



**HAL**  
open science

# Optimal bandwidth criteria for nonparametric trend estimation under stochastic volatility error processes

Karim Benhenni, Didier A. Girard, Sana Louhichi

► **To cite this version:**

Karim Benhenni, Didier A. Girard, Sana Louhichi. Optimal bandwidth criteria for nonparametric trend estimation under stochastic volatility error processes. 2020. hal-02889802v1

**HAL Id: hal-02889802**

**<https://hal.science/hal-02889802v1>**

Preprint submitted on 5 Jul 2020 (v1), last revised 31 Aug 2021 (v2)

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

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

# Optimal bandwidth criteria for nonparametric trend estimation under stochastic volatility error processes

Karim Benhenni\*, Didier A. Girard†, Sana Louhichi‡

July 5, 2020

## Abstract

This paper is concerned with the optimal selection of the smoothing parameter  $h$  in kernel estimation of a trend in nonparametric regression models with (dependent) stochastic volatility errors  $\epsilon_i = \sigma_i Z_i$ ,  $i = 1, \dots, n$ , where  $(\sigma_i)_i$  is referred as the volatility sequences and  $(Z_i)_i$  a sequence of i.i.d random variables. We consider three types of volatility sequences; the log-normal volatility, the Gamma volatility and the log-linear volatility with Bernoulli innovations. In fact, based on three criteria for deriving optimal smoothing parameters, namely the average squared error, the mean average squared error and an adjusted Mallows-type criterion to the dependent case, we show that these three minimizers are first-order equivalent in probability. Moreover, we derive the normal asymptotic distribution of the difference between the minimizer of the average squared error and the minimizer based on the Mallows-type criterion. A Monte-Carlo simulation is conducted for a log-normal stochastic volatility model.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Model, selection criteria and previous results</b>	<b>3</b>
<b>3</b>	<b>Stochastic volatility models</b>	<b>7</b>
3.1	Log-normal SV model	8
3.2	Gamma Stochastic Volatility Models	9
3.3	Other SV model with no mixing properties	11

---

\*Laboratoire Jean Kuntzmann, Université Grenoble Alpes, 700 Avenue Centrale, 38401 Saint-Martin-d'Hères, France. E.mail: karim.benhenni@univ-grenoble-alpes.fr

†CNRS, Laboratoire Jean Kuntzmann. E.mail: didier.girard@univ-grenoble-alpes.fr

‡Corresponding author. Laboratoire Jean Kuntzmann, Université Grenoble Alpes, 700 Avenue Centrale, 38401 Saint-Martin-d'Hères, France. E.mail: sana.louhichi@univ-grenoble-alpes.fr.

<b>4</b>	<b>A Monte-carlo simulation study for a “trend plus a log-normal SV process”</b>	<b>12</b>
<b>5</b>	<b>References</b>	<b>18</b>

# 1 Introduction

Nonparametric trend estimation is a very popular field of research in Statistics and is used in different domain of applications. There are several nonparametric estimate of the trend in time series models or the mean function of stochastic processes. Most of these estimate are constructed from a kernel function that depends on a smoothing parameter  $h$  known also as a bandwidth. The choice of this parameter is crucial since it has an important impact on the performance of the kernel estimate. Some criteria are available for choosing this parameter but there are mostly based on models with independent errors; the plug-in method, see for instance [Ruppert, Sheather and Wand \(1995\)](#) and [Fan, Gijbels, Hu and Huang \(1996\)](#), the cross-validation (CV) and the generalized cross-validation (GCV), see for instance [Rice \(1984\)](#), [Härdle, Hall and Marron \(1988\)](#) and [Girard \(1998\)](#) among others. However, in the case of dependent (correlated) errors, there are a very limited available results in the literature concerning the selection methods of the smoothing parameter, see for instance the review by [Opsomer, Wang and Yang \(2001\)](#). [Hall, Lahiri and Polzehl \(1995\)](#) develop bootstrap and cross-validation methods to select the smoothing parameter under short and long range dependance.

[Benhenni, Girard and Louhichi \(2020\)](#) considered dependent stationary martingale difference errors with an application to ARCH(1) errors. Based on three criteria for deriving optimal smoothing parameters, namely the average squared error, the mean average squared error and the Mallows-type criterion adapted to the dependent case, they showed that these three minimizers are first-order equivalent in probability. Moreover, they give a normal asymptotic behavior of the difference between the minimizer of the average squared error and that of the Mallows-type criterion.

In this paper, we extend these results to other types of strictly stationary dependent errors, namely the time series volatility models.

Volatility has been one of the most active areas of research in time series econometrics and economic forecasting. It may be modeled as an unobserved component following some latent stochastic process, such as auto-regression. The resulting models are called stochastic volatility (SV) models and have been the focus of considerable attention, see for instance [Taylor \(1994\)](#), [Ghysels, Harvey and Renault \(1996\)](#), [Shephard \(1996\)](#) and [Billo and Sartore \(2005\)](#).

More precisely, we concentrate on three families of stochastic volatility models (SV): the log-normal volatility, the Gamma volatility and the log-linear volatility with Bernoulli innovations. The first two SV models enjoy a strong mixing condition with a decreasing power bound whereas the Bernoulli SV models do not satisfy any mixing property.

Likewise, we establish the first-order equivalent in probability of the three minimizers and derive the asymptotic normal distribution of the difference between the minimizer of the average squared error and the minimizer based on the adjusted Mallows-type criterion to (weakly) dependent errors which can be mixing or not with a decreasing power decay of the corresponding (dependence) coefficients.

It should be noticed that [Hall, Lahiri and Polzehl \(1995\)](#) obtained the first order equivalence of the CV criteria instead of the Mallows one for stationary error process with all finite moments but under the Rosenblatt mixing condition with a decreasing exponential decay.

The paper is organized as follows. Section 2 introduces the nonparametric model and defines the different criteria for the selection of the smoothing parameter  $h$ . Section 3 is devoted to the main results when SV models are considered and in particular for some families of SV models: the log-normal SV, the Gamma SV and the log-linear SV with Bernoulli innovations. Finally, a Monte Carlo simulation study is conducted for the log-normal SV error process in order to confirm our theoretical results.

## 2 Model, selection criteria and previous results

In this section, we follow the same notations as in [Benhenni, Girard and Louhichi \(2020\)](#).

Let  $(\epsilon_i)_{i \geq 0}$  be a stationary sequence of centered random variables with finite second moment. Let  $\sigma^2 = \text{Var}(\epsilon_1)$  and  $R$  be the correlation matrix of the vector  $(\epsilon_1, \dots, \epsilon_n)^t$ . Consider the following regression model, defined for  $i = 1, \dots, n$ , by

$$Y_i = r(x_i) + \epsilon_i, \quad x_i = \frac{i}{n}, \quad (1)$$

where  $r$  is an unknown regression function with second order continuous derivative and the  $x_i$ 's are equally spaced fixed design.

We are interested in this paper by the Priestley-Chao estimator of  $r$  defined, for  $x \in \mathbb{R}$ , by

$$\hat{r}(x) = \sum_{i=1}^n l_i(x) Y_i, \quad \text{with} \quad l_i(x) = \frac{1}{nh} K \left( \frac{x - x_i}{h} \right),$$

where  $K$  is a compactly supported even kernel with class  $\mathcal{C}^1([-1, 1])$  and  $h$  is a positive bandwidth less than  $1/2$ .

Other types of Kernel estimate such as the Gasser-Müller, see for instance [Benelmadani, Benhenni and Louhichi \(2019\)](#) or local polynomial, see [Fan, Gijbels, Hu and Huang \(1996\)](#), could also be considered and the optimal selection procedures of the smoothing parameter raised in this section are of the same nature.

The above curve estimator entails the following smoothing, in the matrix form,

$$\hat{r} = LY,$$

with

$$\hat{r} = (\hat{r}(x_1), \dots, \hat{r}(x_n))^t, \quad Y = (Y_1, \dots, Y_n)^t$$

and  $L = (l_j(x_i))_{1 \leq i, j \leq n}$  is known as the smoothing matrix or the hat matrix. Since the estimator  $\hat{r}$  depends on some smoothing parameter  $h$ , we will need some procedure for choosing  $h$ . For this, we recall some known criteria of selecting this parameter  $h$ .

