Statistica Sinica **14**(2004), 1105-1126

ON RANDOM-DESIGN MODEL WITH DEPENDENT ERRORS

Jan Mielniczuk and Wei Biao Wu

Polish Academy of Sciences and University of Chicago

Abstract: We consider random-design nonparametric regression model in which errors depend on predictors as well as on unobservable latent variables. Predictors and latent variables may be short- or long-range dependent. In this setup asymptotic distributions of the Nadaraya-Watson estimate of regression function are studied under various conditions. We prove that their form depends on three factors: amount of smoothing and strength of dependence of both predictors and latent variables. Our results go beyond earlier ones by allowing more general dependence structure.

Key words and phrases: Kernel regression estimators, linear process, long- and short-range dependence, martingale Central Limit Theorem, random-design regression.

1. Introduction

Let $(Y_t, X_t)_{t=1}^{\infty}$ be a bivariate stationary process and suppose that $\mathbb{E}|Y| < \infty$, where $(Y, X) = (Y_1, X_1)$. We consider the problem of nonparametric estimation of the regression function of Y given $X = x$; $g(x) := \mathbb{E}(Y|X = x)$. The problem has been extensively studied when observations (Y_t, X_t) are independent or weakly dependent; see for example Györfi, Härdle, Sarda and Vieu (1989). For recent developments see Nze, Bühlmann and Doukhan (2002). There has also been recent interest in studying properties of nonparametric estimators when observations are long-range dependent (LRD), compare e.g., Hidalgo (1997) and Csörgő and Mielniczuk (1999, 2000). This supplements the much more frequent studies of LRD case when parametric assumptions are made about regression function g (see e.g., Koul (1992) and Robinson and Hidalgo (1997)).

In order to investigate distributional properties of estimators in nonparametric random-design regression model for LRD data, we have to impose some conditions on the structure of errors. Their specific form assumed here is described in the model equation

$$Y_t = g(X_t) + G(Z_t, X_t), \quad t = 1, 2, \ldots \tag{1}$$

Errors $G(Z_t, X_t)$ depend on the explanatory variables X_t, as well as on the latent variables Z_t forming a stationary sequence, and such that $\mathbb{E}(G(Z_t, X_t)|X_t) = 0$ almost surely. This noise structure was first considered by Cheng and Robinson (1994) and yields a substantial relaxation of the assumption of independence between predictors and errors. In the case when the sequence (X_t) is i.i.d. or weakly dependent and independent of the sequence (Z_t), which is assumed to be either Gaussian or a linear process, the properties of kernel estimators of g were studied in Csörgő and Mielniczuk (1999, 2000).

Here we provide substantial generalization of the previous research by dispensing with assumption of weak dependence of predictors *and* their independence from the sequence of latent variables (Z_t). Allowing for more general structure of dependence between and within predictors and errors is desirable e.g., in econometric applications (c.f., Cheng and Robinson (1994)). In order to accommodate this we assume that both (Z_t) and (X_t) are (possibly dependent) linear processes

$$Z_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \qquad X_t = \sum_{i=0}^{\infty} c_i \eta_{t-i},$$

where $(\varepsilon_i, \eta_i)_{-\infty}^{\infty}$ is an i.i.d. sequence having mean 0, $\mathbb{E}(\varepsilon_i^2 + \eta_i^2) < \infty$ and the coefficients $(a_i)_0^{\infty}$ and $(c_i)_0^{\infty}$ are square summable. Moreover, we assume $a_0 = c_0 = 1$. In the paper we focus on the case of univariate predictors (X_t); extensions to the multivariate case will be pursued elsewhere. The strength of dependence of a linear process is determined by the decay rate of the coefficients. If $a_i = L_Z(i) i^{-\beta_Z}$, where $1/2 < \beta_Z < 1$ and $L_Z(\cdot)$ is slowly varying at ∞, routine calculation based on Karamata's theorem implies that $r_Z(i) := \mathrm{Cov}(Z_0, Z_i) \sim C(\beta_Z) L_Z^2(i) i^{-(2\beta_Z - 1)} \mathbb{E}(\varepsilon_i^2)$, where $C(\beta_Z) = \int_0^{\infty} (x + x^2)^{-\beta_Z}\, dx$ and $a_n \sim b_n$ means $\lim_{n\to\infty} a_n/b_n \to 1$. Thus in this case the sum of absolute values of the covariances diverges. This property is called long-range dependence, or long-memory, in contrast to short-range dependence (SRD) that features absolutely summable covariances. Note that if $\sum_{i=0}^{\infty} |a_i| < \infty$, or $\beta > 1$ in the hyperbolic decay condition given above, (Z_t) is SRD.

In the paper we investigate limit laws of kernel estimators of $(g(x_1), \ldots, g(x_l))$ for different points $x_1, \ldots, x_l \in \mathbb{R}$ when the sample $(Y_1, X_1), \ldots, (Y_n, X_n)$ pertaining to the model (1) is available. The main results follow from an asymptotic representation for $\hat{g}_n(x)$, $x \in \mathbb{R}$, given in Proposition 1 of Section 3. It turns out that the correct standardization and asymptotic distribution of $\hat{g}_n(x)$ is determined by three factors: the amount of smoothing and the strength of dependence of the sequences (Z_t) and (X_t). This extends a smoothing dichotomy phenomenon studied previously in the case of dependent predictors. In particular, results of Csörgő and Mielniczuk (2000) are generalized under different set of assumptions.

The results for the random-design model are qualitatively very similar to those for probability density estimates based on LRD data. For the latter case the research goes back to Robinson (1991) (paper submitted in 1988) who showed that asymptotic distributions of a kernel density estimate at different points may be perfectly correlated; this was extended in Cheng and Robinson (1991). Different behaviour of the integrated mean squared error for small and large bandwidths was discussed by Hall and Hart (1990). The smoothing dichotomy of asymptotic distributions of a kernel density estimate was proved by Ho (1996).

The main results of the paper are consequences of an asymptotic representation of $\hat{g}_n(x)$ for various patterns of dependence. In particular it follows that if the amount of smoothing is small in a given sense, the estimators behave asymptotically as if (Z_t) and (X_t) are independent. For a large amount of smoothing, crudely speaking, the standardization and the limiting law is usually determined by strength of dependence of the more strongly dependent sequence among (Z_t) and (X_t). However, the outcome depends also on a pair of integers (l_1, l_2), defined in (18), which yield generalization of a power rank of a function with respect to a given distribution. In the considered context the previous statement holds when $l_1 = l_2$. If l_2 is larger than l_1 it may happen, as it happens for mutually independent sequences (Z_t) and (X_t), that the limiting behavior is determined by the strength of dependence of (Z_t) which is actually *weaker* than that of (X_t). In this context we refer to Choy and Taniguchi (2002) who discuss similar phenomenon in the case of a linear model without intercept. The development depends on the decomposition of a centered Nadaraya-Watson estimate into three terms (cf. equation (4)): a martingale term M_n, a sum of conditional expectations N_n and a term P_n pertaining to bias of the estimator. When both (Z_t) and (X_t) are SRD, only M_n determines the asymptotic law. In the LRD case, all three terms may influence the asymptotic distribution. Analysis of N_n relies on a projection method developed in Wu (2003) and Wu and Woodroofe (2002) to prove limit theorems for linear LRD processes, whereas analysis of P_n is partly based on Wu and Mielniczuk (2002). In Section 2 we state and discuss our assumptions. Section 3 contains the main results and some of their consequences for particular submodels of the regression model (1). Section 4 contains auxiliary lemmas and all proofs.

2. Definitions and Assumptions

The following notation will be used throughout the paper. Let $W_t := (Z_t, X_t)$ and $W_{t,k} = \mathbb{E}(W_t|\widetilde{W}_k)$, where $\widetilde{W}_t = (\ldots, \varepsilon_{t-1}, \eta_{t-1}, \varepsilon_t, \eta_t)$ is a shift process; let f_t be the density of $W_t - W_{t,0} = (\sum_{i=0}^{t-1} a_i \varepsilon_{t-i}, \sum_{i=0}^{t-1} a_i \eta_{t-i})^{\mathbf{T}}$ with $a^{\mathbf{T}}$ denoting transposition of a vector a. Moreover, $f_{t,Z}$ and $f_{t,X}$ denote the marginal densities of f_t. In particular, f_1 stands for the density of (ε_0, η_0) and f_∞ the density of

$(Z, X) = (Z_1, X_1)$ with its marginal densities h and f. Marginal densities of f_1 will be denoted by f_ε and f_η, respectively instead of $f_{1,Z}$ and $f_{1,X}$. Let $\|\xi\| = (\mathbb{E}|\xi|^2)^{1/2}$ be the $\mathcal{L}^2$-norm of a random vector ξ and $\mathcal{P}_k\xi = \mathbb{E}(\xi|\widetilde{W}_k) - \mathbb{E}(\xi|\widetilde{W}_{k-1})$, $k \in \mathbb{N}$, be the projection differences.

