C_1(\theta) P(x) \exp \left(-C_2(\theta) x^2\right)$ where $C_1(\theta)$ and $C_2(\theta)$ are two constants well defined in the parameter space Θ with $C_2(\theta) > 0$, we obtain:

$$\begin{cases} \mathbb{E} \left(\left| Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right|^{2+\delta} \right) < \infty & \text{for some } \delta > 0, \\ \mathbb{E} \left(\sup_{\theta \in \mathcal{U}} \left\| Y_2 u_{\nabla_{\theta} l_{\theta}}^*(Y_1) \right\| \right) < \infty & \text{for some neighbourhood } \mathcal{U} \text{ of } \theta_0. \end{cases}$$

C.2.3 Expression of the covariance matrix

As, the functions $l_{\theta}(x)$ and $\mathbf{P}m_{\theta}$ are the same for the two models, the expressions of the matrix V_{θ_0} and $\Omega_j(\theta_0)$ are given in Lemma 3. We need only to use an estimator of $P_2 = \mathbb{E}[Y_2^2(u_{\nabla l_{\theta}}^*(Y_1))^2]$ since we can just approximate $u_{\nabla l_{\theta}}^*(y)$. A natural and consistent estimator of P_2 is given by:

$$\widehat{P}_2 = \frac{1}{n} \sum_{i=1}^{n-1} \left(Y_{i+1}^2 (u_{\nabla l_{\theta}}^*(Y_i))^2 \right), \tag{39}$$

Remark 7 In some models, the covariance matrix $\Omega_j(\hat{\theta}_n)$ cannot be explicitly computable. We refer the reader to Hayashi (2000) chapter 6 Section 6.6 p.408 for this case.

Appendix D: EM algorithm

We first refer to Dempster et al. (1977) for general details on the EM algorithm. The EM algorithm is an iterative procedure for maximizing the log-likelihood $l(\theta) = \log(f_{\theta}(Y_{1:n}))$. Suppose that after the k^{th} iteration, the estimate for θ is given by θ_k . Since the objective is to maximize $l(\theta)$, we want to compute an updated θ such that:

$$l(\theta) > l(\theta_k)$$

Hidden variables can be introduced for making the ML estimation tractable. Denote the hidden random variables $U_{1:n}$ and a given realization by $u_{1:n}$. The total probability $f_{\theta}(Y_{1:n})$ can be written as:

$$f_{\theta}(Y_{1:n}) = \sum_{u_{1:n}} f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})$$

Hence,

$$\begin{aligned} l(\theta) - l(\theta_k) &= \log(f_{\theta}(Y_{1:n})) - \log(f_{\theta_k}(Y_{1:n})) \\ &= \log \left(\sum_{u_{1:n}} f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n}) \right) - \log(f_{\theta_k}(Y_{1:n})) \\ &= \log \left(\sum_{u_{1:n}} f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n}) \frac{f_{\theta_k}(u_{1:n}|Y_{1:n})}{f_{\theta_k}(u_{1:n}|Y_{1:n})} \right) - \log(f_{\theta_k}(Y_{1:n})) \end{aligned}$$

$$= \log \left(\sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \frac{f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})}{f_{\theta_k}(u_{1:n}|Y_{1:n})} \right) - \log(f_{\theta_k}(Y_{1:n})) \quad (40)$$

$$\geq \sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \log \left(\frac{f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})}{f_{\theta_k}(u_{1:n}|Y_{1:n})} \right) - \log(f_{\theta_k}(Y_{1:n})) \quad (41)$$

$$= \sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \log \left(\frac{f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})}{f_{\theta_k}(u_{1:n}|Y_{1:n})} \right) - \log(f_{\theta_k}(Y_{1:n})) \sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \quad (42)$$

$$= \sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \log \left(\frac{f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})}{f_{\theta_k}(u_{1:n}|Y_{1:n}) f_{\theta_k}(Y_{1:n})} \right) = \Delta(\theta, \theta_k).$$

In going from Eq. (40) to (41) we use the Jensen inequality: $\log \sum_{i=1}^n \lambda_i x_i \geq \sum_{i=1}^n \lambda_i \log(x_i)$ for constants $\lambda_i \geq 0$ with $\sum_{i=1}^n \lambda_i = 1$. And in going from Eq. (41) to (42) we use the fact that $\sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) = 1$. Hence,

$$l(\theta) \geq l(\theta_k) + \Delta(\theta, \theta_k) = \mathcal{L}(\theta, \theta_k) \quad \text{and} \quad \Delta(\theta, \theta_k) = 0 \text{ for } \theta = \theta_k$$

