Ann. Inst. Statist. Math.
Vol. 53, No. 3, 447–468 (2001)
©2001 The Institute of Statistical Mathematics

ASYMPTOTIC NORMALITY OF KERNEL DENSITY ESTIMATORS UNDER DEPENDENCE

ZUDI LU

Insititut de Statistique, Université Catholique de Louvain, B-1348, Belgium and Laboratory of Management Decision and Information Systems, Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, China

(Received May 20, 1997; revised September 21, 1999)

Abstract. In this paper, we study the kernel methods for density estimation of stationary samples under generalized conditions, which unify both the linear and α-mixing processes discussed in the literature and also adapt to the non-linear or/and non-α-mixing processes. Under general, mild conditions, the kernel density estimators are shown to be asymptotically normal. Some specific theorems are derived within various contexts, and their applications and relationship with the relevant references are considered. It is interesting that the conditions on the bandwidth may be very simple, even in the generalized context. The stationary sequences discussed cover a large number of (linear or nonlinear) time series and econometric models (such as the ARMA processes with ARCH errors).

Key words and phrases: Asymptotic normality, α-mixing, linear process, kernel density estimators, stable stationary process, time series.

1. Introduction

Density estimation for dependent observations has received extensive attention in the literature. A number of papers have investigated it under various mixing conditions, including Robinson (1983) , Györfi *et al.* (1989), Roussas (1988), Tran (1989) and Irle (1997). Due to the fact that the mixings cannot cover all linear processes, Chanda (1983), Tran (1992) and Hallin and Tran (1996) explored the consistency and asymptotic normality of density estimators for some linear processes. In this paper, we consider kernel density estimators within a more general context which cover most of the time series models in the literature, unify both the stationary mixing and the linear processes and adapt to the nonlinear or/and non-mixing processes in practice.

We assume that $\{X_t\}$ is a stationary sequence defined on a probability space $(\Omega, \mathcal{F}, P)$. As may be known, α-mixing is the weakest among the widely used ϕ-, ρ-, β- and α-mixings. For later reference, its definition is given below.

DEFINITION 0. The stationary sequence $\{X_t, t = 0, \pm 1, \dots\}$ is α-mixing if

$$\alpha(k) = \sup_{A\in\mathcal{F}_{-\infty}^{n}, B\in\mathcal{F}_{n+k}^{\infty}} |P(AB) - P(A)P(B)| \to 0 \tag{1.1}$$

as $k \to \infty$, where $\mathcal{F}_{-\infty}^{n}$ and $\mathcal{F}_{n+k}^{\infty}$ are two σ-fields generated by $\{X_t, t \le n\}$ and $\{X_t, t \ge n + k\}$, respectively. The $\alpha(k)$ is called the mixing coefficient.

Under some suitable conditions, the stationary solutions of many time series (linear and nonlinear) models are α-mixing (cf., for example, Goródetskii (1977), Pham (1986) and Tong (1990)). From a practical point of view, however, the mixing concept may lead to some undesirable features. Firstly, it is not easy to check the α-mixing of the realizations of a stationary process in practice. Secondly, the requirement of the mixing dependence imposed upon all events (past and future) rules out some stationary processes of interests. As emphasized by Tran *et al.* (1996) and Hallin and Tran (1996), the mixing does not cover all linear processes. Andrews (1984) showed that the stationary solution of a simple linear AR(1) model —$X_t = 1/2X_{t-1}+e_t$ with e_t's being independent symmetric Bernoulli random variables (r.v.'s) taking values -1 and 1— is not α-mixing. Chanda (1983), Tran (1992) and Hallin and Tran (1996) thus developed the density estimation for time series under linear processes covering many important time series models (e.g., linear ARMA models and some long-memory fractional processes),

$$X_t = \sum_{r=0}^{\infty} a_r Z_{t-r}, \tag{1.2}$$

where $\{Z_t\}$ are i.i.d. r.v.'s with $EZ_t = 0$ and $EZ_t^2 = \sigma^2 < \infty$. Finally, a function of a mixing process is not readily the corresponding mixing. For example, even if $\{Z_t\}$ in (1.2) is α -mixing, $\{X_t\}$ in (1.2) needs not be α-mixing. In econometrics, the process $\{X_t\}$ in (1.2), with $\{Z_t\}$ being α-mixing, is of great interest. The ARMA process with ARCH errors discussed in Weiss (1984) and Engle (1982) is, for instance, of this form (note that the ARCH model proposed by Engle (1982) is α-mixing under mild conditions, cf., Lu (1996*a*, 1996*b*)). In addition, it is also noticed that the process in (1.2) with $\{Z_t\}$ being α-mixing may include some long-memory fractional processes with ARCH type errors (cf., Lin and Li (1997)).

Due to the above shortcomings, in this paper we first extend the concept of the α-mixing to a more general context in which a nonlinear form includes the linear process in (1.2) with $\{Z_t\}$ being α-mixing. That is,

$$X_t = g(Z_t, Z_{t-1}, Z_{t-2}, \ldots), \tag{1.3}$$

where $g : R^\infty \longrightarrow R^1$ is a Borel measurable function and $\{Z_t\}$ may be vector-valued. The i.i.d. process $\{Z_t\}$ is one of the simplest α-mixing stationary processes. To our knowledge, the idea of extending a mixing process to a function of the entire mixing process goes back to Ibragimov (1962) and was also formalized by Billingsley (1968) and Bierens (1983). We introduce the formal definition as follows:

Definition 1. The stationary process $\{X_t\}$ is stable in L_2 norm (L_2-stable for simplicity) with respect to (hereafter w.r.t.) the stationary α-mixing process $\{Z_t\}$, if

$$v(m) = E|X_t - X_t^{(m)}|^2 \to 0 \tag{1.4}$$

as $m \to \infty$, where $X_t^{(m)} = g_m(Z_t, \ldots, Z_{t-m+1})$, g_m is a Borel function with m arguments involved and $v(m)$ will be called the stable coefficients.

Remark 1. (a) This definition follows Bierens (1983) except that he defined a stable stationary process w.r.t. ϕ-mixing, a special case of the context here.

(b) As done in Bierens (1983), it is usually taken that $X_t^{(m)} = E(X_t \mid Z_t, \ldots, Z_{t-m+1})$. Clearly, $\{X_t^{(m)}\}$ is α-mixing with mixing coefficients $\alpha^*(k) \le 1$ for $k = 0, 1, \ldots, m$, and equals $\alpha(k-m)$ for $k \ge m+1$, where $\alpha(\cdot)$ is the mixing coefficient of $\{Z_t\}$.

(c) Clearly, we may define a L_p-stable process w.r.t. $\{Z_t\}$ with a p-th-order expectation instead of the second order in (1.4). The results in this paper can be easily adapted to such a situation and so we don't pursue this generality here.

(d) If g in (1.3) is of a linear form and $\{Z_t\}$ is i.i.d. as in (1.2), then $\alpha(k) \le 1$ for $k = 0$, and $= 0$ for $k \ge 1$. Further, assuming that $\sum_{r=0}^{\infty} a_r^2 < \infty$, the stable coefficient $v(m) = E|X_t - X_t^{(m)}|^2 = E|\sum_{r=m}^{\infty} a_r Z_{t-r}|^2 = \sigma^2 \sum_{r=m}^{\infty} a_r^2$. This is also considered in Chanda (1983), Tran (1992) and Hallin and Tran (1996). When $\{Z_t\}$ is α-mixing, our model framework covers the ARMA process with ARCH errors discussed in Weiss (1984) and Engle (1982) as well as some long-memory fractionally integrated processes with ARCH type errors.

(e) If g is of a nonlinear form, then many nonlinear time series models fall into our category, e.g., the bilinear and the random coefficient models in references such as Granger and Andersen (1978), Nicholls and Quinn (1982), Tjøstheim (1986) and Tong (1990).

(f) If $\{X_t\}$ itself is α-mixing stationary with $EX_t = 0$ and $EX_t^2 < \infty$, then setting $Z_t = X_t$ and g the identity function leads to the stable coefficient being $v(m) = 0$ for $m \ge 1$.

In Section 2, we define the kernel density estimators based on the realization of the L_2-stable stationary process and give a general result which shows the mild conditions under which the asymptotic normality of the density estimators is ensured. To shed light on the wide applications of this general result, some specific theorems are derived in Sections 3 and 4. In Section 3, we are concerned with the specific theorems under two contexts related with the relevant references, one covering the α-mixing processes considered in Robinson (1983) and the other deducing a result under linear processes better than Hallin and Tran (1996)'s (cf. Remark 6 below). It should be pointed out that the results obtained in Section 2 can also be applied to density estimation for the stationary sequences which are neither α-mixing nor linear processes. In this case, some mild specific conditions are also derived in Section 4 to guarantee the asymptotic normality of the density estimators. To our knowledge, no one has examined this case. The proofs of the theorems in Section 2 are postponed to Section 5.

