Regression estimation by local polynomial fitting for multivariate data streams

Aboubacar Amiri & Baba Thiam

Table of contents continued on back cover

ONLINE FIRST

STARTING IN 2016 MORE CONTENT

 Springer

 Springer

Springer

CrossMark

Regression estimation by local polynomial fitting for multivariate data streams

Aboubacar Amiri[1] · Baba Thiam[1]

Abstract In this paper we study a local polynomial estimator of the regression function and its derivatives. We propose a sequential technique based on a multivariate counterpart of the stochastic approximation method for successive experiments for the local polynomial estimation problem. We present our results in a more general context by considering the weakly dependent sequence of stream data, for which we provide an asymptotic bias-variance decomposition of the considered estimator. Additionally, we study the asymptotic normality of the estimator and we provide algorithms for the practical use of the method in data streams framework.

1 Introduction

Streaming data, or data streams, are massive data arriving in streams, and if they are not immediately processed or stored, then they are lost forever. Large organizations, such as financial companies in financial markets, generate hundreds of millions of records of transactions in a short time, and many scientific studies also generate gigabytes of

✉ Baba Thiam
baba.thiam@univ-lille3.fr

Aboubacar Amiri
aboubacar.amiri@univ-lille3.fr

1 Université Lille 3, LEM-CNRS (UMR 9221), Domaine universitaire du "pont de bois", Rue du barreau, BP 60149, 59653 Villeneuve d'Ascq Cedex, France

 Springer

data in minutes. In many real situations, such as scientific and real applications, large amounts of raw data can be collected extremely easily such that experiments typically yield a considerable number of data points. Data streams naturally arise in many fields, such as oceanography, climatology, Internet and web traffic, finance, and security. For example, with the advances in modern technology, one can imagine a sensor sending back to a base station a reading of the ocean surface temperature every second or less. The data produced by this sensor are a stream of real numbers. As another example, consider a sensor that reports the surface height of the ocean rather than temperature. The surface height of the ocean varies quite rapidly compared with temperature, and thus, the sensor might return a reading every tenth of a second. Moreover, to learn something about ocean behavior, we might want to deploy many sensors, each sending back to the central node a stream at a rate of ten per second. In these situations, the data arrive so rapidly that it is impossible for the user to store all the data on a disk (as a traditional database) and then interact with them later. Consequently, to address such large data, the traditional nonparametric techniques require a considerable amount of time to be computed and thus become useless in practice if real-time forecasts are expected. Therefore, the development of processing methods and analyzing these data streams effectively and efficiently have become challenging problems in statistics and computational science. A substantial amount of literature has been dedicated to providing specific methods for these particular data. We refer the readers to Aggarwal (2007) and Domingos and Hulten (2003) for a detailed overview. Recent advances on this topic were the focus of the papers of Zhou et al. (2003), Cao et al. (2012), and Xu et al. (2014), which provided kernel density estimation methods over data streams.

In this paper, we are interested in the regression model, and we investigate how to provide simultaneously an estimation of a nonparametric regression function and its derivatives from a locally weighted least squares fit over the data streams. The regression model that we consider is a statistical relation that describes the link between two random variables: a response variable $Y \in \mathbb{R}$ and a regressor $\mathbf{X} \in \mathbb{R}^d$ ($d \geq 1$), that is, $Y = r(\mathbf{X}) + \varepsilon$, where ε is an error random variable. Regression estimation is defined as the construction of an estimate of the regression function r from observed sample data of $(\mathbf{X}, Y)$. The statistics community has devoted many efforts to providing an estimation of the regression function and its derivative using parametric, semi-parametric and nonparametric approaches for univariate and multivariate data. Parametric regression estimation is conducted under the assumption that the regression function admits a known parametric form, such as a polynomial, whereas in nonparametric estimation, no such assumptions are required. Because the statistical link between two random data streams is completely unknown in most applications, nonparametric regression estimation has found a considerable number of applications in many domains. Local polynomial kernel estimation is a classical nonparametric technique for studying the regression model, and it has recently attracted considerable attention among statisticians and econometricians. We refer the readers to the monograph of Fan and Gijbels (1996) for an overview on both the theoretical and practical aspects of the local polynomial method. The principal advantage of this method is its ability to provide simultaneously an estimation of a nonparametric regression function and its derivatives. Moreover, the local polynomial method has better estimation accuracy at the boundary region of the data support and minimax efficiency over

the local constant estimator (Fan and Gijbels 1996). Some significant references on the subject are Fan and Gijbels (1992, 1995), Ruppert and Wand (1994), and Ruppert et al. (1995). The local polynomial estimator for a univariate nonparametric regression model was studied by Fan et al. (1996), who derived a bias-variance representation. In the multivariate case, Masry (1996a, b) derived the rate of convergence and the asymptotic normal distribution of the local polynomial regression estimator for time series data. Gu et al. (2015) extended the result of Masry (1996b) to obtain explicit leading bias terms for the entire vector of the local polynomial estimator. Hansen (2008) provided uniform consistency results for kernel estimators of density with applications to Nadaraya–Watson and local polynomial regression estimators under strong mixing conditions. Recently, Liang and Baek (2016) proposed a Nadaraya–Watson type and local linear estimators of conditional density function and established their asymptotic normality with the left-truncated and α-mixing data. Robust estimation of multivariate regression model using kernel weighted local linear regression was investigated by Li and Zheng (2009). As indicated above, in the data stream framework, many observed data points can be available in a very short period of time. Consequently, the computational time of a standard local polynomial estimator can quickly become large in such situations, and thus, a real-time regression estimation is increasingly more difficult to obtain as the time increases. This situation arises from the fact that such a standard local polynomial estimator has to be completely recomputed when a new observation comes into the picture. This is why we consider in this paper the local polynomial estimation problem in the context of data streams. We propose a sequential technique based on a multivariate counterpart of the stochastic approximation method for successive experiments proposed in Robbins and Monro (1951) for the local polynomial estimation problem. Our estimator is a recursive version of the traditional estimator based on the locally weighted fitting adapted to the multivariate data stream model. Its construction is based on the concepts developed in different contexts by Vilar and Vilar (1998, 2000) for locally weighted smoothers and by Amiri (2012), Amiri et al. (2014), and Huang et al. (2014) for kernel smoothers.

2 Locally weighted regression for data streams

2.1 Presentation of the model

Let $\mathbf{W}_1, \ldots, \mathbf{W}_n$ represent a sequence of observed arrays of the form

$$\mathbf{W}_t := \left\{ (\mathbf{X}_{t1}, Y_{t1}), \ldots, (\mathbf{X}_{tN_t}, Y_{tN_t}) \right\}, \quad t = 1, \ldots, n,$$

where the sub-sample $(\mathbf{X}_{t1}, Y_{t1}), \ldots, (\mathbf{X}_{tN_t}, Y_{tN_t})$ is a sequence of bivariate random variables identically distributed as a stationary stochastic process $(\mathbf{X}, Y)$ in $\mathbb{R}^d \times \mathbb{R}$ ($d \geq 1$). We assume that the $(\mathbf{X}_{tj}, Y_{tj})$'s have a common joint density $f_{(\mathbf{X}, Y)}(\cdot, \cdot)$. In the data stream framework considered here, $\mathbf{W}_t$ is called a window and N_t is its width, which is obviously supposed to be an integer. If n windows are considered, then the sample size is $N^{(n)} := \sum_{t=1}^{n} N_t$. Additionally, when the windows $\mathbf{W}_t$'s have widths $N_t = 1$ for all $t = 1, \ldots, n$, the sample size is equal to the number of windows

n and the windows $\mathbf{W}_1, \ldots, \mathbf{W}_n$ play the role of the observations. The main objective of this paper is to provide an estimator of the regression function $r(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ and its derivatives that can be easily updated when a new set of observations $\mathbf{W}_t$ is available while retaining nice theoretical properties with respect to natural competitors.

2.2 Multivariate regression estimation

Given $\mathbf{x} \in \mathbb{R}^d$ and $p \in \mathbb{N}$, if we assume that the regression function has derivatives of total order $p+1$ at $\mathbf{x}$, then the multivariate Taylor formula provides an approximation of $r(\mathbf{X})$ by a multivariate polynomial of total order p as:

$$r(\mathbf{X}) = r(\mathbf{x}) + \sum_{\{\mathbf{k} \in \mathbb{N}^d \,:\, 1 \le |\mathbf{k}| \le p\}} \frac{1}{\mathbf{k}!} \frac{\partial^{|\mathbf{k}|} r}{\partial \mathbf{x}^{\mathbf{k}}}(\mathbf{x}) (\mathbf{X} - \mathbf{x})^{\mathbf{k}} + \sum_{|\mathbf{k}| = p} \varepsilon_k (\mathbf{X} - \mathbf{x}) (\mathbf{X} - \mathbf{x})^{\mathbf{k}} \quad (1)$$

and $\lim_{\mathbf{X} \to \mathbf{x}} \varepsilon_k(\mathbf{X} - \mathbf{x}) = 0$, where, throughout the paper, for any d-tuples $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{N}^d$ and $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, we note that

$$|\mathbf{k}| = \sum_{i=1}^d k_i; \quad \mathbf{k}! = \prod_{i=1}^d k_i! \quad \text{and} \quad \mathbf{x}^{\mathbf{k}} = \prod_{i=1}^d x_i^{k_i}.$$

Taking into account the observations available in $\mathbf{W}_1, \ldots, \mathbf{W}_n$, we can derive the locally weighted least squares estimators of the quantities

$$\beta_{\mathbf{k}} = \frac{1}{\mathbf{k}!} \frac{\partial^{|\mathbf{k}|} r}{\partial \mathbf{x}^{\mathbf{k}}}(\mathbf{x}), \quad 0 \le |\mathbf{k}| \le p$$

by minimizing the objective function

$$WSS(\beta) = \sum_{t=1}^n \sum_{j=1}^{N_t} \left\{ Y_{tj} - \sum_{\{\mathbf{k} \in \mathbb{N}^d \,:\, 0 \le |\mathbf{k}| \le p\}} \beta_{\mathbf{k}} (\mathbf{X}_{tj} - \mathbf{x})^{\mathbf{k}} \right\}^2 \omega_{tj}^{(n)}(\mathbf{x}), \quad (2)$$

where the weights $\omega_{tj}^{(n)}$ are defined by

$$\omega_{tj}^{(n)}(\mathbf{x}) = \frac{1}{N^{(n)} h_t^d} K\left(\frac{\mathbf{X}_{tj} - \mathbf{x}}{h_t}\right), \quad (3)$$

where K is a kernel function, h_t is a bandwidth parameter, and $\beta = (\beta_{\mathbf{k}}, \ 0 \le |\mathbf{k}| \le p)^{\mathsf{T}}$. For $u = 0, \ldots, p$, let

$$L_u = \left\{ \mathbf{k} \in \mathbb{N}^d, \ |\mathbf{k}| = u \right\} \quad \text{and} \quad q = \sum_{u=0}^p \binom{u+d-1}{d-1}.$$

Define L as the set of q d-tuples obtained by rearranging the elements of the sets $L_0, \ldots, L_p$ with respect to the lexicographic order and concatenating them as a triangular array. Let g represent a continuous bijective function such that:

$$g : L \longrightarrow \{0, \ldots, q-1\}, \qquad \mathbf{k} \longmapsto i \tag{4}$$

where i denotes the index of the d-tuples $\mathbf{k}$ in the set L. For example, for simplicity, consider the basic bivariate case ($d = 2$). Then, we have

$$L = \left\{ \begin{array}{l} (0, 0), \\ (0, 1), \ (1, 0), \\ (0, 2), \ (1, 1), \quad (2, 0), \\ (0, 3), \ (1, 2), \quad (2, 1), \quad (3, 0), \\ \vdots \\ (0, p), \ (1, p-1), \ (2, p-2), \ (3, p-3), \ \ldots, (p, 0) \end{array} \right\},$$

and the function g can be defined as

$$g(k_1, k_2) = \begin{cases} g(0, 0) = 0, \\ g(k_1 - 1, 0) + k_1 + 1 & \text{if } k_2 = 0 \text{ and } k_1 \geq 1, \\ g(0, k_2 - 1) + k_2 & \text{if } k_1 = 0 \text{ and } k_2 \geq 1, \\ g(k_1 - 1, k_2 - 1) + 2(k_1 + k_2) & \text{else.} \end{cases}$$

Because in (4), g is a bijective function, g^{-1} exists, and we note that $g^{-1}(i) = [i]$ for any $i = 0, \ldots, q-1$.