In order to eliminate the boundary effects of the compactly supported kernel  $K$ , we introduce a known function supported on a subinterval of the unit interval. For this, suppose without loss of generality that  $h < \epsilon$  where  $\epsilon$  is a fixed positive real number less than  $1/2$ . Let  $u := u_\epsilon$  be a positive function, of class  $\mathcal{C}^1$  and  $[\epsilon, 1 - \epsilon]$ -compactly supported.

Define the average squared error

$$T_n(h) = \frac{1}{n} \sum_{i=1}^n u(x_i) (\hat{r}(x_i) - r(x_i))^2 = \frac{1}{n} \|U^{1/2}(\hat{r} - r)\|^2,$$

where  $U$  is the diagonal matrix  $U = \text{diag}(u(x_1), \dots, u(x_n))$ . The following lemma from [Benhenni, Girard and Louhichi \(2020\)](#) evaluates its mean,  $\mathbb{E}(T_n(h))$ , for finite variance of stationary errors  $(\epsilon_i)_{i \in \mathbb{N}}$ .

**Lemma 1.** *Suppose that  $\sum_{k=1}^{\infty} k |\text{Cov}(\epsilon_0, \epsilon_k)| < \infty$ . Define,*

$$\begin{aligned} D_n(h) &= \frac{h^4}{4} \int_0^1 u(x) r''^2(x) dx \left( \int_{-1}^1 t^2 K(t) dt \right)^2 \\ &+ \frac{1}{nh} \left( \int_0^1 u(x) dx \right) \int_{-1}^1 K^2(y) dy \left( \sigma^2 + 2 \sum_{k=1}^{\infty} \text{Cov}(\epsilon_0, \epsilon_k) \right). \end{aligned}$$

Then for any  $n \geq 1$  and  $h \in ]0, \epsilon[$ ,

$$\mathbb{E}(T_n(h)) = D_n(h) + O\left(\frac{1}{n}\right) + o(h^4) + O\left(\frac{1}{n^2 h^4}\right) + \frac{\gamma(h)}{nh},$$

where  $O$  is uniformly on  $n$  and  $h$ ,  $\gamma(h)$  depends on  $h$  (but not on  $n$ ) and tends to 0 when  $h$  tends to 0.

Let  $h_n^* = \text{argmin}_{h>0} D_n(h)$ . Clearly, if  $\int_0^1 u(x) r''^2(x) dx \neq 0$  then

$$h_n^* = n^{-1/5} \left( \frac{\left( \int_0^1 u(x) dx \right) \int_{-1}^1 K^2(y) dy \left( \sigma^2 + 2 \sum_{k=1}^{\infty} \text{Cov}(\epsilon_0, \epsilon_k) \right)}{\int_0^1 u(x) r''^2(x) dx \left( \int_{-1}^1 t^2 K(t) dt \right)^2} \right)^{1/5} =: cn^{-1/5}.$$

Let  $H_n$  be a neighborhood of  $h_n^*$ , i.e.,  $H_n = [an^{-1/5}, bn^{-1/5}]$  for some fixed  $a < c < b$ . Define also,

$$h_n = \text{argmin}_{h \in H_n} \mathbb{E}(T_n(h)) \quad \text{and} \quad \hat{h}_n = \text{argmin}_{h \in H_n} T_n(h).$$

Of course these three “optimal” parameters  $h_n$ ,  $h_n^*$  and  $\hat{h}_n$  depend on the unknown function  $r$ , since the criteria that they respectively minimise, depend themselves on the regression function  $r$ . Many authors agree that, among these ones,  $\hat{h}_n$  should be the target (see [Girard \(1998\)](#), page 316). For this reason, an important literature considered minimizers of “good” estimators of  $T_n(h)$  and studied their asymptotic behavior.

A nearly unbiased estimate of  $\mathbb{E}(T_n(h))$  is constructed allowing to define a criterion that selects an observable choice for  $h$ . An early approach was the well known Cross-Validation criterion which was next transformed to various Generalized Cross-Validation (GCV) criteria. All those different forms of the GCV criteria are second order equivalent in the sense that they are asymptotically close in distribution to the following Mallows criterion:

$$C_p := C_p(h) = \frac{1}{\sum_{i=1}^n u(x_i)} \sum_{i=1}^n u(x_i)(Y_i - \hat{r}(x_i))^2 + 2\frac{\nu}{n}\hat{\sigma}^2, \quad (2)$$

where,

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^n u(x_i)} \sum_{i=1}^n u(x_i)(Y_i - \hat{r}(x_i))^2, \quad \nu := n \frac{\text{tr}(UL)}{\text{tr}(U)} = \frac{1}{h}K(0).$$

Let  $\hat{h}$  be a minimizer over  $h \in H_n$  of the function  $C_p$ . [Härdle, Hall and Marron \(1988\)](#) proved, in the context of i.i.d errors  $(\epsilon_i)_{1 \leq i \leq n}$  with all finite moments, that  $\hat{h}, h_n^*, \hat{h}_n, h_n$  are all equivalent in probability and that  $\hat{h} - h_n, h_n - \hat{h}_n$  are also close in distribution as  $n$  tends to infinity.

[Benhenni, Girard and Louhichi \(2020\)](#) extended these appealing theoretical properties of the Mallows criteria (2) to stationary martingale difference sequences of dependent errors with known covariance matrix  $\sigma^2 R$  of the vector  $(\epsilon_1, \dots, \epsilon_n)^t$  (of course  $R = I$  in the case “stationary martingale differences”), by considering the well known “corrected  $C_L$  criterion” (see [Liu \(2001\)](#), [Francisco-Fernandez and Opsomer \(2005\)](#) and [Meilan-Vila, Fernandez-Casal, Crujeiras and Francisco-Fernandez \(2020\)](#) for recent applications)

$$\text{CL}(h) = n^{-1} \|U^{1/2}(I - L)Y\|^2 + 2\sigma^2 n^{-1} \text{tr}(URL). \quad (3)$$

Let us define  $\hat{h}_M$  to be the minimizer of the dependent version of the Mallows criteria (3),

$$\hat{h}_M = \text{argmin}_{h \in H_n} \text{CL}(h).$$

We recall two assumptions that are required to establish the main results in [Benhenni, Girard and Louhichi \(2020\)](#).

**Assumptions (A).** Suppose that both the functions  $h \mapsto T_n(h)$  and  $h \mapsto \text{CL}(h)$  have continuous first derivatives, that  $T'_n(\hat{h}_n) = 0$  and  $\text{CL}'(\hat{h}_M) = 0$  almost surely. Suppose also that the function  $h \mapsto \mathbb{E}(T_n(h))$  is twice differentiable with continuous second derivative and that  $\frac{\partial^2}{\partial h^2} \mathbb{E}(T_n(h)) = \mathbb{E}(T''_n(h))$ .

**Assumptions (B).** Assume that the errors  $(\epsilon_i)_{i \geq 0}$  form a stationary martingale difference sequence (MDS, in short) with respect to some natural filtration  $(\mathcal{F}_i)_{i \geq 1}$ , i.e, for any  $i > 0$ ,  $\epsilon_i$  is  $\mathcal{F}_i$ -measurable and  $\mathbb{E}(\epsilon_i | \mathcal{F}_{i-1}) = 0$ . Suppose also that  $\mathbb{E}(\epsilon_1^{2p}) < \infty$  for some  $p > 8$ .

The first result of [Benhenni, Girard and Louhichi \(2020\)](#) states that for MDS errors, the bandwidths  $h_n, h_n^*, \hat{h}_n$  and  $\hat{h}_M$  are first-order equivalent in probability (and the CL enjoys the "asymptotic optimality" property).

**Proposition 1.** *Suppose that Assumptions (A) and (B) are satisfied. Then*

$$\frac{h_n^*}{h_n}, \frac{\hat{h}_n}{h_n}, \frac{\hat{h}_M}{h_n}$$

*all converge in probability to 1 as  $n$  tends to infinity.*

The second result in [Benhenni, Girard and Louhichi \(2020\)](#) gives, under an additional dependence condition, the rate at which  $\hat{h}_n - \hat{h}_M$  converges in distribution to a centered normal law, and furthermore states that the martingale difference dependence doesn't impact  $\hat{h}_n - \hat{h}_M$  up to second-order.

