



HAL
open science

Nonparametric Estimation in Nonlinear Time-varying Autoregressive Locally Stationary Processes with ARCH-errors

Abderrahim Bourhattas, Naâmane Laïb

► **To cite this version:**

Abderrahim Bourhattas, Naâmane Laïb. Nonparametric Estimation in Nonlinear Time-varying Autoregressive Locally Stationary Processes with ARCH-errors. 2024. hal-04611601v2

HAL Id: hal-04611601

<https://hal.science/hal-04611601v2>

Preprint submitted on 21 Aug 2024

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.

Nonparametric Estimation in Nonlinear Time-varying Autoregressive Locally Stationary Processes with ARCH-errors

Abderrahim Bourhattas^{*,1} and Naâmane Laïb¹

¹CY Cergy Paris Université, Laboratoire AGM, UMR 8088 du CNRS. F-95000 Cergy, France.

Abstract. This paper investigates the asymptotic properties of the Kernel estimators for the conditional mean and variance functions in a time-varying nonlinear autoregressive locally stationary process with ARCH errors (tvNAR) of order $d > 1$. We establish the uniform almost sure consistency for these estimators with rate under assumptions of local stationarity and local ergodicity. Special attention is also paid to the Kernel density estimation of the tvNAR process. The results are derived without assuming a specific type of mixing conditions or a particular physical dependence measure on the data. Consequently, our results are applicable to a wide range of dependent processes, including those satisfying strong mixing conditions.

Key words. Locally stationary processes, Nonparametric estimation, Ergodicity, Nonlinear heteroscedastic models, Uniform convergence rates.

Subject Classifications: 62G05, 62G07, 62G08, 62M10.

1 INTRODUCTION

Stationary time series are commonly used to construct models that accurately fit data and predict random phenomena. The stationarity assumption plays an important role in time series analysis, since it allows particularly to use the ergodic theorem, one of the main tools to derive the consistency of estimators. Unfortunately, stationary time series models do not fully capture the complexities of the real world. Many financial and economic time series exhibit nonstationary behavior, as highlighted by [Mikosch and Starica \(2004\)](#).

Various nonstationarity concepts have been proposed in the past, with one notable approach being local stationarity introduced by [Dahlhaus \(1996a,b, 1997\)](#). Initially, the author defined this concept based on spectral representation and later extended it temporally. This extension, first introduced in [Subba Rao \(2006\)](#) and [Dahlhaus and Subba Rao \(2006\)](#) emphasizes the local approximation of a process by a stationary one near specific dates, while maintaining controlled errors.

[Truquet \(2019\)](#) generalized this notion to discrete state models (and Markov chains).

*Corresponding author: abderrahim.bourhattas@cyu.fr

With the help of the functional dependence measure, originally introduced by [Wu \(2005\)](#) for Bernoulli shift processes, [Dahlhaus et al. \(2019\)](#) started building the foundation for a comprehensive framework designed to address nonlinear locally stationary processes.

Notice also that, [Bardet and Doukhan \(2017\)](#) have considered a class of time varying linear AR(1) processes, with local stationarity and periodic features. They proposed a kernel based non-parametric estimation of the time-varying coefficient and established its limiting law.

In addition, it's worth noting the contributions of [Bardet et al. \(2020, 2022\)](#). The field has been enriched by significant works like those by [Dahlhaus et al. \(1997\)](#), [Nason et al. \(2000\)](#), [Sakiyama and Taniguchi \(2004\)](#), [Zhang and Wu \(2015\)](#), [Wu and Zhou \(2011\)](#), among others. Additional research focusing on bandwidth selection includes the works of [Dahlhaus and Giraitis \(1998\)](#) and as well as [Richter and Dahlhaus \(2019\)](#).

Although locally stationary models have primarily been used within the parametric framework, it is worth noting that [Vogt \(2012\)](#) and the work of [Zhang and Wu \(2015\)](#) have focused on the study of nonparametric time series regression models. We will be particularly interested in [Vogt \(2012\)](#)'s paper, where the author examined a general class of nonparametric time series regression models that can be locally approximated by stationary processes.

This paper deals with the nonparametric estimation of a nonstationary version of the model :

$$X_i = m(\mathbf{X}_{i-1}) + U(\mathbf{X}_{i-1})\epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where $\mathbf{X}_{i-1} = (X_{i-1}, \dots, X_{i-k}, \dots, X_{i-d}) \in \mathbb{R}^d$.

In the context of stationarity, several papers have been published in the past dealing with the probabilistic properties and nonparametric estimation of the functions m and U . Among others, [Masry and Tjøstheim \(1995\)](#) considered kernel estimates m_n and U_n^2 , and demonstrated their strong consistency with rates, as well as their asymptotic normality under a mixing condition. [Härdle and Tsybakov \(1997\)](#) tackled the estimation of m and U^2 using local polynomial regression methods, under strong mixing condition assumption. [Laïb \(2005\)](#) established the uniform convergence rate of the kernel estimators m_n and U_n^2 and derived their asymptotic normality, within the framework of stationarity and ergodicity conditions.

To provide a more structured context, our work concerns [Vogt \(2012\)](#)'s non-stationary time-varying nonlinear autoregressive model (tvNAR) with ARCH errors of order d ($d > 1$) defined by

$$X_{i,n} = m(i/n, \mathbf{X}_{i-1,n}) + U(i/n, \mathbf{X}_{i-1,n})\epsilon_i, \quad i = 1, \dots, n, \quad (1.2)$$

where $\mathbf{X}_{i-1,n} = (X_{i-1,n}, \dots, X_{i-k,n}, \dots, X_{i-d,n}) \in \mathbb{R}^d$.

The functions $m(u, x)$ and $U(u, x)$ are \mathbb{R}^2 -valued smooth functions of rescaled time $u \in]0, 1]$ and $x \in \mathbb{R}^d$. The real random variables ϵ_i 's are assumed to be i.i.d. with mean zero and variance one.

[Vogt \(2012\)](#) extensively investigated the probabilistic properties of the model (1.2), and developed estimation theory for the nonparametric regression function defined by (1.3).

$$Y_{i,n} = m(i/n, \mathbf{X}_{i,n}) + \epsilon_{i,n} \quad \text{for } i = 1, \dots, n \quad (1.3)$$

with $\mathbb{E}(\epsilon_{i,n} | \mathbf{X}_{i,n}) = 0$ almost surely (a.s.), where $Y_{i,n}$ and $\mathbf{X}_{i,n}$ represent random variables of

dimension 1 and d , respectively. Specifically, he derived the uniform convergence rate for a Kernel regression estimate of the function m and established its asymptotic normality under α -mixing condition.

We consider the model (1.2), and define $\mathcal{F}_{i,n} := \sigma(X_{l,n}; l \leq i)$ as the array of σ -fields generated by $(X_{l,n})$.

Thus, for any $i \geq 1$, $\mathbb{E}(X_{i,n} | \mathcal{F}_{i-1,n}) = m(i/n, \mathbf{X}_{i-1,n})$, which can be interpreted as the predicted value given the set of information $\mathcal{F}_{i-1,n}$. Moreover, $\text{Var}(X_{i,n} | \mathcal{F}_{i-1,n}) = U^2(i/n, \mathbf{X}_{i-1,n})$ measures the risk associated with this prediction. This class of models includes many relevant examples discussed in the literature.

We will stipulate that $m(u, x) = m(0, x)$ and $U(u, x) = U(0, x)$ for all $u \leq 0$. This implies that $X_{i,n} = X_i(0)$ for all $i \leq 0$, allowing us to consider only $i \geq 0$.

We investigate the asymptotic properties of the Kernel estimators for the conditional mean and variance functions in model (1.2), as well as the Kernel density estimation.

More precisely, we establish uniform almost sure consistency with rate for the conditional mean and conditional variance appearing in Model (1.2), under the assumptions of local stationarity and local ergodicity. Additionally, we show the uniform convergence of the Kernel density estimate. Here, local ergodicity refers to the property of ergodicity applying to the stationary approximation processes.

Our paper extends the work of Laïb (2005) and Masry and Tjøstheim (1995) to the locally stationary case. It also completes and extends Vogt (2012)'s work in a specific aspect. Without any mixing assumption, we only use ergodicity to investigate the uniform a.s. convergence for both mean and variance, rather than considering the uniform consistency in probability of the mean.

Additionally, we address the issue of the uniform convergence of the Kernel density estimate.

Our approach employs techniques involving martingale differences devices and projections onto suitable σ -fields, as in Zhang and Wu (2015).

Note that despite this approach allowing us to achieve an optimal convergence rate for the conditional mean and variance functions, it may not provide a convergence rate for the Kernel density estimate (see, Remark 4.2).

The paper is organized as follows : Section 2 introduces the model, defines local stationarity, and presents the notations, main hypotheses, and probabilistic properties of the model. In Section 3, we define the kernel estimators of the conditional mean and variance functions, and introduce additional assumptions. Section 4 presents the main results of the study. Section 5 provides the proofs of the main results. For brevity, some of the proofs are deferred to the appendix.

2 LOCALLY STATIONARY PROCESSES AND PROPERTIES OF THE TVNAR-MODEL

2.1 Definition of a Locally stationary process

This subsection provides a definition of a locally stationary process, primarily based on Dahlhaus and Subba Rao (2006) and Vogt (2012).

The functions m and U are identified a.s. on the grid of points $\frac{i}{n}$ for $i = 1, \dots, n$, forming a dense subset of the interval $[0, 1]$, as the value of n tends to infinity. Therefore, m and U are identified a.s. at all rescaled time points $u \in [0, 1]$ since they are assumed to be continuous in the time direction. This allows us to approximate the tvNAR-process defined in (1.2) by a stationary process in a neighborhood of a fixed point i (or in rescaled time u).

First, let's introduce the following definition : For a given $u \in]0, 1]$, the stochastic process $X_i(u) : i \in \mathbb{Z}$, is referred to as a stationary process associated with the tvNAR-process at time point u if it satisfies :

$$X_i(u) = m(u, \mathbf{X}_{i-1}(u)) + U(u, \mathbf{X}_{i-1}(u)) \epsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

where the rescaled time argument is fixed at u and $\mathbf{X}_{i-1}(u) = (X_{i-1}(u), \dots, X_{i-k}(u), \dots, X_{i-d}(u)) \in \mathbb{R}^d$.

