



HAL
open science

Bounds in Wasserstein Distance for Locally Stationary Functional Time Series

Jan Nino G. Tinio, Mokhtar Z. Alaya, Salim Bouzebda

► **To cite this version:**

Jan Nino G. Tinio, Mokhtar Z. Alaya, Salim Bouzebda. Bounds in Wasserstein Distance for Locally Stationary Functional Time Series. 2025. <hal-05026101>

HAL Id: hal-05026101

<https://hal.science/hal-05026101v1>

Preprint submitted on 9 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Bounds in Wasserstein Distance for Locally Stationary Functional Time Series

Jan Nino G. Tinio^{1,2}, Mokhtar Z. Alaya¹, and Salim Bouzebda¹

¹Université de Technologie de Compiègne,
Laboratoire de Mathématiques Appliquées de Compiègne, CS 60 319 - 60 203 Compiègne Cedex

²Department of Mathematics, Caraga State University, Butuan City, Philippines

e-mails: jan-nino.tinio@utc.fr, alayaelm@utc.fr, salim.bouzebda@utc.fr

April 9, 2025

Abstract

Functional time series (FTS) extend traditional methodologies to accommodate data observed as functions/curves. A significant challenge in FTS consists of accurately capturing the time-dependence structure, especially with the presence of time-varying covariates. When analyzing time series with time-varying statistical properties, locally stationary time series (LSTS) provide a robust framework that allows smooth changes in mean and variance over time. This work investigates Nadaraya-Watson (NW) estimation procedure for the conditional distribution of locally stationary functional time series (LSFTS), where the covariates reside in a semi-metric space endowed with a semi-metric. Under small ball probability and mixing condition, we establish convergence rates of NW estimator for LSFTS with respect to Wasserstein distance. The finite-sample performances of the model and the estimation method are illustrated through extensive numerical experiments both on functional simulated and real data.

Keywords: Conditional distribution estimation; Locally stationary functional time series; Mixing condition; Nadaraya-Watson estimation; Wasserstein distance

1 Introduction

In recent years, data collected across various fields increasingly exhibit functional or curve-like characteristics, commonly referred to as functional data whose values come from infinite-dimensional space. This advancement is driven by the proliferation of data collected on progressively refined temporal and spatial grids, for instance, in meteorology, medicine, satellite imagery, economics and finance, environmental science, and many others [65, 16, 18]. Specifically, [25] demonstrated intriguing applications in biometrics, [4] explored the relevance of functional data to environmental science, and [23, 3] focused on applications in econometrics. The statistical modeling of such data, conceptualized as random functions/curves, has given rise to various intricate theoretical and numerical research inquiries. Functional data are investigated for multiple reasons: they assist in developing data representation strategies that highlight significant features, analyzing mean behaviors and deviations, and constructing models when temporal dependency is present in the data [62, 41].

The statistical challenges in studying these data have drawn increasing attention over the last few decades from the statistical community [20, 19, 21]. Much of the existing research in FTS analysis is predicated on stationary models since the stationarity requirement is often adopted in time series modeling, leading to various models and methodologies. Theoretical foundations of function spaces concerning linear processes were explored in [11], which are crucial for modeling and forecasting functional time series. Statistical analysis of autoregressive processes within Banach spaces was emphasized in [12], a specialized class of functional spaces. Similarly, [34] examined asymptotic properties of the sample mean and variance for autoregressive processes of order one in a separable Banach space with independent innovations. Extending this, [2] incorporated wavelet analysis alongside nonparametric kernel regression for FTS forecasting.

However, many FTS exhibit nonstationarity. Even with detrending and deseasonalizing, the stationarity assumption is not always beneficial for modeling functional data. Many time series models, commonly observed in various physical phenomena and economic data, are non-stationary [48, 65, 16]. Conventional approaches become inappropriate when the (weak) stationarity assumption is violated. Specifically, global climate changes in meteorology affect the distribution of a region's daily temperature, precipitation, and cloud cover when viewed interconnectedly. In finance, an option's implied volatility evolves over time depending on its moneyness [68]. To address such nonstationarity, the locally stationary time series (LSTS) framework provides a practical modeling approach. The parameters of LSTS exhibit temporal dependence. These nonstationary processes can be approximated by locally stationary counterparts that remain stationary within smaller time windows. As a result, asymptotic theories can be developed to estimate time-dependent characteristics [26, 29, 28]. Many works have addressed the intuitive concept of local stationarity. The seminal works in [26, 27, 29] provide a strong foundation for the inference of LSTS. Parametric [29, 42, 28] and nonparametric [71, 48, 17] frameworks were developed to analyze LSTS. For instance, [26] introduced a parametric framework by leveraging local periodograms to minimize the generalized Whittle function tailored to locally stationary models. Expanding the parametric framework beyond time-invariant assumptions, [47] proposed a semiparametric model. This approach combined the strengths of parametric and nonparametric methods, allowing for greater flexibility in modeling complex time-varying structures without imposing stringent parametric constraints. Parallel to the parametric advancements, nonparametric approaches have garnered significant attention from researchers aiming to model conditional mean and variance functions without relying on predefined parametric forms. In particular, [48] developed an estimation theory for nonparametric regression problems involving LSFTS. Central to nonparametric approaches in the literature is Nadaraya-Watson (NW) [55, 72] estimation procedure, which has been extensively used to estimate conditional mean functions.

Despite the proven efficacy of NW procedure in estimating conditional mean functions, its usefulness extends to conditional distribution estimation (CDE). Estimating accurately conditional distributions is essential in prediction and forecasting, as it comprehensively describes the conditional law for any given random variable. Due to its critical role in predictive modeling and inference, numerous studies have focused on developing robust estimation theories for CDE within FTS framework [39, 45, 14]. Two distinct approaches for estimating the conditional distribution of a target variable within a prediction set, given a functional covariate, were introduced in [45]. The first method relies on the empirical distribution of the estimated model residuals, while the second involves fitting functional parametric models to the residuals. These approaches provide flexible frameworks for handling complex dependencies between functional covariates and target variables,

particularly in high-dimensional settings where traditional parametric models may be inadequate. Additionally, [14] developed a local polynomial estimator for conditional CDF of a scalar target variable given a functional covariate in the context of stationary strongly mixing processes. They established asymptotic normality property.

In dealing with CDE, optimal transport (OT) theory has emerged as a powerful mathematical framework for quantifying difference between probability distributions. OT measures the minimal cost required to transport one distribution to another, providing a meaningful metric for distributional comparison [59]. This approach addresses shortest-path problems by enabling the concurrent transportation of multiple items along geodesic curves or straight paths. Among various metrics derived from OT, Wasserstein distance stands out for its robustness and versatility in comparing probability distributions [70, 50, 66].

The landscape of LSTS analysis is enriched by parametric and nonparametric approaches, each contributing unique strengths to the modeling of time-evolving data. Parametric approaches offer structured and interpretable models for capturing dynamic behaviors through frameworks like those proposed in [26, 47]. Concurrently, nonparametric techniques, particularly those leveraging NW estimator and OT theory, provide flexible and robust tools for estimating conditional means, variances, and distributions. The integration of these procedures, underpinned by rigorous theoretical advancements, continues to enhance the capacity of statisticians and data scientists to model, predict, and infer from complex, non-stationary time series data.

Contributions. This paper investigates CDE for LSFTS characterized by weakly dependent sequences using NW estimation procedure. We establish convergence rates of NW estimator for conditional distribution of $Y_{t,T} \mid X_{t,T}$ with respect to Wasserstein distance. Here, the response variable $Y_{t,T}$ is scalar, while locally stationary covariate $X_{t,T}$ resides in a semi-metric space \mathcal{H} endowed with a semi-metric $D(\cdot, \cdot)$. We perform numerical experiments on synthetic and real-world data to illustrate our theoretical findings.

Layout. The paper is structured as follows: Section 2 introduces the regression estimation problem and provides an overview of key concepts, including local stationarity, Wasserstein distance, small ball probability, and mixing conditions. In Section 3, we present main theoretical results, which include (i) the formulation of kernelized NW estimator, (ii) the explicit convergence rates of Wasserstein distance between NW estimator and true conditional distributions, and (iii) the proposed method for bandwidth selection. The finite-sample performance of the model and estimation method is presented in Section 4 through the analysis of both simulated and real data. To preserve the flow of the presentation, all proofs are deferred to the appendices.

Notation. Throughout this paper, we adopt the following notations. Dirac measure at a point y is denoted by δ_y . For any real-valued random variable X and any $q \geq 1$, we denote L_q -norm of X by $\|X\|_{L_q}$ and is defined as $\|X\|_{L_q} = (\mathbb{E}[|X|^q])^{1/q}$. We use the notation $a_T \lesssim b_T$ to indicate that there exists a constant C , independent of T , such that $a_T \leq Cb_T$. The constant C may vary unless specified otherwise. Similarly, $a_T \sim b_T$ signifies that both $a_T \lesssim b_T$ and $b_T \lesssim a_T$ hold. For positive sequences $\{a_T\}$ and $\{b_T\}$, we write $a_T = \mathcal{O}(b_T)$ provided that $\lim_{T \rightarrow \infty} \frac{a_T}{b_T} \leq C$ for some constant $C > 0$. Additionally, $a_T = \mathcal{O}(1)$ indicates that a_T is bounded. We denote $a_T = o(b_T)$ if $\lim_{T \rightarrow \infty} \frac{a_T}{b_T} = 0$, and $a_T = o(1)$ when a_T approaches zero. For a sequence of random variables

$\{X_T\}$ and a given sequence $\{a_T\}$, we write $X_T = \mathcal{O}_{\mathbb{P}}(a_T)$ 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, $X_T = o_{\mathbb{P}}(a_T)$ indicates that $\lim_{T \rightarrow \infty} \mathbb{P}[\frac{|X_T|}{a_T} > \epsilon] = 0$ for all $\epsilon > 0$, and $X_T = o_{\mathbb{P}}(1)$ when X_T converges in probability to zero. For any $a, b \in \mathbb{R}$, we define $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$.

2 Background on LSTS and Wasserstein distance

This section presents some preliminaries of LSTS and Wasserstein distance. We then introduce small ball probability and delineate the mixing condition employed to assess weak dependency.

2.1 Locally stationary time series

Let $T \in \mathbb{N}$ and suppose that there exists a process of T random variables $\{Y_{t,T}, X_{t,T}\}_{t=1,\dots,T}$, where $Y_{t,T}$ is real-valued, and $X_{t,T}$ belongs to a semi-metric space \mathcal{H} with a semi-metric $D(\cdot, \cdot)$. The semi-metric space \mathcal{H} can be Banach or Hilbert spaces with norm $\|\cdot\|$. We consider the following regression estimation problem:

$$Y_{t,T} = m^*\left(\frac{t}{T}, X_{t,T}\right) + \varepsilon_{t,T}, \text{ for all } t = 1, \dots, T, \quad (1)$$

where $\{\varepsilon_{t,T}\}_{t \in \mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) random variables independent of $\{X_{t,T}\}_{t=1,\dots,T}$, i.e., $\mathbb{E}[\varepsilon_t | X_{t,T}] = 0$, for all $t = 1, \dots, T$. The response $Y_{t,T}$ is assumed to be integrable. The functional covariate $X_{t,T}$ is assumed to be locally stationary, that dynamically changes slowly over time and hence can be considered approximately stationary at local time. Note that $m^*\left(\frac{t}{T}, X_{t,T}\right) = \mathbb{E}[Y_{t,T} | X_{t,T}]$ is the *oracle* conditional mean function in model 1, and does not depend on real-time t but rather on rescaled time $u = \frac{t}{T}$. As the sample size T goes to infinity, these u -points form a dense subset of the unit interval $[0, 1]$. Hence, at all rescaled u -points, m^* is identified almost surely (a.s.) if it is continuous in the time direction. In LSTS, this rescaled time refers to transforming the original time scale.

Example. Consider the process $X_{t,T} = a\left(\frac{t}{T}\right) + \varepsilon_t$, $t \in \{1, \dots, T\}$, $T \in \mathbb{N}$, where $a(\cdot)$ is a continuous function $a : [0, 1] \rightarrow \mathbb{R}$ and a sequence of i.i.d. random variables $\{\varepsilon_t\}_{t \in \mathbb{N}}$. The process $X_{t,T}$ behaves ‘‘almost’’ stationary for t close to t' , for some $t' \in \{1, \dots, T\}$, that is, $a\left(\frac{t'}{T}\right) \approx a\left(\frac{t}{T}\right)$. However, this process is not weakly stationary. Local stationarity gives a more realistic concept that allows this kind of change [17].

Let us now formally define the notion of LSTS. We adopt the definition given in [48].

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

$$D(X_{t,T}, X_t(u)) \leq \left(\left| \frac{t}{T} - u \right| + \frac{1}{T} \right) U_{t,T}(u) \text{ a.s.}, \quad (2)$$

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

From this definition, if an \mathcal{H} -valued process $\{X_{t,T}\}_{t=1,\dots,T}$ is locally stationary, a strictly stationary process $\{X_t(u)\}_{t=1,\dots,T}$ can always be found around each rescaled time u , which will

be used to approximate $\{X_{t,T}\}_{t=1,\dots,T}$. This approximation will result in a negligible difference between random variables $X_{t,T}$ and $X_t(u)$. Since the ρ -th moments of the positive random variables $U_{t,T}(u)$ are uniformly bounded, $U_{t,T}(u) = O_{\mathbb{P}}(1)$ [71]. Thus, we have

$$D(X_{t,T}, X_t(u)) = O_{\mathbb{P}}\left(\left|\frac{t}{T} - u\right| + \frac{1}{T}\right).$$

If $u = \frac{t}{T}$, then $D(X_{t,T}, X_t(\frac{t}{T})) \leq \frac{C_U}{T}$. The constant ρ can be considered as an indicator of how well this approximation is being made: larger ρ indicates a better approximation of $X_{t,T}$ by $X_t(u)$ and moderate bounds for their absolute difference.

Definition 1 is consistent with the one given in [68, 67] when \mathcal{H} is the Hilbert space $L_{\mathbb{R}}^2([0, 1])$, with inner product L_2 -norm:

$$\|f\|_2 = \sqrt{\langle f, f \rangle}, \quad \langle f, g \rangle = \int_0^1 f(t)g(t)dt,$$

where $f, g \in L_{\mathbb{R}}^2([0, 1])$. Sufficient conditions were also provided so that an $L_{\mathbb{R}}^2([0, 1])$ -valued stochastic process $X_{t,T}$ satisfies (2) with $D(f, g) = \|f - g\|_2$ and $\rho = 2$. In [68], $L_E^p(I, \mu)$ is defined as the Banach space of all strongly measurable functions $f : I \rightarrow E$ with norms

$$\|f\|_p = \|f\|_{L_E^p(I, \mu)} = \left(\int \|f(s)\|_E^p d\mu(s) \right)^{\frac{1}{p}},$$

for $1 \leq p < \infty$, and

$$\|f\|_{\infty} = \|f\|_{L_E^{\infty}(I, \mu)} = \inf_{\mu(N)=0} \sup_{s \in I \setminus N} \|f(s)\|_E,$$

for $p = \infty$.

2.2 Wasserstein distance

Let $\mathcal{P}_r(\mathbb{R})$ be the set of Borel probability measures in \mathbb{R} having finite r -th moment ($r \geq 1$), i.e.,

$$\mathcal{P}_r(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x|^r \mu(dx) < \infty\}.$$

Given probability measures $\mu, \nu \in \mathcal{P}_r(\mathbb{R})$, we calculate the distance between them using the r th-Wasserstein distance, $W_r(\mu, \nu)$, as follows

$$W_r(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{R} \times \mathbb{R}} |u - v|^r \pi(du, dv) \right)^{1/r}, \quad (3)$$

where $\Pi(\mu, \nu)$ denotes the set of probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν . Optimal couplings always exist since \mathbb{R} is a complete and separable metric space, where the infimum is, in fact, a minimum [70]. Equation (3) signifies that $W_r(\mu, \nu)$ is the infimum of expectation of the distance between two random variables over all possible couplings, i.e., $W_r(\mu, \nu) = \left(\inf_{U \sim \mu, V \sim \nu} \mathbb{E}[|U - V|^r] \right)^{1/r}$.

We can represent a simple optimal coupling by a probability inverse transform: given $\mu, \nu \in \mathcal{P}_r(\mathbb{R})$, let $F_{\mu}(\cdot)$ and $F_{\nu}(\cdot)$ be the cumulative distribution functions (CDFs) and $F_{\mu}^{-1}(\cdot)$ and $F_{\nu}^{-1}(\cdot)$

be the respective generalized inverse or quantile functions defined as $F_\mu^{-1}(z) := \inf\{v \in \mathbb{R} : \mu((-\infty, v]) \geq z\}$ for all $z \in [0, 1]$ (similarly for $F_\nu^{-1}(z)$). Then, assuming a random variable Z uniformly distributed on $(0, 1)$, an optimal coupling $(U, V) = (F_\mu^{-1}(Z), F_\nu^{-1}(Z))$ can be established [32, 35]. Hence, the minimization problem (3) can be represented by

$$W_r(\mu, \nu) = \left(\int_0^1 |F_\mu^{-1}(z) - F_\nu^{-1}(z)|^r dz \right)^{1/r},$$

in a one-dimensional context. Specifically, for $r = 1$ and using a change of variable, the 1-Wasserstein distance is represented as

$$W_1(\mu, \nu) = \int_{\mathbb{R}} |F_\mu(v) - F_\nu(v)| dv. \quad (4)$$

Consequently, $W_1(\mu, \nu)$ can be considered as the L_1 -distance between the CDFs $F_\mu(\cdot)$ and $F_\nu(\cdot)$.