2. General result

In the following, we let $X_1, X_2, \ldots, X_n$ be a realization of size n from the L_2-stable stationary process $\{X_t\}$ defined in Definition 1. In time-series analysis, estimating a one-dimensional marginal density is of certain interest (cf., Hallin and Tran (1996)), but even more interesting is to consider the estimation of a marginal joint density because a one-dimensional marginal density cannot capture the dependence of the stationary sequence. For background on estimating the marginal joint density, the reader is referred to, for example, Tjøstheim (1996). Generally speaking, we may consider estimating a d-dimensional marginal joint density of $(X_{i-\tau_{d-1}}, \ldots, X_{i-\tau_1}, X_i)'$, say, where $0 < \tau_1 < \cdots < \tau_{d-1} < +\infty$ are positive integers and Y' denotes the transpose of vector Y. For simplicity of notation, we are concerned with $Y_i = (X_{i-d+1}, \ldots, X_{i-1}, X_i)'$ in this paper (the results obtained can be adapted for general τ_k's). Let $f(y)$ be the marginal joint density of Y_i. The kernel density estimator of $f(y)$ is therefore defined by

$$f_n(y) = (nh_n^d)^{-1} \sum_{i=d}^{n} K((y - Y_i)/h_n). \tag{2.1}$$

Here $K(\cdot)$ is a kernel function defined on R^d, and $h_n > 0$ are bandwidths tending to zero as $n \to \infty$.

Our purpose in this paper is to investigate the asymptotic normality of the kernel density estimators for stable stationary sequences. In this section, we first give a general result which will be applied specifically in Sections 3 and 4, and its usefulness will be seen later on. Throughout the paper, C will denote a generic constant which may differ at different places.

ASSUMPTION 1. $\{Z_t\}$ is α-mixing stationary with the mixing coefficients $\alpha(\cdot)$ satisfying

$$k \sum_{j=k}^{\infty} \alpha(j) \to 0 \quad \text{as} \quad k \to \infty. \tag{2.2}$$

ASSUMPTION 2. $\{X_t\}$ is L_2-stable stationary w.r.t. $\{Z_t\}$ with stable coefficients $v(m)$ as defined in Definition 1. Let $Y_t = (X_{t-d+1}, \dots, X_{t-1}, X_t)'$ have marginal joint density $f(y)$.

ASSUMPTION 3. The kernel function K is a bounded density function with an integrable radial majorant $Q(x)$, that is $Q(x) = \sup\{K(y) : \|y\| \geq \|x\|\}$ is integrable. Assume, in addition, that K satisfies the following Lipschitz condition:

$$|K(x) - K(y)| \leq C\|x - y\|, \tag{2.3}$$

where $\|\cdot\|$ is an Euclidean norm of R^d.

ASSUMPTION 4. (i) The density function f of Y_t is Lipschitz continuous, i.e., for any x, $y \in R^d$,

$$|f(x) - f(y)| \leq C\|x - y\|; \tag{2.4}$$

(ii) The joint density $f_j(x, y)$ of (Y_0, Y_j) is bounded uniformly in j (> 0), that is

$$\sup_j \sup_{(x,y)\in R^d \times R^d} f_j(x, y) \leq C, \tag{2.5}$$

where if $j < d$, (Y_0, Y_j) reads as $(X_{-d+1}, \dots, X_{j-d+1}, \dots, X_0, \dots, X_j) \in R^{j+d}$.

ASSUMPTION 5. The bandwidth h_n satisfies that, as $n \to \infty$,
(i) $h_n \to 0$, $nh_n^d \to \infty$, $nh_n^{2d} \to 0$,
(ii) $nh_n^{2+d} \to 0$.

ASSUMPTION 6. There exist two sequences of positive integers, $p(n)$ and $q(n)$, such that as $n \to \infty$,

(i) $p(n) \to \infty, \quad p^3(n)/n \to 0,$
(ii) $q(n) \to \infty, \quad q(n)/p(n) \to 0,$
(iii) $q(n)h_n^d \to 0,$
(iv) $n\alpha(q(n))/p(n) \to 0,$
(v) $v(q(n)) = O(h_n^{3d+2}).$

Remark 2. Among the above assumptions, (i) Assumption 1 is the condition on the mixing dependence by Robinson ((1983), A3.1, p. 189) for density estimation; (ii) Assumptions 3, 4(i) and 5(ii) with $d = 1$ are the same as Assumptions 1, 4 and 5 in Hallin and Tran ((1996), pages 432, 443) respectively; (iii) Assumption 4(ii) on the joint density is often assumed in the literature (cf., Robinson (1983), A4.5, p. 191) and is easily satisfied by time series models (cf., Lemma 5.1 of Hallin and Tran 1996, p. 446, for the linear processes (1.2)). We assume, without loss of generality, that the mixing and the stable coefficients, $\alpha(\cdot)$ and $v(\cdot)$, in Assumptions 1 and 2 are monotonously decreasing.

Remark 3. Assumption 6 is general but looks cumbersome. In Sections 3 and 4 below, we will specify some more explicit and easily verifiable conditions to ensure Assumption 6. From them, the existence of $p(n)$ and $q(n)$ will become clear.

The asymptotic normality of the kernel density estimators was studied by Robinson (1983) under α-mixing and Hallin and Tran (1996) under linear processes. The following theorems extend their contexts to stable stationary processes w.r.t. α-mixing:

Corresponding to Theorem 3.1 of Hallin and Tran ((1996), p. 443), we have

THEOREM 1. *Suppose that Assumptions* 1–3, 4(ii), 5(i) *and* 6 *hold and* $y_1, \ldots, y_k$ *are* k *distinct points of* R^d, *then*

$$(nh_n^d)^{1/2}(f_n(y_1) - Ef_n(y_1), \ldots, f_n(y_k) - Ef_n(y_k))' \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{C}) \tag{2.6}$$

where $\mathbf{0} = (0, \ldots, 0)' \in R^k$, $\boldsymbol{C}$ *is a diagonal matrix with elements* $C_{ii} = f(y_i) \int_{R^d} K^2(u)du$, $i = 1, \ldots, k$, *and* "$\to^{\mathcal{L}}$" *denotes the convergence in distribution.*

The next theorem corresponds to Theorem 3.2 of Hallin and Tran (1996, p. 443).

THEOREM 2. *If Assumptions* 1–6 *hold and, in addition,*

$$\int_{R^d} \|u\| K(u)du < \infty, \tag{2.7}$$

then for any k and any distinct points $y_1, \ldots, y_k \in R^d$,

$$(nh_n^d)^{1/2}(f_n(y_1) - f(y_1), \ldots, f_n(y_k) - f(y_k))' \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{C}). \tag{2.8}$$

3. Specific theorems related with relevant references

In Section 2, we gave a general result in which $p(n)$ and $q(n)$ are not defined explicitly in Assumption 6. Notice that Assumptions 6(iv) and 6(v) are related to the mixing coefficients, $\alpha(\cdot)$, and the stable coefficients, $v(\cdot)$, respectively. Different definitions of $p(n)$ and $q(n)$ will lead to different conditions on the mixing and the stable coefficients. In this section, we shall derive some specific conditions which will cover the contexts considered in Robinson (1983) (α-mixing) and in Hallin and Tran (1996) (linear processes), respectively. An easily verifiable condition within a generalized context will be provided in Section 4.

3.1 *Case* I *for Robinson* (1983)

Robinson (1983) considered the setting where $X_t = Z_t$ with the α-mixing coefficients satisfying Assumption 1 in Section 2. To keep this condition on the mixing coefficients hold, we may choose $p(n)$ and $q(n)$, as done in the proof of Lemma 7.1 of Robinson ((1983), p. 199), that is

$$p(n) \sim n^{1/3}\eta_n, \qquad q(n) \sim n^{1/3}\eta_n^2, \quad r(n) \sim n/(p(n)+2q(n)),$$

where $a_n \sim b_n$ means that $\lim_{n\to\infty} a_n/b_n = 1$, and η_n is a positive sequence such that $p(n)$, $q(n)$ and $r(n)$ are nondecreasing, $\eta_n \to 0$ and

$$\eta_n > \max(n^{-1/24}, \tilde{\varepsilon}_{n^{1/4}}^{1/6}), \qquad \text{where} \quad \tilde{\varepsilon}_n = \sup_{N\geq n} \varepsilon_N, \ \ \varepsilon_N = N\sum_{j=N}^{\infty} \alpha(j).$$

Thus, as Robinson ((1983), p. 199) proved, Assumptions 6(i), 6(ii) and 6(iv) hold; $q(n)h_n^d \sim n^{1/3}\eta_n^2 h_n^d \to 0$ follows from Assumption 5(i) and, hence, Assumption 6(iii) holds.