The function $\mathcal{L}(\theta, \theta_k)$ is bounded by the log-likelihood function $l(\theta)$ and they are equal when $\theta = \theta_k$. Consequently, any θ which increases $\mathcal{L}(\theta, \theta_k)$ will increase $l(\theta)$. The EM algorithm selects θ such that $\mathcal{L}(\theta, \theta_k)$ is maximized. We denote this updated value θ_{k+1} . Thus,

$$\begin{aligned} \theta_{k+1} &= \arg \max_{\theta} \left\{ l(\theta_k) + \sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \log \left(\frac{f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})}{f_{\theta_k}(u_{1:n}|Y_{1:n}) f_{\theta_k}(Y_{1:n})} \right) \right\} \\ &= \arg \max_{\theta} \left\{ \sum_{u_{1:n}} f_{\theta_k}(u_{1:n}|Y_{1:n}) \log f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n}) \right\} \\ &\quad \text{if we drop the terms which don't depend on } \theta. \\ &= \arg \max_{\theta} \{ \mathbb{E}[\log f_{\theta}(Y_{1:n}|u_{1:n}) f_{\theta}(u_{1:n})] \} \\ &\quad \text{where the expectation is according to } f_{\theta_k}(u_{1:n}|Y_{1:n}). \end{aligned} \quad (43)$$

D.1 Simulated expectation maximization estimator

Here, we describe the SIEMLE proposed by Kim et al. (1994) for the SV model, these authors retain the linear log-transform model given in (13). However, instead of approximating the log-chi-square distribution of ε_i with a Gaussian distribution, they approximate ε_i by a mixture of seven Gaussian. The distribution of the noise is given by:

$$\begin{aligned}
 f_{\varepsilon_i}(x) &\approx \sum_{j=1}^7 q_j \times g_{(m_j, v_j^2)}(x), \\
 &\approx \sum_{j=1}^7 q_j f_{\varepsilon_i|s_i=j}(x)
 \end{aligned}$$

where $g_{(m, v)}(x)$ denotes the Gaussian distribution of ε_i with mean m and variance v , and $f_{\varepsilon_i|s_i=j}(x)$ is a Gaussian distribution conditional to an indicator variable s_i at time i and the variables $q_j, j = 1 \dots, 7$ are the given weights attached to each component and such that $\sum_{j=1}^7 q_j = 1$. Note that, most importantly, given the indicator variable s_i at each time i , the log-transform model is Gaussian. That is:

$$f_{\theta}(Y_i|s_i = j, X_i) \sim g_{(X_i+m_j, v_j^2)}.$$

Then, conditionally to the indicator variable s_i , the SV model becomes a Gaussian state-space model and the Kalman filter can be used in the SIEMLE in order to compute the log-likelihood function given by:

$$\log f_{\theta}(Y_{1:n}|s_{1:n}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log F_i - \frac{1}{2} \sum_{i=1}^n \frac{v_i^2}{F_i},$$

with $v_i = (Y_i - \hat{Y}_i^- - m_{s_i})$ and $F_i = \mathbb{V}_{\theta}[v_i] = P_i^- + v_{s_i}^2$. The quantities $\hat{Y}_i^- = \mathbb{E}_{\theta}[Y_i|Y_{1:i-1}]$ and $P_i^- = \mathbb{V}_{\theta}[(X_i - \hat{X}_i^-)^2]$ are computed by the Kalman filter.

Hence, if we consider that the missing data $u_{1:n}$ for the EM correspond to the indicator variables $s_{1:n}$, then according to Eq. (43) and since $f(s_{1:n})$ do not depend on θ , the Maximization step is:

$$\theta_{k+1} = \arg \max_{\theta} \{\mathbb{E}[\log f_{\theta}(Y_{1:n}|s_{1:n})]\} = \arg \max_{\theta} Q(\theta, \theta_k)$$

where the expectation is according to $f_{\theta_k}(s_{1:n}|Y_{1:n})$. Nevertheless, for the SV model, the problem with the EM algorithm is that the density $f_{\theta}(s_{1:n}|Y_{1:n})$ is unknown. The main idea consists in introducing a Gibbs algorithm to obtain \tilde{M} draws $s_{1:n}^{(1)}, \dots, s_{1:n}^{(\tilde{M})}$ from the law $f_{\theta}(s_{1:n}|Y_{1:n})$. Hence, the objective function $Q(\theta, \theta_k)$ is approximated by:

$$\tilde{Q}(\theta, \theta_k) = \frac{1}{\tilde{M}} \sum_{l=1}^{\tilde{M}} \log f_{\theta}(Y_{1:n}|s_{1:n}^{(l)})$$

Then, the simulated EM algorithm for the SV model is as follows: Let $C > 0$ be a threshold to stop the algorithm and θ_k a given arbitrary value of the parameter. While $|\theta_k - \theta_{k-1}| > C$,

1. Apply the Gibbs sampler as follows:

The Gibbs Sampler: Choose arbitrary starting values $X_{1:n}^{(0)}$, and let $l = 0$.