We estimate g by means of the Nadaraya-Watson estimate

$$\hat{g}_n(x) = \frac{\sum_{t=1}^n K\left(\frac{x-X_t}{b_n}\right) Y_t}{\sum_{t=1}^n K\left(\frac{x-X_t}{b_n}\right)}, \tag{2}$$

where K is a bounded symmetric probability density supported on $[-1, 1]$ and $b_n > 0$ is a sequence of deterministic bandwidths tending to 0 in such a way that $nb_n \to \infty$. Bandwidth $b = b_n$ determines the amount of smoothing employed by the kernel estimator for sample size n. Setting $K_b(x) := b^{-1}K(x/b)$, put $\hat{f}_n(x) := n^{-1}\sum_{t=1}^n K_{b_n}(x - X_t)$ to be a kernel estimate of the marginal density f of X. We further define $\hat{v}_n(x) := \hat{g}_n(x)\hat{f}_n(x)$, $J_t(x) = G(Z_t, X_t)K_b(x - X_t)$ and

$$g_n(x) = \frac{\mathbb{E}\hat{v}_n(x)}{\mathbb{E}\hat{f}_n(x)} = \frac{\mathbb{E}(K_{b_n}(x - X)Y)}{\mathbb{E}(K_{b_n}(x - X))}. \tag{3}$$

Let $\hat{g}_n(x) - g(x) = D_n(x) + g_n(x) - g(x)$, where $D_n(x) = \hat{g}_n(x) - g_n(x)$. As $g_n(x) - g(x)$ is a non-stochastic term which under standard conditions is of order b_n^2 (cf. (16)) we study asymptotic behaviour of $D_n(x)$. Assumptions $\mathcal{C}_1$, $\mathcal{C}_6 - \mathcal{C}_7$ below imply that $\hat{f}_n(x)$ is a weakly consistent estimate of $f(x)$. Thus when $f(x) \neq 0$, in order to investigate asymptotic laws of D_n given (1), it is enough to study laws of $D_n(x)\hat{f}_n(x)$, which is equal to

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n (Y_t - g_n(x))K_b(x - X_t) \\
&= \frac{1}{n}\sum_{t=1}^n [J_t(x) - \mathbb{E}(J_t(x)|\widetilde{W}_{t-1})] + \frac{1}{n}\sum_{t=1}^n \mathbb{E}(J_t(x)|\widetilde{W}_{t-1}) \\
&\quad + \frac{1}{n}\sum_{t=1}^n (g(X_t) - g_n(x))K_b(x - X_t) := M_n(x) + N_n(x) + P_n(x).
\end{aligned} \tag{4}$$

Note that $M_n(x)$ admits a martingale structure. We prove in Lemma 3 of Section 4 that $(nb_n)^{1/2}M_n(x)$ is asymptotically normal regardless of how strongly the dependent the sequences (Z_t) and (X_t) are. The behavior of the terms $N_n(x)$ and $P_n(x)$ is studied in Lemmas 4 and 6 respectively.

Let $\mathcal{C}^{(k)}(U)$ denote the family of k times differentiable functions on an open set U; let $|v| = (\sum_{i=1}^l v_i^2)^{1/2}$ be the norm of a vector $v = (v_1, \ldots, v_l)^{\mathbf{T}} \in \mathbb{R}^l$ and $|\mathcal{A}| = (\sum_{j,k=1}^{l,q} a_{j,k}^2)^{1/2}$ the norm of $(l \times q)$-matrix $\mathcal{A}$. We put $\mathcal{A}_i = \mathrm{diag}(a_i, c_i)$. For an l-dimensional random variable V, $\|V\|$ denotes $\mathcal{L}^2$-norm of $|V|$.

We now state and discuss assumptions under which our results hold.

$\mathcal{C}_1$: f_1 is bounded, twice continuously differentiable with bounded derivatives;

$\mathcal{C}_2$: $\mathbb{E}\{|B_1(y) - B_1(x)|\} \to 0$ as $y \to x$, where

$$B_t(y) := \int G^2(z,y) f_1(z - Z_{t,t-1}, y - X_{t,t-1})\, dz\,;$$

$\mathcal{C}_3$: $\mathbb{E}\bar{G}^2(Z) < \infty$, where $\bar{G}(z) = \sup_{y:|y-x|<\delta_0} |G(z,y)|$ for some $\delta_0 > 0$;

$\mathcal{C}_4$: Let $R_{2,t}(z,y) = f_{t-1}(z - Z_{t,1}, y - X_{t,1}) - f_{t-1}(z - Z_{t,0}, y - X_{t,0}) + \nabla f^{\mathbf{T}}_{t-1}(z - Z_{t,0}, y - X_{t,0})\mathcal{A}_{t-1}\binom{\varepsilon_1}{\eta_1}$. There exist $C > 0$ and $\delta_0 > 0$ such that for sufficiently large $t \in \mathbb{N}$,

$$\sup_{y:|y-x|<\delta_0} \| \int G(z,y)[\nabla f_{t-1}(z - \xi_1, y - \xi_2) - \nabla f_{t-1}(z,y)]\, dz \| \le C \|(\xi_1, \xi_2)\| \quad (5)$$

holds for $(\xi_1, \xi_2) = (Z_{t,0}, X_{t,0})$ and $(\xi_1, \xi_2) = (Z_{t,1}, X_{t,1})$, and

$$\sup_{y:|y-x|<\delta_0} \| \int G(z,y) R_{2,t}(z,y)\, dz \| \le C|\mathcal{A}_{t-1}|^2; \quad (6)$$

$\mathcal{C}_5$: $g \in \mathcal{C}^{(2)}(U(x,\delta_0))$, where $U(x,\delta) = \{y : |y-x| < \delta\}$;

$\mathcal{C}_6$: There exist a sequence $\gamma_t \downarrow 0$ such that

$$\sum_{\iota=0}^{1} \sup_y \|f^{(\iota)}_{t-1,X}(y - X_{t,1}) - f^{(\iota)}_{t-1,X}(y - X_{t,0}) + f^{(\iota+1)}_{t-1,X}(y - X_{t,0}) c_{t-1}\eta_1\| \le \gamma_{t-1}|c_{t-1}|, \quad (7)$$

$$\sum_{\iota=0}^{1} \sup_y [\|f^{(\iota+1)}_{t-1,X}(y - X_{t,0}) - f^{(\iota+1)}_{t-1,X}(y)\| + \|f^{(\iota+1)}_{t-1,X}(y - X_{t,1}) - f^{(\iota+1)}_{t-1,X}(y)\|] \le \gamma_{t-1}; \quad (8)$$

$\mathcal{C}_7$: Innovation coefficients $(c_i)_{i=0}^{\infty}$ satisfy one of the following:

(i) $\sum_{i=0}^{\infty} |c_i| < \infty$;
(ii) $\sum_{i=0}^{\infty} |c_i| = \infty$ and there exists $i_0 \in \mathbb{N}$ and $\tau > 0$ such that for all $j \ge i \ge i_0$ $|c_i + \cdots + c_j| \ge \tau(|c_i| + \cdots + |c_j|)$.

Conditions $\mathcal{C}_2, \mathcal{C}_3$ and the boundedness of f_1 are used to derive asymptotic normality of $M_n(x)$, conditions $\mathcal{C}_1$ and $\mathcal{C}_4$ (respectively, $\mathcal{C}_5 - \mathcal{C}_7$) ensure that suitable approximation of $N_n(x)$ (respectively, $P_n(x)$) holds. Condition $\mathcal{C}_2$ is the basic condition used to prove convergence of the conditional variance part in the martingale CLT for $M_n(x)$. Here we provide some sufficient conditions for it. Observe that, using the triangle inequality, we have

$$\begin{aligned}
&|B_t(y) - B_t(x)| \\
&\leq \int |G^2(v,y) - G^2(v,x)| f_1(v - Z_{t,t-1}, x - X_{t,t-1}) dv \\
&\quad + \int G^2(v,y) |f_1(v - Z_{t,t-1}, y - X_{t,t-1}) - f_1(v - Z_{t,t-1}, x - X_{t,t-1})| dv. \quad (9)
\end{aligned}$$

As we have $\mathbb{E} f_1(z - Z_{t,t-1}, y - X_{t,t-1}) = f_\infty(z,y)$ it is easily seen from (9), using the change of integration order, that the condition $\mathcal{C}_2$ is implied by the conjunction of following two conditions:

$\mathcal{C}_2'$: $\mathbb{E}(|G^2(Z,y) - G^2(Z,x)|\,|X = x) \to 0$ when $y \to x$;

$\mathcal{C}_2''$: $\sup_{y:|y-x|<\delta_0} \int G^2(v,y) \mathbb{E}\tilde{g}(v - Z_{t,t-1})\, dv < \infty$, where $|f_1(v,y) - f_1(v,x)| \leq \tilde{g}(v)\tilde{h}(y - x)$ and $\tilde{h}(z) \to 0$ when $z \to 0$.

In particular when ε_1 and η_1 are independent, and since f_η is Lipschitz continuous in view of $\mathcal{C}_1$, one can take $\tilde{g}(v) = f_\varepsilon(v)$ and $\tilde{h}(y) = Cy$ for some constant C. Condition $\mathcal{C}_2''$ is then implied by $\sup_{y:|y-x|<\delta_0} \mathbb{E}G^2(Z,y) < \infty$ as $\mathbb{E}f_\varepsilon(v - Z_{t,t-1}) = h(v)$. In turn, the last condition is implied by condition $\mathcal{C}_3$. Thus, when predictors and latent variables are independent, $\mathcal{C}_2$ may be replaced by $\mathcal{C}_2'$ which then coincides with $\mathcal{L}^2$-continuity of $G(Z,\cdot)$ at x. In the general case when independence of ε_1 and η_1 is not assumed, condition $\mathcal{C}_2''$ in view of (9) may be always replaced by $\sup_{y:|y-x|<\delta_0} \int G^2(v,y)\, dv < \infty$ because of Lipschitz continuity of f_1. Condition $\mathcal{C}_3$ is used to check Lindeberg's condition in the martingale CLT. It is slightly stronger than finiteness of $\mathbb{E}G^2(Z,y)$ in the neighborhood of x.

We discuss now condition $\mathcal{C}_4$. Notice that the difference $(z - Z_{t,1}, y - X_{t,1}) - (z - Z_{t,0}, y - X_{t,0}) = (-a_{t-1}\varepsilon_1, -c_{t-1}\eta_1)$ and its $\mathcal{L}^2$-norm is equal to $|\mathcal{A}_{t-1}|$. Thus (5) and (6) basically assert the validity of the first and second order Taylor's expansion of the functionals $\int G(z,y) \nabla f_{t-1}(z - \cdot, y - \cdot) dz$ and $\int G(z,y) f_{t-1}(z - \cdot, y - \cdot) dz$, respectively. It is easily seen that $\mathcal{C}_4$ is implied by

$\mathcal{C}_4'$: $\sup_{y:|y-x|<\delta_0} \sup_{t\in\mathbb{N}} \int |G(z,y)| f_{t,i}(z,y)\, dz < \infty$ for $i = 2,3$ and some $\delta_0 > 0$, where

$$f_{t,2}(\mathbf{w}) = \sup_{\mathbf{s}\in\mathbb{R}^2} \frac{|\nabla f_t(\mathbf{w} - \mathbf{s}) - \nabla f_t(\mathbf{w})|}{|\mathbf{s}|} \text{ and } f_{t,3}(\mathbf{w}) = \sup_{\mathbf{s}\in\mathbb{R}^2} \frac{|\mathbf{D}^2 f_t(\mathbf{w} - \mathbf{s}) - \mathbf{D}^2 f_t(\mathbf{w})|}{|\mathbf{s}|}.$$

Moreover, when ∇f_1 and $\mathbf{D}^2 f_1$ are Lipschitz continuous, then

$$f_{t,i}(\mathbf{w}) \leq \int f_{1,i}(\mathbf{w} - \mathbf{t}) \tilde{f}(\mathbf{t})\, d\mathbf{t}, \qquad i = 2,3,$$

where $\tilde{f}$ is the density of $(\sum_{i=1}^{t-1} a_i \varepsilon_{t-i}, \sum_{i=1}^{t-1} c_i \eta_{t-i})^{\mathbf{T}}$. Thus $\mathcal{C}_4$ follows from the following condition on $f_{1,i}$:

$\mathcal{C}_4'' : \sup_{w_2 \in U(x,\delta_0)} \int |G(\mathbf{w})| f_{1,i}(\mathbf{w}-\mathbf{t})\, dw_1 < C < \infty, \quad i = 2, 3$

uniformly in $\mathbf{t} \in \mathbb{R}^2$, where $\mathbf{w} = (w_1, w_2)$.