To verify Assumption 6(v), we impose the following condition on the stable coefficient and the bandwidth:

ASSUMPTION 2'. The stable coefficients, $v(\cdot)$, in Assumption 2 satisfy

$$v(n^{1/4}) = O(h_n^{3d+2}). \tag{3.1}$$

Clearly, if we choose $q(n) \sim n^{1/3}\eta_n^2 > n^{1/3}(n^{-1/24})^2 = n^{1/4}$, then $v(q(n)) \leq v(n^{1/4})$ (recall we assume $v(\cdot)$ is monotonously decreasing). Assumption 6(v) thus follows from Assumption $2'$.

THEOREM 3. *If Assumptions* 1, 2' *and* 3–5 *hold and, in addition,* (2.7) *in Theorem* 2 *is satisfied, then the conclusion of Theorem* 2 *holds.*

If $\{X_t\}$ is α-mixing, then from Remark 1(f) in Section 1, it follows that $v(j) = 0$ for $j \geq 1$. Thus, Assumption $2'$ holds naturally.

COROLLARY 1. *If $\{X_t\}$ is an α-mixing stationary sequence with mixing coefficients, $\alpha(\cdot)$, satisfying* (2.1) *in Assumption* 1 *and, in addition, Assumptions* 3–5 *and* (2.7) (*in Theorem* 2) *all are satisfied, then the conclusion of Theorem* 2 *holds.*

Remark 4. From Corollary 1, Theorem 3 clearly covers the context of α-mixing for density estimation considered in Robinson ((1983), Theorem 4.1, p. 191) with a positive definite matrix-valued bandwidth instead of a scalar bandwidth.

3.2 *Case* II *for Hallin and Tran* (1996)

Hallin and Tran (1996) explored the asymptotic normality of the kernel density estimators under linear processes (1.2). Their results cannot be derived well from Theorem 3 above because Assumption $2'$ does not always hold under their conditions (see Remark

7 below). In this subsection, we slightly strengthen the condition imposed on the mixing coefficients but weaken the condition on the stable coefficients. Thus, the specific theorem obtained will easily produce a better result for linear processes than Hallin and Tran (1996)'s.

ASSUMPTION 1^o. The condition on the mixing coefficients, $\alpha(\cdot)$, in Assumption 1 is strengthened to

$$k^a\alpha(k) \to 0 \quad (k \to \infty) \tag{3.2}$$

for some $2 < a < \infty$.

ASSUMPTION 2^o. The stable coefficients, $v(\cdot)$, in Assumption 2 satisfy

$$v(n^{3/(1+4a)}) = O(h_n^{3d+2}). \tag{3.3}$$

THEOREM 4. *If Assumptions 1^o, 2^o and 3–5 hold and, in addition, (2.7) in Theorem 2 is satisfied, then the conclusion of Theorem 2 holds.*

PROOF. First, take $n/p(n) \sim q^a(n)$, then Assumption 6(iv) follows from (3.2) in Assumption 1^o. To make $p^3(n)/n \to 0$ and $q(n)/p(n) \to 0$ in Assumptions 6(i) and 6(ii) hold simultaneously, we let $q(n)/p(n) \sim p^3(n)/n$. Thus, we may take $q(n) \sim n^{3/(1+4a)}$, $p(n) \sim n/q^a(n) \sim n^{(1+a)/(1+4a)}$. By a simple calculation, $q(n)/p(n) \sim p^3(n)/n \sim n^{-(a-2)/(1+4a)} \to 0$ for $2 < a < \infty$ and, hence, Assumptions 6(i) and 6(ii) clearly hold. Assumption 6(v) follows from Assumption 2^o. Finally, $q(n)h_n^d \sim (nh_n^{(1+4a)d/3})^{3/(1+4a)} \to 0$ follows easily from Assumption 5(i). Now Theorem 4 is clear from Theorem 2 in Section 2.

Hallin and Tran (1996) considered the estimation of one-dimensional marginal density ($d = 1$ in the setting of this paper) for the linear processes (1.2) with $|a_r| = O(r^{-(4+\delta)})$ for some $\delta > 0$ (as $r \to \infty$) in their Assumption 2.

HT'S ASSUMPTION 2. The coefficients of the linear process X_t (in (1.2)) tend to zero sufficiently fast that $|a_r| = O(r^{-(4+\delta)})$ for some $\delta > 0$ as $r \to \infty$. In addition, Z_t (in (1.2)) has mean zero and finite variance and an absolutely integrable characteristic function.

From Theorem 4, we may easily derive the following corollary:

COROLLARY 2. *If HT's Assumption 2 above and our Assumptions 3, 4(i) and 5 and (2.7) in Section 2 with $d = 1$ hold and, in addition,*

$$\liminf_{n\to\infty} n^{3(7+2\delta)/(1+4a)}h_n^5 > 0 \quad \text{for some} \quad a \in (2, (37+12\delta)/20) \text{ and some } \delta > 1/4, \tag{3.4}$$

then the conclusion of Theorem 2 with $d = 1$ holds.

PROOF. First, from Remark 1(d) and HT's Assumption 2, it follows that

$$v(m) = \sigma^2 \sum_{r=m}^{\infty} O(r^{-2(4+\delta)}) = O(m^{-(7+2\delta)}) \tag{3.5}$$

as $m \to \infty$ and, hence, together with (3.4),

$$v(n^{3/(1+4a)}) = O(n^{-3(7+2\delta)/(1+4a)}) = O(h_n^5).$$

Thus, Assumption 2^o with $d = 1$ is met. Assumption 1^o holds naturally since $\{Z_t\}$ is i.i.d. Assumption 4(ii) is clear by Lemma 5.1 of Hallin and Tran ((1996), p. 446) (cf., Remark 2 above). Hence, Corollary 2 follows from Theorem 4.

Remark 5. First, we point out a negligence in Hallin and Tran (1996)'s Assumption 5 and Remark 3.1. From the proof of their Theorem 3.2, it is clear that their Assumption 5 should be $nh_n^3 \to 0$, not $nh_n^3 \to \infty$, as $n \to \infty$. Hence, the assertion in their Remark 3.1 that "their Assumption 3 (i.e., $nh_n^{(13+2\delta)/(3+2\delta)}(\log\log n)^{-1} \to \infty$ as $n \to \infty$) implies their Assumption 5 when $\delta < 1$" is not right.

Remark 6. From Remark 5 above and the conditions of Hallin and Tran (1996)'s Theorem 3.2, we know that $(13 + 2\delta)/(3 + 2\delta) < 3$, i.e., $\delta > 1$, should be needed to guarantee their Assumptions 3 and 5 simultaneously. However, in our Corollary 2, when $\delta > 1/4$, (3.4) may be met for some a (since $\delta > 1/4$, $2 < (37 + 12\delta)/20$, so a may be chosen between 2 and $(37 + 12\delta)/20$, and thus $5(1 + 4a)/[3(7 + 2\delta)] < 2$. Our Assumption 5 with $d = 1$ and (3.4) may hold simultaneously). Hence, our Corollary 2 improves Theorem 3.2 of Hallin and Tran (1996).

Remark 7. From (3.5), $v(n^{1/4}) = O(n^{-(7+2\delta)/4})$. To make Assumptions $2'$ and 5 with $d = 1$ hold simultaneously, it is necessary that $4 \times 5/(7 + 2\delta) < 2$, that is $\delta > 3/2$. Hence, the result following from Theorem 3 cannot always deduce Hallin and Tran (1996)'s Theorem 3.2 ($\delta > 1$, cf., Remark 6).

4. Explicit conditions under the generalized context

Insofar as the incremental contributions cover nonlinear processes, there is more interest. In this section, under the generalized context, we will derive some explicit conditions to ensure Assumption 6, thus the results obtained will apply well practically for the nonlinear or/and non-mixing processes.

By choosing $p(n)$ and $q(n)$ in (iv) and (v) of Assumption 6 to satisfy

$$n/p(n) = q(n)^a, \qquad h_n^{-(3d+2)} = q(n)^b, \tag{4.1}$$

where a and b are some positive constants to be specified below, we found

LEMMA 1. *The conditions in Assumption* 6' *below are sufficient to ensure Assumption* 6.

ASSUMPTION 6'. (i) $j^a\alpha(j) \to 0$, as $j \to \infty$, for some $a > 2$;
(ii) $j^b v(j) = O(1)$, as $j \to \infty$, for some $b > (3d + 2)/d$;
(iii) $nh^{(1+a)(3d+2)/b} \to \infty$, $nh^{3a(3d+2)/(2b)} \to 0$, as $n \to \infty$.

