EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2026, Vol. 19, Issue 1, Article Number 7269
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

Nonparametric Conditional Quantile Estimation for Locally Stationary Functional Time Series: Applications in Financial and Economic Modeling

Jan Nino Tinio[1]

[1] *Department of Mathematics, College of Mathematics and Natural Sciences, Caraga State University, Butuan City, Agusan del Norte, Philippines 8600*

Abstract. Conditional distribution estimation (CDE) is central in nonparametric forecasting and risk analysis. While considerable progress has been made for finite-dimensional and stationary settings, functional data and nonstationary settings pose new challenges. We propose a Nadaraya-Watson (NW) conditional quantile estimator for regularly mixing locally stationary functional time series (LSFTS). It incorporates three kernel functions: one for time rescaling, another for functional covariates, and an integrated kernel that serves as the cumulative distribution function (CDF) of the response variable. A theoretical framework and the estimator's uniform convergence were provided. To demonstrate the estimator's consistency, a numerical experiment was conducted. Finally, we apply the method to financial data, specifically the Nikkei 225.

2020 Mathematics Subject Classifications: 62G08, 62M10, 62R10

Key Words and Phrases: Conditional distribution estimation; Functional data analysis; Locally stationary time series; Nadaraya-Watson estimation; Nonparametric regression

---

1. Introduction

In various nonparametric inferential problems, estimating the conditional distribution is fundamental for prediction and forecasting. It provides an overarching description of the conditional law for any given random variable [1–3]. In most situations, conditional distribution estimation (CDE) is considered a fundamental step in estimating characteristics such as the conditional mode and median, as well as the conditional quantile function, which is essential for detecting outliers in a given dataset. Additionally, the conditional cumulative density function (CDF) enables us to determine conditional hazard functions, which are helpful in reliability and survival analysis [4–6].

Several studies have already been done in the finite-dimensional setting. For instance, [1] proposed two approaches: the local logistic distribution and an adjusted Nadaraya-Watson (NW) procedure, for strictly stationary sequences of i.i.d. random variables.

DOI: https://doi.org/10.29020/nybg.ejpam.v19i1.7269

Email addresses: jgtinio@carsu.edu.ph (J.N. Tinio)

 Copyright: © 2026 The Author(s). (CC BY-NC 4.0)

Analogously, an adaptive weighted NW estimator for strictly stationary processes with varying bandwidths was proposed in [7]. This approach enables dynamic bandwidth selection, which is crucial for capturing the time-varying characteristics of processes.

However, in recent years, many approaches to functional data analysis (FDA) have been developed as the number of applications of functional data from infinite-dimensional spaces has increased. These applications include meteorology, medicine, satellite imagery, econometrics, and many others [8–11]. Numerous studies have examined this data in the context of CDE. For example, [6] proposed a local linear approach in estimating the CDF of mixing data. Additionally, distinct approaches for estimating the conditional distribution of a target variable within a prediction set were proposed in [12]. Moreover, [3] introduced a local polynomial estimator for the conditional CDF for stationary and strongly mixing processes.

Moreover, since many functional data deviate from stationarity, especially in meteorology and finance [13], conventional approaches are considered inappropriate since the assumption of (weak) stationarity is violated [14, 15]. As popularized by [16], the local stationarity framework is an effective modeling approach for addressing nonstationarity. The parameters of a locally stationary process exhibit temporal dependence; however, this nonstationary process can be approximated by a stationary process on finer time grids, enabling the development of asymptotic theories [16–18].

In this paper, a NW estimator for the conditional quantile function of locally stationary functional time series (LSFTS) is considered, which involves a scalar response variable $Y_{t,T}$ and a functional covariate $X_{t,T} \in \mathscr{H}$ that is regularly mixing and locally stationary. Closely related works on LSFTS—such as those of [15, 19] and subsequent single-index LSFTS studies in [20]—primarily focus on conditional mean function estimation. In contrast, our approach targets conditional quantiles, which constitutes the main methodological novelty. The NW estimation procedure is an effective nonparametric local averaging method [14, 21, 22]. The current study provides the uniform convergence of the proposed estimator. Focusing on conditional quantiles offers several advantages over mean-based methods: it provides a more comprehensive description of the conditional distribution, it is robust to outliers and heavy-tailed behavior, and it enables the analysis of heterogeneous dynamics across different parts of the distribution—features that are particularly relevant in financial and other complex functional data. Methodologically, we develop a nonparametric NW estimator tailored to this quantile setting and establish its uniform convergence under LSFTS dependence. We further demonstrate consistency through simulations and illustrate practical relevance via an empirical application, thereby extending the LSFTS literature beyond mean regression to a richer, distribution-sensitive framework.

The remaining part of this paper is organized as follows. It introduces preliminary concepts on LSFTS, small-ball probability, and mixing conditions in Section 2. In Section 3, the proposed NW estimator is defined, the considered assumptions are enumerated, and the theoretical results, together with the proposed bandwidth selection method, are provided. The results of the conducted numerical experiments are shown in Section 4. Lastly, the results of applying the method to the Nikkei 225 dataset are presented in

Section 5.

Notation. Throughout this paper, the following notations are adopted. Let δ_y denote the Dirac measure at a point y. For any real-valued random variable X and any $q \geq 1$, L_q-norm of X is denoted by $\|X\|_{L_q}$ and is defined as $\|X\|_{L_q} = (\mathbb{E}\left[|X|^q\right])^{1/q}$. The notation $a_T \lesssim b_T$ indicates that there exists a constant C, independent of T, such that $a_T \leq Cb_T$, with constant C that may vary unless specified. In addition, $a_T \sim b_T$ signifies that both $a_T \lesssim b_T$ and $b_T \lesssim a_T$ hold. For positive sequences $\{a_T\}$ and $\{b_T\}$, $a_T = \mathcal{O}(b_T)$ if $\lim_{T\to\infty} \frac{a_T}{b_T} \leq C$ for some constant $C > 0$. Additionally, $a_T = \mathcal{O}(1)$ means that a_T is bounded. If $\lim_{T\to\infty} \frac{a_T}{b_T} = 0$, then $a_T = o(b_T)$ and $a_T = o(1)$ when a_T approaches zero. Let $X_T = \mathcal{O}_{\mathbb{P}}(a_T)$ if, for every $\epsilon > 0$, there exist constants $C_\epsilon > 0$ and $T_0(\epsilon) \in \mathbb{N}$ such that for all $T \geq T_0(\epsilon)$, $\mathbb{P}[\frac{|X_T|}{a_T} > C_\epsilon] < \epsilon$. Similarly, when $\lim_{T\to\infty} \mathbb{P}[\frac{|X_T|}{a_T} > \epsilon] = 0$ for all $\epsilon > 0$, then $X_T = o_{\mathbb{P}}(a_T)$, and $X_T = o_{\mathbb{P}}(1)$ if X_T converges in probability to zero. Lastly, for any $a, b \in \mathbb{R}$, write $a \vee b = \max\{a, b\} \quad \text{and} \quad a \wedge b = \min\{a, b\}$.

2. Preliminaries

This section presents preliminary concepts that are deemed essential in this paper.

2.1. Local stationarity

Given a process $\{Y_{t,T}, X_{t,T}\}_{t=1,\ldots,T}$, for $T \in \mathbb{N}$, we consider the same regression estimation problem in [14]:

$$Y_{t,T} = m^\star\big(\frac{t}{T}, X_{t,T}\big) + \varepsilon_{t,T}, \text{ for all } t = 1, \ldots, T, \tag{2.1}$$

where $\{\varepsilon_{t,T}\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables independent of $\{X_{t,T}\}_{t=1,\ldots,T}$, i.e., $\mathbb{E}[\varepsilon_t|X_{t,T}] = 0$, for all $t = 1, \ldots, T$. The integrable response variable $Y_{t,T}$ is real-valued, and the functional covariate $X_{t,T}$ is drawn from some semi-metric space $\mathscr{H}$ with a semi-metric $d(\cdot, \cdot)$. The semi-metric space $\mathscr{H}$ can either be a Banach or a Hilbert space with norm $\|\cdot\|$, so $d(u, v) = \|u-v\|$, for all $u, v \in \mathscr{H}$. Additionally, $X_{t,T}$ is assumed to be locally stationary, meaning it evolves slowly over time and remains approximately stationary at local times. The conditional mean function $m^\star\big(\frac{t}{T}, X_{t,T}\big) = \mathbb{E}[Y_{t,T}|X_{t,T}]$ in model (2.1) does not depend on real-time t but rather on rescaled time $u = \frac{t}{T}$. These u-points form a dense subset of the unit interval $[0, 1]$ as the sample size T goes to infinity. Hence, if it is continuous in the time direction, at all rescaled u-points, $m^\star$ is identified almost surely (a.s.).

As adopted from the definition given in [14], the notion of local stationarity is formally defined as

Definition 2.1. *An $\mathscr{H}$-valued process $\{X_{t,T}\}_{t=1,\ldots,T}$ is locally stationary if there exists an associated $\mathscr{H}$-valued process $\{X_t(u)\}_{t=1,\ldots,T}$, for each rescaled time point $u \in [0, 1]$, satisfying*

$$d\big(X_{t,T}, X_t(u)\big) \leq \Big(\Big|\frac{t}{T} - u\Big| + \frac{1}{T}\Big)U_{t,T}(u) \qquad a.s., \tag{2.2}$$

where $\{U_{t,T}(u)\}_{t=1,\ldots,T}$ *is a positive process verifying* $\mathbb{E}\big[(U_{t,T}(u))^{\rho}\big] < C_U$ *for some* $\rho > 0$ *and* $C_U < \infty$ *independent of* $u, t,$ *and* T.

