Article

Nonparametric Estimation of the Density Function of the Distribution of the Noise in CHARN Models

Joseph Ngatchou-Wandji ^{1,2,*}, Marwa Ltaifa ², Didier Alain Njamen Njomen ³ and Jia Shen ⁴¹ EHESP French School of Public Health, 35043 Rennes, France² Institut Élie Cartan de Lorraine, University of Lorraine, 54052 Vandoeuvre-Lès-Nancy, France; marwa.ltaifa@univ-lorraine.fr³ Department of Mathematics and Computer Science, Faculty of Science, University of Maroua, Maroua P.O. Box 814, Cameroon; didiernjamen1@gmail.com⁴ Department of Statistics, Fudan University, Shanghai 200433, China; jiashen99@gmail.com

* Correspondence: joseph.ngatchou-wandji@univ-lorraine.fr

Abstract: This work is concerned with multivariate conditional heteroscedastic autoregressive non-linear (CHARN) models with an unknown conditional mean function, conditional variance matrix function and density function of the distribution of noise. We study the kernel estimator of the latter function when the former are either parametric or nonparametric. The consistency, bias and asymptotic normality of the estimator are investigated. Confidence bound curves are given. A simulation experiment is performed to evaluate the performance of the results.

Keywords: nonlinear heteroscedastic model; kernel estimation

Citation: Ngatchou-Wandji, J.; Ltaifa, M.; Njamen Njomen, D.A.; Shen, J. Nonparametric Estimation of the Density Function of the Distribution of the Noise in CHARN Models. *Mathematics* **2022**, *10*, 624. <https://doi.org/10.3390/math10040624>

Academic Editor: Andrej Srakar

Received: 31 December 2021

Accepted: 14 February 2022

Published: 17 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The importance of multivariate time series is highlighted in a number of papers. They are widely studied in the econometric and statistical literature as can be seen, for instance, in [1–6]. Because of the interesting features of conditional heteroscedastic models in the analysis of financial time series, a lot of work has been devoted to the study of ARCH models [7–9] which, motivated by many reasons, have been generalized to CHARN models [10]. The CHARN model we deal with here is described as follows:

$$X_t = T(Y_t) + V(Y_t)\varepsilon_t, \quad t \in \mathbb{Z}, \quad (1)$$

where the sequence $\{X_t = (X_{t,1}, \dots, X_{t,k})^\top\}$ is a stationary and ergodic k -dimensional process; the $Y_t = (X_{t-1}^\top, X_{t-2}^\top, \dots, X_{t-p}^\top)^\top \in \mathbb{R}^{pk}$ are a pk -dimensional random vector; the $\varepsilon_t = (\varepsilon_{t,1}, \dots, \varepsilon_{t,k})^\top$, $t \in \mathbb{Z}$ are i.i.d. k -dimensional standard white noise with density function f ; and ε_t is uncorrelated with the σ -algebra spanned by the Y_s 's, $s < t$. The k -dimensional function $T(x) = (T_1(x), \dots, T_k(x))'$ and the $k \times k$ positive definite matrix function $V(x) = (V_{ij}(x) : 1 \leq i, j \leq k)$ are either of unknown or known forms but depend on unknown parameters. Note that the T_j 's and $V_{i,j}$'s, $1 \leq i, j \leq k$ are real-valued functions defined on $\mathbb{R}^{pk}$.

It is well-known (see, for example, [6]) that an estimator of the noise distribution can be used for testing model assumptions. This is often used in time series analysis at the model validation step (see, for example, [11], Section 5.3, p. 164), where one checks whether the residuals from the model fit behave in a consistent manner with the model. In this sense, when the current model is fitted to observations, one should check if the residuals behave like observations from iid random vectors with a specified density $f = f_0$. There are many existing tests, for the iid assumption (see, for example, [11]). If f_0 is the Gaussian density, one can use the Shapiro–Wilk test for testing the Gaussianity of the residuals. For a more general density, a simple test based on confidence bound functions for f can be used if these bounds can be constructed. Then, if the curve of the given f_0 lies between

the bounds, one would accept that $f = f_0$ and that the residuals are from f_0 . Confidence bounds can be obtained if one constructs a consistent and asymptotic normal estimator of f . The present work focuses on the study of the kernel estimator of f based on the residuals. All along the paper, we assume that this function is positive, uniformly continuous and twice differentiable, with continuous and bounded derivatives.

If the ε_t s are observed, the kernel estimator of f would be defined by

$$f_n(x) = \frac{1}{na_n^k} \sum_{t=1}^n K\left(\frac{x - \varepsilon_t}{a_n}\right), \quad x \in \mathbb{R}^k, \quad (2)$$

where a_n is the smoothing parameter and K a Parzen–Rosenblatt kernel ($\|x\|^k K(x) \rightarrow 0$, $\|x\|^k \rightarrow \infty$). Since these are never observed, they have to be substituted for estimates as, for example, residual series from parametric or nonparametric fits. In this paper, we study the asymptotic behavior of such an estimator. In Section 2, we write a short review on the kernel estimation of the density function. In Section 3, we study some asymptotic properties of $f_n(x)$ when the functions T and V are parametric and the ε_t s are substituted for the residuals from the fit. In Section 4, the same is done when the latter functions are nonparametric. We do not study the asymptotic normality in this case, as it is too technical and lengthy. The results are illustrated by a simulation experiment in Section 5, and Section 6 contains our conclusion. In order to have a self-contained paper, the external results used along the current work are gathered in Appendix A.

2. On the Kernel Density Estimation

2.1. A Short Review

During the last four decades, nonparametric functional estimation has undergone considerable developments. In particular, the estimation of the density of a probability law in $\mathbb{R}^k$, $k \geq 1$, has been studied by very different methods, some of which include Refs. [12–19].

The first work in this field goes back to [20,21], which focused on the solution densities of a differential equation:

$$\frac{dg}{dx} = \frac{(x-a)g}{b_0 + b_1x + b_2x^2},$$

where the constants a , b_0 , b_1 and b_2 are expressed as functions of the first four moments of the distribution of the density g . For estimating g , the technique was to estimate the moments of this distribution.

Subsequently, many other methods have been proposed. Because of its ease, the kernel method received the most attention. This method was introduced by [22], from the idea of Fix and Hodges invoked in [23]. The idea consisted in estimating the density g at a point x by counting the number of observations located in the interval $[x-a, x+a]$ for $a > 0$. Rosenblatt proposed a more general estimator defined by

$$g_n(x) = \frac{1}{na_n} \sum_{i=1}^n K\left(\frac{x - U_i}{a_n}\right), \quad x \in \mathbb{R}, \quad (3)$$

where K is a probability density on $\mathbb{R}$, (a_n) is a sequence of positive real numbers, and $U_1, \dots, U_n$ is a sample from g . In the literature, g_n is commonly referred to as the Parzen–Rosenblatt estimator [22,24]. Its extension to the multivariate case was studied by [25] who proposed as the estimator

$$g_n(x) = \frac{1}{na_n^k} \sum_{i=1}^n K\left(\frac{x - U_i}{a_n}\right), \quad x \in \mathbb{R}^k, \quad k \geq 1, \quad (4)$$

where K is a probability density on $\mathbb{R}^k$ and the U_i s are k -dimensional samples.

The scaling factor a_n^k appearing in the denominator in the right-hand side of (4) can be seen as the determinant of a $k \times k$ diagonal matrix with the a_n s on the main diagonal. Moreover, the argument of the function K can also be seen as the transformation of $x - U_i$ by the inverse of this matrix. From these observations [14], instead of considering a diagonal matrix, we rather considered a full matrix and defined the estimator

$$\tilde{g}_n(x) = \frac{1}{n|\det(H_n)|} \sum_{i=1}^n K(H_n^{-1}(x - U_i)), \quad x \in \mathbb{R}^k,$$

where H_n is an invertible $k \times k$ matrix with determinant $\det(H_n)$.

Ref. [26] introduced a new class of kernel estimators of the density of a probability distribution on $\mathbb{R}^k$, $k \geq 1$. The originality of these estimators compared to the previous ones is that the usual bandwidth a_n is a function of the observations and its moments of orders 1 and 2 coincide with the empirical moments of the law to be estimated. However, this study was done for the normal kernel only.

For more on the kernel density estimation, the reader can refer to the monographs of [13,16,18,19,27,28]. A more recent book is that of [29], which also gathers numerous recent references on the subject.

2.2. Some Properties of the Kernel Estimator

The method which seems convenient, simple to implement, robust and does not require a multiple choice of parameters is the kernel method, which has attracted the most interest since [22,24]. The properties of the resulting kernel estimator g_n have been largely studied in the literature.

Several modes of convergence of the kernel estimator g_n to g have been studied. Ref. [24] for $k = 1$ and [25] for $k > 1$, have established the convergence in probability. That is, as $n \rightarrow \infty$, in probability,

$$\sup_{x \in \mathbb{R}^k} |g_n(x) - g(x)| \longrightarrow 0,$$

and the convergence to 0 of the mean square error (MSE)

$$MSE = \mathbb{E}\{[g_n(x) - g(x)]^2\}.$$

Ref. [30] obtained, for $k = 1$, the convergence to 0 of the mean integrated squared error (MISE)

$$MISE = \int_{\mathbb{R}^k} \mathbb{E}\{[g_n(x) - g(x)]^2\} dx.$$

Ref. [18] stated and proved various results on the properties of the kernel estimator g_n . Most of them are about its pointwise and uniform consistency, its asymptotic bias and variance, or its asymptotic normality. Those of these results used to establish our results are recalled in the last section.