We discuss now condition $\mathcal{C}_6$. Observe that it holds when f_1 is three times continuously differentiable with bounded derivatives, γ_{t-1} being a constant multiply of $|c_{t-1}|$ in (7) and of $(\sum_{s=t-1}^{\infty} c_s^2)^{1/2}$ in (8). This follows from Taylor's expansion and an observation analogous to Lemma 1. Conditions $\mathcal{C}_5 - \mathcal{C}_7$ entail asymptotic expansions for certain partial sum processes dealt with in proof of Lemma 6. In particular it follows from the proof that (43), which provides asymptotic representation of centered kernel density estimate for LRD observations, actually holds under *weaker* assumptions than those used in Wu and Mielniczuk (2002), namely it is sufficient to assume that f_η is twice continuously differentiable with bounded derivatives. This is implied by Lemma 7 which is used in place of Ho and Hsing (1996) results (see also Wu (2003) for improvements and generalizations of the last paper).

Condition $\mathcal{C}_7$ is used only to derive suitable approximation of $N_n(x)$ (cf. Lemma 6 in Section 4). Its part (i) corresponds roughly to the assumption that the X_i are SRD, part (ii) is a mild condition assumed for the LRD case: it implies that $\mathrm{Var}(X_1 + \cdots + X_n)/n \to \infty$. It holds when (c_i) satisfies (10) below, with $L_X(\cdot)$ a slowly varying function ultimately of constant sign.

Put $\sigma_{n,Z}^2 = \mathbb{E}(\sum_{t=1}^n Z_t)^2$ and $\sigma_{n,X}^2 = \mathbb{E}(\sum_{t=1}^n X_t)^2$. When

$$a_i = L_Z(i) i^{-\beta_Z}, \;\; c_i = L_X(i) i^{-\beta_X} \text{ with } 1 > \beta_{\mathrm{Z}}, \beta_{\mathrm{X}} > 1/2, \tag{10}$$

where $L_Z(\cdot), L_X(\cdot)$ are slowly varying at infinity, application of Karamata's theorem implies that

$$\sigma_{n,Z}^2 \sim D(\beta_Z) n^{2-(2\beta_Z-1)} L_Z^2(n) \mathbb{E}(\varepsilon_1^2), \quad \sigma_{n,X}^2 \sim D(\beta_X) n^{2-(2\beta_X-1)} L_X^2(n) \mathbb{E}(\eta_1^2), \tag{11}$$

where $D(\beta) = \{(2-2\beta)(3/2-\beta)\}^{-1} C(\beta)$. Then noting that $\sigma_{n,Z}^2 \to \infty$ when $n \to \infty$ it follows from Theorem 18.6.5 in Ibragimov and Linnik (1971) (see also Lemma 8) that

$$\sigma_{n,Z}^{-1} \sum_{t=1}^{n} Z_t \Longrightarrow \mathcal{N}_1, \tag{12}$$

where $\Rightarrow$ denotes convergence in distribution and $\mathcal{N}_1$ is the standard normal random variable. The analogous result holds for $\sum_{t=1}^n X_t$.

3. Results

We first state a crucial approximation of $\hat{g}_n(x) - g_n(x)$ from which asymptotic distributions of $\hat{g}_n(x) - g(x)$ are derived. To this end define

$$S_t(x) = -\int G(z,x) \nabla^T f_\infty(z,x)\, dz \times W_{t,t-1}, \quad \tilde{N}_n(x) = n^{-1} \sum_{t=1}^{n} K_b \star S_t(x),$$

where $\star$ denotes the convolution, and

$$\Xi_n^2 = n\Theta_n^2 + \sum_{i=1}^{\infty}(\Theta_{n+i} - \Theta_i)^2,\ \Theta_n = \sum_{i=1}^{n}\theta_i,\ \theta_i = |\mathcal{A}_{i-1}|\sqrt{A_{i-1}} \text{ and } A_i = \sum_{j=i}^{\infty}|\mathcal{A}_j|^2.$$

Moreover, let

$$C_1(x) = \mu_2[-g'(x)f''(x) + g'(x)f'^2(x)/f(x)], \tag{13}$$

where $\mu_i = \int x^i K(x)\,dx$ for $i \in \mathbb{N}$.

Proposition 1. *Assume that conditions* $\mathcal{C}_1 - \mathcal{C}_7$ *hold and* $f(x) \neq 0$. *Then*

$$\hat{g}_n(x) - g_n(x) = (M_n(x) + \tilde{N}_n(x) + P_n(x))/\hat{f}_n(x) + \mathcal{O}_P(\Xi_n/n), \tag{14}$$

where finite-dimensional distributions of $\sqrt{nb_n}M_n$ *are asymptotically normal,* $P_n(x) = o_P((nb_n)^{-1/2})$ *under* $\mathcal{C}_7$(i) *and*

$$P_n(x) = C_1(x)\frac{b_n^2}{n}\sum_{t=1}^{n}X_{t,t-1} + o_P((nb_n)^{-1/2} + b_n^2\sigma_{n,X}/n) \tag{15}$$

provided $\mathcal{C}_7$(ii) *holds. If* $|\mathcal{A}_n| = \mathcal{O}(L(n)n^{-\beta})$ *for some* $\beta > 1/2$ *and slowly varying function* L, *then* $\Xi_n = \mathcal{O}[\sqrt{n}\sum_{i=1}^{2n} i^{1/2-2\beta}L^2(i) + n^{2-2\beta}L^2(n)]$.

We are now in position to state our main results. Consider distinct points $x_1, \ldots, x_l \in \mathbb{R}$ such that f does not vanish at any of them. We assume that conditions $\mathcal{C}_1 - \mathcal{C}_7$ are satisfied for all $x_i, i = 1, \ldots, l$. Observe that since, in view of $\mathcal{C}_1$ and $\mathcal{C}_5$, $gf(\cdot)$ is two times continuously differentiable in a neighborhood of x, assumptions on kernel K yield

$$g_n(x) - g(x) = C_B(x)b_n^2 + o(b_n^2), \tag{16}$$

where $C_B(x) = \mu_2(2f(x))^{-1}((fg)''(x) - gf''(x))$. Theorems below are direct consequences of Proposition 1, Lemmas 3 and 8 in Section 4, and (16).

3.1. Short-range dependent sequences

We first consider the case when both (Z_t) and (X_t) are short-range dependent. In the following result assume that $nb_n^5 \to C^2$, where C is a nonnegative constant. Thus the result covers the case of bandwidths of order $n^{-1/5}$ which is MSE-optimal order under independence. Although Theorem 1 appears to be new, the asymptotic law of $\hat{g}_n(x)$ is precisely the same as under independence or other weak dependence conditions such as strong mixing (cf. Robinson (1983) for the result under an α-mixing condition). This phenomenon, known as the whitening by windowing principle, is widely known to occur for weakly dependent data. The main results of the paper are thus Theorems 2 and 3 which show when this principle fails under LRD, and how in this case the asymptotic law is affected by strength of dependence.

Theorem 1. *If* $\sum_{i=0}^{\infty} |\mathcal{A}_i| < \infty$, $nb_n^5 \to C^2 \geq 0$ *and* (16) *holds, then*

$$(nb_n)^{1/2} \left(\hat{g}_n(x_i) - g(x_i), 1 \leq i \leq l\right) \Longrightarrow \Big(\frac{\sigma(x_i)}{f(x_i)}\mathcal{N}_i + \mu(x_i), 1 \leq i \leq l\Big), \tag{17}$$

where $\sigma^2(x) = \kappa_2 \int G^2(v,x) f_\infty(v,x) dv$ *with* $\kappa^2 = \int K^2(s) ds$, $\mathcal{N}_i$ *are independent standard normal variables and* $\mu(x) = CC_B(x)$.

The theorem follows from Proposition 1 upon noting that for absolutely summable $\mathcal{A}_i$ we have $\Xi_n^2 = \mathcal{O}(n)$ and $\tilde{N}_n(x)$ and $P_n(x)$ are both $\mathcal{O}_P(n^{-1/2})$. Observe that in view of (16), $(nb_n)^{1/2}(g_n(x) - g(x)) = (nb_n^5)^{1/2} C_B(x) + o((nb_n)^{1/2} b_n^2) \longrightarrow \mu(x)$, which explains the form of the asymptotic mean $\mu(x)$ in (17). The imposed condition on (b_n) is used solely to ensure convergence above. Noting that $\sigma^2(x)/f(x) = \mathbb{E}((Y - g(X))^2 | X = x)$, we see that the limiting distribution is the same as if (Z_i, X_i) were independent.