If an $\mathscr{H}$-valued process of random variables $\{X_{t,T}\}_t$ is locally stationary, around each rescaled time u, it can be approximated by a strictly stationary process $\{X_t(u)\}_t$, resulting in a negligible difference between random variables $X_{t,T}$ and $X_t(u)$. We note that a larger ρ indicates a better approximation of $X_{t,T}$ by $X_t(u)$ and moderate bounds for their absolute difference. This definition agrees with the definition given in [13] and [23] when $\mathscr{H}$ is a Hilbert space $L^2_{\mathbb{R}}([0,1])$, and all real-valued functions are square-integrable with respect to the Lebesgue measure on the interval $[0,1]$ with inner product L_2-norm: $\|f\|_2 = \sqrt{\langle f, f\rangle}$, where $\langle f, g\rangle = \int_0^1 f(t)g(t)\mathrm{d}t$ for $f, g \in L^2_{\mathbb{R}}([0,1])$.

2.2. Small ball probability

In infinite-dimensional spaces, the concept of small ball probability is employed to address the absence of a density function for functional variables, as there is no available universal reference measure, such as the Lebesgue measure. We control the concentration of the probability measure of a functional variable on a small ball using: for $r > 0$ and a fixed $x \in \mathscr{H}$,

$$\mathbb{P}(X \in B(x, r)) =: \phi_x(r) > 0,$$

where the space $\mathscr{H}$ involves a semi-metric $d(\cdot,\cdot)$ and $B(x,r) = \{v \in \mathscr{H} : d(x, v) \leq r\}$ is a ball in space $\mathscr{H}$ with radius r centered at $x \in \mathscr{H}$. If r is a function of T such that $r = r(T) \to 0$ as $T \to \infty$, then $B(x,r)$ is considered as a small ball, hence, $\mathbb{P}(X \in B(x,r))$ is termed a small ball probability [24]. Generally, as $r \to 0$, we suppose

$$\mathbb{P}(X \in B(x, r)) \sim \psi(x)\phi(r), \tag{2.3}$$

where we assume a normalizing restriction $\mathbb{E}[\psi(x)] = 1$ to ensure that this decomposition is identifiable. We conveniently assume (2.3) since the function $\psi(x)$ works as a surrogate density of the functional X and is utilized in different frameworks like in [25–27], where the surrogate density is estimated differently and is used for classification purposes. Additionally, the volumetric term, used to evaluate the complexity of the process's probability law, can be expressed as the function $\phi(r)$ [28].

2.3. Mixing condition

To generalize the law of large numbers for non-i.i.d. stochastic processes, the concept of mixing processes, alongside various mixing coefficients, was introduced [29–31]. It measures the degree of dependence between time-distant observations of a stochastic process, which is essential for effective modeling processes with dependency structure [31–33]. β-mixing, defined below, is one of the mixing criteria typically considered in the context of a stochastic process [34–36].

Definition 2.2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{B}$ and $\mathcal{C}$ be subfields of $\mathcal{A}$, and set $\beta(\mathcal{B}, \mathcal{C}) = \mathbb{E}[\sup_{C \in \mathcal{C}} |\mathbb{P}(C) - \mathbb{P}(C|\mathcal{B})|]$. For any array $\{Z_{t,T} : 1 \leq t \leq T\}$, define the coefficient*

$$\beta(k) = \sup_{1 \leq t \leq T-k} \beta\big(\sigma(Z_{s,T}, 1 \leq s \leq t), \sigma(Z_{s,T}, t+k \leq s \leq T)\big),$$

where $\sigma(Z)$ denotes the σ-algebra generated by Z. The array $\{Z_{t,T}\}$ is said to be β-mixing or absolutely regular mixing if $\beta(k) \to 0$ as $k \to \infty$.

If a process is β-mixing, we attain asymptotic independence as $k \to \infty$. Various forms of β-mixing can be considered, such as exponentially β-mixing: $\beta(k) = \mathcal{O}\big(e^{-\gamma k}\big)$, and arithmetically β-mixing: $\beta(k) = \mathcal{O}\big(k^{-\gamma}\big)$, for $\gamma > 0$ [37, 38].

3. Consistency of Nadaraya-Watson estimator

For a fixed $t \in \{1, \ldots, T\}$ and $x \in \mathscr{H}$, let us denote the conditional CDF by

$$F_t^{\star}(y|x) = \mathbb{P}[Y_{t,T} \leq y | X_{t,T} = x] = \mathbb{E}[\mathbb{1}_{Y_{t,T} \leq y} | X_{t,T} = x],$$

for all $y \in \mathbb{R}$. Let K_1, K_2, and K_3 be 1-dimensional kernel functions and $h = h(T)$ be a bandwidth satisfying $h(T) \to 0$ as $T \to \infty$. Additionally, define $H(z) = \int_{-\infty}^{z} K_3(v) dz$ and set $H_h(\cdot) = H(\frac{\cdot}{h})$, and $K_{h,i}(\cdot) = K_i(\frac{\cdot}{h})$, for $i = 1, 2, 3$, for ease of notation. We estimate $F_t^{\star}(y|x)$ by $\hat{F}_t(y|x)$ defined as follows.

Definition 3.3. *The NW estimator of the conditional CDF $F_t^{\star}(y|x)$ is defined as, for fixed $t \in \{1, \ldots, T\}$ and for all $y \in \mathbb{R}$,*

$$\hat{F}_t(y|x) = \sum_{a=1}^{T} \omega_a(\frac{t}{T}, x) H_h(y - Y_{a,T}), \tag{3.4}$$

where

$$\omega_a(\frac{t}{T}, x) = \frac{K_{h,1}(\frac{t}{T} - \frac{a}{T}) K_{h,2}(d(x, X_{a,T}))}{\sum_{a=1}^{T} K_{h,1}(\frac{t}{T} - \frac{a}{T}) K_{h,2}(d(x, X_{a,T}))}. \tag{3.5}$$

We note that the weights $\{\omega_a(u, x)\}_{a=1,\ldots,T}$ are measurable functions of x, $X_{a,T}$, and u, and do not depend on $Y_{a,T}$. Additionally, this estimator involves three kernel functions: K_1 and K_2 that allow smoothing with respect to the time direction and the space direction of the covariates $X_{t,T}$, respectively, and an integrated kernel H that acts as a CDF for the response variable $Y_{t,T}$. This allows us to analyze the local behavior of the data in the rescaled time $u = \frac{t}{T}$.

As a by-product, for $\alpha \in [0,1]$, we can define the conditional quantile of order α of the conditional CDF $F_t^\star(\cdot|x)$, as $q_\alpha(x) = \inf\{y \in \mathbb{R} : F_t^\star(y|x) \geq \alpha\}$. Since $F_t^\star(\cdot|x)$ is strictly increasing, the uniqueness of the quantile α is guaranteed. We estimate $q_\alpha(x)$ by $\hat{q}_\alpha(x)$ given below.

Definition 3.4. *Since H is strictly increasing, we can define the NW conditional quantile function estimator as*

$$\hat{q}_\alpha(x) = \inf\{y \in \mathbb{R} : \hat{F}_t(y|x) \geq \alpha\}. \tag{3.6}$$

Remark 3.1. *If $\alpha = 0.5$, (3.6) is defined as the NW conditional median estimator.*

We now bring forth the conditions upon which our main theoretical results are based.

3.1. Assumptions

We consider the following assumptions that are standard for local stationarity [14, 19, 38] and CDE [1, 2, 7, 39].

Assumption 3.1 (Local stationarity)**.** *The process of $\mathscr{H}$-value random variables $\{X_{t,T}\}$ for $t = 1, \ldots, T$ is assumed to be locally stationary and is compactly supported by S.*

Assumption 3.2 (Kernel functions)**.** *$K_i(\cdot)$ ($i = 1, 3$) is symmetric about zero, bounded, and has compact support, that is, $K_i(v) = 0$ for all $|v| > C_i$ for some $C_i < \infty$. On the other hand, $K_2(\cdot)$ is non-negative, bounded, and has compact support in $[0,1]$ such that $0 < K_2(0)$ and $K_2(1) = 0$. In addition, $K_2'(v) = \mathrm{d}K_2(v)/\mathrm{d}v$ exists on $[0,1]$, satisfying $C_1' \leq K_2'(v) \leq C_2'$, for real constants $-\infty < C_1' < C_2' < 0$. Moreover, $K_i(\cdot)$, for $i = 1, 2, 3$, is Lipschitz continuous, that is, $|K_i(v) - K_i(v')| \leq L_i|v - v'|$ for some $L_i < \infty$ and all $v, v' \in \mathbb{R}$ and satisfy the following:*

$$\int K_i(z)\mathrm{d}z = 1, \quad \int zK_1(z)\mathrm{d}z = 0 \quad \text{and} \quad \int zK_3(z)\mathrm{d}z = 0. \tag{3.7}$$

In addition, the integrated kernel $H(\cdot)$ is a strictly increasing CDF to the set $\{v \in \mathbb{R}, H(v) \in [0,1]\}$, positive, bounded, and Lipschitzian, satisfying, for $a > 0$,

$$\int H'(\omega)\mathrm{d}\omega = 1, \quad \text{and} \int |\omega|^a H'(\omega)\mathrm{d}\omega < \infty. \tag{3.8}$$

The kernels K_i, $i = 1, 2, 3$, and H are assumed to be compactly supported and Lipschitz, allowing us to obtain upper bounds. We also assume that all involved kernels are probability density functions. Additionally, we assume that $K_1(\cdot)$ and $K_3(\cdot)$ are symmetric around the origin and do not introduce first-order linear bias. Moreover, the conditions (3.8) signify that the integrated kernel H is a CDF with finite moments, which is important to show the convergence rate of the bias term.

Assumption 3.3 (Distribution function). *Let $B(x,h) = \{y \in \mathscr{H} : d(x,y) \le h\}$ denote a ball centered at $x \in \mathscr{H}$ with radius h. For all $u \in [0,1]$, $x \in \mathscr{H}$, and $h > 0$, there exists positive constants $c_d < C_d$, such that*

$$0 < c_d \phi(h) f_1(x) \le \mathbb{P}(X_t(u) \in B(x,h)) =: F_u(h;x) \le C_d \phi(h) f_1(x), \tag{3.9}$$

where $\phi(0) \to 0$, $\phi(u)$ is absolutely continuous in a neighborhood of the origin, and $f_1(x)$ is a nonnegative functional in $x \in \mathscr{H}$.