Comparing (1.2) with (2.1), it seems clear that if $\frac{i}{n}$ is close to u , then $X_{i,n}$ and $X_i(u)$ should be close and the proximity degree should depend both on the rescaling factor n and the deviation $|\frac{i}{n} - u|$ (see Dahlhaus and Subba Rao (2006)). This gives the next definition :

Definition 2.1. [Dahlhaus and Subba Rao (2006), Vogt (2012)] $(X_{i,n})_{i \in \mathbb{N}^*}$ is locally stationary if for each rescaled time point $u \in]0, 1]$ there exists an associated process $(X_i(u))$ such that :

- (i) $(X_i(u))$ is strictly stationary and has a stationary density $f_{X_i(u)}$,
- (ii) it holds that

$$|X_{i,n} - X_i(u)| \leq \left(\left| \frac{i}{n} - u \right| + \frac{1}{n} \right) V_{i,n}(u) \quad a.s. \quad (2.2)$$

where $V_{i,n}(u)$ is a positive process such that

$\mathbb{E}[V_{i,n}(u)^\rho] < C$ for some $\rho > 0$ and $C < \infty$ independent of u, i and n .

This is the definition that will be used throughout this article.

2.2 Assumptions and properties of the tvNAR-model

This subsection mainly recalls the work of [Vogt (2012)] presenting the assumptions that are necessary to establish the local stationarity of the tvNAR process, guarantee the existence of a stationary solution of model (2.1), its density and that of the local stationary $X_{i,n}$, as well as the approximation (2.2).

To facilitate this, we introduce some notations that will be used in the sequel.

We denote \mathcal{C}^1 the space of continuously differentiable functions. We define $I_h := [Ch, 1 - Ch]$, where h is the bandwidth of estimation, C is a positive constant and can take different values from one expression to another, and let S be a fixed compact subset of \mathbb{R}^d . Additionally, $\overset{\circ}{A}$ denotes the interior of a subset A of \mathbb{R}^k . We will now list the conditions on m :

(M1) m is bounded : $\exists C_m < \infty, |m(\cdot, \cdot)| \leq C_m$.

(M2) m is lipschitz with respect to u : $\exists L_m < \infty, |m(u, x) - m(v, x)| \leq L_m |u - v|$ for all $x \in \mathbb{R}^d$.

(M3) m is continuously differentiable with respect to x and

$$\exists C < \infty, \exists \delta < 1, \sup_{u \in [0, 1], \|x\| > C} \left| \frac{\partial m(u, x)}{\partial x_j} \right| \leq \delta.$$

The function U satisfies the following conditions :

- (U1) U is bounded from above and from below : $\exists c_u > 0, \exists C_u < \infty, c_u \leq |U(\cdot, \cdot)| \leq C_u$.
(U2) U is lipschitz with respect to u : $\exists L_u < \infty, |U(u, x) - U(v, x)| \leq L_u |u - v|$ for all $x \in \mathbb{R}^d$.
(U3) U is \mathcal{C}^1 with respect to x and $\exists C < \infty, \sup_{u \in [0,1], \|x\| > C} \left| \frac{\partial U(u,x)}{\partial x_j} \right| \leq \delta < 1$.

The error term is required to satisfy :

- (E1) The random variables (r.v.) ϵ_i have a common positive continuous density f_ϵ .
(E2) The density f_ϵ is bounded and lipschitz : $\exists L_\epsilon < \infty, |f_\epsilon(x) - f_\epsilon(y)| \leq L_\epsilon |x - y|$ for all $x \in \mathbb{R}^d$.
The theorems below summarize the properties of the tvNAR model.

Theorem 2.2. [Vogt (2012), Theorems 3.1-3.2]. *If the assumptions (M1)-(M3), (U1)-(U3) and (E1) are satisfied, then :*

- (i) for each $u \in \mathbb{R}$, the model (2.1) has a strictly stationary solution $X_i(u)$;
(ii) the random variables $X_i(u)$ have a density $f_{X_i(u)}$ w.r.t. Lebesgue measure ;
(iii) the random variables $X_{i,n}$ have a density $f_{X_{i,n}}$ w.r.t. Lebesgue measure.
(iv)

$$|X_{i,n} - X_i(u)| \leq \left(\left| \frac{i}{n} - u \right| + \frac{1}{n} \right) V_{i,n}(u) \quad a.s. \quad (2.3)$$

where $V_{i,n}(u)$ is a positive process such that $\mathbb{E}[|V_{i,n}(u)|^\rho] < C$ for some $\rho > 0$ and $C < \infty$ independent of u, i and n .

Theorem 2.3. [Vogt (2012), Theorem 3.3]. *Let $\mathbf{f}(u, x) := f_{\mathbf{X}_i(u)}$ be the density of $\mathbf{X}_i(u) = (X_i(u), X_{i-1}(u), \dots, X_{i-d+1}(u))$.*

Then we have, under assumptions of Theorem 2.2 and condition (E2), that

$$|\mathbf{f}(u, x) - \mathbf{f}(v, x)| \leq L_x |u - v|^p$$

for some constant $0 < p < 1$ and $L_x < \infty$ continuously depending on x .

3 KERNEL ESTIMATION FOR THE CONDITIONAL MEAN AND VARIANCE

In this section, we introduce kernel estimators of the conditional mean and the variance functions of the model (1.2). For $(u, x) \in [0, 1] \times \mathbb{R}^d$, we define a kernel estimate for the function m as follows:

$$\begin{aligned} m_n(u, x) &= \frac{\sum_{i=d+1}^n X_{i,n} K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n})}{\sum_{i=d+1}^n K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n})} \\ &= \frac{\sum_{i=d+1}^n X_{i,n} K_h(u - \frac{i}{n}) \prod_{j=1}^d K_h(x^j - X_{i-1,n}^j)}{\sum_{i=d+1}^n K_h(u - \frac{i}{n}) \prod_{j=1}^d K_h(x^j - X_{i-1,n}^j)} =: \frac{N_n(u, x)}{D_n(u, x)}, \end{aligned} \quad (3.1)$$

where $x = (x^1, x^2, \dots, x^d)$, K denotes a one-dimensional kernel function and \mathbf{K} a d-dimensional one. We use the notation $K_h(s) = K(s/h)$. For convenience, we work with a product kernel and assume that the bandwidth h is the same in each direction. Our results can however be easily modified to allow for non-product kernels and different bandwidths.

Dividing both the numerator and the denominator by $(n-d)h^{d+1}$, we can also write :

$$m_n(u, x) = \frac{((n-d)h^{d+1})^{-1}N_n(u, x)}{f_n(u, x)}$$

where $f_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i}{n})\mathbf{K}_h(x - \mathbf{X}_{i-1,n})$ is the kernel estimator of the density of $\mathbf{X}_{i-1,n}$.

Notice that $U^2(x) = \text{Var}(X_{i,n}|\mathbf{X}_{i-1,n} = x) = \mathbb{E}[(X_{i,n} - m(\frac{i}{n}, \mathbf{X}_{i-1,n}))^2|X_i = x]$. Therefore, knowing m_n , we can define the kernel estimator for the conditional variance $U^2(u, x)$ as follows:

$$U_n^2(u, x) = \frac{1}{f_n(u, x)} \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i-1}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) [X_{i,n} - m_n(\frac{i}{n}, \mathbf{X}_{i-1,n})]^2 \quad (3.2)$$

3.1 Notations and Assumptions

We begin by introducing some notations that will be used in the sequel. Let $(\Omega, \mathcal{F}, \mu)$ be a σ -finite measure space, where Ω denotes the underlying space, \mathcal{F} the σ -algebra of measurable sets, and μ is the measure on (Ω, \mathcal{F}) . The space of integrable functions on Ω with respect to μ is denoted as L^1 .

We define the triangular array of sigma-algebras $\mathcal{F}_{i,n} = \sigma(X_{0,n}, X_{1,n}, \dots, X_{i,n})$.

We will denote $f_Y^{\mathcal{F}_{i-1,n}}$ the conditional density of Y given $\mathcal{F}_{i-1,n}$. We denote $\mathbb{E}^{\mathcal{F}_{i-1,n}}(Y)$ or $\mathbb{E}(Y|\mathcal{F}_{i-1,n})$ the conditional expectation of Y with respect to $\mathcal{F}_{i-1,n}$. $\mathcal{C}_0(\mathbb{R})$ will be the space of continuous functions that tend to zero at infinity, with the sup norm denoted by $\|\cdot\|$. S is a fixed compact set of \mathbb{R}^d on which the uniform convergence will be proved. I_h is the interval $[Ch, 1 - Ch]$ for some constant $C > 0$. The indicator function of a set E will be noted \mathbb{I}_E .

In order to state our results, we need the following assumptions, inspired by [Vogt \(2012\)](#).

Assumption (P). Regarding the process :

(P1) The process $(X_{i,n})$ is locally stationary in the sense of Definition 2.1 and the stationary approximation process $X_i(u)$ is ergodic.

(P2) The process $V_{i,n}$ controlling the lack of stationarity satisfies $\mathbb{E}(V_{i,n}^\rho | \mathcal{H}_{i-1,n}) \leq C < \infty$ where ρ is the same as in Equation (2.2).

(P3) The density $f(u, x) := f_{X_i(u)}(x)$ satisfies the following :

- (i) $f(u, x)$ is continuously differentiable w.r.t. u for each $x \in \mathbb{R}^d$.
- (ii) $f(u, x)$ is continuously differentiable w.r.t. x for each $u \in [0, 1]$.
- (iii) $f(u, x)$ satisfies $\inf_{u \in [0,1], x \in S} f(u, x) = \delta > 0$.

(P4) $m(u, x)$ is twice continuously differentiable w.r.t. x for each $u \in [0, 1]$.

Assumption (CD) The conditional density of $\mathbf{X}_{i,n}$ is bounded, i.e. $\exists f^* \in \mathbb{R}^+$, independent of i and n such that, $f_{\mathbf{X}_{i,n}}(\cdot | \mathcal{F}_{i-1,n}) \leq f^*$.

