Optimal Transport Guarantees to Nonparametric Regression for Locally Stationary Time Series

Jan Nino G. Tinio[1,2] **Mokhtar Z. Alaya**[1] **Salim Bouzebda**[1]
[1]Université de Technologie de Compiègne, LMAC, Compiègne, France
[2]Department of Mathematics, Caraga State University, Butuan City, Philippines

Abstract

Locally stationary time series (LSTS) represent an essential modeling paradigm for capturing the nuanced dynamics inherent in time series data, whose statistical characteristics, including mean and variance, evolve smoothly over time. In this paper, we propose a conditional probability distribution estimator for LSTS through Nadaraya–Watson (NW) kernel smoothing. NW estimator leverages local kernel smoothing to approximate the conditional distribution of a response variable given its covariates. Under mild conditions, we establish optimal transport convergence guarantees to the proposed NW-based conditional probability estimator. These guarantees are initially proven in the univariate setting using the Wasserstein distance, and subsequently in a multivariate setting employing the sliced Wasserstein distance. To corroborate our theoretical findings, we conduct a wide range of numerical experiments to assess the convergence rates and showcase the practical relevance of the estimator in capturing intricate temporal dependencies in complex nonstationary phenomena.

1 INTRODUCTION

Time series analysis (TSA) aims to study the historical and current behavior of certain variables to predict future patterns. Such analysis is pivotal to forecast and control potential future scenarios, for instance in predicting inflation rates, stock prices, unemployment rates (Weng et al., 2018; Dadashova et al., 2021), and meteorological factors (Jiang et al., 2020; Kolluru et al., 2021), among many others. While classical TSA operates under the assumption of stationarity, it is important to note that many time series display nonstationarity (Aue et al., 2015; Chen et al., 2016; Amato et al., 2020). One approach to model this nonstationarity is through LSTS, where these processes are locally approximated by a strictly stationary process in a finer-grid time interval (Dahlhaus, 1996; Dahlhaus and Subba Rao, 2006; Dahlhaus, 2012). Most of the statistical theoretical guarantees on LSTS in the literature are proposed for both the conditional mean and the variance functions. In a parametric framework, Dahlhaus (1996) obtained estimates by minimizing the generalized Whittle function using local periodograms. Nonparametric approaches rely on NW (Nadaraya, 1964; Watson, 1964) estimation procedure, that is a widely used local averaging method for estimating the conditional mean function (Kristensen, 2009; Vogt, 2012; Truquet, 2019; Kurisu, 2022).

To deviate from estimation of conditional mean function, various works have already been proposed for conditional distribution estimation (Hall et al., 1999; Ferraty and Vieu, 2006). Hall et al. (1999) considered strictly stationary processes and proposed two estimation methods: a local logistic distribution method and an adjusted NW estimation procedure. Ahmed et al. (2020) introduced an adaptive NW estimator for strictly stationary processes using varying bandwidth and proved asymptotic normality of the proposed estimator. In Bouanani and Bouzebda (2024), they developed a local polynomial estimator for the conditional cumulative distribution function (CDF) of a scalar Y_t given a functional X_t. For distributional regression, Dombry et al. (2024) extended Stone's theorem using Wasserstein distance and showed that the conditional CDF estimator with local probability weights is a universally consistent estimator of the true conditional CDF.

It is worthy to note that in conditional distribution estimation, we have to carefully choose a metric measuring

Proceedings of the 29[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

the distance between these distributions. In this work, we consider optimal transport (OT) metrics that have been recognized as effective tools in comparing probability distributions. OT solves problems centered around the shortest path principle (Peyré and Cuturi, 2020). One of the prominent metrics in OT is Wasserstein distance (Villani, 2009). It has gained many applications compared to Total Variation, Hellinger, and Kullback-Leibler divergence since it can be optimally estimated from samples under mild assumptions (Manole et al., 2022). However, the direct computation of Wasserstein distance in high-dimensional setting can be computationally expensive. To overcome this computational burden, sliced Wasserstein distance was proposed to approximate the Wasserstein distance by projecting multivariate distributions onto one-dimensional subspaces and averaging the resulting one-dimensional Wasserstein distances (Bonnotte, 2013; Bayraktar and Guo, 2021).