Assumption 3.4 (Regularity condition on h and $\phi(h)$). *Assume that the bandwidth h and the small ball probability $\phi(h)$ satisfy*

$$\lim_{T\to\infty} h = 0 \quad \textit{and} \quad \lim_{T\to\infty} \frac{\log T}{T h \phi(h)} = 0.$$

Equation (3.9) controls the behavior of the small ball probability around zero, a usual condition on the small ball probability. It can be approximately expressed as the product of two independent functions $\phi(\cdot)$ and $f_1(\cdot)$. This condition corresponds to the assumption made in [25]. On the other hand, the bandwidth h should converge more slowly to zero, for instance, at a polynomial rate, i.e., $h = \mathcal{O}(T^{-\xi})$, for small $\xi > 0$, as indicated in Assumption 3.4. This assumption is consistent with the condition made in [14] that $Th\phi(h) \to \infty$ and is needed to attain the resulting convergence rates. In addition, we suppose $\phi(h)$ converges to zero faster than h.

Assumption 3.5 (Conditional CDF). *The conditional CDF is Lipschitz continuous, that is,*

$$\left|F_a^\star(y|x) - F_t^\star(y'|x')\right| \le L_{F^\star}\big(\left|\frac{a}{T} - \frac{t}{T}\right| + |y-y'| + d(x,x')\big),$$

for some $L_{F^\star} < \infty$, and for all $a,t \in \{1,\dots,T\}$, $y,y' \in \mathbb{R}$, and $x,x' \in \mathscr{H}$.

Assumption 3.6 (Mixing condition). *The process $\{(X_{t,T}, \varepsilon_{t,T})\}$ is arithmetically β-mixing satisfying $\beta(k) \le Ak^{-\gamma}$ for some $A > 0$ and $\gamma > 2$. We also assume that for some $p > 2$ and $\zeta > 1 - \frac{2}{p}$,*

$$\sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} < \infty.$$

Assumption 3.5 states that the conditional CDF $F^\star_\cdot(\cdot|\cdot)$ behaves smoothly, and it changes slowly as the observation changes. Alternatively, one can assume that the conditional CDF is twice differentiable [1, 7]. Assumption 3.6 assumes that the process is β-mixing, a more robust type of independence between far-off observations in a process [31, 36]. It also highlights the decay of $\beta(k)$. We note that common time series models are known to be β-mixing [29, 40–42].

3.2. Uniform convergence

The asymptotic property of the NW conditional quantile function estimator is being studied by establishing its uniform convergence given by Theorem 3.1 below. This convergence rate depends on both the bandwidth h and the small ball probability $\phi(h)$, which is comparable to the rate of uniform convergence of the NW conditional mean function in [14] and the rates of convergence of the NW conditional quantile functions in [2, 4].

Theorem 3.1. *Let Assumptions 3.1 - 3.6 be satisfied, then, for* $I_h = [C_1 h, 1 - C_1 h]$,

$$\sup_{x\in S}\sup_{\frac{t}{T}\in I_h}\left|\hat{q}_\alpha(x) - q_\alpha(x)\right| = \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi(h)}} + h\right).$$

Proof. By Assumption 3.2, we safely say that $\hat{F}_t(y|x)$ is continuous and strictly increasing. Hence, its corresponding inverse function $\hat{q}_\alpha(x)$ exists and is continuous. So, we have, $\forall\epsilon > 0$, $\exists\delta(\epsilon) > 0$, such that

$$\sup_{x\in S}\sup_{\frac{t}{T}\in[0,1]}\left|\hat{F}_t(y|x) - \hat{F}_t(q_\alpha(x)|x)\right| \le \delta(\epsilon) \implies |y - q_\alpha(x)| \le \epsilon,$$

for all $y \in \mathbb{R}$, $x \in S$, and $\frac{t}{T} \in (0,1)$. Consequently, for $y = \hat{q}_\alpha(x)$: $\forall\epsilon > 0$, $\exists\delta(\epsilon) > 0$,

$$\begin{aligned}\mathbb{P}[\sup_{x\in S}\sup_{\frac{t}{T}\in[0,1]}|\hat{q}_\alpha(x) - q_\alpha(x)| > \epsilon] &\le \mathbb{P}[\sup_{x\in S}\sup_{\frac{t}{T}\in[0,1]}|\hat{F}_t(\hat{q}_\alpha(x)|x) - \hat{F}_t(q_\alpha(x)|x)| > \delta(\epsilon)] \\ &= \mathbb{P}[\sup_{x\in S}\sup_{\frac{t}{T}\in[0,1]}|F_t^\star(q_\alpha(x)|x) - \hat{F}_t(q_\alpha(x)|x)| > \delta(\epsilon)],\end{aligned}$$

since $\hat{F}_t(\hat{q}_\alpha(x)|x) = \alpha = F_t^\star(q_\alpha(x)|x)$ by Definition 3.4. Thus, the uniform convergence of $\left|\hat{q}_\alpha(x) - q_\alpha(x)\right|$ is obtained from the uniform convergence of $|\hat{F}_t(y|x) - F_t^\star(y|x)|$. Observe that we can decompose $\hat{F}_t(y|x) - F_t^\star(y|x)$ as

$$\begin{aligned}\hat{F}_t(y|x) - F_t^\star(y|x) &= \frac{\hat{F}_t^N(y|x)}{\hat{F}_t^D(y|x)} - F_t^\star(y|x) = \frac{\hat{F}_t^N(y|x) - F_t^\star(y|x)\hat{F}_t^D(y|x)}{\hat{F}_t^D(y|x)} \\ &= \frac{1}{\hat{F}_t^D(y|x)}\{\hat{F}_t^N(y|x) - \mathbb{E}[\hat{F}_t^N(y|x)] + \mathbb{E}[\hat{F}_t^N(y|x)] - F_t^\star(y|x)\hat{F}_t^D(y|x)\},\end{aligned}$$

where

$$\hat{F}_t^N(y|x) = \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T} - \frac{a}{T})K_{h,2}(d(x, X_{a,T}))H_h(y - Y_{a,T}),$$

and

$$\hat{F}_t^D(y|x) = \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T} - \frac{a}{T})K_{h,2}(d(x, X_{a,T})).$$

The proof is then completed using the following lemmas whose proofs are deferred to Appendix A.

Lemma 3.1. *Assume that Assumptions 3.1 - 3.6 hold. Then*

$$\sup_{x\in S}\sup_{\frac{t}{T}\in I_h}|\hat{F}_t^N(y|x)-\mathbb{E}[\hat{F}_t^N(y|x)]|=\mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi(h)}}\right).$$

Lemma 3.2. *Let Assumptions 3.1 - 3.5 be satisfied. Then*

$$\sup_{x\in S}\sup_{\frac{t}{T}\in I_h}|\mathbb{E}[\hat{F}_t^N(y|x)]-F_t^\star(y|x)\hat{F}_t^D(y|x)|=\mathcal{O}(h).$$

Lemma 3.3. *Assume that we satisfy Assumptions 3.1 - 3.4, and Assumption 3.6. Then*

$$1/\inf_{x\in S}\inf_{\frac{t}{T}\in I_h}\hat{F}_t^D(y|x)=\mathcal{O}_{\mathbb{P}}(1).$$

3.3. Leave-one-out cross-validation bandwidth selection criterion

Since the uniform convergence we have depends on the bandwidth, we propose in this section a method to select h. In nonparametric kernel estimation, especially in local averaging, the bandwidth must be suitably selected for the estimator to perform well. Bandwidth selection methods have already been established and developed in [43, 44]. This paper considers the leave-one-out cross-validation procedure used in [43, 45]. For any fixed $i\in\{1,\ldots,T\}$, we define

$$\hat{m}_i(\frac{t}{T},x)=\sum_{a=1;a\neq i}^{T}\omega_a(\frac{t}{T},x)Y_{a,T}, \tag{3.10}$$

where $\omega_a(\frac{t}{T},x)$ is given by (3.5). Equation (3.10) is regarded as the leave-out-$(X_{i,T},Y_{i,T})$ estimator of $m_i^\star(\frac{t}{T},x)$. To minimize the quadratic loss function, we introduce the following leave-one-out cross-validation (LOOCV) criterion

$$CV(y,x,h):=\frac{1}{T}\sum_{i=1}^{T}\big(Y_{i,T}-\hat{m}_i(\frac{t}{T},x)\big)^2. \tag{3.11}$$

As highlighted in [43], we choose a bandwidth $\hat{h}$ among $h\in[a_T,b_T]$ that minimizes (3.11). As shown in Theorem 3.8 of [46], the ratio $\hat{h}/h_0$ converges to one, where $h_0\asymp T^{-1/(d+4)}$, and thus satisfies Assumption 3.4. Consequently, bandwidths chosen via this cross-validation procedure are guaranteed to obey the convergence rates established in Section 3.2.