According to the above notation, for any $t \in \{1, \ldots, n\}$, one can define the matrices

$$\beta = \begin{pmatrix} \beta_{[0]} \\ \vdots \\ \beta_{[q-1]} \end{pmatrix} \quad \text{and} \quad \mathcal{X}_t = \begin{pmatrix} 1 \ (\mathbf{X}_{t1} - \mathbf{x})^{[1]} \ \ldots \ (\mathbf{X}_{t1} - \mathbf{x})^{[q-1]} \\ 1 \ (\mathbf{X}_{t2} - \mathbf{x})^{[1]} \ \ldots \ (\mathbf{X}_{t2} - \mathbf{x})^{[q-1]} \\ \vdots \\ 1 \ (\mathbf{X}_{tN_t} - \mathbf{x})^{[1]} \ \ldots \ (\mathbf{X}_{tN_t} - \mathbf{x})^{[q-1]} \end{pmatrix} := \begin{pmatrix} \mathcal{X}_{t1}^{\mathsf{T}} \\ \mathcal{X}_{t2}^{\mathsf{T}} \\ \vdots \\ \mathcal{X}_{tN_t}^{\mathsf{T}} \end{pmatrix}.$$

Finally, set

$$\mathcal{Y}_t = \left(Y_{t1}, \ldots, Y_{tN_t} \right)^{\mathsf{T}} \quad \text{and} \quad \Omega_t^{(n)} = \text{diag}\left(\omega_{t1}^{(n)}, \ldots, \omega_{tN_t}^{(n)} \right), \quad t = 1, \ldots, n.$$

Then, the derivative of (2) with respect to β is simply the empirical counterpart of

$$2 \sum_{t=1}^{n} N_t \mathbb{E} \left(\omega_{t1}^{(n)} \mathcal{X}_{t1} \mathcal{X}_{t1}^{\mathsf{T}} \beta - m(Y_{t1}) \omega_{t1}^{(n)} \mathcal{X}_{t1} \right) =: 2F(\beta).$$

By solving the equation $F(\beta) = 0$, we obtain the locally weighted estimator of β, based on $\mathbf{W}_1, \ldots, \mathbf{W}_n$, and then, we can deduce the estimator of the regression function and its derivative as

$$\frac{\partial^{|\mathbf{k}|} \widehat{r}}{\partial \mathbf{x}^{\mathbf{k}}}(\mathbf{x}, n) = \mathbf{k}! \widehat{\beta}^{(n)}_{[g(\mathbf{k})]}, \quad \text{for } 0 \le |\mathbf{k}| \le p,$$

where $\widehat{\beta}^{(n)}_{[g(\mathbf{k})]}$ denotes the $\mathbf{k}$-th component of $\widehat{\beta}_n$. It follows that the problem of estimating the regression and its derivative can be viewed as a particular case of the more general issue addressed by stochastic approximation methods because β is readily the unique root of the function F. Therefore, β can be estimated recursively using the multivariate Newton–Raphson procedure (see Ruppert 1985), as follows. Let $\widehat{\beta}_n = (\widehat{\beta}^{(n)}_{[0]}, \cdots, \widehat{\beta}^{(n)}_{[q-1]})^{\mathsf{T}}$ be an estimator of β based on $\mathbf{W}_1, \ldots, \mathbf{W}_n$. Using the fact that $\dfrac{\partial F}{\partial \beta}(\beta) = \sum_{t=1}^{n} N_t \mathbb{E}(\omega^{(n)}_{t1} \mathcal{X}_{t1} \mathcal{X}^{\mathsf{T}}_{t1})$, the multivariate recursive procedure to estimate β is defined as

$$\widehat{\beta}_n = \widehat{\beta}_{n-1} - D_n \widehat{F}_n \left(\widehat{\beta}_{n-1} \right), \tag{5}$$

where D_n is an estimate of the matrix $[\frac{\partial F}{\partial \beta}(\beta)]^{-1}$ based on $\mathbf{W}_1, \ldots, \mathbf{W}_n$ and $\widehat{F}_n(\cdot)$ is an estimator of $F(\cdot)$ based on the sub-sample $\mathbf{W}_n$ only, that is, the observations received at time n. Easy computations show that the empirical counterparts of $[\frac{\partial F}{\partial \beta}(\beta)]^{-1}$ and $F(\beta)$ are respectively defined by

$$D_n = \left(\sum_{t=1}^{n} \mathcal{X}^{\mathsf{T}}_t \Omega^{(n)}_t \mathcal{X}_t \right)^{-1} \quad \text{and}$$

$$\widehat{F}_n \left(\widehat{\beta}_{n-1} \right) = \mathcal{X}^{\mathsf{T}}_n \Omega^{(n)}_n \mathcal{X}_n \left[\widehat{\beta}_{n-1} - \left(\mathcal{X}^{\mathsf{T}}_n \Omega^{(n)}_n \mathcal{X}_n \right)^{-1} \mathcal{X}^{\mathsf{T}}_n \Omega^{(n)}_n \mathcal{Y}_n \right],$$

which together with (5) indicate that

$$\widehat{\beta}_n = (I_q - \Gamma_n)\widehat{\beta}_{n-1} + \Gamma_n \tilde{\beta}_n, \tag{6}$$

where

$$\Gamma_n = \left(\sum_{t=1}^{n} \mathcal{X}^{\mathsf{T}}_t \Omega^{(n)}_t \mathcal{X}_t \right)^{-1} \mathcal{X}^{\mathsf{T}}_n \Omega^{(n)}_n \mathcal{X}_n, \quad \tilde{\beta}_n = \left(\mathcal{X}^{\mathsf{T}}_n \Omega^{(n)}_n \mathcal{X}_n \right)^{-1} \mathcal{X}^{\mathsf{T}}_n \Omega^{(n)}_n \mathcal{Y}_n$$

and I_q is the unit matrix of size q.

The vector $\tilde{\beta}_n$ is simply the weighted least squares estimator of β based on sub-sample $\mathbf{W}_n := \{(\mathbf{X}_{n1}, Y_{n1}), \ldots, (\mathbf{X}_{nN_n}, Y_{nN_n})\}$. The expression (6) bears a resemblance in its structure to the exponential smoothing scheme, except for the fact that in (6), the smoothing parameter is a matrix. Additionally, this relation can be understood as a multivariate Robbins–Monro recursivity (see Robbins and Monro 1951), with a

step size in a matrix form. Note that easy computations show that the estimator $\widehat{\beta}_n$ appearing in (6) has the form

$$\widehat{\beta}_n = \left(\sum_{t=1}^{n} \mathcal{X}_t^{\mathsf{T}} \Omega_t^{(n)} \mathcal{X}_t \right)^{-1} \sum_{t=1}^{n} \mathcal{X}_t^{\mathsf{T}} \Omega_t^{(n)} \mathcal{Y}_t =: S_n^{-1} T_n. \tag{7}$$

Finally, Eq. (6) indicates how to update the recursive estimator from its immediate past when a new stream $\mathbf{W}_{n+1}$ is recorded. However, the usefulness of (6) presents some drawbacks because it is necessary to recalculate the inverse of the matrix S_n, appearing in the expression of Γ_n. To overcome this issue, by setting $V_t = N^{(n)} \Omega_t^{(n)}$, S_n can be reformulated as $S_n = \dfrac{1}{N^{(n)}} \sum_{t=1}^{n} \mathcal{X}_t^{\mathsf{T}} V_t \mathcal{X}_t$. Using the relation

$$S_{n+1} = \left[1 - \frac{N_{n+1}}{N^{(n+1)}} \right] S_n + \frac{1}{N^{(n+1)}} \mathcal{X}_{n+1}^{\mathsf{T}} V_{n+1} \mathcal{X}_{n+1}$$

and the Woodbury formula (see Woodbury 1950), we found that

$$S_{n+1}^{-1} = \left(1 + \frac{N_{n+1}}{N^{(n)}} \right) \left[S_n^{-1} - \frac{1}{N^{(n)}} S_n^{-1} \mathcal{X}_{n+1}^{\mathsf{T}} V_{n+1}^{1/2} C^{-1} V_{n+1}^{1/2} \mathcal{X}_{n+1} S_n^{-1} \right],$$

where

$$C = I_{N_{n+1}} + \frac{1}{N^{(n)}} V_{n+1}^{1/2} \mathcal{X}_{n+1} S_n^{-1} \mathcal{X}_{n+1}^{\mathsf{T}} V_{n+1}^{1/2}$$

is the so-called capacitance matrix. Finally, note that when $p = 0$, the solution estimator obtained in the optimization problem in (2) coincides with the recursive version of the Nadaraya–Watson kernel estimator defined by

$$\widehat{\beta}_0(\mathbf{x}, n) = \frac{\sum_{t=1}^{n} \sum_{j=1}^{N_t} \frac{Y_t}{h_t^d} K \left(\frac{\mathbf{X}_{tj} - \mathbf{x}}{h_t} \right)}{\sum_{t=1}^{n} \sum_{j=1}^{N_t} \frac{1}{h_t^d} K \left(\frac{\mathbf{X}_{tj} - \mathbf{x}}{h_t} \right)}. \tag{8}$$

3 Asymptotics with respect to the number of streams

For our asymptotic analysis, we will present our results in a more general context by considering the weakly dependent sequence of stream data $\{\mathbf{W}_t, \ t \in \mathbb{N}\}$. The weak dependence is more widely applicable than many existing dependence measures, such as mixing, because it covers a large class of processes. For instance, mixing is considered to be useful for characterizing the dependence between time series data because it is fulfilled for many classes of processes and because it allows deriving the same asymptotic results as in the independent case. However, some classes of processes that are of interest in statistics, such as certain $AR(1)$-process, are not

mixing but fulfill weak dependence conditions (see Nze et al. 2002; Doukhan and Neumann 2008). For an extensive introduction and further information on tools for attacking weakly dependent processes, we refer the readers to Rio (2000); Dedecker et al. (2007). Given a function $\Psi : \mathbb{N}^2 \times \mathbb{R}_+^2 \to \mathbb{R}_+$, we extend the notion of Ψ-weak dependence into the streaming context. First, let us consider for any integers u, v and k the set

$$
\Gamma(u, v, k) = \left\{
\begin{array}{l}
(\mathbf{s}, \mathbf{t}, \mathbf{j}, \mathbf{k}) \in \mathbb{N}^u \times \mathbb{N}^v \times \prod_{p=1}^{u} \left\{1, \ldots, N_{s_p}\right\} \times \prod_{q=1}^{v} \left\{1, \ldots, N_{t_q}\right\}, \\
\text{such that } s_1 < \cdots < s_u \leq s_u + k \leq t_1 < \cdots < t_v
\end{array}
\right\}.
$$

Stream data $\{\mathbf{W}_t, \ t \in \mathbb{N}\}$ are said to be Ψ-weakly dependent if there exists a decreasing sequence $(\epsilon_k)_{k \in \mathbb{N}}$ tending to 0 as $k \to \infty$ for which

$$
\left| \text{Cov}\left[g_0\left(W_{s_1 j_1}, \ldots, W_{s_u j_u}\right), \ g_1\left(W_{t_1 k_1}, \ldots, W_{t_v k_v}\right) \right] \right| \leq \Psi\left(u, v, \mathcal{L}(g_0), \mathcal{L}(g_1)\right)\epsilon_k,
\tag{9}
$$

for any $(s_1, \ldots, s_u, t_1, \ldots, t_v, j_1, \ldots, j_u, k_1, \ldots, k_v) \in \Gamma(u, v, k)$, and arbitrary measurable functions $g_i : \mathbb{R}^{u(1-i)+vi} \to \mathbb{R}$, $i \in \{0, 1\}$. In (9), $\mathcal{L}(g_i)$ denotes the Lipschitz modulus of continuity of g_i, that is,

$$
\mathcal{L}(g_i) := \sup_{\mathbf{x} \neq \mathbf{y}} \frac{|g_i(\mathbf{x}) - g_i(\mathbf{y})|}{\|\mathbf{x} - \mathbf{y}\|_1} < \infty
\tag{10}
$$

and $W_{tj} = \left(\mathbf{X}_{tj}, Y_{tj}\right)$, $j = 1, \ldots, N_t, t = 1, \ldots, n$. In the following, we will denote by $f_{\mathbf{X}}$ the marginal density of $\mathbf{X}$ and $\sigma_Y^2(\mathbf{x}) = \mathbb{E}([Y - r(\mathbf{x})]^2 | \mathbf{X} = \mathbf{x})$ the conditional variance of Y given $\mathbf{X} = \mathbf{x}$.

3.1 Assumptions and main results

We will assume that the following assumptions hold.

(A1) N_n is a positive sequence such that $N_n \to \kappa$ for some (common) constant $\kappa \geq 1$ as $n \to \infty$.

(A2) The kernel function K is bounded and continuously differentiable, with compact support.

(A3) The bandwidth sequence (h_n) is positive such that
   (i) $h_n \downarrow 0$ and $nh_n^d \to \infty$ as $n \to \infty$.
   (ii) For any $j \leq 4p$, $\frac{1}{n} \sum_{t=1}^{n} \left(\frac{h_t}{h_n}\right)^j \to \theta_j < \infty$ as $n \to \infty$.

(A4) The stream $\{\mathbf{W}_t, \ t \in \mathbb{N}\}$ is Ψ-weakly dependent, with $\Psi(u, v, w, x) = uvwx$.