2.3 Small ball probability

The absence of a density function for functional random variables is a technical difficulty in infinite-dimensional spaces since we lack a universal reference such as the Lebesgue measure. We overcome this using the *small ball probability* property. We control the concentration of probability measure of the functional variable on a small ball using a function $\phi(r)$ defined as, for all $r > 0$ and a fixed $x \in \mathcal{H}$,

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

where $B(x, r) = \{v \in \mathcal{H} : D(x, v) \leq r\}$. Assume that r is a function of T such that $r = r(T) \rightarrow 0$ as $T \rightarrow \infty$. If we take T very large, $B(x, r)$ is then considered as a small ball; hence, $\mathbb{P}[X \in B(x, r)]$ is a small ball probability [40]. Unfortunately, obtaining $\mathbb{P}[X \in B(x, r)]$ is complicated [38]. For a survey on the main results on small ball probability, refer to [49]. In most cases, it is fitting to suppose that, as $r \rightarrow 0$,

$$\mathbb{P}[X \in B(x, r)] \sim \psi(x)\phi(r), \quad (5)$$

where $\mathbb{E}[\psi(X)] = 1$, a necessary normalizing restriction to ensure the identifiability of the decomposition. There are two main reasons for conveniently assuming (5). First, the function $\psi(x)$ can be thought as a surrogate density of the functional X and can be utilized in different frameworks where the surrogate density is estimated differently and is used for classification purposes. Second, the function $\phi(r)$ signifies the volumetric term that can be used to evaluate the complexity of the probability law of the process [10]. In the d -dimensional case $X \in \mathbb{R}^d$, we suppose $\phi(r) \sim r^d$, which is commonly referred as the curse of dimensionality [40, 37]. The intrinsic nature of the probability effects, involving small balls, is apparent in infinite-dimensional framework. We give some examples of several forms of $\phi(r)$ that can also be found in [37, 17].

Fractional Brownian Motion. Considering the space $\mathcal{C}([0, 1], \mathbb{R})$ with the supremum norm and its Cameron-Martin associated space $\mathcal{F} = \mathcal{C}([0, 1], \mathbb{R})^{CM}$. Using Theorems 3.1 and 4.6 in [49], for $0 < \eta < 2$, we have

$$\forall x \in \mathcal{F}, \quad C'_x e^{r^{-2/\eta}} \leq \mathbb{P}[\zeta^{FBM} \in B(x, r)] \leq C_x e^{r^{-2/\eta}},$$

where ζ^{FBM} is the usual Fractional Brownian Motion with parameter η and $B(x, r) = \{\zeta^{FBM} \in \mathcal{F} : \|\zeta^{FBM} - x\|_\infty \leq r\}$. In this example, for the Fractional Brownian process, we choose $\phi(r)$ of the form

$$\phi^{FBM}(r) \sim e^{r^{-2/\eta}}.$$

Gaussian process. Next, let us consider the centered Gaussian process $\zeta^{GP} = \{\zeta_t^{GP}, 0 \leq t \leq 1\}$, which can be expanded by Karhunen-Loève decomposition as

$$\zeta_t^{GP} = \sum_{i=1}^{\infty} \sqrt{\lambda_i} Z_i f_i(t),$$

where λ_i 's are the eigenvalues of the covariance operator of ζ^{GP} , f_i 's are the associated orthonormal eigenfunctions, and Z_i 's are independent standard normal real random variables. The orthogonal projection onto the subspace spanned by the eigenfunctions $\{f_1, \dots, f_k\}$ is denoted by Ξ_k , for $k \in \mathbb{N}^*$. Define a semi-metric by

$$D(x, y) = \int_0^1 (\Xi_k(x - y)(t))^2 dt.$$

Using the Karhunen-Loève expansion, we get

$$D(\zeta^{GP}, x) = \sum_{i=1}^k \left(\sqrt{\lambda_i} Z_i - x_i \right)^2 := \sum_{i=1}^k \chi_i^2,$$

where

$$x_i = \int_0^1 x(t) f_i(t) dt, \quad i = 1, \dots, k.$$

$D(\zeta^{GP}, x)$ can be written in terms of the usual Euclidian norm on \mathbb{R}^k of a vector $\chi = (\chi_1, \dots, \chi_k)$. Since χ_i 's are independent real random variables with density with respect to the Lebesgue measure, we have, for $B(x, r) = \{\zeta^{GP} \in \mathcal{F} : D(\zeta^{GP}, x) < r\}$

$$\mathbb{P}[\zeta^{GP} \in B(x, r)] \sim r^k.$$

Ornstein-Uhlenbeck process. Lastly, considering the same space in (i) and the metric $D(\cdot, \cdot)$ associated with the supremum norm

$$\forall x \in \mathcal{C}([0, 1], \mathbb{R}), \quad \|x\|_\infty = \sup_{t \in [0, 1]} |x(t)|.$$

We denote the Wiener measure on $\mathcal{C}([0, 1], \mathbb{R})$ by \mathbb{P}^W and the associated functional Cameron-Martin space of $\mathcal{C}([0, 1], \mathbb{R})$ is given by $\mathcal{F} = \mathcal{C}([0, 1], \mathbb{R})^{CM}$. Moreover, we denote the standard Wiener process by w and let us consider the Ornstein-Uhlenbeck process ζ^{OU} defined by $\zeta_0^{OU} = 0$ and by

$$d\zeta_t^{OU} = dw_t - \frac{1}{2}\zeta_t^{OU}, \quad \forall t, 0 < t \leq 1.$$

By [9], the small centered ball Wiener measures are known to be of the form

$$\mathbb{P}^W[\|x\|_\infty \leq r] \sim \frac{4}{\pi} e^{-\pi^2/8r^2}.$$

The process ζ^{OU} has a probability measure that is absolutely continuous with respect to \mathbb{P}^W , we write

$$\forall x \in \mathcal{F}, \quad \mathbb{P}^W[\|x - \zeta^{OU}\|_\infty \leq r] \sim C_x e^{-\pi^2/8r^2}.$$

For Ornstein-Uhlenbeck process, we choose

$$\phi^{OU}(r) \sim e^{-\pi^2/8r^2}.$$

Since we deal with sequences exhibiting weak dependency, let us formally define the mixing coefficient considered in this paper.

2.4 Mixing condition

The degree of dependence between observations of a stochastic process as they become distant apart in time is measured using mixing coefficients. Mixing processes were introduced to generalize the law of large numbers for non-i.i.d. stochastic processes. For effective modeling and inference, selecting the appropriate mixing condition for a stochastic process is crucial [58, 33, 63]. One of the mixing criteria usually considered is β -mixing. It has been applied to demonstrate moment inequalities and central limit theorems [31, 13, 60].

Definition 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

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

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 $\beta(k) \rightarrow 0$ as $k \rightarrow \infty$.

This definition implies that if a process is β -mixing, asymptotic independence can be attained when $k \rightarrow \infty$. It is a ‘‘just right’’ assumption in analyzing weakly dependent sequences [69]. There are different forms of β -mixing, such as exponentially β -mixing $\beta(k) = \mathcal{O}(e^{-\gamma k})$, for $\gamma > 0$, and arithmetically β -mixing $\beta(k) = \mathcal{O}(k^{-\gamma})$ [41]. Numerous common time series models, such as autoregressive moving average (ARMA) models [54], generalized autoregressive conditional heteroscedastic (GARCH) models [24], and some Markov processes [36], are known to be β -mixing.

3 Nadaraya-Watson estimation with Wasserstein distance

We denote the conditional probability distribution of $Y_{t,T}|X_{t,T} = x$ by $\pi_t^*(\cdot|x)$ and its conditional CDF by $F_t^*(\cdot|x)$, for a fixed $t \in \{1, \dots, T\}$ and $x \in \mathcal{X}$. The mean conditional function reads as

$$m^*\left(\frac{t}{T}, x\right) = \mathbb{E}_{\pi_t^*(\cdot|x)}[Y_{t,T}|X_{t,T} = x] = \int_{-\infty}^{\infty} y d\pi_t^*(y|x).$$

Setting K_1, K_2 two 1-dimensional basic kernel functions and h a bandwidth that depends on the sample size T , i.e., $h = h(T)$ with $h(T) \rightarrow 0$ as $T \rightarrow \infty$. For ease of notation, we set the scaled kernels $K_{h,i}(\cdot) = K_i(\frac{\cdot}{h})$, for $i = 1, 2$. Next, we define the considered NW estimator.

Definition 3. The NW estimator of $\pi_t^*(\cdot|x)$ is given by

$$\hat{\pi}_t(\cdot|x) = \sum_{a=1}^T \omega_a\left(\frac{t}{T}, x\right) \delta_{Y_{a,T}},$$

where

$$\omega_a\left(\frac{t}{T}, x\right) = \frac{K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(D(x, X_{a,T}))}{\sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(D(x, X_{a,T}))}. \quad (6)$$

NW estimator of the conditional CDF $F_t^*(y|x)$ can be written as, for all $y \in \mathbb{R}$,

$$\hat{F}_t(y|x) = \sum_{a=1}^T \omega_a\left(\frac{t}{T}, x\right) \mathbb{1}_{Y_{a,T} \leq y}. \quad (7)$$

This definition extends the estimator considered in [66] to a functional covariate $X_{t,T}$. The weights $\{\omega_a(u, x)\}_{a=1, \dots, T}$ are assumed to be measurable functions of x , $X_{a,T}$, and u but do not depend on $Y_{a,T}$. Note that in [48], the NW estimator of $m^*(u, x)$ is given by

$$\hat{m}(u, x) = \sum_{a=1}^T \omega_a(u, x) Y_{a,T}. \quad (8)$$

Remark. We are using two kernel functions: one with respect to the rescaled time $u = \frac{t}{T}$ and the other in the direction of the functional $X_{t,T}$. To appropriately assign weights $\omega_a(\frac{t}{T}, x)$, we smooth with respect to the rescaled time and the space-direction of the covariates $X_{t,T}$ to analyze the local behavior of the data [71]. We consider a single bandwidth h for the kernels $K_{h,i}(\cdot)$; however, h could also be different for $K_{h,1}(\cdot)$ and $K_{h,2}(\cdot)$ [64].

Next, let us establish the underlying assumptions used for our main results.

3.1 Assumptions

The following assumptions are conventional in the literature of LSTS [71, 48, 17] and CDE [44, 53, 14].

Assumption 1 (Local stationarity). Assume that the \mathcal{H} -valued process $\{X_{t,T}\}_{t=1, \dots, T}$ is locally stationary approximated by $\{X_t(u)\}_{t=1, \dots, T}$ for each time point $u \in [0, 1]$.

Assumption 2 (Kernel functions). The kernel $K_1(\cdot)$ is symmetric about zero, bounded and has compact support, that is, $K_1(v) = 0$ for all $|v| > C_1$ for some $C_1 < \infty$. On the other hand, the kernel $K_2(\cdot)$ is bounded, and has a compact support in $[0, 1]$ such that $0 < K_2(0)$ and $K_2(1) = 0$. In addition, $K_2'(v) = dK_2(v)/dv$ exists on $[0, 1]$, satisfying $C_1' \leq K_2'(v) \leq C_2'$, for real constants $-\infty < C_1' < C_2' < 0$. Moreover, for $i = 1, 2$, $K_i(\cdot)$, is Lipschitz continuous, that is, $|K_i(v) - K_i(v')| \leq L_i|v - v'|$ for some $L_i < \infty$ and all $v, v' \in \mathbb{R}$. We further assume the following:

$$\int K_i(z) dz = 1, \quad \text{and} \quad \int z K_1(z) dz = 0. \quad (9)$$

Assumption 1 formalizes the property of the functional covariate $X_{t,T}$ as locally stationary. Assumption 2 is standard in literature. The conditions that K_i is compactly supported and Lipschitz implies that the kernel function has a bounded rate of change and is essential in obtaining upper bounds. First condition in (9) is a normalization, ensuring that the kernel can be interpreted as a probability density function. We assume that $K_2(\cdot)$ is compactly supported in $[0, 1]$; that is, it is a kernel of type II [41]. Second condition implies that $K_1(\cdot)$ is symmetric around the origin, and it ensures that it does not introduce first-order linear bias when applied to the data.

Assumption 3 (Small ball probability). *Let $B(x, h) = \{v \in \mathcal{H} : D(x, v) \leq h\}$ denote a ball centered at $x \in \mathcal{H}$ with radius h . We assume that for all $u \in [0, 1]$, $x \in \mathcal{H}$, and $h > 0$, there exists positive constants $C' < C$, such that*

$$0 < C' \phi(h) \psi(x) \leq \mathbb{P}[X_t(u) \in B(x, h)] =: F_u(h; x) \leq C \phi(h) \psi(x), \quad (10)$$

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

Assumption 4 (Regularity condition on h and $\phi(h)$). *Assume that as $T \rightarrow \infty$, the bandwidth h satisfies $T^{\frac{1}{2}} h \phi(h) \rightarrow \infty$.*

Assumption 3 gives condition on the distributional behavior of the variables. Equation (10) controls the behavior of the small ball probability around zero. The small ball probability can be approximately expressed as the product of two independent functions $\phi(\cdot)$ and $\psi(\cdot)$. This condition corresponds to the assumption used in [43, 48, 16]. On the other hand, Assumption 4 indicates that h should converge slower to zero, for instance, at a polynomial rate, i.e., $h = \mathcal{O}(T^{-\xi})$, for small $\xi > 0$. As h approaches zero, $\phi(h)$ also goes to zero. Assumption 4 is a strengthening of the condition in [48] that $Th\phi(h) \rightarrow \infty$ and is needed to guarantee our resulting convergence rates. With this, for Fractal-type processes, Assumption 4 holds true when we choose $h \sim T^{-\xi}$ for $0 < \xi < \frac{1}{2(1+\tau_0)}$ and $\phi(h) \sim h^{\tau_0}$ for some $\tau_0 > 1$ [16, 1]. To see different expressions of the function $\phi(h)$, one may refer to [41, 9] for some discussions on fractal-type processes, [52] for diffusion processes, and [49] for general Gaussian processes. We have given examples of the forms of $\phi(\cdot)$ in Subsection 2.3.

Assumption 5 (Conditional CDF). *The conditional CDF $F^*(\cdot|\cdot)$ is Lipschitzian, i.e., $|F_a^*(\cdot|x) - F_t^*(\cdot|x')| \leq L_{F^*} (D(x, x') + |\frac{a}{T} - \frac{t}{T}|)$, for some $L_{F^*} < \infty$, and for all $a, t \in \{1, \dots, T\}$, $x, x' \in \mathcal{H}$.*

The conditional CDF $F^*(\cdot|\cdot)$ should behave smoothly and not change much as the observation does, as assumed in [15, 53, 66]. We do not assume that the conditional CDF is twice differentiable, in contrast to [44, 39, 57].

Assumption 6 (Mixing condition). *The process $\{(X_{t,T}, \varepsilon_{t,T})\}$ is arithmetically β -mixing satisfying $\beta(k) \leq Ak^{-\gamma}$ for some $A > 0$ and $\gamma > 2$. We also assume that for some $p > 2$ and $\zeta > 1 - \frac{2}{p}$,*

$$\sum_{k=1}^{\infty} k^{\zeta} \beta(k)^{1-\frac{2}{p}} < \infty. \quad (11)$$

Assumption 7 (Blocking condition). *There exists a sequence of positive integers $\{q_T\}$ satisfying $q_T \rightarrow \infty$ and $q_T = o(\sqrt{Th\phi(h)})$, as $T \rightarrow \infty$.*

For dependent sequences estimation technique, Assumptions 6 and 7 are helpful. A more robust type of independence between far-off observations in a process is the β -mixing [22, 63, 60]. The decay of the regular mixing coefficient $\beta(k)$ is highlighted by condition (11). Bernstein's blocking approach was utilized to create independent blocks in the proof of Theorem 1 [6]. Assumption 7 defines the big block size as proportional to q_T .

3.2 Convergence in Wasserstein distance

We establish the convergence rate of NW estimator $\hat{\pi}_t(\cdot|x)$ wrt the Wasserstein distance. Theorem 1 below generalizes the convergence results in [66] to the functional or infinite-dimensional setting.

Theorem 1. *Suppose Assumptions 1 - 7 are satisfied and define $I_h = [C_1 h, 1 - C_1 h]$. Then*

$$\sup_{x \in \mathcal{H}, \frac{t}{T} \in I_h} \mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{T^{\frac{1}{2}} h \phi(h)} + h\right).$$

This convergence rate is comparable with Theorem 1 in [66] for the d -dimensional covariate case. However, we do not have the bias term involving ρ that comes from approximating $X_{t,T}$ by a locally stationary $X_t(\frac{t}{T})$. This rate depends on the bandwidth h and the small ball probability $\phi(h)$, where $\phi(h) \rightarrow 0$ as $h \rightarrow 0$, as highlighted in Assumption 4. We defer the proof to Appendix B.1, which follows similar steps of the proof of Theorem 1 in [66].

For the i.i.d. case $(Y_t, X_t)_{t=1, \dots, T}$, where Y_t is scalar, and X_t is functional, [53] provided a convergence result for their proposed conditional CDF estimator with surrogate data. The proposed estimator involves three rescaled kernel functions: a one-dimensional kernel $K(\frac{\cdot}{h_K})$ to account the functional X_t , an integrated Kernel $H(\frac{\cdot}{h_H})$ that acts as a CDF of Y_t , and a two-dimensional kernel $W(\frac{\cdot}{a_T}, \frac{\cdot}{a_T})$ to account for the surrogate variable, with bandwidths h_K , h_H , and a_T , respectively. This estimator converges to the true conditional distribution of order $\mathcal{O}(h_K^{c_1} + h_H^{c_2} + a_T^{c_1}) + \mathcal{O}\left(\sqrt{\frac{\log d_T}{T \phi(a_T)}}\right) + \mathcal{O}\left(\sqrt{\frac{\log T}{T \phi(h_K)}}\right)$, where $c_1, c_2 > 0$ and d_T satisfies $\frac{\log^2 T}{T \phi(a_T)} \leq d_T \leq \frac{T \phi(a_T)}{\log T}$. If there is no surrogate data and the integrated kernel H is replaced by an indicator function, this convergence rate becomes $\mathcal{O}(h_K^{c_1}) + \mathcal{O}\left(\sqrt{\frac{\log T}{T \phi(h_K)}}\right)$, which is comparable to the result above.