4. Numerical experiments on simulated data

To demonstrate the consistency of the proposed NW estimator, we use an example provided in [13, 20] to simulate $(X_{t,T}, Y_{t,T})_{t=1,\dots,T}$. We generate $X_{t,T}$ from a Hilbert space $\mathscr{H} = L^2_{\mathbb{R}}([0,1])$, using:

Gaussian tvFAR(1). Consider the time-varying functional autoregressive process of order 1, tvFAR(1), with Gaussian noise represented by

$$X_{t,T}(\tau) = B_{t/T}(X_{t-1,T})(\tau) + \eta_t(\tau), \quad \tau \in [0,1], \quad t = 1,\dots,T,$$

where $B_{t/T}$ is a linear operator indexed by the rescaled time $u = \frac{t}{T}$ and η_t is a linear combination of the Fourier basis function $(\psi_i)_{i\in\mathbb{N}}$ and generated from a Gaussian distribution with zero mean and variance with ith coefficient $(\pi(i-1.5))^{-2}$. We approximate the functional covariates by

$$\boldsymbol{X}_{t,T} \approx \boldsymbol{B}_{t/T}\boldsymbol{X}_{t-1,T} + \boldsymbol{\eta}_t, \quad t = 1,\dots,T,$$

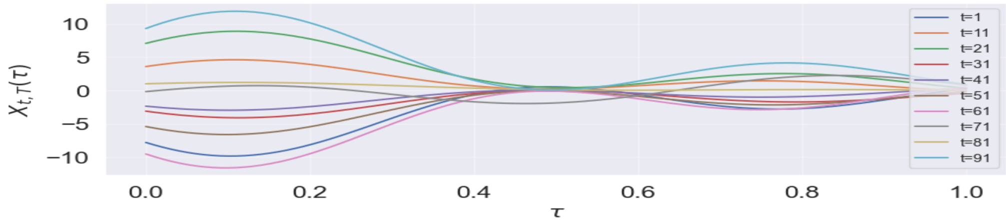
where $\boldsymbol{X}_{t,T} = (\langle X_{t,T}, \psi_1\rangle, \dots, \langle X_{t,T}, \psi_j\rangle)'$, $\boldsymbol{\eta}_t = (\langle \eta_t, \psi_1\rangle, \dots, \langle \eta_t, \psi_j\rangle)'$, and for $1 \le i, j \le J$, $\boldsymbol{B}_{t/T} = (\langle B_{t/T}(\psi_i), \psi_j\rangle)$. The matrix $\boldsymbol{B}_{t/T}$ is set as $\boldsymbol{B}_{t/T} = \frac{0.4\boldsymbol{A}_{t/T}}{\|\boldsymbol{A}_{t/T}\|_\infty}$, where $\boldsymbol{A}_{t/T}$ is a $J \times J$ matrix with entries $A_{t/T}(i,j)$ that are mutually independent Gaussian random variables with zero mean and variance $\frac{t}{Ti^6} + (1-\frac{t}{T})e^{-j-i}$, and $\|\cdot\|_\infty$ is a Schatten ∞-norm given by $\|A\|_\infty = \sup_{\|x\|\le 1} \|Ax\|$. Figure 1 shows the plot of $X_{t,T}(\tau)$ at $T = 100$ using $N = 100$ discretized τ from $[0,1]$ and lower value of $J = 7$, since results do not vary much wrt J [20].

Given the functional covariate above, the response variable $Y_{t,T}$ is generated using (2.1) where $\varepsilon_{t,T} \sim \mathcal{N}[0,1]$ and

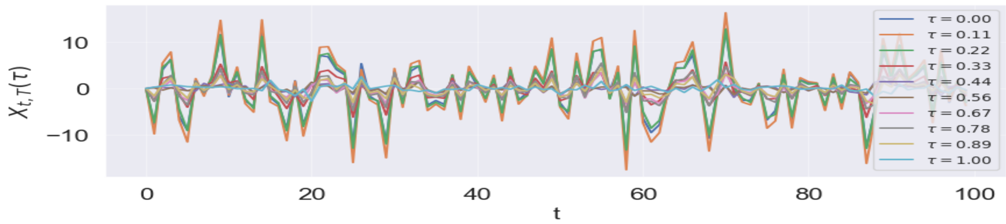
$$m^\star(u,x) = 2.5\sin(2\pi u)\int_0^1 \cos(\pi x(\tau))\mathrm{d}\tau.$$

Figure 2 shows the time plots of $Y_{t,T}$ for $T = 1000$, whose values remain tight with constant mean.

We model the generated $Y_{t,T}$ with its corresponding covariate $X_{t,T}$ using the NW conditional median $\hat{Y}_{t,T} = \hat{q}_{0.5,t}(x)$ using (3.6) with $\alpha = 0.5$. We utilized uniform, triangle, and Gaussian kernels for $K_1(\cdot)$, $K_2(\cdot)$, and $K_3(\cdot)$, respectively. We chose the bandwidth h using a leave-one-out cross-validation method proposed in Section 3.3. To evaluate the performance of this estimation procedure, we calculate the mean absolute error (MAE) between $\hat{Y}_{t,T}$ and $Y_{t,T}$. We show the result of this modeling in Figure 3, where we plot the fitted values of $\hat{Y}_{t,T}$ over the actual values of Gaussian tvFAR(1) $Y_{t,T}$. As depicted, the NW estimator accurately fits the behavior of $Y_{t,T}$, producing an almost negligible error (MAE = 0.000531). We then show its consistency using a Monte Carlo simulation described below.



(a) Behavior of $X_{t,T}(\tau)$ at t given some τ



(b) Behavior of $X_{t,T}(\tau)$ at τ given some t

Figure 1: Realizations of Gaussian tvFAR(1) covariates $X_{t,T}(\tau)$ for some $t \in [1, \ldots, T]$ and $\tau \in [0, 1]$ for $T = 100$ with $J = 7$ and $N = 100$ discretization points of $\tau \in [0, 1]$.

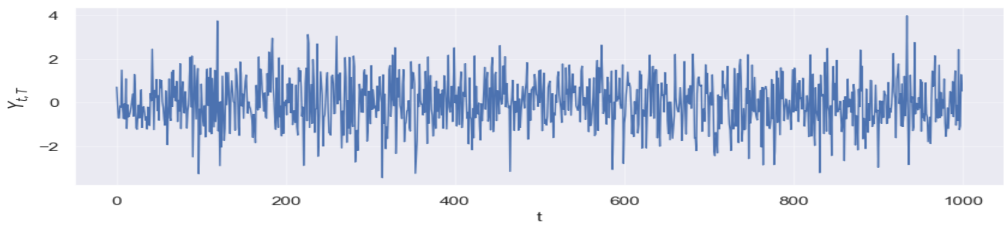


Figure 2: Realizations of Gaussian tvFAR(1) response $Y_{t,T}$ for $T = 1000$.

Monte Carlo Simulation. To illustrate the convergence of the proposed NW estimator for increasing sample size T, we replicate the Gaussian tvFAR(1) functional pairs $(X_{t,T}, Y_{t,T})$ $L = 1000$ times and calculate the corresponding NW conditional median $\hat{Y}_{t,T} = \hat{q}_{0.5,t}(x)$. In this experiment, we set $T = 500, 1000, 3000, 5000, 8000, 10000$. To calculate $\hat{Y}_{t,T}$, we similarly set tricube, triangle, and Gaussian kernel functions for K_1, K_2, and K_3, respectively. We also employed a leave-one-out cross-validation method to select the bandwidths h. Figure 4 reflects the resulting boxplots between the actual values of a Gaussian tvFAR(1) process $Y_{t,T}$ and the fitted values of the NW conditional median $\hat{Y}_{t,T}$. The overall mean values of these MAEs are $0.016049, 0.012940, 0.009749, 0.003657, 0.000723, 0.000188$ for $T = 500, 1000, 3000, 5000, 8000, 10000$, respectively. These values are comparably lower than the recorded mean squares errors (MSEs) in [20], where the authors employed a single index θ in estimating the conditional mean function of the same Gaussian tvFAR(1) simulated data. These findings indicate that the estimation accuracy of the NW conditional median improves with larger sample sizes, as evidenced by decreasing MAE and narrower variability across replications, thereby validating our theoretical result.

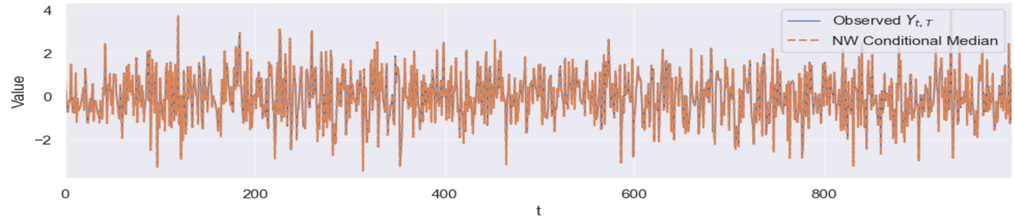


Figure 3: Gaussian tvFAR(1) response $Y_{t,T}$ with fitted values of NW conditional median $\hat{Y}_{t,T}$ for $T = 1000$, using $K_1 =$ tricube, $K_2 =$ triangle, and $K_3 =$ Gaussian; the bandwidth h is selected using a leave-one-out cross-validation method; MAE $= 0.000531$.

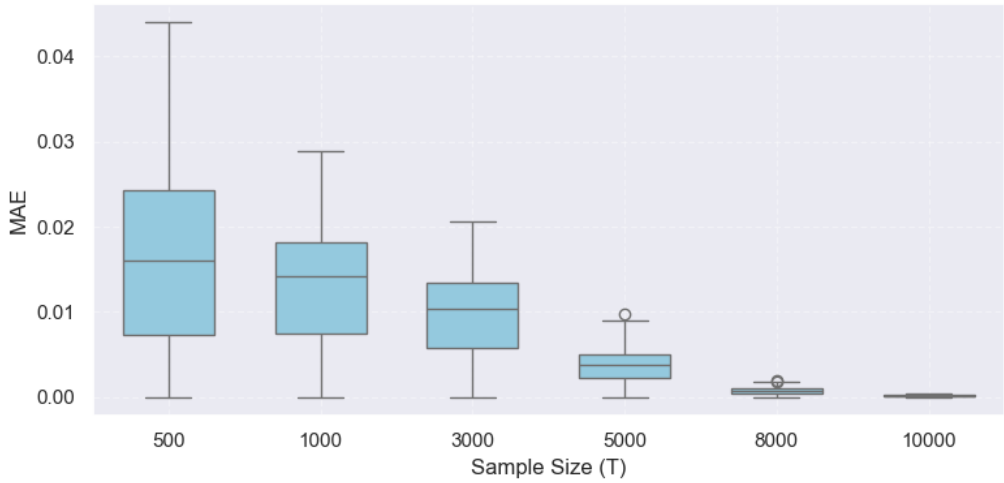


Figure 4: MAEs between the actual values of Gaussian tvFAR(1) $Y_{t,T}$ and the fitted values of NW conditional median $\hat{Y}_{t,T}$ for increasing $T = 500, 1000, 3000, 5000, 8000, 10000$, and $L = 1000$ replications; $K_1 =$ tricube, $K_2 =$ triangle, and $K_3 =$ Gaussian; bandwidth h is selected using a leave-one-out cross-validation method.

