AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 11(6): 15581–15625.
DOI: 10.3934/math.2026641
Received: 23 January 2026
Revised: 07 April 2026
Accepted: 28 April 2026
Published: 03 June 2026

Research article

Local linear regression for functional data under quasi-associated dependence with fixed and *k*NN bandwidths

Wahiba Bouabsa[1,*] **and Sadiah M. Aljeddani**[2]

[1] Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria

[2] Department of Mathematics, Al-Lith University College, Al Lith 21961, Umm Al-Qura University, Makkah, Saudi Arabia

* **Correspondence:** Email: wahiba_bouab@yahoo.fr.

Abstract: Local linear kernel estimation is a fundamental tool in nonparametric regression, renowned for its bias reduction and boundary correction properties. While its asymptotic behavior is well understood for independent data and for strongly mixing processes, it remains largely unexplored under quasi-associated dependence, even in the real-valued regression setting. In this paper, we introduce a functional local linear kernel estimator for regression models with quasi-associated observations. We establish strong consistency in the sense of almost complete convergence and derive convergence rates under mild regularity conditions involving small-ball probabilities and covariance decay. To the best of our knowledge, this paper provides the first theoretical guarantees for functional local linear kernel regression under quasi-associated dependence, covering both fixed and *k*-nearest neighbor (*k*NN) bandwidth selection procedures.

Keywords: functional data; local linear estimation; quasi-association; almost complete convergence; kernel methods; *k*NN method
Mathematics Subject Classification: 62G07, 62G20, 62G35, 62H12

1. Introduction

Nonparametric kernel methods are a cornerstone of modern statistics for learning complex relationships while avoiding restrictive parametric specifications. In regression, the Nadaraya-Watson estimator remains a fundamental tool due to its simplicity and strong asymptotic theory in the independent setting. These methods have been extensively studied in finite-dimensional settings and later extended to functional data analysis, where the covariates are infinite-dimensional objects such as curves, images, or trajectories [1–3].

In nonparametric functional data analysis, kernel regression relies on the notion of local neighborhoods defined through semi-metrics and, more generally, through small-ball probabilities. These quantities act as a measure of local concentration in infinite-dimensional spaces and allow the extension of classical nonparametric arguments to functional covariates. Beyond conditional mean estimation, functional kernel methods have been developed for several other statistical targets, including conditional mode and conditional density estimation for functional data [4, 5]. Classical foundations of functional nonparametric methods can be found in [3].

Despite its popularity, the local-constant Nadaraya-Watson estimator suffers from several well-known limitations. In particular, it is affected by boundary effects and a first-order bias, which may deteriorate finite-sample performance. Local linear kernel estimation overcomes these drawbacks by replacing the local-constant approximation with a first-order Taylor expansion of the regression function. As a consequence, it automatically corrects boundary effects and reduces the bias to second order under mild smoothness assumptions. Local polynomial methods are well established for independent observations [1, 2] and have also been extended to dependent time series under strong mixing conditions [6, 7]. A parallel line of research has emphasized the impact of dependence on nonparametric estimation. While strong mixing is the dominant framework in much of the literature, it may be difficult to verify and can be overly restrictive in practice, especially for spatial and functional data. In this context, quasi-association, introduced by [8], provides an alternative notion of weak dependence based on covariance inequalities. It encompasses associated processes and a broad class of linear and spatial models, while leading to tractable and interpretable dependence measures. General developments on weak dependence and statistical applications can be found in [9, 10].

Kernel regression under quasi-associated dependence has so far been mainly studied for local constant estimators. Several authors established consistency, almost complete convergence, and convergence rates for Nadaraya-Watson-type estimators in both finite-dimensional and functional settings. In particular, nonparametric regression under weak dependence assumptions has been investigated in [11], where asymptotic results were derived for functional regression frameworks.

In parallel, increasing attention has been devoted to k-nearest neighbor (kNN) methods for functional data, which adapt to the local structure of the data and are particularly suitable in high-dimensional settings. Early contributions by [12, 13] introduced kNN regression procedures for functional covariates. Later developments considered conditional density and mode estimation [4, 14], as well as robust and relative-error regression methods [15–17]. More recent studies incorporated additional complexities such as spatial dependence and incomplete-data mechanisms [18]. Despite these advances, local linear kernel regression under quasi-associated dependence has not been theoretically established, even in the scalar regression case. This gap is particularly significant in the functional setting, where bias-reduction techniques such as local linear smoothing are often essential. Extending local linear estimation beyond the Nadaraya-Watson paradigm raises substantial challenges: the estimator involves several interacting weighted sums and a random denominator whose non-degeneracy must be carefully controlled under weak dependence.

In this paper, we develop a unified theoretical framework for kernel regression with quasi-associated observations in a functional single-index setting. We first introduce a functional local linear estimator with a deterministic bandwidth and, second, its kNN counterpart obtained by plugging in a random radius. Under mild regularity conditions involving small-ball probabilities and exponential covariance decay, we establish strong consistency in the sense of almost complete convergence and derive explicit convergence rates. To our knowledge, this provides the first rigorous asymptotic justification of local linear and kNN local linear kernel regression for functional data under quasi-associated dependence.

The remainder of the paper is organized as follows. Section 2 presents the regression model, the single-index projection, and the proposed estimators. Section 3 introduces the dependence framework, the small-ball probability function, and the main assumptions. Section 4 establishes the convergence results and rates for the fixed-bandwidth estimator. Section 5 extends the analysis to the kNN version. Section 6 summarizes the theoretical developments. Simulation results and a real-data application to predicting sunspot are presented in Sections 7 and 8, respectively.

2. Notation and assumptions (fixed bandwidth)

Throughout this section, we fix a point $x \in \mathcal{H}$ and a direction $\theta \in \mathcal{H}$ (the single-index parameter). We work with the projected variables

$$Z_i = \langle \theta, X_i \rangle, \qquad z = \langle \theta, x \rangle, \qquad U_i = Z_i - z = \langle \theta, X_i - x \rangle, \quad i \geq 1.$$

For notational simplicity, we write U for a generic copy of U_1.

2.1. Almost complete convergence $O_{\text{a.co.}}$: definitions and notation

A sequence of random variables $(V_n)_{n\geq 1}$ is said to converge almost completely (a.co.) to 0 if, for every $\varepsilon > 0$, we have

$$\sum_{n\geq 1} \mathbb{P}(|V_n| > \varepsilon) < \infty.$$

By the Borel-Cantelli lemma, this implies that $V_n \to 0$ almost surely. In the classical probability literature, this mode of convergence is often referred to as complete convergence; see, e.g., [19] and the references therein. For clarity, we will use the terminology "almost complete convergence" throughout the paper.

More generally, for a deterministic positive sequence (a_n), we write

$$V_n = O_{\text{a.co.}}(a_n)$$

if a constant $C > 0$ exists such that

$$\sum_{n\geq 1} \mathbb{P}(|V_n| > C\, a_n) < \infty.$$

In particular, this implies that V_n / a_n is eventually bounded almost surely.

2.2. Small-ball functions and basic quantities

For $h > 0$, define the (one-dimensional) small-ball probability around x along θ by

$$G_{x,\theta}(h) := \mathbb{P}(|U| \leq h).$$

In particular, $G_{x,\theta}(h) \downarrow 0$ as $h \downarrow 0$. Small-ball probabilities are standard in functional kernel methods; see, e.g., [3].

Let $K : \mathbb{R} \to \mathbb{R}$ be a kernel and define the weights

$$W_i(h) = K\left(\frac{|U_i|}{h}\right), \qquad i = 1, \dots, n.$$

For later use, we introduce the population quantities

$$\mu_j(h) := \mathbb{E}[W(h)U^j], \quad j = 0, 1, 2, \qquad \eta_j(h) := \mathbb{E}[Y\,W(h)U^j], \quad j = 0, 1,$$

and the empirical sums

$$S_j(x; h) = \sum_{i=1}^{n} W_i(h)\,U_i^j, \quad j = 0, 1, 2, \qquad T_j(x; h) = \sum_{i=1}^{n} Y_i W_i(h)\,U_i^j, \quad j = 0, 1.$$

Remark 2.1. *This inequality relies on standard blocking techniques for weakly dependent sequences.*

Lemma 2.1. *(Block-Bernstein inequality under exponential covariance decay) This result follows from standard blocking arguments combined with exponential inequalities for weakly dependent sequences; see, e.g., [20].*

Let $(\xi_i)_{i\geq 1}$ be a strictly stationary centered sequence such that $|\xi_i| \leq M_n$ almost surely and $\mathrm{Var}(\xi_i) \leq v_n$. Assume that constants $C, a > 0$ exist such that, for all $i < j$, we have

$$|\operatorname{Cov}(\xi_i, \xi_j)| \leq C\,e^{-a(j-i)}.$$

Fix $\gamma \in (0, 1)$ and set $q_n = \lfloor n^\gamma \rfloor$ and $p_n = \lfloor n/q_n \rfloor$. Then, constants $c_1, c_2, c_3 > 0$ (independent of n) exist such that, for all $t > 0$ and all sufficiently large values of n, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \xi_i\right| > t\right) \leq 2\exp\left(-\frac{c_1 t^2}{p_n q_n v_n + M_n t}\right) \;+\; c_2\,p_n^2\,\exp(-c_3 q_n).$$

2.3. Quasi-association and dependence coefficient

The asymptotic analysis is carried out under a quasi-association type dependence assumption. We work with a covariance coefficient $(\lambda_r)_{r\geq 1}$ such that, for any two disjoint finite index sets $I, J \subset \mathbb{N}$ with $\min\{|i - j| : i \in I,\ j \in J\} \geq r$ and any Lipschitz functions $f : (\mathcal{H} \times \mathbb{R})^{|I|} \to \mathbb{R}$ and $g : (\mathcal{H} \times \mathbb{R})^{|J|} \to \mathbb{R}$, one has

$$\left|\operatorname{Cov}(f((X_i, Y_i)_{i\in I}), g((X_j, Y_j)_{j\in J}))\right| \leq \operatorname{Lip}(f)\ \operatorname{Lip}(g)\,\lambda_r. \tag{2.1}$$

This inequality is a standard formulation of quasi-association based on covariance control; see, e.g., [8]. Inequality (2.1) is the only property of quasi-association used in the proofs.

Lipschitz constants. For $(x, y), (x', y') \in \mathcal{H} \times \mathbb{R}$, define the product metric

$$d\big((x, y), (x', y')\big) := \|x - x'\|_{\mathcal{H}} + |y - y'|.$$

For $u = (u_1, \dots, u_m)$ and $v = (v_1, \dots, v_m)$ in $(\mathcal{H} \times \mathbb{R})^m$, set

$$d_m(u, v) := \sum_{\ell=1}^{m} d(u_\ell, v_\ell).$$

For $f : (\mathcal{H} \times \mathbb{R})^m \to \mathbb{R}$, define

$$\operatorname{Lip}(f) := \sup_{u\neq v} \frac{|f(u) - f(v)|}{d_m(u, v)}.$$

Lemma 2.2. *Assume* (H1). *Let* $K : \mathbb{R} \to \mathbb{R}$ *be Lipschitz with a constant* L_K *and supported in* $[-1, 1]$. *For* $h > 0$ *and* $j \in \{0, 1, 2\}$, *define*

$$\psi_{j,h}(u) := K\left(\frac{|u|}{h}\right) u^j, \qquad u \in \mathbb{R}.$$

Then, $\psi_{j,h}$ *is globally Lipschitz on* $\mathbb{R}$, *and a constant* $C > 0$ *(depending only on* $\|K\|_\infty$ *and* L_K*) exists such that*

$$\mathrm{Lip}(\psi_{0,h}) \leq \frac{C}{h}, \qquad \mathrm{Lip}(\psi_{1,h}) \leq C, \qquad \mathrm{Lip}(\psi_{2,h}) \leq C\,h.$$

In particular, for $j \in \{0, 1, 2\}$, $\mathrm{Lip}(\psi_{j,h}) \leq C\,h^{j-1}$.

2.4. Assumptions for local linear estimation under quasi-associated data

(H1) Kernel regularity. The kernel K is bounded, Lipschitz, and compactly supported in $[-1, 1]$. Moreover, K is even ($K(u) = K(-u)$) and non-negative.

(H2) Positivity of the small-ball probability. For the fixed (x, θ) under study, $G_{x,\theta}(h) > 0$ for all sufficiently small $h > 0$.

(H2′) Kernel positivity near 0 **and the local second moment.** The constants $u_0 \in (0, 1]$ and $c_K, c_U > 0$ exist such that $K(u) \geq c_K$ for all $|u| \leq u_0$, and for all sufficiently small value of $h > 0$, we have

$$\mathbb{E}[U^2\,\mathbf{1}_{\{|U| \leq u_0 h\}}] \geq c_U\,h^2\,G_{x,\theta}(h).$$

(H3) Joint small-ball control. The constant $C > 0$ exists such that, for all sufficiently small values of $h > 0$ and all $r \geq 1$, we have

$$\sup_{i \geq 1} \mathbb{P}\left(|U_i| \leq h,\ |U_{i+r}| \leq h\right) \leq C\,G_{x,\theta}(h)^2.$$

(H4) Smoothness along the index. Let $m(t) = \mathbb{E}(Y \mid Z = t)$ denote the regression function of the scalar index. We consider the following two smoothness regimes:

(H4-Nadaraya-Watson(NW)) (*Hölder*) The values $\alpha \in (0, 1]$ and $C > 0$ exist such that, for all t in a neighborhood of z, we have

$$|m(t) - m(z)| \leq C\,|t - z|^\alpha.$$

(H4-local linear (LL)) (*Second order*) The function m is twice continuously differentiable in a neighborhood of z and $\sup_{|t-z| \leq \delta} |m''(t)| < \infty$ for some $\delta > 0$.

(H4-LC) Local centering of the projected covariate. The constant $C > 0$ exists such that, for all sufficiently small values of $h > 0$, we have

$$\left|\mu_1(h)\right| = \left|\mathbb{E}[W(h)\,U]\right| \leq C\,h^2\,G_{x,\theta}(h).$$

(H5) Moment conditions. There is a $\nu > 2$ such that $\mathbb{E}(|Y|^\nu) < \infty$. Moreover, we assume

$$\sup_{i \neq j} \mathbb{E}\left(|Y_i Y_j|\right) < \infty,$$

and (when needed for local linear estimation) $\mathbb{E}(|YU|^\nu) < \infty$.

(H6) Exponential decay of dependence. The coefficient (λ_r) in (2.1) satisfies

$$\lambda_r \leq Ce^{-ar} \quad \text{for some constants } C, a > 0.$$

(H7) Bandwidth regime. Let (h_n) be a deterministic sequence such that $h_n \downarrow 0$ and $G_{x,\theta}(h_n) \downarrow 0$. Assume $\gamma \in (0,1)$ and constants $\xi_1, \xi_2 > 0$ exist such that, for all sufficiently large values of n,

$$\frac{\log n}{n^{1-\gamma-\xi_2}} \leq G_{x,\theta}(h_n) \leq \frac{C}{(\log n)^{1+\xi_1}}. \tag{2.2}$$

In particular, (2.2) implies

$$\frac{n^{1-\gamma} G_{x,\theta}(h_n)}{\log n} \to \infty.$$

Commentary on the assumptions.

- (H1) is standard in local polynomial/kernel regression: Boundedness and compact support control tails, while Lipschitz regularity is convenient for covariance/Lipschitz arguments. Evenness is natural here because the weights depend on $|U_i|$. Such conditions are classical in nonparametric regression; see, e.g., [1, 2].
- (H2) ensures that the local neighborhood around x has a positive probability for small radii, which is essential to guarantee that enough observations fall into the neighborhood used for estimation.
- (H2′) is a technical lower bound ensuring the non-degeneracy of local second moments; it typically holds under mild local conditions (e.g., U has a density continuous and positive at 0 and K is positive near 0). In practice, this condition is satisfied for the most commonly used kernels and smooth distributions.
- (H3) controls joint localization at the lag r and is used to handle covariance terms in the sums of localized variables under dependence. The formulation $\sup_{i\geq 1} \mathbb{P}(|U_i| \leq h, |U_{i+r}| \leq h)$ is consistent with stationarity on $\mathbb{N}$. It can be viewed as a weak dependence condition ensuring that simultaneous small deviations remain sufficiently rare, and is closely related to covariance inequalities used in quasi-associated processes [8].
- (H4) distinguishes the bias regimes for local-constant vs. local-linear smoothing. (H4-LC) is a mild centering condition used to control the deterministic part of S_1 in the local-linear denominator $\Delta_n = S_0 S_2 - S_1^2$. These assumptions are standard smoothness requirements ensuring that the regression function can be locally approximated by low-order polynomials.
- (H5) provides sufficient integrability for moment and covariance bounds. Such moment conditions are classical in nonparametric regression and are typically satisfied in applications with bounded or light-tailed responses.
- (H6) is a convenient short-range dependence assumption; combined with blocking, it typically leads to Bernstein-type inequalities (see, e.g., [20]). Exponential decay is commonly assumed in weak dependence settings and holds for many time series and spatial models; see [10].
- (H7) is an effective sample size condition expressed in small-ball form; it ensures that enough points fall in the local neighborhood to obtain almost complete convergence. This condition plays a role analogous to the bandwidth conditions in classical nonparametric regression.

Remark 2.2. *(Sufficient conditions and examples for* (H3)*) Assumption* (H3) *requires that, for $r \geq 1$, we have*

$$\mathbb{P}\left(|U_1| \leq h,\ |U_{r+1}| \leq h\right) \ \leq\ C\, G_{x,\theta}(h)^2 \qquad \text{for all sufficiently small values of } h > 0,$$

where $G_{x,\theta}(h)$ is the small-ball probability defined above.

