MALAYA JOURNAL OF MATEMATIK
Malaya J. Mat. **9(04)**(2021), 251–258.
http://doi.org/10.26637/mjm904/009

# Nadaraya-Watson estimation of a nonparametric autoregressive model

**Ben Célestin KOUASSI**[*1], **Ouagnina HILI**[1] and **Edoh KATCHEKPELE**[2]

[1] *UMRI Mathématiques et Nouvelles Technologies de l'Information , Institut National Polytechnique Felix Houphouet-Boigny, BP 1093 Yamoussoukro, Cote d'Ivoire.*
[2] *Département de Mathématiques, Faculté des Sciences et Techniques, Université de Kara, Kara, TOGO.*

***R*eceived 29 July 2021; *A*ccepted 27 September 2021**

**Abstract.** We investigate the asymptotic behavior of the Nadaraya-Watson (NW) estimator of the regression function of a $\tau-$mixing process. We prove the strong consistency and the asymptotic normality of this estimator and we illustrate these two properties using simulated data.

**AMS Subject Classifications**: $62E20$, $62G05$, $62G08$, $62G20$.

**Keywords**: Nonparametric autoregression, Nonparametric estimation, Asymptotic normality, Nadaraya-Watson estimator, $\tau-$mixing.

## Contents



## 1. Introduction

From the seminal works by Rosenblatt [20], nonparametric function estimation has been widely investigated. Parzen [19] proposed a family of kernels for nonparametric density function estimation. He obtained the same result as Rosenblatt [20]. These different works allowed Nadaraya [17] and Watson [22] to independently propose a nonparametric estimator of the regression function. This is the Nadaraya-Watson (NW) estimator. Theoretical and practical aspects of this estimator have been studied. Interesting properties have been obtained. For an overview on the question, we refer to Bercu et al. [2], Li et al. [15] and the references therein. The NW estimation method was initially restricted to independent and identically distributed data (see, for example, [16, 18, 21] and the references therein). Then, it has been adapted by several studies to the $\alpha-$, $\beta-$ and $\phi-$mixing processes (see, for example, [5, 7, 12] and the references therein). There are very few studies suitable for $\tau-$mixing processes. This paper presents itself as one of the few contributions on the estimation of the regression function of $\tau-$mixing process. We refer the reader to Dedecker and Prieur [6] for the definition of a $\tau-$mixing process.
More recently, Hong and Linton [13] proposed an infinite dimensional NW type estimator for the regression

[*] **Corresponding author.** Email addresses: **kouassiben5@gmail.com** (Kouassi), **o_hili@yahoo.fr** (Hili), **edohkatchekpele@gmail.com** (Katchekpele)

 

function of an $\alpha-$dependent process. In our paper, we use a NW estimator, as Hong and Linton [13] , to estimate the regression function of a $p-$Markov process. These processes are generally $\beta-$dependant. However, there are some that are neither $\alpha-$dependent nor $\beta-$dependent (but $\tau-$dependent) (see [1]). Among these, we can mention some nonparametric autoregressive (NAR) processes. According to Fan and Yao [10] (p. 19), a sequence $(X_t)_{t\in\mathbb{Z}}$ is a NAR process if it is a solution of (2.1). In our study, we show the strong consistency and the asymptotic normality of the NW estimator of the regression function of NAR process under the assumption of a $\tau-$mixing condition on the sample. Our results go further than those of Hong and Linton ([13] , Theorem 1) since we get the strong consistency.

The remainder of this paper is organized as follows. Section 2 discusses the model and the assumptions. Section 3 contains the main results and their proof. Section 4 is devoted to a small simulation.

## 2. Notations and Assumptions

In this paper, we shall use the following notations : $\|z\| := \sup_{1\leqslant i\leqslant p} |z_i|$, for any $z = (z_1, z_2, \dots, z_p)^{'} \in \mathbb{R}^p$ where $Z^{'}$ denotes the transpose of $Z$. For any $v \in \mathbb{R}$, $[v]$ denotes the largest integer close to $v$;

Let $(X_t)_{t\in\mathbb{Z}}$ be a stochastic process satisfying :

$$X_t = f(Y_t) + \xi_t,\ t \in \mathbb{Z}; \tag{2.1}$$

where $X_t \in \mathbb{R}$, $Y_t = (X_{t-1}, X_{t-2}, ..., X_{t-p})^{'} \in \mathbb{R}^p$, $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of independent identically distributed random variables with $\mathbb{E}(\xi_t) = 0$ and $\sigma^2(\xi_t) > 0$, $t \in \mathbb{Z}$. The random variable $\xi_t$ is independent of $X_i$, for $i < t$ and $f(.) : \mathbb{R}^p \to \mathbb{R}$, $z \longmapsto \mathbb{E}(X_t|Y_t = z)$, $t \in \mathbb{Z}$, is an unknown measurable function.