Contributions. The contributions of the present paper are threefold:

- We consider estimating the conditional probability distribution of LSTS rather than the conditional mean or variance functions, as was largely proposed in the literature.
- Under mixing conditions (Doukhan, 1994; Rio, 2017), we provide the convergence rate of NW conditional distribution estimator with respect to Wasserstein distance for a scalar target $Y_{t,T}$ given a d-dimensional locally stationary covariate $\boldsymbol{X}_{t,T}$.
- We extend the results to a multivariate setting, i.e., $\boldsymbol{Y}_{t,T} \in \mathbb{R}^q (q \geq 1)$, where we derive the convergence rate of NW conditional distribution estimator through sliced Wasserstein distance.

To the best of our knowledge, this is the first work that establishes OT bounds for conditional probability distributions for both univariate and multivariate LSTS. We then illustrate our theoretical findings through numerical experiments on synthetic and real-world datasets.

The structure of this paper is as follows. In Section 2, we present the regression estimation problem in LSTS, a brief background of local stationarity, and Wasserstein distance. We derive the main results in Section 3: we first define the NW kernel estimator, and then provide the rates of convergence of Wasserstein distance between estimated and true conditional distributions. We extend our result to the multivariate case in Section 4. Section 5 shows the results of numerical experiments. All the proofs and additional experiment results are postponed to the appendices.

2 PRELIMINARIES

We start introducing a background of LSTS and optimal transport through Wasserstein distance. We then present the mixing coefficient employed to assess weak dependency between observations in the time series.

2.1 Locally stationary time series

Suppose that we have access to $T \in \mathbb{N}$ random variables $\{Y_{t,T}, \boldsymbol{X}_{t,T}\}_{t=1,\dots,T}$, where $Y_{t,T}$ is real-valued and $\boldsymbol{X}_{t,T} = (X^1_{t,T}, \dots, X^d_{t,T})^\top \in \mathbb{R}^d$. We consider the following regression estimation problem

$$Y_{t,T} = m^\star\big(\frac{t}{T}, \boldsymbol{X}_{t,T}\big) + \varepsilon_{t,T}, \text{ for all } t = 1, \dots, T, \quad (1)$$

where $\{\varepsilon_{t,T}\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables such that $\mathbb{E}[\varepsilon_{t,T}|\boldsymbol{X}_{t,T}] = 0$. We assume that the covariate $\boldsymbol{X}_{t,T}$ is locally stationary and $Y_{t,T}$ is integrable. Note that $m^\star\big(\frac{t}{T}, \boldsymbol{X}_{t,T}\big) = \mathbb{E}[Y_{t,T}|\boldsymbol{X}_{t,T}]$ is the *oracle* conditional mean function in model (1), which does not depend on real-time t but rather on the rescaled time $u = \frac{t}{T}$. It is identified almost surely (a.s.) at all rescaled u-points if it is continuous in the time direction. In LSTS, this rescaled time refers to the transformation of the original time scale.

A wide range of interesting nonlinear process models fit into the general framework (1). An important example is the nonparametric time-varying autoregressive (tvAR) model: $Y_{t,T} = m^\star\big(\frac{t}{T}, Y_{t-1,T}, \dots, Y_{t-d,T}\big) + \varepsilon_{t,T}$, where $\boldsymbol{X}_{t,T} = (Y_{t-1,T}, \dots, Y_{t-d,T})^\top$ is the d-lag of $Y_{t,T}$; for instance, see (Vogt, 2012; Richter and Dahlhaus, 2019). Let us now formally define the notion of LSTS. We adopt the definition given in Vogt (2012).