**Theorem 1.** *Suppose that Assumptions (A) and (B) are satisfied. Moreover, suppose that there exists a positive decreasing function  $\Phi$  defined on  $\mathbb{R}^+$  satisfying*

$$\sum_{s=1}^{\infty} s^4 \Phi(s) < \infty,$$

*and for any positive integer  $q \leq 6$ ,  $1 \leq i_1 \leq \dots \leq i_k < i_{k+1} \leq \dots \leq i_q \leq n$  such that  $i_{k+1} - i_k \geq \max_{1 \leq l \leq q-1} (i_{l+1} - i_l)$ ,*

$$|\text{Cov}(\epsilon_{i_1} \dots \epsilon_{i_k}, \epsilon_{i_{k+1}} \dots \epsilon_{i_q})| \leq \Phi(i_{k+1} - i_k). \quad (4)$$

*Then*

$$n^{3/10}(\hat{h}_n - \hat{h}_M)$$

*converges in distribution to a centered normal law with variance  $\Sigma^2$  given by*

$$\begin{aligned} \Sigma^2 &= \frac{4\sigma^{6/5}}{5^2 A^{8/5} B^{2/5}} \left( \int t^2 K(t) dt \right)^2 \int_0^1 u^2(x) r''^2(x) dx \\ &+ \frac{8\sigma^{6/5}}{5^2 A^{3/5} B^{7/5}} \int_0^1 u^2(x) dx \int (K - G)^2(u) du, \end{aligned}$$

*where  $\sigma^2 = \mathbb{E}(\epsilon_1^2)$ ,  $G$  is the function defined for any  $x \in \mathbb{R}$  by  $G(x) = -xK'(x)$  and*

$$A = \int_0^1 u(x) r''^2(x) dx \left( \int t^2 K(t) dt \right)^2, \quad B = \int_0^1 u(x) dx \int K^2(t) dt.$$

### 3 Stochastic volatility models

A Stochastic Volatility process  $(\epsilon_i)_{i \in \mathbb{N}}$ , SV in short, is defined as

$$\epsilon_i = \sigma_i Z_i, \quad i \in \mathbb{N}, \quad (5)$$

where the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is a strictly stationary sequence of positive random variables which is independent of the i.i.d. centered noise sequence  $(Z_i)_{i \in \mathbb{N}}$ . We refer, for instance, to [Davis and Mikosh \(2009\)](#) for the main properties of SV models.

The following proposition gives conditions under which the requirements of [Proposition 1](#) and [Theorem 1](#) are satisfied for SV error process.

**Proposition 2.** *Let  $(\epsilon_i)_{i \in \mathbb{N}}$  be as defined in (5). Suppose that there exists a sequence of filtration  $(\mathcal{F}_i)_{i \geq 0}$  such that for any  $i \in \mathbb{N}$ ,  $\sigma(Z_0, \dots, Z_i) \subset \mathcal{F}_i$ ,  $\sigma_i$  is  $\mathcal{F}_{i-1}$ -measurable and that almost surely  $\mathbb{E}(Z_i | \mathcal{F}_{i-1}) = 0$ . Suppose moreover that, there exists a positive decreasing function  $\tilde{\Phi}$  defined on  $\mathbb{R}^+$  satisfying*

$$\sum_{s=1}^{\infty} s^4 \tilde{\Phi}(s) < \infty,$$

and for any positive integer  $q \leq 6$ ,  $1 \leq i_1 \leq \dots \leq i_k < i_{k+1} \leq \dots \leq i_q \leq n$  such that  $i_{k+1} - i_k \geq \max_{1 \leq l \leq q-1} (i_{l+1} - i_l)$ ,

$$|\text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q})| \leq \tilde{\Phi}(i_{k+1} - i_k). \quad (6)$$

Suppose that Assumptions (A) are satisfied. If  $\mathbb{E}(\sigma_1^{2p}) < \infty$  and  $\mathbb{E}(Z_1^{2p}) < \infty$  for some  $p > 8$  then the conclusions of [Proposition 1](#) and [Theorem 1](#) hold.

**Proof of Proposition 2.** The sequence  $(\epsilon_i)_{i \in \mathbb{N}}$ , as defined in (5), is strictly stationary. This property is an immediate consequence of the independence of the strictly stationary sequence  $(\sigma_i)_{i \in \mathbb{N}}$  and that of the i.i.d. sequence  $(Z_i)_{i \in \mathbb{N}}$ .

The  $(\epsilon_i)_{i \in \mathbb{N}}$  is MDS since,  $\epsilon_i$  is  $\mathcal{F}_i$ -measurable and  $\mathbb{E}(\epsilon_i | \mathcal{F}_{i-1}) = \sigma_i \mathbb{E}(Z_i | \mathcal{F}_{i-1}) = 0$ , almost surely. We deduce from  $\mathbb{E}(\epsilon_1^{2p}) = \mathbb{E}(Z_1^{2p}) \mathbb{E}(\sigma_1^{2p})$  and the assumptions of [Proposition 2](#) that  $\mathbb{E}(\epsilon_1^{2p}) < \infty$  for some  $p > 8$ . Our task now, is to check Condition (4) of [Theorem 1](#). We have, by definition of SV model, (denoting by  $P$  the distribution of  $Z_1$ ),

$$\begin{aligned} & \text{Cov}(\epsilon_{i_1} \cdots \epsilon_{i_k}, \epsilon_{i_{k+1}} \cdots \epsilon_{i_q}) \\ &= \text{Cov}(\sigma_{i_1} Z_{i_1} \cdots \sigma_{i_k} Z_{i_k}, \sigma_{i_{k+1}} Z_{i_{k+1}} \cdots \sigma_{i_q} Z_{i_q}) \\ &= \int \int z_{i_1} \cdots z_{i_q} \text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q}) dP(z_{i_1}) \cdots dP(z_{i_q}) \\ &= \mathbb{E}(Z_{i_1} \cdots Z_{i_q}) \text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q}). \end{aligned}$$

Consequently,

$$|\text{Cov}(\epsilon_{i_1} \cdots \epsilon_{i_k}, \epsilon_{i_{k+1}} \cdots \epsilon_{i_q})| = |\mathbb{E}(Z_{i_1} \cdots Z_{i_q})| |\text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q})|. \quad (7)$$



Condition (4) is then satisfied from the last equality together with Condition (6). The proof of Proposition 2 is complete since all the requirements of Proposition 1 and Theorem 1 are satisfied.  $\square$

### 3.1 Log-normal SV model

The following corollary gives a log-normal type of SV models satisfying the assumptions of Proposition 2 and thus the conclusions of Proposition 1 and Theorem 1 are true. The log-normal SV models are due to Taylor (1986). For these models, the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is an exponential weight of a Gaussian moving average. They are a basic alternative to ARCH-type processes, since unlike ARCH-type models, their variances always remain positive without the need of further conditions.

Corollary 1 below, proves in particular that for log-normal SV models the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is a stationary strong mixing sequence in the sense of Rosenblatt (1956):  $(\sigma_i)_{i \in \mathbb{N}}$  is a strongly mixing sequence if, its strong mixing coefficient  $\alpha_s$  defined by,

$$\alpha_s = \sup_{k \in \mathbb{N}} \alpha(\sigma(\sigma_i, i \leq k), (\sigma_i, i \geq k + s)),$$

tends to 0 as  $s$  tends to infinity, where for two sigma-fields  $\mathcal{A}$  and  $\mathcal{B}$ ,

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\text{Cov}(\mathbb{1}_A, \mathbb{1}_B)|.$$

**Corollary 1.** *Suppose that the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is defined for  $i \in \mathbb{N}$ , by  $\sigma_i = \exp\left(\sum_{j=0}^{\infty} \gamma^j \eta_{i-j}\right)$  with  $|\gamma| < 1$  and  $(\eta_i)_{i \in \mathbb{N}}$  is an i.i.d. centered sequence distributed as a Gaussian law with finite variance. Suppose also that  $Z_1$  follows a standard Gaussian law. Then the process  $(\epsilon_i)_{i \in \mathbb{N}}$  is stationary MDS, strongly mixing with  $\alpha_s = O(|\gamma|^{\frac{2}{3}s})$ , with finite all integer moments. If Assumptions (A) are satisfied then the conclusions of Proposition 1 and Theorem 1 hold.*