PROOF. First, Assumptions 6(iv) and 6(v) are satisfied by (4.1) and Assumptions $6'$(i) and $6'$(ii). Next, it follows from (4.1) that $q(n) = h_n^{-(3d+2)/b}$ and $p(n) = nh_n^{a(3d+2)/b}$. Thus Assumptions 6(i), 6(ii) and 6(iii) are easily checked by Assumptions $6'$(iii) together with $a > 2$ and $b > (3d+2)/d$ in (i) and (ii) of Assumption 6′. Note that $a > 2$ is to ensure (iii) here, and $b > (3d+2)/d$ is to ensure Assumption 6(iii).

It is noted that Assumption 6′(i) is the same as Assumption 1°. To guarantee the conditions imposed on the bandwidth in Assumptions 5 and 6′(iii) hold simultaneously, the following conditions are necessary:

ASSUMPTION 1*. The bandwidth h_n satisfies that as $n \to \infty$,
(i) $h_n \to 0$,
(ii) $nh_n^{\max\{d,(1+a)(3d+2)/b\}} \to \infty$,
(iii) $nh_n^{\min\{2d,2+d,3a(3d+2)/(2b)\}} \to 0$.

In order to guarantee (ii) and (iii) in Assumption 1*, a and b should be chosen such that

$$3a(3d+2)/(2b) > d, \qquad 2d > (1+a)(3d+2)/b,$$
$$2+d > (1+a)(3d+2)/b, \qquad 3a(3d+2)/(2b) > (1+a)(3d+2)/b.$$

Thus, $a > 2$ and

$$\frac{(1+a)(3d+2)}{\min(2d, 2+d)} < b < \frac{3a(3d+2)}{2d} \tag{4.2}$$

(since $a > 2$ and $d \geq 1$, this is not empty). Hence, Assumption 6′(ii) needs modifying as follows:

ASSUMPTION 2*. The stable coefficients, $v(\cdot)$, in Assumption 2 satisfy $j^b v(j) = O(1)$, as $j \to \infty$, for some b satisfying (4.2).

THEOREM 5. *If Assumptions* 1°, 1*, 2*, 3 *and* 4 *hold and, in addition,* (2.7) *in Theorem* 2 *is satisfied, then the conclusion of Theorem* 2 *holds.*

Furthermore, if we take $b = (1+a)(3d+2)/d$, then Assumptions 1* and 2* become simpler.

ASSUMPTION 1**. The bandwidth h_n satisfies that, as $n \to \infty$, (i) $h_n \to 0$, (ii) $nh_n^d \to \infty$, (iii) $nh_n^{\min\{2+d,3ad/[2(1+a)]\}} \to 0$.

ASSUMPTION 2**. The stable coefficients, $v(\cdot)$, in Assumption 2 satisfy

$$j^{(1+a)(3d+2)/d} v(j) = O(1), \qquad \text{as} \quad j \to \infty.$$

COROLLARY 3. *If Assumptions* 1°, 1**, 2**, 3 *and* 4 *hold and, in addition,* (2.7) in Theorem 2 *is satisfied, then the conclusion of Theorem* 2 *holds.*

Remark 8. It is interesting to note that the conditions on the bandwidth in Assumption 1** are very simple, even under our generalized context, in particular $nh_n^d \to \infty$ in Assumption 1**(ii) is the same as the well-known condition for the i.i.d. samples.

Finally to indicate the application of the results in this section, we consider an example of the ARMA process with ARCH errors, popular in econometrics, defined by

$$\phi(B)X_t = (1-\phi_1 B - \cdots - \phi_p B^p)X_t = \theta(B)Z_t = (1-\theta_1 B - \cdots - \theta_q B^q)Z_t,$$
$$Z_t = e_t h_t^{1/2}, \qquad h_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \cdots + \alpha_r Z_{t-r}^2,$$

where X_t, Z_t and e_t are all adapted to the σ-field $\mathcal{F}_t$ of information collection up to time t, $\{e_t\}$ is an i.i.d. standard normal sequence with e_t independent of $\mathcal{F}_{t-1}$, B is the back shift operator, and p, q and r are positive integers. If the ϕ_i's, θ_i's, and $\alpha_0(>0)$, $\alpha_i(\geq 0)$'s are constants, such that all the roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle and $\alpha_1+\cdots+\alpha_r<1$, then the stationary solution $\{X_t\}$ can be expressed as (1.2) with Z_t being α-mixing. Here, it is difficult to embed $\{X_t\}$ into the context of α-mixing or linear processes, but it easily embeds into the generalized context in this section. The other conditions in Corollary 3 can also be checked carefully, but these are not written out in detail here due to our attention being focused on the generalized context.

5. Proofs for theorems in Section 2

5.1 *Basic lemmas*

For references later on, some basic lemmas are collected in this subsection.

LEMMA 2. *If X and Y are two random variables which are measurable with respect to $\mathcal{A}$ and $\mathcal{B}$, respectively, and there exist two constants C_1, C_2, such that $|X| \leq C_1$, $|Y| \leq C_2$, a.s., then*

$$|EXY - EXEY| \leq 4C_1C_2\alpha(\mathcal{A},\mathcal{B}), \tag{5.1}$$

where $\mathcal{A}$ and $\mathcal{B}$ are two σ-algebras, $\alpha(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B}}|P(AB)-P(A)P(B)|$.

PROOF. See the Appendix of Hall and Heyde (1980).

LEMMA 3. *If Assumption 3 holds and $h_n \to 0$ as $n\to\infty$, then for any $x, y \in R^d$ with $x \neq y$,*

$$h_n^{-d}EK((x-Y_1)/h) \to f(x), \tag{5.2a}$$

$$h_n^{-d}EK^2((x-Y_1)/h) \to f(x)\int K^2(u)du, \tag{5.2b}$$

$$h_n^{-d}EK((x-Y_1)/h)K((y-Y_1)/h) \to 0, \tag{5.2c}$$

as $n\to\infty$. Furthermore, if Assumption 4(ii) also holds, then for $j>0$,

$$h_n^{-d}EK((x-Y_1)/h)K((x-Y_{j+1})/h) = O(h_n^{\min(d,j)}), \tag{5.2d}$$

$$h_n^{-d}EK((x-Y_1)/h)K((y-Y_{j+1})/h) = O(h_n^{\min(d,j)}), \tag{5.2e}$$

where (5.2d) and (5.2e) hold uniformly for $j \geq d$.

PROOF. For the proof of (5.2*a*) and (5.2*b*), see Devroye and Györfi ((1985), Theorem 3, p. 8) (cf., Lemma 2.1 of Hallin and Tran (1996)); (5.2*c*) is the vector version of Lemma 2.6 of Hallin and Tran (1996) (see Masry (1986) for the proof); (5.2*d*) and (5.2*e*) are also easily derived by the conditions. So the proof of this lemma is omitted.

5.2 *Technical lemmas*

Define

$$K_{ni}^{(q)}(x) = K((x - Y_i^{(q)})/h) - EK((x - Y_i^{(q)})/h), \tag{5.3}$$

where $Y_i^{(q)} = (X_{i-d+1}^{(q)}, \ldots, X_{i-1}^{(q)}, X_i^{(q)})'$, $X_i^{(q)}$ is defined in Definition 1 with m replaced by $q = q(n)$ and $q(n)$ are positive integers with $q(n) \to \infty$ as $n \to \infty$.

Throughout this subsection, we all assume that Assumptions 1–4 hold. For convenience of writing, set $h = h_n$, $p = p(n)$ and $q = q(n)$ and $r = r(n) \sim n/(p + 2q)$. All the limits are taken as $n \to \infty$ except specified otherwise.

LEMMA 4. *If $h \to 0$, $h^{-(d+2)}v(q) \to 0$, then for any $x, y \in R^d$ with $x \neq y$,*

$$h^{-d}E(K_{n1}^{(q)}(x))^2 \to f(x)\int K^2(u)du, \tag{5.4a}$$

$$h^{-d}EK_{n1}^{(q)}(x)K_{n1}^{(q)}(y) \to 0, \tag{5.4b}$$

$$h^{-d}E|K_{n1}^{(q)}(x)K_{n,1+j}^{(q)}(x)| = O(h^{\min(d,j)}) + O((h^{-(d+2)}v(q))^{1/2}), \tag{5.4c}$$

$$h^{-d}E|K_{n1}^{(q)}(x)K_{n,1+j}^{(q)}(y)| = O(h^{\min(d,j)}) + O((h^{-(d+2)}v(q))^{1/2}), \tag{5.4d}$$

where (5.4*c*) *and* (5.4*d*) *hold uniformly for $j \geq d$.*

PROOF. Since the proofs are completely similar, we only prove (5.4*a*) in detail as follows:

First observe that

$$\begin{aligned} E(K_{n1}^{(q)}(x))^2 &= E(A_1(x) + K((x - Y_1)/h))^2 \\ &\quad -(EA_1(x) + EK((x - Y_1)/h))^2 \\ &= E(A_1(x))^2 + 2E(A_1(x)K((x - Y_1)/h)) + E(K((x - Y_1)/h))^2 \\ &\quad -(EA_1(x))^2 - 2EA_1(x)EK((x - Y_1)/h) - (EK((x - Y_1)/h))^2, \end{aligned} \tag{5.5}$$

where $A_i(x) = K((x - Y_i^{(q)})/h) - K((x - Y_i)/h)$.