Let $x \in \mathbb{R}^p$, we observe $(X_1, Y_1), ..., (X_T, Y_T)$ and estimate $f(x)$ by :

$$\widehat{f}_T(x) = \begin{cases} \dfrac{\sum_{t=1}^{T} K_t(x)X_t}{\sum_{t=1}^{T} K_t(x)}; & \text{if } \sum_{t=1}^{T} K_t(x) \neq 0 \\ 0, & \text{otherwise}; \end{cases} \tag{2.2}$$

where $K_t(x) = K\left(\left\|h_T^{-1}(x - Y_t)\right\|\right)$, $t = 1, ..., T$; $K(.)$ denotes the kernel function and $h_T > 0$.

Our goal is to establish the consistency and the asymptotic normality of $\widehat{f}_T(x)$. Zhu and Politis [23] have done this for nonparametric functional autoregression models. Hong and Linton [13] also proved it for $\alpha-$dependent processes.

The assumptions needed for the theoretical results are stated below.

$(A_1)$ : There exists an Orlicz function $\Phi(.)$ such that :

$$\Phi(uv) \leqslant \Phi(u)\Phi(v),\ \text{for all } u, v \in \mathbb{R}_+;$$

and for all $y, z \in \mathbb{R}^p$,

$$|f(y) - f(z)| \leqslant \sum_{j=1}^{p} \varpi_j |y_j - z_j|,$$

where $(\varpi_j)_{1\leqslant j\leqslant p}$ is a sequence of nonnegative real numbers such that $\varpi = \sum_{j=1}^{p} \varpi_j < 1$, $|f(0, 0, ..., 0)| + \|\xi_1\|_\Phi < \infty$ and $\|.\|_\Phi$ denotes the Orlicz norm associated with $\Phi(.)$ (see [9] for the definition of the Orlicz norm).

$(A_2)$ : The kernel $K : [0,+\infty[\longrightarrow [0,+\infty[$ is bounded and has compact support, that is, there exists $\lambda > 0$ such that $K(\upsilon) = 0$ for all $\upsilon > \lambda$. There exists two real constants $0 < C_1 < C_2 < \infty$ such that $C_1 \leqslant K(\upsilon) \leqslant C_2$, $\upsilon \in [0,\lambda]$ and $\int_{\mathbb{R}} K(\nu)d\nu = 1$.

$(A_3)$ : For $t = 1,...,T$, $\varphi_x(\lambda h_T) := \mathbb{P}(\|h_T^{-1}(Y_t - x)\| \leqslant \lambda) > 0$ ($\lambda$ is defined in Assumption $(A_2)$) and $h_T \longrightarrow 0$ as $T \longrightarrow \infty$.

From Assumption $(A_1)$, Doukhan and Wintenberger [9] show the existence of a strongly stationary and $\tau-$dependent solution of (2.1) such that $\tau(i) = O(a^i)$, $0 < a < 1$ (see Corollary 3.1 of [9]). According to Remark 3.1 of Doukhan and Wintenberger [9], this solution is an ergodic process. So $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t, Y_t)_{t\in\mathbb{Z}}$ are strongly stationary and ergodic processes (see Theorem 36.4 of [3]). Assumption $(A_1)$ also reflects the continuity of the application $f(.)$. Assumption $(A_2)$ was borrowed from Hong and Linton [13] (Assumption $B3$). Assumption $(A_3)$ expresses the possibility of observing the sample in a neighbourhood of $x$. This is a classic assumption in the nonparametric framework. It naturally extends the hypothesis of the strictly positive density of the explanatory variable.