- (a) Sample $s_{1:n}^{(l+1)} \sim f_{\theta_k}(s_{1:n}|Y_{1:n}, X_{1:n}^{(l)})$.
 - (b) Sample $X_{1:n}^{(l+1)} \sim f_{\theta_k}(X_{1:n}|Y_{1:n}, s_{1:n}^{(l+1)})$.
 - (c) Set $l = l + 1$ and goto (a).
2. $\theta_{k+1} = \arg \max_{\theta} \hat{Q}(\theta, \theta_k)$.

Step (a): to sample the vector $s_{1:n}$ from its full conditional density, we sample each s_i independently. We have:

$$f_{\theta_k}(s_{1:n}|Y_{1:n}, X_{1:n}) = \prod_{r=1}^n f_{\theta_k}(s_r|Y_r, X_r) \propto \prod_{r=1}^n f_{\theta_k}(Y_r|s_r, X_r) f(s_r),$$

and $f_{\theta_k}(Y_r|s_r = j, X_r) \propto g(X_{r+m_j}, v_j^2)$ for $j = 1 \dots, 7$. And the step (b) of the Gibbs sampler is conducted by the Kalman filter since the model is Gaussian.

References

- Andrieu C, Doucet A, Holenstein R (2010) Particle Markov chain Monte Carlo methods. *J R Stat Soc Ser B Stat Methodol* 72(3):269–342. doi:[10.1111/j.1467-9868.2009.00736.x](https://doi.org/10.1111/j.1467-9868.2009.00736.x)
- Cappé O, Moulines E, Rydén T (2005a) Inference in hidden Markov models. Springer Series in Statistics, Springer, New York, with Randal Douc's contributions to Chapter 9 and Christian P. Robert's to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat
- Cappé O, Moulines E, Rydén T (2005b) Inference in hidden Markov models. Springer Series in Statistics, Springer, New York, with Randal Douc's contributions to Chapter 9 and Christian P. Robert's to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat
- Chanda KC (1995) Large sample analysis of autoregressive moving-average models with errors in variables. *J Time Ser Anal* 16(1):1–15. doi:[10.1111/j.1467-9892.1995.tb00220.x](https://doi.org/10.1111/j.1467-9892.1995.tb00220.x)
- Chopin N (2002) A sequential particle filter method for static models. *Biometrika* 89(3):539–551. doi:[10.1093/biomet/89.3.539](https://doi.org/10.1093/biomet/89.3.539)
- Chopin N, Jacob PE, Papaspiliopoulos O (2011) *smc²*: an efficient algorithm for sequential analysis of state-space models. Preprint in arxiv.org/abs/11011528
- Comte F, Lacour C (2011) Data driven density estimation in presence of unknown convolution operator. *J R Stat Soc Ser B Stat Methodol* 73(4):601–627. doi:[10.1111/j.1467-9868.2011.00775.x](https://doi.org/10.1111/j.1467-9868.2011.00775.x), URL:<http://dx.doi.org/10.1111/j.1467-9868.2011.00775.x>
- Comte F, Taupin ML (2001) Semiparametric estimation in the (auto)-regressive β -mixing model with errors-in-variables. *Math Methods Stat* 10(2):121–160
- Comte F, Lacour C, Rozenholc Y (2010) Adaptive estimation of the dynamics of a discrete time stochastic volatility model. *J Econometr* 154(1):59–73. doi:[10.1016/j.jeconom.2009.07.001](https://doi.org/10.1016/j.jeconom.2009.07.001)
- Dedecker J, Doukhan P, Lang G, León RJR, Louhichi S, Prieur C (2007) Weak dependence: with examples and applications, vol 190, Lecture notes in statistics. Springer, New York
- Dedecker J, Samson A, Taupin ML (2011) Estimation in autoregressive model with measurement noise. Preprint: hal-00591114
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Statist Soc Ser B* 39(1):1–38 (with discussion)
- Douc R, Moulines E, Olsson J, van Handel R (2011) Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann Statist* 39(1):474–513. doi:[10.1214/10-AOS834](https://doi.org/10.1214/10-AOS834)
- Doucet A, de Freitas N, Gordon N (eds) (2001) Sequential Monte Carlo methods in practice. Statistics for engineering and information science. Springer, New York
- Fan J (1991) Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser A* 53(1):97–110
- Genon-Catalot V, Jeantheau T, Larédo C (2000) Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* 6(6):1051–1079. doi:[10.2307/3318471](https://doi.org/10.2307/3318471)