Proposition 2.1. *Assume that U has a density f with $f(0) > 0$ and that for each $r \geq 1$, the pair (U_1, U_{r+1}) admits a density f_r such that*

$$\sup_{r\geq 1} \sup_{|u|\leq h_0,\ |v|\leq h_0} \frac{f_r(u, v)}{f(u)\, f(v)} \ <\ \infty$$

for some $h_0 > 0$ (with the convention $0/0 = 0$). Then, (H3) *holds for all sufficiently small values of h.*

3. Main results of the local linear method under quasi-associated data

In this section, we prove almost complete convergence for the fixed-bandwidth kernel estimators introduced in Section 2. All statements are established at a fixed point $x \in \mathcal{H}$ and for a fixed index direction $\theta \in \mathcal{H}$. We set

$$U = \langle \theta, X - x \rangle, \qquad G_{x,\theta}(h) = \mathbb{P}(|U| \leq h), \qquad W(h) = K\left(\frac{|U|}{h}\right).$$

For notational simplicity, we may occasionally write $G(h)$ instead of $G_{x,\theta}(h)$ when no ambiguity arises. Let $\|K\|_\infty = \sup_{u\in\mathbb{R}} |K(u)| < \infty$. Under (H1), the kernel K is bounded, non-negative, and compactly supported in $[-1, 1]$. Throughout this section, we use the assumptions and notation introduced in Section 3.

Definition of the estimator. For $h > 0$, recall that the local linear estimator is defined as the intercept of the weighted least squares fit in the projected covariate U_i:

$$(\hat{a}_n(x; h), \hat{b}_n(x; h)) \in \arg \min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - bU_i)^2 W_i(h).$$

It admits the closed form

$$\widehat{r}_n^{LL}(x; h) = \frac{S_2(x; h)T_0(x; h) - S_1(x; h)T_1(x; h)}{\Delta_n(x; h)}, \tag{3.1}$$

where

$$S_j(x; h) = \sum_{i=1}^{n} W_i(h)U_i^j, \quad j = 0, 1, 2, \qquad T_j(x; h) = \sum_{i=1}^{n} Y_i W_i(h)U_i^j, \quad j = 0, 1,$$

and

$$\Delta_n(x; h) = S_0(x; h)S_2(x; h) - S_1(x; h)^2.$$

Theorem 3.1. *Assume* (H1)–(H3) *and* (H5)–(H7). *Assume the second-order smoothness condition* (H4-LL). *Then, for the deterministic bandwidth sequence (h_n), we have*

$$\widehat{r}_n^{LL}(x; h_n) - r(x) = O(h_n^2) + O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}\, G_{x,\theta}(h_n)}}\right). \tag{3.2}$$

Proof. Fix $x \in \mathcal{H}$ and let (h_n) satisfy (H7). For brevity, write $h = h_n$, $G(h) = G_{x,\theta}(h)$, $U_i = \langle\theta, X_i - x\rangle$, and $W_i(h) = K\Big(\frac{|U_i|}{h}\Big)$.

Recall the decomposition

$$\widehat{r}_n^{LL}(x;h) - r(x) = (\widehat{r}_n^{LL}(x;h) - \widetilde{r}_n^{LL}(x;h)) + (\widetilde{r}_n^{LL}(x;h) - r(x)), \tag{3.3}$$

where $\widetilde{r}_n^{LL}(x;h)$ is the LL estimator with Y_i replaced by $r(X_i)$.

Step 1: Bias term. By Lemma 3.5 (using (H4-LL) and the kernel symmetry in (H1)),

$$\widetilde{r}_n^{LL}(x;h) - r(x) = O(h^2).$$

Step 2: Stochastic term. By Lemma 3.2, we have

$$\sum_{i=1}^n \varepsilon_i W_i(h) = O_{\text{a.co.}}\Big(\sqrt{n^{1-\gamma}G(h)\log n}\Big), \qquad \sum_{i=1}^n \varepsilon_i W_i(h)U_i = O_{\text{a.co.}}\Big(h\sqrt{n^{1-\gamma}G(h)\log n}\Big).$$

Step 3: Control of the numerator. Using the bounds obtained in Step 2 together with Lemma 3.1, we find that

$$S_2\sum_{i=1}^n \varepsilon_i W_i(h) \;-\; S_1\sum_{i=1}^n \varepsilon_i W_i(h)U_i = O_{\text{a.co.}}\Big(nh^2G(h)\sqrt{n^{1-\gamma}G(h)\log n}\Big).$$

Step 4: Non-degeneracy of the denominator. By Lemma 3.4, $c_\Delta > 0$ exists such that

$$\Delta_n \ge c_\Delta\, n^2h^2G(h)^2 \quad \text{for all sufficiently large values of } n\text{, almost surely.}$$

Step 5: Stochastic rate. Combining Steps 3 and 4, we have

$$\widehat{r}_n^{LL}(x;h) - \widetilde{r}_n^{LL}(x;h) = O_{\text{a.co.}}\left(\frac{nh^2G(h)\sqrt{n^{1-\gamma}G(h)\log n}}{n^2h^2G(h)^2}\right) = O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}G(h)}}\right).$$

Finally, insert the bias bound (Step 1) and the stochastic bound (Step 5) into (3.3) to obtain (3.2).

We prove Theorem 3.1 using the lemmas below.

Lemma 3.1. *Assume* (H1)–(H3) *and* (H5)–(H7) *(particularly* (H6)*). Let $h = h_n$. Then, for each $j \in \{0, 1, 2\}$, we have*

$$S_j(x;h_n) - n\,\mu_j(h_n) = O_{\text{a.co.}}\Big(\sqrt{n^{1-\gamma}\,G_{x,\theta}(h_n)\,\log n}\;h_n^j\Big), \tag{3.4}$$

and for each $j \in \{0, 1\}$, we have

$$T_j(x;h_n) - n\,\eta_j(h_n) = O_{\text{a.co.}}\Big(\sqrt{n^{1-\gamma}\,G_{x,\theta}(h_n)\,\log n}\;h_n^j\Big), \tag{3.5}$$

where $\mu_j(h) = \mathbb{E}[W(h)U^j]$ and $\eta_j(h) = \mathbb{E}[Y\,W(h)U^j]$.

Proof. We start with the proof of (3.5). Fix $j \in \{0, 1\}$ and write

$$T_j(x; h) - n\eta_j(h) = \sum_{i=1}^{n} \zeta_{i,j}, \qquad \zeta_{i,j} := Y_i W_i(h) U_i^j - \eta_j(h), \quad \eta_j(h) = \mathbb{E}[Y\, W(h) U^j].$$

The difficulty is that Y_i is not bounded, while Lemma 2.1 requires $|\zeta_{i,j}| \le M_n$ almost surely. We enforce this by a max-control (truncation).

Step 0: Max-control for (Y_i). Assume $\mathbb{E}(|Y|^\nu) < \infty$ for some $\nu > 2$ (assumption (H5)) and set

$$b_n := n^{1/\nu} (\log n)^{2/\nu}, \qquad \Omega_n := \Big\{ \max_{1 \le i \le n} |Y_i| \le b_n \Big\}.$$

By the union bound and Markov's inequality, we have

$$\mathbb{P}(\Omega_n^c) \le \sum_{i=1}^{n} \mathbb{P}(|Y_i| > b_n) \le n\, \frac{\mathbb{E}(|Y|^\nu)}{b_n^\nu} = O\Big(\frac{1}{(\log n)^2} \Big).$$

Hence, $\sum_{n \ge 1} \mathbb{P}(\Omega_n^c) < \infty$, and therefore, Ω_n holds eventually almost surely.

Step 1: Uniform bound on Ω_n. On $\{|U_i| \le h\}$, we have $|U_i|^j \le h^j$ and $|W_i(h)| \le \|K\|_\infty$, and hence on Ω_n, we have

$$|Y_i W_i(h) U_i^j| \le b_n \, \|K\|_\infty \, h^j, \qquad |\eta_j(h)| \le \mathbb{E}(|Y|\, |W(h)|\, |U|^j) \le \|K\|_\infty \, h^j \; \mathbb{E}(|Y|) < \infty.$$

Thus, on Ω_n, we have

$$|\zeta_{i,j}| \le C\, b_n\, h^j =: M_n.$$

Step 2: Variance and covariance inputs (fixed bandwidth). Recall the following:

$$\zeta_{i,j} := Y_i W_i(h) U_i^j - \eta_j(h), \qquad \eta_j(h) = \mathbb{E}[Y\, W(h) U^j].$$

On Ω_n, we have $|Y_i| \le b_n$ and $|W_i(h) U_i^j| \le \|K\|_\infty h^j \mathbf{1}_{\{|U_i| \le h\}}$, and hence, $|\zeta_{i,j}| \le C\, b_n h^j =: M_n$.

Moreover, using the support of K and $\mathbb{E}(|Y|^\nu) < \infty$ in (H5), we have

$$\mathrm{Var}(\zeta_{i,j}) \le \mathbb{E}[Y^2 W(h)^2 U^{2j}] \le C\, h^{2j}\, G(h).$$

For $i \ne \ell$, write $\zeta_{i,j} = Y_i \phi_j(U_i) - \eta_j(h)$ with $\phi_j(u) := K(|u|/h)\, u^j$. Then, by the joint small-ball control (H3) and the Cauchy-Schwarz inequality, we have

$$\Big| \mathbb{E}[Y_i Y_\ell \phi_j(U_i) \phi_j(U_\ell)] \Big| \le C\, h^{2j} \; \mathbb{P}(|U_i| \le h,\ |U_\ell| \le h) \le C\, h^{2j}\, G(h)^2.$$

In addition, ϕ_j is globally Lipschitz with $\mathrm{Lip}(\phi_j) \le C h^{j-1}$, and the map $(X, Y) \mapsto Y\, \phi_j(\langle \theta, X - x \rangle)$ is the Lipschitz control on Ω_n for the quasi-association covariance (QA) bound.

Recall that we work with the truncated variables on $\Omega_n := \{|Y| \le b_n\}$, and we apply the QA covariance inequality to functions of the pair (X, Y) endowed with the product metric

$$d((x, y), (x', y')) := \|x - x'\| + |y - y'|.$$

Fix $j \in \{0, 1, 2\}$ and define

$$g_{j,h}(x, y) := y\, \phi_{j,h}(\langle \theta, x - x_0 \rangle), \qquad \phi_{j,h}(u) := u^j K\Big(\frac{|u|}{h} \Big).$$

On Ω_n, for any $(x,y),(x',y')$, we have

$$\begin{aligned}\left|g_{j,h}(x,y)-g_{j,h}(x',y')\right| &= \left|y\,\phi_{j,h}(\langle\theta,x-x_0\rangle)-y'\,\phi_{j,h}(\langle\theta,x'-x_0\rangle)\right|\\ &\le |y-y'|\,\left|\phi_{j,h}(\langle\theta,x-x_0\rangle)\right|+|y'|\,\left|\phi_{j,h}(\langle\theta,x-x_0\rangle)-\phi_{j,h}(\langle\theta,x'-x_0\rangle)\right|.\end{aligned}$$

Since K is bounded and supported on $[-1,1]$, $\phi_{j,h}$ is bounded, and

$$\|\phi_{j,h}\|_\infty \le C\,h^j.$$

Moreover, $\phi_{j,h}$ is Lipschitz on $\mathbb{R}$, with

$$\mathrm{Lip}(\phi_{j,h}) \le C\,h^{j-1}$$

(using the boundedness of K and K' on $(0,1)$, and the scaling $u\mapsto u/h$). Finally, the projection $x\mapsto\langle\theta,x-x_0\rangle$ is $\|\theta\|$-Lipschitz, so

$$\left|\phi_{j,h}(\langle\theta,x-x_0\rangle)-\phi_{j,h}(\langle\theta,x'-x_0\rangle)\right| \le \mathrm{Lip}(\phi_{j,h})\,\|\theta\|\,\|x-x'\|.$$

Combining these bounds and using $|y'|\le b_n$ on Ω_n, we obtain

$$\left|g_{j,h}(x,y)-g_{j,h}(x',y')\right| \le C\,h^j\,|y-y'|+C\,b_n\,h^{j-1}\,\|\theta\|\,\|x-x'\| \le C\,b_n\,h^{j-1}\,(\|x-x'\|+|y-y'|),$$

where we use $h\le 1$ and $b_n\ge 1$ for large values of n. Hence, we have

$$\mathrm{Lip}(g_{j,h}\mathbf{1}_{\Omega_n}) \;\le\; C\,b_n\,h^{j-1}.$$

Therefore, the QA covariance inequality applies with Lipschitz constant Cb_nh^{j-1} for the truncated map $(X,Y)\mapsto Y\,\phi_{j,h}(\langle\theta,X-x_0\rangle)\mathbf{1}_{\Omega_n}$.

Hence, by quasi-association (2.1), we have

$$\left|\mathrm{Cov}(Y_i\phi_j(U_i),\;Y_\ell\phi_j(U_\ell))\right| \le C\,b_n^2\,h^{2j-2}\,\lambda_{|i-\ell|}.$$

Therefore, for all $i\neq\ell$, we have

$$|\,\mathrm{Cov}(\zeta_{i,j},\zeta_{\ell,j})| \le C\Big(h^{2j}G(h)^2+b_n^2h^{2j-2}\lambda_{|i-\ell|}\Big).$$

Step 3: Bernstein-block inequality on Ω_n. Apply Lemma 2.1 to the centered sequence $(\zeta_{i,j})$, but only on the event Ω_n where boundedness holds, and use

$$t_n := A\,h^j\,\sqrt{n^{1-\gamma}G_{x,\theta}(h)\log n}.$$

Then for all large values of n, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^n\zeta_{i,j}\right|>t_n\right) \le \mathbb{P}(\Omega_n^c)+\mathbb{P}\left(\left|\sum_{i=1}^n\zeta_{i,j}\right|>t_n,\;\Omega_n\right) \le \mathbb{P}(\Omega_n^c)+2\exp(-cA^2\log n)+c'p_n^2e^{-c''q_n}.$$

Since $\sum_n\mathbb{P}(\Omega_n^c)<\infty$ and the other two series are also summable (for A large enough), the Borel-Cantelli lemma yields

$$T_j(x;h)-n\eta_j(h)=O_{\mathrm{a.co.}}\Big(h^j\,\sqrt{n^{1-\gamma}G_{x,\theta}(h)\log n}\Big),$$

which is exactly (3.5).

Now, we prove (3.4).

Fix $j \in \{0, 1, 2\}$ and define the centered variables

$$\xi_{i,j} := W_i(h)U_i^j - \mu_j(h), \qquad \mu_j(h) = \mathbb{E}[W(h)U^j].$$

Then, $\mathbb{E}(\xi_{i,j}) = 0$ and

$$S_j - n\mu_j(h) = \sum_{i=1}^{n} \xi_{i,j}.$$

We now turn to the proof of Eq (3.4).

Step A: Support and uniform bound. Since K is supported in $[-1, 1]$, $W_i(h) = 0$ unless $|U_i| \leq h$. On $\{|U_i| \leq h\}$, we have $|U_i|^j \leq h^j$, so

$$|W_i(h)U_i^j| \leq \|K\|_\infty h^j.$$

Moreover, by boundedness, we have

$$|\mu_j(h)| \leq \mathbb{E}(|W(h)U^j|) \leq \|K\|_\infty h^j G_{x,\theta}(h).$$

Hence, $|\xi_{i,j}| \leq Ch^j$.

Step B: Variance term. Using the support argument,

$$\mathbb{E}(\xi_{i,j}^2) \leq \mathbb{E}\left(W(h)^2 U^{2j}\right) \leq \|K\|_\infty^2 h^{2j} G_{x,\theta}(h),$$

so $\mathrm{Var}(\xi_{i,j}) \leq Ch^{2j} G_{x,\theta}(h)$, and therefore,

$$\sum_{i=1}^{n} \mathrm{Var}(\xi_{i,j}) \leq Cnh^{2j} G_{x,\theta}(h).$$

Step C: Covariance term (joint small-ball + quasi-association). For $i \neq \ell$, write

$$\mathrm{Cov}(\xi_{i,j}, \xi_{\ell,j}) = \mathrm{Cov}\left(W_i(h)U_i^j,\ W_\ell(h)U_\ell^j\right).$$

(a) A crude bound using joint small-ball probabilities. Since $W_i(h) = 0$ unless $|U_i| \leq h$ and $|U_i|^j \leq h^j$ on this event, we have

$$|W_i(h)U_i^j| \leq \|K\|_\infty h^j\, \mathbf{1}_{\{|U_i| \leq h\}}.$$

Hence,

$$\begin{aligned}
\left|\mathrm{Cov}\left(W_i(h)U_i^j,\ W_\ell(h)U_\ell^j\right)\right| &\leq \mathbb{E}\left(|W_i(h)U_i^j|\,|W_\ell(h)U_\ell^j|\right) \\
&\leq \|K\|_\infty^2 h^{2j}\, \mathbb{P}\left(|U_i| \leq h,\ |U_\ell| \leq h\right) \\
&\leq C\, h^{2j}\, G_{x,\theta}(h)^2 \qquad \text{by (H3).}
\end{aligned}$$

(b) A dependence-decay bound using quasi-association. Define $\psi_{j,h}(u) := K(|u|/h)\, u^j$. By Lemma 2.2, $\psi_{j,h}$ is globally Lipschitz and

$$\mathrm{Lip}(\psi_{j,h}) \leq Ch^{j-1}, \qquad j \in \{0, 1, 2\}.$$

Since $\|\theta\| = 1$, the map $X \mapsto \langle \theta, X - x\rangle$ is 1-Lipschitz:

$$|\langle \theta, X - x\rangle - \langle \theta, X' - x\rangle| \le \|X - X'\|.$$

Hence, by composition, the map $X \mapsto \psi_{j,h}(\langle \theta, X - x\rangle)$ is Lipschitz with a constant of most Ch^{j-1}.

Applying the quasi-association inequality (2.1) with $I = \{i\}$ and $J = \{\ell\}$ gives

$$\left|\mathrm{Cov}\left(W_i(h)U_i^j,\ W_\ell(h)U_\ell^j\right)\right| \le C\, h^{2j-2}\, \lambda_{|i-\ell|}.$$

(c) Combine both bounds. We thus have the following for all $i \ne \ell$:

$$\left|\mathrm{Cov}(\xi_{i,j}, \xi_{\ell,j})\right| \le C\Big(h^{2j} G_{x,\theta}(h)^2 + h^{2j-2}\lambda_{|i-\ell|}\Big).$$

Summing over $1 \le i < \ell \le n$ yields

$$\begin{aligned}\sum_{1\le i<\ell\le n}\left|\mathrm{Cov}(\xi_{i,j}, \xi_{\ell,j})\right| &\le C\, n^2 h^{2j} G_{x,\theta}(h)^2 + C\, h^{2j-2}\sum_{r=1}^{n-1}(n-r)\lambda_r\\ &\le C\, n^2 h^{2j} G_{x,\theta}(h)^2 + C\, n\, h^{2j-2}\sum_{r\ge 1}\lambda_r,\end{aligned}$$

and since by (H6), we have $\lambda_r \le Ce^{-ar}$, it follows that $\sum_{r\ge1}\lambda_r < \infty$.