Assumption (K) Regarding the kernel :

(K1) The kernel K is bounded by \overline{K} , $\int K(u)du = 1$, $\int uK(u)du = 0$, $\int u^2K(u)du = \kappa_2 < \infty$, and $\int K^2(u)du = \underline{\kappa}_2 < \infty$.

(K2) K has compact support, that is $K(v) = 0$ for all $|v| > C_k$, for some $C_k < \infty$ and is Lipschitz : $\exists L_K < \infty, |K(u) - K(v)| \leq L_K|u - v|$ for all $u, v \in \mathbb{R}$.

In order to obtain uniform convergence with rate, we need some supplementary assumptions :

Assumption (S1). $\exists s > 2, \exists C_T < \infty, \mathbb{E}(|\epsilon_i|^s) \leq C_T$.

To guarantee the uniform convergence of the variance estimator, assumption **(S1)** should be strengthened to :

Assumption (S2) $\exists s > 4, \exists C_v < \infty, \mathbb{E}(|\epsilon_i|^s) \leq C_v$.

We will also need the additional assumptions, on U :

Assumption (S3) $U(.,.)$ is lipschitz of order $\gamma > 1$, i.e. $\exists L_{u2} < \infty$,

$|U(u, x) - U(v, x)| \leq L_{u2}|u - v|^\gamma$ and $|U(u, x) - U(u, y)| \leq L_{u2}|x - y|^\gamma$ for all $u, v \in [0, 1], x, y \in \mathbb{R}^d$.

And the classical conditions on the bandwidth :

Assumption(H) The bandwidth h satisfies : $h \rightarrow 0$ and $nh^{d+1} \rightarrow \infty$.

The expected rate of convergence is : $O_{a.s.}(a_n)$ with $a_n = \left(\frac{\ln n}{nh^{d+1}}\right)^{1/2}$ together with some extra terms that are due to non stationarity.

Comments on the hypotheses. Conditions **(P)** and **(K)** are the same as those in [Vogt \(2012\)](#) with the addition of the ergodicity of the stationary process $X_i(u)$ and the condition **(P2)** which is stronger than the definition of local stationarity. On the other hand, condition **(P3)(iii)** is needed to provide a uniform lower bound for $\inf f_n(u, x)$ over the set $[0, 1] \times S$. This is essential for deriving the uniform convergence of the regression estimator. Condition **(S1)** is necessary to deal with the tail part after truncation for the mean and **(S2)** for the variance. Condition **(S3)** gives the last term of the rate of convergence of the variance. The parameter γ is at least 1 thanks to assumptions **(U1)**, **(U2)** and **(U3)**.

4 MAIN RESULTS

4.1 Uniform almost sure convergence of the density kernel estimator $f_n(u, x)$

The uniform convergence of the kernel estimator $f_n(u, x)$ of the density $f(u, x) = f_{\mathbf{X}_i(u)}(x)$ of the stationary approximation of $X_{i,n}$ is a crucial requirement for all the results presented in this paper. Below, we state and prove this result and give the proof in Appendix B.

Proposition 4.1. *Under the assumptions of theorem [2.2](#), **(P)**, **(K)**, **(CD)** and **(H)**, we have*

$$\sup_{u \in I_h, x \in S} |f_n(u, x) - f(u, x)| = o_{a.s.}(1) \quad (4.1)$$

Remark 4.2. A similar result of the statement [\(4.1\)](#) is obtained by [Vogt \(2012\)](#), in the context of time-varying regression model and mixing assumption, by splitting the estimator $f_n(u, x)$ into a variance part and a bias part. The bias is non random and its asymptotic behavior is obtained from Taylor expansion of de density $f(u, x)$ under some regularities conditions on f . In our case, the bias is a random variable and the presence of the Kernel relatively to the scale parameter makes its

treatment more complicated. In order to handle its asymptotic behavior, we split it into a sum of three terms. First, we deal with the first two terms using the stationary approximation of $X_{i,n}$ and the properties of the Kernel. The third term, which is the most important part of this decomposition, needs several approximations, along with the properties of the ergodicity (assumption **(P1)**).

Notice that, our results are obtained without any dependence condition. But they do not provide convergence rate for the density estimation.

To estimate the convergence rate in (4.1), additional assumptions, such as the physical dependence measure introduced by Wu (2005), or Laïb and Louani (2019), become necessary.

4.2 Uniform convergence of the conditional mean estimator

The following theorem states the uniform almost sure convergence, with rate, of the estimator m_n for $u \in I_h$ and x in a fixed compact set S of \mathbb{R}^d .

Theorem 4.3. *Under the assumptions of Theorem 2.2, and conditions **(P)**,**(CD)**,**(K)**,**(S1)** and **(H)**, we have :*

$$\sup_{u \in I_h, x \in S} |m_n(u, x) - m(u, x)| = O_{a.s.} \left(\sqrt{\frac{\ln n}{n h^{d+1}}} + \frac{1}{n^r h^d} + h^2 \right), \quad (4.2)$$

with $r = \min(1, \rho)$ and ρ is as in Theorem 2.2.

4.3 Uniform convergence of the volatility estimator

The following theorem states the uniform convergence with rate of the estimator U_n^2 for $u \in I_h$ and x in a fixed compact set S of \mathbb{R}^d .

Theorem 4.4. *Under the assumptions of Theorem 2.2, and conditions **(P)**,**(CD)**,**(K)**,**(H)**,**(S2)** and **(S3)**, we have that :*

$$\sup_{u \in I_h, x \in S} |U_n^2(u, x) - U^2(u, x)| = O_{a.s.} \left(\sqrt{\frac{\ln n}{n h^{d+1}}} \right) + O_{a.s.} \left(\frac{1}{n^{2r} h^{2d}} \right) + O_{a.s.}(h),$$

where γ is defined in **(S3)**, $r = \min(1, \rho)$ and ρ is as in Theorem 2.2.

Remarks on the main results and discussion

Theorem 4.3 generalizes Theorem 4.2 of Vogt (2012), established for the locally stationary mixing tvNAR regression model. We actually don't use Assumption **(E3)** of Vogt (2012) which guarantees the mixing property of the process. Yet this extension provides the same optimal convergence rate as the one found in Vogt (2012), but for almost sure uniform convergence and using only local ergodicity.

The locally stationary behavior of the model changes the asymptotic analysis of the bias component. Specifically, it introduces an additional component of order $1/n^r h^{d+1}$ in the expression of the convergence rate given in (4.2). This component arises from the replacement of the variable $X_{i,n}$ by $X_i(i/n)$ in the bias term, measuring the deviation from stationarity. The larger the value of r , the smaller the deviation of $X_{i,n}$ from its stationary approximation, and hence, the smaller the additional non-stationarity bias.

If we choose $h = O\left(\frac{\log n}{n}\right)^{\frac{d}{(d+1)(d+4)}}$, the convergence rate given in (4.2) becomes

$$O_{a.s.}\left(\sqrt{\frac{\ln n}{n h^{d+1}}} + \frac{1}{n^r h^d}\right) = a.s.\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{d+4}} + \frac{(\ln n)^{-\frac{d^2}{(d+1)(d+4)}}}{n^{r-\frac{d^2}{(d+1)(d+4)}}}\right).$$

Moreover, if $r > \frac{d^2+2d+2}{d^2+5d+4}$, which includes the case when the underlying process is stationary ($r = 1$), then the above rate reduces to $O_{a.s.}(a_n) = O_{a.s.}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{d+4}}\right)$ reaching the classical optimal rate for nonparametric regression estimation (see Masry and Tjøstheim (1995), Corollary 3.1) Theorem 4.4 generalizes Corollary 3.2 of Masry and Tjøstheim (1995), which was stated in the context of a stationary and mixing case, and Theorem 2 of Laïb (2005), established in the framework of stationarity and ergodicity. Similarly, we arrive at the same conclusion as above.

5 PROOFS

5.1 Preliminary results

Before we prove our results, we start from the definition of the Kernel estimate in (3.1) and use equation (1.2), to write

$$\begin{aligned} m_n(u, x) - m(u, x) &= f_n(u, x)^{-1}((n-d)h^{d+1})^{-1} \sum_{i=d+1}^n K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) [X_{i,n} - m(u, x)] \\ &= f_n(u, x)^{-1}((n-d)h^{d+1})^{-1} \sum_{i=d+1}^n K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) \left[m(\frac{i}{n}, \mathbf{X}_{i-1,n}) - m(u, x) \right] \\ &+ f_n(u, x)^{-1}((n-d)h^{d+1})^{-1} \sum_{i=d+1}^n K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i \\ &=: f_n(u, x)^{-1} (T_n(u, x) + V_n(u, x)) \end{aligned} \tag{5.1}$$

$$\text{with } f_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i.$$

Now, in order to separate the bias from the variance part, we define

$$\overline{T}_n(u, x) =: ((n-d)h^{d+1})^{-1} \sum_{i=d+1}^n \mathbb{E} \left(K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) \left[m(\frac{i}{n}, \mathbf{X}_{i-1,n}) - m(u, x) \right] \middle| \mathcal{F}_{i-2,n} \right)$$

and rewrite the term T_n as

$$T_n(u, x) = \overline{T}_n(u, x) + [T_n(u, x) - \overline{T}_n(u, x)] =: B_n(u, x) + W_n(u, x) \tag{5.2}$$

Thus

$$m_n(u, x) - m(u, x) = f_n(u, x)^{-1} (B_n(u, x) + W_n(u, x) + V_n(u, x)). \tag{5.3}$$

We begin by studying the variance term $V_n(u, x) + W_n(u, x)$ defined in (5.3) and in (5.2), through the next proposition :

Proposition 5.1. *Under the assumptions of Theorem 2.4, and conditions (P),(CD),(K),(S1) and (H), we have :*

$$\sup_{u \in [0,1], x \in S} |V_n(u, x) + W_n(u, x)| = O(a_n) = O_{a.s.} \left(\sqrt{\frac{\ln n}{nh^{d+1}}} \right). \quad (5.4)$$

Proof of Proposition 5.1. As $|V_n + W_n| \leq |V_n| + |W_n|$, let's start by the term V_n :

$$V_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i =: \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n Z_{i,n}$$

Observe that $Z_{i,n}$ is $\mathcal{F}_{i-1,n}$ -adapted and that since $\mathbb{E}(\epsilon_i | \mathcal{F}_{i-1,n}) = 0$, we have that :

$$\begin{aligned} \mathbb{E}(Z_{i,n} | \mathcal{F}_{i-2,n}) &= \mathbb{E} \left(K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i | \mathcal{F}_{i-2,n} \right) \\ &= K_h(u - \frac{i}{n}) \mathbb{E} \left(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i | \mathcal{F}_{i-2,n} \right) \\ &= K_h(u - \frac{i}{n}) \mathbb{E} \left(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \mathbb{E}(\epsilon_i | \mathcal{F}_{i-1,n}) | \mathcal{F}_{i-2,n} \right) = 0. \end{aligned}$$

This makes the $Z_{i,n}$ martingale differences with respect to the σ -algebras $\mathcal{F}_{i-1,n}$.