## 3. Main Results

**Theorem 3.1.** *Under Assumptions $(A_1)$, $(A_2)$ and $(A_3)$, for $T$ big enough,*

$$\widehat{f}_T(x) = f(x) + o(1) \ \ \textit{almost surely } (a.s.) \tag{3.1}$$

**Proof**. According to Assumption $(A_3)$; we have, for $t = 1,...,T$, $\mathbb{P}\left(\frac{\|x-Y_t\|}{h_T} \leqslant \lambda\right) > 0$, so $\mathbb{E}\Bigg(K_t(x)\Bigg) > 0$.

Let :

$$\widehat{f}_{1,T}(x) = \frac{\frac{1}{T}\sum_{t=1}^{T} K_t(x)X_t}{\mathbb{E}(K_1(x))} \quad \text{and} \quad \widehat{f}_{2,T}(x) = \frac{\frac{1}{T}\sum_{t=1}^{T} K_t(x)}{\mathbb{E}(K_1(x))}. \tag{3.2}$$

According to Equation (20) of Hong and Linton [13], we can write :

$$\widehat{f}_T(x) - f(x) = \frac{\mathbb{E}\left(\widehat{f}_{1,T}(x)\right) - f(x)}{\widehat{f}_{2,T}(x)} + \frac{\widehat{f}_{1,T}(x) - \mathbb{E}\left(\widehat{f}_{1,T}(x)\right)}{\widehat{f}_{2,T}(x)} - \frac{f(x)\Bigg(\widehat{f}_{2,T}(x) - 1\Bigg)}{\widehat{f}_{2,T}(x)}. \tag{3.3}$$

Let us study the asymptotic behavior of $\widehat{f}_T(x) - f(x)$. To do it, we shall study the asymptotic behaviors of $\widehat{f}_{2,T}(x)$, $\mathbb{E}\left(\widehat{f}_{1,T}(x)\right) - f(x)$ and $\widehat{f}_{1,T}(x) - \mathbb{E}\left(\widehat{f}_{1,T}(x)\right)$.

We start with the asymptotic behavior of $\widehat{f}_{2,T}(x)$.

According to Assumption $(A_1)$, $(X_t)_{t\in\mathbb{Z}}$ is strongly stationary and ergodic. Since $K_t(x)$ is a measurable transformation of $(X_{t-1},...,X_{t-p})'$ and $\mathbb{E}(K_1(x)) < +\infty$ (see Assumption $(A_2)$), we have by Krengel [14], for $T$ big enough,

$$\frac{1}{T}\sum_{t=1}^{T} K_t(x) \longrightarrow \mathbb{E}(K_1(x)) \ a.s.$$

So, we have for $T$ big enough :

$$\widehat{f}_{2,T}(x) \longrightarrow 1 \ a.s. \tag{3.4}$$

According to Assumptions $(A_1)$ and $(A_2)$, $|\mathbb{E}(K_1(x)X_1)| < \infty$. And $K_t(x)X_t$ is a measurable transformation of $(X_t, X_{t-1}, ..., X_{t-p})'$. Therefore, we show as in (3.4), for $T$ big enough :

$$\frac{1}{T}\sum_{t=1}^{T} K_t(x)X_t \longrightarrow \mathbb{E}(K_1(x)X_1)\ a.s.$$

Therefore, for $T$ big enough :

$$\widehat{f}_{1,T}(x) - \mathbb{E}\left(\widehat{f}_{1,T}(x)\right) \longrightarrow 0\ a.s. \tag{3.5}$$

Using the same reasoning as the proof of Equation (53) in Hong and Liton [13] (see also the proof of Lemma 6.2 of [11]), we show, for $T$ big enough :

$$\mathbb{E}\left(\widehat{f}_{1,T}(x)\right) - f(x) \longrightarrow 0. \tag{3.6}$$

Gathering (3.3), (3.4), (3.5) and (3.6), we get (3.1). ■

**Theorem 3.2.** *Under Assumptions $(A_1)$, $(A_2)$ and $(A_3)$, for T big enough,*

$$\varsigma^2 := \lim_{T\to\infty}\frac{1}{T}Var\left(\sum_{t=1}^{T} X_t\right) \ < \ +\infty. \tag{3.7}$$

*And*

$$\sqrt{T}\mathbb{E}(K_1(x))\bigg(\widehat{f}_T(x) - f(x) + o(1)\bigg) + o\bigg(\sqrt{\ln\ln(T)}\bigg) \stackrel{d}{\longrightarrow} N(0,\varsigma^2), \tag{3.8}$$