**Proof of Corollary 1.** The strict stationary property of  $(\epsilon_i)_i$  follows from that of  $(\sigma_i)_i$ . Now, let  $\mathcal{F}_i = \sigma(Z_0, \dots, Z_i, (\eta_l)_{l \leq i+1})$ . Then  $\sigma_i$  is  $\mathcal{F}_{i-1}$ -measurable. We have also, by independence of  $Z_i$  and  $\mathcal{F}_{i-1}$  that  $\mathbb{E}(Z_i | \mathcal{F}_{i-1}) = 0$ , a.s. The process  $(\epsilon_i)_{i \in \mathbb{N}}$  is then stationary and MDS. For  $p \in \mathbb{N}^*$ , all the  $2p$ -moment of  $\epsilon_1$  exist and

$$\mathbb{E}(\epsilon_1^{2p}) = \frac{(2p)!}{2^p p!} \exp\left((2p)^2 \frac{\text{Var}(\eta_1)}{8(1-\gamma^2)}\right),$$

we refer for instance to Cox, Hinkley and Barndorff-Nielsen (1996) (page 22). We have now to check that  $(\epsilon_i)_{i \in \mathbb{N}}$  is strongly mixing and that Condition (6) is satisfied. Since the density of  $\eta_0$  is in  $\mathbb{L}^1$ , the linear process  $\left(\sum_{j=0}^{\infty} \gamma^j \eta_{i-j}\right)_i$  is strongly mixing (see Pham and Tran (1985)) with,

$$\alpha_s \leq K \sum_{j \geq s} \left(\sum_{k \geq j} |\gamma|^k\right)^{2/3} = O(|\gamma|^{\frac{2}{3}s}),$$

for some constant  $K$ . Similarly, the sequence  $(\sigma_i)_{i \geq 0}$  is still strongly mixing with the same mixing coefficients  $(\alpha_s)_s$ . We deduce, from (7), that  $(\epsilon_i)_{i \geq 0}$  is also strongly mixing with mixing coefficient of order  $O(|\gamma|^{\frac{2}{3}s})$ . For  $1 \leq i_1 \leq \dots \leq i_k < i_{k+1} \leq \dots \leq i_q \leq n$  such that  $j := i_{k+1} - i_k \geq \max_{1 \leq l \leq q-1} (i_{l+1} - i_l)$ , for  $s, l, r$  strictly positive reals for which  $1/s + 1/l + 1/r = 1$ ,

$$|\text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q})| \leq 8\alpha_j^{1/r} \|\sigma_{i_1} \cdots \sigma_{i_k}\|_s \|\sigma_{i_{k+1}} \cdots \sigma_{i_q}\|_l$$

(see Davydov (1968)). Condition (6) is then satisfied since  $\sum_{j \geq 1} j^4 |\gamma|^{\frac{2j}{3r}} < \infty$ .

Therefore the requirements of Proposition 2 are satisfied so that the conclusions of Proposition 1 and Theorem 1 hold.  $\square$

### 3.2 Gamma Stochastic Volatility Models

We consider here the Gamma stochastic model which is defined, for  $i \in \mathbb{N}$ , by

$$\epsilon_i = \sigma_i Z_i, \tag{8}$$

where  $(Z_i)_{i \geq 1}$  is a sequence of i.i.d. standard normal random variables and for  $i \in \mathbb{N}$ ,  $\sigma_i = \sqrt{h_i}$  where  $(h_i)_{i \in \mathbb{N}}$  is a positive time-homogeneous strictly stationary Markov chain. We suppose that the marginal distribution of  $(h_i)_{i \in \mathbb{N}}$  is a Gamma  $\Gamma(p, \lambda)$  distribution, i.e., noting by  $\pi$  the invariant measure of this chain,

$$\pi(dx) = f(x)dx, \quad f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbb{1}_{x \geq 0}, \quad p, \lambda > 0,$$

$\Gamma(p)$  is the normalizing constant defined by  $\Gamma(p) = \int_0^\infty u^{p-1} e^{-u} du$ . Suppose that this Markov chain is geometrically ergodic, in the sense that there exists a positive constant  $c$  and a Borel positive function  $a(\cdot)$  such that the following holds for any  $\pi$ -a.e.  $x \in \mathbb{R}$ : for any  $n \in \mathbb{N}^*$ , and Borel set  $B$

$$|P^n(x, B) - \pi(B)| \leq a(x)e^{-cn},$$

recall that the transition probability  $P$  is defined, for suitable set  $A$  and  $x$ , by

$$P(x, A) = \mathbb{P}(h_n \in A | h_{n-1} = x)$$

and for  $n \in \mathbb{N}$

$$P^n(x, A) = \mathbb{P}(h_n \in A | h_0 = x).$$

In this case, it is well known that the Markov chain  $(h_i)_{i \in \mathbb{N}}$  is  $\beta$ -mixing (absolutely regular) with geometrically decaying mixing coefficients  $(\beta_n)_{n \geq 1}$  (cf. for instance Theorem 3.7 in Bradley (2005) and the references therein). Recall that for a sequence  $(X_n)_n$  the  $\beta$ -mixing coefficients  $(\beta_n)_{n \geq 1}$  are defined by (see Doukhan (1994) (Sec 1.1))

$$\beta_n = \sup_{m \in \mathbb{N}} \mathbb{E} \left( \sup_{B \in \sigma(X_i, i \geq m+n)} |\mathbb{P}(B | \sigma(X_0, \dots, X_m)) - \mathbb{P}(B)| \right).$$

The Gamma AR(1) process, stated in the example below, is a Markov chain satisfying all the above assumptions.

**Example.** Let  $h_0$  be distributed as  $\Gamma(p, \lambda)$  distribution. Define, for  $\rho \in ]0, 1[$ ,  $h_n$  recursively by,

$$h_n = \rho h_{n-1} + \xi_n,$$

where  $(\xi_n)_n$  is an i.i.d. sequence of random variables with characteristic function  $\mathbb{E}(e^{-it\xi_1}) = \left(\frac{\lambda-it}{\lambda-it\rho}\right)^{-p}$ . The process  $(h_n)_{n \geq 1}$  is then a stationary Markov chain with Gamma  $\Gamma(p, \lambda)$  univariate marginal distribution (see for instance [Gaver and Lewis \(1980\)](#)). This Markov chain is also geometrically ergodic in the sense of the above definition (see for instance [Kesten \(1974\)](#)).

The following corollary gives conditions under which the Gamma stochastic volatility models satisfy the requirements of [Proposition 1](#) and [Theorem 1](#).

**Corollary 2.** *Suppose that the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is defined for  $i \in \mathbb{N}$ , by  $\sigma_i = \sqrt{h_i}$  where  $(h_i)_{i \in \mathbb{N}}$  is a positive time-homogeneous strictly stationary and geometrically ergodic Markov chain with marginal Gamma  $\Gamma(p, \lambda)$  distribution. Then the process  $(\epsilon_i)_{i \in \mathbb{N}}$ , as defined by [\(8\)](#), is a stationary MDS,  $\beta$ -mixing with  $\beta_n = O(e^{-\rho n})$ , for some  $\rho > 0$ , and with finite all integer moments. If Assumptions (A) are satisfied then the conclusions of [Proposition 1](#) and [Theorem 1](#) hold.*

**Proof of Corollary 2.** This sequence  $(\epsilon_n)_{n \geq 1}$  has finite moments at any order. In particular, for  $r \in \mathbb{N}$ ,

$$\mathbb{E}(\epsilon_1^{2r}) = (2r-1)(2r-3)\cdots 3 \times 1 \frac{\Gamma(p+r)}{\Gamma(p)} \lambda^{-r},$$

(see for instance [Abraham, Balakrishna and Sivakumar \(2007\)](#) and the references therein). Now the sequence  $(\sigma_i)_i$  is also  $\beta$ -mixing so it is also strongly mixing with

$$\alpha_n \leq \beta_n \leq a_1 e^{-a_2 n},$$

for some positive real numbers  $a_1$  and  $a_2$ . It is proved ([Rio \(1993\)](#), see also Lemma 9 in [Doukhan and Louhichi \(1999\)](#)) that, for  $1 \leq i_1 \leq \cdots \leq i_k < i_{k+1} \leq \cdots \leq i_q \leq n$  such that  $s := i_{k+1} - i_k \geq \max_{1 \leq l \leq q-1} (i_{l+1} - i_l)$ ,

$$|\text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q})| \leq 4 \int_0^{\alpha_s} Q^q(u) du, \quad (9)$$

where  $Q$  is the quantile function of  $\sigma_1$ , i.e. the inverse of the tail function  $t \mapsto \mathbb{P}(\sigma_1 > t)$ . We have  $\int_x^\infty t^{a-1} e^{-t} dt \sim x^{a-1} e^{-x}$  as  $x$  tends to infinity. We deduce, from this, that, as  $x$  tends to infinity,

$$\mathbb{P}(h_1 > x) = \frac{\lambda^p}{\Gamma(p)} \int_x^\infty t^{p-1} e^{-\lambda t} dt \sim \frac{\lambda^{p-1}}{\Gamma(p)} x^{p-1} e^{-\lambda x}.$$

Hence  $Q(u) \leq \sqrt{\frac{2}{\lambda} \ln(\frac{c}{u})}$  for some  $c > 1$  and, for some  $a \in ]0, 1[$ .