(A5) For each $(t, s) \in \{1, \ldots, n\}^2$ and $(j, \ell) \in \{1, \ldots, N_t\} \times \{1, \ldots, N_s\}$, the joint probability density function (p.d.f) $f_{(\mathbf{X}_{tj}, \mathbf{X}_{s\ell})}$ of $(\mathbf{X}_{tj}, \mathbf{X}_{s\ell})$ exists and satisfies

$$
\sup_{(t,j) \neq (s,\ell)} \left| f_{(\mathbf{X}_{tj}, \mathbf{X}_{s\ell})}(\mathbf{x}_1, \mathbf{x}_2) - f_{\mathbf{X}}(\mathbf{x}_1) f_{\mathbf{X}}(\mathbf{x}_2) \right| \leq C \quad \text{for some } C > 0 \text{ and for all } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d.
$$

Assumption (A1) is trivially satisfied in the traditional case $N_t = 1$ for all $t = 1, \ldots, n$ such that the framework considered in this paper remains more general. Moreover, if Assumption (A1) holds, then the sequence $\gamma_1, \ldots, \gamma_t$ defined by $\gamma_t := \frac{N_t}{\sum_{s=1}^{t} N_s}$ is such that $\lim_{n \to \infty} \sum_{t=1}^{n} \gamma_t^2 < \infty$ and $\lim_{n \to \infty} \sum_{t=1}^{n} \gamma_t = \infty$. The above condition is a traditional one in the stochastic approximation framework.

Assumptions (A2) and (A3)(i) are usual assumptions for kernel estimation, whereas (A3)(ii) is a technical assumption that arises from the fact that we need to asymptotically control the update of the bandwidth. Moreover, if Assumption (A2) holds, then for any continuous function ϑ, there exist $c_0, c_1, c_2 \in \mathbb{R}$ such that for any $\mathbf{i} \in \mathbb{N}^d$ and $j \in \mathbb{R}$

$$\sup_{\|\mathbf{u}-\mathbf{x}\| \leq c_0 h_t} \left| \mathbf{u}^{\mathbf{i}} K^j(\mathbf{u}) \right| = c_1 \quad \text{and} \quad \sup_{\|\mathbf{u}-\mathbf{x}\| \leq c_0 h_t} |\vartheta(\mathbf{u})| = c_2. \tag{11}$$

Now, before stating the main theoretical results of this paper, let us define

$$s_{n,\mathbf{i}}^{\tau} := \frac{1}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} (\mathbf{X}_{tj} - \mathbf{x})^{\mathbf{i}} \frac{1}{h_t^d} K\left(\frac{\mathbf{X}_{tj} - \mathbf{x}}{h_t}\right) Y_{tj}^{\tau}, \quad \mathbf{i} \in \mathbb{N}^d, \quad \tau \in \{0, 1\}. \tag{12}$$

Then, the matrices S_n and T_n defined in (7) can be rewritten as

$$S_n = \begin{bmatrix} s_{n,[0]+[0]}^0 & s_{n,[0]+[1]}^0 & \cdots & s_{n,[0]+[q-1]}^0 \\ s_{n,[1]+[0]}^0 & s_{n,[1]+[1]}^0 & \cdots & s_{n,[1]+[q-1]}^0 \\ \vdots & \vdots & & \vdots \\ s_{n,[q-1]+[0]}^0 & s_{n,[q-1]+[1]}^0 & \cdots & s_{n,[q-1]+[q-1]}^0 \end{bmatrix} \quad \text{and} \quad T_n = \begin{bmatrix} s_{n,[0]}^1 \\ s_{n,[1]}^1 \\ \vdots \\ s_{n,[q-1]}^1 \end{bmatrix}. \tag{13}$$

In the following result, we provide an asymptotic bias–variance decomposition of (7).

Theorem 3.1 *Assume that Assumptions (A1)–(A3) hold. Then,*

(a) *for every continuity point $\mathbf{x}$ of $f_{\mathbf{X}}$ and for each $\mathbf{i} \in \mathbb{N}^d$ with $0 \leq |\mathbf{i}| \leq 2p$,*

$$\lim_{n \to \infty} h_n^{-|\mathbf{i}|} \mathbb{E}\left[s_{n,\mathbf{i}}^0\right] = \theta_{|\mathbf{i}|} \mu_{\mathbf{i}} f_{\mathbf{X}}(\mathbf{x}), \tag{14}$$

where $\mu_{\mathbf{i}} = \int_{\mathbb{R}^d} \mathbf{u}^{\mathbf{i}} K(\mathbf{u}) d\mathbf{u}$.

(b) *Moreover, if (A4) and (A5) are satisfied and*

$$\lim_{n \to \infty} n^{\alpha} \sum_{k \geq n+1} \epsilon_k = 0 \quad \text{for some } \alpha > 3 + 4|\mathbf{i}|, \tag{15}$$

then

$$\lim_{n \to \infty} n h_n^{d-2|\mathbf{i}|} \operatorname{Var}\left(s_{n,\mathbf{i}}^0\right) = \frac{\theta_{2|\mathbf{i}|-d} \gamma_{2\mathbf{i}}}{\kappa} f_{\mathbf{X}}(\mathbf{x}), \tag{16}$$

where $\gamma_{\mathbf{i}} = \int_{\mathbb{R}^d} \mathbf{u^i} K^2(\mathbf{u}) d\mathbf{u}$.

If we set

$$H_n = \text{diag}\left(1, h_n, h_n^2, \ldots, h_n^{q-1}\right),$$

B and V represent $q \times q$ matrices defined by the entries:

$$b_{ij} = \theta_{[i-1]+[j-1]}^2 \mu_{[i-1]+[j-1]}^2 \quad \text{and}$$
$$v_{ij} = \theta_{2[i-1]+2[j-1]-d} \gamma_{2[i-1]+2[j-1]}, \quad 1 \le i, j \le q;$$

then, Theorem 3.1 can be expressed in the following matrix form.

Corollary 3.1 *Under the assumptions of Theorem 3.1, we obtain*

$$\text{MSE}(S_n) = H_n \left[f_{\mathbf{X}}(\mathbf{x})^2 B + \frac{1}{nh_n^d} \frac{f_{\mathbf{X}}(\mathbf{x})}{\kappa} V \right] H_n + o_p\left(I_q\right), \quad as \to \infty. \quad (17)$$

Note that in (17), the notation $\text{MSE}(S_n)$ refers to a $q \times q$ matrix where the (i, j)-th component is the mean square error of $s_{n,[i-1]+[j-1]}^0$, $1 \le i, j \le q$ (recall that for any $i = 0, \ldots, q-1$, $[i]$ corresponds to the i-th d-tuples in the set L). In other words,

$$H_n^{-1} S_n H_n^{-1} \overset{m.s.}{\longrightarrow} B f_{\mathbf{X}}(\mathbf{x}), \quad (18)$$

where the notation $\overset{m.s.}{\longrightarrow}$ is the convergence in the mean-square sense. Next, to provide the asymptotic properties of T_n, the following additional assumptions and notations are required.

(A7) Given $\left(Y_{tj}, Y_{s\ell}\right)$, $(t, s) \in \left\{1, \ldots, n\right\}^2$ and $(j, \ell) \in \left\{1, \ldots, N_t\right\} \times \left\{1, \ldots, N_s\right\}$, the conditional p.d.f. of $\left(\mathbf{X}_{tj}, \mathbf{X}_{s\ell}\right)$ exists and satisfies

$$f_{(\mathbf{X}_{tj}, \mathbf{X}_{s\ell})|(Y_{tj}, Y_{s\ell})}(\mathbf{u}_1, \mathbf{u}_2 | v_1, v_2) \le C < \infty \quad \text{for any } (\mathbf{u}_1, \mathbf{u}_2) \in \mathbb{R}^d \times \mathbb{R}^d.$$

Moreover, the conditional p.d.f. of $\mathbf{X}$ given Y exists and satisfies $f_{\mathbf{X}|Y}(\mathbf{u}) \le C < \infty$.

(A8) Either

$$\mathbb{E}\left(|Y|^{\delta}\right) < \infty \quad \text{for some } \delta > 2 \quad (19)$$

or there exist $\lambda, \mu > 0$ such that

$$\mathbb{E}[\exp\left(\lambda|Y|^{\mu}\right)] < \infty. \quad (20)$$

(A9) The conditional variance σ_Y^2 is continuous and bounded away form zero at $\mathbf{x}$.
(A10) There exists $\gamma > \frac{2}{\delta-2}$ such that $n^{\gamma} h_n^{dv} \to 0$ as $n \to \infty$ for some $0 < v < 1$.

Assumptions (A7) and (A8) are technical conditions that are imposed to obtain the proofs of our results. Let us consider the centered version of T_n defined by

$$T_n^* = \left(t_{n,[0]}^*, \ldots, t_{n,[q-1]}^*\right)^\top, \tag{21}$$

where

$$t_{n,\mathbf{i}}^* := \frac{1}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} (\mathbf{X}_{tj} - \mathbf{x})^{\mathbf{i}} \frac{1}{h_t^d} K\left(\frac{\mathbf{X}_{tj} - \mathbf{x}}{h_t}\right) \left[Y_{tj} - r(\mathbf{X}_{tj})\right] \quad \text{for } 0 \le |\mathbf{i}| \le p.$$

In the following results, we establish the asymptotic expression of the variance–covariance matrix of T_n^* and its asymptotic normality.

Theorem 3.2 *Assume that (A1)–(A10) hold. If we denote by* $\mathrm{Cov}(T_n^*)$ *the variance–covariance matrix of* T_n^*, *then*

$$\mathrm{Cov}(T_n^*) = \frac{1}{nh_n^d} \frac{\sigma_Y^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\kappa} H_n \widetilde{V} H_n + o_p(I_q), \quad as \to \infty, \tag{22}$$

where H_n *is the diagonal matrix defined in Corollary 3.1 and* $\widetilde{V}$ *is the* $q \times q$ *matrix defined by its* (i, j)*-th component*

$$\tilde{v}_{ij} = \theta_{[i-1]+[j-1]-d} \gamma_{[i-1]+[j-1]}, \quad i, j = 1, \ldots, q.$$

To obtain the asymptotic normality of the estimator, we will assume that the following additional conditions hold:

(A11) The bandwidth sequence (h_n) is such that

    (i) for any sequence of integer numbers (u_n) and (v_n) with $u_n \sim v_n \Rightarrow h_{u_n} \sim h_{v_n}$.

    (ii) there exists $0 < \nu < 2/(5d + 12p + 8)$ such that $n^\nu h_n \to c > 0$ as $n \to \infty$.

Theorem 3.3 *Assume that (A1)–(A11) hold. In addition, suppose that the process* $\{\mathbf{W}_t, \ t \in \mathbb{N}\}$ *is* Ψ*-weakly dependent with*

$$\epsilon_k = O\left(e^{-\rho k}\right) \quad for \ some \ \rho > 4(d + p + 1)/(5d + 12p + 8).$$

Then,

$$\sqrt{nh_n^d} H_n^{-1} T_n^* \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}_q, \frac{\sigma_Y^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\kappa} \widetilde{V}\right), \tag{23}$$

where $\widetilde{V}$ *is given in Theorem 3.2 and* $\mathbf{0}_q$ *is the* q*-dimensional null vector.*

Let us now study the joint asymptotic normality of the estimator $\widehat{\beta}_n$. Because r is $p + 1$ times differentiable, expanding $r(\mathbf{X}_{tj})$ in a Taylor series around $\mathbf{x}$ for $\|\mathbf{X}_{tj} - \mathbf{x}\| \leq h_t$ indicates that for any $t = 1, \ldots, n$ and $j = 1, \ldots, N_t$,

$$
r(\mathbf{X}_{tj}) = \sum_{\{\mathbf{k}\in\mathbb{N}^d \,:\, 0\leq|\mathbf{k}|\leq p\}} \frac{1}{\mathbf{k}!} \frac{\partial^{|\mathbf{k}|} r}{\partial \mathbf{x}^{\mathbf{k}}}(\mathbf{x}) \left(\mathbf{X}_{tj} - \mathbf{x}\right)^{\mathbf{k}}
$$

$$
+ \sum_{\{\mathbf{k}\in\mathbb{N}^d \,:\, |\mathbf{k}|=p+1\}} \frac{1}{\mathbf{k}!} \frac{\partial^{|\mathbf{k}|} r}{\partial \mathbf{x}^{\mathbf{k}}}(\mathbf{x}) \left(\mathbf{X}_{tj} - \mathbf{x}\right)^{\mathbf{k}} + o(h_t^{p+1}).
$$

Arrange the $n_{p+1} := \binom{p+d}{d-1}$ elements of the derivatives $\frac{1}{\mathbf{k}!} \frac{\partial^{|\mathbf{k}|} r}{\partial \mathbf{x}^{\mathbf{k}}}(\mathbf{x})$ such that $|\mathbf{k}| = p + 1$ using the lexicographic order introduced in Sect. 2.2, refer to them as a column vector $b_{p+1}(\mathbf{x})$ and define the matrix A whose (i, j)-th component is

$$
a_{ij} = \theta_{[i-1]+[j-1]} \mu_{[i-1]+[j-1]} \quad \text{with} \quad 1 \leq i \leq q \quad \text{and} \quad q + 1 \leq j \leq q + n_{p+1} + 1.
$$

Then, in the following corollary, we obtain the asymptotic normality of our estimator.

Corollary 3.2 *Under the assumptions of Theorem 3.3 and $h_n = O\left(n^{-1/(d+2p+2)}\right)$, we have*

$$
\sqrt{nh_n^d} \left[H_n \left(\widehat{\beta}_n - \beta\right) - h_n^{p+1} B^{-1} A b_{p+1}(\mathbf{x}) \right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}_q, \frac{\sigma_Y^2(\mathbf{x})}{\kappa_{\mathbf{X}}(\mathbf{x})} B^{-1} \widetilde{V} B^{-1} \right).
$$

In particular, we can deduce from Corollary 3.2 the estimators of individual partial derivatives of the regression function $r(\mathbf{x})$ via the components of the vector $\widehat{\beta}_n$.