3.2. Long-range dependent sequences

Consider the case when either (Z_i) or (X_i) is long-range dependent. Let $\alpha_n(x) := K_{b_n} \star I_1(x)$ and $\beta_n(x) := K_{b_n} \star I_2(x)$, where

$$I_1(x) := -\int G(v,x) f_\infty^{(1,0)}(v,x)\, dv \quad \text{and} \quad I_2(x) := -\int G(v,x) f_\infty^{(0,1)}(v,x)\, dv.$$

Then

$$\widetilde{N}_n(x) = \frac{\alpha_n(x)}{n} \sum_{t=1}^{n} Z_{t,t-1} + \frac{\beta_n(x)}{n} \sum_{t=1}^{n} X_{t,t-1} =: \tilde{N}_{Z,n}(x) + \tilde{N}_{X,n}(x).$$

In the LRD case we need to know the order of magnitude of $\alpha_n(x)$ and $\beta_n(x)$ when $n \to \infty$. To this end define for $i = 1, 2$,

$$l_i(x) = \min\{s \geq 0 : I_i^{(s)}(x) \int x^s K(x)\, dx \neq 0\}, \tag{18}$$

where we assume that derivatives of $I_i(x)$ of sufficient order exist. We let $l_i(x) = \infty$ if the above condition is not fulfilled for any s. Then using standard reasoning it is easy to see that $\alpha_n(x) \sim b_n^{l_1(x)} I_1^{(l_1(x))}(x) \mu_{l_1(x)} / l_1(x)!$ and $\beta_n(x) \sim b_n^{l_2(x)} I_2^{(l_2(x))}(x) \mu_{l_2(x)} / l_2(x)!$. Let $l_i := \min\{l_i(x_1), \ldots, l_i(x_l)\}$. Then $\|\widetilde{N}_{Z,n}(x_k)\|$,

$k = 1, \ldots, l$ is dominated by $\|\widetilde{N}_{Z,n}(x_{k_0})\|$, where k_0 is the index of x_k, $k = 1, \ldots, l$ for which the minimal value l_1 is attained.

3.2.1. Some examples

Below we discuss some examples of the regression model (1) with various dependence structures of (Z, X), functions $G(z, x)$ and $(l_1(x), l_2(x))$. It turns out that the asymptotic distribution of $\hat{g}_n(x)$ depends on the pair $(l_1(x), l_2(x))$.

Example 1. Assume that Z is independent of X. Then $I_2(x) \equiv 0$ regardless of the form of $G(z, x)$, and $l_2(x) = \infty$. Indeed, in this case, $f_\infty(z, y) = h(z)f(y)$ and

$$\begin{aligned} I_2(x) &= -\int G(z, x) f_\infty^{(0,1)}(z, x)\, dz \\ &= -f'(x) \int G(z, x) h(z)\, dz = -f'(x) \mathbb{E} G(Z, x) = 0. \end{aligned}$$

Moreover, if $\lim_{z \to \pm\infty} G(z, x)h(z) = 0$, then $I_1(x) = f(x) G'_\infty(0, x)$, where $G_\infty(z, x) = \mathbb{E}(G(z + Z, x))$. Thus $l_1 = 0$ is equivalent to the fact that the power rank of the function $G(\cdot, x)$ w.r.t. distribution of Z (c.f., Ho and Hsing (1997)) is equal to 1.

Example 2. Consider the case of multiplicative errors $G(z, x) = G_1(z)G_2(x)$. Observe that the situation when errors don't depend on explanatory random variables, i.e., $G(z, x) = G(z)$, is a special case of this example. This also holds for the ARCH-type error function $G(z, x) = G(x)z$. As in Example 1, we have $I_2(y) \equiv 0$ in a neighborhood U_x of x provided $G_2(y) \neq 0$ in U_x. In this case $\int G_1(z) f_\infty(z, y) dz \equiv 0$ in U_x and the remark follows by taking derivatives w.r.t. y of both sides. In particular, assume additionally that Z is independent of X, $\mathbb{E} G_1(Z) = 0$, $\mathbb{E} G'_1(Z) = -\int G_1(z) h'(z) dz \neq 0$ and $(fG_2)(x) = (fG_2)'(x) = \cdots = (fG_2)^{(k-1)}(x) = 0$, $(fG_2)^{(k)}(x) \neq 0$ for some even $k \in \mathbb{N}$. Then $I_1(y) = \mathbb{E} G'_1(Z)(fG_2)(y)$ and $l_1(x) = k$, thus in this case $(l_1(x), l_2(x)) = (k, \infty)$.

Example 3. Suppose that (Z, X) has the bivariate normal distribution $N(0, 0, 1, 1, \rho)$ with $\rho \neq 0$, and that $G(z, x)$ satisfies $\int G(z, x) f_\infty(z, x) dz = 0$. Then $f_\infty^{(1,0)}(z, x) = -f_\infty(z, x)(z - \rho x)/(1 - \rho^2)$ and $f_\infty^{(0,1)}(z, x) = -f_\infty(z, x)(x - \rho z)/(1 - \rho^2)$, which entails $(1 - \rho^2)[I_2(x) + \rho I_1(x)] = -\int G(z, x) f_\infty(z, x)(1 - \rho^2) x dz = 0$ and $\rho I_2(x) + I_1(x) = -\int z G(z, x) f_\infty(z, x) dz$. The last two identities imply that $l_1(x) = l_2(x)$. If the conditional mean $\mathbb{E}[ZG(Z, X) | X = x] \neq 0$ at point x, then $l_1(x) = l_2(x) = 0$. This example shows, interestingly, that in the dependent bivariate normal case $l_1(x)$ and $l_2(x)$ are necessarily the same. One has to consider non-normal models to obtain different l_1 and l_2.

Example 4. We construct an example of a model for which $(l_1(x), l_2(x)) = (0, 2)$ for some $x \in \mathbb{R}$. To this end, let Z be standard normal with the density $\phi(z)$, $X = Z + \Omega$, where Ω is independent of Z with the density $f_\Omega(z)$, and $G(z,x) = z - t(x)$, where $t(x) = \mathbb{E}(Z|X = x)$. Then $f_\infty(z,x) = \phi(z) f_\Omega(x - z)$ and $f_\infty^{(0,1)}(z,x) + f_\infty^{(1,0)}(z,x) = -z\phi(z) f_\Omega(x - z)$. Thus

$$I_1(x) = -\int G(z,x) f_\infty^{(1,0)}(z,x) dz = \int G^{(1,0)}(z,x) f_\infty(z,x) dz = f(x).$$

As $\mathbb{E}(G(Z,X)|X = x) = 0$, $I_1(x) + I_2(x)$ equals

$$-\int [f_\infty^{(0,1)}(z,x) + f_\infty^{(1,0)}(z,x)] G(z,x)\, dz = \int [z - t(x)]^2 \phi(z) f_\Omega(x - z)\, dz.$$

Whence

$$\frac{I_1(x) + I_2(x)}{f(x)} = \mathbb{E}([Z - t(X)]^2 | X = x) = \mathrm{Var}(Z|X = x).$$

By Example 3, to obtain $l_1(x) \neq l_2(x)$, one has to try non-normal random variables Ω. Let $f_\Omega(z) = z^2 \phi(z)$. Then it is easily seen that $I_1(x) = f(x) = e^{-x^2/4}(2 + x^2)/8\sqrt{2\pi} \neq 0$. It follows that $l_1(x) = 0$ and $I_2(x) = (\mathrm{Var}(Z|X = x) - 1) f(x)$. Thus for x_0 such that $\mathrm{Var}(Z|X = x_0) = 1$ and $I_2^{(2)}(x_0) \neq 0$, we have $l_2(x_0) = 2$. It can be checked that $t(x) = x(-2 + x^2)/2(2 + x^2)$ and $\mathrm{Var}(Z|X = x) = (12 + x^4)/2(2 + x^2)^2$, and then $x_0 = \sqrt{\sqrt{20} - 4}$ satisfies both above conditions.

3.2.2. Limit theorems

From now on we assume that coefficients (a_i) and (c_i) decay hyperbolically according to (10) and bandwidths b_n satisfy

$$\frac{b_n^2}{\frac{1}{(nb_n)^{1/2}} + \frac{\sigma_{n,Z} b_n^{l_1}}{n} + \frac{\sigma_{n,X} b_n^{l_2}}{n}} \longrightarrow C \geq 0. \tag{19}$$

Observe that when (Z_i) and (X_i) are short-range dependent, standard deviations $\sigma_{n,Z}$ and $\sigma_{n,X}$ are both of order $\sqrt{n}$ and (19) coincides with the condition on bandwidths imposed in Theorem 1. Note that (19) cannnot hold if, e.g., $\sigma_{n,Z} b_n^{l_1}/n$ is the largest term in the denominator and $l_1 \geq 2$. We consider first the case when $(l_1, l_2) = (0, 0)$. Define

$$D(\beta,\gamma) = \frac{1}{(1-\beta)(1-\gamma)} \int_R [(1-u)_+^{1-\beta} - (-u)_+^{1-\beta}][(1-u)_+^{1-\gamma} - (-u)_+^{1-\gamma}] du, \tag{20}$$

where $1/2<\beta,\gamma<1$. Moreover, let $\tau=\rho D(\beta_Z,\beta_X)/[D^{1/2}(\beta_Z,\beta_Z)D^{1/2}(\beta_X,\beta_X)]$, $\rho=\mathbb{E}(\varepsilon_i\eta_i)/\sqrt{\mathbb{E}\varepsilon_i^2\mathbb{E}\eta_i^2}$. It can be shown that $D(\beta_Z,\beta_Z)$ equals $D(\beta_Z)$ defined in (11).

Theorem 2. *Assume that* $(l_1,l_2)=(0,0)$, $I_i(\cdot)$ *are continuous at* x_k, $k=1,\ldots,l$, $i=1,2$, *and* (19) *is satisfied.*

(a) *Assume that* $\beta_Z<\beta_X$. *If* (i) $\sigma_{n,Z}/n=o((nb_n)^{-1/2})$ *then* (17) *holds; if* (ii) $(nb_n)^{-1/2}=o(\sigma_{n,Z}/n)$, *then*

$$\frac{n}{\sigma_{n,Z}}\left(\hat g_n(x_i)-g(x_i),1\le i\le l\right)\Longrightarrow\left(\mathcal{N}_1\frac{I_1(x_i)}{f(x_i)}+\mu(x_i),1\le i\le l\right),\tag{21}$$

where $\mu(x)$ *is defined in Theorem* 1.

(b) *Assume that* $\beta_Z>\beta_X$. *Then* (a) *holds with the role of* $\sigma_{n,Z}$ *taken over by* $\sigma_{n,X}$ *and* $I_1(x)$ *replaced by* $I_2(x)$.