2.3. Bandwidth Selection

Bandwidth selection is an important issue in the kernel estimation of density functions. It is crucial to the estimation accuracy. However, it is a difficult problem which has been the subject of a large amount of work (see, for example, [31] for a review). Unlike data-driven methods, such as least square or cross-validation selection (see, for example, [32]), the plug-in methods demonstrate their simplicity and adequacy in the symmetric density shape case. The following plug-in method is proposed by [14,33]. Resulted from the optimization of the MISE above in certain cases, their bandwidth a_n is taken to be

$$a_n = \left(\frac{\int K^2(t) dt}{(\int t^2 K(t) dt)^2} \right)^{1/5} \frac{S_n}{(3/8\sqrt{\pi})^{1/5} n^{1/5}},$$

where $S_n^2 = n^{-1} \sum_{i=1}^n (U_i - \bar{U})^2$ is the sample variance. With the Epanechnikov kernel, this bandwidth is approximately defined as

$$a_n = 1.0486 \frac{S_n}{n^{1/5}}.$$

It is this bandwidth that is further used in the numerical study in Section 5 of this paper.

3. The Semi-Parametric Models Case

Here, we consider model (1) with the aim of estimating f by a kernel method in the case $T(x) = M(\theta; x)$ and $V(x) = \Sigma(\theta; x)$ for some k -dimensional vector M and $k \times k$ matrix Σ functions of known forms, with the unknown parameter $\theta = (\theta_1, \dots, \theta_q)'$ lying within a subset Θ of $\mathbb{R}^q$. More precisely, we study some of the asymptotic properties of $f_n(x)$ defined by (2) when the ε_i s are substituted for the residuals from the fit. This requires the estimation of the parameter vector θ . In the following subsection, we briefly review some recent results on this estimation.

3.1. The Parameter Estimation

There is an increasing body of work on the parameter estimation of many classes of models contained within CHARN models (1). In this subsection, we give an overview of some of them.

The main estimators encountered in the time series literature are of least squares (LS) or of the likelihood type. Besides these, are some other less popular ones. Ref. [34] constructed a Godambe's estimator, which they denoted by G , and proved that it is more efficient than the conditional least-squares estimator (CL) for autoregressive conditional heteroscedasticity (ARCH) models. In [35], the efficiency of the CL- and G -estimators for the CHARN model was studied. Ref. [35] deduced the condition according to which the estimator G is asymptotically optimal. The author proved the asymptotic normality of these two estimators and concluded that the G -estimator is less efficient than the CL-estimator.

The aim of [36] was to study the estimation function theory proposed by [37–39] for CHARN models. More precisely, Ref. [36] proposed a general estimator of the estimation function and derived its asymptotic distribution.

Ref. [40] introduced a method for the parameters estimation of ARCH models. They derived the optimal estimating function (EF) by combining linear and quadratic estimating functions. They proved that this estimator is more efficient than the quasi-maximum-likelihood estimator and, under the conditional normality assumption, the constructed EF-estimator is identical to that obtained by using the maximum-likelihood method in finite samples. The simulation results show that the finite-sample properties of the EF-estimators are attractive. The latter tends to display less bias and root mean squared error than the quasi-maximum likelihood estimator.

Ref. [41] studied parameter estimation in a class of CHARN models. He investigated the properties of the conditional least-squares estimators and those of the conditional likelihood estimators. He defined the kernel estimators of noise density and its derivatives, and showed that they are uniformly consistent.

Several papers discussed the problem of estimation and inference in the exponential autoregressive models (EXPAR) introduced by [42]. Among others, one can cite [43–48].

3.2. The Kernel Estimation of the Density Function of the Noise

Denote by $\|U\|$ an appropriate norm of a vector or a matrix U , and by $\frac{\partial U}{\partial \theta}(\theta; z)$ the matrix of the partial derivatives with respect to θ , of the entries of a vector or matrix function $U(\theta; z)$. In the sequel, we restrict ourselves to identifiable models (1) for which the following hold:

(A1) For all $\theta \in \Theta \subset \mathbb{R}^q$ and $z \in \mathbb{R}^{pk}$, $\Sigma(\theta; z)$ is invertible with inverse $\Sigma^{-1}(\theta; z)$.

(A2) The functions $\theta \mapsto M(\theta; z)$, $\theta \mapsto \Sigma(\theta; z)$ and $\theta \mapsto \Sigma^{-1}(\theta; z)$ are each continuously differentiable with respect to $\theta \in \text{int}(\Theta)$. There exists a finite positive number r such that the closed ball $\bar{B}(\theta_0, r)$ is contained within $\text{int}(\Theta)$, and a positive function $\gamma(z)$ with $\mathbb{E}[\gamma^{2+\beta}(Y_0)] < \infty$ such that

$$\max \left\{ \sup_{\theta \in \bar{B}(\theta_0, r)} \left\| \frac{\partial M}{\partial \theta}(\theta; z) \right\|, \sup_{\theta \in \bar{B}(\theta_0, r)} \left\| \frac{\partial \Sigma}{\partial \theta}(\theta; z) \right\|, \sup_{\theta \in \bar{B}(\theta_0, r)} \left\| \frac{\partial \Sigma^{-1}}{\partial \theta}(\theta; z) \right\| \right\} \leq \gamma(z).$$

(A3) The true parameter $\theta_0 \in \Theta$ has a consistent estimator $\tilde{\theta}_n = (\tilde{\theta}_{n,1}, \dots, \tilde{\theta}_{n,q})^\top$ satisfying

$$E[|\sqrt{n}(\tilde{\theta}_n - \theta_0)|^2] \rightarrow \rho < \infty, \quad n \rightarrow \infty$$

and

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, \Delta),$$

where “ $\Rightarrow$ ” denotes the convergence in distribution, and Δ is a $q \times q$ positive definite matrix.

Remark 1. Assumptions (A1)–(A2) hold at least for some usual stationary multivariate time series models, such as vector ARMA, vector EXPAR or vector ARCH models. Assumption (A3) holds for the estimators $\tilde{\theta}_n$ satisfying the so-called Bahadur representation, as the likelihood, pseudo-likelihood, least-squares, conditional likelihood or conditional least-square estimators.

From model (1), by (A1), define for any $t \in \mathbb{Z}$, and for any $\theta \in \Theta$,

$$\varepsilon_t(\theta) = \Sigma^{-1}(\theta; Y_t)[X_t - M(\theta; Y_t)], \quad \theta \in \Theta, \quad \text{and} \quad \tilde{\varepsilon}_t = \varepsilon_t(\tilde{\theta}_n). \quad (5)$$

Proposition 1. Assume that assumptions (A1)–(A3) hold. Then, for any $t \in \mathbb{Z}$,

$$\tilde{\varepsilon}_t \xrightarrow{P} \varepsilon_t(\theta_0), \quad n \rightarrow \infty.$$

Proof. For any $t \in \mathbb{Z}$, one can write

$$\begin{aligned} \tilde{\varepsilon}_t - \varepsilon_t(\theta_0) &= \varepsilon_t(\tilde{\theta}_n) - \varepsilon_t(\theta_0) \\ &= \Sigma^{-1}(\tilde{\theta}_n; Y_t)[X_t - M(\tilde{\theta}_n; Y_t)] - \Sigma^{-1}(\theta_0; Y_t)[X_t - M(\theta_0; Y_t)] \\ &= \Sigma^{-1}(\tilde{\theta}_n; Y_t)[M(\theta_0; Y_t) - M(\tilde{\theta}_n; Y_t)] - [\Sigma^{-1}(\theta_0; Y_t) - \Sigma^{-1}(\tilde{\theta}_n; Y_t)][X_t - M(\theta_0; Y_t)] \\ &= \Sigma^{-1}(\tilde{\theta}_n; Y_t)[M(\theta_0; Y_t) - M(\tilde{\theta}_n; Y_t)] - [\Sigma^{-1}(\theta_0; Y_t) - \Sigma^{-1}(\tilde{\theta}_n; Y_t)]\Sigma(\theta_0; Y_t)\varepsilon_t(\theta_0). \end{aligned}$$

Applying a first-order Taylor expansion to the differences within the above couples of brackets, since $\tilde{\theta}_n$ is consistent to θ_0 , the desired result follows immediately by assumptions (A2) and (A3).

Let $\tilde{K}$ be an almost-everywhere-differentiable Parzen–Rosenblatt kernel with gradient function $\frac{\partial \tilde{K}}{\partial x}(x)$ at x . One considers the kernel estimator of f defined by

$$\tilde{f}_n(x) = \frac{1}{na_n^k} \sum_{t=1}^n \tilde{K}\left(\frac{x - \tilde{\varepsilon}_t}{a_n}\right), \quad x \in \mathbb{R}^k,$$

where the kernel function $\tilde{K}$ and the bandwidth a_n satisfy the following assumptions:

(A4) $\tilde{K}$ is positive, even, and has compact support and bounded variation.