Corollary 1. *Suppose Assumptions 1 - 7 are satisfied and $Y_{t,T}$ is uniformly bounded by $M > 0$. Then, for $r \geq 1$,*

$$\sup_{x \in \mathcal{H}, \frac{t}{T} \in I_h} \mathbb{E}[W_r^r(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{T^{\frac{1}{2}} h \phi(h)} + h\right).$$

Proof of Corollary 1 is shown in Appendix B.2. For the i.i.d case, [8] (Theorem 5.3) showed that $\mathbb{E}[W_r^r(\mu_T, \mu)] = \mathcal{O}((T+2)^{-\frac{r}{2}})$, for $r \geq 1$, where μ_T is the empirical measure of an i.i.d sample $(X_t)_{t \geq 1}$ with common law μ .

Corollary 2. *Let Assumptions 1 - 7 be satisfied. Then*

$$\sup_{x \in \mathcal{H}, \frac{t}{T} \in I_h} \|W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))\|_{L_2} = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{T^{\frac{1}{2}} h \phi(h)} + h\right).$$

Proof of Corollary 2 is based on Minkowski's integral inequality: for any $r \geq 1$,

$$\left\| \int |\hat{F}_t(y|x) - F_t^*(y|x)| dy \right\|_{L_r} \leq \int \|\hat{F}_t(y|x) - F_t^*(y|x)\|_{L_r} dy. \quad (12)$$

The remainder of the proof adheres to the same lines used for Theorem 1, refer to Appendix B.3. The following proposition shows that the NW conditional mean function estimator \hat{m} warrants

Proposition 1. *Let Assumptions 1 - 7 be satisfied and $\hat{m}(\frac{t}{T}, x)$ be defined by (8). Then*

$$\sup_{x \in \mathcal{X}, \frac{t}{T} \in I_h} \mathbb{E}[|\hat{m}(\frac{t}{T}, x) - m^*(\frac{t}{T}, x)|] = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{T^{\frac{1}{2}} h \phi(h)} + h\right).$$

Proof of Proposition 1 is detailed in Appendix B.4. Similar to Proposition 1 in [66], this result indicates that Wasserstein distance can be used to obtain the convergence rate of $\hat{m}(u, x)$. The bound of the Wasserstein distance is slower than $\hat{m}(u, x)$ since we are examining differences between distributions, not just differences between conditional means [66]. This rate is comparable to Theorem 3.1 in [48] with convergence rate of order $\mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log T}{T h \phi(h)}} + h^{2 \wedge \beta}\right)$. We remark that a similar component for the bias term can be obtained if we assume that $F^*(\cdot)$ is twice differentiable and satisfies the Hölder condition.

Proposition 2. *Suppose $X_t(u)$ is a fractal-type process and Assumptions 1 - 7 are satisfied. Let the bandwidth be chosen to be $h = \mathcal{O}(T^{-\xi})$, and the small ball probability take the form $\phi(h) = h^{\tau_0}$, where $0 < \xi < \frac{1}{2(1+\tau_0)}$ and $\tau_0 > 1$. Then*

$$\sup_{x \in \mathcal{X}, \frac{t}{T} \in I_h} \mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{T^{\frac{1}{2}-\xi(1+\tau_0)}} + \frac{1}{T^{\xi}}\right).$$

As demonstrated in Appendix B.5, by setting $h = \mathcal{O}(T^{-\xi})$ and $\phi(h) = h^{\tau_0}$, proof of Proposition 2 follows immediately from the proof of Theorem 1.

3.3 Bandwidth selection criterion

In nonparametric kernel estimation, especially NW, the bandwidth must be suitably selected for the estimator to perform well. Bandwidth selection methods have already been established and developed in [61, 20]. This paper considers the leave-one-out cross-validation procedure used in [5, 61]. For any fixed $i \in \{1, \dots, T\}$, we define

$$\hat{m}_i(\frac{t}{T}, x) = \sum_{a=1; a \neq i}^T \omega_a(\frac{t}{T}, x) Y_{a,T}, \quad (13)$$

where $\omega_a(\frac{t}{T}, x)$ is given by (6). Equation (13) is regarded as the leave-out- $(X_{i,T}, Y_{i,T})$ estimator of $m_i^*(\frac{t}{T}, x)$. To minimize the quadratic loss function, we introduce the following criterion

$$CV(y, x, h) := \frac{1}{T} \sum_{i=1}^T (Y_{i,T} - \hat{m}_i(\frac{t}{T}, x))^2 \tilde{g}(X_{i,T}), \quad (14)$$

for some non-negative weight function $\tilde{g}(\cdot)$. As highlighted in [61], we choose a bandwidth \hat{h} among $h \in [a_T, b_T]$ that minimizes (14). For bandwidths that are locally chosen by data-driven method, according to [5], we replace (14) by

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

In practice, for $i \in \{1, \dots, T\}$, we take the uniform global weights $\tilde{g}(X_{i,T}) = 1$, or the local weights

$$\hat{g}(X_{i,T}, x) = \begin{cases} 1 & \text{if } D(X_{i,T}, x) \leq h, \\ 0 & \text{otherwise.} \end{cases}$$

4 Numerical experiments

To illustrate the convergence of NW estimator wrt Wasserstein distance, we conduct numerical experiments using synthetic and real-world datasets.

4.1 Synthetic data

We generate samples $(X_{t,T}, Y_{t,T})_{t=1, \dots, T}$ using examples provided in [68]. We consider two locally stationary processes.

Generation of functional covariates. We generate the functional covariate from a Hilbert space $\mathcal{H} = L^2_{\mathbb{R}}([0, 1])$, using the following examples:

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

$$X_{t,T}(\tau) = B_{t/T}(X_{t-1,T})(\tau) + \eta_t(\tau), \quad \tau \in [0, 1], \quad t = 1, \dots, T, \quad (15)$$

with a linear operator $B_{t/T}$ indexed by rescaled time $u = \frac{t}{T}$ and innovation function η_t . The innovation η_t is a linear combination of the Fourier basis function $(\psi_j)_{j \in \mathbb{N}}$ with coefficients $\langle \eta_t, \psi_j \rangle$ that are generated from independent zero-mean Gaussian distribution with j th coefficient having variance $(\pi(j - 1.5))^{-2}$, that is,

$$\eta_t = \sum_{j=1}^{\infty} \langle \eta_t, \psi_j \rangle \psi_j \quad \text{with } \langle \eta_t, \psi_j \rangle \sim \mathcal{N}(0, (\pi(j - 1.5))^{-2}),$$

where

$$\psi_j(\tau) = \begin{cases} \sqrt{2} \sin(\pi j \tau), & \text{for odd } j, \\ \sqrt{2} \cos(\pi j \tau), & \text{for even } j. \end{cases}$$

In application, we truncate an infinite-dimensional series at some J basis functions. Now, instead of decomposing $X_{t,T}$ on the basis $(\psi_j)_{j \in \mathbb{N}}$ by $\sum_{j=1}^{\infty} \langle X_{t,T}, \psi_j \rangle \psi_j$, it can be represented by an approximate finite-dimensional $X_{t,T}$:

$$X_{t,T} = \sum_{j=1}^J \langle X_{t,T}, \psi_j \rangle \psi_j.$$

Hence, $\mathbf{X}_{t,T} \approx \mathbf{B}_{t/T} \mathbf{X}_{t-1,T} + \boldsymbol{\eta}_t$, $t = 1, \dots, T$, where $\mathbf{X}_{t,T} = (\langle X_{t,T}, \psi_1 \rangle, \dots, \langle X_{t,T}, \psi_J \rangle)'$, $\boldsymbol{\eta}_t = (\langle \eta_t, \psi_1 \rangle, \dots, \langle \eta_t, \psi_J \rangle)'$, and $\mathbf{B}_{t/T} = (\langle B_{t/T}(\psi_i), \psi_j \rangle)_{1 \leq i, j \leq J}$. In this example, the matrix $\mathbf{B}_{t/T}$ is defined as $\mathbf{B}_{t/T} = \frac{0.4 \mathbf{A}_{t/T}}{\|\mathbf{A}_{t/T}\|_\infty}$, where $\mathbf{A}_{t/T}$ is a $J \times J$ matrix with entries $A_{t/T}(i, j)$ that are mutually independent zero-mean Gaussian random variables with variance $\frac{t}{T^i 6} + (1 - \frac{t}{T})e^{-j-i}$ and $\|A\|_\infty = \sup_{\|x\| \leq 1} \|Ax\|$ is a Schatten ∞ -norm. Figure 1 shows the plot of $X_{t,T}(\tau)$ for $T = 100$. This example was also used in [1].

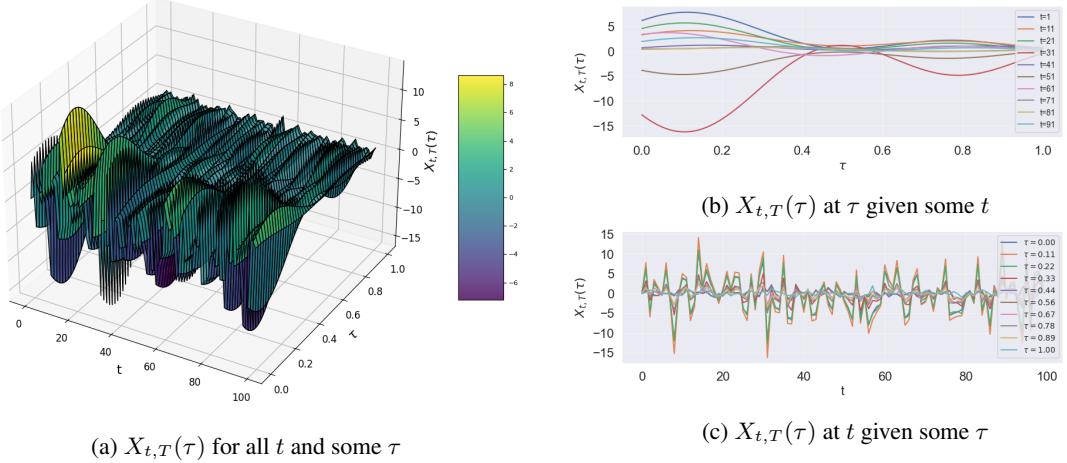


Figure 1: Realizations of Gaussian tvFAR(1) $X_{t,T}(\tau)$ for all t and some τ for $T = 100$ with $J = 7$ and $N = 100$ discretization points of $\tau \in [0, 1]$.

EXAMPLE 2. GAUSSIAN tvFAR(2). We next consider the time-varying functional autoregressive process of order 2, tvFAR(2), with Gaussian noise defined by

$$X_{t,T}(\tau) = B_{t/T,1}(X_{t-1,T})(\tau) + B_{t/T,2}(X_{t-2,T})(\tau) + \eta_t(\tau), \quad \tau \in [0, 1], \quad t = 1, \dots, T, \quad (16)$$

where $B_{t/T,1}$ and $B_{t/T,2}$ are linear operators indexed by the rescaled time $u = \frac{t}{T}$ and innovation function η_t is a linear combination of the Fourier basis function $(\psi_j)_{j \in \mathbb{N}}$. The parameters are set similarly to (15) with $B_{t/T,1} = \frac{0.4 \cos(1.5 - \cos(\pi \frac{t}{T})) \mathbf{A}_{t/T,1}}{\|\mathbf{A}_{t/T,1}\|_\infty}$ and $B_{t/T,2} = \frac{-0.5 \mathbf{A}_{t/T,2}}{\|\mathbf{A}_{t/T,2}\|_\infty}$, where $A_{t/T,1}(i, j)$ and $A_{t/T,2}(i, j)$ are mutually independent-centered Gaussian random variables with variances $e^{-(i-3)-(j-3)}$ and $1/(i^4 + j)$, respectively. Realizations of $X_{t,T}(\tau)$ in this example are depicted in Figure 2.

Generation of response variables. Using locally stationary covariates $X_{t,T}$ in Examples 1 and 2, the response variables $Y_{t,T}$ are generated by (1) with $\varepsilon_{t,T} \sim \mathcal{N}(0, 1)$ and

$$m^*\left(\frac{t}{T}, x\right) = 2.5 \sin\left(2\pi \frac{t}{T}\right) \int_0^1 \cos(\pi x(\tau)) d\tau.$$

Figure 3 shows the time plots of the responses for each process using $T = 1000$, whose values remain tight with constant mean.

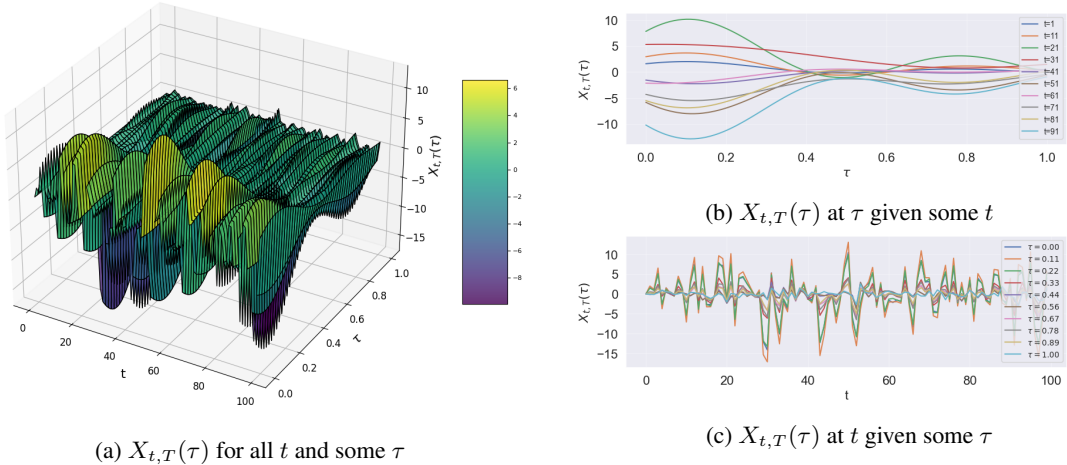


Figure 2: Realizations of Gaussian tvFAR(2) $X_{t,T}(\tau)$ for all t and some τ for $T = 100$ with $J = 7$ and $N = 100$ discretization points of $\tau \in [0, 1]$.

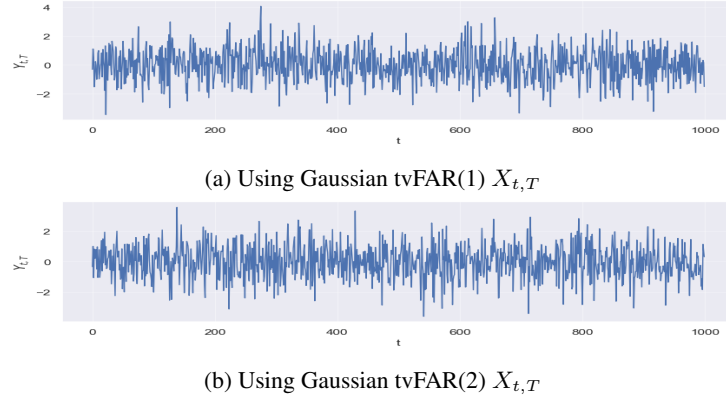


Figure 3: Time plots of response variables for $T = 1000$

Monte Carlo simulations. Using an identical Monte Carlo simulation process in [66], we calculate the NW estimator and true conditional probability distribution for a fixed time $t \in \{1, \dots, T\}$. Each process is replicated using $L = 500$, and as described in Algorithm 1 of [66], for each $l \in \{1, \dots, L\}$, we compute the NW conditional CDF at a given time t . We calculate the average NW and the empirical conditional CDFs using these L replications. We then quantify the corresponding Wasserstein distance.

We obtain the expected Wasserstein distance between the underlying conditional distributions by conducting 50 Monte Carlo runs of Algorithm 1. To produce functional covariates, we select $N = 100$ discretization points of $\tau \in [0, 1]$ and set $J = 7$ since results do not vary much wrt J [1]. As specified in Figure 4, we use different kernels K_1 and K_2 for the chosen processes. Increasing sample sizes $T = 500, 1000, 5000, 10000$ are set. The bandwidths are chosen using the cross-validation method introduced in Section 3.3. Our results are valid when $\frac{t}{T} \in I_h$, hence, we fix t such that $\frac{t}{T} \in I_h = [C_1 h, 1 - C_1 h]$ with constant $C_1 = 1$ for time kernels K_1 belonging to Uniform and Tricube.

Figure 4 depicts the expected Wasserstein distances for each identified process. Wasserstein distance decreases as the sample size T increases. This emphasizes that NW estimator captures the true distribution better as T grows larger; it provides more representative distributions with reduced deviation from the true distribution. Remarkably, the largest sample size, $T = 10000$, consistently achieves the minimum expected Wasserstein distance. This behavior is consistent across both processes under investigation.

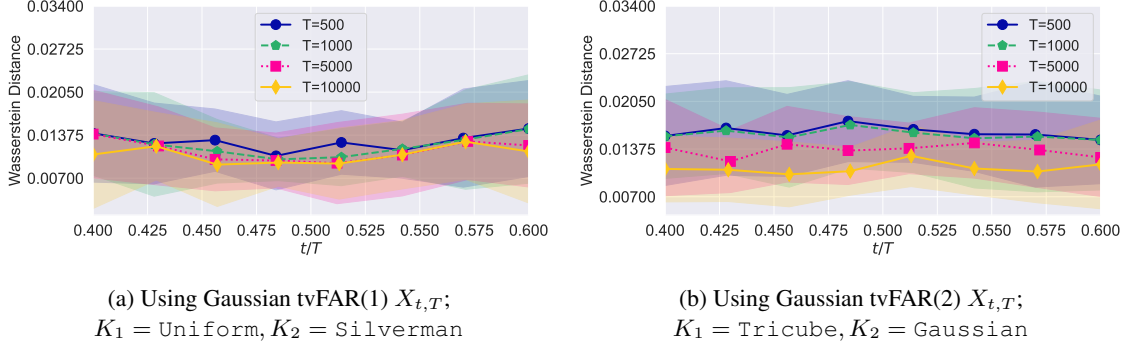


Figure 4: Wasserstein distances \pm standard deviation at different $u = \frac{t}{T}$ for $T = 500, 1000, 5000, 10000$ using $L = 500$ replications and 50 Monte Carlo runs.