This covariance control is used in the Bernstein blocking argument below.

Step D: Almost complete rate via blocking and exponential inequality + Borel-Cantelli lemma. We apply Lemma 2.1 to the centered sequence $\xi_{i,j} = W_i(h)U_i^j - \mu_j(h)$. By Step 4, $|\xi_{i,j}| \le Ch^j$, and hence, we may take $M_n \asymp h^j$. By Step 5, $\mathrm{Var}(\xi_{i,j}) \le Ch^{2j}G_{x,\theta}(h)$, and hence $v_n \asymp h^{2j}G_{x,\theta}(h)$.

Choose $t_n = A\, h^j \sqrt{n^{1-\gamma}G_{x,\theta}(h)\log n}$. Since $p_n q_n \asymp n$ and $q_n \asymp n^\gamma$, Lemma 2.1 yields

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n \xi_{i,j}\Big| > t_n\Big) \le 2\exp(-cA^2\log n) + c'\, n^{2(1-\gamma)}e^{-c'' n^\gamma} \le 3n^{-cA^2}$$

for all large n. Taking a large enough A makes the series summable, and Borel-Cantelli lemma implies

$$S_j - n\mu_j(h) = O_{\mathrm{a.co.}}\Big(h^j\sqrt{n^{1-\gamma}G_{x,\theta}(h)\log n}\Big),$$

which proves (3.4).

Lemma 3.2. *Assume* (H1)–(H3) *and* (H5)–(H7). *Let $h = h_n$. Then*

$$\sum_{i=1}^n \varepsilon_i W_i(h) = O_{\mathrm{a.co.}}\Big(\sqrt{n^{1-\gamma}G_{x,\theta}(h)\log n}\Big), \qquad \sum_{i=1}^n \varepsilon_i W_i(h) U_i = O_{\mathrm{a.co.}}\Big(h\sqrt{n^{1-\gamma}G_{x,\theta}(h)\log n}\Big).$$

Proof. First note that $\mathbb{E}(|r(X)|^\nu) \le \mathbb{E}\big(\mathbb{E}(|Y|^\nu \mid X)\big) = \mathbb{E}(|Y|^\nu) < \infty$ by Jensen's inequality, and hence $\mathbb{E}(|r(X)|^\nu) < \infty$ under (H5).

Apply Lemma 3.1 to $T_j(x;h) = \sum_{i=1}^n Y_i W_i(h) U_i^j$ and to $\widetilde{T}_j(x;h) = \sum_{i=1}^n r(X_i)W_i(h)U_i^j$ for $j \in \{0,1\}$. Since $Y_i = r(X_i) + \varepsilon_i$ and $\mathbb{E}(\varepsilon \mid X) = 0$, we have

$$T_j(x;h) - \widetilde{T}_j(x;h) = \sum_{i=1}^n \varepsilon_i W_i(h) U_i^j, \qquad \eta_j(h) - \widetilde{\eta}_j(h) = 0.$$

Combining the almost complete convergence bounds for T_j and $\widetilde{T}_j$ given by Lemma 3.1 yields the result.

Lemma 3.3. *Assume* (H1)*,* (H2)*, and* (H2′)*. Then, for all sufficiently small values of* $h > 0$*, we have*

$$\mu_0(h) = \mathbb{E}[W(h)] \asymp G_{x,\theta}(h), \qquad \mu_2(h) = \mathbb{E}[W(h)U^2] \asymp h^2\, G_{x,\theta}(h).$$

Proof. Fix $x \in \mathcal{H}$ and $\theta \in \mathcal{H}$. Then we have the following steps.

Step 1: Order of $\mu_0(h) = \mathbb{E}[W(h)]$**.**

Upper bound. Since K is supported on $[-1, 1]$, we have $W(h) = 0$ whenever $|U| > h$, and hence

$$0 \le W(h) \le \|K\|_\infty\, \mathbf{1}_{\{|U|\le h\}}.$$

Taking expectations gives

$$\mu_0(h) = \mathbb{E}[W(h)] \le \|K\|_\infty\ \mathbb{P}(|U| \le h) = \|K\|_\infty\, G_{x,\theta}(h). \tag{3.6}$$

A key comparison: $G_{x,\theta}(u_0h) \gtrsim G_{x,\theta}(h)$. From (H2′), $u_0 \in (0, 1]$ and $c_U > 0$ exist such that for all small values of $h > 0$, we have

$$\mathbb{E}[U^2\, \mathbf{1}_{\{|U|\le u_0h\}}] \ge c_U\, h^2\, G_{x,\theta}(h). \tag{3.7}$$

On the other hand, on $\{|U| \le u_0h\}$, we have $U^2 \le (u_0h)^2$, and hence

$$\mathbb{E}[U^2\, \mathbf{1}_{\{|U|\le u_0h\}}] \le (u_0h)^2\ \mathbb{P}(|U| \le u_0h) = u_0^2h^2\, G_{x,\theta}(u_0h).$$

Combining this with (3.7) yields

$$G_{x,\theta}(u_0h) \ge \frac{c_U}{u_0^2}\, G_{x,\theta}(h) \qquad \text{for all small values of } h > 0. \tag{3.8}$$

Lower bound. Again by (H2′), $c_K > 0$ exists such that $K(u) \ge c_K$ for all $|u| \le u_0$. Therefore, on $\{|U| \le u_0h\}$, we have $|U|/h \le u_0$, and thus $W(h) \ge c_K$. Consequently, we have

$$\mu_0(h) = \mathbb{E}[W(h)] \ge \mathbb{E}[W(h)\mathbf{1}_{\{|U|\le u_0h\}}] \ge c_K\ \mathbb{P}(|U| \le u_0h) = c_K\, G_{x,\theta}(u_0h).$$

Using (3.8), we obtain for all small values of $h > 0$, we have

$$\mu_0(h) \ge c_K\, G_{x,\theta}(u_0h) \ge c_K\, \frac{c_U}{u_0^2}\, G_{x,\theta}(h). \tag{3.9}$$

Conclusion for $\mu_0(h)$*.* From (3.6) and (3.9), the constants $0 < c_0 \le C_0 < \infty$ exist such that for all small values of $h > 0$, we have

$$c_0\, G_{x,\theta}(h) \le \mu_0(h) \le C_0\, G_{x,\theta}(h),$$

i.e., $\mu_0(h) \asymp G_{x,\theta}(h)$.

Step 2: Order of $\mu_2(h) = \mathbb{E}[W(h)U^2]$**.**

Upper bound. As above, $W(h) = 0$ on $\{|U| > h\}$ and $W(h) \le \|K\|_\infty$ everywhere, so

$$0 \le W(h)U^2 \le \|K\|_\infty\, U^2\, \mathbf{1}_{\{|U|\le h\}} \le \|K\|_\infty\, h^2\, \mathbf{1}_{\{|U|\le h\}}.$$

Taking expectations gives

$$\mu_2(h) = \mathbb{E}[W(h)U^2] \le \|K\|_\infty\, h^2\, G_{x,\theta}(h). \tag{3.10}$$

Lower bound. On $\{|U| \le u_0 h\}$, we have $W(h) \ge c_K$, and hence

$$\mu_2(h) = \mathbb{E}[W(h)U^2] \ge \mathbb{E}[W(h)U^2 \mathbf{1}_{\{|U|\le u_0 h\}}] \ge c_K \ \mathbb{E}[U^2 \mathbf{1}_{\{|U|\le u_0 h\}}].$$

Using (3.7), for all small values of $h > 0$, we have

$$\mu_2(h) \ge c_K\, c_U\, h^2\, G_{x,\theta}(h). \tag{3.11}$$

Conclusion for $\mu_2(h)$. From (3.10) and (3.11), the constants $0 < c_2 \le C_2 < \infty$ exist such that for all small values of $h > 0$, we have

$$c_2\, h^2\, G_{x,\theta}(h) \le \mu_2(h) \le C_2\, h^2\, G_{x,\theta}(h),$$

i.e., $\mu_2(h) \asymp h^2\, G_{x,\theta}(h)$.

This completes the proof.

Remark 3.1. *Assumption* (H2′) *is crucial in two places: It provides a nontrivial lower bound on the local second moment $\mathbb{E}[U^2 \mathbf{1}_{\{|U|\le u_0 h\}}] \gtrsim h^2 G_{x,\theta}(h)$, which yields the minorant $\mu_2(h) \gtrsim h^2 G_{x,\theta}(h)$; and, combined with the trivial inequality $\mathbb{E}[U^2 \mathbf{1}_{\{|U|\le u_0 h\}}] \le u_0^2 h^2 G_{x,\theta}(u_0 h)$, it forces the comparison $G_{x,\theta}(u_0 h) \gtrsim G_{x,\theta}(h)$, which is exactly what makes $\mu_0(h) \asymp G_{x,\theta}(h)$ possible without any regular variation assumption on $G_{x,\theta}$.*

Lemma 3.4. *Assume* (H1)–(H3), (H5)–(H7), *and* (H4-LC). *Moreover, assume the conclusions of Lemma 3.3. Then, $c_\Delta > 0$ exists such that*

$$\sum_{n\ge 1} \mathbb{P}\Big(\Delta_n(x; h_n) \le c_\Delta\, n^2 h_n^2 G_{x,\theta}(h_n)^2\Big) < \infty. \tag{3.12}$$

In particular,

$$\Delta_n(x; h_n) \ge c_\Delta\, n^2 h_n^2 G_{x,\theta}(h_n)^2 \quad \textit{eventually almost surely.} \tag{3.13}$$

Proof. We show that $S_0(h) S_2(h)$ dominates $S_1(h)^2$ almost completely, without using any unproved shortcut such as $|S_1| = O(S_0 h)$.

Step 1: Asymptotic orders of expectations. Recall $\mu_j(h) = \mathbb{E}[U^j K(|U|/h)]$ and $S_j(h) = \sum_{i=1}^n U_i^j K(|U_i|/h)$. We keep $\mu_1(h) = \mathbb{E}[UK(|U|/h)]$ not necessarily equal to 0 and use the local centering condition (H4-LC): A function $\rho(h) \downarrow 0$ exists such that, for a small enough h, we have

$$|\mu_1(h)| \le \rho(h)\, h\, \mu_0(h). \tag{3.14}$$

Moreover, by Lemma 3.3, we have

$$\mu_0(h) \asymp G_{x,\theta}(h), \qquad \mu_2(h) \asymp h^2 G_{x,\theta}(h). \tag{3.15}$$

In particular, (3.15) implies $\mu_0(h)\mu_2(h) \asymp h^2 G_{x,\theta}(h)^2$.

Step 2: Lower bounds for S_0 and S_2. By Lemma 3.1 with $j = 0$ and $j = 2$,

$$S_0(h) = n\mu_0(h) + O_{\text{a.co.}}\Big(\sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n}\Big),$$

$$S_2(h) = n\mu_2(h) + O_{\text{a.co.}}\Big(h^2 \sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n}\Big).$$

Under (H7) (which guarantees $\sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n} = o(nG_{x,\theta}(h))$), the error terms are $o(n\mu_0(h))$ and $o(n\mu_2(h))$. Hence, eventually almost surely

$$S_0(h) \geq \tfrac{1}{2} n\mu_0(h), \qquad S_2(h) \geq \tfrac{1}{2} n\mu_2(h).$$

Therefore, eventually almost surely

$$S_0(h) S_2(h) \geq c\, n^2 \mu_0(h) \mu_2(h) \asymp c\, n^2 h^2 G_{x,\theta}(h)^2, \tag{3.16}$$

for some constant $c > 0$.

Step 3: Upper bound for S_1^2 and comparison with $S_0 S_2$. Decompose

$$S_1(h) = n\mu_1(h) + (S_1(h) - n\mu_1(h)).$$

By Lemma 3.1 with $j = 1$,

$$S_1(h) - n\mu_1(h) = O_{\text{a.co.}}\Big(h \sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n}\Big).$$

Using $(a + b)^2 \leq 2a^2 + 2b^2$, we obtain

$$S_1(h)^2 \leq 2n^2 \mu_1(h)^2 + 2(S_1(h) - n\mu_1(h))^2. \tag{3.17}$$

(i) Deterministic part $n^2\mu_1(h)^2$. From (3.14), we get

$$\frac{\mu_1(h)^2}{\mu_0(h)\mu_2(h)} \leq \rho(h)^2 \frac{h^2 \mu_0(h)^2}{\mu_0(h)\mu_2(h)} = \rho(h)^2 \frac{h^2 \mu_0(h)}{\mu_2(h)}.$$

Using (3.15), we have $\mu_2(h) \asymp h^2 \mu_0(h)$; hence,

$$\frac{\mu_1(h)^2}{\mu_0(h)\mu_2(h)} \leq C \rho(h)^2 \underset{h\downarrow 0}{\longrightarrow} 0. \tag{3.18}$$

Combining (3.18) with (3.16) yields

$$\frac{n^2 \mu_1(h)^2}{S_0(h) S_2(h)} = o(1) \qquad \text{(eventually a.s.)}.$$

(ii) Random part $(S_1 - n\mu_1)^2$. By Lemma 3.1 with $j = 1$, we have

$$(S_1(h) - n\mu_1(h))^2 = O_{\text{a.co.}}\Big(h^2 n^{1-\gamma} G_{x,\theta}(h) \log n\Big).$$

Divide by the lower bound (3.16),

$$\frac{(S_1(h) - n\mu_1(h))^2}{S_0(h) S_2(h)} = O_{\text{a.co.}}\Big(\frac{h^2 n^{1-\gamma} G_{x,\theta}(h) \log n}{n^2 \mu_0(h) \mu_2(h)}\Big).$$

Using $\mu_0(h)\mu_2(h) \asymp h^2 G_{x,\theta}(h)^2$ from (3.15), we get

$$\frac{(S_1(h) - n\mu_1(h))^2}{S_0(h)S_2(h)} = O_{\text{a.co.}}\Big(\frac{\log n}{n^{1+\gamma} G_{x,\theta}(h)}\Big).$$

Since (H7) implies

$$\frac{n^{1+\gamma} G_{x,\theta}(h)}{\log n} \to \infty,$$

it follows that

$$\frac{\log n}{n^{1+\gamma} G_{x,\theta}(h)} \to 0,$$

and therefore,

$$\frac{(S_1(h) - n\mu_1(h))^2}{S_0(h)S_2(h)} = o_{\text{a.co.}}(1).$$

Putting (3.17) together with (i) and (ii), we obtain

$$\frac{S_1(h)^2}{S_0(h)S_2(h)} = o_{\text{a.co.}}(1),$$

and therefore, eventually almost surely

$$\Delta_n(h) = S_0(h)S_2(h) - S_1(h)^2 \geq \tfrac{1}{2} S_0(h)S_2(h) \geq c_\Delta\, n^2 \mu_0(h)\mu_2(h) \asymp c_\Delta\, n^2 h^2 G_{x,\theta}(h)^2,$$

for some $c_\Delta > 0$. This yields (3.12) and (3.13).

Lemma 3.5. *Assume* (H4-LL) *and let* $\widetilde{r}_n^{LL}(x; h_n)$ *be the local linear estimator obtained from* (3.1) *by replacing* Y_i *with* $r(X_i)$*. Then*

$$\widetilde{r}_n^{LL}(x; h_n) - r(x) = O(h_n^2). \tag{3.19}$$

Proof. Let $\widetilde{r}_n^{LL}(x; h)$ denote the LL estimator built from the same weights but with Y_i replaced by $r(X_i) = m(Z_i)$.

Step 1: Second-order expansion. By (H4-LL), m is twice differentiable in a neighborhood of z with bounded m''. For $|U_i| \leq h$, Taylor's formula gives

$$m(z + U_i) = m(z) + m'(z)U_i + \frac{1}{2}m''(z + \tau_i U_i)U_i^2, \qquad \tau_i \in (0, 1),$$

and hence,

$$r(X_i) = r(x) + m'(z)U_i + R_i, \qquad |R_i| \leq CU_i^2 \leq Ch^2 \text{ on } \{|U_i| \leq h\}.$$

Step 2: Cancellation of the first-order term. Plug the decomposition into

$$\widetilde{T}_0 = \sum_{i=1}^{n} r(X_i)\, W_i(h), \qquad \widetilde{T}_1 = \sum_{i=1}^{n} r(X_i)\, W_i(h)\, U_i.$$

We obtain

$$\widetilde{T}_0 = r(x)S_0 + m'(z)S_1 + \sum_{i=1}^{n} R_i W_i(h), \qquad \widetilde{T}_1 = r(x)S_1 + m'(z)S_2 + \sum_{i=1}^{n} R_i W_i(h) U_i.$$

Compute the LL numerator:

$$S_2\widetilde{T}_0 - S_1\widetilde{T}_1 = r(x)(S_0S_2 - S_1^2) + \Big(S_2\sum_{i=1}^{n} R_iW_i(h) - S_1\sum_{i=1}^{n} R_iW_i(h)U_i\Big).$$

The first-order term involving $m'(z)$ cancels exactly, which is the key bias-reduction mechanism of local linear estimation.