In view of Assumption 3 and Definition 1,

$$\begin{aligned} E|A_1(x)|^2 \leq Ch^{-2}E\|Y_1^{(q)} - Y_1\|^2 &= Ch^{-2}\sum_{j=1}^{d} E\|X_{i-j+1}^{(q)} - X_{i-j+1}\|^2 \\ &= O(h^{-2}v(q)). \end{aligned} \tag{5.6}$$

Hence, by (5.6) and Lemma 5.3,

$$\begin{aligned} &h^{-d}E(A_1(x))^2 = O(h^{-(2+d)}v(q)), \\ &h^{-d}E(A_1(x)K((x - Y_1)/h)) \leq h^{-d}(E(A_1(x))^2)^{1/2}(EK^2((x - Y_1)/h))^{1/2} \\ &\quad = O((h^{-(2+d)}v(q))^{1/2}), \\ &h^{-d}E(K((x - Y_1)/h))^2 \to f(x)\int K^2(u)du, \\ &h^{-d}(EA_1(x))^2 \leq h^{-d}E(A_1(x))^2 = O(h^{-(2+d)}v(q)), \\ &h^{-d}|EA_1(x)EK((x - Y_1)/h)| \leq h^{-d}(E(A_1(x))^2)^{1/2}EK((x - Y_1)/h) \\ &\quad = O((h^{-2}v(q))^{1/2}), \\ &h^{-d}(EK((x - Y_1)/h))^2 = O(h^d). \end{aligned}$$

Thus, (5.4a) easily follows from (5.5).

Define

$$f_n^{(q)}(y) = (nh_n^d)^{-1} \sum_{i=d}^{n} K((y - Y_i^{(q)})/h_n).$$

LEMMA 5. *If* $nv(q)/h^{2+d} \to 0$, *then*

$$(nh^d)^{1/2} |f_n(y) - f_n^{(q)}(y)| \xrightarrow{P} 0, \tag{5.7a}$$

$$(nh^d)^{1/2} |Ef_n(y) - Ef_n^{(q)}(y)| \to 0. \tag{5.7b}$$

PROOF. First by (5.6), for any $\epsilon > 0$

$$\begin{aligned} P((nh^d)^{1/2}|f_n(y) - f_n^{(q)}(y)| > \epsilon) &\le \epsilon^{-2}(nh^d)E|f_n(y) - f_n^{(q)}(y)|^2 \\ &= (\epsilon^2 nh^d)^{-1} E\left(\sum_{i=d}^{n} A_i(y)\right)^2 \\ &\le (\epsilon^2 h^d)^{-1} \sum_{i=d}^{n} E(A_i(y))^2 \\ &= O(nv(q)/h^{2+d}) \to 0, \end{aligned}$$

from which (5.7a) follows, where $A_i(y)$ was defined in (5.5). (5.7b) may be proved easily.

In the following, we are to try to prove

$$B_n = (nh_n^d)^{1/2}(f_n^{(q)}(y_1) - Ef_n^{(q)}(y_1), \ldots, f_n^{(q)}(y_k) - Ef_n^{(q)}(y_k))' \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{C}), \tag{5.8}$$

where $y_1, \ldots, y_k$ are distinct points in R^d. This is the key step of our proofs of the main results. Since $q = q(n)$ changes with n, Lemma 7.1 of Robinson (1983) cannot be applied directly. For simplicity of exposition, we consider $k = 2$. Set $c = (c_1, c_2)' \in R^2$.

$$\begin{aligned} c'B_n &= (nh_n^d)^{1/2}(c_1(f_n^{(q)}(y_1) - Ef_n^{(q)}(y_1)) + c_2(f_n^{(q)}(y_2) - Ef_n^{(q)}(y_2))) \\ &= (nh_n^d)^{-1/2} \sum_{i=d}^{n} (c_1 K_{ni}^{(q)}(y_1) + c_2 K_{ni}^{(q)}(y_2)), \end{aligned} \tag{5.9}$$

where $K_{ni}^{(q)}(y)$ was defined in (5.3).

For our purpose, we adopt Bernstein's technique.

Set $n - d = r(p + 2q) + s$, where $r = r(n)$ and $p = p(n)$ were defined right before Lemma 4, which tend to ∞ as $n \to \infty$, and $s = s(n)$ is a positive integer satisfying $0 \le s < p + 2q$ and $q = q(n)$ was defined right after (5.3). Denote

$$K_i = c_1 K_{ni}^{(q)}(y_1) + c_2 K_{ni}^{(q)}(y_2),$$

$$U_m = (nh_n^d)^{-1/2} \sum_{i=r_m}^{r_m+p-1} K_i, \qquad r_m = (m-1)(p + 2q) + d,$$

$$V_m = (nh_n^d)^{-1/2} \sum_{i=\ell_m}^{\ell_m+2q-1} K_i, \qquad \ell_m = (m-1)(p+2q)+p+d,$$
$$m = 1, 2, \ldots, r,$$

$$R_r = (nh_n^d)^{-1/2} \sum_{i=r(p+2q)+d}^{n} K_i.$$

Thus (5.9) can be written as

$$c'B_n = \sum_{m=1}^{r} U_m + \sum_{m=1}^{r} V_m + R_r. \tag{5.10}$$

Now we are to prove, as $n \to \infty$,

$$\sum_{m=1}^{r} V_m \xrightarrow{P} 0, \qquad R_r \xrightarrow{P} 0, \tag{5.11a}$$

$$\sum_{m=1}^{r} U_m \xrightarrow{\mathcal{L}} N(0, (c_1^2 f(y_1) + c_2^2 f(y_2)) \int K^2(u)du). \tag{5.11b}$$

LEMMA 6. *If the conditions of Lemma 4 hold, and* $q/p \to 0$, $v(q) = O(h^{3d+2})$, $qh^d \to 0$, $n\sum_{j=n}^{\infty} \alpha(j) \to 0$, *then (5.11a) holds.*

PROOF. By stationarity, it is clear that

$$E\left(\sum_{m=1}^{r} V_m\right)^2 = \sum_{m=1}^{r} EV_m^2 + 2\sum_{m=1}^{r-1}\sum_{j=1}^{r-m} EV_1V_{j+1} \tag{5.12}$$
$$\le rEV_1^2 + 2r\left|\sum_{j=1}^{r} EV_1V_{j+1}\right|.$$

First, consider the first term of (5.12). By stationarity,

$$\begin{aligned} EV_1^2 &= (nh_n^d)^{-1}E\left(\sum_{i=0}^{2q-1} K_i\right)^2 \\ &= (nh_n^d)^{-1}\sum_{i=0}^{2q-1} EK_i^2 + 2(nh_n^d)^{-1}\sum_{i=0}^{2q-2}\sum_{j=i+1}^{2q-1} EK_iK_j \\ &= (n)^{-1}(2q)(h_n^d)^{-1}EK_1^2 + 2n^{-1}(h_n^d)^{-1}\sum_{j=1}^{2q-1}(j-1)EK_1K_{j+1} \\ &\triangleq E_{n1} + E_{n2}. \end{aligned}$$

From (5.4a) and (5.4b), note that

$$\begin{aligned} h_n^{-d}EK_1^2 &= h_n^{-d}[c_1^2E(K_{n1}^{(q)}(y_1))^2 + c_2^2E(K_{n1}^{(q)}(y_2))^2 \\ &\quad +2c_1c_2EK_{n1}^{(q)}(y_1)K_{n1}^{(q)}(y_2)] \\ &= O(1) + O(1) + o(1) = O(1), \end{aligned} \tag{5.13a}$$

and from (5.4c) and (5.4d) that

$$\begin{aligned}(5.13\text{b})\quad h_n^{-d}E|K_1K_{j+1}| &\le h_n^{-d}[c_1^2E|K_{n1}^{(q)}(y_1)K_{n,j+1}^{(q)}(y_1)| + c_2^2E|K_{n1}^{(q)}(y_2)K_{n,j+1}^{(q)}(y_2)| \\ &\quad +|c_1c_2|(E|K_{n1}^{(q)}(y_1)K_{n,j+1}^{(q)}(y_2)| + E|K_{n1}^{(q)}(y_2)K_{n,j+1}^{(q)}(y_1)|)] \\ &= O(h^{\min(d,j)}) + O((h^{-(d+2)}v(q))^{1/2}),\end{aligned}$$

where the last equality holds uniformly for $j \ge d$. Also, since $K(\cdot)$ is bounded, it follows from (5.3) and the definition of K_i, that $|K_i| \le 2k^*(c_1+c_2)$, where $k^* = \sup_x K(x)$, and, hence, from Lemma 2, together with Remark 1(b), that

$$(5.13\text{c})\qquad |EK_1K_{j+1}| \le O(1)\alpha^*(j) = O(1)\alpha(j-q), \qquad \text{for} \quad j > q.$$