Definition 1 *A process $\{\boldsymbol{X}_{t,T}\}_{t=1,\dots,T}$ is locally stationary if for each rescaled time point $u \in [0,1]$, there exists an associated strictly stationary process $\{\boldsymbol{X}_t(u)\}_{t=1,\dots,T}$ verifying*

$$\|\boldsymbol{X}_{t,T} - \boldsymbol{X}_t(u)\| \leq \big(\big|\frac{t}{T} - u\big| + \frac{1}{T}\big)U_{t,T}(u) \quad a.s.,$$

where $\{U_{t,T}(u)\}_{t=1,\dots,T}$ is a positive process such that $\mathbb{E}\big[(U_{t,T}(u))^\rho\big] < C_U$ for some $\rho > 0$ and $C_U < \infty$ independent of u, t, and T. The norm $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^d$.

Definition 1 states that around each rescaled time u, any d-dimensional LSTS $\{\boldsymbol{X}_{t,T}\}_{t=1,\dots,T}$ can be approximated by $\{\boldsymbol{X}_t(u)\}_{t=1,\dots,T}$, which is a strictly stationary process at each fixed u. This suggests that a nonstationary process can be presumed stationary at local time. The exponent ρ can be considered as an indicator of how well the approximation is being done: the larger a ρ, the better approximation of $\boldsymbol{X}_{t,T}$ by $\boldsymbol{X}_t(u)$ and moderate bounds for their absolute difference.

2.2 Optimal transport: 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., $\mathcal{P}_r(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x|^r \mu(\mathrm{d}x) < \infty\}$. We quantify the distance between probability measures $\mu, \nu \in \mathcal{P}_r(\mathbb{R})$ through the rth-Wasserstein distance (Villani, 2009), denoted by $W_r(\mu, \nu)$ and defined as

$$W_r(\mu,\nu) = \Big(\inf_{\pi\in\Pi(\mu,\nu)} \iint_{\mathbb{R}^2} |u-v|^r \pi(\mathrm{d}u,\mathrm{d}v)\Big)^{1/r} \quad (2)$$

where $\Pi(\mu,\nu)$ stands the set of probability measures on $\mathbb{R}\times\mathbb{R}$ with marginals μ and ν. Equation (2) states that $W_r(\mu,\nu)$ is the infimum of the expectation of distance between two random variables over all possible couplings, i.e., $W_r(\mu,\nu) = \big(\inf_{U\sim\mu,\ V\sim\nu} \mathbb{E}[|U-V|^r]\big)^{1/r}$.

A simple optimal coupling can be represented by a probability inverse transform: let $F_\mu(\cdot)$ and $F_\nu(\cdot)$ be the cumulative distribution functions (CDF) and $F_\mu^{-1}(\cdot)$ and $F_\nu^{-1}(\cdot)$ be the respective generalized inverse or quantile functions. Then, for a uniformly distributed random variable Z on $(0,1)$, we can construct an optimal coupling $(U,V) = (F_\mu^{-1}(Z), F_\nu^{-1}(Z))$ (Dedecker and Merlevede, 2017). Hence, in univariate setting, the minimization problem (2) boils down to

$$W_r(\mu,\nu) = \Big(\int_0^1 \big|F_\mu^{-1}(z) - F_\nu^{-1}(z)\big|^r \mathrm{d}z\Big)^{1/r}.$$

For $r = 1$ and using a change of variable, the 1-Wasserstein distance writes as

$$W_1(\mu,\nu) = \int_{\mathbb{R}} |F_\mu(v) - F_\nu(v)| \mathrm{d}v. \quad (3)$$

Clearly, $W_1(\mu,\nu)$ is the L_1-distance between the CDF $F_\mu(\cdot)$ and $F_\nu(\cdot)$.

Now, since we are dealing with sequences exhibiting weak dependency, let us define the mixing coefficient being considered in this paper.

2.3 Mixing condition

The convergence rates of LSTS estimation are established under weakly dependent conditions, often termed mixing conditions (Doukhan, 1994; Rio, 2017). These latter are used to measure the dependency degree between observation sets of a stochastic process when they get far apart in time. In a nutshell, the farthest time distance between observations, the lower dependency. One of the prominent mixing conditions is β-mixing, it has been utilized to prove central limit theorems and moment inequalities (Dedecker et al., 2007; Bosq, 2012; Poinas, 2019).