4.2 Real-world data

We use two real-world datasets: sea surface temperature (SST) and Nikkie225. To handle these datasets, we employ the same method in [39], described below.

EXAMPLE 3. SST DATA¹. This dataset is used for climate monitoring and research, which is continuously updated by the National Centers for Environmental Information (NCEI) [46]. We take the index from Niño 1+2 region with coordinates 0°- 10°South latitude and 90°West - 80°West longitude. This region covers the eastern equatorial Pacific near the coast of South America and is important for monitoring El Niño and La Niña events. SST contains 900 monthly data points from January 1950 to December 2024, depicted in Figure 5a.

Constructing covariates. To construct $X_{t,T}$, we treat the original series as 25 continuous sample curves, each containing 36 monthly observations as plotted in Figure 5b. Particularly, we let the SST observed for $n = 900$ months be $\{Z(s)\}_{s=1,\dots,n}$, and build, $\forall j \in \{1, \dots, 36\}$,

$$z_{t,T}(j) = Z(36(t-1) + j).$$

The covariates are then constructed as $X_{t,T} = (z_{t,T}(1), \dots, z_{t,T}(36))$ that corresponds to the variations of SST for $t = 1, \dots, 25$. We consider the original time series as 25 dependent functional covariates, $X_{1,25}, \dots, X_{25,25}$, which are individually observed at 36 discretized points.

¹Obtained from <https://www.cpc.ncep.noaa.gov/data/indices/>

Constructing response. We construct the response variables by

$$Y_{t,T}(j) = z_{t+1,T}(j) = Z(36t + j),$$

for a fixed $j \in \{1, \dots, 36\}$ and $t = 1, \dots, 24$. This enables us to generate new 24 functional pairs $\{(X_{t,24}, Y_{t,24}(j))\}_{t=1, \dots, 24}$ [39]. Figure 5c presents sample plots of $Y_{t,24}(j)$ for some j .

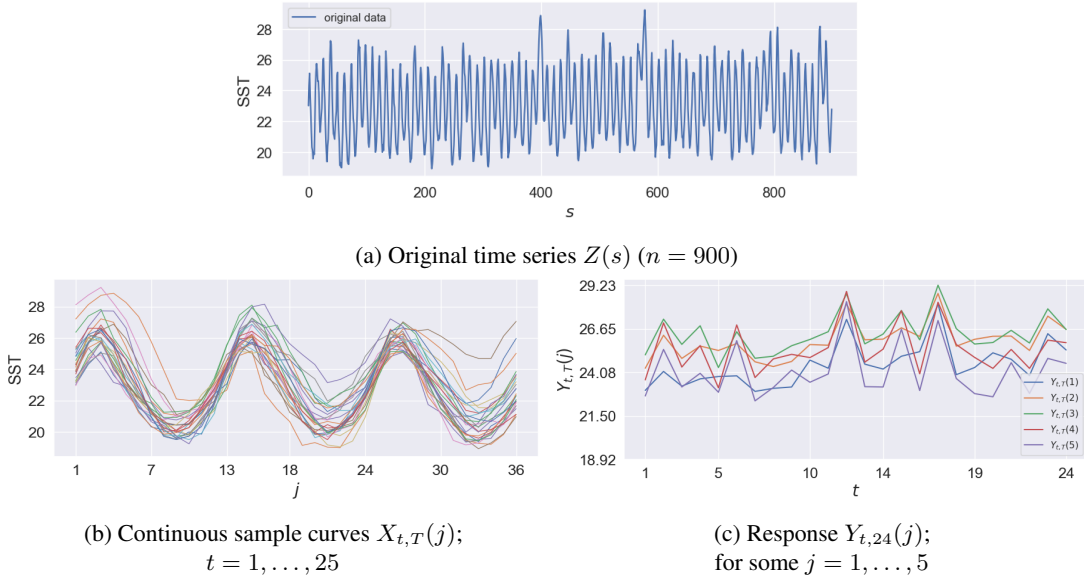


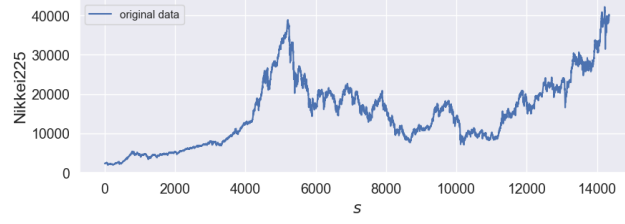
Figure 5: SST monthly time series from Jan. 1950 - Dec. 2024

EXAMPLE 4. NIKKEI225 DATA². We next use the Nikkei stock market index dataset or Nikkei225, a key indicator of the Japanese stock market's overall health. The index tracks the performance of 225 large and active companies listed on the Tokyo Stock Exchange (TSE) [7]. We consider 14340 Nikkei225 data points covering January 14, 1971 to December 31, 2024, plotted in Figure 6a. We construct 239 continuous sample curves by segmenting the original time series $\{Z(s)\}_{s=1, \dots, 14340}$ by 60 observations. Figure 6b reflects 50 examples of the generated continuous sample curves. We use the same method in Example 5 to generate the functional pairs $\{(X_{t,238}, Y_{t,238}(j))\}_{t=1, \dots, 238}$ where $j \in \{1, \dots, 60\}$. The behavior of the response variable is plotted in Figure 6c.

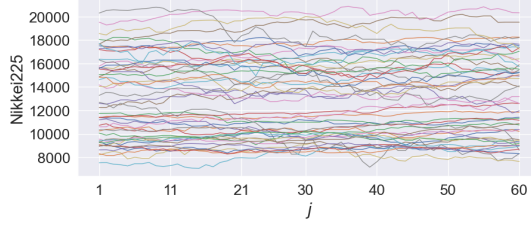
We create copies of these datasets using the same method, Algorithm 2 used in [66], that relies on Gaussian smoothed procedure [56]. For a chosen j th continuous sample curve, we add $Z_{t,T} \sim \mathcal{N}(0, \sigma^2)$ to each data observation $Y_{t,T}$ with $\sigma > 0$, for all $t \in \{1, \dots, T\}$. The Gaussian-smoothed datasets are replicated $L = 500$ times. We then calculate NW conditional CDF for each replicate at a specific time point t . We measure the Wasserstein distance between the average NW and the empirical conditional CDFs.

We refine the segmentation of each dataset to increase the sample size, T . By dividing the 900 monthly SST observations into segments of 12 and 6 months, we generate sample sizes of $T = 74$ and $T = 149$, respectively. Similarly, we split the 14340 observations of Nikkei225 into 30 and 15

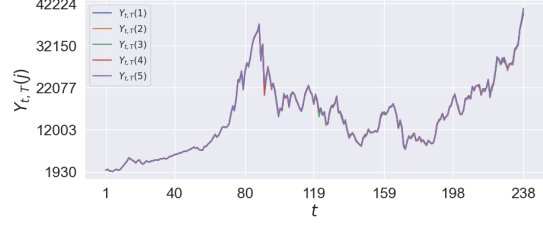
²Obtained from <https://fred.stlouisfed.org/series/NIKKEI225>



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

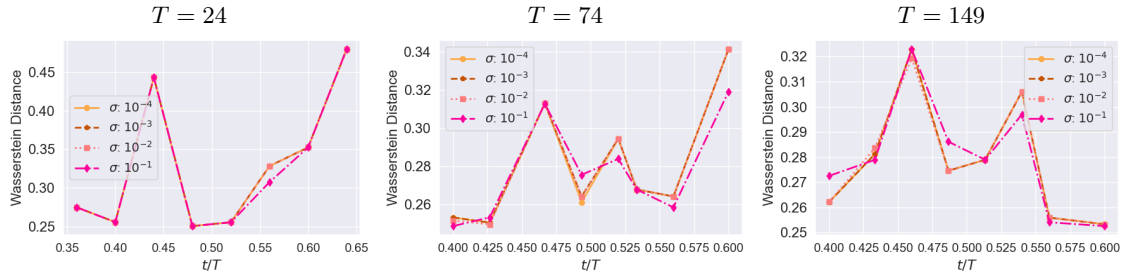


(b) Continuous sample curves $X_{t,T}(j)$;
 $t = 151, \dots, 200$

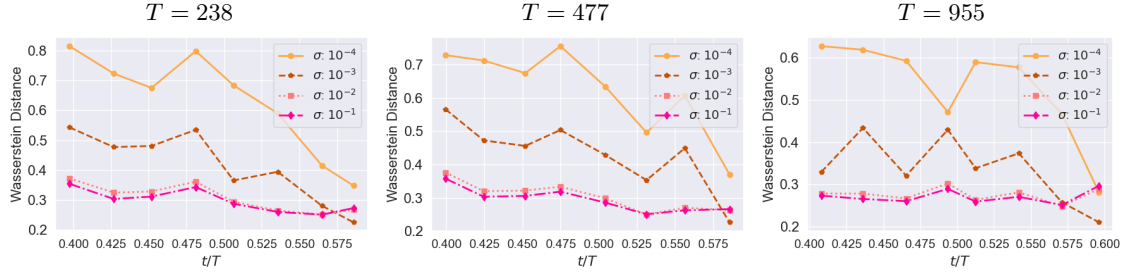


(c) Response $Y_{t,238}(j)$;
for some $j = 1, \dots, 5$

Figure 6: Nikkie225 time series from Jan. 14, 1970 - Dec. 31, 2024



(a) SST



(b) Nikkie225

Figure 7: Wasserstein distance at various $u = \frac{t}{T}$ with different smoothness level σ , $K_1 = \text{Uniform}$ and $K_2 = \text{Silverman}$ at increasing T using $L = 500$ replications.

intervals, yielding sample sizes of $T = 477$ and $T = 955$, respectively. Hence, in this experiment, we set $T = 24, 74, 149$ for SST and $T = 238, 477, 955$ for Nikkie225. We set different values of the smoothing parameter $\sigma \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. We use Uniform and Silverman kernels for K_1 and K_2 , respectively, to quantify NW conditional CDF. Like synthetic data experiments, the bandwidths are selected using a cross-validation method. We select t such that $\frac{t}{T} \in [h, 1 - h]$

since we use a uniform kernel for K_1 . For SST, we selected $j = 21$ from $\{1, \dots, 36\}$, $j = 5$ from $\{1, \dots, 12\}$, and $j = 3$ from $\{1, \dots, 6\}$. Then, for Nikkei225, we fixed $j = 10$ from $\{1, \dots, 60\}$, $j = 10$ from $\{1, \dots, 30\}$, and $j = 5$ from $\{1, \dots, 15\}$. The resulting Wasserstein distances are shown in Figure 7 that depicts similar behavior with the results in [66]. For each dataset, Wasserstein distances for larger sample sizes are slightly lower and are higher for $\sigma \rightarrow 0$.

5 Conclusion

We proposed a NW conditional distribution estimator for LSFTS and established its convergence rates with respect to Wasserstein distance. These rates depend on the bandwidth h and the small ball probability $\phi(h)$. We provided the convergence rates for a fractal-type process with $h = \mathcal{O}(T^{-\xi})$ and $\phi(h) = h^{\tau_0}$, for $0 < \xi < \frac{1}{2(1+\tau_0)}$ and $\tau_0 > 1$. Numerical synthetic and real-world data experiments were conducted, supported by a data-generating algorithm designed to calculate the NW estimator.

This work also outlines promising directions for future research. One avenue involves modifying the basic indicator function to an integrated kernel $H_g(y - Y_{t,T})$, where H is a smooth cumulative distribution function (CDF) and $H_g(y - Y_{t,T})$ serves as a local weighting function with bandwidth g , analogous to h . Another possible extension is amending the NW estimator to handle missing data. While expanding our results to encompass functional ergodic data would be highly valuable, it requires substantial mathematical advancements and lies beyond the current scope of this paper.

Acknowledgements. Mr. Tinio acknowledges the support provided by the Department of Science and Technology - Science Education Institute (DOST-SEI) in partnership with Campus France through the PhilFrance-DOST Scholarship grant.

A Numerical experiment algorithms

We use the following algorithms, which are based on the approach presented in [66], to generate data and calculate NW.

Algorithm 1: Data generating and NW estimation for synthetic data [66]

1. **input** : sample size T , time point $t \in \{1, \dots, T\}$, N spatial discretization points of $\tau \in [0, 1]$, J basis functions, number of replications L , based kernels $K_1(\cdot)$, $K_2(\cdot)$, bandwidth h ;
 2. **for** $l = 1, \dots, L$ **do**
 - # Generate l -th replication process $\{Y_{a,T}^{(l)}\}_{a=1, \dots, T}$ with functional covariates $\{X_{a,T}^{(l)}\}_{a=1, \dots, T}$ constructed using (15) or (16)
 - for** $a = 1, \dots, T$ **do**
 - $Y_{a,T}^{(l)} \leftarrow m^* \left(\frac{a}{T}, X_{a,T}^{(l)} \right) + \varepsilon_{a,T}^{(l)}$;
 - # Calculate l -th NW conditional CDF estimator
 - $\hat{F}_t^{(l)}(y|x) \leftarrow \sum_{a=1}^T \omega_a \left(\frac{t}{T}, x \right) \mathbb{1}_{Y_{a,T}^{(l)} \leq y}$;
 - # Calculate average NW estimator
 3. $\hat{F}_t^L(y|x) \leftarrow \frac{1}{L} \sum_{l=1}^L \hat{F}_t^{(l)}(y|x)$;
 - # Calculate empirical conditional CDF
 4. $F_t^L(y|x) \leftarrow \frac{1}{L} \sum_{l=1}^L \mathbb{1}_{Y_{a,T}^{(l)} \leq y}$;
 5. **return** : $W_1(\hat{F}_t^L(y|x), F_t^L(y|x))$;
-

Algorithm 2: Gaussian smoothed procedure and NW estimation for real datasets [66]

1. **input** : real dataset $\{(X_{a,T}, Y_{a,T}(j))\}_{a=1, \dots, T}$ for fixed j , $\sigma > 0$, time point $t \in \{1, \dots, T\}$, number of replications L , based kernels $K_1(\cdot)$, $K_2(\cdot)$, bandwidth h ;
 2. **for** $l = 1, \dots, L$ **do**
 - # Generate l -th replication $\{Y_{a,T}^{(l)}\}_{a=1, \dots, T}$
 - for** $a = 1, \dots, T$ **do**
 - $Y_{a,T}^{(l)} \leftarrow Y_{a,T}(j) + Z_{a,T}^{(l)}$, where $Z_{a,T}^{(l)} \sim \mathcal{N}(0, \sigma^2)$;
 - # Calculate l -th NW conditional CDF estimator
 - $\hat{F}_t^{(l)}(y|x) \leftarrow \sum_{a=1}^T \omega_a \left(\frac{t}{T}, x \right) \mathbb{1}_{Y_{a,T}^{(l)} \leq y}$;
 - # Calculate average NW estimator
 3. $\hat{F}_t^L(y|x) \leftarrow \frac{1}{L} \sum_{l=1}^L \hat{F}_t^{(l)}(y|x)$;
 - # Calculate empirical conditional CDF
 4. $F_t^L(y|x) \leftarrow \frac{1}{L} \sum_{l=1}^L \mathbb{1}_{Y_{a,T}^{(l)} \leq y}$;
 5. **return** : $W_1(\hat{F}_t^L(y|x), F_t^L(y|x))$;
-

B Proofs of the main results

We begin with the following propositions that will be useful in the succeeding proofs. For the sake of completeness and consistency, the lines of proofs are adapted from [66] where we introduce the semi-metric $D(\cdot, \cdot)$.

Proposition 3. *Let Assumptions 1 to 5 hold. Then, for $a, t \in \{1, \dots, T\}$, the following inequalities hold:*

- (i) $\mathbb{E}[K_{h,2}(D(x, X_{a,T})) - K_{h,2}(D(x, X_a(\frac{a}{T})))] \leq \frac{L_2 C_U}{Th}$.
- (ii) $\mathbb{E}[K_{h,2}(D(x, X_{a,T}))] \leq \frac{L_2 C_U}{Th} + C_d \phi(h) \psi(x)$.
- (iii) $K_{h,1}(\frac{t}{T} - \frac{a}{T}) \mathbb{E}[K_{h,2}(D(x, X_{t,T}))[\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(\cdot|x)]]$
 $\leq (C_1 + C_2) L_{F^*} K_{h,1}(\frac{t}{T} - \frac{a}{T}) \left\{ \frac{L_2 C_U}{T} + C_d h \phi(h) \psi(x) \right\}$,

where $\psi(x)$ is a nonnegative functional in $x \in \mathcal{H}$.

Proposition 4. *Let Assumptions 1 - 4 hold, then*

$$J_{t,T}^{-1}(\frac{t}{T}, x) = \left(\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}(\frac{t}{T} - \frac{a}{T}) K_{h,2}(D(x, X_{t,T})) \right)^{-1} = \mathcal{O}_{\mathbb{P}}(1).$$

Proposition 5. *Let Assumptions 1 - 7 be satisfied. For $y \in \mathbb{R}$ and $x \in \mathcal{H}$, define*

$$Z_{t,T}(y, x) = \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}(\frac{t}{T} - \frac{a}{T}) K_{h,2}(D(x, X_{t,T})) [\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)],$$

then

$$\mathbb{E}[Z_{t,T}^2(y, x)] = \mathcal{O}\left(\frac{1}{Th^2\phi^2(h)} + h^2\right).$$

The proofs of Propositions 3 to 5 are shown in Appendix C.