In order to bound uniformly V_n on the compact set S , we can suppose $S = \{x \in \mathbb{R}^d : \|x\|_\infty \leq C_S\}$.

To cover S , we consider a grid using regions (hypercubes) centered at points (u_k, x_k) of the form $\mathcal{B}_{n,k} = \{(u, x) \in \mathbb{R}^{d+1} : \|(u, x) - (u_k, x_k)\|_\infty \leq a_n\}$ with $a_n = \sqrt{\frac{\ln n}{nh^{d+1}}}$.

Selecting $(u_k, x_k) \in \mathcal{B}_{n,k}$ such that $\mathring{\mathcal{B}}_{n,i} \cap \mathring{\mathcal{B}}_{n,j} = \emptyset$ for each $i \neq j$ and $\mathcal{B}_{n,k} \subset S$, one may say that the set S can be covered by N closed hypercubes $\mathcal{B}_{n,k}$, with $N \leq \frac{C_S^{d+1}}{h^{d+1} a_n^{d+1}}$. Therefore,

$$\begin{aligned} \sup_{(u,x) \in \mathcal{B}_n} |V_n(u, x)| &\leq \max_{1 \leq k \leq N} \sup_{(u,x) \in \mathcal{B}_{n,k}} |V_n(u, x)| \\ &\leq \max_{1 \leq k \leq N} \sup_{(u, x) \in \mathcal{B}_{n,k}} |V_n(u, x) - V_n(u_k, x_k)| + \max_{1 \leq k \leq N} |V_n(u_k, x_k)|. \end{aligned} \quad (5.5)$$

Assumption (K2) implies that for any v_1, v_2 , we have

$$|v_1 - v_2| \leq \delta \leq C_k \implies |K(v_1) - K(v_2)| \leq \delta K^*(v_1),$$

where $K^*(v) = L_K \mathbb{I}_{\|v\|_\infty \leq 2C_k}$ for a kernel of dimension 1 (see, Hansen (2008)) and $\mathbf{K}^*(x) = \Lambda_K \mathbb{I}_{\{\|x\|_\infty \leq 2C_k\}}$ for the $d+1$ -dimensional kernel, with $\Lambda_K = (d+1) \bar{K}^d L_K$.

Note that for any $(u, x) \in \mathcal{B}_{n,k}$, we have $\frac{\|(u, x) - (u_k, x_k)\|}{h} \leq a_n$, we can choose $\delta = a_n \leq C_k$ for n large enough. It follows then, for $(u, x) \in \mathcal{B}_{n,k}$ and n sufficiently large, that

$$\left| K_h \left(u - \frac{i}{n} \right) \prod_{j=1}^n K_h \left(x^j - X_{i,n}^j \right) - K_h \left(u_k - \frac{i}{n} \right) \prod_{j=1}^n K_h \left(x_k^j - X_{i,n}^j \right) \right| \leq a_n \mathbf{K}_h^* \left(u_k - \frac{i}{n}, x_k - \mathbf{X}_{i,n} \right).$$

This allows us to write : $|Z_{i,n}(u, x) - Z_{i,n}(u_k, x_k)| \leq a_n \mathbf{K}_h^* \left(u_k - \frac{i}{n}, x_k - \mathbf{X}_{i,n} \right) |U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i|$.

By replacing $K(u)\mathbf{K}(x)$ with $\mathbf{K}^*(u, x)$ in V_n , we can define

$$\tilde{V}_n(u_k, x_k) := \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n \mathbf{K}_h^*(u_k - \frac{i}{n}, x_k - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i \quad (5.6)$$

Thus,

$$\begin{aligned} \sup_{(u,x) \in \mathcal{B}_{n,k}} |V_n(u, x)| &\leq \sup_{(u,x) \in \mathcal{B}_{n,k}} [|V_n(u, x) - V_n(u_k, x_k)| + |V_n(u_k, x_k)|] \\ &\leq |V_n(u_k, x_k)| + a_n |\tilde{V}_n(u_k, x_k)| \\ &\leq |V_n(u_k, x_k)| + |\tilde{V}_n(u_k, x_k)| \quad \text{since } a_n \leq 1 \quad \text{for } n \text{ large enough.} \end{aligned} \quad (5.7)$$

Consequently,

$$\begin{aligned} &\mathbb{P}\left(\sup_{(u,x) \in \mathcal{B}_n} |V_n(u, x)| \geq 2a_n\right) \\ &\leq N \max_{1 \leq k \leq N} \mathbb{P}(|V_n(u_k, x_k)| \geq a_n) + N \max_{1 \leq k \leq N} \mathbb{P}(|\tilde{V}_n(u_k, x_k)| \geq a_n) =: Q_{1,n} + Q_{2,n}. \end{aligned} \quad (5.8)$$

The two terms can be treated similarly. We focus our attention on the first one and denote by

$$V_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n Z_{i,n}(u, x), \quad (5.9)$$

where $Z_{i,n}(u, x) = K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i$.

In the most general case, the ϵ_i are not bounded. We have to deal with the tail part by truncation.

Let's consider a sequence t_n diverging to ∞ as $n \rightarrow \infty$ and satisfying :

$$t_n a_n = O(1) \quad \text{and} \quad \sum_{n \geq 1} t_n^{-s} < \infty \quad \text{and} \quad t_n^{1-s} = O(a_n). \quad (5.10)$$

Now we can decompose ϵ_i as follows : $\epsilon_i = \epsilon_i \mathbb{I}_{\{|\epsilon_i| < t_n\}} + \epsilon_i \mathbb{I}_{\{|\epsilon_i| \geq t_n\}} =: \epsilon_i^- + \epsilon_i^+$. Therefore,

$$Z_{i,n}(u, x) = Z_{i,n}^-(u, x) + Z_{i,n}^+(u, x), \quad \text{with} \quad Z_{i,n}^\pm(u, x) = K_h(u - \frac{i}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U(\frac{i}{n}, \mathbf{X}_{i-1,n}) \epsilon_i^\pm.$$

Next, we decompose $Z_{i,n}$ in three parts :

$$Z_{i,n}(u, x) = Z_{i,n}^+(u, x) + \left[Z_{i,n}^-(u, x) - \mathbb{E}\left(Z_{i,n}^-(u, x) | \mathcal{F}_{i-2,n}\right) \right] + \mathbb{E}\left(Z_{i,n}^-(u, x) | \mathcal{F}_{i-2,n}\right). \quad (5.11)$$

Summing up for $i = d+1, \dots, n$ and multiplying by $\frac{1}{(n-d)h^{d+1}}$ gives

$$V_n(u, x) = V_{n,1}(u, x) + V_{n,2}(u, x) + V_{n,3}(u, x). \quad (5.12)$$

The next lemmas will deal with each of these three terms beginning by the first one.

Lemma 5.2. *Under assumption (S1) and (5.10), we have :*

$\epsilon_i^+ = 0$, and thus $V_{n,1} = 0$ with probability 1 for sufficiently large n .

Proof. -

$$\mathbb{P}(|\epsilon_i| \geq t_n) \leq \frac{\mathbb{E}(|\epsilon_i|^s)}{t_n^s} \leq \frac{C}{t_n^s} \text{ by Markov inequality.}$$

$\sum_{n \geq 1} t_n^{-s} < \infty \implies |\epsilon_i| \leq t_n$ with probability 1 for sufficiently large n . The lemma follows. \blacksquare

The following lemma concerns the third term.

Lemma 5.3. *Under assumption (S1) and (5.10), we have :*

$$V_{n,3}(u, x) = O_{a.s.}(a_n).$$

Proof. -

For the sake of simplicity, we will omit the arguments (u_k, x_k) below. Since we know that $\mathbb{E}(Z_{i,n} | \mathcal{F}_{i-2,n}) = 0$ we can deduce that : $\mathbb{E}(Z_{i,n}^- | \mathcal{F}_{i-2,n}) = -\mathbb{E}(Z_{i,n}^+ | \mathcal{F}_{i-2,n})$. Thus

$$\left| \mathbb{E}(Z_{i,n}^- | \mathcal{F}_{i-2,n}) \right| = \left| \mathbb{E}(Z_{i,n}^+ | \mathcal{F}_{i-2,n}) \right| \leq \mathbb{E}\left(\left| Z_{i,n}^+ \right| | \mathcal{F}_{i-2,n}\right).$$

Using the boundedness of U , gives

$$\begin{aligned} \left| \mathbb{E}(Z_{i,n}^- | \mathcal{F}_{i-2,n}) \right| &\leq CK_h\left(u - \frac{i}{n}\right) \mathbb{E}(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) | \epsilon_i^+ | | \mathcal{F}_{i-2,n}) \\ &\leq CK_h\left(u - \frac{i}{n}\right) \mathbb{E}(\mathbb{E}(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) | \epsilon_i^+ | | \mathcal{F}_{i-1,n}) | \mathcal{F}_{i-2,n}) \\ &\leq CK_h\left(u - \frac{i}{n}\right) \mathbb{E}(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) \mathbb{E}(|\epsilon_i^+| | \mathcal{F}_{i-1,n}) | \mathcal{F}_{i-2,n}) \\ &\leq CK_h\left(u - \frac{i}{n}\right) \mathbb{E}(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) \mathbb{E}(|\epsilon_i^+|) | \mathcal{F}_{i-2,n}). \end{aligned}$$