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 $\beta(k) = \sup_{1\leq t\leq T-k} \beta\big(\sigma(Z_{s,T}, 1\leq s\leq t), \sigma(Z_{s,T}, t+k \leq s \leq T)\big)$, where $\sigma(Z)$ denotes the σ-algebra generated by Z. The array $\{Z_{t,T}\}$ is said to be β-mixing or absolutely regular mixing if $\beta(k) \to 0$ as $k \to \infty$.*

Definition 2 entails asymptotic independence as $k \to \infty$ for a β-mixing process. As argued in Vidyasagar (1997), β-mixing is a "just right" assumption in analyzing weakly dependent sequences. β-mixing is highly desirable in practice, as many commonly used time series models exhibit this property (McDonald et al., 2011). Various types of β-mixing include exponentially β-mixing where $\beta(k) = \mathcal{O}\big(e^{-\gamma k}\big)$ for $\gamma > 0$ (Masuda, 2007; Lee, 2012). It can also be arithmetically β-mixing, i.e., $\beta(k) = \mathcal{O}\big(k^{-\gamma}\big)$ (Ferraty and Vieu, 2006; Vogt, 2012; Bouzebda, 2024).

3 WASSERSTEIN BOUNDS FOR UNIVARIATE LSTS

For a fixed sample time $t \in \{1,\dots,T\}$ and a covariate $\boldsymbol{x} \in \mathbb{R}^d$, we denote the conditional probability distribution of $Y_{t,T}|\boldsymbol{X}_{t,T} = \boldsymbol{x}$ by $\pi_t^\star(\cdot|\boldsymbol{x})$ and its associated conditional CDF by $F_t^\star(\cdot|\boldsymbol{x})$. The mean conditional regression function is then given by

$$m^\star(\frac{t}{T}, \boldsymbol{x}) = \mathbb{E}_{\pi_t^\star(\cdot|\boldsymbol{x})}[Y_{t,T}|\boldsymbol{X}_{t,T} = \boldsymbol{x}] = \int_{-\infty}^{\infty} y\, \mathrm{d}\pi_t^\star(y|\boldsymbol{x}).$$

Let K_1, K_2 be two 1-dimensional based kernel functions and h be a T-dependent bandwidth, i.e., $h = h(T)$ satisfying $h(T) \to 0$ as $T \to \infty$. Setting the scaled kernels $K_{h,i}(\cdot) = K_i(\frac{\cdot}{h})$, for $i = 1, 2$, we define.

Definition 3 *The NW estimator of $\pi_t^\star(\cdot|\boldsymbol{x})$ reads as $\hat\pi_t(\cdot|\boldsymbol{x}) = \sum_{a=1}^T \omega_a(\frac{t}{T}, \boldsymbol{x})\delta_{Y_{a,T}}$, where*

$$\omega_a(\frac{t}{T}, \boldsymbol{x}) = \frac{K_{h,1}(\frac{t}{T} - \frac{a}{T}) \prod_{j=1}^d K_{h,2}(x^j - X^j_{a,T})}{\sum_{a=1}^T K_{h,1}(\frac{t}{T} - \frac{a}{T}) \prod_{j=1}^d K_{h,2}(x^j - X^j_{a,T})}.$$

The associated conditional CDF to $\hat\pi_t(\cdot|\boldsymbol{x})$ is defined as, for all $y \in \mathbb{R}$,

$$\hat F_t(y|\boldsymbol{x}) = \sum_{a=1}^T \omega_a(\frac{t}{T}, \boldsymbol{x}) \mathbb{1}_{Y_{a,T} \leq y}. \quad (4)$$

The weights $\{\omega_a(u, \boldsymbol{x})\}_{a=1,\dots,T}$ are measurable functions of $\boldsymbol{x}$, $\boldsymbol{X}_{a,T}$, and u but do not depend on $Y_{a,T}$.