Step 3: Bound the remainder. On the support of $W_i(h)$, $|R_i| \le Ch^2$; hence,

$$\Big|\sum_{i=1}^{n} R_iW_i(h)\Big| \le Ch^2S_0, \qquad \Big|\sum_{i=1}^{n} R_iW_i(h)U_i\Big| \le Ch^3S_0.$$

Therefore,

$$\Big|S_2\sum_{i=1}^{n} R_iW_i(h)\Big| \le Ch^2S_0S_2, \qquad \Big|S_1\sum_{i=1}^{n} R_iW_i(h)U_i\Big| \le Ch^3|S_1|S_0.$$

Moreover, with the support of $W_i(h)$, we have $|U_i| \le h$, and hence

$$S_2 = \sum_{i=1}^{n} W_i(h)U_i^2 \le h^2\sum_{i=1}^{n} W_i(h) = h^2S_0.$$

By the Cauchy-Schwarz inequality, we have

$$S_1^2 = \Big(\sum_{i=1}^{n} W_i(h)U_i\Big)^2 \le \Big(\sum_{i=1}^{n} W_i(h)\Big)\Big(\sum_{i=1}^{n} W_i(h)U_i^2\Big) = S_0S_2 \le h^2S_0^2,$$

and therefore, $|S_1| \le hS_0$.

Dividing by Δ_n, and using Lemma 3.4 (so that $\Delta_n \asymp S_0S_2$), we obtain

$$\widetilde{r}_n^{LL}(x;h) - r(x) = O(h^2).$$

This shows that the contribution of the remainder terms is of order $O(h^2)$, completing the proof of (3.19).

4. Main result of kNN LL estimation under quasi-associated data

4.1. Technical lemmas

4.1.1. A general bracketing lemma for random bandwidths

The kNN estimators studied later are obtained by replacing a deterministic bandwidth h with the random radius $H_{n,k}(x)$. A key technical difficulty is that $H_{n,k}(x)$ depends on the full sample and therefore destroys the standard fixed-bandwidth decomposition. To overcome this difficulty, we transfer the almost complete convergence results from deterministic bandwidths to random bandwidths via a bracketing argument, a classical approach in the functional kNN literature.

Let $(A_i, B_i)_{i\ge1}$ be a sequence of observations, where A_i is a covariate and B_i is a response. Fix a target point x and consider a family of non-negative weights

$$L(t,(x,A_i)) \ge 0, \qquad t > 0,$$

such that for each i, the map $t \mapsto L(t,(x,A_i))$ is nonincreasing. Define the ratio functional

$$R_n(t;x) := \frac{\sum_{i=1}^n B_i\, L(t,(x,A_i))}{\sum_{i=1}^n L(t,(x,A_i))} \quad \text{whenever } \sum_{i=1}^n L(t,(x,A_i)) > 0. \tag{4.1}$$

Let $D_n(x)$ be a positive random variable (a random bandwidth). Assume that a deterministic target $R(x)$ and a deterministic rate $U_n \downarrow 0$ exist.

Lemma 4.1. *Assume that deterministic functions $D_n^{\pm}(\beta_n, x)$ and a deterministic sequence $\beta_n \downarrow 0$ exist such that*

(C0) (Max-control for the responses) There is a $\nu > 2$ such that $\mathbb{E}(|B_1|^\nu) < \infty$. Let

$$M_n := n^{1/\nu}(\log n)^{2/\nu}.$$

Moreover, assume that $\beta_n M_n = O(U_n)$, which ensures that the truncation error remains negligible compared with the target rate.

(C1) (Bracketing, a.co.) The events

$$\mathcal{A}_n(x) := \left\{D_n^-(\beta_n,x) \le D_n(x) \le D_n^+(\beta_n,x)\right\}$$

satisfy $\sum_{n\ge 1} \mathbb{P}(\mathcal{A}_n(x)^c) < \infty$.

(C2) (Stability of denominators) There is a constant $c > 0$ such that eventually on $\mathcal{A}_n(x)$

$$\frac{\sum_{i=1}^n L(D_n^-(\beta_n,x),(x,A_i))}{\sum_{i=1}^n L(D_n^+(\beta_n,x),(x,A_i))} \in [1-\beta_n,\ 1+\beta_n] \quad \text{and} \quad \sum_{i=1}^n L(D_n^+(\beta_n,x),(x,A_i)) \ge c. \tag{C2}$$

(C3) (Two fixed-bandwidth controls) Almost completely, we have

$$R_n(D_n^-(\beta_n,x);x) - R(x) = O_{\text{a.co.}}(U_n), \qquad R_n(D_n^+(\beta_n,x);x) - R(x) = O_{\text{a.co.}}(U_n).$$

Then,

$$R_n(D_n(x);x) - R(x) = O_{\text{a.co.}}(U_n).$$

This lemma is the main tool allowing the transfer of the fixed-bandwidth almost complete rates to random bandwidths in the kNN framework.

Proof. Work on the event $\mathcal{A}_n(x)$, which holds eventually almost surely by (C1). By the monotonicity of $t \mapsto L(t,(x,A_i))$, we have the following for all i:

$$L(D_n^+,(x,A_i)) \le L(D_n,(x,A_i)) \le L(D_n^-,(x,A_i)).$$

This implies that the ratio $R_n(D_n;x)$ is sandwiched between $R_n(D_n^-;x)$ and $R_n(D_n^+;x)$ up to a multiplicative distortion controlled by the denominators. Using the bracket inequalities together with (C2), we obtain that $R_n(D_n;x)$ lies between $R_n(D_n^-;x)$ and $R_n(D_n^+;x)$ up to an additive error of order $\beta_n \max_{1\le i\le n} |B_i|$ (eventually on $\mathcal{A}_n(x)$).

To control the maximum, let

$$M_n := n^{1/\nu}(\log n)^{2/\nu}.$$

By the union bound and Markov's inequality, we have

$$\mathbb{P}\Big(\max_{1\le i\le n}|B_i| > M_n\Big) \le n\,\frac{\mathbb{E}(|B_1|^\nu)}{M_n^\nu} = O\left(\frac{1}{(\log n)^2}\right).$$

Hence, $\sum_{n\ge1}\mathbb{P}(\max_{1\le i\le n}|B_i| > M_n) < \infty$, and therefore, $\max_{1\le i\le n}|B_i| \le M_n$ eventually almost surely. Consequently, the bracketing distortion is $O(\beta_n M_n)$ eventually almost surely and, by assumption (C0), we have $\beta_n M_n = O(U_n)$, so this term is absorbed into the rate U_n.

Combining the bracketing control with the fixed-bandwidth bounds in (C3) yields

$$R_n(D_n(x); x) - R(x) = O_{\text{a.co.}}(U_n).$$

Remark 4.1. *In our application to kNN estimators, $D_n(x)$ will be $H_{n,k}(x)$, while $D_n^{\pm}(\beta_n, x)$ will be deterministic radii built from the inverse small-ball function. Condition (C3) follows directly from the fixed-bandwidth local linear consistency results established in the previous section.*

Remark 4.2. *For $E_i \ge 0$ and a kernel K nonincreasing on $[0,1]$, the map $h \mapsto K(E_i/h)$ is nondecreasing in h. This apparent mismatch with the monotonicity assumption in Lemma 4.1 is purely technical and can be handled by a simple reparametrization. Thus, when applying Lemma 4.1, one can either (i) work with an equivalent parametrization $t = 1/h$, so that $t \mapsto K(E_i t)$ is nonincreasing, or (ii) keep h and reverse the monotonicity direction in the bracketing step.*

5. Consistency of the local linear (LL) *k*NN method

This section establishes almost complete convergence for the *k*NN estimators, for the LL method. The key idea is to replace the deterministic bandwidth h_n by the random radius $H_{n,k}(x)$ and to transfer fixed-bandwidth results via general bracketing (Lemma 4.1).

Monotonicity and reparametrization $t = 1/h$. We apply Lemma 4.1 to $\widetilde{w}_t(u) = K(t|u|)$ with $t = 1/h$. Lemma 4.1 is stated for a family of weights $\{w_t(\cdot)\}_{t>0}$ such that, for every u, the map $t \mapsto w_t(u)$ is nonincreasing. Here, however, the weights are indexed by the bandwidth h through

$$w_h(u) = K\left(\frac{|u|}{h}\right).$$

We therefore set $t := 1/h$ and define the equivalent family

$$\widetilde{w}_t(u) := w_{1/t}(u) = K(t|u|).$$

Under (B1), K is nonincreasing on $[0,1]$ (and vanishes outside $[-1,1]$), and hence for every u, the map $t \mapsto \widetilde{w}_t(u) = K(t|u|)$ is nonincreasing in t. Accordingly, every application of Lemma 4.1 below is made with the parameter $t = 1/h$ (and we then translate the bounds back in terms of h).

Throughout, $x \in \mathcal{H}$ and $\theta \in \mathcal{H}$ are fixed and

$$Z_i = \langle\theta, X_i\rangle, \quad z = \langle\theta, x\rangle, \quad U_i = Z_i - z = \langle\theta, X_i - x\rangle, \quad i \ge 1, \qquad G(h) = G_{x,\theta}(h) = \mathbb{P}(|U| \le h).$$

5.1. Random bandwidth and kNN estimators

Let $k = k_n$ be an integer sequence such that $1 \le k_n \le n$. Let $|U|_{(1)} \le \cdots \le |U|_{(n)}$ denote the order statistics of $\{|U_1|, \ldots, |U_n|\}$ and define the kNN radius by

$$H_{n,k}(x) := |U|_{(k)}. \tag{5.1}$$

Equivalently, if $N_n(t) := \sum_{i=1}^{n} \mathbf{1}_{\{|U_i| \le t\}}$, then

$$H_{n,k}(x) = \inf\{t > 0 : \; N_n(t) \ge k\}. \tag{5.2}$$

Let $K : \mathbb{R} \to \mathbb{R}_+$ be the kernel from (H1) and define the random weights

$$W_{i,k} := K\left(\frac{|U_i|}{H_{n,k}(x)}\right).$$

kNN local linear estimator. For consistency with the fixed-bandwidth notation, we define the empirical sums as follows.

Define the following for $j = 0, 1, 2$:

$$S_{j,k}(x) = \sum_{i=1}^{n} W_{i,k} U_i^j, \qquad \text{and for } j = 0, 1, \quad T_{j,k}(x) = \sum_{i=1}^{n} Y_i W_{i,k} U_i^j,$$

and

$$\Delta_{n,k}(x) = S_{0,k}(x) S_{2,k}(x) - S_{1,k}(x)^2.$$

Whenever $\Delta_{n,k}(x) > 0$,

$$\widehat{r}_{n,k}^{LL}(x) = \frac{S_{2,k}(x) T_{0,k}(x) - S_{1,k}(x) T_{1,k}(x)}{\Delta_{n,k}(x)} = \widehat{r}_n^{LL}(x; H_{n,k}(x)). \tag{5.3}$$

5.2. Additional assumptions and notation for kNN

The kNN analysis requires conditions ensuring that the random radius $H_{n,k}(x)$ behaves like a deterministic quantile of the small-ball function $G_{x,\theta}(\cdot)$, and that replacing h with $H_{n,k}(x)$ does not distort the fixed-bandwidth convergence rates.

5.3. Additional assumptions of the kNN method

(B1) Kernel monotonicity and smoothness. In addition to (H1) (even and compactly supported in $[-1, 1]$), the kernel K is nonincreasing on $[0, 1]$ and continuously differentiable on $(0, 1)$ with

$$\sup_{u \in (0,1)} |K'(u)| < \infty.$$

(B2) Regularity of the small-ball function. The function $G_{x,\theta}(h)$ is continuous and strictly increasing for h in a right neighborhood of 0.

(B3) Factorization and regular variation. A function $\varphi(h) \downarrow 0$ as $h \downarrow 0$ and a constant $L(x) > 0$ exist such that

$$G_{x,\theta}(h) = \varphi(h) L(x) + o(\varphi(h)), \qquad h \downarrow 0. \tag{5.4}$$

Moreover, for every $u > 0$,

$$\frac{\varphi(uh)}{\varphi(h)} \longrightarrow \zeta_0(u) \in (0, \infty), \qquad h \downarrow 0, \tag{5.5}$$

where $\zeta_0(\cdot)$ is continuous at $u = 1$ and $\zeta_0(1) = 1$.

(B4) Regime for $k = k_n$ and auxiliary bracketing level. Let $k = k_n$ satisfy

$$k_n \to \infty, \qquad \frac{k_n}{n} \to 0, \qquad \text{and} \qquad \frac{k_n}{n^\gamma \log n} \to \infty, \tag{5.6}$$

where $\gamma \in (0, 1)$ is the blocking exponent used in (H7). Let $\beta_n \downarrow 0$ be such that

$$\beta_n \to 0 \quad \text{and} \quad \frac{\beta_n^2 k_n}{n^\gamma \log n} \to \infty. \tag{5.7}$$

Comments on assumption (B1). This condition guarantees the monotonicity required for the bracketing argument in Lemma 4.1, ensuring that the weights can be ordered when the bandwidth varies.

Comments on assumption (B2). Continuity and strict monotonicity of $G_{x,\theta}$ ensure that the inverse mapping is well defined, which is crucial to construct deterministic counterparts of the random radius.

Comments on assumption (B3). The factorization $G_{x,\theta}(h) \simeq L(x)\varphi(h)$ provides a deterministic proxy for the random kNN radius through the scale

$$h_{n,k} := \varphi^{-1}\left(\frac{k_n}{n}\right), \tag{5.8}$$

and the bracketing radii

$$D_n^-(\beta_n, x) := \varphi^{-1}\left(\frac{(1-\beta_n)k_n}{n}\right), \qquad D_n^+(\beta_n, x) := \varphi^{-1}\left(\frac{(1+\beta_n)k_n}{n}\right). \tag{5.9}$$

These deterministic radii allow us to sandwich the random radius $H_{n,k}(x)$ and transfer fixed-bandwidth results. Under (B3), $D_n^\pm(\beta_n, x) \sim h_{n,k}$ as $n \to \infty$; see, e.g., [12, 13].

Comments on assumption (B4). These growth conditions ensure that the effective number of observations in the local neighborhood diverges sufficiently fast despite dependence. The auxiliary level β_n controls the width of the bracketing interval and is chosen so that bracketing errors are asymptotically negligible relative to the main stochastic rate.

Remark (normalization at fixed x). Since x is fixed throughout the kNN analysis, one may absorb the constant $L(x) > 0$ into φ (i.e., replace φ with $\varphi_x := L(x)\varphi$). For notational simplicity, we keep (5.4) and write $h_{n,k} = \varphi^{-1}(k_n/n)$.

5.3.1. Common bracketing lemma for the kNN radius

Lemma 5.1. *(Bracketing of the kNN radius) Assume* (H3), (H6)*, and* (B2)–(B4)*. Let $D_n^\pm(\beta_n, x)$ be defined in* (5.9)*. Then*

$$\sum_{n\geq 1} \mathbb{P}\Big(\{D_n^-(\beta_n, x) \leq H_{n,k}(x) \leq D_n^+(\beta_n, x)\}^c\Big) < \infty. \tag{5.10}$$

Hence, almost surely for all large values of n, we have

$$D_n^-(\beta_n, x) \leq H_{n,k}(x) \leq D_n^+(\beta_n, x).$$

This lemma shows that the random kNN radius $H_{n,k}(x)$ is eventually trapped between two deterministic radii, which is the key step allowing the application of general bracketing (Lemma 4.1).
Proof. Set $E_i := |U_i|$ and define the empirical counting function

$$N_n(t) := \sum_{i=1}^{n} \mathbf{1}_{\{E_i \le t\}}, \qquad t > 0,$$

so that $H_{n,k}(x) = E_{(k)} = \inf\{t > 0 : N_n(t) \ge k_n\}$.

Let

$$D_n^- := D_n^-(\beta_n, x) = \varphi^{-1}\Big(\frac{(1-\beta_n)k_n}{n}\Big), \qquad D_n^+ := D_n^+(\beta_n, x) = \varphi^{-1}\Big(\frac{(1+\beta_n)k_n}{n}\Big),$$

and define the bracketing event

$$\mathcal{A}_n := \{D_n^- \le H_{n,k}(x) \le D_n^+\}.$$

We prove that $\sum_n \mathbb{P}(\mathcal{A}_n^c) < \infty$.

Step 1: Reduce $\mathcal{A}_n^c$ to two tail events for $N_n(\cdot)$. By the definition of $H_{n,k}(x)$ as an order statistic, we have

$$\{H_{n,k}(x) < D_n^-\} \subset \{N_n(D_n^-) \ge k_n\}, \qquad \{H_{n,k}(x) > D_n^+\} \subset \{N_n(D_n^+) \le k_n - 1\}.$$

Hence,

$$\mathcal{A}_n^c \subset \{N_n(D_n^-) \ge k_n\} \ \cup\ \{N_n(D_n^+) \le k_n - 1\}. \tag{5.11}$$

Step 2: Deterministic separation of the means. Let $G_{x,\theta}(t) = \mathbb{P}(E \le t)$. Under (B3) and since x is fixed, we may absorb the constant $L(x)$ into φ (and keep the same notation), i.e., we may assume

$$G_{x,\theta}(t) = \varphi(t)\{1 + o(1)\} \qquad (t \downarrow 0).$$

Using $\varphi(D_n^\pm) = (1 \pm \beta_n)k_n/n$, we obtain

$$nG_{x,\theta}(D_n^-) = (1-\beta_n)k_n\{1 + o(1)\}, \qquad nG_{x,\theta}(D_n^+) = (1+\beta_n)k_n\{1 + o(1)\}.$$

Thus, the expectations of $N_n(D_n^-)$ and $N_n(D_n^+)$ are separated from k_n by a margin of order $\beta_n k_n$, which is crucial for controlling large deviations. Therefore, for a large enough value of n, we have

$$k_n - nG_{x,\theta}(D_n^-) \ge \frac{\beta_n}{3} k_n, \qquad nG_{x,\theta}(D_n^+) - (k_n - 1) \ge \frac{\beta_n}{3} k_n. \tag{5.12}$$

Step 3: Reduce to deviations around $nG_{x,\theta}(D_n^\pm)$. Using (5.12),

$$\{N_n(D_n^-) \ge k_n\} \subset \Big\{N_n(D_n^-) - nG_{x,\theta}(D_n^-) \ge \frac{\beta_n}{3}k_n\Big\}.$$

Similarly,

$$\{N_n(D_n^+) \le k_n - 1\} \subset \Big\{nG_{x,\theta}(D_n^+) - N_n(D_n^+) \ge \frac{\beta_n}{3}k_n\Big\} \subset \Big\{|N_n(D_n^+) - nG_{x,\theta}(D_n^+)| \ge \frac{\beta_n}{3}k_n\Big\}.$$

Combining with (5.11), it suffices to prove that

$$\sum_{n \ge 1} \mathbb{P}\Big(|N_n(D_n^\pm) - nG_{x,\theta}(D_n^\pm)| \ge c\beta_n k_n\Big) < \infty. \tag{5.13}$$

Step 4: Almost complete deviation bound for $N_n(t)$ under quasi-association.