Now, by (5.13a), it is easily known that

$$E_{n1} = (n)^{-1}(2q)O(1) = O(q/n);$$

taking $N \sim h_n^{-d}$, by (5.13b) and (5.13c),

$$\begin{aligned}|E_{n2}| &\le 2n^{-1}(2q)(h_n^d)^{-1}\sum_{j=1}^{\infty}|EK_1K_{j+1}| \\ &= 2n^{-1}(2q)\sum_{j=1}^{N}|(h_n^d)^{-1}EK_1K_{j+1}| + 2n^{-1}(2q)(h_n^d)^{-1}\sum_{j=N+1}^{\infty}|EK_1K_{j+1}| \\ &= 2n^{-1}(2q)\left(\sum_{j=1}^{d}+\sum_{j=d+1}^{N}\right)((h^{-(d+2)}v(q))^{1/2} + h^{\min(d,j)}) \\ &\quad +2n^{-1}(2q)(h_n^d)^{-1}O(1)\sum_{j=N+1}^{\infty}\alpha(j-q) \\ &= O(q/n)[dh + (N-d)h^d + (h^{-(3d+2)}v(q))^{1/2} + N\sum_{j=N-q+1}^{\infty}\alpha(j)] \\ &= O(q/n),\end{aligned}$$

where the last equality follows from the conditions of this lemma and the fact that $N-q+1 \sim N$ due to the condition $qh^d \to 0$. Thus,

$$(5.14\text{a})\qquad EV_1^2 \le O(q/n) + O(q/n) = O(q/n).$$

Next, consider the second term of (5.12). By the definition of V_i and the stationarity,

$$\begin{aligned}\sum_{j=1}^{r}|EV_1V_{j+1}| &= (nh_n^d)^{-1}\sum_{j=1}^{r}\left|\sum_{i=p+d}^{p+d+2q-1}\;\sum_{\ell=j(p+2q)+p+d}^{j(p+2q)+p+d+2q-1}EK_iK_\ell\right| \\ &\le (nh_n^d)^{-1}\sum_{j=1}^{r}\;\sum_{i=p+d}^{p+d+2q-1}\;\sum_{\ell=j(p+2q)+p+d}^{j(p+2q)+p+d+2q-1}|EK_iK_\ell|\end{aligned}$$

$$= (nh_n^d)^{-1} \sum_{j=1}^{r} \sum_{i=p+d}^{p+d+2q-1} \sum_{\ell=j(p+2q)+p+d-i}^{j(p+2q)+p+d+2q-1-i} |EK_i K_{\ell+i}|$$

$$= (nh_n^d)^{-1} \sum_{j=1}^{r} \sum_{i=p+d}^{p+d+2q-1} \sum_{\ell=j(p+2q)+p+d-i}^{j(p+2q)+p+d+2q-1-i} |EK_1 K_{\ell+1}|.$$

Set $i' = i - (p+d)$ and $\ell' = \ell - j(p+2q)$. Then the right-hand side of the last equality becomes (for simplicity of notation, i' and ℓ' are still denoted by i and ℓ below)

$$(nh_n^d)^{-1} \sum_{j=1}^{r} \sum_{i=0}^{2q-1} \sum_{\ell=-i}^{2q-1-i} |EK_1 K_{j(p+2q)+\ell+1}|.$$

By this, together with (5.13b) and (5.13c), taking $N \sim h_n^{-d} p^{-1}$ and similarly treating E_{n2} as in the above,

$$\begin{aligned}
\sum_{j=1}^{r} |EV_1 V_{j+1}| &\le (nh_n^d)^{-1} \left(\sum_{j=1}^{N} + \sum_{j=N+1}^{\infty} \right) \sum_{i=0}^{2q-1} \sum_{\ell=-i}^{2q-1-i} |EK_1 K_{j(p+2q)+\ell+1}| \\
&= O(1)(n)^{-1} \sum_{j=1}^{N} \sum_{i=0}^{2q-1} \sum_{\ell=-i}^{2q-1-i} [h^{\min\{d, j(p+2q)+\ell\}} + O((h^{-(d+2)} v(q))^{1/2})] \\
&\quad + O(1)(nh_n^d)^{-1} \sum_{j=N+1}^{\infty} \sum_{i=0}^{2q-1} \sum_{\ell=-i}^{2q-1-i} \alpha(j(p+2q)+\ell-q) \\
&\triangleq E_{n3} + E_{n4},
\end{aligned}$$

where if $N < 1$, the sum $\sum_{j=1}^{N}$ is taken as 0 (i.e., without this sum), so it is assumed below that $N \ge 1$ in E_{n3}; for E_{n3}, it is clear that

$$\begin{aligned}
E_{n3} &= O(1) n^{-1} N q^2 [h^d + O(h^{-(d+2)} v(q))^{1/2}] \\
&= O(1)(np)^{-1} q^2 [1 + O(h^{-(3d+2)} v(q))^{1/2}] = O(q^2/(np)),
\end{aligned}$$

for E_{n4}, recalling from Remark 2 that $\alpha(\cdot)$ is decreasing, we have

$$\begin{aligned}
&\alpha(j(p+2q)+\ell-q) \le \alpha(j(p+2q)-3q), \quad \text{for} \\
&-i \le \ell \le 2q-1-i \text{ and } 0 \le i \le 2q-1, \\
&(p+2q)\alpha(j(p+2q)-3q) \le \sum_{\ell=(j-1)(p+2q)-3q}^{j(p+2q)-3q} \alpha(j),
\end{aligned}$$

and, hence,

$$\begin{aligned}
E_{n4} &\le O(1)(nh_n^d)^{-1} \sum_{j=N+1}^{\infty} \sum_{i=0}^{2q-1} \sum_{\ell=-i}^{2q-1-i} \alpha(j(p+2q)-3q) \\
&= O(1)(nh_n^d)^{-1} \sum_{j=N+1}^{\infty} (2q)^2 \alpha(j(p+2q)-3q)
\end{aligned}$$

$$
\begin{aligned}
&\le O(q^2/n)h_n^{-d} \sum_{j=N+1}^{\infty} (p+2q)^{-1} \sum_{\ell=(j-1)(p+2q)-3q}^{j(p+2q)-3q} \alpha(j) \\
&\le O(q^2/(np))h_n^{-d} \sum_{j=N(p+2q)-3q}^{\infty} \alpha(j) \\
&= O(q^2/(np))(Np) \sum_{j=N(p+2q)-3q}^{\infty} \alpha(j) \\
&= O(q^2/(np)),
\end{aligned}
$$

where $Np \sim h^{-d} \to \infty$. Thus, clearly,

(5.14b)
$$\sum_{j=1}^{r} |EV_1V_{j+1}| = O(q^2/(np)).$$

Finally, it follows from (5.12), (5.14a), and (5.14b) that

$$
\begin{aligned}
E\left(\sum_{m=1}^{r} V_m\right)^2 &= O(rq/n) + O(rq^2/(np)) \\
&= O(q/p) + O(q^2/p^2) \to 0,
\end{aligned}
$$

from which the first expression of (5.11a) is derived. The second of (5.11a) may be proved similarly.

To prove (5.11b), we first notice the following lemma:

LEMMA 7. *If* $n\alpha(q)/p \to 0$, *then*

(5.15)
$$\left| E\exp\left\{iu\sum_{m=1}^{r} U_m\right\} - \prod_{m=1}^{r} E\exp\{iuU_m\}\right| \to 0,$$

where $i = \sqrt{-1}$.

PROOF. The left-hand side of (5.15) is bounded by

$$
\begin{aligned}
&\left| E\exp\left\{iu\sum_{m=1}^{r} U_m\right\} - E\exp\{iuU_1\}E\exp\left\{iu\sum_{m=2}^{r} U_m\right\}\right| \\
&\quad + \sum_{\ell=2}^{r}\left| \left(\prod_{m=1}^{\ell-1} E\exp\{iuU_m\}\right) E\exp\left\{iu\sum_{m=\ell}^{r} U_m\right\} \right. \\
&\qquad \left. - \left(\prod_{m=1}^{\ell} E\exp\{iuU_m\}\right) E\exp\left\{iu\sum_{m=\ell+1}^{r} U_m\right\}\right| \\
&\le 4r\alpha^*(2q) \le 4n\alpha(q)/p,
\end{aligned}
$$

from which (5.15) follows clearly, where Lemma 2 is applied in the last second inequality.