Corollary 3.3 *Under the assumptions of Theorem 3.3 and $h_n = O\left(n^{-1/(d+2p+2)}\right)$, for any $i \in \{0, \cdots, q-1\}$ such that $g^{-1}(i) = [i]$, we have*

$$
\sqrt{nh_n^{d+2i}} \left[H_n \left(\widehat{\beta}_{[i]}^{(n)} - \beta_{[i]}\right) - h_n^{p+1-i} \left(B^{-1} A b_{p+1}(\mathbf{x})\right)_i \right]
$$

$$
\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma_Y^2(\mathbf{x})}{\kappa f_{\mathbf{X}}(\mathbf{x})} \left(B^{-1} \widetilde{V} B^{-1}\right)_{i,i} \right),
$$

where $\left(B^{-1} \widetilde{V} B^{-1}\right)_{i,i}$ is the (i, i)-th diagonal element of the matrix $B^{-1} \widetilde{V} B^{-1}$ and $\left(B^{-1} A b_{p+1}(\mathbf{x})\right)_i$ is the ith element of the vector $B^{-1} A b_{p+1}(\mathbf{x})$.

4 Applications

4.1 An algorithm for numerical computations

In this section, we present an algorithm for numerically computing the value of the local polynomial kernel regression.

RLL estimator

1. | *Fix an integer $n_0 \geq 1$ (resp. $n > n_0$) as the starting (resp. the ending) time of the estimation procedure;*
2. | *Choose a tolerance level $\epsilon > 0$ and a kernel K;*
3. | *Initialization: $k \longrightarrow n_0$*
   (a) | *observe the windows $W_1, \ldots, W_k$;*
   (b) | *compute $N^{(k)}$, the total number of observations available at the time k.*
   (c) | *compute the bandwidth h_k;*
   (d) | *for $t = 1, \ldots, k$:*
      i. | *compute the sample size N_t of the sub-sample W_t;*
      ii. | *extract the design matrix $\mathcal{X}_t$ and the response vector $\mathcal{Y}_t$;*
      iii. | *for $j = 1, \ldots N_t$:*
         | *compute the weights $\omega_{tj}^{(k)}(x)$*
         | *end for*
      iv. | *define the diagonal matrix of weights $\Omega_t^{(k)} = \text{diag}\left(\omega_{t1}^{(k)}(x), \ldots, \omega_{tN_t}^{(k)}(x)\right)$;*
      | *end for;*
   (e) | *concatenate the matrices of weights in a quasi-diagonal matrix $\Omega_k^{\dagger} = \text{diag}\left(\Omega_1^{(k)}, \ldots, \Omega_k^{(k)}\right)$;*
   (f) | *define the initial design matrix and response vector:*

$$\mathcal{X}_k^{\dagger} = \begin{pmatrix} \mathcal{X}_1 \\ \vdots \\ \mathcal{X}_k \end{pmatrix} \quad and \quad \mathcal{Y}_k^{\dagger} = \begin{pmatrix} \mathcal{Y}_1 \\ \vdots \\ \mathcal{Y}_k \end{pmatrix};$$

   (g) | *compute the matrices $T_k = \mathcal{X}_k^{\dagger \mathsf{T}} \Omega_k^{\dagger} \mathcal{Y}_k^{\dagger}, \quad S_k = \mathcal{X}_k^{\dagger \mathsf{T}} \Omega_k^{\dagger} \mathcal{X}_k^{\dagger} \quad and \quad S_k^{-1}$;*
   (h) | *compute the recursive estimator as $\widehat{\beta}_k = S_k^{-1} T_k$*
   (i) | *save $N_k, \quad h_k, \quad S_k^{-1}, \quad \widehat{\beta}_k$.*
4. | *if $k + 1 < n$:*
   (a) | *$k \longrightarrow k + 1$*
      i. | *observe W_k*
      ii. | *compute the sample size N_k of the sub-sample W_k;*
      iii. | *update the total number of observations: $N^{(k)} = N^{(k-1)} + N_k$.*
      iv. | *extract the current design matrix $\mathcal{X}_k$ and the response vector $\mathcal{Y}_k$;*
      v. | *update the bandwidth: $h_{k-1} \longrightarrow h_k$;*
      vi. | *$t \longrightarrow k$*
         | *repeat steps 3(d)i−3(d)iv to obtain $V_k = N^{(k)} \Omega_k^{(k)}$;*
      vii. | *compute the capacitance matrix*

$$C_k = I_{N_k} + \frac{1}{N^{(k-1)}} V_k^{1/2} \mathcal{X}_k S_{k-1}^{-1} \mathcal{X}_k^{\mathsf{T}} V_k^{1/2},$$

   its inverse C_k^{-1} and the matrix $T_k = \frac{1}{N^{(k)}} \mathcal{X}_k^{\mathsf{T}} V_k \mathcal{X}_k$.

    viii. update S_k^{-1}:

$$S_k^{-1} = \left(1 + \frac{N_k}{N^{(k-1)}}\right)\left[S_{k-1}^{-1} - \frac{1}{N^{(k-1)}}S_{k-1}^{-1}\mathcal{X}_k^{\mathsf{T}}V_k^{1/2}C_k^{-1}V_k^{1/2}\mathcal{X}_kS_{k-1}^{-1}\right],$$

    ix. compute the step-size matrix $\Gamma_k = S_k^{-1}T_k$.
    x. update the local linear estimator $\widehat{\beta}_k = (I_q - \Gamma_k)\widehat{\beta}_{k-1} + \mathcal{X}_k^{\mathsf{T}}\Omega_k^{(k)}\mathcal{Y}_k$
    xi. $\lfloor$save $N_k,\ N^{(k)},\ h_k,\ S_k^{-1},\ \widehat{\beta}_k.$
  (b) if $\|\widehat{\beta}_k - \widehat{\beta}_{k-1}\| > \epsilon$
      repeat 4a
    end if
  end if
end RLL estimator

4.2 Bandwidth selector

The algorithm *RLL estimator* requires an update of the bandwidth at step 4(a)v. As is well known in the nonparametric framework, the estimation of the bandwidth is a crucial issue that notably affects the level of smoothing of the resulting estimator. Large values of the bandwidth provide an estimator with a large bias and a small variability, while values that are too small produce an estimator with a small bias and a large variability. Therefore, one can consider two methods according to algorithms *bandwidth 1* and *bandwidth 2* described below.

  bandwidth 1
    1. At step 3c of *RLL estimator*, compute h_k using the subsample $\mathbf{W}_k$;
    2. At step 4(a)v of *RLL estimator*, update the bandwidth using a recursive estimator given by the convex combination

$$\widehat{h}_k = \left(1 - \frac{1}{k}\right)\widehat{h}_{k-1} + \frac{1}{k}\tilde{h}(\mathbf{W}_k),$$

    where $\tilde{h}(\mathbf{W}_k)$ is the bandwidth selected based on the data available in the window $\mathbf{W}_k$
  end bandwidth 1

  bandwidth 2
    1. At step 3c of *RLL estimator*, compute h_k using the subsample $\mathbf{W}_k$;
    2. At step 4(a)v of *RLL estimator*, update the bandwidth using an estimator $\widehat{h}_k$ based on the data available in the windows $\mathbf{W}_1, \ldots, \mathbf{W}_k$.
  end bandwidth 2

5 Proofs

In all the proofs of our results, C denotes a constant whose value is unimportant and may vary from line to line. Additionally, for any $\mathbf{i} \in \mathbb{N}^d$, $t \in \mathbb{N}$, and $(\mathbf{x}, \mathbf{u}, v) \in$

$\mathbb{R}^{2d} \times \mathbb{R}$, set

$$\phi_{t,\mathbf{i}}(\mathbf{u}) = \frac{1}{h_t^d}\left(\frac{\mathbf{u} - \mathbf{x}}{h_t}\right)^{\mathbf{i}} K\left(\frac{\mathbf{u} - \mathbf{x}}{h_t}\right). \tag{24}$$

5.1 Proof of Theorem 3.1

Noting that $|\mathbf{u}^{\mathbf{i}}| \leq \|\mathbf{u}\|^{|\mathbf{i}|}$ and applying Bochner's Lemma with Assumptions (A2) and (A3)(i), we obtain

$$\lim_{t\to\infty}\mathbb{E}\Big[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\Big] = \lim_{t\to\infty}\frac{1}{h_t^d}\int_{\mathbb{R}^d}\left(\frac{\mathbf{u}-\mathbf{x}}{h_t}\right)^{\mathbf{i}} K\left(\frac{\mathbf{u}-\mathbf{x}}{h_t}\right) f_{\mathbf{X}}(\mathbf{u})d\mathbf{u} = f_{\mathbf{X}}(\mathbf{x})\mu_{\mathbf{i}}, \tag{25}$$

and similarly,

$$\lim_{t\to\infty}\mathbb{E}\Big[h_t^d \phi_{t,\mathbf{i}}^2(\mathbf{X}_{tj})\Big] = f_{\mathbf{X}}(\mathbf{x})\gamma_{2\mathbf{i}}. \tag{26}$$

Then, by (12), we have

$$\lim_{n\to\infty}\mathbb{E}\Big[h_n^{-|\mathbf{i}|}s_{n,\mathbf{i}}^0\Big] = \lim_{n\to\infty}\frac{1}{(N^{(n)}/n)}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{h_t}{h_n}\right)^{|\mathbf{i}|} N_t \mathbb{E}\Big[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\Big],$$

and Assumptions (A1) and (A3)(ii) combined with Toeplitz's and Cesaro's lemmas give (14). For the variance term in (16), define

$$A_{t,s} = h_t^{|\mathbf{i}|}h_s^{|\mathbf{i}|}\operatorname{Cov}\left(\sum_{j=1}^{N_t}\phi_{t,\mathbf{i}}(\mathbf{X}_{tj}), \sum_{\ell=1}^{N_s}\phi_{s,\mathbf{i}}(\mathbf{X}_{t\ell})\right); \tag{27}$$

then, we have

$$\operatorname{Var}\left(h_n^{-|\mathbf{i}|}s_{n,\mathbf{i}}^0\right) = \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{t=1}^{n}A_{t,t} + \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{t\neq s}A_{t,s} := E_1 + E_2. \tag{28}$$

Again, E_1 can be decomposed into variance and covariance terms as

$$\begin{aligned}
E_1 &= \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{t=1}^{n}\sum_{j=1}^{N_t}h_t^{2|\mathbf{i}|}\operatorname{Var}\Big[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\Big] \\
&\quad + \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{t=1}^{n}\sum_{j\neq\ell}h_t^{2|\mathbf{i}|}\operatorname{Cov}\Big[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj}), \phi_{t,\mathbf{i}}(\mathbf{X}_{t\ell})\Big] \\
&:= E_{11} + E_{12}.
\end{aligned}$$

Using (25) and (26), we obtain

$$\lim_{t\to\infty} h_t^d \, \mathrm{Var}\left[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\right] = f_{\mathbf{X}}(\mathbf{x})\gamma_{2\mathbf{i}}.$$

Therefore, similar arguments as those used to prove (14) ensure that

$$E_{11} = \frac{1}{nh_n^d}\frac{\theta_{2|\mathbf{i}|-d}\gamma_{2\mathbf{i}}}{\kappa} f_{\mathbf{X}}(\mathbf{x})\left[1 + o(1)\right], \quad \text{as} \ \ n\to\infty. \tag{29}$$

Turning to E_{12}, by Assumption (A5), we have for any $(t, s) \in \{1, \dots, n\}^2$ and $(j, \ell) \in \{1, \dots, N_t\} \times \{1, \dots, N_s\}$, the following bound of the covariance term.

$$\left|\mathrm{Cov}\left[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj}), \phi_{t,\mathbf{i}}(\mathbf{X}_{s\ell})\right]\right| \leq \frac{1}{h_t^d h_s^d}\int_{\mathbb{R}^{2d}} \left|\left(\frac{\mathbf{u}_1 - \mathbf{x}}{h_t}\right)^{\mathbf{i}}\left(\frac{\mathbf{u}_2 - \mathbf{x}}{h_s}\right)^{\mathbf{i}}\right| K\left(\frac{\mathbf{u}_1 - \mathbf{x}}{h_t}\right) K\left(\frac{\mathbf{u}_2 - \mathbf{x}}{h_s}\right)$$
$$\times \left|f_{(\mathbf{X}_{tj},\mathbf{X}_{s\ell})}(\mathbf{u}_1, \mathbf{u}_2) - f_{\mathbf{X}}(\mathbf{u}_1) f_{\mathbf{X}}(\mathbf{u}_2)\right| d\mathbf{u}_1 d\mathbf{u}_2$$
$$\leq C\left(\int_{\mathbb{R}^d} |\mathbf{u}|^{\mathbf{i}} K(\mathbf{u}) d\mathbf{u}\right)^2. \tag{30}$$

Using (30) with $s = t$, we obtain

$$|E_{12}| \leq C\frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{t=1}^{n} h_t^{2\mathbf{i}} N_t(N_t - 1).$$