B.1 Proof of Theorem 1

Recall that $\pi_t^*(\cdot|x)$ is the probability measure of the random variable $Y_{t,T}|X_{t,T} = x$ with conditional CDF $F_t^*(y|x) = \mathbb{P}[Y_{t,T} \leq y|X_{t,T} = x]$. Observe that, by the definition of W_1 given in (4),

$$\begin{aligned} \mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] &= \mathbb{E}\left[\int |\hat{F}_t(y|x) - F_t^*(y|x)| dy\right] \\ &= \int \mathbb{E}[|\hat{F}_t(y|x) - F_t^*(y|x)|] dy, \end{aligned}$$

using Fubini's theorem. Observe that, using (6) and (7),

$$\begin{aligned}
\hat{F}_t(y|x) - F_t^*(y|x) &= \frac{\sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T})) \mathbf{1}_{Y_{a,T} \leq y} - F_t^*(y|x)}{\sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T}))} \\
&= \frac{\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T})) [\mathbf{1}_{Y_{a,T} \leq y} - F_t^*(y|x)]}{\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T}))}.
\end{aligned} \tag{17}$$

Further, by applying Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] \\
&= \int \mathbb{E} \left[\left| \frac{\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T})) [\mathbf{1}_{Y_{a,T} \leq y} - F_t^*(y|x)]}{\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T}))} \right| \right] dy \\
&\leq \int \left(\mathbb{E} \left[\left(\frac{1}{\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T}))} \right)^2 \right] \right)^{\frac{1}{2}} \\
&\quad \times \left(\mathbb{E} \left[\left(\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T})) [\mathbf{1}_{Y_{a,T} \leq y} - F_t^*(y|x)] \right)^2 \right] \right)^{\frac{1}{2}} dy.
\end{aligned} \tag{18}$$

Let $J_{t,T}\left(\frac{t}{T}, x\right) = \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T}))$. Using Proposition 4, $J_{t,T}^{-1}\left(\frac{t}{T}, x\right) = \mathcal{O}_{\mathbb{P}}(1)$. Hence, the first term in (18) becomes

$$\left(\mathbb{E} \left[\left(\frac{1}{\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}(\mathbf{D}(x, X_{a,T}))} \right)^2 \right] \right)^{\frac{1}{2}} = \mathcal{O}_{\mathbb{P}}(1). \tag{19}$$

Additionally, from Proposition 5, the second term in (18) is shown to be $\mathcal{O}\left(\frac{1}{T^{\frac{1}{2}}h\phi(h)} + h\right)$. Therefore, from (18) and combining (41) and (19), we have

$$\mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{T^{\frac{1}{2}}h\phi(h)} + h\right).$$

B.2 Proof of Corollary 1

Using the definition of W_1 and noting that $y \in [-M, M]$, we have

$$W_r^r(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x)) \leq (2M)^{r-1} \int_{-M}^M |\hat{F}_t(y|x) - F_t^*(y|x)| dy.$$

So,

$$\begin{aligned}
\mathbb{E}[W_r^r(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] &\leq (2M)^{r-1} \mathbb{E} \left[\int_{-M}^M |\hat{F}_t(y|x) - F_t^*(y|x)| dy \right] \\
&\leq (2M)^{r-1} \mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))].
\end{aligned}$$

By Theorem 1, we get the desired result.

B.3 Proof of Corollary 2

Again, we used the definition of W_1 given by (4). Additionally, by using Minkowski's integral inequality given by (12), for $r = 2$ we have

$$\begin{aligned} \|W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))\|_{L_2} &= \left\| \int_{\mathbb{R}} |\hat{F}_t(y|x) - F_t^*(y|x)| dy \right\|_{L_2} \\ &\leq \int_{\mathbb{R}} \|\hat{F}_t(y|x) - F_t^*(y|x)\|_{L_2} dy \\ &= \int_{\mathbb{R}} (\mathbb{E}[(\hat{F}_t(y|x) - F_t^*(y|x))^2])^{\frac{1}{2}} dy \\ &= \int_{\mathbb{R}} \left(\mathbb{E} \left[\left(\frac{Z_{t,T}(y,x)}{J_{t,T}(\frac{t}{T}, x)} \right)^2 \right] \right)^{\frac{1}{2}} dy, \end{aligned}$$

using (17) and (21). However, using Proposition 4, $J_{t,T}^{-1}(\frac{t}{T}, x) = \mathcal{O}_{\mathbb{P}}(1)$. So

$$\begin{aligned} \|W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))\|_{L_2} &\lesssim \int_{\mathbb{R}} (\mathbb{E}[Z_{t,T}^2(y,x)])^{\frac{1}{2}} dy \\ &\lesssim \int_{\mathbb{R}} \left(\frac{1}{Th^2\phi^2(h)} + h^2 \right)^{\frac{1}{2}} dy, \end{aligned}$$

by Proposition 5. Therefore,

$$\|W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))\|_{L_2} = \mathcal{O}_{\mathbb{P}} \left(\frac{1}{T^{\frac{1}{2}} h \phi(h)} + h \right).$$

B.4 Proof of Proposition 1

Observe that

$$\begin{aligned} |\hat{m}(\frac{t}{T}, x) - m^*(\frac{t}{T}, x)| &= |\mathbb{E}[\hat{Y}_{t,T}|X_{t,T} = x] - \mathbb{E}[Y_{t,T}|X_{t,T} = x]| \\ &= \left| \int_{\mathbb{R}} \hat{y} d\hat{\pi}_t(\cdot|x) - \int_{\mathbb{R}} y d\pi_t^*(\cdot|x) \right| \\ &\leq \sup_{f \in \mathcal{F}} \left| \int_{\mathbb{R}} f d\hat{\pi}_t(\cdot|x) - \int_{\mathbb{R}} f d\pi_t^*(\cdot|x) \right| \\ &= W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x)). \end{aligned}$$

The duality formula of the Kantorovich-Rubinstein distance is used in the last equality (see Remark 6.5 in [70]), where \mathcal{F} is the set of all continuous functions satisfying the Lipschitz condition $\|f\|_{Lip} \leq 1$, i.e., $\sup_{y \neq y'} \frac{|f(y) - f(y')|}{|y - y'|} \leq 1$. Hence,

$$\mathbb{E}[|\hat{m}(\frac{t}{T}, x) - m^*(\frac{t}{T}, x)|] \leq \mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))].$$

This finishes the proof.

B.5 Proof of Proposition 2

If $h = \mathcal{O}(T^{-\xi})$ and $\phi(h) = h^{\tau_0}$, for $\tau_0 > 1$, then directly from Theorem 1,

$$\begin{aligned} \mathbb{E}[W_1(\hat{\pi}_t(\cdot|x), \pi_t^*(\cdot|x))] &\lesssim \frac{1}{T^{\frac{1}{2}}h\phi(h)} + h \\ &\lesssim \frac{1}{T^{\frac{1}{2}}T^{-\xi}T^{-\xi\tau_0}} + \frac{1}{T^\xi} \\ &\lesssim \frac{1}{T^{\frac{1}{2}-\xi(1+\tau_0)}} + \frac{1}{T^\xi}, \end{aligned}$$

which goes to zero if $0 < \xi < \frac{1}{2(1+\tau_0)}$.

C Proofs of Propositions 3, 4, and 5

C.1 Proof of Proposition 3

(i) Using Assumption 2, we note that K_2 is Lipschitz. In addition, by Assumption 1 and when $u = \frac{t}{T}$, $\mathbb{D}(X_{a,T}, X_a(\frac{a}{T})) \leq \frac{1}{T}U_{t,T}(\frac{a}{T})$, where $\mathbb{E}[(U_{t,T}(\frac{a}{T}))^\rho] < C_U$. So,

$$\begin{aligned} &\mathbb{E}[K_{h,2}(\mathbb{D}(x, X_{a,T})) - K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))] \\ &= \mathbb{E}[K_2(\frac{\mathbb{D}(x, X_{a,T})}{h}) - K_2(\frac{\mathbb{D}(x, X_a(\frac{a}{T}))}{h})] \\ &\leq \frac{L_2}{h} \mathbb{E}[|\mathbb{D}(x, X_{a,T}) - \mathbb{D}(x, X_a(\frac{a}{T}))|] \\ &\leq \frac{L_2}{h} \mathbb{E}[|\mathbb{D}(X_{a,T}, X_a(\frac{a}{T}))|] \\ &\leq \frac{L_2}{h} \mathbb{E}[|\frac{1}{T}U_{t,T}(\frac{a}{T})|] \\ &\leq \frac{L_2 C_U}{Th}. \end{aligned}$$

(ii) We have

$$\begin{aligned} &\mathbb{E}[K_{h,2}(\mathbb{D}(x, X_{a,T}))] \\ &= \mathbb{E}[K_{h,2}(\mathbb{D}(x, X_{a,T})) - K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T}))) + K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))] \\ &= \mathbb{E}[K_{h,2}(\mathbb{D}(x, X_{a,T})) - K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))] + \mathbb{E}[K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))] \\ &\leq \frac{L_2 C_U}{Th} + \mathbb{E}[K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))], \end{aligned}$$

using (i). Now using Assumption 3,

$$\begin{aligned} \mathbb{E}[K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))] &\leq \mathbb{E}[\mathbf{1}_{\mathbb{D}(x, X_a(\frac{a}{T})) \leq h}] = \mathbb{P}[X_a(\frac{a}{T}) \in B(x, h)] \\ &= F_{t/T}(h; x) \leq C_d \phi(h) \psi(x). \end{aligned}$$

Hence,

$$\mathbb{E}[K_{h,2}(D(x, X_{t,T}))] \leq \frac{L_2 C_U}{Th} + C_d \phi(h) \psi(x).$$

(iii) Note that using Assumption 5, $|F_a^*(y|X_{a,T}) - F_t^*(y|x)| \leq L_{F^*}(D(x, X_{a,T}) + |\frac{a}{T} - \frac{t}{T}|)$. Now see that

$$\begin{aligned} & K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))[\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)]\right] \\ & \leq K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))\mathbb{E}\left[(\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)) \middle| X_{a,T}\right]\right] \\ & \leq K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))|F_a^*(y|X_{a,T}) - F_t^*(y|x)|\right] \\ & \leq L_{F^*} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))\left(D(x, X_{a,T}) + \left|\frac{a}{T} - \frac{t}{T}\right|\right)\right]. \end{aligned}$$

However, using Assumption 2, $D(x, X_{a,T}) \leq C_2 h$ otherwise, $K_{h,2}(D(x, X_{a,T})) = 0$. Additionally, $|\frac{a}{T} - \frac{t}{T}| \leq C_1 h$ otherwise, $K_{h,1}(|\frac{a}{T} - \frac{t}{T}|) = 0$. So,

$$\begin{aligned} & K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))[\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)]\right] \\ & \leq L_{F^*} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \left\{ C_2 h \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))\right] + C_1 h \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))\right] \right\} \\ & \leq (C_1 + C_2) L_{F^*} h K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))\right] \\ & \leq (C_1 + C_2) L_{F^*} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \left\{ \frac{L_2 C_U}{T} + C_d h \phi(h) \psi(x) \right\}, \end{aligned}$$

using (ii).

C.2 Proof of Proposition 4

By applying Theorem 3.1 in [48], $|J_{t,T}(\frac{t}{T}, x) - \mathbb{E}[J_{t,T}(\frac{t}{T}, x)]| = \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log T}{Th\phi(h)}}\right)$. Additionally, using Assumption 1, $J_{t,T}(\frac{t}{T}, x)$ can be decomposed as $J_{t,T}(\frac{t}{T}, x) = \tilde{J}_{t,T}(\frac{t}{T}, x) + \bar{J}_{t,T}(\frac{t}{T}, x)$. So,

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

where

$$\tilde{J}_{t,T}\left(\frac{t}{T}, x\right) = \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}\left(\mathbb{D}\left(x, X_a\left(\frac{a}{T}\right)\right)\right),$$

and

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

Now, let us first observe $\mathbb{E}[\bar{J}_{t,T}(\frac{t}{T}, x)]$. Using Assumptions 1 and 2 together with Proposition 3.i, we have

$$\begin{aligned} \mathbb{E}[\bar{J}_{t,T}\left(\frac{t}{T}, x\right)] &= \mathbb{E}\left[\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \left\{ K_{h,2}\left(\mathbb{D}\left(x, X_{a,T}\right)\right) - K_{h,2}\left(\mathbb{D}\left(x, X_a\left(\frac{a}{T}\right)\right)\right) \right\}\right] \\ &= \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[\left\{ K_{h,2}\left(\mathbb{D}\left(x, X_{a,T}\right)\right) - K_{h,2}\left(\mathbb{D}\left(x, X_a\left(\frac{a}{T}\right)\right)\right) \right\}\right] \\ &\leq \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \left(\frac{L_2 C_U}{Th}\right) \\ &\leq \frac{L_2 C_U}{Th\phi(h)} \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right)}_{\mathcal{O}(1)} \\ &\leq \frac{L_2 C_U}{Th\phi(h)} \\ &\lesssim \frac{1}{Th\phi(h)}, \end{aligned}$$

which converges to zero using Assumption 4. In the lines above, $\frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) = \mathcal{O}(1)$ since, using Lemma B.2 in [71], for $I_h = [C_1 h, 1 - C_1 h]$,

$$\begin{aligned} \frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) &\leq \sup_{u \in I_h} \left| \frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(u - \frac{a}{T}\right) \right| \\ &\leq \sup_{u \in I_h} \left| \frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(u - \frac{a}{T}\right) - 1 \right| + 1 \\ &= \mathcal{O}\left(\frac{1}{Th^2}\right) + o(h) + 1 = \mathcal{O}(1). \end{aligned} \tag{20}$$

On the other hand,

$$\mathbb{E}[\tilde{J}_{t,T}\left(\frac{t}{T}, x\right)] = \mathbb{E}\left[\frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,2}\left(\mathbb{D}\left(x, X_a\left(\frac{a}{T}\right)\right)\right)\right]$$

$$= \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}[K_{h,2}(\mathbb{D}(x, X_a(\frac{a}{T})))].$$

Using equation (4.3) in [41], we have

$$\begin{aligned} \mathbb{E}[\tilde{J}_{t,T}(\frac{t}{T}, x)] &= \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}[\mathbb{1}_{(\mathbb{D}(x, X_a(\frac{a}{T}))) \leq h}] \\ &= \frac{1}{Th\phi(h)} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) F_{t/T}(h; x) \\ &\geq \frac{1}{\phi(h)} \frac{1}{Th} \underbrace{\sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right)}_{\mathcal{O}(1)} c_d \phi(h) \psi(x) \quad (\text{using Assumption 3}) \\ &\sim \psi(x) > 0, \end{aligned}$$

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

$$\begin{aligned} \frac{1}{J_{t,T}(\frac{t}{T}, x)} &= \frac{1}{o_{\mathbb{P}}(1) + o(1) + \mathbb{E}[\tilde{J}_{t,T}(\frac{t}{T}, x)]} \\ &= \mathcal{O}_{\mathbb{P}}(1). \end{aligned}$$

C.3 Proof of Proposition 5

Let

$$Z_{t,T}(y, x) := \frac{1}{Th^{d+1}} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) Z_{a,t,T}(y, x), \quad (21)$$

where

$$Z_{a,t,T}(y, x) = K_{h,2}(\mathbb{D}(x, X_{a,T})) [\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)].$$

Applying Bernstein's big-block and small-block procedure on $Z_{t,T}(y, x)$, we partition the set $\{1, \dots, T\}$ into $2v_T + 1$ independent subsets: v_T big blocks of size r_T , v_T small blocks of size s_T , and a remainder block of size $T - v_T(r_T + s_T)$, where $v_T = \lfloor \frac{T}{r_T + s_T} \rfloor$. To establish independence between the blocks, we need to place the asymptotically negligible small blocks in between two consecutive big blocks. This procedure was also used in [51, 14]. So, we decompose $Z_{t,T}(y, x)$ as

$$\begin{aligned} Z_{t,T}(y, x) &= \Lambda_{t,T}(y, x) + \Pi_{t,T}(y, x) + \Xi_{t,T}(y, x) \\ &:= \sum_{l=0}^{v_T-1} \Lambda_{l,t,T}(y, x) + \sum_{l=0}^{v_T-1} \Pi_{l,t,T}(y, x) + \Xi_{t,T}(y, x), \end{aligned} \quad (22)$$

where

$$\Lambda_{l,t,T}(y, x) = \frac{1}{Th^{d+1}} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) Z_{a,t,T}(y, x),$$

$$\Pi_{l,t,T}(y, x) = \frac{1}{Th^{d+1}} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) Z_{a,t,T}(y, x),$$

and

$$\Xi_{t,T}(y, x) = \frac{1}{Th^{d+1}} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) Z_{a,t,T}(y, x).$$

Let us define the size of the big blocks as $r_T = \lfloor \sqrt{Th\phi(h)}/q_T \rfloor$, where q_T satisfies Assumption 7, i.e., $q_T = o(\sqrt{Th\phi(h)})$. This further implies that there exists a sequence of positive integers $\{q_T\}$, $q_T \rightarrow \infty$, such that $q_T s_T = o(\sqrt{Th\phi(h)})$. Additionally, as $T \rightarrow \infty$,

$$\frac{s_T}{r_T} \rightarrow 0, \quad \text{and} \quad \frac{r_T}{T} \rightarrow 0. \quad (23)$$

Note that defining $r_T = \lfloor \sqrt{Th\phi(h)}/q_T \rfloor$ immediately implies that $r_T = o(\sqrt{Th\phi(h)})$. Additionally, note that $s_T = o(r_T)$ and $v_T = o(q_T \sqrt{Th\phi(h)})$. Now,

$$\begin{aligned} \mathbb{E}[Z_{t,T}^2(y, x)] &= \mathbb{E}[\Lambda_{t,T}^2(y, x)] + \mathbb{E}[\Pi_{t,T}^2(y, x)] + \mathbb{E}[\Xi_{t,T}^2(y, x)] \\ &+ 2\left\{ \mathbb{E}[\Lambda_{t,T}(y, x)\Pi_{t,T}(y, x)] + \mathbb{E}[\Lambda_{t,T}(y, x)\Xi_{t,T}(y, x)] + \mathbb{E}[\Pi_{t,T}(y, x)\Xi_{t,T}(y, x)] \right\}. \end{aligned}$$

However, the defined size of big blocks and the relation (23) ensure that the blocks are asymptotically independent and the sums of small blocks and the remainder block are asymptotically negligible. Consequently, we can neglect the last terms in the previous equation. Hence, we have

$$\mathbb{E}[Z_{t,T}^2(y, x)] \approx \mathbb{E}[\Lambda_{t,T}^2(y, x)] + \mathbb{E}[\Pi_{t,T}^2(y, x)] + \mathbb{E}[\Xi_{t,T}^2(y, x)].$$

For convenience of notation, in the succeeding steps, we let the dependency on y and x be implicit.