(c) *Assume that* $\beta_Z=\beta_X$. *If* (i) $\sigma_{n,Z}/n=o((nb_n)^{-1/2})$ *then* (17) *holds; if* (ii) $(nb_n)^{-1/2}=o(\sigma_{n,Z}/n)$ *and* $lim_{t\to\infty}L_X(t)/L_Z(t)\to A$, *then*

$$\frac{n}{\sigma_{n,Z}}\left(\hat g_n(x_i)-g(x_i),1\le i\le l\right)$$
$$\Longrightarrow\left(\mathcal{N}_1^0\frac{I_1(x_i)}{f(x_i)}+\mathcal{N}_2^0\frac{AI_2(x_i)}{f(x_i)}+(1+A)\mu(x_i),1\le i\le l\right),$$

where $(\mathcal{N}_1^0,\mathcal{N}_2^0)$ *has bivariate normal distribution with standard normal marginals and correlation* τ *defined below* (20).

The proof of Theorem 2 (c)(ii) follows from Lemma 8. We consider now cases when only one of l_i is equal to 0. Observe that as K is symmetric, the smallest possible nonzero value of l_i is 2.

Theorem 3. *Assume that* $(l_1,l_2)=(0,2)$ *and* $I_i(\cdot)\in\mathcal{C}^{(2)}(U(x_k,\delta_0))$ *for some* $\delta_0>0$ *and* $k=1,\ldots,l$, $i=1,2$, *and that* (19) *is satisfied. Moreover, let* $b_n=\bar L(n)n^{-\gamma}$ *for some* $\gamma>0$ *and slowly varying function* $\bar L(\cdot)$.

(a) *Assume that* $\beta_Z<\min(\beta_X+2\gamma,2\beta_X-1/2)$. *If* (i) $\sigma_{n,Z}/n=o((nb_n)^{-1/2})$ *then* (17) *holds; if* (ii) $(nb_n)^{-1/2}=o(\sigma_{n,Z}/n)$ *then* (21) *holds.*

(b) *Assume that* $\beta_X+2\gamma<\min(\beta_Z,2\beta_X-1/2)$. *If* $\sigma_{n,X}b_n^2/n=o((nb_n)^{-1/2})$ *then* (17) *holds.*

When $(l_1,l_2)=(2,0)$, an analogue of Theorem 3 is true with $\sigma_{n,Z}$ (respectively $\sigma_{n,X}$) replaced by $\sigma_{n,X}$ (respectively $\sigma_{n,Z}$) and $\beta_Z\,(\beta_X)$ replaced by $\beta_X\,(\beta_Z)$.

If $(nb_n)^{-1/2}=o(\sigma_{n,X}b_n^2/n)$ under conditions of Theorem 3 (b), one would expect the normalization $n/(\sigma_{n,X}b_n^2)$ to yield non-degenerate asymptotic law for centered $\hat g_n$. However, this is the case when we center at $g_n(x)$ instead of $g(x)$, as $n/(\sigma_{n,X}b_n^2)(g_n(x)-g(x)) \to \infty$ under (16) since (19) is not satisfied. Let $\bar{I}_2(x) = C_1(x) + 2^{-1}I_2^{(2)}(x)\mu_2$, where $C_1(x)$ is defined in (13) and assume that there exists $k \in \{1,\ldots,l\}$ such that $\bar{I}_2(x_k) \neq 0$. Moreover, assume that $\beta_X + 2\gamma < \min(\beta_Z, 2\beta_X - 1/2)$. Then if $(nb_n)^{-1/2} = o(\sigma_{n,X}b_n^2/n)$,

$$\frac{n}{\sigma_{n,X}b_n^2}\left(\hat{g}_n(x_i) - g_n(x_i),\, 1 \le i \le l\right) \Longrightarrow \mathcal{N}_1\left(\frac{\bar{I}_2(x_i)}{f(x_i)},\, 1 \le i \le l\right). \tag{22}$$

Analogously, for the case $(l_1,l_2) = (2,0)$ when $I_i(\cdot) \in \mathcal{C}^{(2)}(U(x_k,\delta_0))$ for some $\delta_0 > 0$ and $k = 1,\ldots,l$, $i = 1,2$, $\beta_Z + 2\gamma < \min(\beta_X, 2\beta_Z - 1/2)$ and $(nb_n)^{-1/2} = o(\sigma_{n,Z}b_n^2/n)$, we have

$$\frac{n}{\sigma_{n,Z}b_n^2}\left(\hat{g}_n(x_i) - g_n(x_i),\, 1 \le i \le l\right) \Longrightarrow \mathcal{N}_1\frac{\mu_2}{2}\left(\frac{I_1^{(2)}(x)}{f(x_1)},\, 1 \le i \le l\right). \tag{23}$$

The difference between the cases (l_1,l_2) equal to (0,2) and (2,0) exhibited by the different scaling constants in the last two asymptotic laws is due to the fact that in the asymptotic representation of $\hat{g}_n(x) - g_n(x)$ in Proposition 1, the sum $\sum X_{t,t-1}$ appears in both $\widetilde{N}_n$ and P_n whereas here there is only one term $\widetilde{N}_n$ involving $\sum Z_{t,t-1}$.

Consider the case $(l_1,l_2) = (2,2)$. If $(nb_n)^{-1/2} + \sigma_{n,X}b_n^2/n = o(\sigma_{n,Z}b_n^2/n)$, then (23) holds. In this case (19) is violated. If $\sigma_{n,Z}b_n^2/n + \sigma_{n,X}b_n^2/n = o((nb_n)^{-1/2})$, then we have (17) under condition (19). If $(nb_n)^{-1/2} + \sigma_{n,Z}b_n^2/n = o(\sigma_{n,X}b_n^2/n)$, then (22) is valid.

By a standard approach one may consider a kernel of order k higher than 2 to center at $g(x_i)$, $i = 1,\ldots,l$, in the two last results. This would require imposing a modified (19) with b_n^2 replaced by b_n^k.

In view of Theorem 3 (a) it may happen that, although dependence of (Z_t) is weaker than that of (X_t), the asymptotic law of $\hat{g}_n(x)$ is determined by the dependence of (Z_t) alone. For similar phenomena in the linear model with no intercept, see Choy and Taniguchi (2002). Note that for b_n defined in Theorem 3, $\sigma_{n,Z}/n = o((nb_n)^{-1/2})$ holds for $\gamma > 2(1-\beta_Z)$, whereas $\sigma_{n,Z}b_n^2/n = o((nb_n)^{-1/2})$ holds for $\gamma > (2/5)(1-\beta_Z)$. In Theorem 2 (a)(ii) (19) reduces to $nb_n^2/\sigma_{n,Z} \to C$, and this together with $(nb_n)^{-1/2} = o(\sigma_{n,Z}/n)$ implies $\beta_Z < 9/10$.

In case $I_2(\cdot) \equiv 0$, as in Examples 1 and 2, the asymptotic law of $\hat{g}_n(x) - g(x)$ may be analogously described by comparing the magnitudes of $M_n(x), I_1(x)$ and $P_n(x)$. We omit the statement of this result.

The question of how to use the presented results in inference is a challenging open problem. In particular, note that since Z_t are latent variables, no obvious estimate of $\sigma_{n,Z}$ in (11) exists. Csörgő and Mielniczuk (1999, p.216) proposed a heuristic approach to this problem when explanatory variables are independent;

see also Robinson (1997) for a solution to an analogous problem in fixed-design regression.

4. Auxiliary lemmas and proofs

Proof of Proposition 1. Clearly (14) follows from (4) and Lemmas 3, 4 and 6 below, since $\hat{f}_n(x) \to f(x) > 0$ in probability. Indeed, $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) \to 0$ in probability in view of (43) and $\mathbb{E}\hat{f}_n(x) \to f(x)$ as $f(\cdot)$ is continuous under $\mathcal{C}_1$.

By Karamata's theorem, we have $A_n = \mathcal{O}[\sum_{i=n}^{\infty}(L(i)i^{-\beta})^2] = \mathcal{O}[n^{1-2\beta}L^2(n)]$, and

$$\sum_{i=n+1}^{\infty}(\Theta_{n+i}-\Theta_i)^2 = \sum_{i=n+1}^{\infty}\mathcal{O}[(n\theta_i)^2] = n^2\sum_{i=n}^{\infty}\mathcal{O}[i^{1/2-2\beta}L^2(i)]^2 = n^{4-4\beta}L^4(n).$$

Note that $\Xi_n^2 = \mathcal{O}[2n\Theta_{2n}^2 + \sum_{i=n+1}^{\infty}(\Theta_{n+i}-\Theta_i)^2]$. Thus $\Xi_n^2 = \mathcal{O}[n\Theta_{2n}^2 + n^{4-4\beta}L^4(n)]$ proves the theorem.

In order to prove auxiliary lemmas we note first that easy reasoning along the lines of the proof of Lemma 1 in Wu and Mielniczuk (2002) implies that the property listed in the condition $\mathcal{C}_1$ for f_1 is inherited by f_t for $t = 2, 3, \ldots$, and f_∞. We state this fact as a separate lemma for future reference.

Lemma 1. (a) *If $\mathcal{C}_1$ holds, then f_t is twice continuously differentiable with bounded derivatives for $t = 2, \ldots, \infty$. In particular, ∇f_t is Lipschitz continuous for $t = 2, \ldots, \infty$. Moreover, $\nabla f_\infty(z, y) = \mathbb{E}\nabla f_t(z - Z_{t,0}, y - X_{t,0})$.*

We now state and prove a crucial auxiliary lemma which is used to find an approximation for N_n.

Lemma 2. *Let U_k be a stationary sequence such that U_k is $\widetilde{W}_k$-measurable and $\mathbb{E}U_k = 0$. Then*

$$\|\sum_{t=1}^{n} U_t\|^2 \le \sum_{k=-\infty}^{n}(\sum_{t=1}^{n}\|\mathcal{P}_1 U_{t-k+1}\|)^2. \tag{24}$$

Proof of Lemma 2. Noticing that $\mathcal{P}_k$, $k \in \mathbb{Z}$ are orthogonal, (24) follows from $\|\sum_{t=1}^{n} U_t\|^2 = \|\sum_{k=-\infty}^{n}\mathcal{P}_k\sum_{t=1}^{n} U_t\|^2 = \sum_{k=-\infty}^{n}\|\mathcal{P}_k\sum_{t=1}^{n} U_t\|^2$ and $\|\mathcal{P}_k\sum_{t=1}^{n} U_t\| \le \sum_{t=1}^{n}\|\mathcal{P}_k U_t\| = \sum_{t=1}^{n}\|\mathcal{P}_1 U_{t-k+1}\|$ by the triangle inequality and stationarity.