Note that NW estimator of $m^\star(u,\boldsymbol{x})$ is given by $\hat{m}(u,\boldsymbol{x}) = \sum_{a=1}^T \omega_a(u,\boldsymbol{x}) Y_{a,T}$, that involves two kernel functions: one is in the direction of the d-dimensional covariate $\boldsymbol{X}_{t,T}$ and the other is with respect to the rescaled time $u = \frac{t}{T}$. This means that we do not only smooth in the space-direction of $\boldsymbol{X}_{t,T}$ but also in the time-direction, allowing us to properly assign weights $\omega_a(\frac{t}{T},\boldsymbol{x})$ and then consider local behavior of the data in the rescaled time $\frac{t}{T}$.

Next, we present the assumptions about the underlying process in model (1) and NW estimator given in Definition 3.

3.1 Assumptions

Our main results are based on the following assumptions that are classical in LSTS (Fan and Masry, 1992; Hansen, 2008; Vogt, 2012) and conditional distribution estimation (Owen, 1986; Hall et al., 1999; Veraverbeke et al., 2014; Otneim and Tjøstheim, 2018).

Assumption 1 (Local stationarity) *Assume that $\{\boldsymbol{X}_{t,T}\}_{t=1,\ldots,T}$ has compact support $\mathcal{X}$ and is locally stationary approximated by $\{\boldsymbol{X}_t(u)\}$ for each time point $u \in [0,1]$. The density $f(u,\boldsymbol{x})$ of $\boldsymbol{X}_t(u)$ has continuous partial derivative, $\partial_j f(u,\boldsymbol{x}) := \frac{\partial}{\partial x^j} f(u,\boldsymbol{x})$, with respect to $\boldsymbol{x}$ for each $u \in [0,1]$.*

Assumption 1 establishes the smoothness of the density $f(u,\boldsymbol{x})$ w.r.t $\boldsymbol{x}$, allowing to use Taylor expansion in the proofs.

Assumption 2 (Kernel functions) *The based kernel $K_i(\cdot)$, $i = 1,2$, is symmetric about zero, bounded, and has compact support, that is, $K_i(z) = 0$ for all $|z| > C_i$ for some $C_i < \infty$. Additionally, it fulfills a Lipschitz condition with a positive constant $L_i < \infty$, such that $|K_i(z) - K_i(z')| \le L_i|z - z'|$, for all $z, z' \in \mathbb{R}$, and $\int K_i(z)\mathrm{d}z = 1$, $\int zK_i(z)\mathrm{d}z = 0$, and $\int z^2 K_i(z)\mathrm{d}z = \kappa < \infty$.*

Assumption 2 signifies that the kernel function has a bounded rate of change. By assuming that K_i is symmetric about zero, we allow either or both kernel functions to be box, triangle, quadratic, or Gaussian kernels. The kernels follow bounded second-moment regularity, leading each kernel to have finite variance and limiting influence of outliers.

Assumption 3 (Bandwidth regularity condition) *The bandwidth h satisfies $T^{-(\nu \wedge \frac{1}{2})} h^{-d-1} = o(1)$, where $\nu = \rho \wedge 1$, for $\rho > 0$ as introduced in Definition 1.*

Assumption 3 indicates that h converges slower to zero, for instance at a polynomial rate, i.e., $h = \mathcal{O}(T^{-\xi})$, for small $\xi > 0$. It is worth noting that the choice of bandwidth is crucial for the bias-variance tradeoff (Silverman, 1998): small h leads to over-fitting, producing an estimator with high variance and low bias, while large h may cause under-fitting. Assumption 3 is a strengthening of the usual condition $Th^{d+1} \to \infty$, needed to guarantee convergence to zero of our resulting bounds.