Step 1. Control of the big blocks. First, let us start by dealing with $\mathbb{E}[\Lambda_{t,T}^2]$. One has

$$\begin{aligned} \mathbb{E}[\Lambda_{t,T}^2] &= \sum_{l=0}^{v_T-1} \mathbb{E}[\Lambda_{l,t,T}^2] + \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \mathbb{E}[\Lambda_{l,t,T}] \mathbb{E}[\Lambda_{l',t,T}] \\ &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \mathbb{E}\left[\left(\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) Z_{a,t,T} \right)^2 \right] \\ &+ \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\ &\quad \times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}]. \end{aligned}$$

Observe that

$$\begin{aligned}
\mathbb{E}[\Lambda_{t,T}^2] &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}[Z_{a,t,T}^2] \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{a=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} \sum_{b=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\
&\quad \quad \quad \times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}] \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\
&\quad \quad \quad \times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}] \\
&=: \Sigma_1^\Lambda + \Sigma_2^\Lambda + \Sigma_3^\Lambda.
\end{aligned}$$

Step 1.1. Control of Σ_1^Λ . Considering Σ_1^Λ , we have

$$\begin{aligned}
\Sigma_1^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}[Z_{a,t,T}^2] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}^2(\mathbb{D}(x, X_{a,T})) (\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x))^2\right].
\end{aligned}$$

By Proposition 3.iii, we have

$$\begin{aligned}
&K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}^2(\mathbb{D}(x, X_{a,T})) (\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x))^2\right] \\
&\leq 2C_2 K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \mathbb{E}\left[K_{h,2}(\mathbb{D}(x, X_{a,T})) |\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)|\right] \\
&\leq 2C_2(C_1 + C_2) L_{F^*} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \left\{ \frac{L_2 C_U}{T} + C_d h \phi(h) \psi(x) \right\} \\
&\lesssim K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \left(\frac{1}{T} + h\phi(h)\right).
\end{aligned}$$

So

$$\begin{aligned}
\Sigma_1^\Lambda &\lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right) \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right) \\
&\leq \frac{C_1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right) \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right)}_{\mathcal{O}(1)},
\end{aligned}$$

using (20). So we have

$$\begin{aligned}
\Sigma_1^\Lambda &\lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right) \\
&\lesssim \frac{1}{T^2h\phi^2(h)} + \frac{1}{T\phi(h)} \\
&\lesssim \frac{1}{Th\phi^2(h)}.
\end{aligned} \tag{24}$$

Step 1.2. Control of Σ_2^Λ . On the other hand,

$$\begin{aligned}
\Sigma_2^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{\substack{b=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{\substack{b=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \text{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{\substack{b=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\
&\quad \quad \quad \times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\
&:= \Sigma_{21}^\Lambda + \Sigma_{22}^\Lambda.
\end{aligned}$$

Step 1.2.1. Control of Σ_{21}^Λ . Looking at Σ_{21}^Λ , we have

$$\begin{aligned}
\Sigma_{21}^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{\substack{b=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \\
&\quad \quad \quad \times \text{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\quad \quad \quad \times \text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}) \\
&\leq \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\quad \quad \quad \times |\text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T})|,
\end{aligned}$$

where $\lambda = l(r_T + s_T)$. Note that, by Assumption 6, $\{X_{t,T}, \varepsilon_{t,T}\}$ is regularly mixing. So using Davydov's inequality [30] and Lemma 1 in [66], $\beta(\sigma(X_{\lambda+n_1,t,T}), \sigma(X_{\lambda+n_2,t,T})) \leq \beta(|n_1 - n_2|)$.

Then, for $p > 2$, we have

$$\begin{aligned}
& K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right)\left|\text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T})\right| \\
& \leq 8K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right)\|Z_{\lambda+n_1,t,T}\|_{L_p}\|Z_{\lambda+n_2,t,T}\|_{L_p}\beta(\sigma(X_{\lambda+n_1,t,T}), \sigma(X_{\lambda+n_2,t,T}))^{1-\frac{2}{p}} \\
& \leq 8K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right) \\
& \quad \times \left(\mathbb{E}\left[\left|K_{h,2}(\text{D}(x, X_{\lambda+n_1,T}))(\mathbb{1}_{Y_{\lambda+n_1,T} \leq y} - F_t^*(y|x))\right|^p\right]\right)^{\frac{1}{p}} \\
& \quad \times \left(\mathbb{E}\left[\left|K_{h,2}(\text{D}(x, X_{\lambda+n_2,T}))(\mathbb{1}_{Y_{\lambda+n_2,T} \leq y} - F_t^*(y|x))\right|^p\right]\right)^{\frac{1}{p}}\beta(|n_1 - n_2|)^{1-\frac{2}{p}} \\
& \leq K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right) \\
& \quad \times \left(C_2^{p-1}2^{p-1}\mathbb{E}\left[\left|K_{h,2}(\text{D}(x, X_{\lambda+n_1,T}))\right|\mathbb{1}_{Y_{\lambda+n_1,T} \leq y} - F_t^*(y|x)\right]\right)^{\frac{1}{p}} \\
& \quad \times \left(C_2^{p-1}2^{p-1}\mathbb{E}\left[\left|K_{h,2}(\text{D}(x, X_{\lambda+n_2,T}))\right|\mathbb{1}_{Y_{\lambda+n_2,T} \leq y} - F_t^*(y|x)\right]\right)^{\frac{1}{p}}\beta(|n_1 - n_2|)^{1-\frac{2}{p}} \\
& \lesssim K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)\left(\frac{1}{T} + h\phi(h)\right)^{\frac{1}{p}}K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right)\left(\frac{1}{T} + h\phi(h)\right)^{\frac{1}{p}}\beta(|n_1 - n_2|)^{1-\frac{2}{p}} \\
& \lesssim K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right)\left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}}\beta(|n_1 - n_2|)^{1-\frac{2}{p}}, \tag{25}
\end{aligned}$$

using Proposition 3.iii. In consequence,

$$\begin{aligned}
\Sigma_{21}^\Lambda & \lesssim \frac{1}{(Th\phi(h))^2}\left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{r_T}K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right)K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_2}{T}\right)\beta(|n_1 - n_2|)^{1-\frac{2}{p}} \\
& \leq \frac{C_1^2}{T^2h^2\phi^2(h)}\left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{r_T}\beta(|n_1 - n_2|)^{1-\frac{2}{p}}.
\end{aligned}$$

Using Assumption 6, $\sum_{k=1}^\infty k^\zeta\beta(k)^{1-\frac{2}{p}} < \infty$, which can be expressed as $\sum_{k=1}^{r_T} k^\zeta\beta(k)^{1-\frac{2}{p}} + \sum_{k=r_T+1}^\infty k^\zeta\beta(k)^{1-\frac{2}{p}}$. Now, observe that letting $k = |n_1 - n_2|$ yields

$$\begin{aligned}
\sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{r_T}\sum_{n_2=1}^{r_T}\beta(|n_1 - n_2|)^{1-\frac{2}{p}} & = \sum_{n_1=1}^{r_T}\left(\sum_{n_2>n_1}^{r_T}\beta(n_2 - n_1)^{1-\frac{2}{p}} + \sum_{n_2<n_1}^{r_T}\beta(n_1 - n_2)^{1-\frac{2}{p}}\right) \\
& = \sum_{n_1=1}^{r_T}\sum_{k>0}^{r_T-n_1}\beta(k)^{1-\frac{2}{p}} + \sum_{n_2=1}^{r_T}\sum_{k>0}^{r_T-n_2}\beta(k)^{1-\frac{2}{p}} \\
& = 2\sum_{n=1}^{r_T}\sum_{k>0}^{r_T-n}\beta(k)^{1-\frac{2}{p}} \leq 2r_T\sum_{k=1}^{r_T}\beta(k)^{1-\frac{2}{p}} \\
& \lesssim r_T\sum_{k=1}^{r_T}k^\zeta\beta(k)^{1-\frac{2}{p}} \leq r_T\sum_{k=1}^\infty k^\zeta\beta(k)^{1-\frac{2}{p}},
\end{aligned}$$

since $k^\zeta \geq 1$ for $\zeta > 1 - \frac{2}{p}$, where $p > 2$. Hence

$$\begin{aligned}
\Sigma_{21}^\Lambda &\leq \frac{C_1^2 r_T}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{l=0}^{v_T-1} \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\
&\lesssim \frac{v_T r_T}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\
&\lesssim \frac{1}{T h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}}, \quad \text{since } v_T r_T \leq \frac{T}{r_T} r_T = T, \\
&= \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \left(\frac{1}{T^2} + h^2 \phi^2(h) \right) \right)^{\frac{1}{p}} \\
&\lesssim \left(\frac{1}{T^{2+p} h^{2p} \phi^{2p}(h)} + \frac{1}{T^p h^{2p-2} \phi^{2p-2}(h)} \right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \right)^{\frac{1}{p}} \lesssim \frac{1}{T h^2 \phi^2(h)}. \quad (26)
\end{aligned}$$

Step 1.2.2. Control of Σ_{22}^Λ . Considering Σ_{22}^Λ , see that

$$\begin{aligned}
\Sigma_{22}^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{\substack{b=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{k_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \mathbb{E}[Z_{\lambda+n_1,t,T}] \mathbb{E}[Z_{\lambda+n_2,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\quad \times \mathbb{E}\left[K_{h,2}(\mathbb{D}(x, X_{\lambda+n_1,T})) (\mathbb{1}_{Y_{\lambda+n_1,T} \leq y} - F_t^*(y|x))\right] \\
&\quad \times \mathbb{E}\left[K_{h,2}(\mathbb{D}(x, X_{\lambda+n_2,T})) (\mathbb{1}_{Y_{\lambda+n_2,T} \leq y} - F_t^*(y|x))\right].
\end{aligned}$$

By Proposition 3.iii, for $i = 1, 2$, $K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_i}{T}\right) \mathbb{E}\left[K_{h,2}(\mathbb{D}(x, X_{\lambda+n_i,T})) (\mathbb{1}_{Y_{\lambda+n_i,T} \leq y} - F_t^*(y|x))\right] \lesssim K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_i}{T}\right) \left(\frac{1}{T} + h\phi(h)\right)$, then

$$\begin{aligned}
\Sigma_{22}^\Lambda &\lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h) \right)^2 \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\leq \frac{C_1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right)}_{\mathcal{O}(1)} \lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \\
&\lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T^2} + h^2 \phi^2(h) \right) \lesssim \frac{1}{T^3 h \phi^2(h)} + \frac{h}{T} \lesssim \frac{1}{Th\phi^2(h)}. \quad (27)
\end{aligned}$$

Step 1.2.2. Control of Σ_{22}^Λ . Considering Σ_{22}^Λ , see that

$$\begin{aligned}
\Sigma_3^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \text{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \\
&\hspace{15em} \times \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\
&=: \Sigma_{31}^\Lambda + \Sigma_{32}^\Lambda.
\end{aligned}$$

Step 1.3.1 Control of Σ_{31}^Λ . Looking at Σ_{31}^Λ , we have

$$\begin{aligned}
\Sigma_{31}^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \text{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda' + n_2}{T}\right) \text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda'+n_2,t,T}),
\end{aligned}$$

where $\lambda = l(r_T + s_T)$ and $\lambda' = l'(r_T + s_T)$, however, for $l \neq l'$, see that

$$\begin{aligned}
|\lambda - \lambda' + n_1 - n_2| &\geq |l(r_T + s_T) - l'(r_T + s_T) + n_1 - n_2| \\
&\geq |(l - l')(r_T + s_T) + n_1 - n_2| \\
&> s_T,
\end{aligned}$$

since $n_1, n_2 \in \{1, \dots, r_T\}$. So if we let $m = \lambda + n_1$ and $m' = \lambda' + n_2$, we have

$$\begin{aligned}
\Sigma_{31}^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{m=1}^{v_T(r_T+s_T)-s_T} \sum_{\substack{m'=1 \\ |m-m'| > s_T}}^{v_T(r_T+s_T)-s_T} K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \text{Cov}(Z_{m,t,T}, Z_{m',t,T}) \\
&\leq \frac{1}{(Th\phi(h))^2} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'| > s_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) |\text{Cov}(Z_{m,t,T}, Z_{m',t,T})|.
\end{aligned}$$

Now, using (25), we have

$$\begin{aligned}
&K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) |\text{Cov}(Z_{m,t,T}, Z_{m',t,T})| \\
&\lesssim K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \beta(|m - m'|)^{1-\frac{2}{p}}.
\end{aligned}$$

Thus

$$\begin{aligned}\Sigma_{31}^\Lambda &\lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'| > s_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \beta(|m-m'|)^{1-\frac{2}{p}} \\ &\leq \frac{C_1^2}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'| > s_T}}^T \beta(|m-m'|)^{1-\frac{2}{p}}.\end{aligned}$$

By Assumption 6, $\sum_{k=1}^\infty k^\zeta \beta(k)^{1-\frac{2}{p}} < \infty$. Now, observe that letting $k = |m-m'|$ yields

$$\begin{aligned}\sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'| > s_T}}^T \beta(|m-m'|)^{1-\frac{2}{p}} &\leq C \sum_{k=s_T+1}^T \beta(k)^{1-\frac{2}{p}} \lesssim \frac{1}{k^\zeta} \sum_{k=s_T+1}^T k^\zeta \beta(k)^{1-\frac{2}{p}} \\ &\leq \frac{1}{s_T^\zeta} \sum_{k=s_T+1}^T k^\zeta \beta(k)^{1-\frac{2}{p}}, \quad \text{since } k > s_T, \\ &\leq \frac{1}{s_T^\zeta} \sum_{k=s_T+1}^\infty k^\zeta \beta(k)^{1-\frac{2}{p}},\end{aligned}$$

since $\beta(k) \geq 0$ and $\left(\frac{k}{s_T}\right)^\zeta \geq 1$ for $\zeta > 1 - \frac{2}{p}$, where $p > 2$. So

$$\begin{aligned}\Sigma_{31}^\Lambda &\leq \frac{C_1^2}{s_T^\zeta T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{k=s_T+1}^\infty k^\zeta \beta(k)^{1-\frac{2}{p}} \\ &\lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}}, \quad \text{since } \frac{1}{s_T^\zeta} \leq 1, \\ &\lesssim \left(\frac{1}{T^{2p} h^{2p} \phi^{2p}(h)} \left(\frac{1}{T} + h\phi(h)\right)^2\right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^{2p} h^{2p} \phi^{2p}(h)} \left(\frac{1}{T^2} + h^2 \phi^2(h)\right)\right)^{\frac{1}{p}} \\ &\lesssim \left(\frac{1}{T^{2p+2} h^{2p} \phi^{2p}(h)} + \frac{1}{T^{2p} h^{2p-2} \phi^{2p-2}(h)}\right)^{\frac{1}{p}} \\ &\lesssim \left(\frac{1}{T^{2p} h^{2p} \phi^{2p}(h)}\right)^{\frac{1}{p}} \lesssim \frac{1}{T^2 h^2 \phi^2(h)}.\end{aligned}\tag{28}$$

Step 1.3.2 Control of Σ_{32}^Λ . In view of Σ_{32}^Λ , observe that

$$\begin{aligned}\Sigma_{32}^\Lambda &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \\ &\quad \times \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\ &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{r_T} \sum_{n_2=1}^{r_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda'+n_2}{T}\right)\end{aligned}$$

$$\times \mathbb{E}[Z_{\lambda+n_1,t,T}] \mathbb{E}[Z_{\lambda'+n_2,t,T}].$$

Similarly, for $l \neq l'$, $|\lambda - \lambda' + n_1 - n_2| > s_T$, then

$$\begin{aligned} \Sigma_{32}^\Lambda &\leq \frac{1}{(Th\phi(h))^2} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>s_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \mathbb{E}[Z_{m,t,T}] \mathbb{E}[Z_{m',t,T}] \\ &= \frac{1}{(Th\phi(h))^2} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>s_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{m,T}))(\mathbf{1}_{Y_{m,T} \leq y} - F_t^*(y|x))\right] \\ &\quad \times \mathbb{E}\left[K_{h,2}(D(x, X_{m',T}))(\mathbf{1}_{Y_{m',T} \leq y} - F_t^*(y|x))\right]. \end{aligned}$$

Using Proposition 3.iii, $K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{m,T}))(\mathbf{1}_{Y_{m,T} \leq y} - F_t^*(y|x))\right] \lesssim K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) \left(\frac{1}{T} + h\phi(h)\right)$, then

$$\begin{aligned} \Sigma_{32}^\Lambda &\lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h)\right)^2 \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>s_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \\ &\leq \frac{1}{\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^2 \underbrace{\frac{1}{Th} \sum_{m=1}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right)}_{\mathcal{O}(1)} \underbrace{\frac{1}{Th} \sum_{m'=1}^T K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right)}_{\mathcal{O}(1)} \\ &\lesssim \frac{1}{\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^2 \lesssim \frac{1}{\phi^2(h)} \left(\frac{1}{T^2} + h^2\phi^2(h)\right) \\ &\lesssim \frac{1}{T^2\phi^2(h)} + h^2, \end{aligned} \tag{29}$$

which goes to zero as $T \rightarrow \infty$ using Assumption 4. Hence, comparing (24), (26), (27), (28), and (29), we have

$$\mathbb{E}[\Lambda_{t,T}^2] \lesssim \frac{1}{Th^2\phi^2(h)} + h^2. \tag{30}$$

Step 2. Control of the small blocks. Next, we deal with the small blocks. See that

$$\begin{aligned} \mathbb{E}[\Pi_{t,T}^2] &= \mathbb{E}\left[\sum_{l=0}^{v_T-1} \Pi_{l,t,T}^2 + \sum_{\substack{l=0 \\ l \neq l'}}^{v_T-1} \sum_{l'=0}^{v_T-1} \Pi_{l,t,T} \Pi_{l',t,T}\right] \\ &= \mathbb{E}\left[\frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \left(\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) Z_{a,t,T}\right)^2\right] \\ &\quad + \frac{1}{(Th\phi(h))^2} \sum_{\substack{l=0 \\ l \neq l'}}^{v_T-1} \sum_{l'=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} \sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \end{aligned}$$

$$\times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right)\mathbb{E}[Z_{a,t,T}Z_{b,t,T}].$$