In Lemma 3 we derive asymptotic finite-dimensional distributions for $M_n(x)$, while asymptotic representations for $N_n(x)$ and $P_n(x)$ are stated in Lemmas 4 and 6.

Lemma 3. *Assume that $\mathcal{C}_2$ and $\mathcal{C}_3$ for any $x = x_i$, $i = 1, \ldots, l$, and that f_1 is bounded. Then*

$$\sqrt{nb_n}\,(M_n(x_1), M_n(x_2), \ldots, M_n(x_l)) \Longrightarrow (\mathcal{N}_1\sigma(x_1), \mathcal{N}_2\sigma(x_2), \ldots, \mathcal{N}_l\sigma(x_l))\,, \tag{25}$$

where $\sigma^2(x) = \kappa_2 \int G^2(v,x) f_\infty(v,x)\,dv$ *with* $\kappa^2 = \int K^2(s)\,ds$, *and the* $\mathcal{N}_i$ *are independent standard normal.*

Proof of Lemma 3. We prove the lemma for $l = 1$, the extension to the case $l > 1$ is routinely obtained by the Crámer-Wold device. Recall $J_t(x) = G(Z_t, X_t)K_b(x - X_t)$. Let $M_{n,t} = J_t(x) - \mathbb{E}[J_t(x)|\widetilde{W}_{t-1}]$. Since the summands of M_n form (triangular-array) martingale differences, it suffices to check conditions for the martingale CLT, namely Lindeberg's condition and

$$\mathbb{E}\left|\frac{b_n}{n}\sum_{t=1}^{n}\mathbb{E}(M_{n,t}^2|\widetilde{W}_{t-1}) - \sigma^2(x)\right| \to 0. \tag{26}$$

In order to prove (26) note that since f_1 is bounded, using reasoning as in proof of Lemma 2 in Wu and Mielniczuk (2002), it is enough to check (26) with $M_{n,t}$ replaced by $J_t(x)$. Thus we show $\mathbb{E}|n^{-1}\sum_{t=1}^{n} p_t(x) - \sigma^2(x)| \to 0$, where

$$\begin{aligned} p_t(x) &:= b_n\mathbb{E}[J_t^2(x)|\tilde{W}_{t-1}] \\ &= \int K^2(u)G^2(v, x - ub_n) f_1(v - Z_{t,t-1}, x - X_{t,t-1} - ub_n)\,du\,dv, \end{aligned} \tag{27}$$

recalling that $a_0 = c_0 = 1$. Let $s_t(x) = \int K^2(u)G^2(v,x)f_1(v - Z_{t,t-1}, x - X_{t,t-1})\,dudv$. Noting that integrand of $s_t(x)$ is nonnegative, we have

$$\begin{aligned} \mathbb{E}s_t(x) &= \kappa_2 \int G^2(v,x)\mathbb{E}f_1(v - Z_{t,t-1}, x - X_{t,t-1})\,dv \\ &= \kappa_2 \int G^2(v,x) f_\infty(v,x) dv \end{aligned}$$

by changing the order of integration. Write

$$\mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n} p_t(x) - \sigma^2(x)\right| \le \mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n} p_t(x) - \frac{1}{n}\sum_{t=1}^{n} s_t(x)\right| + \mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n} s_t(x) - \sigma^2(x)\right|.$$

Note that $s_t(x)$ is ergodic as an instantaneous transformation of a linear process is ergodic (c.f., Theorem 1.3.3 in Taniguchi and Kakizawa (2000)). By the Ergodic Theorem, the second term in the above bound tends to 0. The first term is bounded by

$$\begin{aligned} \mathbb{E}|p_1(x) - s_1(x)| &= \mathbb{E}\left|\int K^2(u)(B_1(x - ub_n) - B_1(x))\,du\right| \\ &\le \int K^2(u)\mathbb{E}|B_1(x - ub_n) - B_1(x)|\,du \\ &\le \kappa_2 \sup_{y \in U(x,b_n)} \mathbb{E}|B_1(y) - B_1(x)| \to 0 \end{aligned}$$

in view of $\mathcal{C}_2$, since the support of K is contained in $[-1,1]$ and $b_n \to 0$.

To check Lindeberg's condition, it is enough to verify that

$$\frac{1}{b_n}\int_{\{(s,t):\, G^2(s,t)K^2((x-t)b_n^{-1})\geq \varepsilon n b_n\}} G^2(s,t)K^2\left(\frac{x-t}{b_n}\right) f_\infty(s,t)\,dsdt \to 0$$

by Corollary 9.5.2 in Chow and Teicher (1988). Since K is bounded and compactly supported, for sufficiently large n the left hand side is bounded by

$$C\int_{\{(s,t):\, \bar{G}^2(s)\geq \varepsilon C^{-1}nb_n,\, t\in R\}} \bar{G}^2(s) f_\infty(s,t)\,dsdt = C\int_{\{s:\bar{G}(s)\geq \varepsilon C^{-1}nb_n\}} \bar{G}^2(s)h(s)\,ds$$

with $C=\sup K^2(\cdot)$, and the bound tends to 0 under $\mathcal{C}_3$ as $nb_n\to\infty$.

Lemma 4. *Assume $\mathcal{C}_1$ and $\mathcal{C}_4$. Then*

$$\|N_n(x)-\tilde{N}_n(x)\| = \mathcal{O}(\Xi_n/n), \tag{28}$$

where $\tilde{N}_n(x)$ and Ξ_n are defined in Proposition 1.

Proof of Lemma 4. The summands of N_n can be written as $\mathbb{E}(J_t(x)|\widetilde{W}_{t-1}) = K_b\star T_t(x)$, where $T_t(x)=\int G(z,x)f_1(z-Z_{t,t-1},x-X_{t,t-1})\,dz$. Observe that $\|\mathcal{P}_i[K_b\star T_t(x)-K_b\star S_t(x)]\|=0$ for $i\geq t$ as $T_t(\cdot)$ and $S_t(\cdot)$ are $\widetilde{W}_{t-1}$-measurable. Thus (28) follows from Lemma 2 provided

$$\|\mathcal{P}_1[K_b\star T_t(x)-K_b\star S_t(x)]\|\leq C\theta_t \tag{29}$$

for sufficiently large t. Indeed, it is easy to see that

$$\begin{aligned}\sum_{k=-\infty}^{n}\Big(\sum_{t=1+k^+}\theta_{t-k+1}\Big)^2 &\leq \sum_{k=-\infty}^{0}\Big(\sum_{t=1}^{n}\theta_{t-k+1}\Big)^2+\sum_{k=1}^{n}\Big(\sum_{t=1+k^+}^{n}\theta_{t-k+1}\Big)^2\\ &\leq \sum_{k=1}^{\infty}(\Theta_{n+k}-\Theta_k)^2+n\Theta_n^2=\Xi_n^2,\end{aligned}$$

where $k^+=\max(0,k)$. Now we verify (29). Taking into account compactness of support of K and the fact that K is bounded it is easy to see that for any $(\xi_v)_{v\in R}$,

$$\begin{aligned}&\mathbb{E}\left[\int K_b(x-v)\xi_v\,dv\right]^2\\ &\leq \int\int |K_b(x-v)||K_b(x-v')|\mathbb{E}(|\xi_v||\xi_{v'}|)\,dv\,dv'\\ &\leq \int\int |K_b(x-v)||K_b(x-v')|\|\xi_v\|\|\xi_{v'}\|\,dv\,dv' \leq \sup_{v:|v-x|\leq b_n}\|\xi_v\|^2.\end{aligned}$$

Observe that $-\mathcal{P}_1 K_b \star S_t(x) = \int K_b(x-y)G(z,y)\nabla^{\mathbf{T}} f_\infty(z,y)\mathcal{A}_{t-1}\binom{\varepsilon_1}{\eta_1}\,dz$ and $\mathcal{P}_1[K_b \star T_t(x) - K_b \star S_t(x)]$ equals

$$\int\int K_b(x-y)G(z,y)[\mathcal{P}_1 f_1(z-Z_{t,t-1},x-X_{t,t-1})+\nabla^{\mathbf{T}} f_\infty(z,y)\mathcal{A}_{t-1}\begin{pmatrix}\varepsilon_1\\ \eta_1\end{pmatrix}]\,dzdy.$$

Thus the following Lemma 5 entails (29).

Lemma 5. *Assume $\mathcal{C}_1$ and $\mathcal{C}_4$. Then for sufficiently large t,*

$$\sup_{y:|y-x|<\delta_0} \|\int G(z,y)[\mathcal{P}_1 f_1(z-Z_{t,t-1},y-X_{t,t-1})+\nabla^{\mathbf{T}} f_\infty(z,y)\mathcal{A}_{t-1}\begin{pmatrix}\varepsilon_1\\ \eta_1\end{pmatrix}]dz\| \le C\theta_t. \tag{30}$$

Proof of Lemma 5. Let $R_{1,t}(z,y) = \nabla f_{t-1}(z-Z_{t,0},y-X_{t,0}) - \nabla f_{t-1}(z,y)$ and $R'_{1,t}(z,y) = \nabla f_{t-1}(z-Z_{t,1},y-X_{t,1}) - \nabla f_{t-1}(z,y)$. By Lemma 1, $\mathbb{E}R'_{1,t}(z,y) = \nabla f_\infty(z,y) - \nabla f_{t-1}(z,y)$. Using $|\mathbb{E}\xi| \le \|\xi\|$ and (5), we get $\sup_{y:|y-x|<\delta_0} |\int G(z,y)[\nabla f_\infty(z,y) - \nabla f_{t-1}(z,y)]\,dz| \le C\sqrt{A_{t-1}}$. Using the last inequality and (5) again, we obtain via the triangle inequality that

$$\sup_{y:|y-x|<\delta_0} \|\int G(z,y)[\nabla f_{t-1}(z-Z_{t,0},y-X_{t,0})-\nabla f_\infty(z,y)]dz\| \le C\sqrt{A_t}+C\sqrt{A_{t-1}}. \tag{31}$$