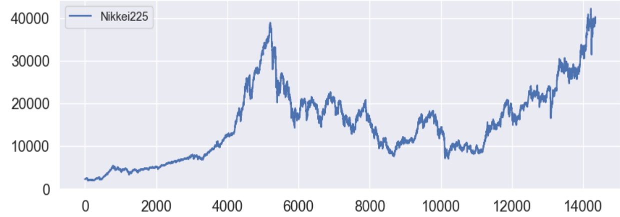
5. Application to stock market data

In this illustration, we use the Nikkei Stock Market Index dataset (Nikkei 225), available at `https://fred.stlouisfed.org/series/NIKKEI225`. It serves as a key indicator of the Japanese stock market's overall status by tracking the performance of 225 large and active companies listed on the Tokyo Stock Exchange [47]. It contains 14340 data points from January 14, 1971, to December 31, 2024, plotted in Figure 5a.

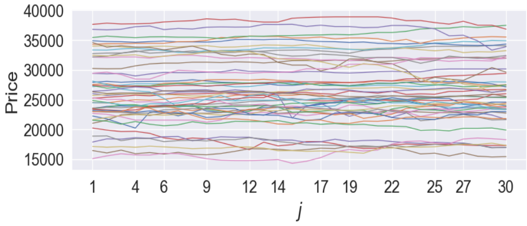
To treat this dataset, we employ the procedure used in [2, 39]. That is, the functional covariate $X_{t,T}$ is generated by segmenting the original time series $\{Z(s)\}_{s=1,\ldots,14340}$ by 30 observations, i.e., we construct 478 continuous sample curves:

$$z_{t,T}(j) = Z(30(t-1)+j), \quad \forall j \in \{1,\ldots,30\}.$$

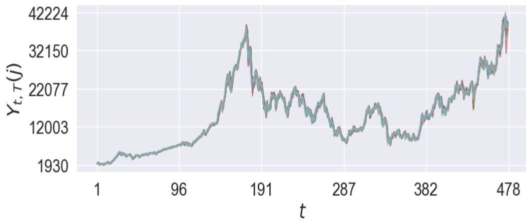
The covariates are then constructed as $X_{t,478} = (z_{t,76}(1),\ldots,z_{t,478}(30))$. The response is



(a) Original time series $Z(s)$ $(n = 14340)$



(b) Continuous sample curves $X_{t,T}$; $t = 150, \ldots, 199$



(c) Response $Y_{t,477}(j)$; for all $j = 1, \ldots, 30$

Figure 5: Actual observations $Z(s)$, functional covariate $X_{t,T}$, and response $Y_{t,T}(j)$ of Nikkei 225 dataset from Jan. 14, 1970 - Dec. 31, 2024

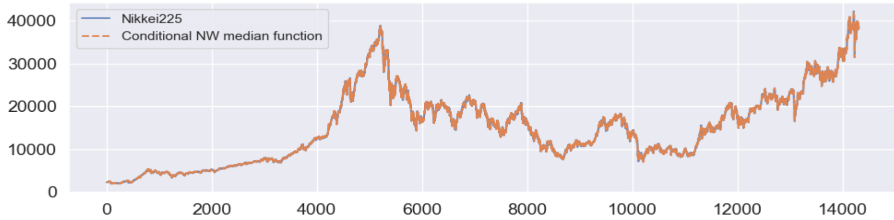
constructed as

$$Y_{t,T}(j) = z_{t+1,T}(j) = Z(30t + j).$$

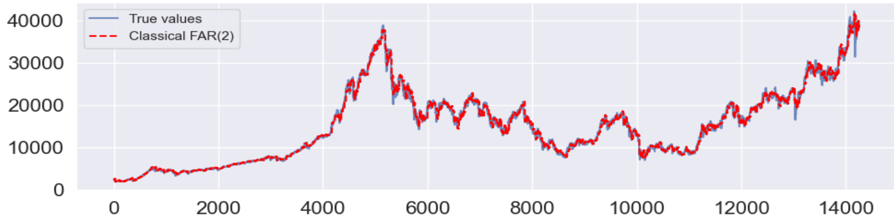
With this construction, we then generate new functional pairs $\{(X_{t,477}, Y_{t,477}(j))\}_{t=1,\ldots,477}$, $j \in \{1, \ldots, 30\}$. Figure 5b reflects 50 examples of the generated continuous sample curves. The behavior of the response variable $Y_{t,T}(j)$ is plotted in Figure 5c.

Similar to the numerical experiment above, we model Nikkei 225 by calculating the NW conditional median $\hat{Y}_{t,T}(j) = \hat{q}_{0.5,t}(x)$ using (3.6) with $\alpha = 0.5$. We use the functional pairs $\{(X_{t,477}, Y_{t,477}(j))\}_{t=1,\ldots,477}$, $j \in [1, \ldots, 30]$ and set a tricube kernel for K_1, triangle for K_2, and Gaussian kernel for K_3. We again used a leave-one-out cross-validation procedure to select the bandwidth h. As a parametric counterpart to the nonparametric NW estimator, we fit a classical functional autoregressive model of order FAR(p). Unlike the locally adaptive, kernel-based NW approach, FAR(p) assumes linear dependence in a functional space. Curves are first reduced via functional principal component analysis, the resulting scores are modeled with a VAR(p), and one-step-ahead forecasts are obtained by reconstructing curves from the predicted scores.

As illustrated in Figure 6, both the proposed NW conditional median estimator and the FAR(2) model track the overall dynamics of the observed Nikkei 225 series $Y_{t,T}$. However, the NW estimator provides a substantially more accurate fit, achieving a much lower error (MAE = 0.47643032) than the FAR(2) model (MAE = 497.52067804). This contrast highlights the advantage of the locally adaptive, nonparametric NW approach over the global linear FAR(2) specification for capturing complex dynamics in the Nikkei 225.



(a) Fitted values of the NW conditional median $\hat{Y}_{t,T}$ using K_1 = Tricube, K_2 = Triangle, and K_3 = Gaussian; h is selected using a leave-one-out cross-validation procedure; MAE = 0.47643032.



(b) Fitted values of the classical FAR(p) model, with p = 2; **MAE** = 497.52067804.

Figure 6: Actual values of Nikkei 225 fitted with estimates generated from using NW conditional median estimator and the classical FAR(p) model.

Additionally, the NW conditional median estimator outperforms the model used in analyzing Nikkei 225 in [47]. Note that, as a major Asian index, movements in the Nikkei 225 often influence other international markets. We have demonstrated a more efficient method for analyzing such data, which is essential for both domestic and global financial decision-making. Overall, the results indicate that the NW conditional median estimator is more effective for modeling LSFTS arising from stock market data and related financial applications.

6. Conclusion

This study develops and investigates a Nadaraya–Watson (NW) estimator for the conditional quantile function of locally stationary functional time series (LSFTS) with a scalar response and a functional covariate. Theoretical results establish the uniform convergence of the proposed estimator under suitable regularity conditions, ensuring its asymptotic reliability. Through simulation experiments, the estimator demonstrates strong finite-sample performance and consistency. Furthermore, the analysis of the Nikkei 225 dataset confirms its practical applicability for modeling complex dynamic processes. Overall, the proposed NW estimation framework provides a flexible and robust nonparametric approach for conditional inference in LSFTS, with potential extensions to broader functional and time-dependent data settings.

Acknowledgements

The author extends his appreciation to the support provided by the Department of Science and Technology - Science Education Institute (DOST-SEI) and Campus France through the PhilFrance-DOST Scholarship grant.

References

[1] P. Hall, R. C. L. Wolff, and Q. Yao. Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94:154–163, 1999.

[2] F. Ferraty, A. Rabhi, and P. Vieu. Conditional quantiles for dependent functional data with application to the climatic "El Niño" phenomenon. *Sankhyā: The Indian Journal of Statistics*, pages 378–398, 2005.

[3] O. Bouanani and S. Bouzebda. Limit theorems for local polynomial estimation of regression for functional dependent data. *AIMS Math.*, 9(9):23651–23691, 2024.

[4] S. Attaoui and N. Ling. Asymptotic results of a nonparametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications. *Metrika*, 79:485–511, 2016.

[5] O. Bouanani, A. Laksaci, M. Rachdi, and S. Rahmani. Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics. *Behaviormetrika*, 46:199–233, 2018.

[6] O. Bouanani, S. Rahmani, and L. Ait-Hennani. Local linear conditional cumulative distribution function with mixing data. *Arabian Journal of Mathematics*, 9:289–307, 2019.

[7] H.I.E.S. Ahmed, B. Salha, R, and H.O. EL-Sayed. Adaptive weighted nadaraya-watson estimation of the conditional quantiles by varying bandwidth. *Communications in Statistics - Simulation and Computation*, 49(5):1105–1117, 2020.

[8] A. Aue, D. D. Nourinho, and S. Hörmann. On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.*, 110:378–392, 2015.

[9] S.X. Chen, L. Lei, and Y. Tu. Functional coefficient moving average model with applications to forecasting Chinese CPI. *Statistica Sinica*, 26:1649–1672, 2016.

[10] F. A. Bugni, P. Hall, J. Horowitz, and G. R. Neumann. Goodness-of-fit tests for functional data. *The Econometrics Journal*, 12:S1–S18, 2009.

[11] A. Aue and A. van Delft. Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.*, 48:2505–2547, 2020.

[12] S. Hörmann, T. Kuenzer, and G. Rice. Estimating the conditional distribution in functional regression problems. *Electronic Journal of Statistics*, 16:5751–5778, 2022.