Observe that

$$\begin{aligned}\mathbb{E}[\Pi_{t,T}^2] &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right)\mathbb{E}[Z_{a,t,T}^2] \\ &\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} \sum_{\substack{b=l(r_T+s_T)+r_T+1 \\ a \neq b}}^{(l+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\ &\quad \quad \quad \times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right)\mathbb{E}[Z_{a,t,T}Z_{b,t,T}] \\ &\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} \sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\ &\quad \quad \quad \times K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right)\mathbb{E}[Z_{a,t,T}Z_{b,t,T}] \\ &=: \Sigma_1^\Pi + \Sigma_2^\Pi + \Sigma_3^\Pi.\end{aligned}$$

Step 2.1. Control of Σ_1^Π First, let us consider Σ_1^Π .

$$\begin{aligned}\Sigma_1^\Pi &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right)\mathbb{E}[Z_{a,t,T}^2] \\ &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right)\mathbb{E}\left[K_{h,2}^2(D(x, X_{a,T}))(\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x))^2\right] \\ &\leq \frac{2C_2}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right)\mathbb{E}\left[K_{h,2}(D(x, X_{a,T}))|\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)|\right].\end{aligned}$$

By Proposition 3.iii, we get

$$\begin{aligned}\Sigma_1^\Pi &\lesssim \frac{1}{T^2 h \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right) \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\left(\frac{t}{T} - \frac{a}{T}\right) \\ &\leq \frac{C_1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right) \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) \\ &\leq \frac{C_1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right) \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right)}_{\mathcal{O}(1)} \\ &\lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)\end{aligned}$$

$$\begin{aligned}
&\lesssim \frac{1}{T^2 h \phi^2(h)} + \frac{1}{T \phi(h)} \\
&\lesssim \frac{1}{T h \phi^2(h)}.
\end{aligned} \tag{31}$$

Step 2.2. Control of Σ_2^Π . On the other hand,

$$\begin{aligned}
\Sigma_2^\Pi &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{a=l(r_T+s_T)+r_T+1 \\ a \neq b}}^{(l+1)(r_T+s_T)} \sum_{b=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\quad \times \{ \text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}) + \mathbb{E}[Z_{\lambda+n_1,t,T}] \mathbb{E}[Z_{\lambda+n_2,t,T}] \},
\end{aligned}$$

where $\lambda = l(r_T + s_T) + r_T$. So

$$\begin{aligned}
\Sigma_2^\Pi &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}) \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\quad \times \mathbb{E}[Z_{\lambda+n_1,t,T}] \mathbb{E}[Z_{\lambda+n_2,t,T}] \\
&=: \Sigma_{21}^\Pi + \Sigma_{22}^\Pi.
\end{aligned}$$

Step 2.2.1. Control of Σ_{21}^Π . Taking Σ_{21}^Π into consideration, we have

$$\Sigma_{21}^\Pi = \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}).$$

Using (25),

$$\begin{aligned}
&K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) |\text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T})| \\
&\lesssim K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \beta(|n_1 - n_2|)^{1-\frac{2}{p}}.
\end{aligned}$$

Thus

$$\Sigma_{21}^\Pi \lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \beta(|n_1 - n_2|)^{1-\frac{2}{p}}$$

$$\leq \frac{C_1^2}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} \beta(|n_1 - n_2|)^{1-\frac{2}{p}}.$$

Using Assumption 6, $\sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} < \infty$, which can be expressed as $\sum_{k=1}^{s_T} k^\zeta \beta(k)^{1-\frac{2}{p}} + \sum_{k=s_T+1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}}$. In addition, letting $k = |n_1 - n_2|$ yields

$$\begin{aligned} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} \beta(|n_1 - n_2|)^{1-\frac{2}{p}} &= \sum_{n_1=1}^{s_T} \left(\sum_{n_2>n_1}^{s_T} \beta(n_2 - n_1)^{1-\frac{2}{p}} + \sum_{n_2<n_1}^{s_T} \beta(n_1 - n_2)^{1-\frac{2}{p}} \right) \\ &= \sum_{n_1=1}^{s_T} \sum_{k>0}^{s_T-n_1} \beta(k)^{1-\frac{2}{p}} + \sum_{n_2=1}^{s_T} \sum_{k>0}^{s_T-n_2} \beta(k)^{1-\frac{2}{p}} \\ &= 2 \sum_{n=1}^{s_T} \sum_{k>0}^{s_T-n} \beta(k)^{1-\frac{2}{p}} \leq 2s_T \sum_{k=1}^{s_T} \beta(k)^{1-\frac{2}{p}} \\ &\lesssim s_T \sum_{k=1}^{s_T} k^\zeta \beta(k)^{1-\frac{2}{p}} \leq s_T \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}}, \end{aligned}$$

since $\beta(k) \geq 0$ and $k^\zeta \geq 1$ for $\zeta > 1 - \frac{2}{p}$, where $p > 2$. So

$$\begin{aligned} \Sigma_{21}^{\Pi} &\leq \frac{C_1^2 s_T}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{l=0}^{v_T-1} \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\ &\lesssim \frac{v_T s_T}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\ &\lesssim \frac{1}{T h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}}, \quad \text{since } v_T s_T \leq \frac{T}{s_T} s_T = T, \\ &= \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \left(\frac{1}{T^2} + h^2 \phi^2(h) \right) \right)^{\frac{1}{p}} \\ &\lesssim \left(\frac{1}{T^{2+p} h^{2p} \phi^{2p}(h)} + \frac{1}{T^p h^{2p-2} \phi^{2p-2}(h)} \right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \right)^{\frac{1}{p}} \lesssim \frac{1}{T h^2 \phi^2(h)}. \quad (32) \end{aligned}$$

Step 2.2.2. Control of Σ_{22}^{Π} . Next, looking at Σ_{22}^{Π} , we have

$$\begin{aligned} \Sigma_{22}^{\Pi} &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{a=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} \sum_{b=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\ &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\ &\quad \times \mathbb{E}\left[K_{h,2}(D(x, X_{\lambda+n_1, T})) (\mathbb{1}_{Y_{\lambda+n_1, T} \leq y} - F_t^*(y|x)) \right] \end{aligned}$$

$$\times \mathbb{E} \left[K_{h,2}(\mathbf{D}(x, X_{\lambda+n_2, T})) (\mathbb{1}_{Y_{\lambda+n_2, T} \leq y} - F_t^*(y|x)) \right].$$

Using Proposition 3.iii, for $i = 1, 2$, $K_{h,1}(\frac{t}{T} - \frac{\lambda+n_i}{T}) \mathbb{E} [K_{h,2}(\mathbf{D}(x, X_{\lambda+n_i, T})) (\mathbb{1}_{Y_{\lambda+n_i, T} \leq y} - F_t^*(y|x))] \lesssim K_{h,1}(\frac{t}{T} - \frac{\lambda+n_i}{T}) (\frac{1}{T} + h\phi(h))$, then

$$\begin{aligned} \Sigma_{22}^{\Pi} &\lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h) \right)^2 \sum_{l=0}^{v_T-1} \sum_{\substack{n_1=1 \\ |n_1-n_2|>0}}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_1}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_2}{T} \right) \\ &\leq \frac{C_1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right)}_{\mathcal{O}(1)} \\ &\lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \\ &\lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T^2} + h^2\phi^2(h) \right) \\ &\lesssim \frac{1}{T^3 h\phi^2(h)} + \frac{h}{T} \\ &\lesssim \frac{1}{Th\phi^2(h)}. \end{aligned} \tag{33}$$

Step 2.3. Control of Σ_3^{Π} . Now, let us deal with Σ_3^{Π} .

$$\begin{aligned} \Sigma_3^{\Pi} &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} \sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{b}{T} \right) \\ &\quad \times \mathbf{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\ &+ \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} \sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{b}{T} \right) \\ &\quad \times \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\ &= \Sigma_{31}^{\Pi} + \Sigma_{32}^{\Pi}. \end{aligned}$$

Step 2.3.1 Control of Σ_{31}^{Π} . Looking at Σ_{31}^{Π} , see that

$$\begin{aligned} \Sigma_{31}^{\Pi} &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_1}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{\lambda'+n_2}{T} \right) \\ &\quad \times \mathbf{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda'+n_2,t,T}), \end{aligned}$$

where $\lambda = l(r_T + s_T) + r_T$ and $\lambda' = l'(r_T + s_T) + r_T$, however, for $l \neq l'$,

$$|\lambda - \lambda' + n_1 - n_2| \geq |l(r_T + s_T) + r_T - l'(r_T + s_T) - r_T + n_1 - n_2|$$

$$\begin{aligned} &\geq |(l - l')(r_T + s_T) + n_1 - n_2| \\ &> r_T, \end{aligned}$$

since $n_1, n_2 \in \{1, \dots, s_T\}$. So if we let $q = \lambda + n_1$ and $q' = \lambda' + n_2$, we have

$$\begin{aligned} \Sigma_{31}^{\Pi} &= \frac{1}{(Th\phi(h))^2} \sum_{q=r_T+1}^{v_T(r_T+s_T)} \sum_{\substack{q'=r_T+1 \\ |q-q'|>r_T}}^{v_T(r_T+s_T)} K_{h,1}\left(\frac{t}{T} - \frac{q}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{q'}{T}\right) \text{Cov}(Z_{q,t,T}, Z_{q',t,T}) \\ &= \frac{1}{(Th\phi(h))^2} \sum_{m=1}^{v_T(r_T+s_T)-r_T} \sum_{\substack{m'=1 \\ |m-m'|>r_T}}^{v_T(r_T+s_T)-r_T} K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \\ &\quad \times \text{Cov}(Z_{m,t,T}, Z_{m',t,T}) \\ &\leq \frac{1}{(Th\phi(h))^2} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>r_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) |\text{Cov}(Z_{m,t,T}, Z_{m',t,T})|, \end{aligned}$$

where $m = q - r_T$ and $m' = q' - r_T$. Now, using (25), we have

$$\begin{aligned} &K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) |\text{Cov}(Z_{m,t,T}, Z_{m',t,T})| \\ &\lesssim K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \beta(|m - m'|)^{1-\frac{2}{p}}. \end{aligned}$$

Thus

$$\begin{aligned} \Sigma_{31}^{\Pi} &\lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>r_T}}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right) \beta(|m - m'|)^{1-\frac{2}{p}} \\ &\leq \frac{C_1^2}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>r_T}}^T \beta(|m - m'|)^{1-\frac{2}{p}}. \end{aligned}$$

By Assumption 6, $\sum_{k=1}^{\infty} k^{\zeta} \beta(k)^{1-\frac{2}{p}} < \infty$, which can be expressed as $\sum_{k=1}^{r_T} k^{\zeta} \beta(k)^{1-\frac{2}{p}} + \sum_{k=r_T+1}^{\infty} k^{\zeta} \beta(k)^{1-\frac{2}{p}}$. Additionally, observe that letting $k = |m - m'|$ yields

$$\begin{aligned} \sum_{m=1}^T \sum_{\substack{m'=1 \\ |m-m'|>r_T}}^T \beta(|m - m'|)^{1-\frac{2}{p}} &\leq C \sum_{k=r_T+1}^T \beta(k)^{1-\frac{2}{p}} \\ &\lesssim \frac{1}{k^{\zeta}} \sum_{k=r_T+1}^T k^{\zeta} \beta(k)^{1-\frac{2}{p}} \\ &\leq \frac{1}{r_T^{\zeta}} \sum_{k=r_T+1}^T k^{\zeta} \beta(k)^{1-\frac{2}{p}}, \quad \text{since } k > r_T, \end{aligned}$$

$$\leq \frac{1}{r_T^\zeta} \sum_{k=r_T+1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}},$$

since $\beta(k) \geq 0$ and $\left(\frac{k}{r_T}\right)^\zeta \geq 1$ for $\zeta > 1 - \frac{2}{p}$, where $p > 2$. So

$$\begin{aligned} \Sigma_{31}^\Pi &\lesssim \frac{1}{r_T^\zeta T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}} \sum_{k=r_T+1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\ &\lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^{\frac{2}{p}}, \text{ since } \frac{1}{r_T^\zeta} \leq 1, \\ &\lesssim \left(\frac{1}{T^{2p} h^{2p} \phi^{2p}(h)} \left(\frac{1}{T} + h\phi(h)\right)^2\right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^{2p} h^{2p} \phi^{2p}(h)} \left(\frac{1}{T^2} + h^2 \phi^2(h)\right)\right)^{\frac{1}{p}} \\ &\lesssim \left(\frac{1}{T^{2p+2} h^{2p} \phi^{2p}(h)} + \frac{1}{T^{2p} h^{2p-2} \phi^{2p-2}(h)}\right)^{\frac{1}{p}} \lesssim \left(\frac{1}{T^{2p} h^{2p} \phi^{2p}(h)}\right)^{\frac{1}{p}} \\ &\lesssim \frac{1}{T^2 h^2 \phi^2(h)}. \end{aligned} \tag{34}$$

Step 2.3.2 Control of Σ_{32}^Π . In dealing with Σ_{32}^Π , observe that

$$\begin{aligned} \Sigma_{32}^\Pi &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda' + n_2}{T}\right) \\ &\quad \times \mathbb{E}[Z_{\lambda+n_1, t, T}] \mathbb{E}[Z_{\lambda'+n_2, t, T}] \\ &= \frac{1}{(Th\phi(h))^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda' + n_2}{T}\right) \\ &\quad \times \mathbb{E}\left[K_{h,2}(D(x, X_{\lambda+n_1, T}))(\mathbf{1}_{Y_{\lambda+n_1, T} \leq y} - F_t^*(y|x))\right] \\ &\quad \times \mathbb{E}\left[K_{h,2}(D(x, X_{\lambda'+n_2, T}))(\mathbf{1}_{Y_{\lambda'+n_2, T} \leq y} - F_t^*(y|x))\right]. \end{aligned}$$

Using Proposition 3.iii, $K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) \mathbb{E}\left[K_{h,2}(D(x, X_{\lambda+n_1, T}))(\mathbf{1}_{Y_{\lambda+n_1, T} \leq y} - F_t^*(y|x))\right] \lesssim K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) \left(\frac{1}{T} + h\phi(h)\right)$, then

$$\Sigma_{32}^\Pi \lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h)\right)^2 \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda + n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda' + n_2}{T}\right).$$

Similarly, for $l \neq l'$, $|\lambda - \lambda' + n_1 - n_2| > r_T$, then

$$\Sigma_{32}^\Pi \lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h)\right)^2 \sum_{\substack{m=1 \\ |m-m'| > r_T}}^T \sum_{m'=1}^T K_{h,1}\left(\frac{t}{T} - \frac{m}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{m'}{T}\right)$$

$$\begin{aligned}
&\leq \frac{1}{\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \underbrace{\frac{1}{Th} \sum_{m=1}^T K_{h,1} \left(\frac{t}{T} - \frac{m}{T} \right)}_{\mathcal{O}(1)} \underbrace{\frac{1}{Th} \sum_{m'=1}^T K_{h,1} \left(\frac{t}{T} - \frac{m'}{T} \right)}_{\mathcal{O}(1)} \\
&\lesssim \frac{1}{\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \lesssim \frac{1}{\phi^2(h)} \left(\frac{1}{T^2} + h^2\phi^2(h) \right) \lesssim \frac{1}{T^2\phi^2(h)} + h^2, \tag{35}
\end{aligned}$$

which goes to zero as $T \rightarrow \infty$ using Assumption 4. Now, comparing (31), (32), (33), (34), and (35), we get

$$\mathbb{E}[\Pi_{t,T}^2] \lesssim \frac{1}{Th^2\phi^2(h)} + h^2. \tag{36}$$

Step 3. Control of the remainder block. Now, let us deal with $\mathbb{E}[\Xi_{t,T}^2]$. See that

$$\begin{aligned}
\mathbb{E}[\Xi_{t,T}^2] &= \mathbb{E} \left[\left(\frac{1}{Th\phi(h)} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) Z_{a,t,T} \right)^2 \right] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}^2 \left(\frac{t}{T} - \frac{a}{T} \right) \mathbb{E}[Z_{a,t,T}^2] \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{b}{T} \right) \mathbb{E}[Z_{a,t,T} Z_{b,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}^2 \left(\frac{t}{T} - \frac{a}{T} \right) \mathbb{E}[Z_{a,t,T}^2] \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{b}{T} \right) \text{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\
&\quad + \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{b}{T} \right) \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\
&=: \Sigma_1^{\Xi} + \Sigma_2^{\Xi} + \Sigma_3^{\Xi}.
\end{aligned}$$

Step 3.1. Control of Σ_1^{Ξ} . Considering Σ_1^{Ξ} , we have

$$\begin{aligned}
\Sigma_1^{\Xi} &= \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}^2 \left(\frac{t}{T} - \frac{a}{T} \right) \mathbb{E}[Z_{a,t,T}^2] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}^2 \left(\frac{t}{T} - \frac{a}{T} \right) \mathbb{E} \left[K_{h,2}^2(\mathbb{D}(x, X_{a,T})) (\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x))^2 \right] \\
&\leq \frac{2C_2}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}^2 \left(\frac{t}{T} - \frac{a}{T} \right) \mathbb{E} \left[K_{h,2}(\mathbb{D}(x, X_{a,T})) |\mathbb{1}_{Y_{a,T} \leq y} - F_t^*(y|x)| \right].
\end{aligned}$$

Using Proposition 3.iii, we have

$$\begin{aligned}\Sigma_1^{\Xi} &\lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right) \sum_{a=v_T(r_T+s_T)+1}^T K_{h,1}^2 \left(\frac{t}{T} - \frac{a}{T} \right) \\ &\leq \frac{C_1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right) \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right)}_{\mathcal{O}(1)},\end{aligned}$$

using 20. So

$$\Sigma_1^{\Xi} \lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right) \lesssim \frac{1}{T^2 h\phi^2(h)} + \frac{1}{T\phi(h)} \lesssim \frac{1}{Th\phi^2(h)}. \quad (37)$$