Using assumptions (S1) and (CV) to obtain

$$\begin{aligned} \left| \mathbb{E}(Z_{i,n}^- | \mathcal{F}_{i-2,n}) \right| &\leq CK_h\left(u - \frac{i}{n}\right) \mathbb{E}\left(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) \times \mathbb{E}\left(\frac{|\epsilon_i^+|^s}{t_n^{s-1}}\right) | \mathcal{F}_{i-2,n}\right) \\ &\leq \frac{C}{t_n^{s-1}} K_h\left(u - \frac{i}{n}\right) \mathbb{E}(\mathbf{K}_h(x - \mathbf{X}_{i-1,n}) | \mathcal{F}_{i-2,n}) \\ &\leq \frac{C}{t_n^{s-1}} K_h\left(u - \frac{i}{n}\right) \int_{\mathbb{R}^d} \mathbf{K}_h(x - t) |f_{\mathbf{X}_{i-1,n}}^{\mathcal{F}_{i-2,n}}(t)| dt \\ &\leq \frac{C}{t_n^{s-1}} K_h\left(u - \frac{i}{n}\right) h^d \int_{\mathbb{R}^d} \mathbf{K}(y) |f_{\mathbf{X}_{i-1,n}}^{\mathcal{F}_{i-2,n}}(x - hy)| dy \leq C \frac{h^d}{t_n^{s-1}} K_h\left(u - \frac{i}{n}\right) \end{aligned}$$

Next we obtain :

$$V_{n,3}(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n \left| \mathbb{E}(Z_{i,n}^-(u, x) | \mathcal{F}_{i-2,n}) \right| \leq \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n C \frac{h^d}{t_n^{s-1}} K_h\left(u - \frac{i}{n}\right).$$

With the help of lemma 6.1, we then get : $|V_{n,3}(u, x)| \leq \frac{C}{t_n^{s-1}} = O(a_n)$. \blacksquare

We go back to the variance term V_n . To summarize, we have :

$$V_{n,1} = 0 \text{ with probability one for } n \text{ large enough and thus } V_n = V_{n,2} + V_{n,3} \text{ for } n \text{ large enough.}$$

On the other hand, we have $V_{n,3} = O_{a.s.}(a_n)$ and we can write for some $\alpha > 0$ and sufficiently large n , that, with probability one, we have

$$|V_n(u, x)| \leq |V_{n,2}(u, x)| + |V_{n,3}(u, x)| \leq |V_{n,2}(u, x)| + \alpha a_n.$$

Consequently, $\mathbb{P}(|V_n(u, x)| \geq 2\alpha a_n) \leq \mathbb{P}(|V_{n,2}(u, x)| \geq \alpha a_n)$.

Thus, we can replace $V_n(u_k, x_k)$ by $V_{n,2}(u_k, x_k)$ in the expression of $Q_{1,n}$.

Now, recall that $V_{n,2}(u_k, x_k) := \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n M_{i,n}$ where $M_{i,n} = Z_{i,n}^-(u_k, x_k) - \mathbb{E}\left(Z_{i,n}^-(u_k, x_k) | \mathcal{F}_{i-1,n}\right)$, and $Q_{1,n} = N \max_{1 \leq k \leq N} \mathbb{P}(|V_n(u_k, x_k)| \geq a_n)$.

We will state and prove the next lemma :

Lemma 5.4. *Under the assumptions of Theorem 2.4, conditions **(P)**, **(K1)**-**(K2)**, **(H1)**, then, we have for a fixed compact set $S \subset \mathbb{R}^d$, that*

$$\sum_{n \geq d+1} Q_{1,n} < \infty. \quad (5.13)$$

Proof. -

Recall that since $|\epsilon_i^-| \leq t_n$, U and K are bounded, we have $|Z_{i,n}^-| \leq Ct_n$ and also $|M_{i,n}| \leq Ct_n$. In order to apply the [Freedman \(1975\)](#) inequality we need to give an upper bound for the conditional variance of $M_{i,n}$ with the help of assumption **(CV)**. We will omit the arguments (u_k, x_k) and denote $\overline{Z_{i,n}^-} = \mathbb{E}\left(Z_{i,n}^-(u_k, x_k) | \mathcal{F}_{i-2,n}\right)$. We can thus write

$$\begin{aligned} v_i &= \mathbb{E}(M_{i,n}^2 | \mathcal{F}_{i-2,n}) = \mathbb{E}\left(\left(Z_{i,n}^- - \overline{Z_{i,n}^-}\right)^2 | \mathcal{F}_{i-2,n}\right) = \mathbb{E}\left(\left(Z_{i,n}^-\right)^2 | \mathcal{F}_{i-2,n}\right) - \left(\overline{Z_{i,n}^-}\right)^2 \\ &\leq \mathbb{E}\left(\left(Z_{i,n}^-\right)^2 | \mathcal{F}_{i-2,n}\right) \leq K_h^2\left(u - \frac{i}{n}\right) \mathbb{E}\left(\mathbf{K}_h^2(x_k - \mathbf{X}_{i-1,n}) \left(U\left(\frac{i}{n}, \mathbf{X}_{i-1,n}\right) \epsilon_i\right)^2 | \mathcal{F}_{i-2,n}\right) \\ &\leq K_h^2\left(u - \frac{i}{n}\right) \mathbb{E}\left(\mathbb{E}\left(\mathbf{K}_h^2(x_k - \mathbf{X}_{i-1,n}) \left(U\left(\frac{i}{n}, \mathbf{X}_{i-1,n}\right) \epsilon_i\right)^2 | \mathcal{F}_{i-1,n}\right) | \mathcal{F}_{i-2,n}\right) \\ &\leq C_U^2 K_h^2\left(u - \frac{i}{n}\right) \mathbb{E}\left(\mathbf{K}_h^2(x_k - \mathbf{X}_{i-1,n}) | \mathcal{F}_{i-2,n}\right) \leq C_U^2 K_h^2\left(u - \frac{i}{n}\right) \int_{\mathbb{R}^d} \mathbf{K}_h^2(x_k - t) f_{\mathbf{X}_{i-1,n}^{\mathcal{F}_{i-2,n}}}(t) dt. \end{aligned}$$

After the usual change of variable : $y = \frac{x-t}{h}$, we get

$$v_i \leq C_U^2 K_h^2\left(u - \frac{i}{n}\right) h^d \int_{\mathbb{R}^d} \mathbf{K}^2(y) f_{\mathbf{X}_{i-1,n}^{\mathcal{F}_{i-2,n}}}(x_k - hy) dy \leq C_U^2 f^* K_h^2\left(u - \frac{i}{n}\right) h^d.$$

Summing for $i = d+1, \dots, n$ leads to

$$\sum_{i=d+1}^n v_i = \sum_{i=d+1}^n \mathbb{E}(M_{i,n}^2 | \mathcal{F}_{i-2,n}) \leq C_U^2 f^* h^d \sum_{i=d+1}^n K_h^2\left(u - \frac{i}{n}\right)$$

We then apply Lemma 6.2 to get

$$\sum_{i=d+1}^n \mathbb{E}\left(\left(M_{i,n}^-\right)^2 | \mathcal{F}_{i-2,n}\right) \leq C_U f^* (n-d) h^{d+1} \left(\underline{\kappa}_2 + O\left(\frac{1}{nh^2}\right)\right) \leq C (n-d) h^{d+1}$$

for some constant $C > 0$.

We are now ready to apply [Freedman \(1975\)](#) inequality to the sequences $S_{n,k}^- = \sum_{i=d+1}^n Z_{i,n}^-(u_k, x_k)$

$$\begin{aligned} \mathbb{P}\left(|V_{n,2}(u_k, x_k)| > \alpha a_n\right) &= \mathbb{P}\left(|S_{n,k}^-| > (n-d)h^{d+1}\alpha a_n\right) \\ &\leq d \exp\left[-\frac{(n-d)^2\alpha^2 a_n^2 h^{2(d+1)}}{2C(n-d)h^{d+1} + \frac{2}{3}Ct_n(n-d)\alpha a_n h^{d+1}}\right] \\ &\leq d \exp\left[-\frac{(n-d)\alpha^2 a_n^2 h^{2(d+1)}}{2Ch^{d+1} + \frac{2}{3}t_n\alpha a_n h^{d+1}}\right] \end{aligned}$$

Knowing the value of a_n , the assumption **(H1)** and $t_n a_n = O(1)$ allow us to write

$$\mathbb{P}\left(|V_{n,2}(u_k, x_k)| > \alpha a_n\right) \leq d \exp\left[-\frac{(n-d)\alpha^2 a_n^2 h^{2(d+1)}}{C\alpha h^{d+1}}\right]$$

taking α large enough leads to :

$$\mathbb{P}\left(|V_{n,2}(u_k, x_k)| > \alpha a_n\right) \leq d \exp[-C\alpha \ln n] = dn^{-C\alpha}.$$

Back to term $Q_{1,n}$. Observe that

$$Q_{1,n} = N \max_{1 \leq k \leq N} \mathbb{P}\left(|V_{n,2}(u_k, x_k)| \geq \alpha a_n\right) \leq Ndn^{-C\alpha}$$

As $N = \left(\frac{C_S}{a_n h}\right)^{d+1}$, a choice of α sufficiently large leads to a general term of a convergent series.

This ends the proof of lemma [5.4](#). ■

The same techniques give the same result for $Q_{1,n}$ and as we have

$$\mathbb{P}\left(\sup_{(u,x) \in \mathcal{B}_n} |V_n(u, x)| \geq 2a_n\right) \leq Q_{1,n} + Q_{2,n} \quad (5.14)$$

we can conclude, thanks to the Borel-Cantelli lemma that

$$\sup_{(u,x) \in \mathcal{B}_n} |V_n(u, x)| = O_{p.s.}(a_n) \quad (5.15)$$

To deal with $W_n(u, x)$, recall that if $\omega_{i,n}(u, x) := K_h\left(u - \frac{i}{n}\right) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) \left[m\left(\frac{i}{n}, \mathbf{X}_{i-1,n}\right) - m(u, x) \right]$.

$$\text{Then } T_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n \omega_{i,n}(u, x).$$

Now, thanks to the assumptions **(M1)**-**(M3)** on m and **(K)** on the kernel K , we observe that $\omega_{i,n}(u, x)$ is bounded, has compact support and is L_ω -lipschitz with : $L_\omega = (d+1)L_K C_m + \bar{K}^{d+1} L_m$.