(A5) $a_n \rightarrow 0$, $\sqrt{na_n^k} \rightarrow \infty$, $n \rightarrow \infty$.

(A6) For all $\ell \geq 0$ and $x, y \in \mathbb{R}^k$, $a_n^{-\ell} y^\top \frac{\partial \tilde{K}}{\partial x}(xa_n^{-1}) \rightarrow 0$, as $n \rightarrow \infty$, and $y \rightarrow y_0$, and the sequence of functions $a_n^{-\ell} \frac{\partial \tilde{K}}{\partial x}(xa_n^{-1})$ is bounded.

□

Remark 2. It can be easily verified that assumptions (A4) and (A6) hold at least for Gaussian and Epanechnikov kernels. As a consequence of (A4), $\int \|y\|^2 \tilde{K}(y) dy < \infty$. Assumption (A5) implies that $na_n^k / \log(n) \rightarrow \infty$ as $n \rightarrow \infty$.

3.2.1. The Consistency of $\tilde{f}_n$ to f

Proposition 2. Assume that the assumptions (A1)–(A6) are satisfied. Then,

$$\sup_{x \in \mathbb{R}^k} |\tilde{f}_n(x) - f(x)| = o_P(1).$$

Proof. For any $t \in \mathbb{Z}$, recall the form of the $k \times q$ matrix $\frac{\partial \varepsilon_t}{\partial \theta}$:

$$\frac{\partial \varepsilon_t}{\partial \theta} = \begin{pmatrix} \frac{\partial^\top \varepsilon_{t,1}}{\partial \theta} \\ \vdots \\ \frac{\partial^\top \varepsilon_{t,k}}{\partial \theta} \end{pmatrix},$$

where for any, $t \in \mathbb{Z}$, $1 \leq j \leq k$,

$$\frac{\partial^\top \varepsilon_{t,j}}{\partial \theta} = \left(\frac{\partial \varepsilon_{t,j}}{\partial \theta_1}, \dots, \frac{\partial \varepsilon_{t,j}}{\partial \theta_q} \right).$$

For any $x \in \mathbb{R}^k$,

$$\begin{aligned} |\tilde{f}_n(x) - f_n(x)| &= \left| \frac{1}{na_n^k} \sum_{t=1}^n \left[\tilde{K}\left(\frac{x - \tilde{\varepsilon}_t}{a_n}\right) - \tilde{K}\left(\frac{x - \varepsilon_t}{a_n}\right) \right] \right| \\ &\leq \frac{1}{na_n^k} \sum_{t=1}^n \left| \tilde{K}\left(\frac{x - \tilde{\varepsilon}_t}{a_n}\right) - \tilde{K}\left(\frac{x - \varepsilon_t}{a_n}\right) \right| \\ &\leq \frac{1}{n} \sum_{t=1}^n \left| \frac{1}{a_n^{k+1}} (\tilde{\varepsilon}_t - \varepsilon_t)^\top \frac{\partial \tilde{K}}{\partial x}\left(\frac{x - \dot{\varepsilon}_t}{a_n}\right) \right|, \end{aligned}$$

where $\dot{\varepsilon}_t$ lies between $\tilde{\varepsilon}_t$ and ε_t .

Since for $t \in \mathbb{Z}$, $\tilde{\varepsilon}_t$ tends in probability to ε_t as n tends to infinity, in view of our assumptions, one has that, in probability,

$$\left| \frac{1}{a_n^{k+1}} (\tilde{\varepsilon}_t - \varepsilon_t)^\top \frac{\partial \tilde{K}}{\partial x}\left(\frac{x - \dot{\varepsilon}_t}{a_n}\right) \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

By a stochastic version of the Cesàro theorem (see Appendix A), one has that, in probability, the right-hand side of the last inequality tends to 0 as n tends to infinity. Consequently, in probability,

$$\sup_{x \in \mathbb{R}^k} |\tilde{f}_n(x) - f_n(x)| \rightarrow 0, \quad n \rightarrow \infty.$$

Now, by Theorem 1 in Appendix A.1, it follows that, in probability,

$$\sup_{x \in \mathbb{R}^k} |f_n(x) - f(x)| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Whence, in probability,

$$\sup_{x \in \mathbb{R}^k} |\tilde{f}_n(x) - f(x)| \longrightarrow 0 \text{ as } n \rightarrow \infty.$$

□

3.2.2. The Bias Study

Proposition 3. Assume that assumptions (A1)–(A6) hold. Then,

$$\forall x \in \mathbb{R}^k, \quad \mathbb{E}[\tilde{f}_n(x)] - f(x) \longrightarrow 0, \quad n \rightarrow \infty.$$

Proof. For all $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, define the function

$$\chi_k(x) = \sum_{1 \leq i, j \leq k} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int t_i t_j K(t_1, \dots, t_k) dt_1 \dots dt_k. \quad (6)$$

Then for all $x = (x_1, \dots, x_k) \in \mathbb{R}^k$,

$$\begin{aligned} \mathbb{E}[\tilde{f}_n(x)] - f(x) &= \mathbb{E}[\tilde{f}_n(x) - f_n(x)] + [\mathbb{E}(f_n(x)) - f(x)] \\ &= \frac{1}{na_n^k} \sum_{t=1}^n \mathbb{E} \left[\tilde{K} \left(\frac{x - \tilde{\varepsilon}_t}{a_n} \right) - \tilde{K} \left(\frac{x - \varepsilon_t}{a_n} \right) \right] \\ &\quad + \frac{a_n^2}{2} \chi_k(x) + o(a_n^2). \end{aligned}$$

For all $t \in \mathbb{Z}$ and $x \in \mathbb{R}^k$, define the functions

$$\begin{aligned} d_t(x) &= \frac{1}{a_n^k} \mathbb{E} \left[\tilde{K} \left(\frac{x - \tilde{\varepsilon}_t}{a_n} \right) - \tilde{K} \left(\frac{x - \varepsilon_t}{a_n} \right) \right] \\ w_n(x) &= \frac{1}{n} \sum_{t=1}^n d_t(x). \end{aligned}$$

With these, one can write, for all $x \in \mathbb{R}^k$,

$$\mathbb{E}[\tilde{f}_n(x)] - f(x) = w_n(x) + \frac{a_n^2}{2} \chi_k(x) + o(a_n^2).$$

Now, we show that for all $x \in \mathbb{R}^k$, in probability, $w_n(x) \longrightarrow 0$ as n tends to ∞ .

By a first-order Taylor expansion of the function $\tilde{K}$, one obtains for all $x \in \mathbb{R}^k$ and $t \in \mathbb{Z}$,

$$\tilde{K} \left(\frac{x - \tilde{\varepsilon}_t}{a_n} \right) - \tilde{K} \left(\frac{x - \varepsilon_t}{a_n} \right) = \frac{1}{a_n} (\tilde{\varepsilon}_t - \varepsilon_t)^\top \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right),$$

where for all $t \in \mathbb{Z}$, $\dot{\varepsilon}_t$ lies between ε_t and $\tilde{\varepsilon}_t$.

By another first-order Taylor expansion applied to $\varepsilon_t = \varepsilon_t(\theta)$, one then obtains, for some $\dot{\theta}_n = (\dot{\theta}_{1n}, \dots, \dot{\theta}_{qn})'$ lying between $\tilde{\theta}_n$ and θ ,

$$\tilde{K} \left(\frac{x - \tilde{\varepsilon}_t}{a_n} \right) - \tilde{K} \left(\frac{x - \varepsilon_t}{a_n} \right) = \frac{1}{a_n} \left(\frac{\partial \varepsilon_t}{\partial \theta}(\dot{\theta}_n) (\tilde{\theta}_n - \theta) \right)^\top \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right)$$

It results that, for all $t \in \mathbb{Z}$ and $x \in \mathbb{R}^k$,

$$d_t(x) = \frac{1}{\sqrt{na_n^k}} \mathbb{E} \left[\frac{1}{a_n} \left(\frac{\partial \varepsilon_t}{\partial \theta}(\dot{\theta}_n) \sqrt{n} (\tilde{\theta}_n - \theta) \right)^\top \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right) \right].$$

By the Cauchy–Schwarz inequality, one has for all $t \in \mathbb{Z}$ and $x \in \mathbb{R}^k$,

$$|d_t(x)| \leq \frac{1}{\sqrt{na_n^k}} \mathbb{E}^{1/2} \left\{ \left[\frac{\partial^\top \varepsilon_t}{\partial \theta} (\dot{\theta}_n) \frac{1}{a_n} \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right) \right]^2 \right\} \mathbb{E}^{1/2} \left(\|\sqrt{n}(\tilde{\theta} - \theta)\|^2 \right).$$

By our assumptions and the dominated convergence theorem, it is easy to see that for all $t \in \mathbb{Z}$, $x \in \mathbb{R}^k$,

$$d_t(x) \longrightarrow 0, \quad n \rightarrow \infty.$$

Whence, by Cesàro's theorem (see Appendix A), one has that for all $x \in \mathbb{R}^k$,

$$w_n(x) = \frac{1}{n} \sum_{t=1}^n d_t(x) \longrightarrow 0, \quad n \rightarrow \infty.$$