Under (A1), similar arguments as in the proof of (14) yield that

$$nh_n^d E_{12} = O\left(h_n^d\right) = o(1). \tag{31}$$

Next, let

$$c_n = \left\lfloor h_n^{-\frac{2+2|\mathbf{i}|+2d}{1+\alpha}}\right\rfloor \quad \text{with} \ \ \alpha > 1 + \frac{2(1 + |\mathbf{i}|)}{d}.$$

We can write

$$E_2 = \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{\substack{0<s-t\leq c_n \\ t<s}} A_{t,s} + \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{\substack{s-t>c_n \\ t<s}} A_{t,s} := E_{21} + E_{22}. \tag{32}$$

For $t < s$, setting $k = s - t$ and using (30), one can write

$$E_{21} \leq C\frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}\sum_{k=1}^{c_n}\sum_{t=1}^{n} h_{k+t}^{|\mathbf{i}|} h_t^{|\mathbf{i}|} \leq C\frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2}c_n\sum_{t=1}^{n} h_t^{2|\mathbf{i}|}.$$

Arguing as in the proof of (14), we may obtain

$$nh_n^d E_{21} = O\left(h_n^d c_n\right) = o(1) \quad \text{as } n \to \infty. \tag{33}$$

For E_{22}, we have

$$E_{22} \leq \frac{h_n^{-2|\mathbf{i}|}}{(N^{(n)})^2} \sum_{k=c_n+1}^{n-1} \sum_{t=1}^{n-k} h_{k+t}^{|\mathbf{i}|} h_t^{|\mathbf{i}|} \sum_{j=1}^{N_{k+t}} \sum_{\ell=1}^{N_t} \left| \text{Cov}\left[\phi_{t,\mathbf{i}}(\mathbf{X}_{(k+t)j}), \phi_{t,\mathbf{i}}(\mathbf{X}_{t\ell})\right] \right|.$$

Because K has compact support and $h_n \downarrow 0$, then using (A1) and (A4), the choice of c_n and (15), we obtain

$$
\begin{aligned}
nh_n^d E_{22} &\leq C \frac{h_n^{d-2|\mathbf{i}|}}{(N^{(n)}/n)^2} \frac{1}{n} \sum_{k=c_n+1}^{n-1} \sum_{p=1}^{n-k} h_{k+p}^{-2(d+1)} \epsilon_k \\
&\leq C \frac{h_n^{-(2+d+2|\mathbf{i}|)}}{(N^{(n)}/n)^2} \sum_{k=c_n+1}^{\infty} \epsilon_k \frac{1}{n} \sum_{p=1}^{n} \left(\frac{h_p}{h_n}\right)^{-2(d+1)} \\
&= O\left(h_n^{-(d+2+2|\mathbf{i}|)} \sum_{k=c_n+1}^{\infty} \epsilon_k\right) = O\left(h_n^{-(2+d+2|\mathbf{i}|)} c_n^{-\alpha}\right) = o(1). \quad (34)
\end{aligned}
$$

Finally, the asymptotic variance in (16) follows from (29), (31), (33) and (34).

5.2 Proofs of Theorems 3.2 and 3.3

Before we present the proofs of Theorems 3.2 and 3.3, we state some auxiliary lemmas. Because some arguments for the proofs are similar to those employed in the proof of Theorem 3.1, we shall only present the main differences. First, for any $\mathbf{i} \in \mathbb{N}^d$, $t \in \mathbb{N}$, $1 \leq j \leq N_t$, and $(\mathbf{x}, \mathbf{u}, v) \in \mathbb{R}^{2d} \times \mathbb{R}$, we set

$$\varphi(\mathbf{u}, v) = v - r(\mathbf{u}), \quad \psi(v) = v \mathbb{1}_{\{|v| > b_n\}} \quad \text{and} \quad U_{tj,\mathbf{i}} = \phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\left[\varphi(\mathbf{X}_{tj}, Y_{tj}) - \psi(Y_{tj})\right]. \tag{35}$$

For each $\mathbf{i} \in \mathbb{N}^d$, define the truncated version of $t_{n,\mathbf{i}}^*$:

$$\tilde{t}_{n,\mathbf{i}} := \frac{1}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} h_t^{\mathbf{i}} \phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\left[Y_{tj} \mathbb{1}_{\{|Y_{tj}| \leq b_n\}} - r(\mathbf{X}_{tj})\right], \tag{36}$$

where $b_n \to \infty$ as $n \to \infty$ and $\phi_{t,\mathbf{i}}$ is defined in (24).

Lemma 5.1 *Let Assumptions (A1)–(A10) hold and* $\mathbf{i}, \mathbf{k} \in \mathbb{N}^d$ *such that* $0 \leq |\mathbf{i}|, |\mathbf{k}| \leq 2p$. *If* $f_{\mathbf{X}}$ *is continuous on* $\mathbf{x}$, *then*

$$\text{Cov}\left(\tilde{t}_{n,\mathbf{i}}, \tilde{t}_{n,\mathbf{k}}\right) = \frac{1}{nh_n^{d-|\mathbf{i}|-|\mathbf{k}|}} \frac{\theta_{|\mathbf{i}|+|\mathbf{k}|-d}\sigma_Y^2(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\gamma_{\mathbf{i}+\mathbf{k}}}{\kappa}[1+o(1)] \qquad (37)$$

and

$$\text{Cov}\left(t_{n,\mathbf{i}}^* - \tilde{t}_{n,\mathbf{i}}, t_{n,\mathbf{k}}^* - \tilde{t}_{n,\mathbf{k}}\right) = o\left(\frac{1}{nh_n^{d-|\mathbf{i}|-|\mathbf{k}|}}\right). \qquad (38)$$

Proof By conditioning on $\mathbf{X}_{tj}$, a simple computation shows that

$$\begin{aligned}
\text{Cov}\left(U_{tj,\mathbf{i}}, U_{tj,\mathbf{k}}\right) &= \mathbb{E}\left[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\phi_{t,\mathbf{k}}(\mathbf{X}_{tj})\left(\varphi^2\left(\mathbf{X}_{tj}, Y_{tj}\right) - 2\varphi\left(\mathbf{X}_{tj}, Y_{tj}\right)\psi\left(Y_{tj}\right) + \psi^2\left(Y_{tj}\right)\right)\right] \\
&\quad - \mathbb{E}\left[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\psi\left(Y_{tj}\right)\right]\mathbb{E}\left[\phi_{t,\mathbf{k}}(\mathbf{X}_{tj})\psi\left(Y_{tj}\right)\right] \\
&:= E_1 - E_2 + E_3 - E_4.
\end{aligned}$$

On the one hand,

Because σ_Y^2 and $f_{\mathbf{X}}$ are continuous, it follows by applying Bochner's Lemma with the help of (A3) and (A9) and the first part of (11) that

$$\begin{aligned}
E_1 &= \int_{\mathbb{R}^{2d}} \phi_{t,\mathbf{i}}(\mathbf{u})\phi_{t,\mathbf{k}}(\mathbf{u})\left[\mathbb{E}\left(Y^2|\mathbf{X}=\mathbf{u}\right) - r^2(\mathbf{u})\right]f_{\mathbf{X}}(\mathbf{u})d\mathbf{u} \\
&= \frac{\sigma_Y^2(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\gamma_{\mathbf{i}+\mathbf{k}}}{h_t^d}(1+\tau_t), \quad \tau_t \to 0 \text{ as } t \to \infty.
\end{aligned} \qquad (39)$$

On the other hand, using again (11), Cauchy–Schwarz and Markov's inequalities, we have

$$\begin{aligned}
|E_2| &\leq \left\{\sup_{\|\mathbf{u}-\mathbf{x}\| \leq c_0 h_t} \mathbb{E}\left[\phi_{t,\mathbf{k}}^2(\mathbf{u})\phi_{t,\mathbf{i}}^2(\mathbf{u})\left(Y_{tj}-r(\mathbf{u})\right)^2 Y_{tj}^2\right]\mathbb{P}\left[|Y_{tj}|>b_n\right]\right\}^{1/2} \\
&\leq c_1^2 h_t^{-2d}\left[\max\left\{\mathbb{E}\left(Y_{tj}^4\right), 2c_2\mathbb{E}|Y_{tj}|^3, c_2^2\mathbb{E}\left(Y_{tj}^2\right)\right\}\mathbb{P}\left(|Y_{tj}|>b_n\right)\right]^{1/2} \\
&\leq Ch_t^{-2d}(\ln n)^{2/\mu}e^{-\frac{\lambda b_n}{2}}.
\end{aligned}$$

Similarly, one can prove that $|E_3| + |E_4| \leq Ch_t^{-2d}(\ln n)^{2/\mu}e^{-\frac{\lambda b_n}{2}}$; it follows that

$$\left|\text{Cov}\left(U_{tj,\mathbf{i}}, U_{tj,\mathbf{k}}\right)\right| = \frac{\sigma_Y^2(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\gamma_{\mathbf{i}+\mathbf{k}}}{h_t^d}(1+\tau_t) + A_{n,t}, \qquad (40)$$

where $A_{n,t} = O\left(\dfrac{C(\ln n)^{2/\mu} e^{-\frac{\lambda b_n}{2}}}{h_t^{2d}}\right)$ and $\tau_t \to 0$ as $t \to \infty$. Next, we write

$$
\begin{aligned}
\mathrm{Cov}\left(\tilde{t}_{n,\mathbf{i}}, \tilde{t}_{n,\mathbf{k}}\right) &= \frac{1}{(N^{(n)})^2} \sum_{t=1}^{n} h_t^{|\mathbf{i}|+|\mathbf{k}|} \left[\sum_{j=1}^{N_t} \mathrm{Cov}\left(U_{tj,\mathbf{i}}, U_{tj,\mathbf{k}}\right) + \sum_{j \neq \ell} \mathrm{Cov}\left(U_{tj,\mathbf{i}}, U_{t\ell,\mathbf{k}}\right)\right] \\
&\quad + \frac{1}{(N^{(n)})^2} \sum_{t \neq s} h_t^{|\mathbf{i}|} h_s^{|\mathbf{k}|} \sum_{j=1}^{N_t} \sum_{\ell=1}^{N_s} \mathrm{Cov}\left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}}\right) \\
&:= F_1 + F_2 + F_3.
\end{aligned}
$$

For F_1, using (40) and applying Toeplitz's and Cesaro's lemmas with the help of Assumption (A1), we obtain

$$
F_1 = \frac{1}{n h_n^{d-|\mathbf{i}|-|\mathbf{k}|}} \theta_{|\mathbf{i}|+|\mathbf{k}|-d} \frac{\sigma_Y^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \gamma_{\mathbf{i}+\mathbf{k}}}{\kappa} + o(1) \quad \text{as } n \to \infty. \tag{41}
$$

Now, let us study F_2 and F_3. To this end, from (35), using (A2), (A7) and Bochner's lemma, we obtain for any $(t, s) \in \{1, \ldots, n\}^2$ and $(j, \ell) \in \{1, \ldots, N_t\} \times \{1, \ldots, N_s\}$,

$$
\begin{aligned}
\left|\mathbb{E}\left(U_{tj,\mathbf{i}} U_{s\ell,\mathbf{k}}\right)\right| &= \left|\int_{\mathbb{R}^{2d}} \left\{\int_{\mathbb{R}^2} \left[v_1 - \psi(v_1) - r(\mathbf{u}_1)\right]\left[v_2 - \psi(v_2) - r(\mathbf{u}_2)\right]\right.\right. \\
&\qquad \times \left. f_{(Y_{tj}, Y_{s\ell})}(v_1, v_2) dv_1 dv_2\right\} \\
&\qquad \times \left. \phi_{t,\mathbf{i}}(\mathbf{u}_1) \phi_{s,\mathbf{k}}(\mathbf{u}_2) f_{(\mathbf{X}_{tj}, \mathbf{X}_{s\ell})|(Y_{tj}, Y_{s\ell})}(\mathbf{u}_1, \mathbf{u}_2) d\mathbf{u}_1 d\mathbf{u}_2\right| \\
&\leq \sup_{\substack{\|\mathbf{u}_1 - \mathbf{x}\| \leq c_0 h_t \\ \|\mathbf{u}_2 - \mathbf{x}\| \leq c_0 h_s}} \left[|b_n| + |r(\mathbf{u}_1)|\right]\left[|b_n| + |r(\mathbf{u}_2)|\right] \\
&\qquad \times \int_{\mathbb{R}^{2d}} \left|\phi_{t,\mathbf{i}}(\mathbf{u}_1) \phi_{s,\mathbf{k}}(\mathbf{u}_2)\right| f_{(\mathbf{X}_{tj}, \mathbf{X}_{s\ell})|(Y_{tj}, Y_{s\ell})}(\mathbf{u}_1, \mathbf{u}_2) d\mathbf{u}_1 d\mathbf{u}_2 \leq C b_n^2,
\end{aligned}
\tag{42}
$$

and similarly,

$$
\left|\mathbb{E}\left(U_{tj,\mathbf{i}}\right)\right| = \left|\int_{\mathbb{R}^{d+1}} \phi_{t,\mathbf{i}}(\mathbf{u}) \left[v - \psi(v) - r(\mathbf{u})\right] f(\mathbf{u}, v) d\mathbf{u} dv\right| \leq C b_n. \tag{43}
$$