From now on, let  $cst$  denotes a constant that may be different from line to line  
We have,

$$\begin{aligned} \int_0^{\alpha_s} Q^q(u) du &\leq cst \int_0^{\alpha_s} \left( \ln\left(\frac{c}{u}\right) \right)^{q/2} du \\ &\leq cst \int_{c/\alpha_s}^{\infty} \frac{(\ln(v))^{q/2}}{v^2} dv \leq cst \int_{c/\alpha_s}^{\infty} \frac{1}{v^{1+a}} dv \leq cst \alpha_s^a \leq cst e^{-a\rho s}. \end{aligned}$$

Consequently,

$$\sum_{s \geq 1} s^4 \int_0^{\alpha_s} Q^q(u) du < \infty. \quad (10)$$

Condition (6) is satisfied from (9) and (10). As in the proof of Corollary 1, we deduce that all the requirements of Proposition 2 are satisfied and therefore the conclusions of Proposition 1 and Theorem 1 also hold.  $\square$

### 3.3 Other SV model with no mixing properties

The following corollary studies another SV model for which the conclusions of Proposition 1 and Theorem 1 still hold. In this model, we suppose that  $\log(\sigma_i)$  is a linear process with Bernoulli innovation having coefficients  $(2^{-k})_k$ . Unlike the log-normal SV, in this case the volatility sequence  $(\sigma_i)_i$  is not strongly mixing (see Bradley (1986)) but it's associated in the sense of Esary, Proschan and Walkup (1967). Recall that the sequence  $(\sigma_i)_i$  is associated if for any non decreasing and bounded functions  $f$  and  $g$ ,

$$\text{Cov}(f(\sigma_1, \dots, \sigma_n), g(\sigma_1, \dots, \sigma_n)) \geq 0. \quad (11)$$

**Corollary 3.** *Suppose that the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is defined for  $i \in \mathbb{N}$ , by  $\sigma_i = \exp\left(\sum_{j=0}^{\infty} 2^{-j} \eta_{i-j}\right)$  where  $(\eta_i)_{i \in \mathbb{Z}}$  is an i.i.d. centered sequence distributed as a Bernoulli  $\mathcal{B}(1/2)$  distribution. Suppose also that  $Z_1$  follows a standard normal law. Then the volatility sequence  $(\sigma_i)_{i \in \mathbb{N}}$  is associated. If Assumptions (A) are satisfied then the conclusions of Proposition 1 and Theorem 1 hold.*

**Proof of Corollary 3.** Let  $\mathcal{F}_i = \sigma(Z_0, \dots, Z_i, (\eta_l)_{l \leq i+1})$ . Then  $\sigma_i$  is  $\mathcal{F}_{i-1}$ -measurable. We have also, by independence of  $Z_i$  and  $\mathcal{F}_{i-1}$  that  $\mathbb{E}(Z_i | \mathcal{F}_{i-1}) = 0$ , almost surely. The sequence  $(\sigma_i)_i$  is associated since it is a nondecreasing of independent random variables (see Esary, Proschan and Walkup (1967)). The random variable  $\sigma_i$  is bounded,  $|\sigma_i| \leq e^2$  and, by (see Birkel (1988)),

$$|\text{Cov}(\sigma_i, \sigma_l)| \leq e^2 \sum_{j=0}^{\infty} 2^{-j} \sum_{k=0}^{\infty} 2^{-k} \text{Cov}(\eta_{i-j}, \eta_{l-k}) \leq cst 2^{-|i-l|}.$$

For  $1 \leq i_1 \leq \dots \leq i_k < i_{k+1} \leq \dots \leq i_q \leq n$  such that  $s := i_{k+1} - i_k \geq \max_{1 \leq l \leq q-1} (i_{l+1} - i_l)$ , we have using [Birkel \(1988\)](#),

$$|\text{Cov}(\sigma_{i_1} \cdots \sigma_{i_k}, \sigma_{i_{k+1}} \cdots \sigma_{i_q})| \leq e^{2(q-2)} \sum_{i=i_1}^{i_k} \sum_{l=i_{k+1}}^{i_q} \text{Cov}(\sigma_i, \sigma_l) \leq cst 2^{-s}.$$

Condition (6) is then satisfied. All the requirements of [Proposition 2](#) are satisfied and therefore the conclusions of [Proposition 1](#) and [Theorem 1](#) hold.  $\square$

## 4 A Monte-carlo simulation study for a “trend plus a log-normal SV process”

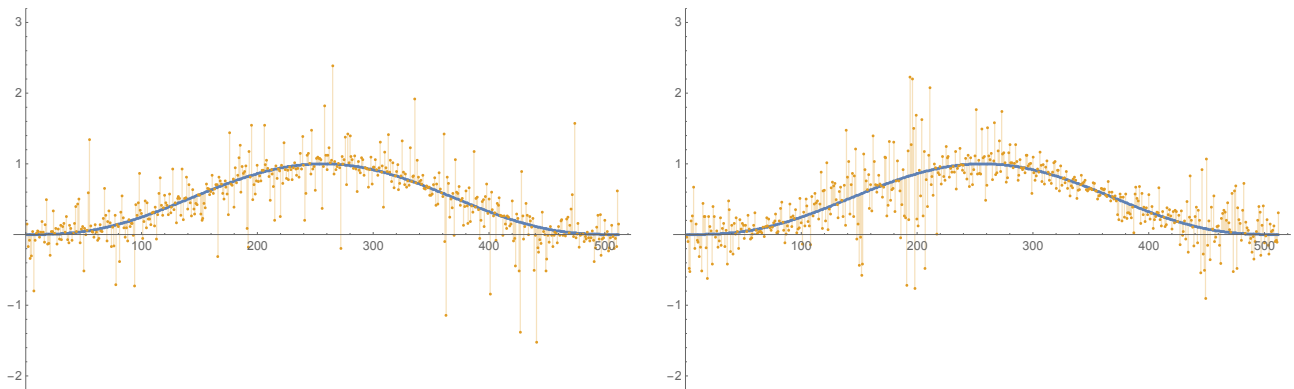


Figure 1 :  $n = 2^9$ . Each of these 2 panels displays one data set  $Y$  and the “smooth” deterministic trend  $r(x)$  when the noise is a log-normal SV sequence. In the 2 panels  $\tau = 0.75$ , and they only differ by  $\gamma = 0.01$  (left) and  $\gamma = 0.98$  (right); see (12) for the definition of these two parameters

Our experiments use the same designs as the ones used for the experiments in [Benhenni, Girard and Louhichi \(2020\)](#); the only difference is that the noise sequence is now a common log-normal SV sequence, as analyzed in [Corollary 1](#), instead of an ARCH(1). We thus consider the same smooth function  $r(x) = (4x(1-x))^3$  as “deterministic trend”, an equispaced design and a noise level  $\sigma$  for which the noise-to-signal ratio is “moderate” (precisely  $0.32^2$ ). The kernel regression technique employs a periodic kernel smoothing (the chosen  $r$  being “smoothly periodic”, such a periodic processing is appropriate and makes very affordable extensive large-scale simulations). We use  $u(t) \equiv 1$  as a weight function, which is possible in this periodic setting. See [Benhenni, Girard and Louhichi \(2020\)](#) for more details.