Step 3.2. Control of Σ_2^{Ξ} . Taking Σ_2^{Ξ} into account, we have

$$\begin{aligned}\Sigma_2^{\Xi} &= \frac{1}{(Th\phi(h))^2} \sum_{a=v_T(r_T+s_T)+1}^T \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^T K_{h,1} \left(\frac{t}{T} - \frac{a}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{b}{T} \right) \text{Cov}(Z_{a,t,T}, Z_{b,t,T}) \\ &= \frac{1}{(Th\phi(h))^2} \sum_{n_1=1}^{T-v_T(r_T+s_T)} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{T-v_T(r_T+s_T)} K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_1}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_2}{T} \right) \text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}),\end{aligned}$$

where $\lambda = v_T(r_T + s_T)$. Now, using (25), we have

$$\begin{aligned}&K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_1}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_2}{T} \right) |\text{Cov}(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T})| \\ &\lesssim K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_1}{T} \right) K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_2}{T} \right) \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \beta(|n_1 - n_2|)^{1-\frac{2}{p}}.\end{aligned}$$

Thus

$$\begin{aligned}\Sigma_2^{\Xi} &\lesssim \frac{1}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{n_1=1}^{T-v_T(r_T+s_T)} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{T-v_T(r_T+s_T)} K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_1}{T} \right) \\ &\quad \times K_{h,1} \left(\frac{t}{T} - \frac{\lambda+n_2}{T} \right) \beta(|n_1 - n_2|)^{1-\frac{2}{p}} \\ &\leq \frac{C_1^2}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{n_1=1}^{T-v_T(r_T+s_T)} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{T-v_T(r_T+s_T)} \beta(|n_1 - n_2|)^{1-\frac{2}{p}}.\end{aligned}$$

Assumption 6 entails $\sum_{k=1}^{\infty} k^{\zeta} \beta(k)^{1-\frac{2}{p}} < \infty$. Moreover, letting $k = |n_1 - n_2|$ and $w_T = T - v_T(r_T + s_T)$ yields

$$\sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} \beta(|n_1 - n_2|)^{1-\frac{2}{p}} = \sum_{n_1=1}^{w_T} \left(\sum_{n_2>n_1}^{w_T} \beta(n_2 - n_1)^{1-\frac{2}{p}} + \sum_{n_2<n_1}^{w_T} \beta(n_1 - n_2)^{1-\frac{2}{p}} \right)$$

$$\begin{aligned}
&= \sum_{n_1=1}^{w_T} \sum_{k>0}^{w_T-n_1} \beta(k)^{1-\frac{2}{p}} + \sum_{n_2=1}^{w_T} \sum_{k>0}^{w_T-n_2} \beta(k)^{1-\frac{2}{p}} \\
&= 2 \sum_{n=1}^{w_T} \sum_{k>0}^{w_T-n} \beta(k)^{1-\frac{2}{p}} \leq 2w_T \sum_{k=1}^{w_T} \beta(k)^{1-\frac{2}{p}} \\
&\lesssim w_T \sum_{k=1}^{w_T} k^\zeta \beta(k)^{1-\frac{2}{p}} \leq w_T \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}},
\end{aligned}$$

since $\beta(k) \geq 0$ and $k^\zeta \geq 1$ for $\zeta > 1 - \frac{2}{p}$, where $p > 2$. So

$$\begin{aligned}
\Sigma_2^{\Xi} &\leq \frac{C_1^2 w_T}{T^2 h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}} \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\
&\lesssim \frac{1}{T h^2 \phi^2(h)} \left(\frac{1}{T} + h\phi(h) \right)^{\frac{2}{p}}, \quad \text{since } w_T \ll T, \\
&= \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \left(\frac{1}{T} + h\phi(h) \right)^2 \right)^{\frac{1}{p}} \\
&\lesssim \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \left(\frac{1}{T^2} + h^2 \phi^2(h) \right) \right)^{\frac{1}{p}} \\
&\lesssim \left(\frac{1}{T^{2+p} h^{2p} \phi^{2p}(h)} + \frac{1}{T^p h^{2p-2} \phi^{2p-2}(h)} \right)^{\frac{1}{p}} \\
&\lesssim \left(\frac{1}{T^p h^{2p} \phi^{2p}(h)} \right)^{\frac{1}{p}} \\
&\lesssim \frac{1}{T h^2 \phi^2(h)}. \tag{38}
\end{aligned}$$

Step 3.3. Control of Σ_3^{Ξ} . Lastly, let us look at Σ_3^{Ξ} .

$$\begin{aligned}
\Sigma_3^{\Xi} &= \frac{1}{(Th\phi(h))^2} \sum_{\substack{a=v_T(r_T+s_T)+1 \\ a \neq b}}^T \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{b}{T}\right) \mathbb{E}[Z_{a,t,T}] \mathbb{E}[Z_{b,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \mathbb{E}[Z_{\lambda+n_1,t,T}] \mathbb{E}[Z_{\lambda+n_2,t,T}] \\
&= \frac{1}{(Th\phi(h))^2} \sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\
&\quad \times \mathbb{E} \left[\prod_{j=1}^d K_{h,2}(x^j - X_{\lambda+n_1,T}^j) (\mathbb{1}_{Y_{\lambda+n_1,T} \leq y} - F_t^*(y|x)) \right] \\
&\quad \times \mathbb{E} \left[\prod_{j=1}^d K_{h,2}(x^j - X_{\lambda+n_2,T}^j) (\mathbb{1}_{Y_{\lambda+n_2,T} \leq y} - F_t^*(y|x)) \right].
\end{aligned}$$

Using Proposition 3.iii, for $i = 1, 2$, $K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_i}{T}\right)\mathbb{E}[K_{h,2}(D(x, X_{\lambda+n_i,T}))(\mathbb{1}_{Y_{\lambda+n_i,T} \leq y} - F_t^*(y|x))]$ $\lesssim K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_i}{T}\right)\left(\frac{1}{T} + h\phi(h)\right)$, then

$$\begin{aligned} \Sigma_3^{\Xi} &\lesssim \frac{1}{(Th\phi(h))^2} \left(\frac{1}{T} + h\phi(h)\right)^2 \sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_1}{T}\right) K_{h,1}\left(\frac{t}{T} - \frac{\lambda+n_2}{T}\right) \\ &\leq \frac{C_1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^2 \underbrace{\frac{1}{Th} \sum_{a=1}^T K_{h,1}\left(\frac{t}{T} - \frac{a}{T}\right)}_{\mathcal{O}(1)}, \end{aligned}$$

using 20. So

$$\begin{aligned} \Sigma_3^{\Xi} &\lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T} + h\phi(h)\right)^2 \lesssim \frac{1}{Th\phi^2(h)} \left(\frac{1}{T^2} + h^2\phi^2(h)\right) \\ &\lesssim \frac{1}{T^3h\phi^2(h)} + \frac{h}{T} \lesssim \frac{1}{Th\phi^2(h)}. \end{aligned} \quad (39)$$

Now, comparing (37), (38), and (39), we have

$$\mathbb{E}[\Xi_{t,T}^2] \lesssim \frac{1}{Th^2\phi^2(h)}. \quad (40)$$

Therefore, following (30), (36), and (40), we get

$$\mathbb{E}[Z_{t,T}^2] = \mathcal{O}\left(\frac{1}{Th^2\phi^2(h)} + h^2\right). \quad (41)$$

References

- [1] B. M. Agua and S. Bouzebda. Single index regression for locally stationary functional time series. *AIMS Mathematics*, 9(12):36202–36258, 2024. (Cited on pages: 10, 14, and 15.)
- [2] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet-kernel approach for time series prediction. *J. Roy. Statist. Soc. Ser. B*, 86:837–857, 2006. (Cited on page: 2.)
- [3] A. Aue, D. D. Nourinho, and S. Hörmann. On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.*, 110:378–392, 2015. (Cited on page: 1.)
- [4] A. Aue and A. van Delft. Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.*, 48:2505–2547, 2020. (Cited on page: 1.)
- [5] K. Benhenni, F. Ferraty, M. Rachdi, and P. Vieu. Local smoothing regression with functional data. *Comput. Statist.*, 22(3):353–369, 2007. (Cited on pages: 12, 13.)
- [6] S. N. Bernstein. Sur l’extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, 97:1–59, 1927. (Cited on page: 11.)

- [7] F. Björgevinnsson. 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. (Cited on page: 17.)
- [8] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and kantovich transport distances. *Mem. Amer. Math. Soc.*, 261, 2019. (Cited on page: 11.)
- [9] V. I. Bogachev. *Gaussian measures, volume 62 of Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998. (Cited on pages: 7, 10.)
- [10] E.G. Bongiorno, A. Goia, and P. Vieu. Evaluating the complexity of some families of functional data. *SORT*, 42(1):27–44, 2018. (Cited on page: 6.)
- [11] D. Bosq. *Linear processes in function spaces*. Springer, New York, USA, 2000. (Cited on page: 2.)
- [12] D. Bosq. Estimation of mean and covariance operator of autoregressive processes in banach spaces. *Stat. Inference Stoch. Process*, 5:287–306, 2002. (Cited on page: 2.)
- [13] D. Bosq. *Nonparametric statistics for stochastic processes: estimation and prediction*. Springer Science and Business Media, 2012. (Cited on page: 8.)
- [14] O. Bouanani and S. Bouzebda. Limit theorems for local polynomial estimation of regression for functional dependent data. *AIMS Math.*, 9(9):23651–23691, 2024. (Cited on pages: 2, 3, 9, and 27.)
- [15] 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. (Cited on page: 10.)
- [16] S. Bouzebda. Limit Theorems in the Nonparametric Conditional Single-Index U-Processes for Locally Stationary Functional Random Fields under Stochastic Sampling Design. *Mathematics*, 12(13):1–81, 2024. (Cited on pages: 1, 2, and 10.)
- [17] S. Bouzebda. Weak convergence of the conditional single index U -statistics for locally stationary functional time series. *AIMS Math.*, 9(6):14807–14898, 2024. (Cited on pages: 2, 4, 6, and 9.)
- [18] S. Bouzebda and M. Chaouch. Uniform limit theorems for a class of conditional Z -estimators when covariates are functions. *J. Multivariate Anal.*, 189:Paper No. 104872, 21, 2022. (Cited on page: 1.)
- [19] S. Bouzebda and B. Nemouchi. Weak-convergence of empirical conditional processes and conditional U -processes involving functional mixing data. *Stat. Inference Stoch. Process.*, 26(1):33–88, 2023. (Cited on page: 2.)
- [20] 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. (Cited on pages: 2, 12.)

- [21] S. Bouzebda and A. Nezzal. Uniform in number of neighbors consistency and weak convergence of k NN empirical conditional processes and k NN conditional U -processes involving functional mixing data. *AIMS Math.*, 9(2):4427–4550, 2024. (Cited on page: 2.)
- [22] R. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Survey*, 2:107–144, 2005. (Cited on page: 11.)
- [23] 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. (Cited on page: 1.)
- [24] M. Carrasco and X Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002. (Cited on page: 8.)
- [25] 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. (Cited on page: 1.)
- [26] R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37, 1996. (Cited on pages: 2, 3.)
- [27] R. Dahlhaus. A likelihood approximation for locally stationary processes. *Ann. Statist.*, 25:1762–1794, 2000. (Cited on page: 2.)
- [28] 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. (Cited on page: 2.)
- [29] R. Dahlhaus and S. Subba Rao. Statistical inference for time-varying ARCH processes. *Ann. Statist.*, 34:1075–1114, 2006. (Cited on page: 2.)
- [30] J.A. Davydov. Mixing conditions for markov chains. *Teor. Veroyatnost. i Primenen.*, pages 321–338, 1973. (Cited on page: 30.)
- [31] 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. (Cited on page: 8.)
- [32] J. Dedecker and F. Merlevede. Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α -dependent sequences. *Bernoulli*, 23(3):2083–2127, 2017. (Cited on page: 6.)
- [33] J. Dedecker and C. Prieur. New dependence coefficients: examples and applications to statistics. *Probab. Theory Relat. Fields*, 132:203–236, 2005. (Cited on page: 8.)
- [34] H. Dehling and O.S. Sharipov. Estimation of mean and covariance operator for banach space valued autoregressive processes with independent innovations. *Stat. Inference Stoch. Process*, 8:37–149, 2005. (Cited on page: 2.)
- [35] C. Dombry, T. Modeste, and R. Pic. Stone’s theorem for distributional regression in Wasserstein distance. *J. Nonparametr. Stat.*, 0(0):1–23, 2024. (Cited on page: 6.)

- [36] P. Doukhan. *Mixing: properties and examples*. Springer-Verlag New York, 1994. (Cited on page: 8.)
- [37] F. Ferraty, A. Laksaci, and P. Vieu. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat Inference Stoch Process*, 9(1):47–76, 2006. (Cited on page: 6.)
- [38] F. Ferraty, A. Mas, and P. Vieu. Nonparametric regression on functional data: inference and practical aspect. *Aust N Z J Stat*, 49(3):267–286, 2007. (Cited on page: 6.)
- [39] 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. (Cited on pages: 2, 10, 16, and 17.)
- [40] 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. (Cited on page: 6.)
- [41] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis*. Springer, 233 Spring Street, New York, NY 10013, USA, 2006. (Cited on pages: 1, 8, 10, and 27.)
- [42] P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.*, 36(2):742–786, 2008. (Cited on page: 2.)
- [43] 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. (Cited on page: 10.)
- [44] 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. (Cited on pages: 9, 10.)
- [45] S. Hörmann, T. Kuenzer, and G. Rice. Estimating the conditional distribution in functional regression problems. *Electronic Journal of Statistics*, 16:5751–5778, 2022. (Cited on page: 2.)
- [46] K. A. Kilpatrick, G. Podestá, S. Walsh, E. Williams, V. Halliwell, M. Szczodrak, O.B. Brown, P.J. Minnett, and R. Evans. A decade of sea surface temperature from modis. *Remote Sensing of Environment*, 165:27–41, 2015. (Cited on page: 16.)
- [47] B. Koo and O. Linton. Estimation of semiparametric locally stationary diffusion models. *Journal of Econometrics*, 170:210–233, 2012. (Cited on pages: 2, 3.)
- [48] D. Kurisu. Nonparametric regression for locally stationary functional time series. *Electronic Journal of Statistics*, 16:3973–3995, 2022. (Cited on pages: 2, 4, 9, 10, 12, and 25.)
- [49] W. V. Li and Q.M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *In Stochastic processes: theory and methods*, volume 19 of *Handbook of Statistics*, pages 533–597. North-Holland, Amsterdam, 2001. (Cited on pages: 6, 10.)
- [50] T. Manole, S. Balakrishnan, and L. Wasserman. Minimax confidence intervals for the sliced Wasserstein distance. *Electron. J. Stat.*, 16(1):2252–2345, 2022. (Cited on page: 3.)

- [51] E. Masry. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications*, 115:155–177, 2005. (Cited on page: 27.)
- [52] E. Mayer-Wolf and O. Zeitouni. The probability of small gaussian ellipsoids and associated conditional moments. *Ann. Probab.*, 21(1):14–24, 1993. (Cited on page: 10.)
- [53] I. Metmous, M. K. Attouch, B. Mechab, and T. Merouan. Nonparametric estimation of the conditional distribution function for surrogate data by the regression model. *Applications and Applied Mathematics: An International Journal (AAM)*, 16(1):4, 2021. (Cited on pages: 9, 10, and 11.)
- [54] A. Mokkadem. Mixing properties of ARMA processes. *Stochastic processes and their applications*, 29:309–315, 1998. (Cited on page: 8.)
- [55] È. A. Nadaraya. On a regression estimate. *Teor. Veroyatnost. i Primenen.*, 9:157–159, 1964. (Cited on page: 2.)
- [56] S. Nietert, Z. Goldfeld, and K. Kato. Smooth p -Wasserstein Distance: structure, empirical approximation, and statistical applications. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139. PMLR, 2021. (Cited on page: 17.)
- [57] H. Otneim and D. Tjøstheim. Conditional density estimation using the local Gaussian correlation. *Stat. Comput.*, 28(2):303–321, 2018. (Cited on page: 10.)
- [58] M. Peligrad. Some remarks on coupling of dependent random variables. *Statistics and Probability Letters*, 60(2):201–209, 2002. (Cited on page: 8.)
- [59] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11:355–607, 2020. (Cited on page: 3.)
- [60] 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. (Cited on pages: 8, 11.)
- [61] M. Rachdi and P. Vieu. Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, 137(9):2784–2801, 2007. (Cited on pages: 12, 13.)
- [62] J.O. Ramsay and B.W. Silverman. *Applied functional data analysis: methods and case studies*. Springer-Verlag New York, Inc., 175 Fifth Avenue, NY 10010, USA, 2002. (Cited on page: 1.)
- [63] E. Rio. *Asymptotic theory of weakly dependent random processes*. Springer, Berlin Heidelberg, 2017. (Cited on pages: 8, 11.)
- [64] B.W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall/CRC, FL, 1998. (Cited on page: 9.)
- [65] 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. (Cited on pages: 1, 2.)

- [66] J. N. Tinio, M. Z. Alaya, and S. Bouzebda. Bounds in Wasserstein distance for locally stationary processes, 2024. (Cited on pages: 3, 9, 10, 11, 12, 15, 17, 19, 20, 21, and 30.)
- [67] 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. (Cited on page: 5.)
- [68] A. van Delft and M. Eichler. Locally stationary functional time series. *Electronic Journal of Statistics*, 12:107–170, 2018. (Cited on pages: 2, 5, and 13.)
- [69] M. Vidyasagar. *A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems*. Springer Verlag, Berlin, 1997. (Cited on page: 8.)
- [70] C. Villani. *Optimal transport: old and new*. Springer, Verlag Berlin Heidelberg, 2009. (Cited on pages: 3, 5, and 23.)
- [71] M. Vogt. Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40(5):2601–2633, 2012. (Cited on pages: 2, 5, 9, and 26.)
- [72] G. S. Watson. Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372, 1964. (Cited on page: 2.)