- Ghysels E, Harvey AC, Renault E (1996) Stochastic volatility. In: Statistical methods in finance, handbook of statistics, vol 14. North-Holland, Amsterdam, pp 119–191. doi:[10.1016/S0169-7161\(96\)14007-4](https://doi.org/10.1016/S0169-7161(96)14007-4), URL:[http://dx.doi.org/10.1016/S0169-7161\(96\)14007-4](http://dx.doi.org/10.1016/S0169-7161(96)14007-4)
- Hansen BE, Horowitz JL (1997) Handbook of econometrics, vol. 4 robert f. engle and daniel l. mcFadden, editors elsevier science b. v., 1994. *Econometr Theory* 13(01):119–132, URL http://ideas.repec.org/a/cup/etheor/v13y1997i01p119-132_00.html
- Harvey A, Ruiz E, Shephard N (1994) Multivariate stochastic variance models. *Rev Econ Stud* 61(2):247–264
- Hayashi F (2000) *Econometrics*. Princeton University Press, Princeton
- Ionides EL, Bhadra A, Atchadé Y, King A (2011) Iterated filtering. *Ann Statist* 39(3):1776–1802. doi:[10.1214/11-AOS886](https://doi.org/10.1214/11-AOS886)
- Jacquier E, Polson NG, Rossi PE (2002) Bayesian analysis of stochastic volatility models. *J Bus Econom Statist* 20(1):69–87. doi:[10.1198/073500102753410408](https://doi.org/10.1198/073500102753410408), twentieth anniversary commemorative issue
- Johannes MS, Polson NG, Stroud JR (2009) Optimal filtering of jump diffusions: extracting latent states from asset prices. *Rev Financial Stud* 22(7):2559–2599
- Johansen AM, Doucet A, Davy M (2008) Particle methods for maximum likelihood estimation in latent variable models. *Stat Comput* 18(1):47–57. doi:[10.1007/s11222-007-9037-8](https://doi.org/10.1007/s11222-007-9037-8)
- Jones GL (2004) On the Markov chain central limit theorem. *Probab Surv* 1:299–320. doi:[10.1214/154957804100000051](https://doi.org/10.1214/154957804100000051)
- Kalman RE (1960) A new approach to linear filtering and prediction problems URL <http://www.cs.unc.edu/welch/kalman/media/pdf/Kalman1960.pdf>
- Kim S, Shephard N (1994) Stochastic volatility: likelihood inference and comparison with arch models. *Economics papers* 3, Economics Group, Nuffield College, University of Oxford, URL <http://ideas.repec.org/p/nuf/econwp/0003.html>
- Liu J, West M (2001) Combined parameter and state estimation in simulation-based filtering. in arnaud doucet, nando de freitas, and neil gordon, editors. *Sequential Monte Carlo Methods in Practice*
- Mathieu RJ, Schotman PC (1998) An empirical application of stochastic volatility models. *Rev Econ Stud* 13(4):333–360
- Melino A, Turnbull SM (1990) Pricing foreign currency options with stochastic volatility. *J Econometr* 45(1–2):239–265 URL <http://ideas.repec.org/a/eee/econom/v45y1990i1-2p239-265.html>
- Newey WK (1987) Advanced econometrics by takeshi amemiya, harvard university press, 1986. *Econometr Theory* 3(01):153–158 URL http://ideas.repec.org/a/cup/etheor/v3y1987i01p153-158_00.html
- Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. In: *Handbook of econometrics*, vol IV, *Handbooks in Econometrics*, vol 2, North-Holland, Amsterdam, pp 2111–2245
- Peters GW, Hosack GR, Hayes KR (2010) Ecological non-linear state space model selection via adaptive particle markov chain monte carlo (adpmcmc). Preprint:arXiv-10052238v1
- Pitt MK, Shephard N (1999) Filtering via simulation: auxiliary particle filters. *J Am Statist Assoc* 94(446):590–599. doi:[10.2307/2670179](https://doi.org/10.2307/2670179)
- Poyiadjis G, Doucet A, Singh SS (2011) Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika* 98(1):65–80 URL <http://ideas.repec.org/a/oup/biomet/v98y2011i1p65-80.html>
- Robert CP, Rydén T, Titterton DM (2000) Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J R Stat Soc Ser B Stat Methodol* 62(1):57–75. doi:[10.1111/1467-9868.00219](https://doi.org/10.1111/1467-9868.00219)
- Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol* 71(2):319–392. doi:[10.1111/j.1467-9868.2008.00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x)
- Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J Roy Statist Soc Ser B* 55(1):3–23. URL [http://links.jstor.org/sici?sici=0035-9246\(1993\)55:1<3:BCVTGS>2.0.CO;2-&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1993)55:1<3:BCVTGS>2.0.CO;2-&origin=MSN)
- Taylor S (2005) *Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices* vol 1, Oxford University Press, pp 203–226. URL <http://eprints.lancs.ac.uk/29900/>
- Van der Vaart AW (1998) *Asymptotic statistics*, Cambridge series in statistical and probabilistic mathematics, vol 3. Cambridge University Press, Cambridge