Given that for all $x \in \mathbb{R}^k$, $\frac{a_n^2}{2} \chi_k(x) + o(a_n^2)$ tends to 0 as n tends to ∞ , one can then conclude that for all $x \in \mathbb{R}^k$,

$$\mathbb{E}[\tilde{f}_n(x)] - f(x) = w_n(x) + \frac{a_n^2}{2} \chi_k(x) + o(a_n^2) \longrightarrow 0, \quad n \rightarrow \infty.$$

This shows that for all $x \in \mathbb{R}^k$, $\tilde{f}_n(x)$ is an asymptotically unbiased estimator of $f(x)$. $\square$

3.2.3. Asymptotic Normality

Proposition 4. Assume that (A1)–(A6) hold. Then

$$\forall x \in \mathbb{R}^k, \quad \sqrt{na_n^k} \left(\frac{\tilde{f}_n(x) - f(x)}{(f(x) \int \tilde{K}^2(u) du)^{1/2}} \right) \Longrightarrow \mathcal{N}(0, 1).$$

Proof. One has for all $x \in \mathbb{R}^k$,

$$\sqrt{na_n^k} [\tilde{f}_n(x) - f(x)] = \sqrt{na_n^k} [\tilde{f}_n(x) - f_n(x)] + \sqrt{na_n^k} [f_n(x) - f(x)].$$

We first study the first term in the right-hand side of the above equality.

For all $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, by applications of Taylor expansions to $\tilde{K}$ and ε_t respectively, one can write

$$\begin{aligned} \sqrt{na_n^k} [\tilde{f}_n(x) - f_n(x)] &= \frac{1}{\sqrt{na_n^k}} \sum_{t=1}^n \left[\tilde{K} \left(\frac{x - \tilde{\varepsilon}_t}{a_n} \right) - \tilde{K} \left(\frac{x - \varepsilon_t}{a_n} \right) \right] \\ &= \frac{1}{\sqrt{na_n^k}} \sum_{t=1}^n \frac{1}{a_n} (\varepsilon_t - \tilde{\varepsilon}_t)^\top \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right). \\ &= \frac{1}{\sqrt{na_n^k}} \sum_{t=1}^n \frac{1}{a_n} \left(\frac{\partial \varepsilon_t}{\partial \theta} (\dot{\theta}_n) (\tilde{\theta} - \theta) \right)^\top \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right). \\ &= \frac{1}{n} \sum_{t=1}^n \frac{1}{a_n^{1+k/2}} \left(\frac{\partial \varepsilon_t}{\partial \theta} (\dot{\theta}_n) \sqrt{n} (\tilde{\theta} - \theta) \right)^\top \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right), \end{aligned}$$

where $\dot{\varepsilon}_t$ lies between ε_t and $\tilde{\varepsilon}_t$, and $\dot{\theta}_n$ lies between $\tilde{\theta}_n$ and θ .

Define for all $t \in \mathbb{Z}$ and all $x \in \mathbb{R}^k$,

$$l_t(x) = \frac{1}{a_n^{1+k}} \left| \frac{\partial \varepsilon_t}{\partial \theta} (\dot{\theta}_n) \frac{\partial \tilde{K}}{\partial x} \left(\frac{x - \dot{\varepsilon}_t}{a_n} \right) \right|.$$

By our assumptions, for any $x \in \mathbb{R}^k$, $l_t(x)$ tends in probability to 0 as n tends to infinity. Now, it is easy to see that

$$\left| \sqrt{na_n^k} [\tilde{f}_n(x) - f_n(x)] \right| \leq \frac{1}{n} \sum_{t=1}^n l_t(x) \|\sqrt{n}(\tilde{\theta}_n - \theta)\| \sqrt{a_n}.$$

By assumption (A3), $\|\sqrt{n}(\tilde{\theta}_n - \theta)\|$ is tight. Thus, for all $t \in \mathbb{Z}$ and all $x \in \mathbb{R}^k$, $l_t(x) \|\sqrt{n}(\tilde{\theta}_n - \theta)\| \sqrt{a_n}$ tends in probability to 0 as n tends to infinity. By a stochastic version of the Cesàro theorem (see Appendix A.4), for all $x \in \mathbb{R}^k$, in probability,

$$\left| \sqrt{na_n^k} [\tilde{f}_n(x) - f_n(x)] \right| \longrightarrow 0, \quad n \rightarrow \infty.$$

Since from Theorem A3 in Appendix A.3, one has that for all $x \in \mathbb{R}^k$,

$$\sqrt{na_n^k} [f_n(x) - f(x)] \Longrightarrow \mathcal{N}\left(0, f(x) \int \tilde{K}^2(u) du\right),$$

it results that for all $x \in \mathbb{R}^k$, as n tends to infinity,

$$\sqrt{na_n^k} [\tilde{f}_n(x) - f(x)] \Longrightarrow \mathcal{N}\left(0, f(x) \int \tilde{K}^2(u) du\right).$$

Then one concludes that for all $x \in \mathbb{R}^k$, as n tends to infinity,

$$\sqrt{na_n^k} \left(\frac{\tilde{f}_n(x) - f(x)}{(f(x) \int \tilde{K}^2(u) du)^{1/2}} \right) \Longrightarrow \mathcal{N}(0, 1).$$

□

Proposition 5. Assume that assumptions (A1)–(A6) hold. Then,

$$\sqrt{na_n^k} \left(\frac{\tilde{f}_n(t_i) - f(t_i)}{(f(t_i) \int \tilde{K}^2(u) du)^{1/2}}, 1 \leq i \leq \ell \right) \Longrightarrow \mathcal{N}_\ell \equiv \mathcal{N}(\mathbf{0}, I_\ell),$$

where $\mathcal{N}(\mathbf{0}, I_\ell)$ denotes the standard ℓ -dimensional Gaussian random vector, $\ell \in \mathbb{N}$, $t_i \in \mathbb{R}^k$, $1 \leq i \leq \ell$.

Proof. For all $\ell \in \mathbb{N}$, one can write

$$\begin{aligned} \sqrt{na_n^k} \left(\frac{\tilde{f}_n(t_i) - f(t_i)}{(f(t_i) \int \tilde{K}^2(u) du)^{1/2}}, 1 \leq i \leq \ell \right) &= \sqrt{na_n^k} \left(\frac{\tilde{f}_n(t_i) - f_n(t_i)}{(f(t_i) \int \tilde{K}^2(u) du)^{1/2}}, 1 \leq i \leq \ell \right) \\ &+ \sqrt{na_n^k} \left(\frac{f_n(t_i) - f(t_i)}{(f(t_i) \int \tilde{K}^2(u) du)^{1/2}}, 1 \leq i \leq \ell \right). \end{aligned}$$

From Theorem A3 in Appendix A.3, as $n \rightarrow \infty$, one has that

$$\sqrt{na_n^k} \left(\frac{f_n(t_i) - f(t_i)}{(f(t_i) \int \tilde{K}^2(u) du)^{1/2}}, 1 \leq i \leq \ell \right) \Longrightarrow \mathcal{N}_\ell.$$

It follows from the proof of the above Proposition 4 that for $t \in \mathbb{R}^k$, fixed,

$$\sqrt{na_n^k} \left(\frac{\tilde{f}_n(t) - f_n(t)}{\left(f(t) \int \tilde{K}^2(u) du \right)^{\frac{1}{2}}} \right) \rightarrow 0, n \rightarrow \infty.$$

It results from the above that as $n \rightarrow \infty$, one has that

$$\sqrt{na_n^k} \left(\frac{\tilde{f}_n(t_i) - f_n(t_i)}{\left(f(t_i) \int \tilde{K}^2(u) du \right)^{\frac{1}{2}}}, 1 \leq i \leq \ell \right) \rightarrow (0, \dots, 0).$$

So, as $n \rightarrow \infty$, one has

$$\sqrt{na_n^k} \left(\frac{\tilde{f}_n(t_i) - f(t_i)}{\left(f(t_i) \int \tilde{K}^2(u) du \right)^{\frac{1}{2}}}, 1 \leq i \leq \ell \right) \Rightarrow \mathcal{N}_\ell.$$

A direct consequence of the above proposition is given by the following result. $\square$

Proposition 6. Assume that assumptions (A1)–(A6) hold. Then, for all $x \in \mathbb{R}^k$, a confidence band for $f(x)$ at risk $\alpha \in [0, 1]$ is given by

$$CB_\alpha(x) = \left[\tilde{f}_n(x) - \frac{q_{1-\frac{\alpha}{2}}(\tilde{f}_n(x) \int \tilde{K}^2(u) du)^{1/2}}{\sqrt{na_n^k}}; \tilde{f}_n(x) + \frac{q_{1-\frac{\alpha}{2}}(\tilde{f}_n(x) \int \tilde{K}^2(u) du)^{1/2}}{\sqrt{na_n^k}} \right],$$

where $q_{1-\frac{\alpha}{2}}$ stands for the $(1 - \frac{\alpha}{2})$ -quantile of the standard Gaussian distribution.