Assumption 4 (Conditional CDF) *The conditional CDF $F^\star_\cdot(\cdot|\cdot)$ is Lipschitzian, i.e., $\left|F^\star_a(\cdot|\boldsymbol{x}) - F^\star_t(\cdot|\boldsymbol{x'})\right| \le L_{F^\star}\left(\|\boldsymbol{x} - \boldsymbol{x'}\| + \left|\frac{a}{T} - \frac{t}{T}\right|\right)$, for some constant $L_{F^\star} < \infty$, and for all $a, t \in \{1,\ldots,T\}$, $\boldsymbol{x}, \boldsymbol{x'} \in \mathbb{R}^d$.*

Assumption 4 entails $F^\star_\cdot(\cdot|\cdot)$ to not change rapidly as the observation changes. This differs from the assumption used in (Hall et al., 1999; Veraverbeke et al., 2014; Otneim and Tjøstheim, 2018; Ahmed et al., 2020) where the conditional CDF is assumed to be twice differentiable.

Assumption 5 (Mixing condition) *The process $\{(\boldsymbol{X}_{t,T}, \varepsilon_{t,T})\}_{t=1,\ldots,T}$ is arithmetically β-mixing, that is, $\beta(k) \le Ak^{-\gamma}$ for some $A > 0$ and $\gamma > 2$. We further assume $\sum_{k=1}^\infty k^\zeta \beta(k)^{1-\frac{2}{p}} < \infty$, for some $p > 2$ and $\zeta > 1 - \frac{2}{p}$.*

Assumption 6 (Blocking condition) *There exists a sequence of positive integers $\{q_T\}$ satisfying $q_T \to \infty$ and $q_T = o\big(\sqrt{Th^{d+1}}\big)$, as $T \to \infty$.*

Assumptions 5 and 6 are useful for dependent sequence estimation procedures. Assumption 5 highlights the decay of β-mixing coefficient $\beta(k)$. In the proof of Theorem 1, Bernstein's blocking technique was used to create independent blocks (Bernstein, 1927).

3.2 Convergence rate in Wasserstein distance

We investigate the error between NW estimator $\hat{\pi}_t(\cdot|\boldsymbol{x})$ and true conditional distribution $\pi^\star_t(\cdot|\boldsymbol{x})$ by establishing the rate of convergence w.r.t Wasserstein distance.

Theorem 1 *Let Assumptions 1 - 6 hold and define $I_h = [C_1 h, 1 - C_1 h]$. Then,*

$$\sup_{\boldsymbol{x}\in\mathcal{X}, \frac{t}{T}\in I_h} \mathbb{E}\big[W_1\big(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi^\star_t(\cdot|\boldsymbol{x})\big)\big] = \mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\Big).$$

Theorem 1 ensures that the expectation of Wasserstein distance between the underlying conditional probability distributions converges to zero with nonstandard components of orders $\mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}} h^{d+1-\frac{1}{p}(1-\nu)}}\Big)$ and $\mathcal{O}\big(\frac{1}{T^\nu h^{d+\nu-1}}\big)$, and a standard component of order $\mathcal{O}(h)$. Generally, this convergence depends on the bandwidth h; as discussed in Assumption 3, it should slowly approach zero for this result to be valid. The first and second components, which depend on ν and p, are respective results of approximating $\{\boldsymbol{X}_{t,T}\}_{t=1,\ldots,T}$ by a locally stationary $\{\boldsymbol{X}_t(\frac{t}{T})\}_{t=1,\ldots,T}$ and by assuming that $\{\boldsymbol{X}_{t,T}\}_{t=1,\ldots,T}$ is β-mixing. More precisely, the dependency on $\nu = \rho \wedge 1$ comes directly from Definition 1, i.e., the difference between $\boldsymbol{X}_{t,T}$ and $\boldsymbol{X}_t(\frac{t}{T})$ is bounded by a product of a negligible term and a positive process $U_{t,T}(\frac{t}{T})$ having finite ρ-th moment. Additionally, $p > 2$ corresponds to the rate of decay of the β-mixing coefficient in Assumption 5. While the last component is obtained through Lipschitz continuity property of the conditional CDF $F^\star_\cdot(\cdot|\cdot)$. If $\nu = 1$, this convergence becomes $\mathcal{O}\big(\frac{1}{T^{\frac{1}{2}} h^{d+1}} + h\big)$.