We deduce that

$$
\left|\mathrm{Cov}\left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}}\right)\right| \leq C b_n^2, \tag{44}
$$

which together with (A1), Toeplitz's and Cesaro's lemmas and the choice of b_n leads to

$$nh_n^{d-|\mathbf{i}|-|\mathbf{k}|} F_2 \leq \frac{Cb_n^2 h_n^d}{\left(N^{(n)}/n\right)^2} \frac{1}{n} \sum_{t=1}^n \left(\frac{h_t}{h_n}\right)^{|\mathbf{i}|+|\mathbf{k}|} N_t (N_t - 1) = O\left(b_n^2 h_n^d\right) = o(1) \tag{45}$$

as $n \to \infty$. Next, to study the term F_3, we use a truncation device. Let

$$c_n = \left\lfloor h_n^{-\frac{2+|\mathbf{i}|+|\mathbf{k}|+2d}{1+\alpha}} \right\rfloor \quad \text{with} \quad \alpha > \frac{d\nu + 2 + |\mathbf{i}| + |\mathbf{k}|}{d(1-\nu)}, \tag{46}$$

where ν is given in (A10), and write

$$\begin{aligned} F_3 &= \frac{2}{(N^{(n)})^2} \sum_{\substack{0 < s-t \leq c_n \\ t < s}} h_t^{|\mathbf{i}|} h_s^{|\mathbf{k}|} \sum_{j=1}^{N_t} \sum_{\ell=1}^{N_s} \mathrm{Cov}\left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}}\right) \\ &+ \frac{2}{(N^{(n)})^2} \sum_{\substack{s-t > c_n \\ t < s}} h_t^{|\mathbf{i}|} h_s^{|\mathbf{k}|} \sum_{j=1}^{N_t} \sum_{\ell=1}^{N_s} \mathrm{Cov}\left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}}\right) \\ &:= F_{31} + F_{32}. \end{aligned} \tag{47}$$

Now, by (44), using (A1) and (A3) with the choice (46), combined with Cesaro's and Toeplitz's lemmas and the definition of b_n, we have

$$\begin{aligned} nh_n^{d-|\mathbf{i}|-|\mathbf{k}|} |F_{31}| &\leq Cb_n^2 nh_n^{d-|\mathbf{i}|-|\mathbf{k}|} \frac{1}{(N^{(n)})^2} \sum_{\substack{0 < s-t \leq c_n \\ t < s}} h_t^{|\mathbf{i}|} h_s^{|\mathbf{k}|} N_t N_s \\ &\leq Cb_n^2 nh_n^{d-|\mathbf{i}|-|\mathbf{k}|} \frac{1}{(N^{(n)})^2} \sum_{k=1}^{c_n} \sum_{p=1}^n h_{k+p}^{|\mathbf{i}|} h_p^{|\mathbf{k}|} N_{k+p} N_p \\ &= O\left(b_n^2 c_n h_n^d\right) = o(1), \end{aligned}$$

as $n \to \infty$. About F_{32}, we go back to the decomposition (47).

For any $t \in \mathbb{N}$ and $\mathbf{i} \in \mathbb{N}^d$, let $g_{t,\mathbf{i}}$ be the function defined by

$$g_{t,\mathbf{i}}(\mathbf{u}, v) = \phi_{t,\mathbf{i}}(\mathbf{u})(v - r(\mathbf{u})) \mathbb{1}_{\left\{\|\mathbf{u}-\mathbf{x}\| \leq c_0 h_t, \ |v| \leq b_n\right\}}(\mathbf{u}, v),$$

where c_0 is defined in (11). If $\mathbf{u} = (u_1, \ldots, u_d)^{\mathsf{T}}$, then using (A2), we have for any $j = 1, \ldots, d$

$$
\left| \frac{\partial g_{t,\mathbf{i}}}{\partial u_j}(\mathbf{u}, v) \right| \leq C \left(b_n + \sup_{\{\|\mathbf{u}-\mathbf{x}\| \leq c_0 h_t\}} \left| \frac{\partial r}{\partial u_j}(\mathbf{u}) \right| \right) \quad \text{and} \quad \left| \frac{\partial g_{t,\mathbf{i}}(\mathbf{u}, v)}{\partial v} \right| \leq C.
$$

$$(48)$$

Therefore, by Taylor's formula, we obtain

$$
\sup_{(\mathbf{u},v) \neq (\mathbf{u}',v')} \left| g_{t,\mathbf{i}}(\mathbf{u}, v) - g_{t,\mathbf{i}}(\mathbf{u}', v') \right| \leq C b_n h_t^{-(d+1)} \left(\left\| \mathbf{u} - \mathbf{u}' \right\|_1 + h_t |v - v'| \right).
$$

Consequently, for sufficiently large t, the Lipschitz modulus of continuity of $g_{t,\mathbf{i}}$ and $g_{s,\mathbf{k}}$ can be bounded as

$$
\mathcal{L}(g_{t,\mathbf{i}}) \mathcal{L}(g_{s,\mathbf{k}}) \leq C b_n^2 h_t^{-(d+1)} h_s^{-(d+1)},
$$

$$(49)$$

which together with Assumption (A4), the second part of (44), leads to

$$
\left| \operatorname{Cov} \left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}} \right) \right| \leq C b_n^2 h_t^{-(d+1)} h_s^{-(d+1)} \epsilon_{t-s}.
$$

$$(50)$$

Therefore, Assumptions (A1) and (A3)(i), the condition $\lim_{n \to \infty} n^\alpha \sum_{k \geq n+1} \epsilon_k = 0$ combined with the choice (46) and the definition of b_n ensure that

$$
\begin{aligned}
nh_n^{d-|\mathbf{i}|-|\mathbf{k}|} |F_{32}| &\leq \frac{C b_n^2 n h_n^{d-|\mathbf{i}|-|\mathbf{k}|}}{(N^{(n)})^2} \sum_{\substack{s-t > c_n \\ t < s}} h_t^{|\mathbf{i}|-(d+1)} h_s^{|\mathbf{k}|-(d+1)} N_t N_s \epsilon_{t-s} \\
&= O \left(b_n^2 h_n^{-(2+d+|\mathbf{i}|+|\mathbf{k}|)} \sum_{k=c_n+1}^{\infty} \epsilon_k \right) \\
&= O \left(b_n^2 h_n^{-(2+d+|\mathbf{i}|+|\mathbf{k}|)} c_n^{-\alpha} \right) = o(1),
\end{aligned}
$$

as $n \to \infty$. We deduce that

$$
nh_n^{d-|\mathbf{i}|-|\mathbf{k}|} |F_3| = o(1) \quad \text{as } n \to \infty,
$$

$$(51)$$

which proves (37). Regarding (38), using (11), (A8) and the Cauchy–Schwarz inequality, we have

$$
\mathbb{E} \left| \phi_{t,\mathbf{i}}(X_{tj}) \phi_{s,\mathbf{k}}(X_{s\ell}) \psi \left(Y_{tj} \right) \psi \left(Y_{s\ell} \right) \right| \leq \frac{C}{h_s^d h_t^d} \mathbb{E} \left| \psi \left(Y_{tj} \right) \psi \left(Y_{s\ell} \right) \right| \leq \frac{C b_n^{2-\delta}}{h_t^d h_s^d} \mathbb{E} \left| Y \right|^\delta,
$$

and in the same manner,

$$\mathbb{E}\left|\phi_{t,\mathbf{i}}(\mathbf{X}_{tj})\psi\left(Y_{tj}\right)\right|\mathbb{E}\left|\phi_{s,\mathbf{k}}(\mathbf{X}_{s\ell})\psi\left(Y_{s\ell}\right)\right| \leq \frac{Cb_n^{2-\delta}}{h_t^d h_s^d}\mathbb{E}|Y|^\delta.$$

Therefore, Assumption (A1) and Toeplitz and Cesaro's lemmas indicate that

$$nh_n^{d-|\mathbf{i}|-|\mathbf{k}|}\,\mathrm{Cov}\left(t_{n,\mathbf{i}}-\tilde{t}_{n,\mathbf{i}},\,t_{n,\mathbf{k}}-\tilde{t}_{n,\mathbf{k}}\right)=O\left(b_n^{2-\delta}nh_n^{-d}\right),$$

and (38) follows from Assumption (A10) and the choice of b_n, which achieves the proof of Lemma 5.1.

Theorem 3.2 follows now from Lemma 5.1, the Cauchy–Schwarz inequality and the fact that

$$\left|\mathrm{Cov}\left(t_{n\mathbf{i}}^*,t_{n\mathbf{k}}^*\right)\right| \leq \left|\mathrm{Cov}\left(t_{n,\mathbf{i}}^*-\tilde{t}_{n,\mathbf{i}},\,t_{n,\mathbf{k}}^*-\tilde{t}_{n,\mathbf{k}}\right)\right| + \left|\mathrm{Cov}\left(\tilde{t}_{n,\mathbf{i}},\tilde{t}_{n,\mathbf{k}}\right)\right|$$
$$+\sqrt{\mathrm{Var}\left(t_{n,\mathbf{i}}^*-\tilde{t}_{n,\mathbf{i}}\right)\mathrm{Var}\left(\tilde{t}_{n,\mathbf{k}}\right)}+\sqrt{\mathrm{Var}\left(t_{n,\mathbf{k}}^*-\tilde{t}_{n,\mathbf{k}}\right)\mathrm{Var}\left(\tilde{t}_{n,\mathbf{i}}\right)}.$$

Next, Theorem 3.3 is a consequence of Lemmas 5.2–5.4 below and the fact that an arbitrary linear combination of $h_n^{-\mathbf{i}}t_{n,\mathbf{i}}^*$ can be represented as

$$\sum_{0\leq|\mathbf{i}|\leq p}c_{\mathbf{i}}h_n^{-\mathbf{i}}t_{n,\mathbf{i}}^*=A_n+B_n+C_n,$$

with

$$A_n:=\sum_{0\leq|\mathbf{i}|\leq p}\frac{c_{\mathbf{i}}}{h_n^{\mathbf{i}}}\left(\tilde{t}_{n,\mathbf{i}}-\mathbb{E}\left(\tilde{t}_{n,\mathbf{i}}\right)\right),\quad B_n:=\sum_{0\leq|\mathbf{i}|\leq p}\frac{c_{\mathbf{i}}}{h_n^{\mathbf{i}}}\mathbb{E}\left(\tilde{t}_{n,\mathbf{i}}\right)\text{ and}$$
$$C_n:=\sum_{0\leq|\mathbf{i}|\leq p}\frac{c_{\mathbf{i}}}{h_n^{\mathbf{i}}}\left(t_{n,\mathbf{i}}^*-\tilde{t}_{n,\mathbf{i}}\right),$$

where $t_{n,\mathbf{i}}^*$ and $\tilde{t}_{n,\mathbf{i}}$ are defined in (21) and (36), respectively, and $b_n=(\delta\ln n)^{1/\mu}$ for $\delta,\mu>0$.

Lemma 5.2 *Under Assumptions (A1)–(A11), we obtain*

$$\sqrt{nh_n^d}A_n\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma_A^2(\mathbf{x})\right),$$

where

$$\sigma_A^2(\mathbf{x})=\sigma_Y^2(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\kappa^{-1}\sum_{0\leq|\mathbf{i}|,|\mathbf{k}|\leq p}c_{\mathbf{i}}c_{\mathbf{k}}\theta_{|\mathbf{i}|+|\mathbf{k}|-d}\gamma_{\mathbf{i}+\mathbf{k}}.$$

Proof Set

$$W_{nt} = \sqrt{nh_n^d}\left(N^{(n)}\right)^{-1} \sum_{0 \le |\mathbf{i}| \le p} c_{\mathbf{i}} h_n^{-\mathbf{i}} h_t^{\mathbf{i}} \sum_{j=1}^{N_t} U_{tj,\mathbf{i}} \quad \text{and} \quad Z_{nt} = W_{nt} - \mathbb{E}\left(W_{nt}\right).$$

Then, $\sqrt{nh_n^d}\,A_n$ can be rewritten as

$$\sqrt{nh_n^d}\,A_n = \sum_{t=1}^{n} Z_{nt}. \tag{52}$$

Now, let us define the sequences of real numbers $(b_n)_{n\in\mathbb{N}}$, $(p_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ such that

$$b_n \to \infty, \quad b_n^2 h_n^d \to 0, \quad p_n \to \infty, \quad q_n \to \infty, \quad p_n/n \to 0 \quad \text{and} \quad q_n/p_n \to 0. \tag{53}$$

We employ Bernstein's big-block and small-block procedure. Partition the set $\{1, 2, \cdots, n\}$ into $r_n + 1$ subsets with large blocks of size p_n and small blocks of size q_n and set

$$r_n = \left\lfloor \frac{n}{p_n + q_n} \right\rfloor. \tag{54}$$

For any $m = 1, \ldots, r$, let $T_{n,m}$, $T'_{n,m}$ and $T''_{n,m}$ be defined as follows:

$$T_{n,m} = \sum_{t=k_m}^{k_m+p_n-1} Z_{nt}, \quad \text{with } k_m := (m-1)(p_n+q_n)+1, \tag{55}$$

$$T'_{n,m} = \sum_{t=l_m}^{l_m+q_n-1} Z_{nt}, \quad \text{with } l_m := (m-1)(p_n+q_n)+p_n+1, \tag{56}$$

$$T''_{n,m} = \sum_{t=r_n(p_n+q_n)+1}^{n} Z_{nt}. \tag{57}$$

We can write

$$\sqrt{nh_n^d}\,A_n = \sum_{m=1}^{r_n} T_{n,m} + \sum_{m=1}^{r_n} T'_{n,m} + T''_{n,m} := S_{n,1} + S_{n,2} + S_{n,3}. \tag{58}$$