*where $\stackrel{d}{\longrightarrow}$ denotes convergence in distribution.*

**Proof**. According to (3.3), (3.4) and (3.6), we have *a.s.*, for $T$ big enough :

$$\begin{aligned}
\widehat{f}_T(x) - f(x) &= \widehat{f}_{1,T}(x) - \mathbb{E}\left(\widehat{f}_{1,T}(x)\right) + o(1)\\
&= \frac{1}{T\mathbb{E}(K_1(x))}\sum_{t=1}^{T}\bigg(K_t(x)X_t - \mathbb{E}(K_1(x)X_1)\bigg) + o(1)\\
&= \frac{1}{T\mathbb{E}(K_1(x))}\sum_{t=1}^{T}\bigg\{(K_t(x)-1)X_t - \mathbb{E}\bigg((K_1(x)-1)X_1\bigg)\bigg\}\\
&+ \frac{1}{T\mathbb{E}(K_1(x))}\sum_{t=1}^{T}\bigg(X_t - \mathbb{E}(X_1)\bigg) + o(1).
\end{aligned} \tag{3.9}$$

Since $(K_t(x)-1)X_t$ is a measurable transformation of $(X_t, X_{t-1}, ..., X_{t-p})'$, so we have, for $T$ big enough :

$$\frac{1}{T}\sum_{t=1}^{T}\bigg\{(K_t(x)-1)X_t - \mathbb{E}\bigg((K_1(x)-1)X_1\bigg)\bigg\} \longrightarrow 0\ a.s.$$

we have *a.s.*, for $T$ big enough :

$$\widehat{f}_T(x) - f(x) = \frac{1}{T\mathbb{E}(K_1(x))}\sum_{t=1}^{T}\bigg(X_t - \mathbb{E}(X_1)\bigg) + o(1). \tag{3.10}$$

The function $s \longmapsto |s|^2 \ln(1+|s|)$ is measurable. So $(|X_t - \mathbb{E}(X_1)|^2 \ln(1+|X_t-\mathbb{E}(X_1)|))_t$ is stationary because $(X_t)_t$ is strongly stationary and ergodic. Therefore $\mathbb{E}(|X_t - \mathbb{E}(X_1)|^2 \ln(1+|X_t-\mathbb{E}(X_1)|)) < \infty$.
According to the Hypothesis $(A_1)$, the mixing coefficient $\tau(.)$ of the process $(X_t)_{t\in\mathbb{Z}}$ is such that $\tau(i) = O(a^i)$, $0 < a < 1$.
From item 3 of Corollary 2 of Dedecker and Prieur [6], we have (3.7) and there exists a sequence $(Z_t)_{1\leqslant t \leqslant T}$ of independent $N(0;\varsigma^2)-$distributed random variables such that :

$$\sum_{t=1}^{T}\Big(X_t - \mathbb{E}(X_1)\Big) = \sum_{t=1}^{T} Z_t + o\Big(\sqrt{T\ln\ln(T)}\Big) \; a.s.; \tag{3.11}$$

where $\varsigma^2$ is defined in (3.7).
According to (3.10) and (3.11), we have, for $T$ big enough :

$$\sqrt{T}\mathbb{E}(K_1(x))\Big(\widehat{f}_T(x) - f(x) + o(1)\Big) + o\Big(\sqrt{\ln\ln(T)}\Big) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \;\; a.s. \tag{3.12}$$

From the Central Limit Theorem, we have for $T$ big enough :

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \xrightarrow{d} N(0,\varsigma^2).$$

Back to (3.12), we get (3.8). ■

## 4. Simulation study

In this section we present some results of our simulation study. We first (Section 4.1) focus on the strong consistency of estimator of regression function defined in (2.2). And we verify numerically the asymptotic normality of this estimator in Section 4.2. The simulation study was performed using R software and the results presented in these simulations correspond to 200 replications. Here, the Orlicz space is $L^1(\mathbb{R})$ and we use the absolute value function as Orlicz function.
Let $f$ be the function from $\mathbb{R}$ to $\mathbb{R}$ defined by :

$$f : x \longmapsto 0.2x. \tag{4.1}$$

We consider :

$$X_t = f(X_{t-1}) + \xi_t, \; t = 1, ..., T; \tag{4.2}$$

where $X_0 = 0$ and $(\xi_t)_t$ is a sequence of independent identically uniformly distributed on $[-0.3, 0.3]$.