As parameters, in order to define the noise process, in addition to  $\sigma$  and  $\gamma$  (the serial correlation introduced in [Corollary 1](#)), we introduce

$$\tau := \sqrt{\text{Var}(\eta_1)/(1-\gamma^2)}. \quad (12)$$

The advantage of using  $\tau$  instead of  $\text{Var}(\eta_1)$  is, as it is immediate to see, that  $\tau$  is the unique shape-parameter for the marginal density of the sequence of conditional variances  $\sigma_i^2$ 's. Such a parameterization is common (see e.g. Taylor (1994)). Let us remark that, to generate a noise process of variance  $\sigma^2$ , it is easy to check that the variance of  $Z_i$  (used in definition 8 of the  $\epsilon_i$ 's) has to be set to  $\sigma^2 \exp(-2\tau^2)$ .

We consider three values for  $\tau$ , and three values for the serial-correlation parameter  $\gamma$ , precisely

$$\tau \in \{0.2, 0.4, 0.75\}, \quad \gamma \in \{0.01, 0.9, 0.98\},$$

with a common value  $\sigma = 0.32$ . Note that the intermediate value 0.4 for  $\tau$  is representative of values often obtained by fitting such a log-normal SV model to real econometric series; see Taylor (1994) (especially its Section 3.4, where  $\tau$  is denoted by  $\beta$ ) for an interesting review. Any large value of  $\tau$  (say, greater than 1) implies a very fat tail for the marginal density of the amplitude of the noise  $|\epsilon_i|$  which may cause a large instability of the classical kernel curve-estimate (a “robust” version kernel smoothing would be much more appropriate in such case). On the other hand, recall that a value very close to 0 for  $\tau$  would imply that the density of the conditional variance  $\sigma_i^2$  is concentrated around 1 and thus the serial correlation would have virtually no impact on the dependence between the  $\epsilon_i$ 's (which is then a “quasi-iid-normal” sequence). Thus we restrict the present study to the range  $[0.2, 0.75]$  for  $\tau$ .

**The “a.o.” property.** First, let us analyze the asymptotic optimality (a.o.) result. As is well known, a result like Proposition 1 generally stems from a uniform relative accuracy result which states that  $\text{CL}(h) - n^{-1} \|U^{1/2}(Y - r)\|^2$  uniformly approximates  $T_n(h)$  (or its expectation  $\text{MASE}(h)$ ) with a small (in probability and in *sup* norm over the domain of candidate  $h$ 's) error, “small” being defined relatively to  $\text{MASE}(h)$ .

We resume in Figure 2 that a uniform relative accuracy is well observed for all the considered values of  $\tau$  and  $\gamma$  (note that the results for  $\gamma = 0.01$  and  $\tau \in \{0.2, 0.4, 0.75\}$  have produced plots very similar to the plot for  $\gamma = 0.9, \tau = 0.2$  (top-left panel in Figure 2) so they are not included in Figure 2). Only a slight deterioration is observed for the largest  $\tau$  and  $\gamma$ .

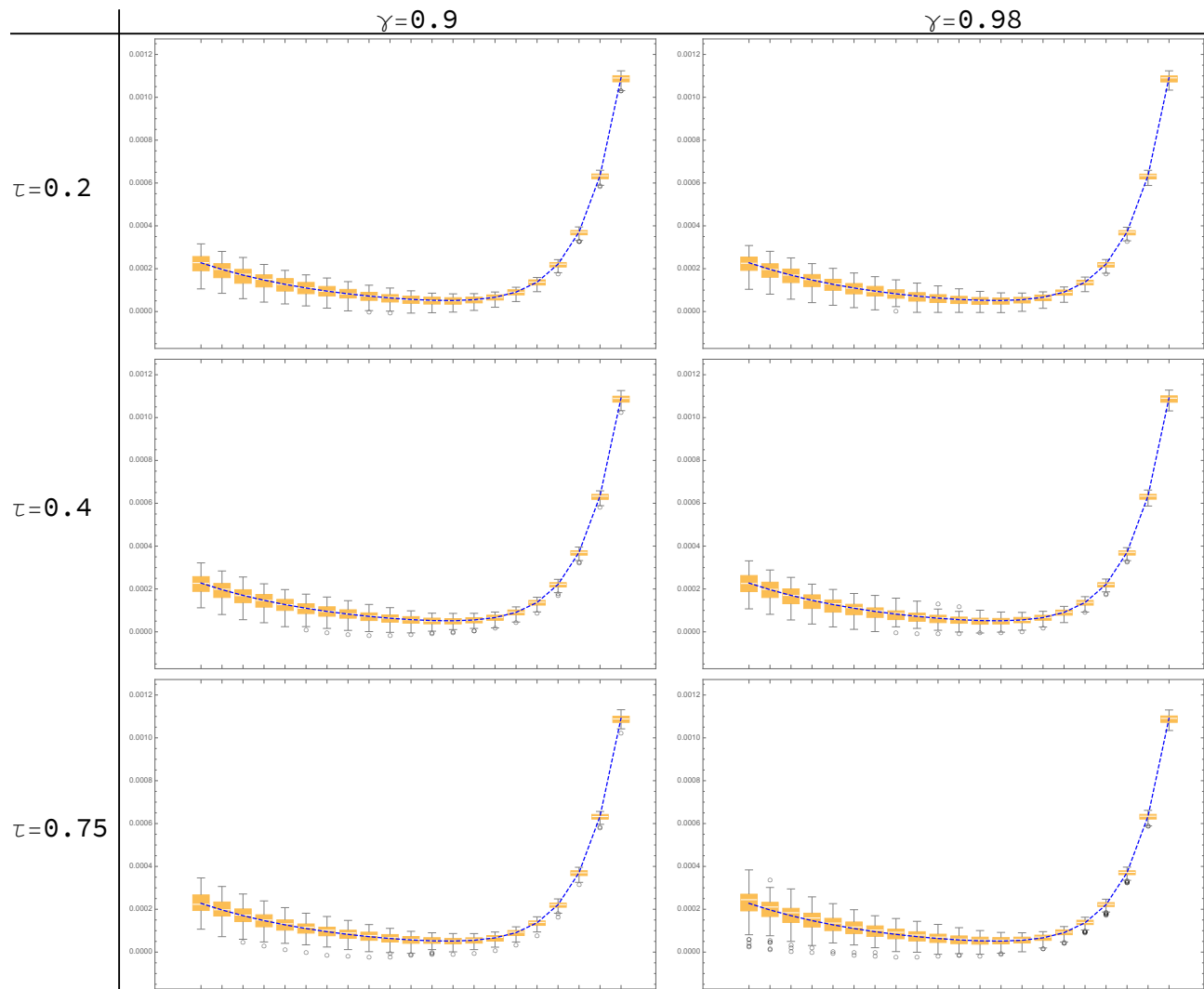


Figure 2 :  $n = 2^{15}$ . These 6 panels only differ by  $(\tau, \gamma)$  varying in  $\{0.2, 0.4, 0.75\} \times \{0.9, 0.98\}$ . In each panel, the dashed blue curve is the “empirical MASE”, precisely the average (over the 3000 replicates) of the  $T_n(h)$  curves. Each of the 21 boxplots (located at 21 fixed discrete values for  $h$ ) are built from the first 100 replicates of  $CL(h) - n^{-1} \|U^{1/2}(Y - r)\|^2$ .

**Asymptotic normal distribution.** This Section aims to assess the usefulness of the theoretical asymptotic normal approximation stated in Theorem 1 for reasonable dataset sizes  $n$ . We are going to demonstrate that both  $\tau$  and  $\gamma$  have an impact on the rate of convergence (with  $n$ ) toward this approximation.