To obtain the asymptotic normality of $\sqrt{nh_n^d}\,A_n$, we have to show that

$$\mathbb{E}\left[S_{n,2}^2\right] \to 0, \quad \mathbb{E}\left[S_{n,3}^2\right] \to 0, \tag{59}$$

$$\left|\mathbb{E}\left[e^{it S_{n,1}}\right] - \prod_{m=1}^{r_n} \mathbb{E}\left[e^{it T_{n,m}}\right]\right| \to 0, \tag{60}$$

$$s_n^2 := \sum_{m=1}^{r_n} \mathbb{E}(T_{n,m}^2) \to \sigma_A^2(\mathbf{x}), \tag{61}$$

and for every $\varepsilon > 0$,

$$\frac{1}{s_n^2} \sum_{m=1}^{r_n} \mathbb{E}\left[T_{n,m}^2 \mathbb{1}_{|T_{n,m}|>\varepsilon s_n^2}\right] \to 0. \tag{62}$$

Relation (59) implies that $S_{n,2}$ and $S_{n,3}$ are asymptotically negligible, (60) shows that the random variables $T_{n,m}$ in $S_{n,1}$ are asymptotically independent, and (61) and (62) are the standard Lindeberg–Feller conditions for the asymptotic normality of $S_{n,1}$ under independence. Let us now prove (59). From (56) and (58), we have

$$\mathbb{E}\left[S_{n,2}^2\right] = \sum_{m=1}^{r_n} \sum_{t=l_m}^{l_m+q_n-1} \mathrm{Var}\,(W_{nt}) + 2\sum_{m=1}^{r_n} \sum_{l_m \leq t < s \leq l_m+q_n-1} \mathrm{Cov}\,(W_{nt}, W_{ns})$$

$$+ 2\sum_{1 \leq m < m' \leq r} \sum_{t=l_m}^{l_m+q_n-1} \sum_{s=l_{m'}}^{l_{m'}+q_n-1} \mathrm{Cov}\,(W_{nt}, W_{ns}) := \Delta_1 + \Delta_2 + \Delta_3.$$

Using (35), we can write

$$\Delta_1 = nh_n^d \left(N^{(n)}\right)^{-2} \sum_{m=1}^{r_n} \sum_{t=l_m}^{l_m+q_n-1} \mathrm{Var}\left\{\sum_{0 \leq |\mathbf{i}| \leq p_n} c_{\mathbf{i}} h_n^{-\mathbf{i}} h_t^{|\mathbf{i}|} \sum_{j=1}^{N_t} U_{tj,\mathbf{i}}\right\}.$$

Note that, by (44), because (h_n) is decreasing, using Assumption (A1), we can bound

$$\mathrm{Var}\,(W_{nt}) = nh_n^d \left(N^{(n)}\right)^{-2} \sum_{0 \leq |\mathbf{i}|,|\mathbf{k}| \leq p} c_{\mathbf{i}} c_{\mathbf{k}} \left(\frac{h_t}{h_n}\right)^{\mathbf{i}+\mathbf{k}} \sum_{j=1}^{N_t} \sum_{\ell=1}^{N_t} \mathrm{Cov}\,(U_{tj,\mathbf{i}}, U_{t\ell,\mathbf{k}})$$

$$\leq Cnh_n^d \left(N^{(n)}\right)^{-2} b_n^2 \sum_{0 \leq |\mathbf{i}|,|\mathbf{k}| \leq p} c_{\mathbf{i}} c_{\mathbf{k}} \left(\frac{h_t}{h_n}\right)^{\mathbf{i}+\mathbf{k}} N_t^2 \leq Cn \left(N^{(n)}\right)^{-2} b_n^2 h_n^{d-2p}.$$

Therefore,

$$\Delta_1 \le C\frac{b_n^2 r_n q_n}{nh_n^{2p-d}} = O\left(\frac{q_n}{p_n h_n^{2p}}\right) \quad \text{as } n \to \infty, \tag{63}$$

and similarly, using the Cauchy–Schwarz inequality, we have

$$\Delta_2 \le C\frac{b_n^2 r_n q_n^2}{nh_n^{2p-d}} = O\left(\frac{q_n^2}{p_n h_n^{2p}}\right) \quad \text{as } n \to \infty. \tag{64}$$

For the last term of $\mathbb{E}[S_{n,2}^2]$, we can write

$$\Delta_3 \le 2\frac{nh_n^d}{\left(N^{(n)}\right)^2} \sum_{1 \le m < m' \le r_n}^{r_n} \sum_{t=l_m}^{l_m+q_n-1} \sum_{s=l_{m'}}^{l_{m'}+q_n-1} \sum_{0 \le |\mathbf{i}|, |\mathbf{k}| \le p} c_{\mathbf{i}} c_{\mathbf{k}} \left(\frac{h_t}{h_n}\right)^{\mathbf{i}} \left(\frac{h_s}{h_n}\right)^{\mathbf{k}}$$
$$\times \sum_{j=1}^{N_t} \sum_{\ell=1}^{N_s} \left|\text{Cov}\left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}}\right)\right|.$$

By the boundedness of the covariance term in (50) and because ϵ_t and (h_n) are decreasing, we obtain

$$\Delta_3 \le C\frac{nb_n^2 h_n^d}{\left(N^{(n)}\right)^2} \sum_{\tau=1}^{r_n-1} \sum_{m=1}^{r_n} \sum_{t=l_m}^{l_m+q_n-1} \sum_{s=l_{m'}}^{l_{m'}+q_n-1} \sum_{0 \le |\mathbf{i}|, |\mathbf{k}| \le p} c_{\mathbf{i}} c_{\mathbf{k}} \left(\frac{h_t}{h_n}\right)^{\mathbf{i}} \left(\frac{h_s}{h_n}\right)^{\mathbf{k}} \frac{\epsilon_{\tau(p_n+q_n)}}{(h_t h_s)^{(d+1)}}$$
$$\le C\frac{nq_n^2 r_n b_n^2}{\left(N^{(n)}\right)^2 h_n^{d+2p+2}} \sum_{\tau=1}^{r_n-1} e^{-\tau q_n} = O\left(\frac{q_n^2}{p_n h_n^{2(d+p+1)}}\right) \quad \text{as } n \to \infty. \tag{65}$$

Therefore, the proof of (59) can be deduced from (63), (64) and (65), and the sequences p_n and q_n are chosen such that

$$\frac{q_n}{p_n h_n^{2p}} \to 0, \quad \frac{q_n^2}{p_n h_n^{2p}} \to 0 \quad \text{and} \quad \frac{q_n^2}{p_n h_n^{2(d+p+1)}} \to 0 \quad \text{as } n \to \infty. \tag{66}$$

Next, consider the contribution of $S_{n,3}$. Setting $M_n = r_n(p_n + q_n)$,

$$\mathbb{E}\left(S_{n,3}^2\right) = \sum_{t=M_n+1}^{n} \text{Var}\left(W_{nt}\right) + 2 \sum_{M_n+1 \le t < s \le n} \text{Cov}\left(W_{nt}, W_{ns}\right) := \Theta_1 + \Theta_2.$$

By (36) and (37) in the second part of Lemma 5.1, we can write

$$nh_n^d \text{Var}\left(\sum_{0 \le |\mathbf{i}| \le p} c_{\mathbf{i}} h_n^{-\mathbf{i}} \tilde{t}_{n,\mathbf{i}}\right) \sim \sum_{t=1}^{n} \text{Var}\left(Z_{nt}\right) = \sum_{t=1}^{M_n} \text{Var}\left(Z_{nt}\right) + \sum_{t=M_n+1}^{n} \text{Var}\left(Z_{nt}\right).$$

Because $M_n \sim n$, Assumption (A11)(i) ensures that $\frac{nh_n^d}{M_n h_{M_n}^d} \to 1$ as $n \to \infty$. It follows that

$$\sum_{t=1}^{M_n} \text{Var}\,(Z_{nt}) \sim \sum_{t=1}^{n} \text{Var}\,(Z_{nt}) = \sigma_A^2(\mathbf{x})\,[1+o(1)] \quad \text{as } n \to \infty.$$

Therefore, $\Theta_1 = o(1)$ as $n \to \infty$. For Θ_2, using (51), it is easy to see that

$$\Theta_2 \le \sum_{0 \le |\mathbf{i}|, |\mathbf{k}| \le p} c_{\mathbf{i}} c_{\mathbf{k}} \left| \frac{nh_n^{d-\mathbf{i}-\mathbf{k}}}{\left(N^{(n)}\right)^2} \sum_{t \ne s} h_t^{\mathbf{i}} h_s^{\mathbf{k}} \sum_{j=1}^{N_t} \sum_{\ell=1}^{N_s} \text{Cov}\,\left(U_{tj,\mathbf{i}}, U_{s\ell,\mathbf{k}}\right) \right|$$
$$= o(1) \quad \text{as } n \to \infty.$$

Consequently, (57) follows. To prove (60), observe that for any $m = 1, \dots, r_n$, if we let

$$\Xi_m = \left(\mathbf{X}_{k_m 1}, \dots, \mathbf{X}_{k_m N_{k_m}}, \dots, \mathbf{X}_{k_m+p-11}, \dots, \mathbf{X}_{k_m+p-1 N_{k_m+p-1}}\right),$$

$$\Upsilon_m = \left(Y_{k_m 1}, \dots, Y_{k_m N_{k_m}}, \dots, Y_{k_m+p-11}, \dots, Y_{k_m+p-1 N_{k_m+p-1}}\right),$$

and $\Lambda_m = \sum_{t=k_m}^{k_m+p_n-1} N_t$, then using (48), it can be shown that there exists a measurable function $G_m : \mathbb{R}^{d\Lambda_m} \times \mathbb{R}^{\Lambda_m} \to \mathbb{R}$ such that

$$\exp\left(it T_{n,r_n}\right) = G_{r_n}\left(\Xi_{r_n}, \Upsilon_{r_n}\right) \quad \text{and} \quad \exp\left(it \sum_{m=1}^{r_n-1} T_{n,m}\right) = \prod_{m=1}^{r_n} G_m\left(\Xi_m, \Upsilon_m\right),$$

where

$$\sup_{1 \le i \le d\Lambda_m} \left\| \frac{\partial G_m}{\partial u_i} \right\|_\infty \le C \frac{|t| b_n \sqrt{nh_n^d}}{N^{(n)}} \sum_{s=k_m}^{k_m+p_n-1} \sum_{0 \le |\mathbf{i}| \le p} c_{\mathbf{i}} c_{\mathbf{k}} \left(\frac{h_s}{h_n}\right)^{\mathbf{i}},$$

which together with (24) and the decrease of (h_n) indicates that

$$\mathcal{L}(G_m) \le C \frac{|t| b_n \sqrt{nh_n^d}\, p_n}{N^{(n)} h_n^{d+p+1}}.$$

Therefore,

$$\left| \text{Cov}\left[\exp\left(it \sum_{m=1}^{r_n-1} T_{n,m}\right), \exp\left(it T_{n,r_n}\right)\right] \right| \le C \frac{t^2 b_n^2 nh_n^d\, p_n^2 r_n \epsilon_{q_n}}{(N^{(n)})^2 h_n^{2(d+p+1)}}.$$

Using (53) and arguing as in Doukhan and Louhichi (2001), we obtain from the last relation

$$\left| \mathbb{E}\left[e^{it \sum_{m=1}^{r_n} T_{n,m}} \right] - \prod_{m=1}^{r_n} \mathbb{E}\left[e^{it T_{n,m}} \right] \right| \leq \frac{t^2 p_n r_n e^{-\rho q_n}}{h_n^{2(d+p+1)}}. \tag{67}$$

Then, the proof of (60) follows by taking $t/\sqrt{r_n p_n}$ rather than t in the above relation as soon as

$$\frac{e^{-\rho q_n}}{h_n^{2(d+p+1)}} \to 0 \quad \text{as } n \to \infty. \tag{68}$$

For the proof of (61), using (58), we can write

$$\operatorname{Var}(S_{n,1}) = \operatorname{Var}\left(\sum_{t=1}^n Z_{nt} \right) - \operatorname{Var}(S_{n,2}) - \operatorname{Var}(S_{n,3})$$
$$- 2\operatorname{Cov}(S_{n,1}, S_{n,2}) - 2\operatorname{Cov}(S_{n,2}, S_{n,3}) - 2\operatorname{Cov}(S_{n,1}, S_{n,3}).$$

Then, Cauchy–Schwarz's inequality combined with (59) and (37) in Lemma 5.1 ensures that

$$\lim_{n\to\infty} \operatorname{Var}(S_{n,1}) = \lim_{n\to\infty} \operatorname{Var}\left(\sum_{t=1}^n Z_{nt} \right) = \sigma_A^2(\mathbf{x}). \tag{69}$$

Now, using the arguments employed before for bounding Δ_3, we find that

$$\sum_{m\neq m'} \operatorname{Cov}(T_{n,m}, T_{n,m'}) \to 0 \quad \text{as } n \to \infty. \tag{70}$$

Then, (61) follows from (69), (70) and the fact that

$$\operatorname{Var}(S_{n,1}) = s_n^2 + 2 \sum_{1\leq m < m' \leq r_n} \operatorname{Cov}(T_{n,m}, T_{n,m'}).$$

For the proof of (62), first note that

$$|T_{n,m}| \leq C \frac{p_n b_n}{h_n^p \sqrt{nh_n^d}}.$$

It follows from Markov's inequality that

$$\frac{1}{s_n^2} \sum_{m=1}^{r_n} \mathbb{E}\left[T_{n,m}^2 \mathbb{1}_{|T_{n,m}| > \varepsilon s_n^2} \right] = O\left(\frac{p_n^2 b_n^2}{nh_n^{d+2p}} \right).$$

Therefore, (62) is satisfied as soon as

$$\lim_{n\to\infty} \frac{p_n^2 b_n^2}{nh_n^{d+2p}} = 0. \tag{71}$$

By Assumption (A11)(ii) and taking $p_n = \lfloor n^a \rfloor$, $q_n = \lfloor \log n \rfloor$ with $a > 2\nu(d + p + 1)$, we obtain (66), (68) and (71).