We can then write :

$$W_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n Y_{i,n} \quad \text{with} \quad Y_{i,n} = \omega_{i,n}(u, x) - \mathbb{E}\left[\omega_{i,n}(u, x) \middle| \mathcal{F}_{i-2,n}\right].$$

The terms $Y_{i,n}$ are triangular arrays of bounded martingale differences with respect to the sequences of triangular array σ -fields $(\mathcal{F}_{i-1,n})_{i \geq 2}$.

Proceeding as for the term $V_n(u, x)$, and without the need for truncation, one may write:

$$\sup_{u \in I_h, x \in S} |W_n(u, x)| = O\left(\sqrt{\frac{\ln n}{n h^{d+1}}}\right).$$

This ends the proof of Proposition 5.1.

We turn now to the study of the bias term $B_n(u, x)$ defined in (5.2)-(5.3). The asymptotic behavior of $B_n(u, x)$ is the subject of the following Lemma.

Lemma 5.5. *Define $I_h = [C_1 h, 1 - C_1 h]$, $r = \min(\rho, 1)$ and consider a fixed compact subset S . Then, under the assumptions of Theorem 2.4, and conditions **(P)**, **(CD)**, **(K)**, **(S1)** and **(H)**, we have :*

$$\sup_{u \in I_h, x \in S} |B_n(u, x)| = O_{a.s.}\left(\frac{1}{n^r h^d} + h^2\right).$$

Proof. -

We will start by using the same arguments as (Vogt (2012)).

Recall that : $B_n = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n \left\{ \mathbb{E} \left[\omega_{i,n}(u, x) \middle| \mathcal{F}_{i-2,n} \right] \right\}$.

As in Vogt's proof of part (iii) of Theorem 4.2 Vogt (2012), we consider the Lipschitz continuous function \tilde{K} defined on \mathbb{R} with support $[-qC_1, qC_1]$ for some constant $q > 1$, such that $\tilde{K}(x) = 1$ for all $x \in [-C_1, C_1]$. Let $I_h = [C_1 h, 1 - C_1 h]$. We note $\tilde{\mathbf{K}}$ the d -dimensional product kernel derived from \tilde{K} .

In order to give an upper bound of $|B_{n,2}(u, x)|$, we will use the following decomposition

$$B_n(u, x) = R_1(u, x) + R_2(u, x) + R_3(u, x) + R_4(u, x),$$

where $R_\ell(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - i/n) \mathbb{E}^{\mathcal{F}_{i-1,n}} \{q_\ell(u, x)\}$ with

$$\begin{aligned} q_1(u, x) = & \tilde{\mathbf{K}}_{\mathbf{h}}(x - \mathbf{X}_{i-1,n}) \left[\mathbf{K}_{\mathbf{h}}(x - \mathbf{X}_{i-1,n}) - \mathbf{K}_{\mathbf{h}}\left(x - \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) \right] \\ & \times \left[m\left(\frac{i-1}{n}, \mathbf{X}_{i-1,n}\right) - m(u, x) \right] \end{aligned}$$

$$\begin{aligned} q_2(u, x) = & \tilde{\mathbf{K}}_{\mathbf{h}}(x - \mathbf{X}_{i-1,n}) \mathbf{K}_{\mathbf{h}}\left(x - \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) \\ & \times \left[m\left(\frac{i-1}{n}, \mathbf{X}_{i-1,n}\right) - m\left(\frac{i-1}{n}, \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) \right] \end{aligned}$$

$$\begin{aligned} q_3(u, x) = & \left[\tilde{\mathbf{K}}_{\mathbf{h}}(x - \mathbf{X}_{i-1,n}) - \tilde{\mathbf{K}}_{\mathbf{h}}\left(x - \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) \right] \\ & \times \mathbf{K}_{\mathbf{h}}\left(x - \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) \left[m\left(\frac{i-1}{n}, \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) - m(u, x) \right] \end{aligned}$$

$$q_4(u, x) = \mathbf{K}_{\mathbf{h}}\left(x - \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) \left[m\left(\frac{i-1}{n}, \mathbf{X}_{i-1} \left(\frac{i-1}{n}\right)\right) - m(u, x) \right].$$

Following the proof of part (iii) of Vogt's Theorem 3 (which used the approximation of $\mathbf{X}_{i,n}$ by the stationary process $X_i(i/n)$, and the fact that K is a bounded Lipschitz function), we get, with $r = \min(1, \rho)$, that

$$\begin{aligned} \left| \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) - \mathbf{K}_h\left(x - \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right) \right| &\leq C \left| \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) - \mathbf{K}_h\left(x - \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right) \right|^r \\ &\leq CL_K^r \left\| \frac{\mathbf{X}_{i-1,n} - \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)}{h} \right\|^r \leq CL_K^r \left| \frac{1}{nh} V_{i,n} \right|^r \end{aligned}$$

On the other hand, the factor : $\widetilde{\mathbf{K}}_h(x - \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)) [m\left(\frac{i-1}{n}, \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right) - m(u, x)]$ can be bounded by Ch . Thus, after applying the conditional expectation, an upper bound for R_1 may be obtained:

$$\sup_{u \in I_h, x \in S} |R_1(u, x)| \leq \frac{1}{(n-d)h^{d+1}} (Ch) \sum_{i=d+1}^n \left(\frac{C}{nh}\right)^r \leq \frac{C}{n^r h^{d+r}} \quad \text{for some } 0 < C < \infty.$$

To deal with the term q_2 , observe that it is non equal to zero only if

$$\frac{x - \mathbf{X}_{i-1,n}}{h} \in [-qC_1, qC_1]^d \implies \mathbf{X}_{i-1,n} \in [x - qC_1h, x + qC_1h]$$

and

$$\frac{x - \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)}{h} \in [-C_1, C_1]^d \implies \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right) \in [x - C_1h, x + C_1h].$$

Using the fact that m is continuous, we can establish that the factor $[m\left(\frac{i-1}{n}, \mathbf{X}_{i-1,n}\right) - m\left(\frac{i-1}{n}, \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right)]$ is bounded. This allows us to write :

$$\left| m\left(\frac{i-1}{n}, \mathbf{X}_{i-1,n}\right) - m\left(\frac{i-1}{n}, \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right) \right| \leq C \left| m\left(\frac{i-1}{n}, \mathbf{X}_{i-1,n}\right) - m\left(\frac{i-1}{n}, \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right) \right|^r.$$

Now using the fact that m is Lipschitz, we get :

$$\left| m\left(\frac{i-1}{n}, \mathbf{X}_{i-1,n}\right) - m\left(\frac{i-1}{n}, \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right)\right) \right| \leq C \left\| \mathbf{X}_{i-1,n} - \mathbf{X}_{i-1}\left(\frac{i-1}{n}\right) \right\|^r \leq C \left| \frac{1}{nh} V_{i,n} \right|^r.$$

Thus, we can state that :

$$\sup_{u \in I_h, x \in S} |R_2(u, x)| \leq \frac{C}{n^r h^d}.$$

Using analogous arguments as for $q_1(u, x)$, we can further show that

$$\sup_{u, x} |R_3(u, x)| \leq \frac{C}{n^r h^{d+r}}.$$

Finally, applying Lemmas B1 and B2 and exploiting the smoothness conditions on m and f , we obtain that

$$R_4(u, x) = h^2 \frac{\kappa_2}{2} \sum_{i=0}^d [2\partial_i m(u, x) \partial_i f(u, x) + \partial_i^2 m(u, x) f(u, x)] + O(h^2)$$

uniformly in u and x . This ends the proof of Lemma (5.5). ■

5.2 Proof of Theorem 4.3

Using assumption (P2)(iii) together with Proposition 4.1, and the fact that

$$\frac{1}{\sup_{u \in I_h, x \in S} (\widehat{f}_n(u, x))^{-1}} = \inf_{u \in I_h, x \in S} |\widehat{f}_n(u, x)| \geq \inf_{u \in I_h, x \in S} f(u, x) - \sup_{u \in I_h, x \in S} |\widehat{f}_n(u, x) - f(u, x)|$$

we can write $\sup_{u \in I_h, x \in S} (\widehat{f}_n(u, x))^{-1} \leq \frac{1}{\delta - o_{a.s.}(1)} = O_{a.s.}(1)$.

Next, adding the result of Proposition 5.1 to the result of lemma (5.5) ends the proof of Theorem 4.3.

5.3 Proof of Theorem 4.4

Write:

$$\begin{aligned} U_n^2(u, x) - U^2(u, x) &= [U_n^2(u, x) - U^2(\frac{i}{n}, \mathbf{X}_{i-1, n})] + [U^2(\frac{i}{n}, \mathbf{X}_{i-1, n}) - U^2(u, x)] =: \frac{\tilde{N}_n(u, x)}{D_n(u, x)} \end{aligned} \quad (5.16)$$

where $U_n^2(u, x)$ is defined in (3.2) and $D_n(u, x) = f_n(u, x)$. It follows that

$$\tilde{N}_n(u, x) = f_n(u, x) (U_n^2(u, x) - U^2(u, x)).$$

After replacing $X_{i, n}$ by $m(\frac{i}{n}, \mathbf{X}_{i-1, n}) + U(\frac{i}{n}, \mathbf{X}_{i-1, n}) \epsilon_i$, we can rewrite $\tilde{N}_n(u, x)$ as the sum of 4 terms as follows

$$\begin{aligned} \tilde{N}_n(u, x) &= \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i-1}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1, n}) [m(\frac{i}{n}, \mathbf{X}_{i-1, n}) - m_n(\frac{i}{n}, \mathbf{X}_{i-1, n})]^2 \\ &+ \frac{2}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i-1}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1, n}) [m(\frac{i}{n}, \mathbf{X}_{i-1, n}) - m_n(\frac{i}{n}, \mathbf{X}_{i-1, n})] U(\frac{i}{n}, \mathbf{X}_{i-1, n}) \epsilon_i \\ &+ \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i-1}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1, n}) U^2(\frac{i}{n}, \mathbf{X}_{i-1, n}) (\epsilon_i^2 - 1) \\ &+ \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h(u - \frac{i-1}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1, n}) [U^2(\frac{i}{n}, \mathbf{X}_{i-1, n}) - U^2(u, x)] \\ &=: A_{1n} + A_{2n} + A_{3n} + A_{4n}. \end{aligned}$$

We need to evaluate the asymptotic behavior of each term in the last equality.