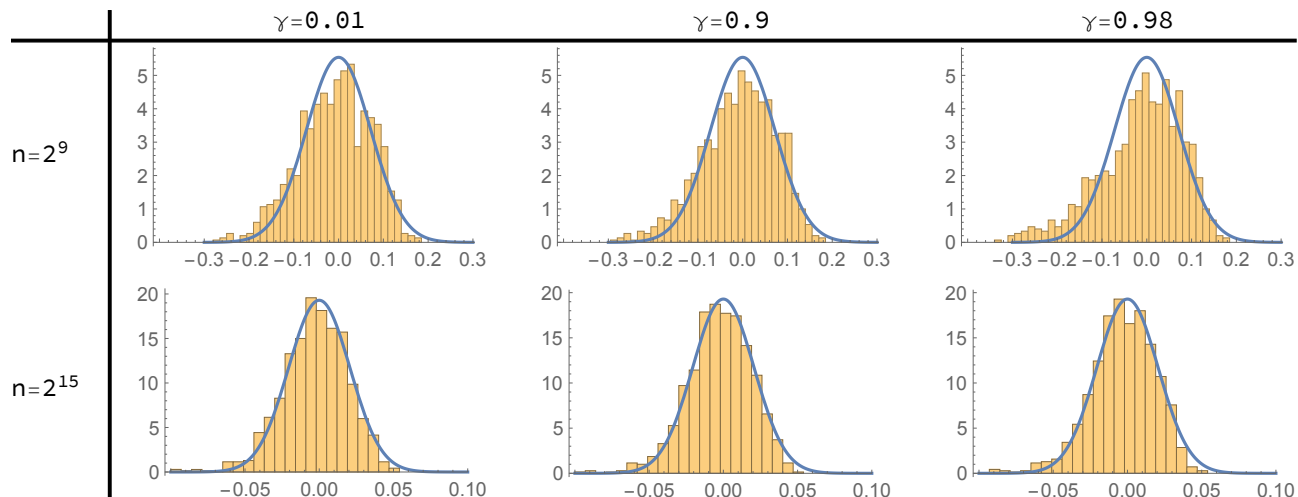


Figure 3 :  $\tau = 0.2$ . . These 6 panels only differ by  $n$  ( $= 2^9$  in the top row and  $2^{15}$  in the bottom row) and by  $\gamma$  varying in  $\{0.01, 0.9, 0.98\}$ . In each panel, the displayed normalized histogram is that of the 3000 replicates of  $\hat{h}_M - \hat{h}_n$ . The superposed blue curve is the normal distribution of  $\hat{h}_M - \hat{h}_n$  predicted by the asymptotic theory.

Let us first consider  $\tau = 0.2$ . By inspecting Figure 3, we clearly see, in the three top panels, that the asymptotic approximation fits rather well already for  $n = 2^9$  and for any  $\gamma \in \{0.01, 0.9, 0.98\}$ . For  $n = 2^{15}$  the three bottom panels illustrate that the asymptotic theory provides a very accurate prediction of the finite sample “truth”. Notice that, as expected the accuracy for  $n = 2^{12}$  (not displayed here) is observed to be intermediary between the one for  $n = 2^9$  and that for  $n = 2^{15}$ , and is thus also quite good.

Notice that, again as expected, the range of the abscissae ( $h$ -differences) increases by moving from  $n = 2^{15}$  (bottom) to  $n = 2^9$  (top).

It is good news that the approximation given by Theorem 1 is very useful for  $n$  as small as 512.



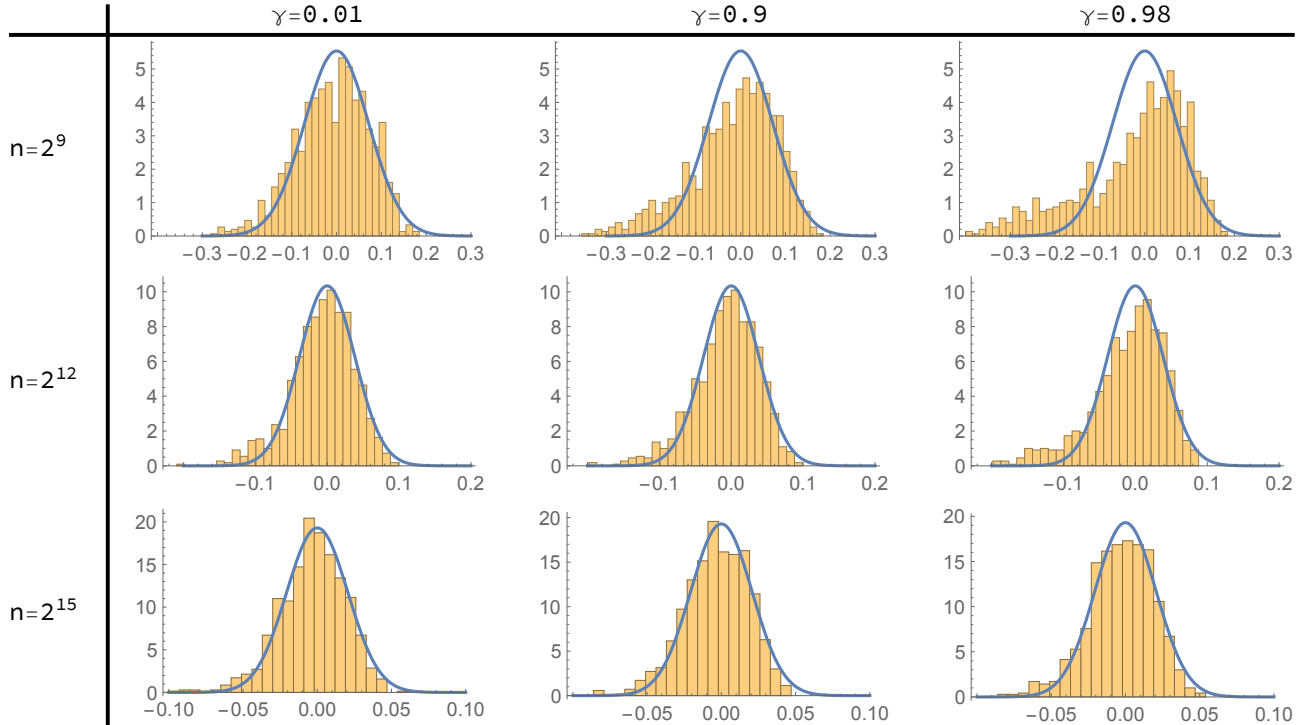


Figure 4 :  $\tau = 0.4$ . These 9 panels only differ by  $n$  ( $= 2^9$  in the top row,  $2^{12}$  (top) and  $n = 2^{15}$  (bottom). ) and by  $\gamma$  varying in  $\{0.01, 0.9, 0.98\}$ . In each panel, the displayed normalized histogram is that of the 3000 replicates of  $\hat{h}_M - \hat{h}_n$ . The superposed blue curve is the normal distribution of  $\hat{h}_M - \hat{h}_n$  predicted by the asymptotic theory.

The simulation results for  $\tau = 0.4$  are described in Figure 4. Here we add the three panels corresponding to  $n = 2^{12}$ . The analog figure for  $\tau = 0.75$  is Figure 5. Now one clearly sees that, for  $\gamma = 0.01$  (first column in these two  $3 \times 3$  arrays of histograms) the smallest value of  $n$  ( $= 2^9$ ) is always sufficient for the usefulness of the asymptotic normal approximation - although there is a slight degradation for  $\tau = 0.75$  (precisely the histogram in the top-left panel exhibits a non-negligible proportion of large negative values for  $\hat{h}_M - \hat{h}_n$  which almost always are associated with too-small  $\hat{h}_M$ 's). One observes that the latter degradation is softened if  $n$  is increased to  $2^{12}$  (middle panel of first column of Figure 5). Next, an inspection of the second column (thus  $\gamma = 0.9$ ) of both these two arrays shows that  $n = 2^9$  is "just sufficient" only for the smaller  $\tau = 0.4$  and provided one accepts a slight inaccuracy of the same type as the one mentioned above. But  $n = 2^{12}$  is clearly required for  $\tau = 0.75$ . Notice that, because of the observed jump in the observed accuracy by passing from  $(\tau, \gamma) = (0.75, 0.01)$  to  $(0.75, 0.9)$ , we also repeated the same simulations for the case  $(\tau, \gamma) = (0.75, 0.5)$ : they produced a histogram rather close to the one for  $(0.75, 0.01)$ ; this demonstrates that it is only for "large"  $\gamma$  (that is, near 0.9 or above) that the asymptotic approximation is not accurate for  $n = 2^9$ .

Next, the third columns (that is, for  $\gamma = 0.98$ ) shows that  $n = 2^{12}$  is required for  $\tau = 0.4$ , and  $n = 2^{15}$  is required for  $\tau = 0.75$ .