4. The Nonparametric Models Case

Here, considering model (1), we estimate f by a kernel method in the case that the vector and matrix functions T and $V(x)$ are unknown with unknown forms. That is, they are nonparametric functions. Because the study is lengthy and too technical, unlike in the preceding section, we only study the consistency and the bias of $\tilde{f}_n(x)$ defined by (2) when the ε_t s are substituted for the residuals from the fit. In the first subsection, we give a brief review of some recent results on nonparametric estimation of T and V . The consistency and the bias studies are performed in the second subsection.

4.1. The Conditional Mean and Variance Functions Estimation

In the literature, the conditional mean (autoregression function) and the conditional variance (volatility) are estimated in various nonparametric methods. The estimators of the autoregression function are generally constructed and studied in the same ways as those of the well-known regression function. The pointwise convergence of this function was studied by [49–51]. The uniform convergence due to [52] generalizes the earlier results of [53–56]. The speed of convergence was treated by [57,58]. The L^p convergence was obtained by [59–62], Györfi (1981). The assessment of bias and asymptotic variance were obtained by [49] for $k = 1$ and [63,64] for $k \geq 1$.

Ref. [65] considered model (1) in which both T and V are unknown functions of the past. Nonparametric estimators of these functions are constructed based on local polynomial fitting. They examined the rate of convergence of these estimators and established their asymptotic normality. Indeed, they generalized the result of [9] to a wider class of conditional mean and variance functions that can be seen as a limit of CHARN models, which themselves are generalizations of the ARCH structure.

Ref. [66] studied kernel methods based on weighted averages of response variables. The resulting estimators are very sensitive to large fluctuations in the data. To overcome this drawback, Ref. [67] used a robust estimation. This allowed the authors to extend the results of [66] to a large class of processes under stationarity and ergodicity. They proposed a family of robust nonparametric estimators of the regression or autoregression functions based on the kernel method. They subsequently established the uniform convergence of this family of estimators in the case that the observations are not bounded and belong to an increasing series of compacts. Ref. [68] considered model (1). They proposed an efficient and adaptive method for estimating the conditional variance. Their basic idea is to apply local linear regression to the squared residuals. They proved that, without knowing the regression function, they can estimate the conditional variance asymptotically as well as if the regression were given. Ref. [69] surveyed the semi-parametric and nonparametric methods in univariate and multivariate ARCH/GARCH models. They introduced some specific semi-parametric models and investigated the semi-parametric and nonparametric estimation techniques applied to the error density, the functional form of the volatility functions, the relationships between mean and variance, long memory processes, locally stationary processes, continuous time processes and multivariate models. Ref. [70] applied bootstrap methods to CHARN models and gave non-parametric estimators of the trend and volatility function. Ref. [71] studied the estimation of regression and volatility functions of nonlinear autoregressive models with ARCH errors. In [72], the nonparametric local exponential estimator was applied to estimate conditional volatility functions, ensuring its non-negativity. It is proved there that the obtained estimator is asymptotically fully adaptive to unknown conditional mean functions. Ref. [73] proposed a kernel weighted version of the standard realized integrated volatility estimator and gave sample properties. In addition, Ref. [74] constructed a spot volatility estimator for high-frequency financial data which contain market microstructure noise, proved the consistency of this estimator and derived its asymptotic distribution. The aim of [75] was to establish asymptotic results for estimators of functions and functionals linked to various models. The studied estimators are built on the basis of nonparametric methods. Ref. [76] considered a multidimensional nonparametric additive regression model with dependent observations. the authors used the marginal integration technique and wavelets methodology to develop a new adaptive estimator for a component of the additive regression function. Refs. [67,77,78] proposed a family of robust nonparametric estimators for regression or autoregression functions in a univariate or multivariate context.

4.2. The Kernel Estimation of the Density of the Noise

In this subsection, we assume the functions T and V unknown and nonparametric with unknown forms. We study the kernel estimator defined by (2) when the ε_t s are substituted for the residuals from the fit. For the reasons mentioned in the introduction, we do not study the asymptotic normality of this estimator.

Let $\hat{T} = (\hat{T}_1, \dots, \hat{T}_k)^\top$, $\hat{V} = (\hat{V}_{ij})$ and $\hat{V}^{-1} = (\hat{V}_{ij}^{-1})$ be nonparametric estimators of $T = (T_1, \dots, T_k)^\top$, $V = (V_{ij})$ and $V^{-1} = (V_{ij}^{-1})$, respectively, satisfying, for any $t \in \mathbb{Z}$,

$$(A7) \quad \hat{T}(Y_t) - T(Y_t) = o_P(1), \quad \hat{V}(Y_t) - V(Y_t) = o_P(1), \quad \hat{V}^{-1}(Y_t) - V^{-1}(Y_t) = o_P(1).$$

Remark 3. (A7) is readily satisfied by estimators derived in some of the references cited above. $\hat{T}$ can be obtained by kernel methods as in [66,67] or by local polynomial methods as in [6], or by adaptive methods as in [76]. $\hat{V}$ can be obtained by local polynomial approaches as in [1,6,65] or by a local exponential method as in [72].

For any $t \in \mathbb{Z}$, define

$$\varepsilon_t = \varepsilon_t(T, V) = V^{-1}(Y_t)[X_t - T(Y_t)] \quad \text{and} \quad \hat{\varepsilon}_t = \varepsilon_t(\hat{T}, \hat{V}).$$

It is easy to see from assumption (A7) that for any $t \in \mathbb{Z}$,

$$\widehat{\varepsilon}_t = \varepsilon_t + o_p(1).$$

In the sequel, the kernel considered is denoted by $\widehat{K}$ rather than $\widetilde{K}$ as in the preceding section. We study the asymptotic behavior of the following kernel estimator:

$$\widehat{f}_n(x) = \frac{1}{na_n^k} \sum_{t=1}^n \widehat{K}\left(\frac{x - \widehat{\varepsilon}_t}{a_n}\right), \quad x \in \mathbb{R}^k.$$

Proposition 7. Assume that (A4)–(A7) hold. Then, in probability,

$$\sup_{x \in \mathbb{R}^k} |\widehat{f}_n(x) - f(x)| \rightarrow 0, \quad n \rightarrow \infty.$$

Proof. For all $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, one can write

$$\begin{aligned} |\widehat{f}_n(x) - f_n(x)| &= \left| \frac{1}{na_n^k} \sum_{t=1}^n \left[\widehat{K}\left(\frac{x - \widehat{\varepsilon}_t}{a_n}\right) - \widehat{K}\left(\frac{x - \varepsilon_t}{a_n}\right) \right] \right| \\ &= \left| \frac{1}{na_n^k} \sum_{t=1}^n \left[\widehat{K}\left(\frac{x - \widehat{V}^{-1}(Y_t)[X_t - \widehat{T}(Y_t)]}{a_n}\right) - \widehat{K}\left(\frac{x - V^{-1}(Y_t)[X_t - T(Y_t)]}{a_n}\right) \right] \right| \\ &\leq \frac{1}{na_n^k} \sum_{t=1}^n \left| \widehat{K}\left(\frac{x - \widehat{V}^{-1}(Y_t)[X_t - \widehat{T}(Y_t)]}{a_n}\right) - \widehat{K}\left(\frac{x - V^{-1}(Y_t)[X_t - T(Y_t)]}{a_n}\right) \right|. \end{aligned}$$

Recalling that $\frac{\partial \widehat{K}}{\partial x}^\top(x)$ denotes the transpose of $\frac{\partial \widehat{K}}{\partial x}(x)$, by a Taylor expansion of $\widehat{K}$, one has

$$\begin{aligned} &\widehat{K}\left(\frac{x - \widehat{V}^{-1}(Y_t)[X_t - \widehat{T}(Y_t)]}{a_n}\right) - \widehat{K}\left(\frac{x - V^{-1}(Y_t)[X_t - T(Y_t)]}{a_n}\right) \\ &= \frac{1}{a_n} \frac{\partial \widehat{K}}{\partial x}^\top\left(\frac{\dot{\varepsilon}_{t,x}}{a_n}\right) \left\{ V^{-1}(Y_t)[X_t - T(Y_t)] - \widehat{V}^{-1}(Y_t)[X_t - \widehat{T}(Y_t)] \right\}, \end{aligned}$$

where for any $t \in \mathbb{Z}$, $x \in \mathbb{R}^k$, $\dot{\varepsilon}_{t,x}$ is a random vector lying between $x - \varepsilon_t$ and $x - \widehat{\varepsilon}_t$.