• **Study of the first term.** Observe that

$$\begin{aligned} |A_{1n}| &\leq \left[\sup_{u \in I_h, x \in S} |m(u, x) - m_n(u, x)| \right]^2 \frac{1}{(n-d)h^{d+1}} \left| \sum_{i=d+1}^n K_h(u - \frac{i-1}{n}) \mathbf{K}_h(x - \mathbf{X}_{i-1, n}) \right| \\ &\leq \left[\sup_{u \in I_h, x \in S} |m(u, x) - m_n(u, x)| \right]^2 |f_n(u, x)|. \end{aligned}$$

Now, with the help of Theorem 4.1, we can conclude that:

$$A_{1n} = O_{a.s.} \left(\left(\sqrt{\frac{\ln n}{n h^{d+1}}} + \frac{1}{n^r h^d} + h^2 \right)^2 \right) |f_n(u, x)|.$$

• **Study of the second term defined by**

$$A_{2n} = \frac{2}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h \left(u - \frac{i-1}{n} \right) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) [m \left(\frac{i}{n}, \mathbf{X}_{i-1,n} \right) - m_n \left(\frac{i}{n}, \mathbf{X}_{i-1,n} \right)] U \left(\frac{i}{n}, \mathbf{X}_{i-1,n} \right) \epsilon_i.$$

We have

$$|A_{2n}| \leq 2 \sup_{u \in I_h, x \in S} |m(u, x) - m_n(u, x)| \times |V_n|, \text{ where}$$

$$V_n(u, x) = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h \left(u - \frac{i-1}{n} \right) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U \left(\frac{i}{n}, \mathbf{X}_{i-1,n} \right) \epsilon_i.$$

Making use of Theorem 4.1. and Lemma 5.1, we can conclude that :

$$A_{2n} = O_{a.s.} \left(\sqrt{\frac{\ln n}{n h^{d+1}}} \right) \times O_{a.s.} \left(\sqrt{\frac{\ln n}{n h^{d+1}}} + \frac{1}{n^r h^d} + h^2 \right)$$

• **Study of the third term :**

$$A_{3n} = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h \left(u - \frac{i-1}{n} \right) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) U^2 \left(\frac{i}{n}, \mathbf{X}_{i-1,n} \right) (\epsilon_i^2 - 1).$$

With the help of the additionnal assumption **(S2)** and the fact that this term is a sum of martingale differences, we can proceed in the same way as in Lemma 5.1, and get :

$$A_{3n} = O_{a.s.} \left(\sqrt{\frac{\ln n}{n h^{d+1}}} \right)$$

• **Study of the fourth term :**

$$A_{4n} = \frac{1}{(n-d)h^{d+1}} \sum_{i=d+1}^n K_h \left(u - \frac{i-1}{n} \right) \mathbf{K}_h(x - \mathbf{X}_{i-1,n}) [U^2 \left(\frac{i}{n}, \mathbf{X}_{i-1,n} \right) - U^2(u, x)].$$

Assumption **(S3)**, combined with the compact support of K , implies

$$A_{4n} = O_{a.s.}(h^\gamma) |f_n(u, x)|.$$

The assumption $\inf_{u \in [0,1], x \in S} f(u, x) = \delta > 0$ and Proposition 4.1 guarantee that $\sup_{u \in I_h, x \in S} (f_n(u, x))^{-1} = O_{a.s.}(1)$, which completes the proof. ■

References

- Bardet, J.M. and Doukhan, P. (2017). Non-parametric estimation of time varying AR(1)–processes with local stationarity and periodicity. *Electronic Journal of Statistics*. 0: 1-8.
- Bardet, J. M, Doukhan, P. and Wintenberger, O. (2020). Contrast estimation of general locally stationary processes using coupling. *arXiv: Statistics Theory*.
- Bardet, J.M, Doukhan, P. and Wintenberger, O. (2022). Contrast estimation of time-varying infinite memory processes. *Stochastic Processes and their Applications*. 152: 10.1016/j.spa.2022.06.005.
- Dahlhaus, R. (1996a). On the Kullback-Leibler information divergence of locally stationary process. *Stoch. Proc. Appl.* 62: 139–68.

- Dahlhaus, R. (1996b). Maximum likelihood estimation and model selection for locally stationary processes. *J.Nonpar. Statist.* 6: 168–71.
- Dahlhaus, R.(1996c). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. *In Athens Conference on Applied Probability and Time Series Analysis, (Eds. P. M. Robinson and M. Rosenblatt) 2*: Springer, New York.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ain. Statist.* 25: 1–37.
- Dahlhaus, R., & Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis*, 19(6): 629-655.
- Dahlhaus, R., Neumann, M., and Von Sachs, R. (1997). Nonlinear Wavelet Estimation of Time-Varying Autoregressive Processes. SFB 373 Discussion Papers 1997,34, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.
- Dahlhaus, R. and Subba Rao, S. (2006). Statistical Inference for time-varying ARCH processes. *The Annals of Statistics*. **34**, No. 3, 1075–1114
- Dahlhaus, R. (2012). Locally Stationary Processes. *Handbook of Statistics*. 30: 351–413
- Dahlhaus, R., Richter, S. and Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*. 25(2): 1013–1044.
- Delecroix, M. and Rosa, A. (1996). Nonparametric estimation of a regression function and its derivatives under an ergodic hypothesis. *Journal of Nonparametric Statistics*. 6(4):367–382.
- Draghicesu, D., Guillas, S. and Wu, W.B. (2009). Quantile curve estimation and visualization for nonstationary time series. *J. Comput. Graph. Statist.* 18: 1–20
- Doukhan, P. 2018. Conference on Nonstationarity June 4–6. <http://urlz.fr/6CNu> is organized, AGM laboratory.
- Freedman, David A. (1975). On Tail Probabilities for Martingales. *The Annals of Probability*. 1: 100–178, 19.
- Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalised least squares estimation in time-varying ARCH models. *Annals of Statistics*. 36: 742–786.
- Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (2013). Nonparametric curve estimation from time series. *Nonparametric curve estimation from time series*. 60. Springer.
- Bruce E. Hansen. 2008. Uniform Convergence rates for kernel estimation with dependent data. *Econometric Theory*. **24** 726–748.
- Härdle, W., and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics*. 81(1): 223-242.

- Krengel, U. (1985). Ergodic theorems, *De gruyter studies in mathematics*, 6, Walter De Gruyter Berlin. New-York.
- Laïb, N. (2005). Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *Journal of Statistical Planning and Inference*. 134: 116–139.
- Laïb, N. and Louani, D. (2019). Asymptotic normality of kernel density function estimator from continuous time stationary and dependent processes. *Statistics and Probability Letters*. 145: 187–196.
- Laïb, N. and Louani, D. (2010). Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *Journal of Multivariate Analysis*. 01: 2266–2281. Saki2004
- Masry, E., and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear arch time series strong convergence and asymptotic normality: Strong convergence and asymptotic normality. *Econometric theory*. 11(2): 258-289.
- Mikosch, T. and Stărică , C. (2004). Non-stationarities in financial time series, the long-range dependence, and IGARCH effects. *The Review of Economics and Statistics*. 86: 378–390.
- Nason, G.P., von Sachs, R. and Kroisandt, G. (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. Series B*, 62: 271-292
- Richter, S., and Dahlhaus, R. (2019). Cross validation for locally stationary processes. *The Annals of Statistics*, 47(4): 2145–2173.
- Sakiyama, K., and Taniguchi, M. (2004). Discriminant analysis for locally stationary processes. *Journal of Multivariate Analysis*. 90(2): 282-300.
- Subba Rao, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* 38: 1155–1172.
- Truquet, L. (2019). Local stationrity and time-inhomogeneous Markov chains. *The Annals of Statistics.*, 47(4): 2023–2050.
- Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*. 40: No. 5, 2601-2633
- Wu, W. B., and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *The Annals of statistics*. 30(5): 1441-1459.
- Zhou, Z., and Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* 37: 2696–2729
- Wu, W.B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*. 21: 1397–1413

Zhou, Z. (2014). Nonparametric specification for non-stationary time series regression. *Bernoulli*. 20: 78-108.

Zhang, T. and Wu, W.B. (2015). Time-varying nonlinear regression models: nonparametric estimation and model selection. *The Annals of Statistics*. 43: No. 2, 741-768.

6 ANNEX A

In order to prove Proposition 4.1 as well as Theorem 4.3 and Theorem 4.4, we need the 2 following auxiliary Lemmas :

Lemma 6.1. *Suppose the Kernel K satisfies (K1)-(K2) and let $I_h = [C_k h, 1 - C_k h]$. Then, we have for $k = 0, 1, 2$,*

$$\sup_{u \in I_h} \left| \frac{1}{nh} \sum_{i=1}^n K_h \left(u - \frac{i}{n} \right) \left(\frac{u - i/n}{h} \right)^k - \int_0^1 \frac{1}{h} K_h(u - v) \left(\frac{u - v}{h} \right)^k dv \right| = O \left(\frac{1}{nh^2} \right).$$

Lemma 6.2. *Suppose the Kernel K satisfies (K1)-(K2) and let $g : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$, $(u, x) \mapsto g(u, x)$ be continuous differentiable w.r.t. u . Then, we have for a fixed compact set $S \subset \mathbb{R}^d$,*

$$\sup_{u \in I_h, x \in S} \left| \frac{1}{nh} \sum_{i=1}^n K_h \left(u - \frac{i}{n} \right) g \left(\frac{i}{n}, x \right) - g(u, x) \right| = O \left(\frac{1}{nh^2} \right) + o(h).$$

These lemmas are specially needed to prove the last part of Theorem 4.3 and to prove the consistency of the kernel estimator of the density (see annex B).