[13] A. van Delft and M. Eichler. Locally stationary functional time series. *Electronic Journal of Statistics*, 12:107–170, 2018.

[14] D. Kurisu. Nonparametric regression for locally stationary functional time series. *Electronic Journal of Statistics*, 16:3973–3995, 2022.

[15] I. Soukarieh and S. Bouzebda. Weak convergence of the conditional U-statistics for

locally stationary functional time series. *Statistical Inference for Stochastic Processes*, 27(2):227–304, 2024.

[16] R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37, 1996.

[17] R. Dahlhaus and S. Subba Rao. Statistical inference for time-varying ARCH processes. *Ann. Statist.*, 34:1075–1114, 2006.

[18] R. Dahlhaus. Locally stationary processes. In Tata Subba Rao, Suhasini Subba Rao, and C.R. Rao, editors, *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, pages 351–413. Elsevier, 2012.

[19] D. Kurisu, R. Fukami, and Y. Koike. Adaptive deep learning for nonlinear time series models. *Bernoulli*, 31(1):240–270, 2025.

[20] B. M. Agua and S. Bouzebda. Single index regression for locally stationary functional time series. *AIMS Mathematics*, 9(12):36202–36258, 2024.

[21] È. A. Nadaraya. On a regression estimate. *Teor. Verojatnost. i Primenen.*, 9:157–159, 1964.

[22] G. S. Watson. Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372, 1964.

[23] A. van Delft and H. Dette. A general framework to quantify deviations from structural assumptions in the analysis of nonstationary function-valued processes. *The Annals of Statistics*, 52(2):550 – 579, 2024.

[24] F. Ferraty and P. Vieu. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *J Nonparametr Stat*, 16(1-2):111–125, 2004.

[25] T. Gasser, P. Hall, and B.. Presnell. Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc. Ser. B.*, 60:681–691, 1998.

[26] F. Ferraty, N. Kudraszow, and P. Vieu. Nonparametric estimation of a surrogate density function in infinite dimensional spaces. *J Nonparametr Stat*, 24(2):447–464, 2012.

[27] E.G. Bongiorno and A. Goia. Classification methods for Hilbert data based on surrogate density. *Comput Stat Data Anal*, 99:204–222, 2016.

[28] E.G. Bongiorno, A. Goia, and P. Vieu. Evaluating the complexity of some families of functional data. *SORT*, 42(1):27–44, 2018.

[29] P. Doukhan. *Mixing: properties and examples.* Springer-Verlag New York, 1994.

[30] M.E. Ahsen and M. Vidyasagar. Mixing coefficients between discrete and real random variables: Computation and properties. *IEEE Transactions on Automatic Control*, 59(1):34–47, 2014.

[31] E. Rio. *Asymptotic theory of weakly dependent random processes.* Springer, Berlin Heidelberg, 2017.

[32] M. Peligrad. Some remarks on coupling of dependent random variables. *Statistics and Probability Letters*, 60(2):201–209, 2002.

[33] J. Dedecker and C. Prieur. New dependence coefficients: examples and applications to statistics. *Probab. Theory Relat. Fields*, 132:203–236, 2005.

[34] J. Dedecker, P. Doukhan, G. Lang, J.R. Leon, S. Louhichi, and C. Prieur. *Weak dependence: with examples and applications.* Springer Science and Business Media,

2007.

[35] D. Bosq. *Nonparametric statistics for stochastic processes: estimation and prediction.* Springer Science and Business Media, 2012.

[36] A. Poinas. A bound of the β-mixing coefficient for point processes in terms of their intensity functions. *Statist. Probab. Lett.*, 148:88–93, 2019.

[37] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis.* Springer, 233 Spring Street, New York,NY10013, USA, 2006.

[38] M. Vogt. Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40(5):2601–2633, 2012.

[39] N. Ling, L. Cheng, P. Vieu, and H. Ding. Missing responses at random in functional single index model for time series data. *Statistical Papers*, 63:665–692, 2022.

[40] D.J. McDonald, C.R. Shalizi, and M. Schervish. Estimating beta-mixing coefficients. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15, Fort Lauderdale, FL, USA, 2011. JMLR: W and CP.

[41] A. Mokkadem. Mixing properties of ARMA processes. *Stochastic processes and their applications*, 29:309–315, 1998.

[42] M. Carrasco and X Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002.

[43] M. Rachdi and P. Vieu. Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, 137(9):2784–2801, 2007.

[44] S. Bouzebda and A. Nezzal. Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data. *Jpn. J. Stat. Data Sci.*, 5(2):431–533, 2022.

[45] K. Benhenni, F. Ferraty, M. Rachdi, and P. Vieu. Local smoothing regression with functional data. *Comput. Statist.*, 22(3):353–369, 2007.

[46] S. Richter and R. Dahlhaus. Cross validation for locally stationary processes. *Ann. Statist.*, 47(4):2145–2173, 2019.

[47] F. Björgvinsson. Introducing a novel method to identify the future trend of nikkei 225 stock price in order to reduce investment risk. *Journal of Artificial Intelligence and System Modelling*, 2(01):60–73, 2024.

[48] E. Masry. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications*, 115:155–177, 2005.

A. Proofs of the main results

A.1. Proof of Lemma 3.1

Proof. Set $B = [0,1]$, $a_T = \sqrt{\log T / Th\phi(h)}$, and $\tau_T = \rho_T T^{1/\zeta}$ with $\rho_T = (\log T)^{\zeta_0}$ for some $\zeta_0 > 0$. Now, we consider the decomposition below

$$\hat{F}_t^N(y|x) - \mathbb{E}[\hat{F}_t^N(y|x)] = (\hat{F}_{t,1}^N(y|x) - \mathbb{E}[\hat{F}_{t,1}^N(y|x)]) + (\hat{F}_{t,2}^N(y|x) - \mathbb{E}[\hat{F}_{t,2}^N(y|x)]),$$

where

$$\hat{F}_{t,1}^N(y|x) = \frac{1}{Th\phi(h)} \sum_{a=1}^{T} K_{h,1}(u - \frac{a}{T}) K_{h,2}(d(x, X_{a,T})) H_h(y - Y_{a,T}) \mathbb{1}_{|H_h(y-Y_{a,T})| \le \tau_T}; \quad \text{and}$$

$$\hat{F}_{t,2}^N(y|x) = \frac{1}{Th\phi(h)} \sum_{a=1}^{T} K_{h,1}(u - \frac{a}{T}) K_{h,2}(d(x, X_{a,T})) H_h(y - Y_{a,T}) \mathbb{1}_{|H_h(y-Y_{a,T})| > \tau_T}.$$

Then, the next steps of the proof are outlined as:

1. $\sup_{x\in S}\sup_{u\in B} |\hat{F}_{t,2}^N(y|x) - \mathbb{E}[\hat{F}_{t,2}^N(y|x)]| = \mathcal{O}_{\mathbb{P}}(a_T)$; and
2. $\sup_{x\in S}\sup_{u\in B} |\hat{F}_{t,1}^N(y|x) - \mathbb{E}[\hat{F}_{t,1}^N(y|x)]| = \mathcal{O}_{\mathbb{P}}(a_T)$.

Step 1. Observe that

$$\begin{aligned}
&\mathbb{P}\big[\sup_{x\in S}\sup_{u\in B} |\hat{F}_{t,2}^N(y|x)| > a_T\big] \\
&= \mathbb{P}\Big[\Big\{\sup_{x\in S}\sup_{u\in B} |\hat{F}_{t,2}^N(y|x)| > a_T\Big\} \cup \\
&\qquad \Big\{\Big\{\sup_{x\in S}\sup_{u\in B} \bigcup_{a=1}^{T} |H_h(y - Y_{a,T})| > \tau_T\Big\} \cap \Big\{\sup_{x\in S}\sup_{u\in B} \bigcup_{a=1}^{T} |H_h(y - Y_{a,T})| > \tau_T\Big\}^c\Big\}\Big] \\
&\le \mathbb{P}\Big[\Big\{\sup_{x\in S}\sup_{u\in B} |\hat{F}_{t,2}^N(y|x)| > a_T\Big\} \cap \Big\{\sup_{x\in S}\sup_{u\in B} \bigcup_{a=1}^{T} |H_h(y - Y_{a,T})| > \tau_T\Big\}\Big] \\
&\qquad + \mathbb{P}\Big[\Big\{\sup_{x\in S}\sup_{u\in B} |\hat{F}_{t,2}^N(y|x)| > a_T\Big\} \cap \Big\{\sup_{x\in S}\sup_{u\in B} \bigcup_{a=1}^{T} |H_h(y - Y_{a,T})| > \tau_T\Big\}^c\Big] \\
&\le \mathbb{P}\big[\sup_{x\in S}\sup_{u\in B} |H_h(y - Y_{a,T})| > \tau_T\big] + \mathbb{P}[\emptyset] \quad \text{for some } a = 1, \ldots, T \\
&\le \tau_T^{-\zeta} \sum_{a=1}^{T} \mathbb{E}\big[\sup_{x\in S}\sup_{u\in B} |H_h(y - Y_{a,T})|^\zeta\big] \le \tau_T^{-\zeta} T = \rho_T^{-\zeta} \to 0 \text{ as } T \to \infty.
\end{aligned}$$

Now, let us consider

$$\mathbb{E}[|\hat{F}_{t,2}^N(y|x)|] \le \frac{1}{Th\phi(h)} \sum_{a=1}^{T} K_{h,1}(u - \frac{a}{T}) \mathbb{E}\big[K_{h,2}(d(x, X_{a,T})) |H_h(y - Y_{a,T})| \mathbb{1}_{|H_h(y-Y_{a,T})| > \tau_T}\big].$$