Let $\{U_i^*\}_{i=1}^r$ be independent random variables with U_i^* identically distributed as U_i, $i = 1, \ldots, r$. By Lemma 7, to prove (5.11*b*), it suffices to check that

$$\sum_{m=1}^{r} U_m^* \xrightarrow{\mathcal{L}} N\left(0, (c_1^2 f(y_1) + c_2^2 f(y_2)) \int K^2(u)du\right). \tag{5.16}$$

According to Lindeberg's CLT, we need to check that

$$s_n^2 = \sum_{m=1}^{r} \mathrm{Var}(U_m^*) \to \tau^2 = (c_1^2 f(y_1) + c_2^2 f(y_2)) \int K^2(u)du; \tag{5.17a}$$

$$s_n^{-2} \sum_{m=1}^{r} E(U_m^*)^2 I_{(|U_m^*|/s_n \geq \epsilon)} \to 0 \tag{5.17b}$$

for any $\epsilon > 0$.

LEMMA 8. *Under the conditions of Lemma* 6, (5.17a) *holds.*

PROOF. By the stationarity,

$$\begin{aligned} s_n^2 &= r\,\mathrm{Var}(U_1^*) = rEU_1^2 \\ &= r/(nh^d)E\left(\sum_{i=d}^{d+p-1} K_i\right)^2 \\ &= r/(nh^d)\left[pEK_1^2 + 2\sum_{j=1}^{p}(j-1)EK_1K_{j+1}\right] \\ &= rp/(nh^d)EK_1^2 + 2O(rp/n)h^{-d}\sum_{j=1}^{p}|EK_1K_{j+1}| \\ &\triangleq s_{n1} + s_{n2}. \end{aligned} \tag{5.18}$$

First consider s_{n2}. Similarly as treating E_{n2} in the proof of Lemma 6, taking $N = \varepsilon h^{-d} = q\varepsilon/(qh^d) > 2q$ for n large enough due to $qh^d \to 0$, where ε is a positive constant, since $rp \sim n$, it follows from (5.13b) and (5.13c) that

$$\begin{aligned} s_{n2} &= O(1)\left(\sum_{j=1}^{d} + \sum_{j=d+1}^{N}\right)((h^{-(d+2)}v(q))^{1/2} + h^{\min(d,j)}) \\ &\quad + O(1)(h_n^d)^{-1}\sum_{j=N+1}^{\infty}\alpha(j-q) \\ &= O(1)\left(\sum_{j=1}^{d} h^j + (N^2h^{-(d+2)}v(q))^{1/2} + (N-d)h^d + h_n^{-d}\sum_{j=N+1}^{p}\alpha(j-q)\right) \\ &= O(1)\left(O(h) + \varepsilon(h^{-(3d+2)}v(q))^{1/2} + \varepsilon + \varepsilon^{-1}N\sum_{j=N+1}^{\infty}\alpha(j-q)\right). \end{aligned}$$

Thus, first letting $n \to \infty$ and then $\varepsilon \to 0$, gets $s_{n2} \to 0$ as $n \to \infty$.

Next consider s_{n1}. By (5.4a) and (5.4b) together with the first equality of (5.13a), it clearly follows that

$$s_{n1} \to (c_1^2 f(y_1) + c_2^2 f(y_2)) \int K^2(u)du \equiv \tau^2.$$

Thus, (5.17a) follows from (5.18).

To get (5.17b), we prove a stronger result.

LEMMA 9. *Under the conditions of Lemma* 6, *if* $nh^d \to \infty$, $p^3/n \to 0$, *then* $\sum_{m=1}^r E(U_m^*)^4 \to 0$, *and, hence*, (5.17$b$) *holds.*

PROOF. The proof of this lemma is similar to the proof of (7.9) of Robinson ((1983), p. 198), but since q here depends on n, it needs to be treated more carefully by (5.13a), and (5.13b). By stationarity,

$$\sum_{m=1}^r E(U_m^*)^4 = rE(U_1)^4 = r/(nh^d)^2 E\left(\sum_{i=1}^p K_i\right)^4. \tag{5.19}$$

It is well known (cf., the equality at the end of p. 198 of Robinson (1983)) that

$$\begin{aligned} E\left(\sum_i K_i\right)^4 &= E\sum_i \Big\{ K_i^4 + \sum_{u\neq i}[K_i^2 K_u(K_i + K_u) \\ &\qquad + \sum_{v\neq u\neq i}(K_i^2 K_u K_v + \sum_{w\neq v\neq u\neq i} K_i K_u K_v K_w)]\Big\} \\ &= E\sum_i K_i^4 + E\sum_i \sum_{u\neq i} K_i^2 K_u (K_i + K_u) \\ &\quad + E\sum_i \sum_{u\neq i} \sum_{v\neq u\neq i} K_i^2 K_u K_v \\ &\quad + E\sum_i \sum_{u\neq i} \sum_{v\neq u\neq i} \sum_{w\neq v\neq u\neq i} K_i K_u K_v K_w \\ &\triangleq A_{1n} + A_{2n} + A_{3n} + A_{4n}, \end{aligned} \tag{5.20}$$

where all i, u, v, w take their values, respectively, from 1 to p. Since A_{jn}, $j = 1, 2, 3, 4$, can be treated similarly, we mainly focus on the most complex term A_{4n} below.

Note that by the definition of K_i and the boundedness of the kernel function $K(\cdot)$ (bearing in mind that $O(1)$ may be different at different places below),

$$\begin{aligned} &E|K_i K_u K_v K_w| \le O(1) E|K_i K_u|, \qquad E|K_i K_u K_v K_w| \le O(1) E|K_v K_w|, \\ &E|K_i K_u K_v K_w| \le O(1) E|K_u K_v|. \end{aligned}$$

Then,

$$|A_{4n}| \le \sum_i \sum_{u\neq i} \sum_{v\neq u\neq i} \sum_{w\neq v\neq u\neq i} E|K_i K_u K_v K_w| \tag{5.21}$$

$$\leq O(1)\Bigg\{\sum_{i=1}^{p-3}\sum_{u=i+1}^{i+d}\sum_{v=u+1}^{u+d}\sum_{w=v+1}^{v+d} E|K_iK_u|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+1}^{i+d}\sum_{v=u+1}^{u+d}\sum_{w=v+d+1}^{p} E|K_vK_w|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+1}^{i+d}\sum_{v=u+d+1}^{p-1}\sum_{w=v+1}^{v+d} E|K_uK_v|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+1}^{i+d}\sum_{v=u+d+1}^{p-1}\sum_{w=v+d+1}^{p} E|K_uK_v|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+d+1}^{p-2}\sum_{v=u+1}^{u+d}\sum_{w=v+1}^{v+d} E|K_iK_u|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+d+1}^{p-2}\sum_{v=u+1}^{u+d}\sum_{w=v+d+1}^{p} E|K_vK_w|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+d+1}^{p-2}\sum_{v=u+d+1}^{p-1}\sum_{w=v+1}^{v+d} E|K_iK_u|$$

$$+\sum_{i=1}^{p-3}\sum_{u=i+d+1}^{p-2}\sum_{v=u+d+1}^{p-1}\sum_{w=v+d+1}^{p} E|K_iK_u|\Bigg\}$$

$$\triangleq O(1)\left\{\sum_{j=1}^{8} B_{jn}\right\}.$$

Now we treat $B_{jn}, j = 1, \ldots, 8$, respectively. For B_{1n}, setting $u' = u - i$ (u' is still denoted by u below) and by the stationarity and (5.13b),

$$\begin{aligned}
B_{1n} &= \sum_{i=1}^{p-3}\sum_{u=1}^{d}\sum_{v=i+u+1}^{i+u+d}\sum_{w=v+1}^{v+d} E|K_iK_{i+u}| \\
&= \sum_{i=1}^{p-3}\sum_{u=1}^{d}\sum_{v=i+u+1}^{i+u+d}\sum_{w=v+1}^{v+d} E|K_1K_{1+u}| \\
&= \sum_{i=1}^{p-3}\sum_{u=1}^{d}\sum_{v=i+u+1}^{i+u+d}\sum_{w=v+1}^{v+d} h^d[O(h^{\min(d,u)}) + O((h^{-(d+2)}v(q))^{1/2})] \\
&\leq d^3ph^d[O(h) + O((h^{-(d+2)}v(q))^{1/2})];
\end{aligned}$$

for B_{2n}, setting $w' = w - v$ (w' is still denoted by w below) and similarly to treating B_{1n},

$$B_{2n} = \sum_{i=1}^{p-3}\sum_{u=i+1}^{i+d}\sum_{v=u+1}^{u+d}\sum_{w=d+1}^{p-v} E|K_vK_{v+w}|$$