The proofs of these lemmas are straightforward using integral approximation by Riemann sums.

7 ANNEX B (PROOF OF PROPOSITION 4.1)

7.1 A lemma

Before proving theorem 4.1, we begin by introducing a useful lemma and give its proof.

Lemma 7.1. *Under assumptions (P), (K), (CD) and (H), we have :*

$$\mathbb{E}(\mathbf{K}_h(x - \mathbf{X}_{i,n}) | \mathcal{F}_{i-1,n}) = h^d \int_{\mathbf{X}_i(\frac{i}{n})}^{\mathcal{F}_{i-1,n}}(x) + O_{a.s.}(h^{d+2} + \frac{1}{n^r h^r}) \quad (7.1)$$

$$\mathbb{E}(\mathbf{K}_h^2(x - \mathbf{X}_{i,n}) | \mathcal{F}_{i-1,n}) = \kappa_2 h^d \int_{\mathbf{X}_i(\frac{i}{n})}^{\mathcal{F}_{i-1,n}}(x) + O_{a.s.}(h^{d+2} + \frac{1}{n^r h^r}) \quad (7.2)$$

where $r = \min(1, \rho)$, and ρ is given in Definition 2.1.

Proof. As in Vogt's proof of part (iii) of Theorem 4.2 Vogt (2012), we consider the Lipschitz continuous function \tilde{K} defined on \mathbb{R} with support $[-qC_k, qC_k]$ for some constant $q > 1$, such that $\tilde{K}(x) = 1$ for all $x \in [-C_k, C_k]$. We note $\tilde{\mathbf{K}}$ the d -dimensional product kernel derived from \tilde{K} .

We can thus decompose $\mathbf{K}_h(x - \mathbf{X}_{i,n})$ as the sum of 3 terms $\mathbf{K}_h(x - \mathbf{X}_{i,n}) = b_1 + b_2 + b_3$ with

$$\begin{aligned}
b_1 &= \widetilde{\mathbf{K}}_h(x - \mathbf{X}_{i,n}) [\mathbf{K}_h(x - \mathbf{X}_{i,n}) - \mathbf{K}_h(x - \mathbf{X}_i(\frac{i}{n}))] \\
b_2 &= \left[\widetilde{\mathbf{K}}_h(x - \mathbf{X}_{i,n}) - \widetilde{\mathbf{K}}_h(x - \mathbf{X}_i(\frac{i}{n})) \right] \mathbf{K}_h(x - \mathbf{X}_i(\frac{i}{n})) \\
b_3 &= \mathbf{K}_h(x - \mathbf{X}_i(\frac{i}{n}))
\end{aligned}$$

Using the approximation of $\mathbf{X}_{i,n}$ by the stationary process $X_i(i/n)$, and the fact that K is a bounded Lipschitz function), we get, with $r = \min(1, \rho)$, that

$$\begin{aligned}
\left| \mathbf{K}_h(x - \mathbf{X}_{i,n}) - \mathbf{K}_h\left(x - \mathbf{X}_i\left(\frac{i}{n}\right)\right) \right| &\leq C \left| \mathbf{K}_h(x - \mathbf{X}_{i,n}) - \mathbf{K}_h\left(x - \mathbf{X}_i\left(\frac{i}{n}\right)\right) \right|^r \\
&\leq C \left\| \frac{\mathbf{X}_{i,n} - \mathbf{X}_i\left(\frac{i}{n}\right)}{h} \right\|^r \leq C \left| \frac{1}{nh} V_{i,n} \right|^r
\end{aligned}$$

On the other hand, the factor : $\widetilde{\mathbf{K}}_h(x - \mathbf{X}_i(\frac{i}{n}))$ is bounded. Thus $|b_1| \leq C \left| \frac{1}{nh} V_{i,n} \right|^r$.

We obtain in the same way that $|b_2| \leq C \left| \frac{1}{nh} V_{i,n} \right|^r$.

Now for b_3 , thanks to the properties of the kernel, and using the classical change of variables, we have

$$\begin{aligned}
\mathbb{E}(b_3 | \mathcal{H}_{i-1,n}) &= \mathbb{E}\left(\mathbf{K}_h(x - \mathbf{X}_i\left(\frac{i}{n}\right)) | \mathcal{H}_{i-1,n}\right) = \int_{\mathbb{R}^d} \mathbf{K}_h(x - t) f_{\mathbf{X}_i(\frac{i}{n})}^{\mathcal{H}_{i-1,n}}(t) dt \\
&= h^d \int_{\mathbb{R}^d} \mathbf{K}(z) f_{\mathbf{X}_i(\frac{i}{n})}^{\mathcal{H}_{i-1,n}}(x - hz) dy = h^d f_{\mathbf{X}_i(\frac{i}{n})}^{\mathcal{H}_{i-1,n}}(x) + O_{a.s.}(h^{d+1})
\end{aligned}$$

Since $\mathcal{F}_{i,n} \subset \mathcal{H}_{i,n}$, the tower rule and the addition of the 3 terms gives the first result, and the same technique applied to \mathbf{K}^2 concludes the proof of the lemma. \blacksquare

7.2 Proof of theorem 4.1

In order to decompose $\widehat{f}_n(u, x) - f(u, x)$ into a sum of martingale differences and a bias term we introduce :

$$\overline{f}_n(u, x) = \frac{1}{(n-d+1)h^{d+1}} \sum_{i=d}^n \mathbb{E}^{\mathcal{H}_{i-1,n}} \left\{ K_h\left(u - \frac{i}{n}\right) K_h(x - \mathbf{X}_{i,n}) \right\}$$

This allows to write

$$\widehat{f}_n(u, x) - f(u, x) = \widehat{f}_n(u, x) - \overline{f}_n(u, x) + [\overline{f}_n(u, x) - f(u, x)] =: A_n + B_n. \text{ Thus}$$

$$\begin{aligned}
A_n &= \frac{1}{(n-d+1)h^{d+1}} \sum_{i=d}^n K_h\left(u - \frac{i}{n}\right) [K_h(x - \mathbf{X}_{i,n}) - \mathbb{E}^{\mathcal{H}_{i-1,n}}\{K_h(x - \mathbf{X}_{i,n})\}] \\
B_n &= \frac{1}{(n-d+1)h^{d+1}} \sum_{i=d}^n K_h\left(u - \frac{i}{n}\right) \mathbb{E}^{\mathcal{H}_{i-1,n}}\{K_h(x - \mathbf{X}_{i,n})\}
\end{aligned}$$

The term A_n can be handled in the same way as V_n in theorem 4.2, but without the need for truncation.

Thus we obtain $A_n = O_{a.s.}(a_n) = O_{a.s.}\left(\sqrt{\frac{\ln n}{n h^{d+1}}}\right) = o_{a.s.}(1)$.

The term B_n , on the other hand, can be decomposed, thanks to lemma 7.1, as

$$B_n = B_{n,1} + B_{n,2} + B_{n,3} - f(u, x)$$

with

$$B_{n,i} = \frac{1}{(n-d+1)h^{d+1}} \sum_{i=d}^n K_h \left(u - \frac{i}{n} \right) \mathbb{E}(b_i | \mathcal{H}_{i-1,n})$$

where the b_i are the same as in the preceding lemma.

Thanks to the proof of lemma 7.1, we can conclude that

$$B_n = \frac{1}{(n-d+1)h} \sum_{i=d}^n K_h \left(u - \frac{i}{n} \right) \left[f_{\mathbf{X}_i(\frac{i}{n})}^{\mathcal{H}_{i-1,n}}(x) + O_{a.s.}(h^2 + \frac{1}{n^r h^{d+r}}) \right]$$

With regard to theorem 2.3 and the compact support of K , we can say that

$$B_n = \frac{1}{(n-d+1)h} \sum_{i=d}^n K_h \left(u - \frac{i}{n} \right) \left[f_{\mathbf{X}_i(u)}^{\mathcal{H}_{i-1,n}}(x) + O_{a.s.}(h^p + h^2 + \frac{1}{n^r h^{d+r}}) \right]$$

The stationarity of $\mathbf{X}_i(u)$ allows to denote its density $f^{\mathcal{H}_{i-1,n}}(u, x) = f_{\mathbf{X}_i(u)}^{\mathcal{H}_{i-1,n}}(x) = f_{\mathbf{X}_0(u)}^{\mathcal{H}_{i-1,n}}(x)$

The first part of the sum defining B_n can be written

$$\begin{aligned} & \frac{1}{(n-d+1)h} \sum_{i=d}^n K_h \left(u - \frac{i}{n} \right) f^{\mathcal{H}_{i-1,n}}(u, x) \\ &= \frac{1}{(n-d+1)h} \sum_{i=d}^n \left[K_h \left(u - \frac{i}{n} \right) - h \right] f^{\mathcal{H}_{i-1,n}}(u, x) + \frac{1}{(n-d+1)} \sum_{i=d}^n f^{\mathcal{H}_{i-1,n}}(u, x) \\ & \quad =: P_n + Q_n \end{aligned}$$

Using the boundedness of $f^{\mathcal{H}_{i-1,n}}(u, x)$ we obtain

$$|P_n| = \left| \frac{1}{(n-d+1)h} \sum_{i=d}^n \left[K_h \left(u - \frac{i}{n} \right) - h \right] \right| f^{\mathcal{H}_{i-1,n}}(u, x) \leq C \left| \frac{1}{(n-d+1)h} \sum_{i=d}^n K_h \left(u - \frac{i}{n} \right) - 1 \right|.$$

It follows from lemmas 6.1, 6.2 that

$$P_n = O \left(\frac{1}{nh^2} \right) + o(h) = o(1).$$

On the other hand, the ergodicity of $\mathbf{X}_i(u)$ guarantees that

$$Q_n = \frac{1}{(n-d+1)} \sum_{i=d}^n f^{\mathcal{H}_{i-1,n}}(u, x) = f(u, x) + o_{a.s.}(1)$$

The conclusion is that $B_n = f(u, x) + o_{a.s.}(1)$. And finally :

$$\widehat{f}_n(u, x) - f(u, x) = o_{a.s.}(1), \quad \text{which completes the proof.} \quad \blacksquare$$