Lemma 5.3 *Assume that (A1)–(A3) and (A8) hold. Then,*

$$\lim_{n \to \infty} \sqrt{nh_n^d} B_n = 0.$$

Proof Note that by (36),

$$\tilde{t}_{n,\mathbf{i}} := \frac{1}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} h_t^{\mathbf{i}} \phi_{t,\mathbf{i}}(\mathbf{X}_{tj}) \left[Y_{tj} - r(\mathbf{X}_{tj}) - Y_{tj} \mathbb{1}_{\{|Y_{tj}| > b_n\}} \right].$$

Then, by conditioning on $\mathbf{X}_{tj}$, we obtain

$$\left| \mathbb{E}(\tilde{t}_{n,\mathbf{i}}) \right| \le \frac{1}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} h_t^{\mathbf{i}} \left| \mathbb{E}\left[\phi_{t,\mathbf{i}}(\mathbf{X}_{tj}) Y_{tj} \mathbb{1}_{\{|Y_{tj}| > b_n\}} \right] \right|.$$

Cauchy–Schwarz and Markov's inequalities ensure that

$$\left| \mathbb{E}(\tilde{t}_{n,\mathbf{i}}) \right| \le \frac{C}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} h_t^{\mathbf{i}-d} \left\{ \mathbb{E}\left[Y_{tj}^2 \right] \mathbb{P}\left[|Y_{tj}| > b_n \right] \right\}^{1/2}$$

$$\le \frac{C}{N^{(n)}} \sum_{t=1}^{n} \sum_{j=1}^{N_t} h_t^{\mathbf{i}-d} \left\{ \mathbb{E}\left[Y_{tj}^2 \right] \mathbb{E}\left[e^{\lambda |Y_{tj}|^\mu} \right] e^{-\lambda b_n^\mu} \right\}^{1/2}$$

$$\le \frac{Cn^{-\lambda\delta/2}(\ln n)^{1/\mu}}{n} \sum_{t=1}^{n} h_t^{\mathbf{i}-d}.$$

Therefore, using (A3)(ii) with the decrease of the bandwidth h_n, we have

$$\sqrt{nh_n^d} B_n \le C \frac{n^{1-\lambda\delta/2}(\ln n)^{1/\mu}}{\sqrt{nh_n^d}} \sum_{0 \le |\mathbf{i}| \le p} c_{\mathbf{i}} \frac{1}{n} \sum_{t=1}^{n} \left(\frac{h_t}{h_n} \right)^{\mathbf{i}}$$

$$= o(1) \quad \text{as soon as } \delta > 2/\lambda,$$

and Lemma 5.3 follows.

Lemma 5.4 *Under Assumption (A8),*

$$\sqrt{nh_n^d} C_n \to 0 \quad a.s.$$

Proof Observe that for any $\varepsilon > 0$,

$$\mathbb{P}\left[\sqrt{nh_n^d}C_n > \varepsilon\right] \leq \mathbb{P}\left[\bigcup_{t=1}^{n}\{|Y_{tj}| > b_n\}\right] \leq \mathbb{E}\left[e^{\lambda|Y|^\mu}\right]n^{1-\lambda\delta},$$

where the last inequality follows by setting $b_n = (\delta \ln n)^{\frac{1}{\mu}}$, with the help of Markov's inequality. Therefore, Assumption (A8) ensures that for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\sqrt{nh_n^d}C_n > \varepsilon\right] < \infty \quad \text{if } \delta > 2/\lambda,$$

and Lemma 5.4 follows by applying Borel–Cantelli's Lemma. $\qquad\square$

5.3 Proof of Corollary 3.2

Recall that n_{p+1} is the number of derivative elements $\frac{1}{\mathbf{k}!}\frac{\partial^{|\mathbf{k}|}r}{\partial\mathbf{x}^{\mathbf{k}}}(\mathbf{x})$ such that $|\mathbf{k}| = p+1$, and for any $t \in \{1, \ldots, n\}$, define

$$M_t = \left(r(\mathbf{X}_{t1}), \ldots, r(\mathbf{X}_{tN_t})\right)^\mathsf{T}, \quad \beta_n^* = S_n^{-1}\sum_{t=1}^{n}\mathcal{X}_t^\mathsf{T}\Omega_t^{(n)}M_t, \tag{72}$$

$$\Lambda_t = \begin{pmatrix} (\mathbf{X}_{t1} - \mathbf{x})^{[q]} & \cdots & (\mathbf{X}_{t1} - \mathbf{x})^{[q+n_{p+1}]} \\ (\mathbf{X}_{t2} - \mathbf{x})^{[q]} & \cdots & (\mathbf{X}_{t2} - \mathbf{x})^{[q+n_{p+1}]} \\ \vdots & & \vdots \\ \left(\mathbf{X}_{tN_t} - \mathbf{x}\right)^{[q]} & \cdots & \left(\mathbf{X}_{tN_t} - \mathbf{x}\right)^{[q+n_{p+1}]} \end{pmatrix}.$$

Then, we can write for sufficiently large t,

$$M_t = \mathcal{X}_t\beta + \Lambda_t b_{p+1}(\mathbf{x}) + o\left(h_t^{p+1}\mathbb{1}_{N_t}\right), \quad \text{with } \mathbb{1}_{N_t} = (1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^{N_t}.$$

Therefore, setting

$$D_n = \sum_{t=1}^{n}\mathcal{X}_t^\mathsf{T}\Omega_t^{(n)}\Lambda_t,$$

we deduce from (72) and Toeplitz's Lemma that

$$\beta_n^* = \beta + S_n^{-1}\left[D_n b_{p+1}(\mathbf{x}) + o_p\left(H_n\mathbb{1}_q h_n^{p+1-d}\right)\right].$$

Next, a simple computation shows that

$$D_n = H_n A_n h_n^{p+1},$$

where A_n is a $q \times n_{p+1}$ matrix whose (i,j)-th component is $a_{n,[i],[j]} = h_n^{-([i-1]+[j-1])} s_{n,[i-1]+[j-1]}^0$, $1 \le i \le q$ and $q+1 \le j \le q + n_{p+1} + 1$. Similar to (18), the matrix A_n converges in mean square error to $f(\mathbf{x})A$. Then, we can deduce that

$$\begin{aligned}
\beta_n^* &= \beta + S_n^{-1} H_n A_n b_{p+1}(\mathbf{x}) h_n^{p+1} + o_p \left(S_n^{-1} H_n h_n^{p+1} \mathbb{1}_q \right) \\
&= \beta + H_n^{-1} H_n S_n^{-1} H_n A_n b_{p+1}(\mathbf{x}) h_n^{p+1} + o_p \left(H_n^{-1} H_n S_n^{-1} H_n h_n^{p+1} \mathbb{1}_q \right) \\
&= \beta + H_n^{-1} \left[B^{-1} A b_{p+1}(\mathbf{x}) h_n^{p+1} + o_p \left(h_n^{p+1} \mathbb{1}_q \right) \right].
\end{aligned} \tag{73}$$

Because

$$S_n^{-1} T_n^* = \beta_n - \beta_n^*,$$

by (73), we have

$$H_n S_n^{-1} T_n^* = H_n (\beta_n - \beta) - \left[B^{-1} A b_{p+1}(\mathbf{x}) h_n^{p+1} + o_p \left(h_n^{p+1} \mathbb{1}_q \right) \right]. \tag{74}$$

Now, by Slutsky's Theorem and (18), we obtain

$$\sqrt{n h_n^d} H_n S_n^{-1} T_n^* \xrightarrow{\mathcal{D}} \mathcal{N} \left(\mathbf{0}_q, \frac{\sigma_Y^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\kappa} B^{-1} \widetilde{V} B^{-1} \right). \tag{75}$$

Therefore, Corollary 3.2 follows from expressions (74) and (75).

References

Aggarwal CC (2007) Data streams: models and algorithms. Springer, New York

Amiri A (2012) Recursive regression estimators with application to nonparametric prediction. J Nonparametr Stat 24:169–186

Amiri A, Crambes C, Thiam B (2014) Recursive estimation of nonparametric regression with functional covariate. Comput Stat Data Anal 69:154–172

Cao Y, He H, Man H (2012) SOMKE: kernel density estimation over data streams by sequences of selforganizing maps. IEEE Trans Neural Netw Learn Syst 23(8):1254–1268

Dedecker J, Doukhan P, Lang G, Leon JR, Louhichi S, Prieur C (2007) Weak dependence: with examples and applications. Lecture Notes in Statistics. Springer, New York

Domingos P, Hulten G (2003) A general framework for mining massive data stream. J Comput Graph Stat 12(4):945–949

Doukhan P, Louhichi S (2001) Functional estimation of a density under a new weak dependence condition. Scand J Stat 28(2):325–341

Doukhan P, Neumann MH (2008) The notion of ψ-weak dependence and its applications to bootstrapping time series. Probab Surv 5:146–168

Fan J, Gijbels I (1992) Variable bandwidth and local linear regression smoothers. Ann Stat 2:2008–2036

Fan J, Gijbels I (1995) Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. J R Stat Soc B 57(2):371–394

Fan J, Gijbels I (1996) Local polynomial modeling and its applications. Chapman and Hall, London

Fan J, Gijbels I, Hu TC, Huang LS (1996) A study of variable bandwidth selection for local polynomial regression. Stat Sin 6:113–127

Gu J, Li Q, Yang JC (2015) Multivariate local polynomial kernel estimators: leading bias and asymptotic distribution. Econom Rev 34(6–10):979–1010

Hansen B (2008) Uniform convergence rates for kernel estimation with dependent data. Econom Theory 24:726–748

Huang Y, Chen X, Wu WB (2014) Recursive nonparametric estimation for time series. IEEE Trans Inf Theory 60(2):1301–1312

Li J, Zheng M (2009) Robust estimation of multivariate regression model. Stat Papers 50(1):81–100

Liang HY, Baek JI (2016) Asymptotic normality of conditional density estimation with left-truncated and dependent data. Stat Papers 57(1):1–20

Masry E (1996a) Multivariate regression estimation local polynomial fitting for time series. Stoch Process Appl 65:81–101

Masry E (1996b) Multivariate local polynomial regression for time series: uniform strong consistency and rates. J Time Ser Anal 17:571–599

Nze AP, Bühlmann P, Doukhan P (2002) Weak dependence beyond mixing and asymptotic for nonparametric regression. Ann Stat 30(2):397–430

Rio E (2000) Théorie asymptotique des processus aléatoires faiblement dépendants. Springer, Berlin

Robbins R, Monro SA (1951) A stochastic approximation method. Ann Stat 22(3):400–407

Ruppert D (1985) A Newton-Raphson version of the multivariate Robbins-Monro procedure. Ann Stat 13(1):236–245

Ruppert D, Wand P (1994) Multivariate locally weighted least squares regression. Ann Stat 22:1346–1370

Ruppert D, Sheather SJ, Wand P (1995) An effective bandwidth selector for local least squares regression. J Am Stat Assoc 90:1257–1270

Vilar JA, Vilar JM (1998) Recursive estimation of regression functions by local polynomial fitting. Ann Inst Stat Math 50(4):729–754

Vilar JA, Vilar JM (2000) Recursive local polynomial regression under dependence conditions. Test 9(1):209–232

Woodbury MA (1950) Inverting modified matrices. Statistical Research Group, Memorandum Report No. 42. Princeton University, Princeton

Xu M, Ishibuchi H, Gu X, Wang S (2014) Dm-KDE: dynamical kernel density estimation by sequences of KDE estimators with fixed number of components over data streams. Front Comput Sci 8(4):563–580

Zhou A, Cai Z, Wei L, Qian W (2003) M-kernel merging: towards density estimation over data streams. In: Proceedings of the 18th international conference on database systems for advanced applications, pp 285–292