For a strictly stationary stochastic process $\{Y_t, X_t\}$, where Y_t and X_t are scalar, Hall et al. (1999) (Theorem 1.*ii*) had shown the pointwise convergence of their proposed adjusted NW conditional distribution function estimator to be $\mathcal{O}\big(\frac{1}{\sqrt{Th}} + h^2\big)$.

We further bound the second moment of W_1 in Appendix B. The next succeeding results give the bounds for the expectation of general W_r. The proofs are deferred to Appendices C.2–C.5, respectively.

Corollary 1 *Assume that $Y_{t,T}$ is uniformly bounded by $M > 0$ and that Assumptions 1–6 hold. Then, for every $r \geq 1$,*

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}[W_r^r(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x}))] = \mathcal{O}\left(\frac{1}{T^{1/2} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\right).$$

In particular, for $r = 2$,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}\big[W_2^2(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x}))\big] = \mathcal{O}\left(\frac{1}{T^{1/2} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\right).$$

Remark 1 *Corollary 1 shows that the expected squared 2-Wasserstein distance between the estimated conditional distribution $\hat{\pi}_t(\cdot \mid \boldsymbol{x})$ and the oracle conditional distribution $\pi_t^\star(\cdot \mid \boldsymbol{x})$ converges to zero with the same rate as in Theorem 1. In this sense, $\mathbb{E}[W_2^2(\hat{\pi}_t, \pi_t^\star)]$ can be regarded as a* distributional analogue *of the mean squared error, with W_2 playing the role of an L_2-distance between full conditional laws, rather than between their means only.*

Under finite r-th moments, r–Wasserstein adheres to the following convergence.

Corollary 2 *Assume that Assumptions 1–6 hold and that, for some $r \geq 1$, the conditional distributions satisfy the uniform moment condition*

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, t,T} \int_{\mathbb{R}} |y|^r \, \pi_t^\star(\mathrm{d}y \mid \boldsymbol{x}) < \infty, \quad \textit{and}$$
$$\sup_{\boldsymbol{x}\in\mathcal{X},\, t,T} \int_{\mathbb{R}} |y|^r \, \hat{\pi}_t(\mathrm{d}y \mid \boldsymbol{x}) < \infty.$$

Then,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}[W_r^r(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x}))] = \mathcal{O}\left(\frac{1}{T^{1/2} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\right).$$

Furthermore, under integrability and tail conditions, the convergence in expected W_r of the underlying conditional distributions is provided as follows.

Corollary 3 *Let $r \geq 1$ and assume Assumptions 1–6 hold. Suppose furthermore that the conditional laws $\pi_t^\star(\cdot \mid \boldsymbol{x})$ and $\hat{\pi}_t(\cdot \mid \boldsymbol{x})$ have uniformly integrable r-th moments, i.e.*

$$\lim_{R\to\infty} \sup_{\boldsymbol{x}\in\mathcal{X},\, t,T} \int_{\{|y|>R\}} |y|^r \, \pi_t^\star(\mathrm{d}y \mid \boldsymbol{x}) = 0, \quad \textit{and}$$
$$\lim_{R\to\infty} \sup_{\boldsymbol{x}\in\mathcal{X},\, t,T} \int_{\{|y|>R\}} |y|^r \, \hat{\pi}_t(\mathrm{d}y \mid \boldsymbol{x}) = 0.$$

Then, as $T \to \infty$,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}\Big[W_r^r\big(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})\big)\Big] \longrightarrow 0.$$

In particular, this holds for $r = 2$, yielding a Wasserstein analogue of the mean squared error.

Moreover, we provide below the refined rate of convergence for W_r^r given an optimized choice of R.