Now, observing that

$$\begin{aligned} &V^{-1}(Y_t)[X_t - T(Y_t)] - \widehat{V}^{-1}(Y_t)[X_t - \widehat{T}(Y_t)] \\ &= [V^{-1}(Y_t) - \widehat{V}^{-1}(Y_t)]\varepsilon_t(T, V) + \widehat{V}^{-1}(Y_t)[\widehat{T}(Y_t) - T(Y_t)] \end{aligned}$$

It results from this that

$$\begin{aligned} \sup_{x \in \mathbb{R}^k} |\widehat{f}_n(x) - f_n(x)| &\leq \frac{1}{n} \sum_{t=1}^n \sup_{x \in \mathbb{R}^k} \left| \frac{1}{a_n^{k+1}} \frac{\partial \widehat{K}}{\partial x}^\top\left(\frac{\dot{\varepsilon}_{t,x}}{a_n}\right) \right. \\ &\quad \left. \left\{ [V^{-1}(Y_t) - \widehat{V}^{-1}(Y_t)]\varepsilon_t(T, V) + \widehat{V}^{-1}(Y_t)[\widehat{T}(Y_t) - T(Y_t)] \right\} \right|. \end{aligned}$$

By our assumptions, the terms in the summand in the right-hand side of the above inequality tends to 0 as n tends to infinity. Then, by a stochastic version of the Cesàro theorem (see Appendix A.4), in probability,

$$\sup_{x \in \mathbb{R}^k} |\hat{f}_n(x) - f_n(x)| \longrightarrow 0, \quad n \rightarrow \infty.$$

It is well known (see Theorem A1, Appendix A.1) that, in probability,

$$\sup_{x \in \mathbb{R}^k} |f_n(x) - f(x)| \longrightarrow 0, \quad n \rightarrow \infty.$$

Consequently, in probability,

$$\sup_{x \in \mathbb{R}^k} |\hat{f}_n(x) - f(x)| \longrightarrow 0, \quad n \rightarrow \infty.$$

□

Bias Study of $\hat{f}_n$

Proposition 8. Assume that (A4)–(A7) hold, and that $\hat{K}$ is bounded by a positive number. Assume that $\int \|x\| f(x) dx < \infty$. Then, for all $x \in \mathbb{R}^k$, $\hat{f}_n(x)$ is an asymptotically unbiased estimator of $f(x)$.

Proof. For all $x \in \mathbb{R}^k$,

$$\mathbb{E}[\hat{f}_n(x)] - f(x) = \mathbb{E}[\hat{f}(x) - f_n(x)] + \{\mathbb{E}[f_n(x)] - f(x)\}.$$

By Theorem A2 in Appendix A.2, as $n \rightarrow \infty$,

$$\mathbb{E}[f_n(x)] - f(x) \longrightarrow 0.$$

Now, we show that for all $x \in \mathbb{R}^k$, as $n \rightarrow \infty$,

$$\mathbb{E}[\hat{f}_n(x) - f_n(x)] \longrightarrow 0.$$

Indeed, for $x \in \mathbb{R}^k$,

$$\begin{aligned} \mathbb{E}[\hat{f}_n(x) - f_n(x)] &= \frac{1}{na_n^k} \sum_{t=1}^n \mathbb{E} \left[\hat{K} \left(\frac{x - \hat{\varepsilon}_t}{a_n} \right) - \hat{K} \left(\frac{x - \varepsilon_t}{a_n} \right) \right] \\ &= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[\frac{1}{a_n^{k+1}} (\hat{\varepsilon}_t - \varepsilon_t)^\top \frac{\partial \hat{K}^\top}{\partial x} \left(\frac{x - \hat{\varepsilon}_t}{a_n} \right) \right] \\ &= \frac{1}{n} \sum_{t=1}^n L_t(x), \end{aligned}$$

where, for any $t \in \mathbb{Z}$ and $x \in \mathbb{R}^k$,

$$\begin{aligned} L_t(x) &= \mathbb{E} \left[\frac{1}{a_n^{k+1}} (\hat{\varepsilon}_t - \varepsilon_t)^\top \frac{\partial \hat{K}^\top}{\partial x} \left(\frac{x - \hat{\varepsilon}_t}{a_n} \right) \right] \\ &= \mathbb{E} \left\{ \frac{1}{a_n^{k+1}} \frac{\partial \hat{K}^\top}{\partial x} \left(\frac{x - \hat{\varepsilon}_t}{a_n} \right) \right. \\ &\quad \left. \left[V^{-1}(Y_t) - \hat{V}^{-1}(Y_t) \right] \varepsilon_t(T, V) + \hat{V}^{-1}(Y_t) \left[\hat{T}(Y_t) - T(Y_t) \right] \right\}, \end{aligned}$$

where for all $t \in \mathbb{Z}$, $\hat{\varepsilon}_t$ lies between $\hat{\varepsilon}_t$ and ε_t .

Since for all $t \in \mathbb{Z}$, $T(Y_t) - \hat{T}(Y_t)$ and $V^{-1}(Y_t) - \hat{V}^{-1}(Y_t) \rightarrow 0$ in probability as $n \rightarrow \infty$. The above terms in the pairs of brackets tend to 0 in probability as n tends to infinity. It follows from our assumptions that for all $t \in \mathbb{Z}$ and all $x \in \mathbb{R}^k$,

$$L_t(x) \rightarrow 0, \quad n \rightarrow \infty.$$

Thus, by the Cesàro Theorem, for all $x \in \mathbb{R}^k$,

$$\mathbb{E}[\hat{f}_n(x) - f_n(x)] = \frac{1}{n} \sum_{t=1}^n L_t(x) \rightarrow 0, \quad n \rightarrow \infty.$$

This concludes the proof of the result. $\square$

5. Simulation Experiments

In this section, we conduct a simulation experiment to evaluate the finite sample properties of the estimators studied in the preceding sections. We restrict to the cases $k = 1, 2$. For $k = 1$, we study the situation where the conditional mean and variance functions are parametric, and where they are nonparametric. For $k = 2$, we only study the case that these functions are parametric.

Although we do not study the asymptotic normality of the kernel estimator in the case that the functions T and V are nonparametric, we construct heuristically asymptotic confidence-bound functions in the lines of those obtained in the case that T and V are parametric.

5.1. Unidimensional Case

Let $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)^\top$. For several values of θ , we generate n observations from model (1) with

$$T(x) = (\theta_0 + \theta_1 e^{-0.03x^2})x \text{ and } V(x) = \sqrt{\theta_2 + \theta_3 x^2} \quad (7)$$

and f is the standard Gaussian density function. Once the data are generated, in order to apply our results, we first assume that functions T and V have the above forms with unknown θ , estimated by a maximum likelihood method. Next, we assume that the forms of these functions are unknown and they are estimated by a kernel method.

For all the kernel methods used, the bandwidth is the same $a_n = \sigma_n n^{-1/5}$, with σ_n^2 standing for the sample variance of the simulated observations. The kernel considered is the Gaussian kernel. For any $x \in \mathbb{R}$, the kernel estimator of $f(x)$ and its associated confidence bounds are computed on the basis of 1000 replications for a sample size of $n = 200$ at each replication.

Figures 1 and 2 respectively exhibit the results corresponding to the parametric and nonparametric cases. They both show the graph of the true density (black color), the upper-bound graph (red), the lower-bound graph (green) and that of the estimator $\hat{f}_n$ (blue). Either of these graphs indicate that the Gaussian density of the noise is globally well estimated for the examples considered.

Figure 1. Kernel estimation of the density of the noise for $k = 1$ and parametric conditional mean and variance functions. (a–d): exact density, kernel estimator, lower and upper confidence bounds curves.

Figure 2. Kernel estimation of the density of the noise for $k = 1$ and nonparametric conditional mean and variance functions. (a–d): exact density, kernel estimator, lower and upper confidence bounds curves.

5.2. Bidimensional Case

Let $\theta = (\theta_1, \theta_2)^\top$. For $k = 2$, we generate n observations $X_1 = (X_{1,1}, X_{1,2})^\top, \dots, X_n = (X_{n,1}, X_{n,2})^\top$ from model (1) with $T(x) = (T_1(x), T_2(x))^\top$, where for any $x = (x_1, x_2)^\top \in \mathbb{R}^2$, $T_1(x) = \theta_1 e^{0.03x_1^2} x_1$, $T_2(x) = \theta_2 x_2$, $V(x) = 1$, and f is the standard bidimensional density function. We do this for $\theta_1 = 0.5$, $\theta_2 = 0.75$.

As mentioned above, we only treat the parametric case. That is, we now assume that the functions T and V have the above parametric form with unknown θ estimated by a maximum likelihood method. Then, the kernel estimator of f is computed with a bidimensional standard Gaussian kernel. Since this can be written as a product of two univariate standard Gaussian kernels, we consider, for each of these kernels, the smoothing parameters $a_{n,j} = \sigma_{n,j} n^{-\frac{1}{5}}$, where $\sigma_{n,j}^2$ is the sample variance of the $X_{1,j}, \dots, X_{n,j}$, $j = 1, 2$.

Here, we take $n = 800$, although smaller values give satisfactory results. For any $x \in \mathbb{R}^2$, $f_n(x)$ and its confidence bounds are computed on the basis of 1000 replications. Figure 3 shows, for the single model considered, the graph of the true density (blue color), the upper-bound graph (red), the lower-bound graph (green) and that of the estimator f_n (blue). The four graphs indicate that the bidimensional Gaussian density of the noise is well estimated for the example considered.

Figure 3. Kernel estimation of the density of the noise for $k = 2$ and parametric conditional mean and variance functions. (a): lower confidence bound curve; (b): upper confidence bound curve; (c): kernel estimate; (d): exact density.