We choose the uniform kernel on $[0,1]$; for the bandwidth, we choose $h_T = T^{-1/6}$. We numerically verify (3.1) and (3.8) at point $0$.

### 4.1. Simulation of strong consistency of $\widehat{f}_T(0)$

The samples are taken with size which varies between 100 and 500 observations. Table 4.1 reports the root mean square error (RMSE). The RMSE is calculated from the following formula :

$$RMSE = \sqrt{\frac{1}{r}\sum_{i=1}^{r}(\widehat{f}_{T,r}(0) - f(0))^2},$$

where $r$ denotes the number of replications (here $r = 200$) and $\widehat{f}_{T,r}(0)$, the value of $\widehat{f}_T(0)$ at the $r^{th}$ replication (see (2.2) for the definition of $\widehat{f}_T(0)$).
As it can be seen in Table 4.1, the RMSE decreases when the sample size increases. This corroborates the convergence of estimator.

| $T$ | *RMSE* |
|---|---|
| 100 | 0.018896 |
| 200 | 0.011016 |
| 500 | 0.007764 |

Table 4.1 : *RMSE values*

### 4.2. Simulation of asymptotic normality of $\widehat{f}_T(0)$

The purpose of this subsection is to illustrate the asymptotic normality of estimator $\widehat{f}_T(0)$ (see (3.8)). To this purpose, we randomly generate samples of size $T \in \{100, 300, 500\}$ of $\widehat{f}_T(0)$. Figure 4.1 shows the histogram and the $Q-Q$ plot of the estimator $\widehat{f}_{500}(0)$. In addition to these graphical representations, we performed a Shapiro-Wilk normality test. The results of the test are presented in Table 4.2 where $W$ refers to the test statistic.

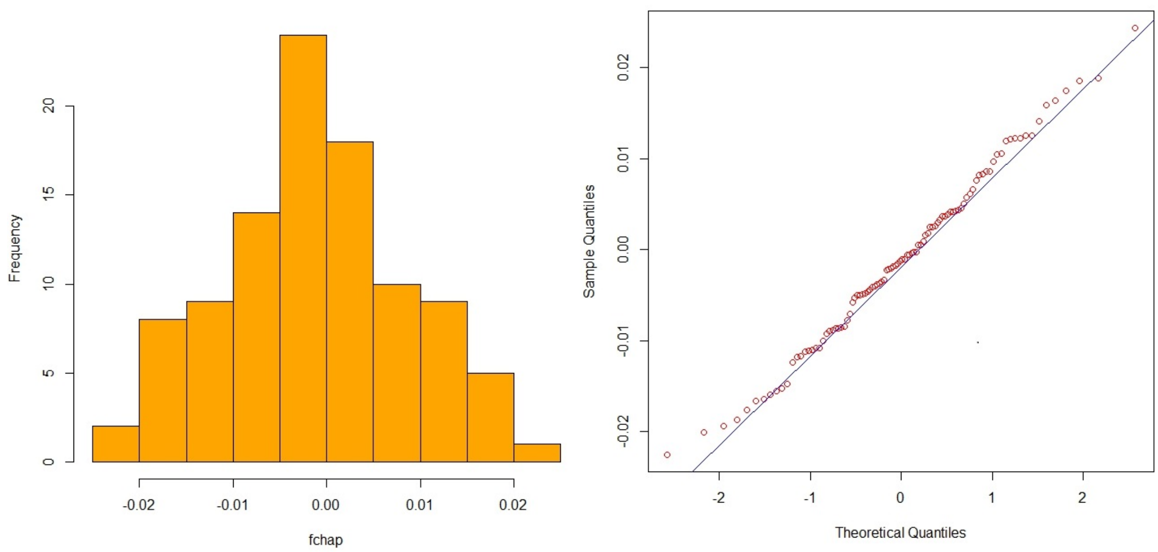


Figure 4.1 : Graphical illustration of the normality of $\widehat{f}_{500}(0)$.