Note that

$$\begin{aligned}K_{h,2}(d(x,X_{a,T})) &\leq \Big|K_{h,2}(d(x,X_{a,T})) - K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\Big| + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\\ &= \Big|K_2\Big(\frac{d(x,X_{a,T})}{h}\Big) - K_2\Big(\frac{d\big(x,X_a\big(\frac{a}{T}\big)\big)}{h}\Big)\Big| + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\\ &\leq h^{-1}\Big|d(x,X_{a,T}) - d\big(x,X_a\big(\frac{a}{T}\big)\big)\Big| + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\\ &\leq h^{-1}d\big(X_{a,T},X_a\big(\frac{a}{T}\big)\big) + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\\ &\leq \frac{1}{Th}U_{a,T}\big(\frac{a}{T}\big) + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big). \qquad \text{(A.12)}\end{aligned}$$

So,

$$\begin{aligned}&\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))|H_h(y-Y_{a,T})|\mathbb{1}_{|H_h(y-Y_{a,T})|>\tau_T}\big]\\ &\lesssim \tau_T^{-(\zeta-1)}\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))|H_h(y-Y_{a,T})|^{\zeta}\big]\\ &\lesssim \tau_T^{-(\zeta-1)}\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))\big]\\ &\leq \tau_T^{-(\zeta-1)}\mathbb{E}\big[\frac{1}{Th}U_{a,T}\big(\frac{a}{T}\big) + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big]\\ &\lesssim \frac{1}{Th\tau_T^{\zeta-1}}\mathbb{E}\big[\frac{1}{Th}U_{a,T}\big(\frac{a}{T}\big)\big] + \tau_T^{-(\zeta-1)}\mathbb{E}\big[K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big]\\ &\lesssim \frac{1}{Th\tau_T^{\zeta-1}} + \tau_T^{-(\zeta-1)}\mathbb{E}\Big[\mathbb{1}_{\left(d\left(x,X_a\left(\frac{a}{T}\right)\right)\right)\leq h}\Big]\\ &\lesssim \frac{1}{Th\tau_T^{\zeta-1}} + \tau_T^{-(\zeta-1)}F_{a/T}(h;x) \lesssim \frac{1}{\tau_T^{\zeta-1}}\phi(h).\end{aligned}$$

Then, we have

$$\begin{aligned}\mathbb{E}[|\hat{F}^N_{t,2}(y|x)|] &\lesssim \frac{1}{\tau_T^{\zeta-1}}\phi(h)\frac{1}{Th\phi(h)}\sum_{a=1}^{T}K_{h,1}(u-\frac{a}{T})\\ &\lesssim \frac{1}{\tau_T^{\zeta-1}}\underbrace{\frac{1}{Th}\sum_{a=1}^{T}K_{h,1}(u-\frac{a}{T})}_{\mathcal{O}(1)}\\ &\lesssim \frac{1}{\tau_T^{\zeta-1}} = \rho_T^{-(\zeta-1)}T^{-\frac{\zeta-1}{\zeta}} \lesssim a_T.\end{aligned}$$

In the lines above, $\frac{1}{Th}\sum_{a=1}^{T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big) = \mathcal{O}(1)$ since, Lemma B.2 in [14], for $I_h = [C_1h, 1-C_1h]$,

$$\frac{1}{Th}\sum_{a=1}^{T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big) \leq \sup_{u\in I_h}\Big|\frac{1}{Th}\sum_{a=1}^{T}K_{h,1}\big(u-\frac{a}{T}\big)\Big|$$

$$\leq \sup_{u\in I_h}\Big|\frac{1}{Th}\sum_{a=1}^{T}K_{h,1}\big(u-\frac{a}{T}\big)-1\Big|+1$$
$$=\mathcal{O}\Big(\frac{1}{Th^2}\Big)+o(h)+1=\mathcal{O}(1).$$

Hence

$$\sup_{x\in S}\sup_{u\in B}|\hat{F}^N_{t,2}(y|x)-\mathbb{E}[\hat{F}^N_{t,2}(y|x)]|=\mathcal{O}_{\mathbb{P}}(a_T).$$

Step 2. Now, we need to show $\sup_{x\in S}\sup_{u\in B}|\hat{F}^N_{t,1}(y|x)-\mathbb{E}[\hat{F}^N_{t,1}(y|x)]|=\mathcal{O}_{\mathbb{P}}(a_T)$. First, we assume that S is a compact support of $\mathcal{H}$. Suppose that $N_{S,T}$ is the minimal number of balls in $\mathcal{H}$ needed to cover S, i.e.,

$$N_{S,T}\leq C\frac{1}{a_T}\quad \text{balls } B(z,h)=\{z\in\mathcal{H}:d(x,z)\leq h\},$$

with radius h and centers $x_1,\dots,x_{N_{S,T}}$. We also suppose that we cover the region B with

$$N_{B_{i,T}}\leq C\frac{1}{ha_T}\quad \text{balls } B_{i,T}=\{u\in\mathbb{R}:|u-u_i|\leq a_Th\},$$

with $u_i=\frac{t_i}{T}$ as the midpoint of $B_{i,T}$. Additionally, we assume that for $(w,v)\in\mathbb{R}^2$,

$$K^*(w,v)=C\mathbb{1}_{|w|\leq 2C_1}K_2(v).$$

Now, for $u\in B_{i,T}$ and sufficiently large T, we have

$$\Big|K_{h,1}\Big(u-\frac{a}{T}-K_{h,1}\Big(u_i-\frac{a}{T}\Big)\Big|K_{h,2}(d(x,X_{a,T}))\leq a_TK^*_h\Big(u_i-\frac{a}{T},d(x,X_{a,T})\Big),$$

where $K^*_h(v)=K^*(v/h)$. Now, let us define

$$\bar{F}^N_{t,1}(y|x)=\frac{1}{Th\phi(h)}\sum_{a=1}^{T}K^*_h\Big(u-\frac{a}{T},d(x,X_{a,T})\Big)|H_h(y-Y_{a,T})|\mathbb{1}_{|H_h(y-Y_{a,T})|\leq\tau_T}.$$

Note that for sufficiently large M, $\mathbb{E}[|\bar{F}^N_{t,1}(y|x)|]\leq M<\infty$. Then we get

$$\begin{aligned}&\sup_{x\in S}\sup_{u\in B_{i,T}}|\hat{F}^N_{t,1}(y|x)-\mathbb{E}[\hat{F}^N_{t,1}(y|x)]|\\&\leq|\hat{F}^N_{t_i,1}(y|x)-\mathbb{E}[\hat{F}^N_{t_i,1}(y|x)]|+a_T\big(|\bar{F}^N_{t_i,1}(y|x)|+\mathbb{E}[|\bar{F}^N_{t_i,1}(y|x)|]\big)\\&\leq|\hat{F}^N_{t_i,1}(y|x)-\mathbb{E}[\hat{F}^N_{t_i,1}(y|x)]|+|\bar{F}^N_{t_i,1}(y|x)-\mathbb{E}[\bar{F}^N_{t_i,1}(y|x)]|+2Ma_T.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{P}\Big[\sup_{x\in S}\sup_{u\in B}|\hat{F}^N_{t,1}(y|x)-\mathbb{E}[\hat{F}^N_{t,1}(y|x)]|>4Ma_T\Big]\\&\leq N_{S,T}N_{B_{i,T}}\max_{1\leq i\leq N_{B_{i,T}}}\mathbb{P}\Big[\sup_{x\in S}\sup_{u\in B_{i,T}}|\hat{F}^N_{t,1}(y|x)-\mathbb{E}[\hat{F}^N_{t,1}(y|x)]|>4Ma_T\Big]\end{aligned}$$

$$\leq Q_{1,T} + Q_{2,T},$$

where

$$Q_{1,T} = N_{S,T} N_{B_{i,T}} \max_{1 \leq i \leq N_{B_{i,T}}} \mathbb{P}\Big[|\hat{F}^N_{t_i,1}(y|x) - \mathbb{E}[\hat{F}^N_{t_i,1}(y|x)]| > M a_T \Big]$$

$$Q_{2,T} = N_{S,T} N_{B_{i,T}} \max_{1 \leq i \leq N_{B_{i,T}}} \mathbb{P}\Big[|\bar{F}^N_{t_i,1}(y|x) - \mathbb{E}[\bar{F}^N_{t_i,1}(y|x)]| > M a_T \Big].$$

Hereafter, we only show $Q_{1,T}$ since $Q_{2,T}$ can be shown similarly. Define

$$Z_{a,T}(u,x) = K_{h,1}(u - \frac{a}{T}) \Big\{ K_{h,2}(d(x, X_{a,T})) H_h(y - Y_{a,T}) \mathbb{1}_{|H_h(y - Y_{a,T})| \leq \tau_T} - \mathbb{E}[K_{h,2}(d(x, X_{a,T})) H_h(y - Y_{a,T}) \mathbb{1}_{|H_h(y - Y_{a,T})| \leq \tau_T}] \Big\}.$$

Take note that for each fixed (u, x), the array $\{Z_{a,T}(u, x)\}$ is β-mixing with coefficients $\beta_{Z,T}(k)$ satisfying

$$\beta_{Z,T}(k) \leq \beta(k).$$

Using Lemma B.3 in [14], for sufficiently large $C > 0$ and $S_T = a_T^{-1} \tau_T^{-1}$, we set

$$\varepsilon = M a_T T h \phi(h) \quad \text{and} \quad b_T = C \tau_T.$$

Furthermore, for C' independent of (u, x), by Theorem 2 in [48], we get

$$\sigma^2_{S_T,T} \leq C' S_T h \phi(h).$$

Now, we take sufficiently large $M > 0$ such that $C' < M$. Hence, for any fixed (u, x) and sufficiently large T, we get

$$\begin{aligned} \mathbb{P}\left[\left| \sum_{a=1}^T Z_{a,T}(u,x) \right| > M a_T T h \phi(h) \right] &\leq 4 \exp\left(-\frac{\varepsilon^2}{64 \sigma_{S_T,T} \frac{T}{S_T} + \frac{8}{3} \varepsilon b_T S_T} \right) + 4 \frac{T}{S_T} \beta(S_T) \\ &\leq 4 \exp\left(\frac{-M^2 \log T}{64 C' + \frac{8}{3} C M} \right) + 4 \frac{T}{S_T} S_T^{-\gamma} \\ &\lesssim \exp\left(\frac{-M \log T}{64 \frac{C'}{M} + \frac{8}{3} C} \right) + T S_T^{-\gamma - 1} \\ &\leq \exp\left(\frac{-M \log T}{64 + \frac{8}{3} C} \right) + T S_T^{-\gamma - 1} \\ &= T^{-\frac{M}{64 + 3C}} + T a_T^{\gamma+1} \tau_T^{\gamma+1}. \end{aligned}$$