6. Conclusions

We studied the kernel estimation of the density f of the noise in model (1) when the functions T and V are both parametric and when they are both nonparametric. In the former case, the consistency of the estimator as well as its pointwise asymptotic normality

are established. This led to the construction of pointwise confidence-bound functions. It is interesting to note that from Proposition 5, Wald-type tests can be constructed for goodness-of-fit tests on f . In the case that T and V are nonparametric, only the consistency of the estimator is studied. The main reason for this is that the study seemed to us too long and too technical.

The simulation experiment conducted shows that the estimator behaves quite well on the examples considered. It can be seen from the graphics associated with the results obtained for the parametric functions T and V that the curve of the true density function generally lies within the curves of the pointwise confidence-bound functions. Despite the fact that the pointwise asymptotic normality of $\hat{f}_n$ was not established (T and V are nonparametric), in light of the case where they are parametric, we wrote down the expressions of the confidence bound functions for f . It seems from the simulation results (see Figure 2) that the curve of true density is usually between those of the confidence-bound curves.

Author Contributions: Investigation, D.A.N.N. and J.S.; Methodology, J.N.-W. and M.L.; Supervision, J.N.-W.; Writing—original draft, J.N.-W., M.L. and D.A.N.N.; Writing—review & editing, J.N.-W. and M.L. Software, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the referees for their comments and suggestions, which led to an improvement of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Theorem III.3 of Bosq and Lecoutre (1987), p. 65

Definition A1. A kernel function $K(\cdot)$ is said to be a Geffroy kernel function if the following hold:

1. The set of discontinuity points of $K(\cdot)$ has zero measure;
2. The function $x \mapsto \sup\{|K(u)| : \|u - x\| < 1\}$ is integrable on $\mathbb{R}^k$.

Some examples are the following:

1. The indicator function $\mathbb{I}_{[-1/2, 1/2]}$ is a Geffroy kernel;
2. Any bounded variation kernel is a Geffroy kernel.

Theorem A1. Suppose that the density function f is uniformly continuous. Let f_n be the density estimator associated with any kernel $K(\cdot)$. Then the convergence in probability

$$\sup_{x \in \mathbb{R}^k} |f_n(x) - f(x)| \rightarrow 0, \quad n \rightarrow \infty$$

implies

$$h_n \rightarrow 0, \quad \frac{nh_n^k}{\log n} \rightarrow \infty.$$

Conversely, if $K(\cdot)$ is a Geffroy kernel, the condition

$$h_n \rightarrow 0, \quad \frac{nh_n^k}{\log n} \rightarrow \infty$$

implies the almost complete convergence

$$\sup_{x \in \mathbb{R}^k} |f_n(x) - f(x)| \longrightarrow 0, \quad n \rightarrow \infty.$$

Remark A1. When a Geffroy kernel $K(\cdot)$ is used, the above condition on h_n is equivalent to the convergence of the density estimator f_n in probability, almost surely and almost completely.

Appendix A.2. Theorem V.2 of Bosq and Lecoutre (1987), p. 75

Theorem A2. Suppose that the density function f is twice differentiable, all its partial derivatives are continuous and bounded. Let $K(\cdot)$ be an even kernel with $\int \|y\|^2 K(y) dy < \infty$. Denote $\chi_k(x) = \sum_{1 \leq i, j \leq k} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int t_i t_j K(t) dt$. Then, under the condition $h_n \rightarrow 0$, $\frac{nh_n^k}{\log n} \rightarrow \infty$, we have following asymptotic developments:

$$\begin{aligned} E[f_n(x)] - f(x) &= \frac{h_n^2}{2} \chi_k(x) + o(h_n^2), \\ \text{Var}(f_n(x)) &= \frac{f(x)}{nh_n^k} \int K^2 + o\left(\frac{1}{nh_n^k}\right). \end{aligned}$$

In consequence, we have

$$E\{[f_n(x) - f(x)]^2\} = \frac{h_n^4}{4} \chi_k^2(x) + \frac{f(x)}{nh_n^k} \int K^2 + o\left(h_n^4 + \frac{1}{nh_n^k}\right).$$

Appendix A.3. Theorem VIII.2 of Bosq and Lecoutre (1987), p. 86

Theorem A3. Suppose that the density function f is continuous and strictly positive at distinct points $t_1, \dots, t_m \in \mathbb{R}^k$. Let f_n be the density estimator associated to a kernel $K(\cdot)$. Then the condition $h_n \rightarrow 0$, $\frac{nh_n^k}{\log n} \rightarrow \infty$ implies the asymptotic normality

$$\sqrt{nh_n^k} \left(\frac{f_n(t_i) - Ef_n(t_i)}{(\int K^2)^{1/2}}; 1 \leq i \leq m \right) \Longrightarrow \mathcal{N}(\mathbf{0}, I_m),$$

where $\mathcal{N}(\mathbf{0}, I_m)$ is the m -dimensional standard normal distribution.

Appendix A.4. Cesàro Means

Theorem A4. Let (a_n) be a sequence of numbers. Let

$$b_n = \frac{a_1 + a_2 + \dots + a_n}{n}.$$

If $a_n \rightarrow a$, then

$$\lim_{n \rightarrow \infty} b_n = a.$$

References

1. Ruppert, D.; Wand, M.; Holst, U.; Hösler, O. Local polynomial variance-function estimation. *Technometrics* **1997**, *39*, 262–273. [\[CrossRef\]](#)
2. Härdle, W.; Tsybakov, A. Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econom.* **1997**, *81*, 223–242. [\[CrossRef\]](#)
3. Neumeyer, N.; Pablo, A.; Perri, F. Business cycles in emerging economies: the role of interest rates. *J. Monet. Econ.* **2005**, *52*, 345–380. [\[CrossRef\]](#)
4. Pardo-Fernández, J.; Van Keilegom, I.; González-Manteiga, W. Testing for the equality of k regression curves. *Stat. Sin.* **2007**, *17*, 1115–1137.
5. Dette, H.; Neumeyer, N.; Van Keilegom, I. A new test for the parametric form of the variance function in non-parametric regression. *J. R. Stat. Soc. Ser. B* **2007**, *69*, 903–917. [\[CrossRef\]](#)