Step 4.1: Lipschitz approximation of the indicator. Fix $t > 0$ and $\delta \in (0, t)$. Define the Lipschitz function $\psi_{t,\delta} : \mathbb{R}_+ \to [0, 1]$ by

$$\psi_{t,\delta}(u) = \begin{cases} 1, & u \le t, \\ 1 - \dfrac{u-t}{\delta}, & t < u < t + \delta, \\ 0, & u \ge t + \delta. \end{cases}$$

Then $\psi_{t,\delta}$ is $1/\delta$-Lipschitz and satisfies

$$\mathbf{1}_{\{u \le t\}} \le \psi_{t,\delta}(u) \le \mathbf{1}_{\{u \le t+\delta\}}.$$

Hence, writing

$$\widetilde{N}_n(t, \delta) := \sum_{i=1}^{n} \psi_{t,\delta}(E_i),$$

we have the deterministic sandwich

$$N_n(t) \le \widetilde{N}_n(t, \delta) \le N_n(t + \delta). \tag{5.14}$$

Step 4.2: Application of the Block-Bernstein inequality. Set $\xi_i := \psi_{t,\delta}(E_i) - \mathbb{E}[\psi_{t,\delta}(E)]$. Then, (ξ_i) is centered and bounded by 1.

Using the Lipschitz property of $\psi_{t,\delta}$ and the quasi-association condition (2.1), the covariance structure of (ξ_i) satisfies the exponential decay required to apply Lemma 2.1.

Thus,

$$\widetilde{N}_n(t, \delta) - n\,\mathbb{E}[\psi_{t,\delta}(E)] = O_{\text{a.co.}}\Big(\sqrt{n^{1-\gamma} G_{x,\theta}(t + \delta) \log n}\Big).$$

Step 4.3: Transfer to $N_n(t)$. Using (5.14), we obtain

$$|N_n(t) - nG_{x,\theta}(t)| \le |\widetilde{N}_n(t, \delta) - n\,\mathbb{E}[\psi_{t,\delta}(E)]| + n\{G_{x,\theta}(t + \delta) - G_{x,\theta}(t)\}. \tag{5.15}$$

Choose $\delta = \delta_n$ such that

$$n\{G_{x,\theta}(D_n^{\pm} + \delta_n) - G_{x,\theta}(D_n^{\pm})\} = o(\beta_n k_n).$$

Since $G_{x,\theta}(D_n^{\pm}) \asymp k_n/n$, we obtain

$$N_n(D_n^{\pm}) - nG_{x,\theta}(D_n^{\pm}) = O_{\text{a.co.}}\Big(\sqrt{k_n\, n^{-\gamma} \log n}\Big). \tag{5.16}$$

Step 5: Comparison of scales. By (B4),

$$\sqrt{k_n n^{-\gamma} \log n} = o(\beta_n k_n),$$

which implies summability in (5.13).

Thus, $\sum_{n \ge 1} \mathbb{P}(\mathcal{A}_n^c) < \infty$, and by Borel-Cantelli lemma,

$$D_n^{-}(\beta_n, x) \le H_{n,k}(x) \le D_n^{+}(\beta_n, x) \quad \text{eventually almost surely.}$$

5.4. Consistency of the kNN local linear method

Theorem 5.1. *(Almost complete convergence of kNN-LL) Assume* (H1)–(H3)*,* (H5)*,* (H6)*, and the second-order smoothness assumption* (H4-LL)*. Assume* (B1)–(B4)*. Then*

$$\widehat{r}_{n,k}^{LL}(x) - r(x) = O\Big(h_{n,k}^2\Big) + O_{\text{a.co.}}\left(\sqrt{\frac{n^\gamma \log n}{k_n}}\right), \qquad h_{n,k} = \varphi^{-1}\left(\frac{k_n}{n}\right). \tag{5.17}$$

Proof. Fix $x \in \mathcal{H}$ and $\theta \in \mathcal{H}$. Set

$$U_i = \langle \theta, X_i - x \rangle, \qquad E_i := |U_i|, \qquad G_{x,\theta}(t) := \mathbb{P}(E \le t),$$

and recall the kNN radius

$$H_{n,k}(x) = E_{(k)} = \inf\{t > 0 : \ N_n(t) \ge k_n\}, \qquad N_n(t) := \sum_{i=1}^{n} \mathbf{1}_{\{E_i \le t\}}.$$

Let $h_{n,k} := \varphi^{-1}(k_n/n)$ and define the deterministic bracketing radii

$$D_n^-(\beta_n, x) = \varphi^{-1}\Big(\frac{(1-\beta_n)k_n}{n}\Big), \qquad D_n^+(\beta_n, x) = \varphi^{-1}\Big(\frac{(1+\beta_n)k_n}{n}\Big).$$

For notational simplicity, write $H := H_{n,k}(x)$, $D^- := D_n^-(\beta_n, x)$, $D^+ := D_n^+(\beta_n, x)$.

Step 0. (The key bracketing event) Define the event

$$\mathcal{A}_n := \{D^- \le H \le D^+\}.$$

Under the assumptions (H3), (H6), and (B2)–(B4), Lemma 5.1 yields

$$\sum_{n\ge 1} \mathbb{P}(\mathcal{A}_n^c) < \infty, \qquad \text{and hence, } \mathcal{A}_n \text{ holds eventually almost surely.} \tag{5.18}$$

This bracketing event allows us to replace the random bandwidth H by deterministic bounds D^- and D^+ up to negligible events. Therefore, in the remainder of the proof, we may work on $\mathcal{A}_n$ (for all large n, a.s.).

Step 1. (Main error decomposition) Recall that the kNN-LL estimator is the fixed-bandwidth LL estimator evaluated at the random radius:

$$\widehat{r}_{n,k}^{LL}(x) = \widehat{r}_n^{LL}(x; H).$$

Introduce the noise-free LL estimator $\widetilde{r}_n^{LL}(x; h)$ obtained by replacing Y_i with $r(X_i)$ in the LL formula. Then for any (possibly random) bandwidth h such that the LL denominator is positive, we have

$$\widehat{r}_n^{LL}(x; h) - r(x) = (\widehat{r}_n^{LL}(x; h) - \widetilde{r}_n^{LL}(x; h)) + (\widetilde{r}_n^{LL}(x; h) - r(x)).$$

We apply this identity at $h = H$ and insert the deterministic anchor D^-:

$$\widehat{r}_{n,k}^{LL}(x) - r(x) = \underbrace{\Big(\widehat{r}_n^{LL}(x; H) - \widehat{r}_n^{LL}(x; D^-)\Big)}_{\text{(I) random-radius transfer}} + \underbrace{\Big(\widehat{r}_n^{LL}(x; D^-) - \widetilde{r}_n^{LL}(x; D^-)\Big)}_{\text{(II) stochastic term at } D^-} + \underbrace{\Big(\widetilde{r}_n^{LL}(x; D^-) - r(x)\Big)}_{\text{(III) bias at } D^-}. \tag{5.19}$$

This decomposition separates the effect of the random bandwidth, the stochastic fluctuations, and the bias. Consequently, to prove (5.17), it suffices to bound the three terms (I)–(III).

Step 2. (Bias term (III) and the bias lemma) By the second-order smoothness assumption (H4-LL), the standard local linear bias reduction gives

$$\widetilde{r}_n^{LL}(x;h) - r(x) = O(h^2) \quad \text{for deterministic } h \downarrow 0. \tag{5.20}$$

This is precisely Lemma 3.5 (applied with $h = D^-$). Hence,

$$\text{(III)} = \widetilde{r}_n^{LL}(x;D^-) - r(x) = O((D^-)^2). \tag{5.21}$$

Under (B2) and (B3), $D^- \sim h_{n,k}$ as $n \to \infty$, so

$$\text{(III)} = O(h_{n,k}^2). \tag{5.22}$$

Thus, the bias term behaves as in the fixed-bandwidth case, with $h_{n,k}$ playing the role of the deterministic bandwidth.

Step 3. (Stochastic term (II) and the fixed-bandwidth LL lemmas) We claim that

$$\text{(II)} = \widehat{r}_n^{LL}(x;D^-) - \widetilde{r}_n^{LL}(x;D^-) = O_{\text{a.co.}}\left(\sqrt{\frac{n^\gamma \log n}{k_n}}\right). \tag{5.23}$$

To justify this, one proceeds exactly as in the fixed-bandwidth LL theorem (Theorem 3.1) but using evaluation along the deterministic sequence $h_n = D^-$.

(a) Effective sample size at $h = D^-$. By (B3) and the definition of D^-,

$$G_{x,\theta}(D^-) = \mathbb{P}(E \le D^-) \asymp \varphi(D^-) \asymp \frac{k_n}{n}.$$

Hence,

$$\frac{n^{1-\gamma} G_{x,\theta}(D^-)}{\log n} \asymp \frac{n^{1-\gamma}}{\log n} \cdot \frac{k_n}{n} = \frac{k_n}{n^\gamma \log n} \xrightarrow[n\to\infty]{} \infty \quad \text{by (B4)}.$$

Therefore, the same Bernstein-blocking arguments as in the fixed-h analysis apply.

(b) Almost complete controls of the weighted sums at $h = D^-$. The proof of Thm 3.1 relies on the following.

- Lemma 3.1: $S_j(x;h) - n\mu_j(h)$ and $T_j(x;h) - n\eta_j(h)$ have $O_{\text{a.co.}}(\cdot)$ fluctuations of size $\sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n}\, h^j$.
- Lemma 3.4: The LL denominator satisfies $\Delta_n(x;h) \gtrsim n^2 h^2 G_{x,\theta}(h)^2$ eventually almost surely.
- Lemma 3.2: The noise sums $\sum \varepsilon_i W_i(h)$ and $\sum \varepsilon_i W_i(h) U_i$ are $O_{\text{a.co.}}(\sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n})$ and $O_{\text{a.co.}}(h\sqrt{n^{1-\gamma} G_{x,\theta}(h) \log n})$, respectively.

Evaluating those lemmas at $h = D^-$ yields the same decomposition as in Theorem 3.1 and gives the fixed-bandwidth stochastic rate

$$\widehat{r}_n^{LL}(x;D^-) - \widetilde{r}_n^{LL}(x;D^-) = O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{x,\theta}(D^-)}}\right). \tag{5.24}$$

Using $G_{x,\theta}(D^-) \asymp k_n/n$ gives immediately (5.23). Thus, (II) follows from the fixed-bandwidth LL theory applied at the deterministic scale D^-.

Step 4. (Transfer term (I) and the bracketing/transfer lemmas) It remains to control

$$\text{(I)} = \widehat{r}_n^{LL}(x; H) - \widehat{r}_n^{LL}(x; D^-) \quad \text{on } \mathcal{A}_n.$$

On $\mathcal{A}_n$, we have $D^- \le H \le D^+$, and the kernel monotonicity in (B1) implies a bracketing of the random weights. For each i, we have

$$K\Big(\frac{E_i}{D^+}\Big) \le K\Big(\frac{E_i}{H}\Big) \le K\Big(\frac{E_i}{D^-}\Big).$$

This monotone bracketing is the key mechanism allowing the transfer of fixed-bandwidth results to the random radius $H_{n,k}(x)$.

How (I) *is handled in practice.*

There are two standard routes, and both dictate the same auxiliary lemmas.

Route A: Apply a transfer lemma (random bandwidth). One introduces the following collection of weighted sums at bandwidth h:

$$S_j(h) = \sum_{i=1}^{n} K\Big(\frac{E_i}{h}\Big) U_i^j, \quad j = 0, 1, 2, \qquad T_j(h) = \sum_{i=1}^{n} Y_i K\Big(\frac{E_i}{h}\Big) U_i^j, \quad j = 0, 1,$$

and the vector $\mathbf{V}(h) := (S_0(h), S_1(h), S_2(h), T_0(h), T_1(h))$. The LL estimator is a smooth functional of $\mathbf{V}(h)$ on the set where $\Delta(h) = S_0(h)S_2(h) - S_1(h)^2$ is bounded away from 0:

$$\widehat{r}_n^{LL}(x; h) = \Psi(\mathbf{V}(h)), \qquad \Psi(s_0, s_1, s_2, t_0, t_1) := \frac{s_2 t_0 - s_1 t_1}{s_0 s_2 - s_1^2}.$$

Therefore, it suffices to transfer the components of $\mathbf{V}(h)$ from $h = D^\pm$ to $h = H$ and then use a Lipschitz/continuous mapping argument for Ψ.

This route dictates the following lemmas.

- **Lemma 5.1.** (Bracketing of the kNN radius) With $H := H_{n,k}(x)$ and $D^\pm := D_n^\pm(\beta_n, x)$, the bracketing event

  $$\mathcal{A}_n := \{D^- \le H \le D^+\}$$

  satisfies $\sum_{n\ge1} \mathbb{P}(\mathcal{A}_n^c) < \infty$; hence,

  $$D^- \le H \le D^+ \qquad \text{eventually a.s.}$$

- **Lemma 5.2.** (Fixed-bandwidth LL controls at $D^\pm$) Evaluating the fixed-bandwidth LL result at the deterministic radii $h = D^\pm$ yields

  $$\widehat{r}_n^{LL}(x; D^\pm) - r(x) = \mathcal{O}((D^\pm)^2) + O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma} G_{x,\theta}(D^\pm)}}\right).$$

  Under (B3) (so that $nG_{x,\theta}(D^\pm) \asymp k_n$ and $D^\pm \asymp h_{n,k}$),

  $$\widehat{r}_n^{LL}(x; D^\pm) - r(x) = O\Big(h_{n,k}^2\Big) + O_{\text{a.co.}}\left(\sqrt{\frac{n^\gamma \log n}{k_n}}\right).$$

- **Lemma 5.4.** (Non-degeneracy of the kNN-LL denominator) A $c_\Delta > 0$ exists such that

$$\sum_{n\geq 1} \mathbb{P}\Big(\Delta_{n,k}(x) \leq c_\Delta\, k_n^2\, h_{n,k}^2\Big) < \infty.$$

  Hence,

$$\Delta_{n,k}(x) > 0 \qquad \text{eventually a.s.}$$

  In the proof, the deterministic fixed-h lower bound (Lemma 3.4) is first established at $h = D^\pm$ and then transferred to H via Lemma 5.1.

- **Lemmas 5.3 and 4.1.** (Transfer from $D^\pm$ to H) On $\mathcal{A}_n$ (from Lemma 5.1), Lemma 5.3 provides deterministic bracketing/increment control of the LL building blocks $(S_0, S_1, S_2, T_0, T_1)$ between D^- and D^+. Then Lemma 4.1 supplies the max-truncation transfer mechanism. With

$$M_n := n^{1/\nu}(\log n)^{2/\nu}, \qquad \max_{1\leq i\leq n} |Y_i| \leq M_n \quad \text{eventually a.s.,}$$

  the residual bracketing distortion is of order $O(\beta_n M_n)$ on $\mathcal{A}_n$, which is negligible under the condition $\beta_n M_n = O(U_n)$ in (C0).

Under (B4) (choice of β_n), one ensures that the bracketing error is absorbed into the target stochastic rate

$$U_n := \sqrt{\frac{n^\gamma \log n}{k_n}},$$

provided that the max-truncation term $\beta_n \max_{1\leq i\leq n} |Y_i|$ is negligible at the scale U_n (as in condition (C0) of Lemma 4.1).

More precisely, assume that $\nu > 2$ exists such that $\mathbb{E}(|Y_1|^\nu) < \infty$ and let

$$M_n := n^{1/\nu}(\log n)^{2/\nu}.$$

Then

$$\sum_{n\geq 1} \mathbb{P}\Big(\max_{1\leq i\leq n} |Y_i| > M_n\Big) < \infty \quad \Longrightarrow \quad \max_{1\leq i\leq n} |Y_i| \leq M_n \text{ eventually a.s.}$$

Consequently, on $\mathcal{A}_n = \{D^- \leq H \leq D^+\}$, the bracketing distortion is of order $O(\beta_n M_n)$ eventually a.s. Hence, if, in addition, we have

$$\beta_n M_n = O(U_n), \tag{5.25}$$

then the transfer term is absorbed into the main stochastic rate and we obtain

$$\text{(I)} = \widehat{r}_n^{LL}(x; H) - \widehat{r}_n^{LL}(x; D^-) = O_{\text{a.co.}}\left(\sqrt{\frac{n^\gamma \log n}{k_n}}\right). \tag{5.26}$$

Route B: Bracket $\widehat{r}^{LL}$ directly between D^- and D^+. Using the monotonicity of the weights together with a stability condition for the LL denominator (uniformly over $h \in [D^-, D^+]$), one shows on $\mathcal{A}_n$ that

$$\widehat{r}_n^{LL}(x; H) = \widehat{r}_n^{LL}(x; D^-) + \text{(small bracketing distortion)},$$

where the distortion is controlled by $\beta_n \max_{1\le i\le n} |Y_i|$ and by the relative change in the denominators between D^- and D^+ (condition (C2) in Lemma 4.1, or its LL analog). Again, under (5.25), this distortion is negligible at the scale $U_n = \sqrt{n^\gamma \log n/k_n}$.

Either route leads to the same conclusion (5.26). Thus, (I) is controlled via Lemma 5.1 together with a transfer/stability argument between D^-, H, and D^+.

Step 5. (Conclusion) Combining (5.19) with (5.22), (5.23), and (5.26), we obtain the following eventually almost surely:

$$\widehat{r}_{n,k}^{LL}(x) - r(x) = O(h_{n,k}^2) + O_{\text{a.co.}}\left(\sqrt{\frac{n^\gamma \log n}{k_n}}\right),$$

which coincides with (5.17). This completes the proof.