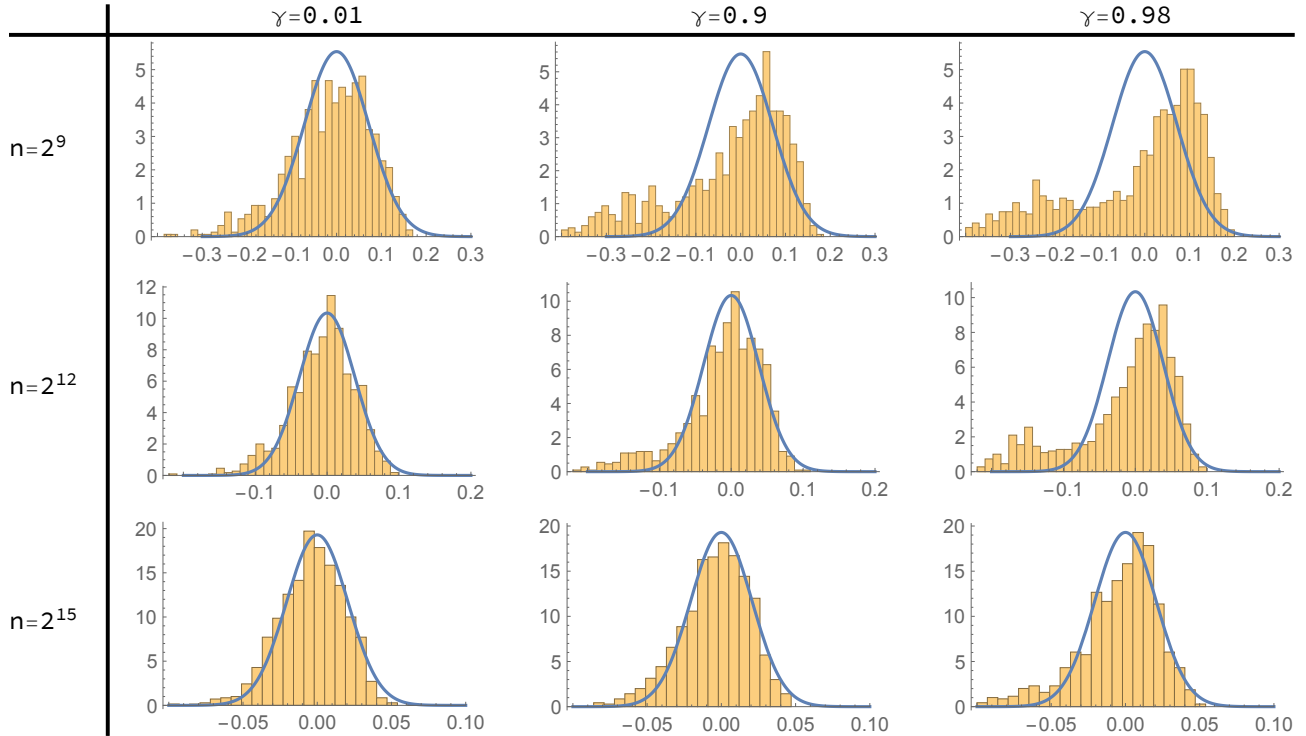


Figure 5 :  $\tau = 0.75$ .  $n = 2^9$  (top),  $n = 2^{12}$  (middle) and  $n = 2^{15}$  (bottom). These 9 panels only differ by  $n$  and by  $\gamma$  varying in  $\{0.01, 0.9, 0.98\}$ . In each panel, the displayed normalized histogram is that of the 3000 replicates of  $\hat{h}_M - \hat{h}_n$ . The superposed blue curve is the normal distribution of  $\hat{h}_M - \hat{h}_n$  predicted by the asymptotic theory.

All these experiments are thus well in agreement with Theorem 1. But for certain settings, which are not un-common in practice (see Taylor (1994)) this normal approximation is accurate only for quite large  $n$  (for example,  $n$  larger than  $2^{12}$  is required for  $(\tau, \gamma) = (0.4, 0.98)$ ). And this so-required value for  $n$  is shown to be an increasing function of both  $\tau$  and  $\gamma$ .

## Acknowledgements

This paper was developed in the framework of Grenoble Alpes Data Institute (ANR-15-IDEX-02).

## 5 References

- Abraham, B., Balakrishna, N., and Sivakumar, R. (2007). Gamma stochastic volatility models. *Journal of Forecasting*. 25, 153-171.
- Benelmadani, D., Benhenni, K. and Louhichi, S. (2019). The reproducing kernel Hilbert space approach in nonparametric regression problems with correlated observations. *Ann. Inst. Stat. Math.* Available at <https://doi.org/10.1007/s10463-019-00733-3>
- Benhenni, K. Girard, D. and Louhichi, S. (2020). On bandwidth selection problems in non-parametric trend estimation under martingale difference errors. Available at <https://arxiv.org/abs/2003.10164>
- Billo, M. and Sartore D. (2005). Stochastic Volatility Models: A Survey with Applications to Option Pricing and Value at Risk DOI: 10.1002/0470013265.ch8 In book: Applied Quantitative Methods for Trading and Investment.
- Birkel, T. (1988). The invariance principle for associated processes. *Stochastic Process. Appl.* 27, 57-71.
- Bradley, R. C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probab. Surveys*, 2 107-144.
- Bradley, R. C. (1986). Basic properties of strong mixing conditions, in: Eberlein E., Taqqu M.S. (Eds.), *Dependence in Probability and Statistics*, Birkhäuser, Boston, 165-192.
- Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (1996). *Time Series Models: In econometrics, finance and other fields*. Chapman & Hall.
- Davis, R. A., Mikosch, T. (2009). Probabilistic properties of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) *Handbook of Financial Time Series*. Springer, 255-268.
- Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.*, 13-4, 691-696.
- Doukhan, P. (1994). *Mixing: Properties and Examples*. Lecture Notes in Statistics. New York. Springer-Verlag.
- Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84-2, 313-342.

- Esary, J., Proschan, F., Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Stat.* 38, 1466-1476.
- Fan, J., Gijbels, I., Hu, T.C. and Huang, L.S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statist. Sinica* 6, 113-127.
- Francisco-Fernandez, M. and Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian J. Statist.*, 33, 279-295.
- Gaver, D. P. and Lewis, P. A. W. (1980). First-Order Autoregressive Gamma Sequences and Point Processes. *Adv. Appl. Prob.* 12-3, 727-745.
- Ghysels, E., A. Harvey and E. Renault (1996). Stochastic volatility, in: *Statistical Methods in Finance*, Rao C. and Maddala G., eds., North-Holland, Amsterdam.
- Girard, D. (1998). Asymptotic comparison of (partial) cross-validation, GCV and randomized GCV in non-parametric regression. *Ann. Stat.*, 26, 315-334.
- Härdle, W., Hall, P. and Marron, J.S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association* 83, 86-95.
- Hall, P., Lahiri, S. N. and Polzehl, J. (1995). On bandwidth choice in nonparametric regression with both short and long-range dependent errors. *Ann. Statist.* 23-6, 1921-1936.
- Kesten, H. (1974). Renewal Theory for Functionals of a Markov Chain with General State Space. *Ann. Probab.* 2-3, 355-386.
- Liu, X. (2001). Kernel smoothing for spatially correlated data. Ph. D. thesis, Department of Statistics, Iowa State University.
- Meilan-Vila, A., Fernandez-Casal, R., Crujeiras, R.M. and Francisco-Fernandez, M. (2020). A computational validation for nonparametric assessment of spatial trends. Available at <https://arxiv.org/abs/2002.05489>
- Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.*, 61, 85-109.
- Pham, T. D., Tran, L. T. (1985). Some mixing properties of time series models. *Stoch. Proc. Appl.* 19, 297-303.
- Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* 12, 1215-1230.

- Rio, E. (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29, 587-597.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA*, 42, 43-47.
- Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* 90, 1257-1270.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility, in *Time Series Models with Econometric, Finance and Other Applications*, Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (eds.), Chapman & Hall, London, 1-677.
- Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley and Sons, Ltd., Chichester.
- Taylor, S.J. (1994). Modelling stochastic volatility: a review and comparative study. *Mathematical Finance*, 4, 183-204.