6. Neumeyer, N.; Van Keilegom, I. Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivar. Anal.* **2010**, *101*, 1067–1078. [\[CrossRef\]](#)
7. Engle, R. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **1982**, *50*, 987–1007. [\[CrossRef\]](#)
8. Bollerslev, T.; Chou, R.; Kroner, K. ARCH modeling in finance: A review of the theory and empirical evidence. *J. Econom.* **1992**, *52*, 5–59. [\[CrossRef\]](#)
9. Gouriéroux, C.; Monfort, A. Qualitative threshold ARCH models. *J. Econom.* **1992**, *52*, 159–199. [\[CrossRef\]](#)
10. Härdle, W.; Mammen, E.; Müller, M. Testing parametric versus semiparametric modeling in generalized linear models. *J. Am. Stat. Assoc.* **1998**, *93*, 1461–1474. [\[CrossRef\]](#)
11. Brockwell, P.; Davis, R. *Introduction to Time Series and Forecasting*; Springer: Berlin/Heidelberg, Germany, 2002.
12. Geffroy, J. Sur la convergence uniforme des estimateurs d'une densité de probabilité. *Séminaire de Statistique ISUP* **1973**.
13. Scott, D.; Tapia, R.; Thompson, J. Multivariate density estimation by discrete maximum penalized likelihood methods. In *Graphical Representation of Multivariate Data*; Elsevier: Amsterdam, The Netherlands, 1978; pp. 169–182.
14. Deheuvels, P. Estimation non paramétrique de la densité par histogrammes généralisés. *Rev. Stat. Appliquée* **1977**, *25*, 5–42.
15. Deheuvels, P. Non Parametric Tests of Independence. In *Statistique non Paramétrique Asymptotique*; Springer: Berlin/Heidelberg, Germany, 1980; pp. 95–107.
16. Prakasa-Rao, B. Asymptotic theory for non-linear least squares estimator for diffusion processes. *J. Theor. Appl. Stat.* **1983**, *14*, 195–209. [\[CrossRef\]](#)
17. Devroye, L.; Györfi, L. *Nonparametric density estimation: The L_1 View*; Wiley: New York, NY, USA, 1985.
18. Bosq, D.; Lecoutre, J.P. *Théorie de L'estimation Fonctionnelle*; Economica: Paris, France, 1987.
19. Silverman, B. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.
20. Pearson, K. On the systematic fitting of curves to observations and measurements I, II. *Biometrika* **1902**, *1*, 265–303. [\[CrossRef\]](#)
21. Pearson, K. On the systematic fitting of curves to observations and measurements I, II. *Biometrika* **1902**, *2*, 1–23. [\[CrossRef\]](#)
22. Rosenblatt, M. Remark on some nonparametric estimates of a density function. *Ann. Math. Statist.* **1956**, *27*, 832–837. [\[CrossRef\]](#)
23. Silverman, B.; Jones, M. E. Fix and J.L. Hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation. *Int. Stat. Rev.* **1989**, *57*, 233–247. [\[CrossRef\]](#)
24. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [\[CrossRef\]](#)
25. Cacoullos, T. On a Class of Admissible Partitions. *Ann. Math. Stat.* **1966**, *37*, 189–195. [\[CrossRef\]](#)
26. Bierens, H. Uniform consistency of kernel estimators of a regression function under generalized conditions. *J. Am. Stat. Assoc.* **1983**, *78*, 699–707. [\[CrossRef\]](#)
27. Devroye, L.; Penrod, C. Distribution-free lower bounds in density estimation. *Ann. Stat.* **1984**, *12*, 1250–1262. [\[CrossRef\]](#)
28. Abdous, B. Étude d'une classe d'estimateurs à noyaux de la densité une loi de probabilité. Ph.D. Thesis, Paris, France, 1986.
29. Tsybakov, A. *Introduction to Nonparametric Estimation*; Springer: Berlin/Heidelberg, Germany, 2009.
30. Tiago de Oliveira, M. Chromatographic isolation of monofluoroacetic acid from *Palicourea marcgravii* St. Hil. *Experientia* **1963**, *19*, 586–587. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Heidenreich, N.B.; Schindler, A.; Sperlich, S. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *Adv. Stat. Anal.* **2013**, *97*, 403–433. [\[CrossRef\]](#)
32. Li, Q.; Racine, J. *Nonparametric Econometrics: Theory and Practice*; Princeton University Press: Princeton, NJ, USA, 2007.
33. Lecoutre, J.P. Convergence et Optimisation de Certains Estimateurs des Densités de Probabilité. Ph.D. Thesis, 1975.
34. Chandra, S.; Taniguchi, M. Estimating functions for nonlinear time series models. *Ann. Inst. Stat. Math.* **2001**, *53*, 125–141. [\[CrossRef\]](#)
35. Amano, T. Asymptotic Optimality of Estimating Function Estimator for CHARN Model. *Adv. Decis. Sci.* **2012**, *2012*, 515494. [\[CrossRef\]](#)
36. Kanai, H.; Ogata, H.; Taniguchi, M. Estimating function approach for CHARN models. *Metron* **2010**, *68*, 1–21. [\[CrossRef\]](#)
37. Godambe, V. An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* **1960**, *31*, 1208–1211. [\[CrossRef\]](#)
38. Godambe, V. The foundations of finite sample estimation in stochastic processes. *Biometrika* **1985**, *72*, 419–428. [\[CrossRef\]](#)
39. Hansen, L. Large sample properties of generalized method of moments estimators. *J. Econom. Soc.* **1982**, *50*, 1029–1054. [\[CrossRef\]](#)
40. Li, D.; Turtle, H. Semiparametric ARCH models: An estimating function approach. *J. Bus. Econ. Stat.* **2000**, *18*, 174–186.
41. Ngatchou-Wandji, J. Estimation in a class of nonlinear heteroscedastic time series models. *Electron. J. Stat.* **2008**, *2*, 40–62. [\[CrossRef\]](#)
42. Ozaki, T. Non-linear time series models for non-linear random vibrations. *J. Appl. Probab.* **1980**, *17*, 84–93. [\[CrossRef\]](#)
43. Chan, K.S.; Tong, H. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **1985**, *17*, 666–678. [\[CrossRef\]](#)
44. Al-Qassam, M.; Lane, J. Forecasting exponential autoregressive models of order 1. *J. Time Ser. Anal.* **1989**, *10*, 95–113. [\[CrossRef\]](#)
45. Koul, H.L.; Schick, A. Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **1997**, *3*, 247–277. [\[CrossRef\]](#)
46. Ismail, M. Bayesian analysis of exponential AR models. *Far East J. Stat.* **2001**, *5*, 1–15.
47. Baragona, R.; Battaglia, F.; Cucina, D. A note on estimating autoregressive exponential models. *Quad. Stat.* **2002**, *4*, 71–88.
48. Ghosh, H.; Gurung, B.; Gupta, P. Fitting EXPAR models through the extended Kalman filter. *Sankhya B* **2015**, *77*, 27–44. [\[CrossRef\]](#)
49. Rosenblatt, M. *Conditional Probability Density and Regression Estimator*; Academic Press: New York, NY, USA, 1969.

50. Noda, K. Estimation of a regression function by the Parzen kernel type density estimators. *Ann. Inst. Statist. Math.* **1976**, *28*, 221–234. [\[CrossRef\]](#)
51. Greblicki, W.; Krzyzak, A. Asymptotic properties of kernel estimates of a regression function. *J. Stat. Plan. Inference* **1980**, *4*, 81–90. [\[CrossRef\]](#)
52. Collomb, G. Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. *CRAS* **1979**, *A*, 161–163.
53. Nadaraya, E.A. Some new estimates for distribution functions. *Theory Probab. Appl.* **1964**, *9*, 497–500. [\[CrossRef\]](#)
54. Nadaraya, E. On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* **1965**, *10*, 186–190. [\[CrossRef\]](#)
55. Devroye, L. The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Trans. Inf. Theory* **1978**, *24*, 142–151. [\[CrossRef\]](#)
56. Devroye, L.P. The uniform convergence of the nadaraya-watson regression function estimate. *Can. J. Stat.* **1978**, *6*, 179–191. [\[CrossRef\]](#)
57. Schuster, E.; Yakowitz, S. Contributions to the theory of nonparametric regression, with application to system identification. *Ann. Stat.* **1979**, *7*, 139–149. [\[CrossRef\]](#)
58. Härdle, W.; Luckhaus, S. Uniform consistency of a class of regression function estimators. *Ann. Stat.* **1984**, *12*, 612–623. [\[CrossRef\]](#)
59. Devroye, L.; Wagner, T. Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Stat.* **1980**, *8*, 231–239. [\[CrossRef\]](#)
60. Devroye, L.; Wagner, T. The strong uniform consistency of kernel density estimates. In Proceedings of the fifth International Symposium on Multivariate Analysis, New York, NY, USA, 1 January 1980; Volume 5, pp. 59–77.
61. Spiegelman, C.; Sacks, J. Consistent window estimation in nonparametric regression. *Ann. Stat.* **1980**, *8*, 240–246. [\[CrossRef\]](#)
62. Devroye, L. On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Stat.* **1981**, *9*, 1310–1319. [\[CrossRef\]](#)
63. Collomb, G. Estimation non paramétrique de la régression par la méthode du noyau: Propriété de convergence asymptotiquement normale indépendante. *Mathematics* **1977**, *65*, 24–46.
64. Collomb, G. Estimation non-paramétrique de la régression: revue bibliographique. *Int. Stat. Rev.* **1981**, *49*, 75–93. [\[CrossRef\]](#)
65. Härdle, W.; Tsybakov, A.; Yang, L. Nonparametric vector autoregression. *J. Stat. Plan. Inference* **1998**, *68*, 221–245. [\[CrossRef\]](#)
66. Collomb, G.; Härdle, W. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stoch. Process. Their Appl.* **1986**, *23*, 77–89. [\[CrossRef\]](#)
67. Laïb, N.; Ould-Saïd, E. A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Can. J. Stat.* **2000**, *28*, 817–828. [\[CrossRef\]](#)
68. Fan, J.; Yao, Q. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **1998**, *85*, 645–660. [\[CrossRef\]](#)
69. Linton, O.; Yan, Y. Semi-and nonparametric arch processes. *J. Probab. Stat.* **2011**, *2011*, 906212. [\[CrossRef\]](#)
70. Franke, J.; Neumann, M.; Stockis, J.P. Bootstrapping nonparametric estimators of the volatility function. *J. Econom.* **2004**, *118*, 189–218. [\[CrossRef\]](#)
71. Laïb, N. Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *J. Stat. Plan. Inference* **2005**, *134*, 116–139. [\[CrossRef\]](#)
72. Ziegelmann, F. Nonparametric estimation of volatility functions: the local exponential estimator. *Econom. Theory* **2002**, *18*, 985–991. [\[CrossRef\]](#)
73. Kristensen, D. Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econom. Theory* **2010**, *26*, 60–93. [\[CrossRef\]](#)
74. Zu, Y.; Boswijk, H. Estimating spot volatility with high-frequency financial data. *J. Econom.* **2014**, *181*, 117–135. [\[CrossRef\]](#)
75. Maillot, B. Propriétés Asymptotiques de Quelques Estimateurs Non-Paramétriques Pour des Variables Vectorielles et Fonctionnelles. Ph.D. Thesis, Université du Littoral, Paris, France, 2008.
76. Chesneau, C.; Fadili, J.; Maillot, B. Adaptive estimation of an additive regression function from weakly dependent data. *J. Multivar. Anal.* **2015**, *133*, 77–94. [\[CrossRef\]](#)
77. Crambes, C.; Delsol, L.; Laksaci, A. Robust nonparametric estimation for functional data. *J. Nonparametr. Stat.* **2008**, *20*, 573–598. [\[CrossRef\]](#)
78. Gheriballah, A.; Laksaci, A.; Rouane, R. Robust nonparametric estimation for spatial regression. *J. Stat. Plan. Inference* **2010**, *140*, 1656–1670. [\[CrossRef\]](#)