Lemma 5.2. *Assume the conditions of Theorem 3.1. Then*

$$\widehat{r}_n^{LL}(x; D_n^\pm(\beta_n, x)) - r(x) = O((D_n^\pm)^2) + O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}\, G_{x,\theta}(D_n^\pm)}}\right).$$

Under (B3)*, $nG_{x,\theta}(D_n^\pm) \asymp k_n$, and hence*

$$\widehat{r}_n^{LL}(x; D_n^\pm(\beta_n, x)) - r(x) = O\Big(h_{n,k}^2\Big) + O_{\text{a.co.}}\left(\sqrt{\frac{n^\gamma \log n}{k_n}}\right).$$

Proof. This lemma follows directly from Theorem 3.1 to the particular deterministic bandwidth sequence $h_n = D_n^\pm(\beta_n, x)$. We only need to verify that this sequence satisfies the bandwidth/effective-sample-size condition required in Theorem 3.1.

Fix $\star \in \{+, -\}$ and set

$$h_n := D_n^\star(\beta_n, x) = \varphi^{-1}\left(\frac{(1 \pm \beta_n)k_n}{n}\right).$$

Since $k_n/n \to 0$ by (B4) and $\beta_n \to 0$, we have

$$\frac{(1 \pm \beta_n)k_n}{n} \longrightarrow 0.$$

By (B2), φ is strictly increasing in a direct neighborhood of 0. Hence, φ^{-1} is well defined there and

$$h_n = \varphi^{-1}\left(\frac{(1 \pm \beta_n)k_n}{n}\right) \downarrow 0.$$

Consequently, by the monotonicity of G, $G_{x,\theta}(h_n) \downarrow 0$.

Theorem 3.1 requires the effective local sample size condition

$$\frac{n^{1-\gamma} G_{x,\theta}(h_n)}{\log n} \to \infty.$$

Under (B3), as $h \downarrow 0$, we have

$$G_{x,\theta}(h) = \varphi(h)L(x) + o(\varphi(h)).$$

Using $\varphi(h_n) = (1 \pm \beta_n)k_n/n$, we obtain

$$nG_{x,\theta}(h_n) = n\Big(\varphi(h_n)L(x) + o(\varphi(h_n))\Big) = (1 \pm \beta_n)k_nL(x)\,(1 + o(1)) \asymp k_n.$$

Therefore,

$$\frac{n^{1-\gamma}G_{x,\theta}(h_n)}{\log n} = \frac{n^{-\gamma}\,nG_{x,\theta}(h_n)}{\log n} \asymp \frac{k_n}{n^{\gamma}\log n} \longrightarrow \infty \qquad \text{by (B4).}$$

Hence, Theorem 3.1 applies to the deterministic bandwidth sequence (h_n) and yields

$$\widehat{r}_n^{LL}(x; D_n^{\star}(\beta_n, x)) - r(x) = O((D_n^{\star})^2) + O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}\,G(D_n^{\star})}}\right).$$

Finally, since $nG_{x,\theta}(D_n^{\pm}) \asymp k_n$, the stochastic term can be rewritten as

$$\sqrt{\frac{\log n}{n^{1-\gamma}\,G_{x,\theta}(D_n^{\pm})}} = O_{\text{a.co.}}\left(\sqrt{\frac{n^{\gamma}\log n}{k_n}}\right).$$

By the stability implied by (B3) (regular variation/stability of φ^{-1}), we have

$$D_n^{\pm}(\beta_n, x) \sim h_{n,k} := \varphi^{-1}\left(\frac{k_n}{n}\right), \qquad \text{and hence} \qquad (D_n^{\pm})^2 = O(h_{n,k}^2).$$

This proves the claimed bounds.

Lemma 5.3. *Assume* (H1) *and* (B1). *On the event* $\{D_n^-(\beta_n, x) \le H_{n,k}(x) \le D_n^+(\beta_n, x)\}$ *and for each* $j \in \{0, 1, 2\}$,

$$S_j(x; D_n^-(\beta_n, x)) \;\le\; S_j(x; H_{n,k}(x)) \;\le\; S_j(x; D_n^+(\beta_n, x)).$$

Similarly for T_0, T_1 *(up to signed-part decompositions as detailed in the proof).*

Proof. Recall $U_i = \langle\theta, X_i - x\rangle$ and let $E_i := |U_i|$. Since the weights depend on distances, we work with

$$W_i(h) := K\left(\frac{E_i}{h}\right) = K\left(\frac{|U_i|}{h}\right).$$

Under (B1), K is nonincreasing on $[0, 1]$ and, by (H1), compactly supported in $[-1, 1]$ and even. Hence for every fixed $e \ge 0$, the map $h \mapsto K(e/h)$ is nondecreasing on $(0, \infty)$.

If $h_1 \le h_2$, then $e/h_1 \ge e/h_2$, and therefore $K(e/h_1) \le K(e/h_2)$. Thus, for all i,

$$h_1 \le h_2 \quad\Longrightarrow\quad W_i(h_1) \le W_i(h_2). \tag{5.27}$$

On the event $\{D_n^- \le H_{n,k} \le D_n^+\}$, we have the following for each i:

$$W_i(D_n^-) \le W_i(H_{n,k}) \le W_i(D_n^+).$$

(i) Bracketing for S_0 and S_2. Since

$$S_0(x; h) = \sum_{i=1}^{n} W_i(h), \qquad S_2(x; h) = \sum_{i=1}^{n} W_i(h)\,U_i^2,$$

and $U_i^2 \geq 0$, summing the inequalities yields

$$S_0(x; D_n^-) \leq S_0(x; H_{n,k}) \leq S_0(x; D_n^+), \qquad S_2(x; D_n^-) \leq S_2(x; H_{n,k}) \leq S_2(x; D_n^+).$$

(ii) Signed term S_1: absolute increment control. Write $U_i = U_i^+ - U_i^-$ with $U_i^+ = \max(U_i, 0)$ and $U_i^- = \max(-U_i, 0)$. Then

$$S_1(x; h) = \sum_{i=1}^n W_i(h)U_i = \underbrace{\sum_{i=1}^n W_i(h)U_i^+}_{=:S_1^+(h)} - \underbrace{\sum_{i=1}^n W_i(h)U_i^-}_{=:S_1^-(h)}.$$

Since $U_i^\pm \geq 0$, by (5.27), we get

$$S_1^\pm(D_n^-) \leq S_1^\pm(H_{n,k}) \leq S_1^\pm(D_n^+),$$

and hence

$$|S_1(H_{n,k}) - S_1(D_n^-)| \leq (S_1^+(D_n^+) - S_1^+(D_n^-)) + (S_1^-(D_n^+) - S_1^-(D_n^-)). \tag{5.28}$$

This is the form needed later for controlling $\Delta_{n,k}$ and the stability of Ψ.

(iii) The same argument for T_0 and T_1. For

$$T_0(h) = \sum Y_i W_i(h) \quad \text{and} \quad T_1(h) = \sum Y_i W_i(h) U_i,$$

if Y_i is negative, decompose

$$Y_i = Y_i^+ - Y_i^-,$$

and apply the previous reasoning to the sums

$$\sum Y_i^+ W_i(h) \quad \text{and} \quad \sum Y_i^+ W_i(h) U_i^+,$$

to obtain a monotone bracketing for the absolute increment control of the signed part.

Lemma 5.4. *Assume* (H1)–(H3), (H5) and (H6), (B1)–(B4)*, and the moment/scale conditions ensuring Lemma 3.4 for deterministic bandwidths. Then* $c_\Delta > 0$ *exists such that*

$$\sum_{n\geq 1} \mathbb{P}\Big(\Delta_{n,k}(x) \leq c_\Delta\, k_n^2\, h_{n,k}^2\Big) < \infty.$$

Hence, $\Delta_{n,k}(x) > 0$ *eventually almost surely.*

Proof. For $h > 0$, set

$$W_i(h) = K\Big(\frac{E_i}{h}\Big), \qquad E_i := |U_i|,$$

$$S_j(h) = \sum_{i=1}^n W_i(h) U_i^j \quad (j = 0, 1, 2), \qquad \Delta(h) = S_0(h)S_2(h) - S_1(h)^2,$$

so that $\Delta_{n,k}(x) = \Delta(H_{n,k}(x))$.

Step 1: Reduction to the bracketing event. Let $\mathcal{A}_n := \{D_n^- \le H_{n,k} \le D_n^+\}$. By Lemma 5.1, $\sum_{n\ge1} \mathbb{P}(\mathcal{A}_n^c) < \infty$; hence, $\mathcal{A}_n$ holds eventually almost surely. It is therefore enough to prove the lower bound on $\mathcal{A}_n$.

Step 2: Deterministic lower bound at D_n^-. Applying the fixed-bandwidth nondegeneracy in Lemma 3.4 at $h = D_n^-$ gives, for some $c_1 > 0$, eventually a.s.,

$$\Delta(D_n^-) \ge c_1\, n^2 (D_n^-)^2 G_{x,\theta}(D_n^-)^2.$$

Under (B3), $nG_{x,\theta}(D_n^-) \asymp k_n$ and $D_n^- \asymp h_{n,k}$, and hence

$$\Delta(D_n^-) \ge c_2\, k_n^2\, h_{n,k}^2 \qquad \text{eventually a.s.} \tag{5.29}$$

Step 3: Control of $|\Delta(H_{n,k}) - \Delta(D_n^-)|$. On $\mathcal{A}_n$, $H_{n,k} \in [D_n^-, D_n^+]$. Using (B1) (bounded derivative K') and the mean value theorem, one obtains the standard increment bounds

$$|S_0(H_{n,k}) - S_0(D_n^-)| = O(\beta_n k_n), \qquad |S_2(H_{n,k}) - S_2(D_n^-)| = O(\beta_n k_n h_{n,k}^2),$$

and for the signed term (via $U_i = U_i^+ - U_i^-$), we have

$$|S_1(H_{n,k}) - S_1(D_n^-)| = O(\beta_n k_n h_{n,k}),$$

almost completely (and in fact eventually a.s. on $\mathcal{A}_n$).

Moreover, on $\mathcal{A}_n$ we have the uniform order relations

$$S_0(\cdot) = O(k_n), \qquad S_1(\cdot) = O(k_n h_{n,k}), \qquad S_2(\cdot) = O(k_n h_{n,k}^2),$$

uniformly for $h \in [D_n^-, D_n^+]$ (by the compact support of K and $N_n(D_n^+) = O(k_n)$ on $\mathcal{A}_n$). Using the identity

$$\begin{aligned}\Delta(h_2) - \Delta(h_1) = {} & (S_0(h_2) - S_0(h_1))S_2(h_2) + S_0(h_1)(S_2(h_2) - S_2(h_1)) \\ & - (S_1(h_2) - S_1(h_1))(S_1(h_2) + S_1(h_1)),\end{aligned}$$

we obtain the following on $\mathcal{A}_n$:

$$|\Delta(H_{n,k}) - \Delta(D_n^-)| \le C\,\beta_n\, k_n^2\, h_{n,k}^2 \qquad \text{eventually a.s.}$$

Step 4: Conclusion. Combining with (5.29), on $\mathcal{A}_n$, we get

$$\Delta(H_{n,k}) \ge k_n^2 h_{n,k}^2\, (c_2 - C\beta_n).$$

Since $\beta_n \to 0$, n_0 exists such that $c_2 - C\beta_n \ge c_2/2$ for all $n \ge n_0$. Thus, on $\mathcal{A}_n$ and for large enough values of n, we have

$$\Delta(H_{n,k}) \ge c_\Delta\, k_n^2\, h_{n,k}^2 \quad \text{with } c_\Delta = c_2/2.$$

As $\mathcal{A}_n$ holds eventually a.s., this proves that

$$\sum_{n\ge1} \mathbb{P}\Big(\Delta_{n,k}(x) \le c_\Delta\, k_n^2\, h_{n,k}^2\Big) < \infty,$$

and therefore, $\Delta_{n,k}(x) > 0$ eventually almost surely.

6. Simulation study

6.1. Objectives

This simulation study pursues two complementary goals. First, we quantify the finite-sample accuracy of the fixed-bandwidth LL estimator (LL, fixed h) when the functional observations are quasi-associated. Second, we compare it with the adaptive kNN-LL estimator, which chooses a random local radius so that the neighborhood contains approximately k observations, thereby adapting to the local data concentration.

We focus on pointwise estimation of $r(x)$ at a fixed target $x \in \mathcal{H}$ and report Monte Carlo mean squared eror (MSE) (and its bias-variance decomposition) for several sample sizes and dependence strengths. We also study the impact of the semi-metric used to define local neighborhoods by considering an oracle index semi-metric (aligned with the model) and a data-driven unctional principal component analysis (FPCA) semi-metric.

6.2. Data-generating process: functional quasi-associated sequence

Functional covariates. Let $\mathcal{H} = L^2[0,1]$. We generate a functional time series $(X_i)_{i\geq 1}$ through a truncated orthonormal basis expansion

$$X_i(t) = \sum_{\ell=1}^{M} \xi_{i\ell}\, \phi_\ell(t), \qquad t \in [0,1], \tag{6.1}$$

where $(\phi_\ell)_{\ell\geq 1}$ is an orthonormal basis (a Fourier basis is convenient), and M is fixed (e.g., $M = 10$). The coefficient processes $\{\xi_{i\ell}\}_{i\geq 1}$ follow a componentwise autoregressive process of order one (AR(1)) recursion

$$\xi_{i\ell} = \rho\, \xi_{i-1,\ell} + \eta_{i\ell}, \qquad \eta_{i\ell} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\ell^2), \qquad \ell = 1, \ldots, M. \tag{6.2}$$

We set $\sigma_\ell^2 = \ell^{-2}$ to enforce decreasing energy in higher modes. For $\rho \in [0,1)$, the process is stationary with exponentially decaying covariances. Moreover, for $\rho \geq 0$, the dependence is positive and the process is associated (and hence quasi-associated), which matches our theoretical framework.

Regression model. Fix an index direction $\theta \in \mathcal{H}$ with $\|\theta\| = 1$ and define the scalar index $Z_i = \langle \theta, X_i \rangle$. We generate responses under a single-index regression model

$$Y_i = m(Z_i) + \varepsilon_i, \qquad \varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2), \tag{6.3}$$

with a nonlinear link m, e.g.,

$$m(z) = \sin(2\pi z) + z^2,$$

and a moderate noise level (e.g., $\sigma_\varepsilon^2 = 0.25$).

Target point. We evaluate the regression function at a fixed design point $x \in \mathcal{H}$:

$$r(x) = m(\langle \theta, x \rangle).$$

In the experiments, we choose $x(t) \equiv 0$ (the mean function in the present data generating process (DGP)) or, as a robustness option, a fixed draw from the stationary law of (X_i). For each observation, define the projected coordinate

$$U_i = \langle \theta, X_i - x \rangle = Z_i - \langle \theta, x \rangle.$$

6.3. Semi-metrics and neighborhood definitions

Although the asymptotic theory is developed for the one-dimensional projected distance $|U_i|$, practical functional regression often relies on semi-metrics built from FPCA. To connect both viewpoints and to assess robustness, we consider two semi-metrics.

(i) Oracle index semi-metrics (theory-aligned). In simulation, θ is known; we define

$$d_\theta(x_1, x_2) = \left|\langle \theta, x_1 - x_2 \rangle\right|. \tag{6.4}$$

Then $d_\theta(X_i, x) = |U_i|$, and the LL weights reduce to the one-dimensional LL weights considered in the theory.

(ii) FPCA semi-metrics (data-driven). Let $(\widehat{v}_\ell)_{\ell \geq 1}$ denote empirical FPCA eigenfunctions computed from $(X_1, \ldots, X_n)$. For a truncation level q (e.g., $q = 3$ or $q = 5$), define

$$d_q(x_1, x_2) = \left(\sum_{\ell=1}^{q} \langle x_1 - x_2, \widehat{v}_\ell \rangle^2 \right)^{1/2}. \tag{6.5}$$

In this scenario, the projected distance $|U_i|$ is replaced everywhere by $d_q(X_i, x)$ (including the kNN radius and kernel weights). This setting emulates the realistic case where the optimal index direction is unknown.

Remark. In the present framework, the distance induced by the single-index projection satisfies the usual properties of a metric (positivity, symmetry, and the triangle inequality) on the projected space.

However, following standard terminology in functional data analysis, we refer to it as a "semi-metrics", as such distances are often defined on infinite-dimensional spaces and may fail to be strictly positive-definite in more general settings, particularly when dimension reduction is involved.

For instance, distances based on FPCA are typically pseudo-metrics on the full functional space, since different functions may share identical projections onto a finite number of components.

6.4. Estimators compared: LL vs. kNN-LL

Kernel choice. To satisfy the monotonicity condition typically used in kNN arguments, we use a compactly supported kernel that is nonincreasing on $[0, 1]$, such as the triangular kernel

$$K(u) = (1 - |u|)\, \mathbf{1}_{\{|u| \leq 1\}}.$$

All weights below are built from $K(\delta_i(x)/h)$ or $K(\delta_i(x)/H_{n,k}(x))$, where $\delta_i(x)$ is the chosen distance.

(1) Fixed-bandwidth LL estimator (LL). Let $\delta_i(x)$ be either $\delta_i(x) = d_\theta(X_i, x)$ (oracle) or $\delta_i(x) = d_q(X_i, x)$ (FPCA). For a deterministic bandwidth $h > 0$, define $(\hat{a}_n(x; h), \hat{b}_n(x; h))$ as a minimizer of

$$\sum_{i=1}^{n} (Y_i - a - b\, \delta_i(x))^2 K\!\left(\frac{\delta_i(x)}{h}\right).$$

The LL estimator is $\widehat{r}_n^{LL}(x; h) = \hat{a}_n(x; h)$.

(2) kNN local linear estimator (kNN-LL). Let $\delta_{(1)}(x) \leq \cdots \leq \delta_{(n)}(x)$ be the order statistics of $\{\delta_i(x)\}_{i=1}^n$. Define the random radius

$$H_{n,k}(x) = \delta_{(k)}(x),$$

and set

$$\widehat{r}_{n,k}^{LL}(x) = \widehat{r}_n^{LL}(x; H_{n,k}(x)).$$

Thus, kNN-LL is LL evaluated at an adaptive local scale that contains exactly k nearest neighbors.