We lastly show that $Q_{1,T} \lesssim \mathcal{O}(R_{1,T}) + \mathcal{O}(R_{2,T}) = o(1)$, where for sufficiently large M,

$$R_{1,T} = h^{-1} a_T^{-1} T^{-\frac{M}{64+3C}} = o(1),$$

and

$$\begin{aligned} R_{2,T} &= h^{-1}a_T^{-1}Ta_T^{\gamma+1}\tau_T^{\gamma+1} \\ &= h^{-1}T\left(\sqrt{\frac{\log T}{Th\phi(h)}}\right)^{\frac{\gamma}{2}}\rho_T^{\gamma+1}T^{\frac{\gamma+1}{\zeta}} \\ &= \frac{(\log T)^{\frac{\gamma}{2}+\zeta_0(\gamma+1)}}{T^{\frac{\gamma}{2}-1-\frac{\gamma+1}{\zeta}}h^{\frac{\gamma}{2}+1}\phi(h)^{\frac{\gamma}{2}}} = o(1). \end{aligned}$$

This ends the proof.

A.2. Proof of Lemma 3.2

Proof. First,

$$\begin{aligned} \mathbb{E}[\hat{F}_t^N(y|x)] &= \mathbb{E}\left[\frac{1}{Th\phi(h)}\sum_{a=1}^{T}K_{h,1}(\frac{t}{T}-\frac{a}{T})K_{h,2}(d(x,X_{a,T}))H_h(y-Y_{a,T})\right] \\ &= \frac{1}{Th\phi(h)}\sum_{a=1}^{T}K_{h,1}(\frac{t}{T}-\frac{a}{T})\mathbb{E}[K_{h,2}(d(x,X_{a,T}))H_h(y-Y_{a,T})] \\ &= \frac{1}{Th\phi(h)}\sum_{a=1}^{T}K_{h,1}(\frac{t}{T}-\frac{a}{T})\mathbb{E}[K_{h,2}(d(x,X_{a,T}))\mathbb{E}[H_h(y-Y_{a,T})|X_{a,T}]], \end{aligned}$$

by conditioning on $X_{a,T}$. Now see that

$$\begin{aligned} \mathbb{E}[H_h(y-Y_{a,T})|X_{a,T}] &= \int H_h(y-z)f_{X_{a,T}}(z)\mathrm{d}z \\ &= \int H\big(\frac{y-z}{h}\big)f_{X_{a,T}}(z)\mathrm{d}z. \end{aligned}$$

By integration by parts and change of variables: $\omega = \frac{y-z}{h}$, we have

$$\mathbb{E}[H_h(y-Y_{a,T})|X_{a,T}] = \int H'(\omega)F_t^{\star}(y-h\omega|X_{a,T})\mathrm{d}\omega.$$

Now, by Assumptions 3.2 and 3.5,

$$\begin{aligned} \mathbb{E}[H_h(y-Y_{a,T})|X_{a,T}] - F_t^{\star}(y|x) &\leq \int H'(\omega)|F_t^{\star}(y-h\omega|X_{a,T}) - F_t^{\star}(y|x)|\mathrm{d}\omega \\ &\leq \int H'(\omega)|d(x,X_{a,T}) + h\omega|\mathrm{d}\omega \\ &\lesssim h\int H'(\omega)\mathrm{d}\omega + h\int|\omega|H'(\omega)\mathrm{d}\omega \\ &\lesssim h. \end{aligned}$$

So, again by Assumptions 3.2 and 3.4, and using (A.12), we have

$$\begin{aligned}
&\big|\mathbb{E}[\hat{F}_t^N(y|x)] - F_t^\star(y|x)\hat{F}_t^D(y|x)\big| \\
&\leq \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})\big|\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))H_h(y-Y_{a,T})\big] - K_{h,2}(d(x,X_{a,T}))F_t^\star(y|x)\big| \\
&\leq \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})\big|\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))\mathbb{E}[H_h(y-Y_{a,T})|X_{a,T}]\big] - F_t^\star(y|x)\big| \\
&\leq \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})\big|\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))\big|\mathbb{E}[H_h(y-Y_{a,T})|X_{a,T}] - F_t^\star(y|x)\big|\big]\big| \\
&\lesssim \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})\mathbb{E}\big[K_{h,2}(d(x,X_{a,T}))\big]h \\
&\leq \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})h\mathbb{E}\big[\frac{1}{Th}U_{a,T}\big(\frac{a}{T}\big) + K_{h,2}\big(d\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big] \\
&\lesssim \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})h\big(\frac{1}{Th}+\phi(h)\big) \\
&\lesssim \frac{1}{T\phi(h)} + h\cdot\underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}(\frac{t}{T}-\frac{a}{T})}_{\mathcal{O}(1)} = \mathcal{O}(h).
\end{aligned}$$

Therefore,

$$\sup_{x\in S}\sup_{\frac{t}{T}\in I_h}|\mathbb{E}[\hat{F}_t^N(y|x)] - F_t^\star(y|x)\hat{F}_t^D(y|x)| = \mathcal{O}(h).$$

A.3. Proof of Lemma 3.3

Proof. By applying Lemma 3.1, for $H_h(y-Y_{a,T})=1$,

$$\sup_{x\in\mathscr{H},\frac{t}{T}\in I_h}\Big|J_{t,T}(\frac{t}{T},x) - \mathbb{E}\Big[J_{t,T}(\frac{t}{T},x)\Big]\Big| = \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th\phi(h)}}\Big).$$

Additionally, using Assumption 3.1, we use the decomposition $J_{t,T}(\frac{t}{T},x) = \widetilde{J}_{t,T}(\frac{t}{T},x) + \bar{J}_{t,T}(\frac{t}{T},x)$, where

$$\widetilde{J}_{t,T}(\frac{t}{T},x) = \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,2}\big(\mathsf{D}\big(x,X_a\big(\frac{a}{T}\big)\big)\big),$$

and

$$\bar{J}_{t,T}(\frac{t}{T},x) = \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\big\{K_{h,2}(\mathsf{D}(x,X_{a,T})) - K_{h,2}\big(\mathsf{D}\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big\}.$$

So,

$$\begin{aligned}
\Big|J_{t,T}(\frac{t}{T},x)\Big| &= \Big|J_{t,T}(\frac{t}{T},x) - \mathbb{E}[J_{t,T}(\frac{t}{T},x)] + \mathbb{E}[J_{t,T}(\frac{t}{T},x)]\Big| \\
&\leq \Big|J_{t,T}(\frac{t}{T},x) - \mathbb{E}[J_{t,T}(\frac{t}{T},x)]\Big| + \Big|\mathbb{E}[J_{t,T}(\frac{t}{T},x)]\Big| \\
&\leq \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th\phi(h)}}\Big) + \Big|\mathbb{E}[J_{t,T}(\frac{t}{T},x)]\Big| \\
&\leq \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th\phi(h)}}\Big) + \Big|\mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T},x) + \bar{J}_{t,T}(\frac{t}{T},x)]\Big| \\
&\leq \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th\phi(h)}}\Big) + \Big|\mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T},x)]\Big| + \Big|\mathbb{E}[\bar{J}_{t,T}(\frac{t}{T},x)]\Big|,
\end{aligned}$$

Now, let us first observe $\mathbb{E}[\bar{J}_{t,T}(\frac{t}{T},x)]$. Using Assumptions 3.1 and 3.2, we have

$$\begin{aligned}
\mathbb{E}[\bar{J}_{t,T}(\frac{t}{T},x)] &= \mathbb{E}\big[\frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\big\{K_{h,2}(\mathsf{D}(x,X_{a,T})) - K_{h,2}\big(\mathsf{D}\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big\}\big] \\
&= \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\big[\big\{K_{h,2}(\mathsf{D}(x,X_{a,T})) - K_{h,2}\big(\mathsf{D}\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big\}\big] \\
&\leq \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\big(\frac{L_2C_U}{Th}\big) \\
&\leq \frac{L_2C_U}{Th\phi(h)}\underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)}_{\mathcal{O}(1)} \\
&\leq \frac{L_2C_U}{Th\phi(h)} \lesssim \frac{1}{Th\phi(h)}.
\end{aligned}$$

which converges to zero using Assumption 3.4. On the other hand,

$$\begin{aligned}
\mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T},x)] &= \mathbb{E}\big[\frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,2}\big(\mathsf{D}\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big] \\
&= \frac{1}{Th\phi(h)}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\big[K_{h,2}\big(\mathsf{D}\big(x,X_a\big(\frac{a}{T}\big)\big)\big)\big]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{Th\phi(h)} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)\mathbb{E}\big[\mathbb{1}_{(\mathsf{D}(x, X_a(\frac{a}{T}))) \le h}\big] \\
&= \frac{1}{Th\phi(h)} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)F_{t/T}(h;x) \\
&\ge \frac{1}{\phi(h)} \underbrace{\frac{1}{Th} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} c_d\phi(h)f_1(x) \quad \text{(using Assumption 3.3)} \\
&\sim f_1(x) > 0,
\end{aligned}
$$

which implies that $\mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T}, x)] > 0$. Therefore,

$$
\frac{1}{\inf_{x\in S} \inf_{\frac{t}{T}\in I_h} J_{t,T}(\frac{t}{T}, x)} = \frac{1}{o_{\mathbb{P}}(1) + o(1) + \mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T}, x)]} = \mathcal{O}_{\mathbb{P}}(1).
$$