Let $\mathbf{w} = \mathcal{A}_{t-1}(\varepsilon_1,\eta_1)^{\mathbf{T}}$. The last inequality implies

$$\begin{aligned} &\sup_{y:|y-x|<\delta_0} \|\int G(z,y)[\nabla^{\mathbf{T}} f_{t-1}(z-Z_{t,0},y-X_{t,0}) - \nabla^{\mathbf{T}} f_\infty(z,y)]\mathbf{w}\,dz\| \\ &\le 2C|\mathcal{A}_{t-1}|\sqrt{A_{t-1}}. \end{aligned} \tag{32}$$

Let $(\varepsilon'_i,\eta'_i)_{-\infty}^{\infty}$ be an i.i.d. copy of $(\varepsilon_i,\eta_i)_{-\infty}^{\infty}$, $Z^*_{t,1} = Z_{t,1} - a_{t-1}\varepsilon_1 + a_{t-1}\varepsilon'_1$ and $X^*_{t,1} = X_{t,1} - c_{t-1}\eta_1 + c_{t-1}\eta'_1$. Namely $Z^*_{t,1}$ and $X^*_{t,1}$ are $Z_{t,1}$ and $X_{t,1}$ with ε_1 and η_1 replaced by ε'_1 and η'_1, respectively. Let $R^*_{2,t}(z,y)$ be $R_{2,t}(z,y)$ with ε_1 and η_1 replaced by ε'_1 and η'_1, respectively. Hence (6) entails $\sup_{y:|y-x|<\delta_0} \|\int G(z,y) R^*_{2,t}(z,y)\,dz\| \le C|\mathcal{A}_{t-1}|^2$ and

$$\sup_{y:|y-x|<\delta_0} \|\int G(z,y)[R_{2,t}(z,y) - R^*_{2,t}(z,y)]\,dz\| \le 2C|\mathcal{A}_{t-1}|^2. \tag{33}$$

Observe that $\mathcal{P}_1 f_1(z-Z_{t,t-1},y-X_{t,t-1}) = f_{t-1}(z-Z_{t,1},y-X_{t,1}) - f_t(z-Z_{t,0},y-X_{t,0})$ and $f_t(z-Z_{t,0},y-X_{t,0}) = \mathbb{E}[f_{t-1}(z-Z^*_{t,1},y-X^*_{t,1})|\widetilde{W}_1]$. We have $\mathbb{E}[R_{2,t}(z,y) - R^*_{2,t}(z,y)|\widetilde{W}_1] = \mathcal{P}_1 f_1(z-Z_{t,t-1},y-X_{t,t-1}) + \nabla^{\mathbf{T}} f_{t-1}(z-Z_{t,0},y-X_{t,0})\mathbf{w}$, which implies (30) by (33), (32) and the definition of $\theta_t = |\mathcal{A}_{t-1}|\sqrt{A_{t-1}}$ as $|\mathcal{A}_{t-1}|^2 = \mathcal{O}(|\mathcal{A}_{t-1}|\sqrt{A_{t-1}})$.

Lemma 6. *Assume $\mathcal{C}_5$ and $\mathcal{C}_6$. If* (i) *$\mathcal{C}_7$ holds, then*

$$P_n(x) = \mathcal{O}_P\Big(\frac{b_n}{\sqrt{n}}\Big); \tag{34}$$

If (ii) *$\mathcal{C}_7$ holds, then*

$$P_n(x) = C_1(x)\frac{b_n^2}{n}\sum_{t=1}^n X_{t,t-1} + \mathcal{O}_P(\sqrt{b_n/n}) + o_P(b_n^2\sigma_{n,X}/n), \tag{35}$$

where $C_1(x)$ is defined in (13).

Note that the $\mathcal{O}_P(\sqrt{b_n/n})$ and $\mathcal{O}_P(b_n/\sqrt{n})$ terms appearing in the approximations of $P_n(x)$ are $o_P(1/\sqrt{nb_n})$, and consequently $o_P(M_n(x))$.

The following lemma plays a crucial role in the proof of Lemma 6.

Lemma 7. *Let $H_n(y) = \sum_{t=1}^n f_\eta(y - X_{t,t-1}) - f(y)$. Assume $\mathcal{C}_6$ and $\mathcal{C}_7$. Then*

$$\text{(i)} \qquad \sum_{\iota=0}^1 \sup_y \|\mathcal{P}_1[f_\eta^{(\iota)}(y - X_{t,t-1}) - f^{(\iota)}(y) + f^{(\iota+1)}(y)X_{t,t-1}]\| = \mathcal{O}(\gamma_{t-1}|c_{t-1}|), \tag{36}$$

$$\text{(ii)} \qquad \sup_y \|H_n(y) + \sum_{t=1}^n f'(y)X_{t,t-1}\| + \sup_y \|H_n'(y) + \sum_{t=1}^n f''(y)X_{t,t-1}\| = o(\sigma_{n,X}), \tag{37}$$

when condition $\mathcal{C}_7$ (ii) *holds and $o(\sigma_{n,X})$ is replaced by $\mathcal{O}(\sqrt{n})$ under $\mathcal{C}_7$* (i).

Proof of Lemma 7. (i) The proof is similar to that of Lemma 5. Recall that $X_{t,1}^* = X_{t,1} - c_{t-1}\eta_1 + c_{t-1}\eta_1'$ as in the proof of Lemma 5. Similarly as in (33), (7) implies that

$$\sum_{\iota=0}^1 \sup_y \|f_{t-1,X}^{(\iota)}(y - X_{t,1}) - f_{t-1,X}^{(\iota)}(y - X_{t,1}^*) + f_{t-1,X}^{(\iota+1)}(y - X_{t,0})c_{t-1}(\eta_1 - \eta_1')\|$$
$$\leq 2\gamma_{t-1}|c_{t-1}|.$$

Observe that $\mathcal{P}_1 f_\eta^{(\iota)}(y - X_{t,t-1}) = f_{t-1,X}^{(\iota)}(y - X_{t,1}) - f_{t,X}^{(\iota)}(y - X_{t,0})$ and $f_{t,X}^{(\iota)}(y - X_{t,0}) = \mathbb{E}f_{t-1}^{(\iota)}(z - X_{t,1}^*)|\widetilde{W}_1)$. Reasoning as in Lemma 5 we get from the last displayed inequality,

$$\sum_{\iota=0}^1 \sup_y \|\mathcal{P}_1[f_\eta^{(\iota)}(y - X_{t,t-1}) + f_{t-1,X}^{(\iota+1)}(y - X_{t,0})c_{t-1}\eta_1]\| \leq 2\gamma_{t-1}|c_{t-1}|.$$

To establish (36), it suffices to verify that $\sup_y \|f_{t-1,X}^{(\iota+1)}(y - X_{t,0}) - f^{(\iota+1)}(y)\| \leq 2\gamma_{t-1}$. Since $\mathbb{E}f_{t-1,X}^{(\iota+1)}(y - X_{t,1}) = f^{(\iota+1)}(y)$, (8) implies $\sup_y |f^{(\iota+1)}(y) - f_{t-1,X}^{(\iota+1)}(y)| \leq$

γ_{t-1}, and as at (31), $\sup_y \|f^{(\iota+1)}_{t-1,X}(y-X_{t,0})-f^{(\iota+1)}(y)\| \le \sup_y \|f^{(\iota+1)}_{t-1,X}(y-X_{t,0})-f^{(\iota+1)}_{t-1,X}(y)\| + \sup_y |f^{(\iota+1)}(y) - f^{(\iota+1)}_{t-1,X}(y)| \le 2\gamma_{t-1}$.

(ii) By Lemma 2 and part (i), $\|H_n(y) + \sum_{t=1}^n f'(y)X_{t,t-1}\|^2 = \mathcal{O}[\sum_{k=-\infty}^n(\sum_{t=1}^n \gamma_{t-k}|c_{t-k}|)^2]$. Notice that $\sigma^2_{n,X} = \sum_{k=-\infty}^n(\sum_{t=1}^n c_{t-k})^2$. Assume that $\mathcal{C}_7$ (ii) holds (the proof in the other case is similar but simpler). Let $s_k = \sum_{i=0}^k c_i$. As $\sigma^2_{n,X} \ge \sum_{k=1}^n(\sum_{t=1}^n c_{t-k})^2 \ge \sum_{k=1}^n s^2_{n-k}$, then, since $s_n \to \infty$, it follows that $n = o(\sigma^2_{n,X})$. For any fixed integer $\kappa \ge i_0$, $\sum_{t=\kappa}^m |\gamma_t c_t| \le \gamma_\kappa |\sum_{t=\kappa}^m c_t|/\tau$, so we have by elementary manipulations and $n = o(\sigma^2_{n,X})$ that $\limsup_{n\to\infty}\sigma^{-2}_{n,X}\sum_{k=-\infty}^n(\sum_{t=1}^n|\gamma_{t-k}c_{t-k}|)^2 \le \gamma_\kappa^2/\tau^2$, which proves the lemma since $\gamma_t \downarrow 0$ and κ is arbitrarily chosen. The other inequality is proved similarly.