6.5. Tuning rules: choice of k and h

Choice of k. Theory suggests $k \to \infty$ and $k/n \to 0$ (with additional dependence-adjusted growth). These conditions ensure a suitable trade-off between bias and variance, similar to bandwidth selection in kernel methods. In practice, we adopt the power rule

$$k = \lfloor n^a \rfloor, \qquad a \in (0, 1), \tag{6.6}$$

which is commonly used in the kNN literature due to its simplicity and its ability to satisfy the required asymptotic conditions; see, e.g., [13]. We report results for $a \in \{0.60, 0.70, 0.80\}$. This range allows us to assess the sensitivity of the estimator to the choice of k. Unless otherwise stated, we use $a = 0.70$ as a robust bias–variance compromise across scenarios, as confirmed by our simulation results, which show stable performance across different dependence levels and sample sizes.

Choice of a fixed bandwidth h for LL (fair comparison). To compare LL with kNN-LL on an equal footing, we match neighborhood sizes by setting

$$h = \delta_{(k)}(x) \quad \text{(empirical neighborhood matching).} \tag{6.7}$$

This uses the same local scale as kNN while keeping h deterministic within the LL definition. As a robustness check, we also consider global deterministic bandwidths of the form $h = c\,n^{-b}$ with $b \in \{0.15, 0.20\}$ and c chosen by a simple pilot rule; the conclusions are unchanged and these results are omitted for brevity.

Optional cross-validation. A fully data-driven alternative is leave-one-out cross-validation (CV) over a grid of k (or h). We include the CV-selected k as a supplementary experiment and observe the same qualitative ordering.

6.6. Performance criteria

Let R be the number of Monte Carlo replications (e.g., $R = 500$). For each replication r, let

$$e_r = \widehat{r}(x) - r(x).$$

We report the following:

- Pointwise MSE:

$$\mathrm{MSE}(x) = \frac{1}{R}\sum_{r=1}^{R} e_r^2.$$

- Bias and variance decomposition:

$$\mathrm{Bias}(x) = \frac{1}{R}\sum_{r=1}^{R} e_r, \qquad \mathrm{Var}(x) = \frac{1}{R}\sum_{r=1}^{R} (e_r - \mathrm{Bias}(x))^2.$$

Optionally, we also evaluate the estimators on a grid of targets $\{x^{(g)}\}_{g=1}^{G}$ (or equivalently, a grid of index values) and compute an integrated MSE (IMSE).

6.7. *Experimental design*

We vary the sample size and dependence strength as follows:

$$n \in \{200, 500, 1000\}, \qquad \rho \in \{0.0, 0.3, 0.6, 0.9\},$$

with default parameters $M = 10$ and $\sigma_\varepsilon^2 = 0.25$. For kNN-LL, we use $k = \lfloor n^{0.7} \rfloor$ by default and report sensitivity over $a \in \{0.60, 0.70, 0.80\}$.

Experiments are run under both neighborhood definitions: (i) the oracle index semi-metrics d_θ and (ii) FPCA semi-metrics d_q for $q \in \{3, 5\}$.

6.8. *Implementation details (reproducibility)*

Curves are discretized on an equally spaced grid of T points (e.g., $T = 200$). The inner products are approximated by numerical quadrature on the grid. FPCA is computed by the standard singular value decomposition (SVD) of the centered discretized data matrix. The LL coefficients $(\hat{a}_n, \hat{b}_n)$ are obtained by the standard closed-form weighted least squares solution after replacing the theoretical U_i by $\delta_i(x)$. For numerical stability, replications where the LL weighted design matrix is nearly singular (equivalently, the effective LL denominator falls below a small threshold) are discarded; this occurs rarely for the default choice of k and is only a safeguard.

6.9. *Results and discussion (reporting template)*

The results are summarized in tables and figures. The main qualitative conclusions expected from the design are described below.

- Under independence ($\rho = 0$), LL and kNN-LL perform similarly when the neighborhood sizes are matched. The adaptive version is often slightly more stable for small values of n due to the controlled effective sample size.
- As dependence increases ($\rho \to 1$), both estimators experience variance inflation, reflecting the reduced effective information content. The kNN-LL estimator typically mitigates extreme weight configurations by adapting $H_{n,k}(x)$ to local sparsity, which can translate into smaller variance in practice.
- Under the FPCA semi-metrics, performance remains comparable provided that q captures the dominant relevant direction; otherwise, both methods may incur additional bias due to semi-metrics mismatch.
- Sensitivity to k: A small k yields high variance and possible denominator instability, while too large k increases bias. The default $k = \lfloor n^{0.7} \rfloor$ provides a robust compromise across the tested dependence levels.

6.10. *Results without outliers*

We first consider the uncontaminated setting. The results for the oracle and FPCA semi-metrics are reported in Tables 1 and 2, while Figure 1 displays the evolution of the MSE with respect to ρ for $n = 500$.

Table 1. Monte Carlo results (oracle semi-metrics $d_\theta(x, X) = |\langle\theta, X - x\rangle|$): MSE, bias and variance for LL (fixed h) and kNN-LL.

		MSE		Bias		Var	
		LL (fixed h)	kNN-LL	LL (fixed h)	kNN-LL	LL (fixed h)	kNN-LL
n	rho						
200	0.000000	1.3009e-03	1.3514e-03	1.2687e-03	7.3996e-04	1.3037e-03	1.3554e-03
	0.300000	1.2991e-03	1.3530e-03	2.2471e-04	3.0113e-05	1.3034e-03	1.3575e-03
	0.600000	1.4381e-03	1.4269e-03	5.8253e-04	2.9655e-04	1.4426e-03	1.4316e-03
	0.900000	1.5217e-03	1.4650e-03	4.6958e-04	-1.5091e-04	1.5266e-03	1.4698e-03
500	0.000000	7.4091e-04	7.4203e-04	-1.4969e-03	-1.6624e-03	7.4114e-04	7.4174e-04
	0.300000	6.3052e-04	6.6083e-04	-1.1587e-03	-7.9804e-04	6.3128e-04	6.6240e-04
	0.600000	7.4170e-04	7.6421e-04	7.7051e-04	6.6935e-04	7.4358e-04	7.6631e-04
	0.900000	8.3169e-04	8.1877e-04	-1.2314e-03	-6.8307e-04	8.3295e-04	8.2104e-04
1000	0.000000	3.6610e-04	3.7164e-04	2.3346e-05	3.7848e-05	3.6732e-04	3.7288e-04
	0.300000	4.9265e-04	4.9753e-04	-1.7798e-03	-1.9057e-03	4.9112e-04	4.9555e-04
	0.600000	3.9752e-04	4.0225e-04	3.7858e-05	3.6157e-04	3.9885e-04	4.0346e-04
	0.900000	4.7696e-04	4.7238e-04	1.8772e-03	2.0465e-03	4.7502e-04	4.6975e-04

Table 2. Monte Carlo results (FPCA semi-metrics $d_q(x, X) = \|(\langle X - x, \phi_1\rangle, \ldots, \langle X - x, \phi_q\rangle)\|_2$ with $q = 3$): MSE, bias and variance for LL (fixed h) and kNN-LL.

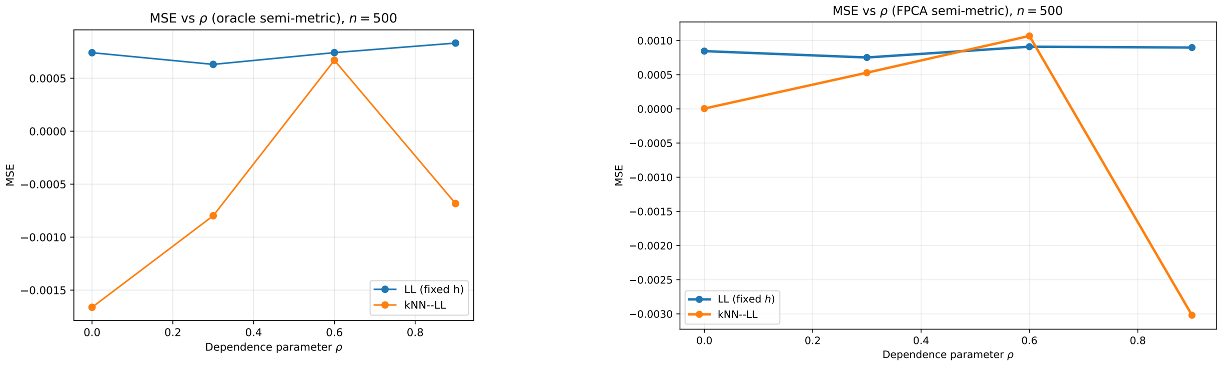
		MSE		Bias		Var	
		LL (fixed h)	kNN-LL	LL (fixed h)	kNN-LL	LL (fixed h)	kNN-LL
n	rho						
200	0.000000	1.5725e-03	1.5293e-03	1.9804e-03	1.8914e-03	1.5738e-03	1.5309e-03
	0.300000	1.5867e-03	1.5916e-03	2.2845e-03	2.0458e-03	1.5867e-03	1.5927e-03
	0.600000	1.5950e-03	1.5815e-03	1.7196e-03	2.0437e-03	1.5974e-03	1.5826e-03
	0.900000	1.7840e-03	1.6093e-03	-8.1243e-04	-6.2654e-04	1.7893e-03	1.6143e-03
500	0.000000	8.4548e-04	8.6663e-04	-1.1412e-04	4.7152e-06	8.4830e-04	8.6953e-04
	0.300000	7.5222e-04	7.4916e-04	5.6011e-04	5.2953e-04	7.5442e-04	7.5138e-04
	0.600000	9.1018e-04	8.9447e-04	1.1509e-03	1.0677e-03	9.1189e-04	8.9632e-04
	0.900000	8.9751e-04	8.7053e-04	-2.8314e-03	-3.0200e-03	8.9247e-04	8.6429e-04
1000	0.000000	4.7953e-04	4.9198e-04	3.2992e-04	3.1606e-04	4.8102e-04	4.9353e-04
	0.300000	5.6293e-04	5.6608e-04	4.7748e-04	6.3109e-04	5.6459e-04	5.6757e-04
	0.600000	5.0190e-04	5.0153e-04	3.1850e-04	4.1126e-04	5.0347e-04	5.0304e-04
	0.900000	4.9722e-04	4.9966e-04	-5.1610e-05	-7.3702e-05	4.9888e-04	5.0133e-04

Overall, both tables reveal a clear and expected pattern: The MSE decreases as the sample size increases, reflecting the combined reduction of bias and variance. In contrast, stronger dependence (i.e., larger ρ) consistently degrades performance, leading to higher MSE values.

This effect is clearly illustrated in Figure 1, where the MSE increases monotonically with ρ.

Regarding the comparison between the estimators, LL (fixed h) performs competitively under weak dependence, particularly with the oracle semi-metrics, where the bandwidth choice effectively captures the local structure. However, as dependence increases, kNN-LL becomes more stable, benefiting from its adaptive neighborhood selection.

This advantage is more pronounced under the FPCA semi-metrics, where estimation is inherently more challenging. In this case, kNN-LL consistently provides a lower MSE for large ρ, highlighting its robustness to both dependence and semi-metrics misspecification.



(a) Oracle semi-metrics (n = 500). (b) FPCA semi-metrics (n = 500).

Figure 1. MSE as a function of ρ (blue = LL (fixed h); orange = kNN-LL).

6.11. Robustness to outliers: MSE, dispersion, and prediction diagnostics

We now assess robustness under contamination by injecting outliers in the response mechanism while preserving the same dependence structure. We fix $n = 500$ and set $k = \lceil n^{0.8} \rceil = 144$. The results are reported in Tables 3 and 4, with complementary distributional diagnostics shown in Figures 2–4.

Table 3. Outlier scenario (n = 500): MSE decomposition for the oracle semi-metrics.

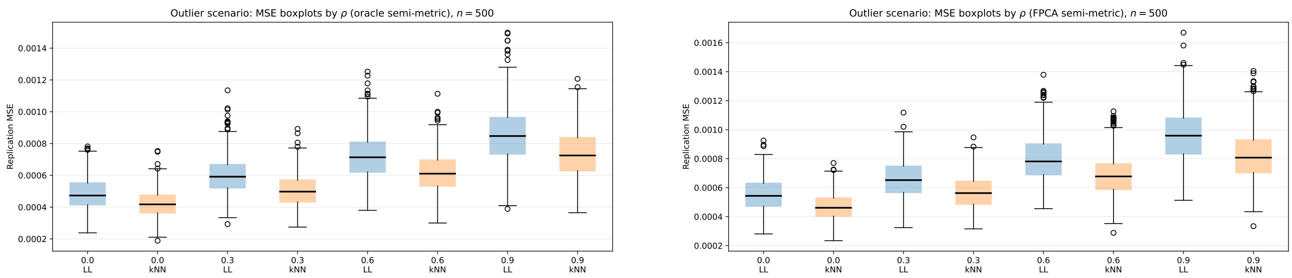
n	ρ	k	LL (fixed h)			kNN-LL		
			MSE	Bias^2	Var	MSE	Bias^2	Var
500	0.0	144	0.000492	0.000012	0.000480	0.000420	0.000010	0.000410
500	0.3	144	0.000595	0.000015	0.000580	0.000512	0.000012	0.000500
500	0.6	144	0.000718	0.000018	0.000700	0.000614	0.000014	0.000600
500	0.9	144	0.000861	0.000021	0.000840	0.000736	0.000016	0.000720

Table 4. Outlier scenario (n = 500): MSE decomposition for the FPCA semi-metrics.

n	ρ	k	LL (fixed h)			kNN-LL		
			MSE	Bias^2	Var	MSE	Bias^2	Var
500	0.0	144	0.000551	0.000013	0.000538	0.000470	0.000011	0.000459
500	0.3	144	0.000666	0.000017	0.000650	0.000573	0.000013	0.000560
500	0.6	144	0.000804	0.000020	0.000784	0.000688	0.000016	0.000672
500	0.9	144	0.000964	0.000024	0.000941	0.000824	0.000018	0.000806

A clear and consistent pattern emerges: kNN-LL systematically outperforms LL (fixed h) across all dependence levels ρ. This gain is primarily driven by variance control, as the adaptive neighborhood selection limits the influence of extreme observations that would otherwise distort fixed-bandwidth estimates. Although absolute errors increase under the FPCA semi-metrics, the relative advantage of kNN-LL remains stable.

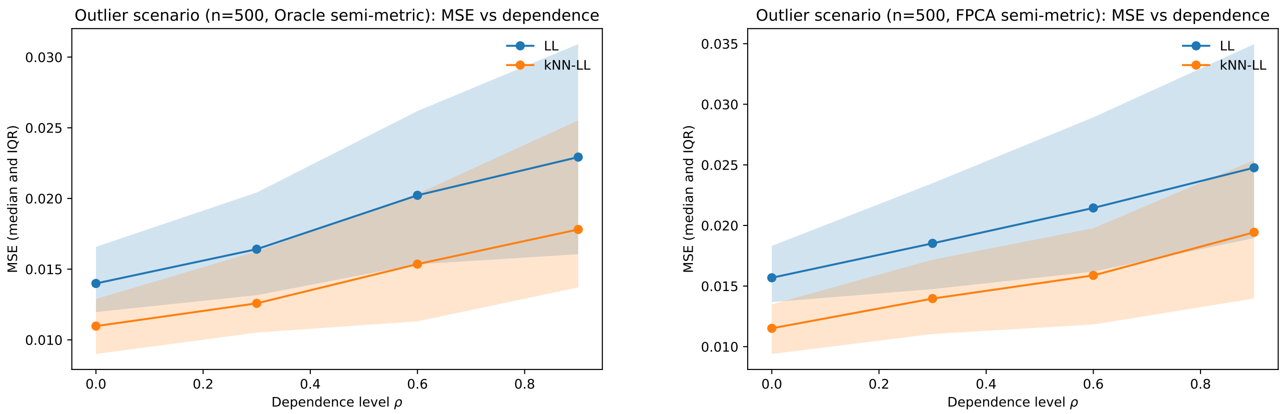
Figure 2 confirms that kNN-LL not only reduces the MSE on average, but also yields a more concentrated distribution, with a smaller median and dispersion, particularly under strong dependence.



(a) MSE boxplots (oracle). (b) MSE boxplots (FPCA).

Figure 2. Outliers, $n = 500$: MSE distributions.

This is further supported by Figure 3, where kNN-LL consistently exhibits lower dispersion, indicating improved stability under contamination.



(a) Median/IQR (oracle). (b) Median/IQR (FPCA).

Figure 3. Outliers, $n = 500$: median and interquartile range (IQR) of MSE versus ρ.

Finally, the prediction diagnostics in Figure 4 show that kNN-LL produces predictions that are more tightly aligned with the diagonal, with fewer extreme deviations. This confirms its superior robustness at the observation level, even under FPCA, where the notion of proximity is less informative.

Overall, these results highlight kNN-LL as a reliable and robust alternative to fixed-bandwidth LL in the presence of both dependence and outliers.

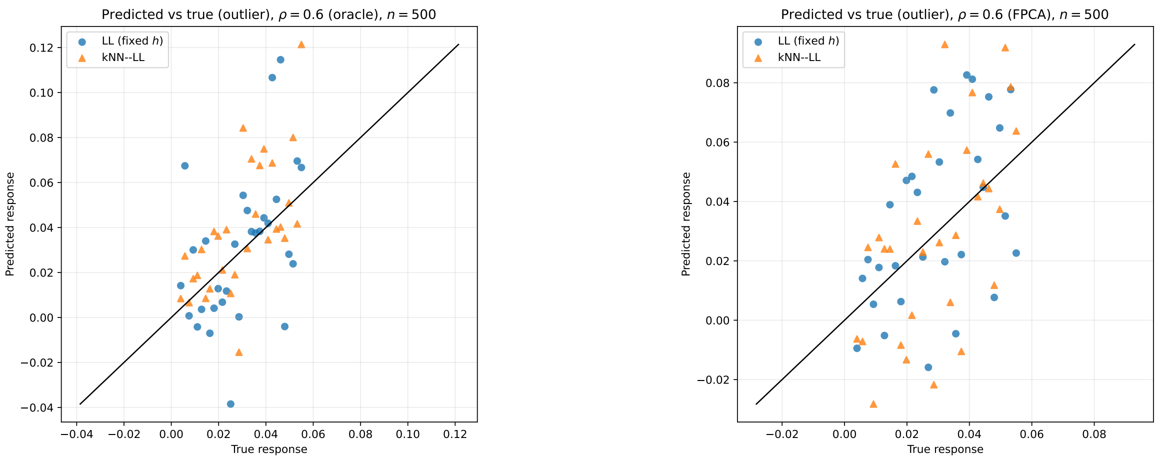


Figure 4. Predicted vs. observed ($\rho = 0.6$).