$$= \sum_{i=1}^{p-3} \sum_{u=i+1}^{i+d} \sum_{v=u+1}^{u+d} \sum_{w=d+1}^{p-v} E|K_1 K_{1+w}|$$
$$= \sum_{i=1}^{p-3} \sum_{u=i+1}^{i+d} \sum_{v=u+1}^{u+d} \sum_{w=d+1}^{p-v} h^d[O(h^{\min(d,w)}) + O((h^{-(d+2)}v(q))^{1/2})]$$
$$\leq d^2p^2h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})];$$

similarly to B_{2n},

$$B_{3n} = d^2p^2h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})],$$
$$B_{5n} = d^2p^2h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})];$$

for B_{4n}, setting $v' = v - u$ (v' is still denoted by v below) and similarly to treating B_{1n},

$$B_{4n} = \sum_{i=1}^{p-3} \sum_{u=i+1}^{i+d} \sum_{v=d+1}^{p-1-u} \sum_{w=u+v+d+1}^{p} E|K_u K_{u+v}|$$
$$= \sum_{i=1}^{p-3} \sum_{u=i+1}^{i+d} \sum_{v=d+1}^{p-1-u} \sum_{w=u+v+d+1}^{p} E|K_1 K_{1+v}|$$
$$= \sum_{i=1}^{p-3} \sum_{u=i+1}^{i+d} \sum_{v=d+1}^{p-1-u} \sum_{w=u+v+d+1}^{p} h^d[O(h^{\min(d,v)}) + O((h^{-(d+2)}v(q))^{1/2})]$$
$$\leq dp^3h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})];$$

similarly to B_{4n},

$$B_{6n} = dp^3h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})],$$
$$B_{7n} = dp^3h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})];$$

finally, for B_{8n}, similarly to treating B_{1n},

$$B_{8n} = \sum_{i=1}^{p-3} \sum_{u=d+1}^{p-2-i} \sum_{v=i+u+d+1}^{p-1} \sum_{w=v+d+1}^{p} E|K_i K_{i+u}|$$
$$= \sum_{i=1}^{p-3} \sum_{u=d+1}^{p-2-i} \sum_{v=i+u+d+1}^{p-1} \sum_{w=v+d+1}^{p} E|K_1 K_{1+u}|$$
$$= \sum_{i=1}^{p-3} \sum_{u=d+1}^{p-2-i} \sum_{v=i+u+d+1}^{p-1} \sum_{w=v+d+1}^{p} h^d[O(h^{\min(d,u)}) + O((h^{-(d+2)}v(q))^{1/2})]$$
$$\leq p^4h^d[O(h^d) + O((h^{-(d+2)}v(q))^{1/2})].$$

Thus, it follows from (5.21) that

$$\text{(5.22a)} \quad |A_{4n}| \leq O(ph^d)(h + (h^{-(d+2)}v(q))^{1/2}) + O(p^2h^d)(h^d + (h^{-(d+2)}v(q))^{1/2})$$
$$+O(p^2h^d)(h^d + (h^{-(d+2)}v(q))^{1/2}) + O(p^3h^d)(h^d + (h^{-(d+2)}v(q))^{1/2})$$
$$+O(p^2h^d)(h^d + (h^{-(d+2)}v(q))^{1/2}) + O(p^3h^d)(h^d + (h^{-(d+2)}v(q))^{1/2})$$

$$+O(p^3h^d)(h^d + (h^{-(d+2)}v(q))^{1/2}) + O(p^4h^d)(h^d + (h^{-(d+2)}v(q))^{1/2})$$
$$= O(ph^d)(h + (h^{-(d+2)}v(q))^{1/2}) + O(p^4h^d)(h^d + (h^{-(d+2)}v(q))^{1/2})$$
$$= O(ph^d) + O(p^4h^d)(h^d + (h^{-(d+2)}v(q))^{1/2}).$$

Similarly to A_{4n}, by (5.13a), and (5.13b),

(5.22b) $$|A_{1n}| = O(ph^d),$$

(5.22c) $$|A_{2n}| = O(ph^d) + O(p^2h^d)(h^d + (h^{-(d+2)}v(q))^{1/2}),$$

(5.22d) $$|A_{3n}| = O(ph^d) + O(p^3h^d)(h^d + (h^{-(d+2)}v(q))^{1/2}).$$

Now it follows from (5.19), (5.20) and (5.22) that

$$\sum_{m=1}^{r} E(U_m^*)^4 = O(r/(nh^d)^2)[ph^d + p^2((h^{-(d+2)}v(q))^{1/2} + h^d)h^d$$
$$+p^3((h^{-(d+2)}v(q))^{1/2} + h^d)h^d + p^4((h^{-(d+2)}v(q))^{1/2} + h^d)h^d]$$
$$= O(1)[(nh^d)^{-1} + (p^3/n)((h^{-(3d+2)}v(q))^{1/2} + 1)] \to 0,$$

which is the desired result.

5.3 *Proofs of Theorems* 1 *and* 2

PROOF OF THEOREM 1. First, by Assumptions 5(i) and 6(v), $nv(q)/h^{2+d} = O(nh^{2d}) \to 0$ as $n \to \infty$, thus the condition of Lemma 5 is satisfied. The conditions of the other lemmas in Subsection 5.2 are obviously verified by Assumptions 1, 5 and 6. Hence, (5.8) is deduced from Lemmas 6–9. Finally, Theorem 1 follows from (5.7) and (5.8).

PROOF OF THEOREM 2. By Assumption 4(i), (2.7) and then Assumption 5(ii), it easily follows that

$$(nh_n^d)^{1/2}|Ef_n(y) - f(y)| = O((nh_n^{d+2})^{1/2}) \to 0$$

as $n \to \infty$. Thus, together with Theorem 1, Theorem 2 holds.

Acknowledgements

Financial support from a Postdoctoral Research Fellowship of Université Catholiqué de Louvain, the Special Research Fund for Mathematics in the Chinese Academy of Sciences and the Natural Science Foundation of China are gratefully acknowledged. Here I would especially like to express my deep gratitude to both referees and the editor, Prof. Shinto Eguchi, for their invaluable comments, suggestions and help which greatly improved the presentation of the paper. Thanks are also due to Prof. Marc Hallin, Prof. Lah Tat Tran and Dr. Steven X. Wei for their comments, encouragement, and kind help.

REFERENCES

Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes, *J. Appl. Probab.*, **21**, 930–934.
Bierens, H. J. (1983). Uniform consistency of kernel estimators of a regression function under generalized conditions, *J. Amer. Statist. Assoc.*, **78**, 699–707.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Chanda, K. C. (1983). Density estimation for linear processes, *Ann. Inst. Statist. Math.*, **35**, 439–446.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*, Wiley, New York.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Goródetskii, V. V. (1977). On the strong mixing properties for linear sequences, *Theory Probab. Appl.*, **22**, 411–413.

Granger, C. W. J. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statist., No. 60, Springer, New York.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*, Academic Press, New York.

Hallin, M. and Tran, L. T. (1996). Kernel density estimation for linear processes: asymptotic normality and optimal bandwidth derivation. *Ann. Inst. Statist. Math.*, **48**, 429–449.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes, *Theory Probab. Appl.*, **7**, 349–382.

Irle, A. (1997). On consistency in nonparametric estimation under mixing conditions, *J. Multivariate Anal.*, **60**, 123–147.

Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity, *J. Amer. Statist. Assoc.*, **92**, 1184–1194.

Lu, Z. D. (1996*a*). A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model, *Statist. Probab. Lett.*, **30**, 305–311.

Lu, Z. D. (1996*b*). Geometric ergodicity of a general ARCH type model with applications to some typical models, *Chinese Sci. Bull.*, **41**, 1630.

Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, *IEEE Trons. Inform. Theory*, **32**, 254–267.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models*, Lecture Notes in Statist, No. 11. Springer, New York.

Pham, D. T. (1986). The mixing property of bilinear and generalized random coefficient models, *Stochastic Process. Appl.*, **23**, 291–300.

Robinson, P. M. (1983). Nonparametric estimators for time series, *J. Time Ser. Anal.*, **4**, 185–197.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables, *J. Statist. Plann. Inference*, **18**, 135–149.

Tjøstheim, D. (1986). Some doubly stochastic time series models. *J. Time Ser. Anal.*, **7**, 51–72.

Tjøstheim, D. (1996). Measures of dependence and tests of independence, *Statistics*, **28**, 249–284.

Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

Tran, L. T. (1989). Recursive density estimation under dependence, *IEEE Trans. Inform. Theory*, **35**, 1103–1108.

Tran, L. T. (1992). Kernel density estimation for linear processes, *Stochastic Process. Appl.*, **41**, 281–296.

Tran, L. T., Roussas, G., Yakowitz, S. and Truong Van, B. (1996). Fixed-design regression for linear time series, *Ann. Statist.*, **24**, 975-991

Weiss, A. A. (1984). ARMA model with ARCH errors, *J. Time Ser. Anal.*, **5**, 129–143.