Figure 4.1 is composed of two sub-figures: an histogram (on the left) and a $Q-Q$ plot (on the right). On the left side of Figure 4.1, we have plotted the histogram of $\widehat{f}_{500}(0)$ (orange colour). The shape of the histogram reminds us of the graphical representation of the density of normal distribution. This presumption is accentuated with the quantile cloud of dots. On the right side of Figure 4.1, we have plotted $Q-Q$ plot in red and Henry's line in blue. Most of the points seem to line up with Henry's line. And the extremities of the cloud seem to move away from it. Figure 4.1 therefore shows a presumption of normality of the sample. To confirm the normality of sample, we have performed the Shapiro-Wilk test. The test results show high values of $p-value$. This value increases when the sample size increases. In view of results, we can confirm the normality of these samples .

| $T$ | $W$ | *p-value* |
|---|---|---|
| 100 | 0.98869 | 0.5604 |
| 300 | 0.98961 | 0.6334 |
| 500 | 0.99248 | 0.8547 |

Table 4.2 : Shapiro-Wilk normality test on $\widehat{f}_T(0)$, $T \in \{100, 300, 500\}$.

## References


[1] D.W.K. Andrews , Nonstrong mixing autoregressive processes, *Journal of Applied Probability,* **21(4)**(1984), 930—934.

[2] B. Bercu, T. M. N. Nguyen and J. Saracco, On the asymptotic behaviour of the Nadaraya-Watson estimator associated with the recursive sliced inverse regression method, *A journal of Theoretical and Applied Statistics,* **49(3)**(2015), 660–679.

[3] P. Billingsley , Probability and Measure : Third Edition, *WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS ,* 1995.

[4] Z. Cai, Weighted Nadaraya–Watson regression estimation. *Statistics and Probability Letters*, **51(3)**(2001), 307–318.

[5] G. Collomb and W. Hardle , Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Processes and their Applications*, **23(1)**(1986), 77–89.

[6] J. Dedecker and C. Prieur , Coupling for $\tau-$Dependent Sequences and Applications. *Journal of Theoretical Probability*, **17(4)**(2004), 861–885.

[7] H. Dette, J. C. Pardo-Fernández and I. Van Keilegom , Goodness-of-Fit Tests for Multiplicative Models with Dependent Data. *Scandinavian Journal of Statistics*, **36(4)** (2009), 782—799.

[8] L. P. Devroye , The uniform convergence of the Nadaraya-Watson regression function estimate. *The Canadian Journal of Statistics*. **6(2)** (1978), 179–191.

[9] P. Doukhan and O. Wintenberger , Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications*, **118(11)**(2008), 1997—2013.

[10] J. Fan and Q. Yao, Nonlinear Time Series : Nonparametric and Parametric Methods, *Springer-Verlag New York, Inc.,* 2003.

[11] F. Ferraty and P. Vieu , Nonparametric Functional Data Analysis: Theory and Practice, *Springer,* 2006.

[12] L. Györfi, W. Härdle, P. Sarda and Vieu , Nonparametric Curve Estimation From Time Series,*Springer-Verlag New York*, **60(1)** 1989.

[13] S. Y. Hong and O. Linton , Nonparametric estimation of infinite order regression and its application to the risk-return tradeoff. *Journal of Econometrics*, **219(2)** (2020), 389–424.

[14] U. Krengel , Ergodic Theorems, *De Gruyter, Berlin,* 1985.

[15] P. Li, X. Li and L. Chen , The asymptotic normality of internal estimator for nonparametric regression, *Journal of Inequalities and Applications* 2018, 231.

[16] Y. P. MACK AND B. W. SILVERMAN , Weak and strong uniform consistency of kernel regression estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **61(3)** (1982), 405—415.

[17] E. A. NADARAYA , On estimating Regression. *Theory of Probability and its Applications*, **9(1)** (1964), 141–142.

[18] K. NODA , Estimation of a regression function by the Parzen kernel-type density estimators. *Annals of the Institute of Statistical Mathematics* **28(1)** (1976), 221–234.

[19] E. PARZEN , On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33(3)** (1962), 1065–1076.

[20] M. ROSENBLATT , Remarks on some non-parametric estimates of a density function. *Annals of Mathematical Statistics*, **27(3)** (1956), 832–837.

[21] E. F. SCHUSTER , Joint asymptotic distribution of the estimated regression function at a finite number of distinct points, *Annals of Mathematical Statistics* **43(1)** (1972), 84—88.

[22] G. S. WATSON, Smooth Regression Analysis. *Sankhya Series A* **26** (1964) 359–372.

[23] T. ZHU AND D. N. POLITIS, Kernel estimates of nonparametric functional autoregression models and their bootstrap approximation, *Electronic Journal of Statistics* **11 (2)** (2017), 2876-2906.