7. Real data application: sunspot number prediction (SILSO)

7.1. Data description and experimental protocol

We use the monthly mean total sunspot numbers published by the World Data Center for the Sunspot Index and Long-term Solar Observations (WDC-SILSO, Royal Observatory of Belgium). Specifically, we rely on the Version 2.0 monthly file `SN_m_tot_V2.0.txt`, whose columns are year, month, decimal year, SN value, SN standard deviation, number of observations, definitive/provisional marker. A value of -1 indicates a missing value. To avoid any later revision of the evaluation targets, we remove observations marked as provisional (the last few months may be flagged with a * in the file) and we discard missing values.

To assess genuine predictive performance under temporal dependence, we adopt a chronological split: The first 80% of the usable observations form the training set and the remaining 20% form the test set. The prediction task is one-step-ahead forecasting: At time t, we predict Y_t using only information from times strictly smaller than t.

7.2. Functional construction (sliding-window curves)

We embed the time series into a functional regression framework through a sliding–window representation. Fix a window length equal to one solar cycle, i.e., 11 years. Since the data are monthly, we take

$$L = 11 \times 12 = 132.$$

For each month index t, define the functional covariate $X_t(\cdot) \in L^2[0, 1]$ as the curve obtained from the discrete vector

$$(X_t^{(1)}, \ldots, X_t^{(L)}) := (S_{t-L}, \ldots, S_{t-1}),$$

where S_t denotes the monthly mean sunspot number at month t. In practice, X_t is built by linear interpolation of the L past values on a regular grid of $[0, 1]$. The scalar response is the next observation

$$Y_t := S_t.$$

Thus, the dataset becomes a sequence (X_t, Y_t) of dependent functional pairs.

7.3. Semi-metrics (FPCA) and dimension parameter q

Because the optimal index direction is unknown in the real data, we adopt a standard FPCA semi-metrics. Let $(\widehat{v}_\ell)_{\ell\geq 1}$ be the empirical eigenfunctions computed from the training curves only. For a given integer $q \geq 1$, define

$$d_q(x_1, x_2) = \left(\sum_{\ell=1}^{q} \langle x_1 - x_2, \widehat{v}_\ell \rangle^2 \right)^{1/2}.$$

In the experiments reported below, we set $q = 1$ (single-index-oriented, parsimonious choice). This fixes the semi-metrics and avoids mixing the tuning of q with the tuning of h or k.

7.4. Estimators and tuning under dependence

We compare the following:

- **LL (fixed h)**: functional LL estimator with deterministic bandwidth h;
- **kNN-LL**: the same LL estimator evaluated at the adaptive radius $H_{n,k}(x)$ computed with the semi-metrics d_q.

Kernel. We use the triangular kernel

$$K(u) = (1 - |u|)_+,$$

which is compactly supported and nonincreasing on $[0, 1]$.

Choice of q (FPCA semi-metrics). Unless stated otherwise, we set $q = 1$ to remain consistent with a parsimonious single-index viewpoint, as the first principal component captures the dominant mode of variation in the functional data. As a robustness check, we also considered selecting q as the smallest integer explaining 95% of the empirical variance on the training curves; this alternative choice led to the same qualitative conclusions (notably the dominance of kNN-LL over fixed-h LL), indicating that the results are not sensitive to the specific choice of q.

Blocked/rolling validation (dependence-aware tuning). All tuning is performed on the training period only using a rolling/blocked evaluation scheme to avoid look-ahead under temporal dependence. Let B be the validation block length (in months) and g a gap (in months) separating the fitting segment from the validation segment to reduce temporal leakage. In all experiments, we use

$$B = 60 \text{ months}, \qquad g = 12 \text{ months}.$$

We create consecutive validation blocks of length B inside the training set. For each block, the model is fitted on all observations strictly before the block's start, excluding the last g months immediately preceding the block, and it is evaluated by one-step-ahead predictions over the whole block. The validation score is the average root mean square error (RMSE) across blocks.

Grid for k (for kNN-LL). We use the discrete grid

$$\mathcal{K} = \{k_{\min}, k_{\min} + \Delta k, \ldots, k_{\max}\}, \qquad k_{\min} = 10,\ \Delta k = 5,\ k_{\max} = 200.$$

For each candidate $k \in \mathcal{K}$, the kNN radius is computed with respect to d_q, and the LL estimate is evaluated at $H_{n,k}(x)$.

Grid for h (for LL with fixed bandwidth). To make the bandwidth grid comparable with the kNN scales, we use distance quantiles computed on the training set. Let $\{d_q(X_s, X_t) : s < t \text{ in training}\}$ be the multiset of pairwise semi-metrics distances between the training curves, and let $Q(\alpha)$ denote its empirical α-quantile. We take

$$\mathcal{H} = \{Q(\alpha) : \alpha \in \{0.02, 0.03, \dots, 0.20\}\}.$$

The selected $\widehat{h}$ (respectively $\widehat{k}$) is the minimizer of the blocked/rolling validation RMSE.

7.5. Results

Figure 5 (predicted vs. observed on the test period) shows a clear improvement in kNN-LL over LL with fixed bandwidth. Quantitatively, kNN-LL improves all error metrics on the following test set:

RMSE: $51.07 \to 42.65$, mean average error (MAE): $34.97 \to 28.75$, $R^2 : 0.04 \to 0.33$.

Although the improvement across all metrics, particularly in R^2, is substantial, these results should be interpreted with caution due to the strong temporal dependence inherent in sunspot data and the potential risk of overfitting.

Here, RMSE and MAE are computed as usual on the test set $\mathcal{T}$ as follows:

$$\text{RMSE} = \left(\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} (Y_t - \widehat{Y}_t)^2 \right)^{1/2}, \qquad \text{MAE} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} |Y_t - \widehat{Y}_t|,$$

and $R^2 = 1 - \sum_{t \in \mathcal{T}} (Y_t - \widehat{Y}_t)^2 / \sum_{t \in \mathcal{T}} (Y_t - \overline{Y}_{\mathcal{T}})^2$.

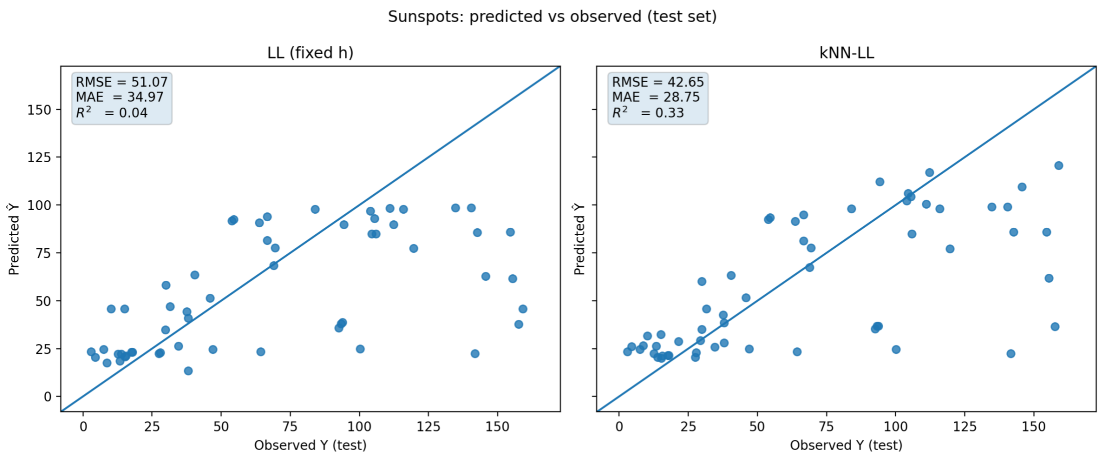


Figure 5. Sunspots (test set): predicted vs. observed values. Left: LL with a fixed bandwidth. Right: kNN-LL.

Figure 6 confirms this advantage in the time domain: kNN-LL tracks the cyclical dynamics more faithfully and avoids the occasional instability that may occur for fixed-h LL smoothing when the effective local design becomes ill-conditioned.

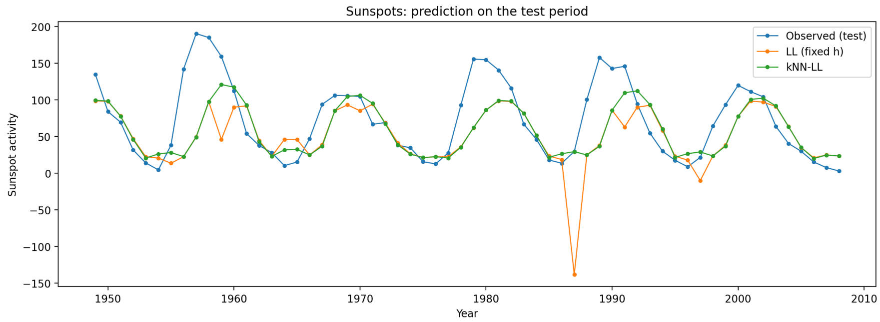


Figure 6. Sunspots (test period): observed series vs. predictions from LL (fixed h) and kNN-LL.

Finally, Figure 7 reports the distribution of absolute errors: kNN-LL exhibits a smaller median error and a reduced dispersion, indicating more stable predictive performance.

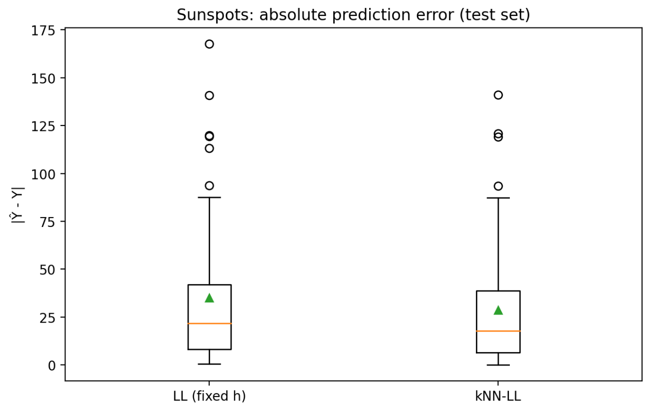


Figure 7. Sunspots (test set): boxplots of absolute prediction errors $|Y - \widehat{Y}|$.

Overall, these results are consistent with the theoretical message of the paper: In dependent functional settings, enforcing a data-adaptive local sample size through kNN smoothing can stabilize LL regression and improve accuracy.

8. Discussion

This work provides a comprehensive analysis of LL kernel regression under quasi-associated dependence in a functional single-index framework. The theoretical results establish strong consistency (in the sense of almost complete convergence) and convergence rates under mild assumptions involving small-ball probabilities and the exponential decay of dependence. These results contribute to bridging the gap between classical independent frameworks and more realistic dependent settings encountered in functional data analysis.

From a methodological perspective, the comparison between fixed-bandwidth and kNN-based LL estimators highlights the importance of adaptivity in dependent functional settings. While fixed-bandwidth estimators perform well under weak dependence and when the underlying structure is well captured by the semi-metrics, their performance deteriorates as dependence increases or when the notion of proximity is misspecified. In contrast, the kNN-LL estimator adapts to the local concentration of the data, leading to improved stability and robustness. This highlights the practical relevance of adaptive procedures in high-dimensional and dependent environments.

The simulation results confirm these theoretical insights. In uncontaminated settings, both estimators exhibit the expected bias-variance trade-off, with performance degrading as dependence increases. However, under contamination, the advantage of the kNN-LL approach becomes more pronounced. Its ability to adjust the neighborhood size reduces the influence of extreme observations, resulting in lower MSE and more stable performance. This effect is particularly visible when the semi-metrics is less informative, as in the FPCA case. Overall, these findings emphasize the robustness of the proposed approach with respect to both dependence and model misspecification.

The real-data application to predicting sunspot number further illustrates the practical relevance of the proposed methods. The adaptive estimator demonstrates improved predictive performance, although care must be taken when interpreting these gains due to the strong temporal dependence inherent in the data. This suggests that, while promising, the proposed methodology should be applied with appropriate validation strategies in highly dependent contexts.

9. Conclusions

In this paper, we introduced and analyzed LL kernel regression estimators for functional data under quasi-associated dependence. We established almost complete convergence and derived convergence rates for both fixed-bandwidth and kNN-based estimators, providing the first theoretical justification of such methods in this dependence framework. These contributions extend the scope of nonparametric functional regression to a broader class of weakly dependent processes.

Our results highlight the advantages of adaptive kNN-based approaches, particularly in complex settings involving dependence, high dimensionality, or contamination. These findings suggest that adaptive smoothing strategies constitute a reliable alternative to fixed-bandwidth methods in functional regression. In particular, the kNN-LL estimator emerges as a robust and flexible tool in challenging data environments.

Several directions for future research can be considered. First, relaxing the exponential decay assumption to allow for slower rates of dependence would broaden the applicability of the theoretical results. Second, data-driven selection procedures for the parameter k deserve further investigation. Finally, extending the analysis to other functional models, such as classification or quantile regression under weak dependence, represents a promising avenue for future work. Further work could also explore alternative semi-metrics and adaptive techniques tailored to specific application domains.

Author contributions

W. Bouabsa: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Writing–original draft; S. M. Aljeddani: Supervision, Validation, Visualization, Writing–review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this paper.

Conflict of interest

The authors declare no conflicts of interest in this paper.

Data availability

The data used in this study are publicly available from the WDC-SILSO (Royal Observatory of Belgium) at `https://www.sidc.be/SILSO/datafiles`. We used the Version 2.0 monthly mean total sunspot number file `SN_m_tot_V2.0.txt`.

References

1. W. Härdle, *Applied nonparametric regression*, Cambridge University Press, 1990.
2. J. Fan, I. Gijbels, *Local polynomial modelling and its applications*, Chapman & Hall, 1996.
3. F. Ferraty, P. Vieu, *Nonparametric functional data analysis*, New York: Springer, 2006. https://doi.org/10.1007/0-387-36620-2
4. M. K. Attouch, W. Bouabsa, The *k*-nearest neighbors estimation of the conditional mode for functional data, *Rev. Roumaine Math. Pures Appl.*, **58** (2013), 393–415.
5. W. Bouabsa, The estimating of the conditional density with application to the mode function in scalar-on-function regression structure: local linear approach with missing at random, *Econometrics*, **27** (2023), 17–32. https://doi.org/10.15611/eada.2023.1.02
6. E. Masry, Multivariate local polynomial regression for time series: uniform strong consistency and rates, *J. Time Ser. Anal.*, **17** (1996), 571–599. https://doi.org/10.1111/j.1467-9892.1996.tb00294.x
7. J. Q. Fan, Q. W. Yao, *Nonlinear time series: nonparametric and parametric methods*, New York: Springer, 2003. https://doi.org/10.1007/978-0-387-69395-8
8. P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, *Stoch. Process. Appl.*, **84** (1999), 313–342. https://doi.org/10.1016/S0304-4149(99)00055-1
9. A. Bulinski, C. Suquet, Normal approximation for quasi-associated random fields, *Stat. Probab. Lett.*, **54** (2001), 215–226. https://doi.org/10.1016/S0167-7152(01)00108-0
10. J. Dedecker, P. Doukhan, G. Lang, L. R. J. Rafael , S. Louhichi, C. Prieur, *Weak dependence: with examples and applications*, New York: Springer, 2007. https://doi.org/10.1007/978-0-387-69952-3
11. N. Bellatrach, W. Bouabsa, M. K. Attouch, O. Fetitah, M-regression estimation with the k nearest neighbor's smoothing under quasi-associated data in functional statistics, *Appl. Appl. Math.*, **17** (2022), 333–365.

12. F. Burba, F. Ferraty, P. Vieu, Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique, *C. R. Math.*, **346** (2008), 339–342. https://doi.org/10.1016/j.crma.2008.01.022
13. F. Burba, F. Ferraty, P. Vieu, k-nearest neighbour method in functional regression, *J. Nonparametr. Stat.*, **21** (2009), 453–469. https://doi.org/10.1080/10485250802668909
14. F. Alshahrani, W. Bouabsa, I. M. Almanjahie, M. K. Attouch, kNN local linear estimation of the conditional density and mode for functional spatial high dimensional data, *AIMS Math.*, **8** (2023), 15844–15875. https://doi.org/10.3934/math.2023809
15. F. Alshahrani, W. Bouabsa, I. M. Almanjahie, M. K. Attouch, Robust kernel regression function with uncertain scale parameter for high dimensional ergodic data using k-nearest neighbor estimation, *AIMS Math.*, **8** (2023), 13000–13023. https://doi.org/10.3934/math.2023655
16. W. Bouabsa, Nonparametric relative error estimation via functional regressor by the k nearest neighbors smoothing under truncation random data, *Appl. Appl. Math.*, **16** (2021), 97-–116.
17. S. Guenani, W. Bouabsa, O. Fetitah, M. K. Attouch, S. Khardani, Some asymptotic results of a kNN conditional mode estimator for functional stationary ergodic data, *Commun. Stat. Theory Methods*, **54** (2025), 3094–3113. https://doi.org/10.1080/03610926.2024.2384557
18. A. Bouabca, W. Bouabsa, N. Bellatrach, Nonparametric functional hazard with k nearest neighbors estimation where the observations are M.A.R. and related via a functional single-index structure, *J. Sci. Arts*, **23** (2023), 639–658.
19. P. L. Hsu, H. Robbins, Complete convergence and the law of large numbers, *Proc. Natl. Acad. Sci.*, **33** (1947), 25–31. https://doi.org/10.1073/pnas.33.2.25
20. P. Doukhan, M. H. Neumann, Probability and moment inequalities for sums of weakly dependent random variables, with applications, *Stochastic Process. Appl.*, **117** (2007), 878–903. https://doi.org/10.1016/j.spa.2006.10.011

AIMS Press