Proof of Lemma 6. We first prove part (ii). Let $Q_t = [g(X_t)-g(x)]K_b(x-X_t)$, $W_n = n^{-1}\sum_{t=1}^n Q_t$ and $\widetilde{X}_t = (\ldots,\eta_{t-1},\eta_t)$. Let $D_n(g-g(x)) = n^{-1}\sum_{t=1}^n[Q_t - \mathbb{E}(Q_t|\widetilde{X}_{t-1})]$ and $B_n(g-g(x)) = n^{-1}\sum_{t=1}^n[\mathbb{E}(Q_t|\widetilde{X}_{t-1}) - \mathbb{E}Q_t]$. Then $W_n - \mathbb{E}W_n = D_n(g-g(x)) + B_n(g-g(x))$ and

$$\begin{aligned} P_n &= \frac{1}{n}\sum_{t=1}^n (g(X_t) - g_n(x))K_b(x-X_t) \\ &= D_n(g-g(x)) + B_n(g-g(x)) - (\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x))(g_n(x) - g(x)). \end{aligned} \tag{38}$$

As the summands of $D_n(g-g(x))$ are uncorrelated, we have in view of $\mathcal{C}_1$ and $\mathcal{C}_5$ that

$$\begin{aligned} \|D_n(g-g(x))\|^2 &= \frac{1}{n}\|Q_t - \mathbb{E}(Q_t|\widetilde{X}_{t-1})\|^2 \le \frac{1}{n}\|Q_1\|^2 \\ &= \frac{1}{nb_n}\int (g(x-ub_n) - g(x))^2 f_\eta(x-ub_n)K^2(u)\,du = \mathcal{O}(b_n/n). \end{aligned}$$

Recall that $H_n(y) = \sum_{t=1}^n f_\eta(y - X_{t,t-1}) - f(y)$ as in Lemma 7. Observe that $B_n(g-g(x)) = n^{-1}\int(g(x-ub_n) - g(x))K(u)H_n(x-ub_n)\,du$. As g is two times continuously differentiable in the neighborhood of x we have that, uniformly,

$$g(x-ub_n) - g(x) = -b_n u g'(x) + \frac{1}{2}b_n^2 u^2 g''(x) + o(b_n^2) \tag{39}$$

for $|u| \le 1$. By (ii) of Lemma 7, $\sup_y \mathbb{E}|H_n(y)| \le \sup_y \|H_n(y)\| = \mathcal{O}(\sigma_{n,X})$. Hence

$$nB_n(g-g(x)) = \int [-b_n u g'(x) + \frac{1}{2}b_n^2 u^2 g''(x)]K(u)H_n(x-ub_n)\,du + o_P(b_n^2\sigma_{n,X}). \tag{40}$$

Again by (ii) of Lemma 7, since K has support within $[-1,1]$ and $\mathbb{E}|\xi| \le \|\xi\|$,

$$\mathbb{E}\left|\int u^2K(u)H_n(x-ub_n)\,du + \int u^2K(u)f'(x-ub_n)\,du\sum_{t=1}^n X_{t,t-1}\right| = o(\sigma_{n,X}) \tag{41}$$

and, since f'' is continuous at x,

$$\begin{aligned}
&\sup_{|y-x|\le b_n} \mathbb{E}|H_n'(y) - H_n'(x)| \\
&\le \sup_y \mathbb{E}|H_n'(y) + \sum_{t=1}^n f''(y)X_{t,t-1}| \\
&\quad + \sup_{|y-x|\le b_n} \mathbb{E}|\sum_{t=1}^n [f''(y) - f''(x)]X_{t,t-1}| + \mathbb{E}|H_n'(x) + \sum_{t=1}^n f''(x)X_{t,t-1}| \\
&= o(\sigma_{n,X}).
\end{aligned}$$

Notice that $H_n(x - ub_n) - H_n(x) = \int_0^{-ub_n} H_n'(x+v)dv$. Thus we have

$$\begin{aligned}
&\mathbb{E}\left|\int uK(u)[H_n(x-ub_n) - H_n(x) + ub_nH_n'(x)]du\right| \\
&\le \int |uK(u)| \int_{-|ub_n|}^{|ub_n|} \mathbb{E}|H_n'(x+v) - H_n'(x)|dvdu = o(b_n\sigma_{n,X}). \qquad (42)
\end{aligned}$$

Collecting (40), (41) and (42), we have by another application of (ii) of Lemma 7 that

$$nB_n(g - g(x)) = C_n^* b_n^2 \sum_{t=1}^n X_{t,t-1} + o(b_n^2\sigma_{n,X}) = Cb_n^2 \sum_{t=1}^n X_{t,t-1} + o(b_n^2\sigma_{n,X}),$$

where $C = \mu_2[-g''(x)f'(x)/2 - g'(x)f''(x)]$ in view of $C_n^* = -(g''(x)/2)\int u^2K(u)$ $f'(x-ub_n)du - g'(x)f''(x)\mu_2 = C + o(1)$. By the proof of Theorem 2 Wu and Mielniczuk (2002) and Lemma 7, we have under $\mathcal{C}_6$ and $\mathcal{C}_7$ that

$$\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x) = M_n'(x) - n^{-1}f'(x)\sum_{t=1}^n X_{t,t-1} + o_P(\max((nb_n)^{-1/2}, \sigma_{n,X}/n)), \qquad (43)$$

where $M_n' = n^{-1}\sum_{i=1}^n K_b(x - X_i) - \mathbb{E}(K_b(x-X_i)|\widetilde{X}_{i-1})$, with $\widetilde{X}_t = (\ldots, \eta_{t-1}, \eta_t)$ a martingale such that $(nb_n)^{1/2}M_n' = \mathcal{O}_P(1)$. The lemma is proved using (16) and (43), and noting that $C_1(x) = C_Bf'(x) + C$ and $b_n^2M_n' = o((b_n/n)^{1/2})$.

Part (i) follows from (40) and Lemma 7 by using $\sup_y ||H_n(y)|| = \mathcal{O}(\sqrt{n})$.

The last lemma is used to investigate the boundary case $\beta_Z = \beta_X$ in Theorem 2.

Lemma 8. *We have* $(\sigma_{n,Z}^{-1}\sum_{t=1}^n Z_t, \sigma_{n,X}^{-1}\sum_{t=1}^n X_t) \Longrightarrow N(0,\Sigma)$*, where* Σ *is a* 2×2 *matrix with* $\Sigma_{11} = \Sigma_{22} = 1$ *and* $\Sigma_{12} = \Sigma_{21} = \tau$ *defined above Theorem* 2.

Proof of Lemma 8. We use the Crámer-Wold device. For $n \ge 1$, let $u_n = \sum_{j=0}^n a_j \sim L_Z(n)n^{1-\beta_Z}/(1-\beta_Z)$ and $v_n = \sum_{j=0}^n c_j \sim L_X(n)n^{1-\beta_X}/(1-\beta_X)$;

put $u_i = v_i = 0$ for $i < 0$. Then for $c_1, c_2 \in \mathbb{R}$, we have $c_1\sigma_{n,Z}^{-1}\sum_{t=1}^{n} Z_t + c_2\sigma_{n,X}^{-1}\sum_{t=1}^{n} X_t = \sum_{j=-\infty}^{n}[c_1\sigma_{n,Z}^{-1}(u_{n-j} - u_{-j})\varepsilon_j + c_2\sigma_{n,X}^{-1}(v_{n-j} - v_{-j})\eta_j]$. Using the Lindeberg-Feller CLT, we need to verify (i) $\sum_{j=-\infty}^{n}\mathbb{E}[c_1\sigma_{n,Z}^{-1}(u_{n-j} - u_{-j})\varepsilon_j + c_2\sigma_{n,X}^{-1}(v_{n-j} - v_{-j})\eta_j]^2 \to c_1^2 + c_2^2 + 2c_1c_2\Sigma_{12}$ and (ii) the Lindeberg condition. For (i) we focus on the cross-product term. Using Karamata's theorem, we can show that

$$\begin{aligned}
&\sum_{j=-\infty}^{n}(u_{n-j} - u_{-j})(v_{n-j} - v_{-j})\\
&\sim \frac{L_Z(n)L_X(n)}{(1-\beta_Z)(1-\beta_X)}\int_{-\infty}^{n}[(n-x)_+^{1-\beta_Z} - (-x)_+^{1-\beta_Z}][(n-x)_+^{1-\beta_X} - (-x)_+^{1-\beta_X}]dx\\
&\sim \frac{L_Z(n)L_X(n)}{(1-\beta_Z)(1-\beta_X)}n^{1-\beta_Z}n^{1-\beta_X}nD(\beta_Z, \beta_X),
\end{aligned}$$

where $a_+ = \max(a, 0)$. The Lindeberg condition follows from the proof of Theorem 18.6.5 in Ibragimov and Linnik (1971) which asserts that $\sigma_{n,Z}^{-1}\sum_{t=1}^{n} Z_t \Longrightarrow \mathcal{N}_1$. We omit the details.

A functional weak convergence result for multivariate linear processes under a different set of conditions can be found in Marinucci and Robinson (2000).

Acknowledgement

Comments of an associate editor which helped to improve a previous version of this paper are gratefully acknowledged.

References

Cheng, B. and Robinson, P. M. (1991). Density estimation in strongly dependent non-linear time series. *Statist. Sinica* **1**, 335-359.

Cheng, B. and Robinson, P. M. (1994). Semiparametric estimation from time series with long-range dependence. *J. Econometrics* **64**, 335-353.

Chow, Y. S. and Teicher, S. (1988). *Probability Theory.* 2nd edition. Springer, New York.

Choy, K. and Taniguchi, M. (2002). Stochastic regression model with dependent disturbances. *J. Time Ser. Anal.* **22**, 175-196.

Csörgő, S. and Mielniczuk, J. (1999). Random-design regression under long-range dependence. *Bernoulli* **5**, 209-224.

Csörgő, S. and Mielniczuk, J. (2000). The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. *Statist. Sinica* **10**, 771-787.

Györfi, L, Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series.* Springer, New York.

Hall, P. and Hart, J. D. (1990). Convergence rates in density estimation for data from infinite-order moving average process. *Probab. Theory Related Fields* **87**, 253-274.

Hidalgo, F. J. (1997). Nonparametric estimation with strongly dependent multivariate time series. *J. Time Ser. Anal.* **8**, 95-122.

Ho, H.-C. (1996). On central and non-central limit theorems in density estimation for sequences of long-range dependence. *Stochastic Process. Appl.* **63**, 153-174.

Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24**, 992-1024.

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25**, 1636-1669.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters, Noordhoff-Groningen.

Koul, H. S. (1992). M-estimators in linear models with long range dependent errors. *Statist. Probab. Lett.* **14**, 153-164.

Marinucci, D. and Robinson, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Process. Appl.* **86**, 103-120

Nze, P. A., Bühlmann, P. and Doukhan, P. (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.* **30**, 397-430.

Robinson, P. M. (1983). Nonparametric estimates for time series. *J. Time Ser. Anal.* **4**, 185-207.

Robinson, P. M. (1991). Nonparametric function estimation for long memory time series. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (Edited by W. Barnett, J. Powell and G. Tauchen), 437-457. Cambridge University Press.

Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25**, 2054-2083.

Robinson, P. M. and Hidalgo, F. J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25**, 77-104.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30**, 1441-1459.

Wu, W. B. and Woodroofe, M. (2003). Asymptotic theory for multiple linear processes. Technical Report 406, Department of Statistic, University of Michigan.

Wu, W. B. (2003). Empirical Processes of Long-memory Sequences. *Bernoulli* **9**, 809-831.

Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01–237 Warsaw, Poland.

E-mail: miel@ipipan.waw.pl

Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, U.S.A.

E-mail: wbwu@galton.uchicago.edu

(Received November 2002; accepted December 2003)