Proposition 1 *Let $r \geq 1$ and suppose that Assumptions 1–6 hold. Assume furthermore that there exists $s > r$ such that*

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, t,T} \int_{\mathbb{R}} |y|^s \, \pi_t^\star(\mathrm{d}y \mid \boldsymbol{x}) < \infty, \quad \textit{and}$$
$$\sup_{\boldsymbol{x}\in\mathcal{X},\, t,T} \int_{\mathbb{R}} |y|^s \, \hat{\pi}_t(\mathrm{d}y \mid \boldsymbol{x}) < \infty. \tag{5}$$

Define

$$A_T := \sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}\Big[W_1\big(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})\big)\Big]. \tag{6}$$

By Theorem 1, we have

$$A_T = \mathcal{O}\left(\frac{1}{T^{1/2} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\right).$$

Then there exists a constant $C > 0$ such that, for all sufficiently large T,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}\Big[W_r^r\big(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})\big)\Big] \;\leq\; C\, A_T^{\frac{s-r}{s-1}}. \tag{7}$$

In particular, for $r = 2$,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \mathbb{E}\Big[W_2^2\big(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})\big)\Big] = \mathcal{O}\bigg(A_T^{\frac{s-2}{s-1}}\bigg),$$

which can be interpreted as a Wasserstein analogue of the mean squared error with a rate slowed down by the tail parameter s.

Proposition 1 shows that, under a uniform s-moment condition with $s > r$, one can choose the truncation level optimally to obtain a refined convergence rate for the expected r-Wasserstein distance, bounding it by a fractional power of A_T; in particular, for $r = 2$, the squared Wasserstein distance behaves like a mean–squared–error analogue with a rate slowed according to the tail parameter s.

On a different note, as a result of Theorem 1, the NW conditional mean estimator $\hat{m}$ of $m^\star$ verifies

Proposition 2 *One has,*

$$\begin{aligned}\sup_{\boldsymbol{x}\in\mathcal{X},\frac{t}{T}\in I_h} &\mathbb{E}\big[|\hat{m}(\frac{t}{T},\boldsymbol{x}) - m^\star(\frac{t}{T},\boldsymbol{x})|\big] \\ &\leq \sup_{\boldsymbol{x}\in\mathcal{X},\frac{t}{T}\in I_h} \mathbb{E}\big[W_1\big(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x})\big)\big]. \end{aligned}\tag{8}$$

Proposition 2 signifies that convergence rate of NW regression function estimator $\hat{m}(u,\boldsymbol{x})$ can also be obtained through Wasserstein distance. This latter is comparable with the uniform convergence rate of the NW mean function estimator in Vogt (2012) (Theorem 4.2), of order $\mathcal{O}_{\mathbb{P}}\big(\sqrt{\frac{\log T}{Th^{d+1}}} + \frac{1}{T^\nu h^d} + h^2\big)$. If we assume that $F_\cdot^\star(\cdot|\cdot)$ is twice differentiable, then we get a similar convergence rate for the bias component. The bound in the right hand side of (8), given in Theorem 1, is slower than that of $\hat{m}(u,\boldsymbol{x})$ given in Vogt (2012) since we are measuring the disparity between underlying distributions, taking into account all aspects of distributional differences, not just discrepancies between conditional means.

For a bandwidth that goes to zero at a polynomial rate, the explicit Wasserstein bound is given below.

Proposition 3 *Assume Assumptions 1 - 6 hold and let $h \asymp T^{-\xi}$, where $0 < \xi < \frac{\frac{1}{2}\wedge\nu}{d+1}$. Then,*

$$\begin{aligned}&\sup_{\boldsymbol{x}\in\mathcal{X},\frac{t}{T}\in I_h} \mathbb{E}\big[W_1\big(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x})\big)\big] \\ &= \mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}-\xi(d+1-\frac{1}{p}(1-\nu))}} + \frac{1}{T^{\nu-\xi(d+\nu-1)}} + \frac{1}{T^\xi}\Big).\end{aligned}$$

Proof of Proposition 3 follows the same lines of Theorem 1's proof, by setting $h \asymp T^{-\xi}$.

4 SLICED WASSERSTEIN BOUND FOR MULTIVARIATE LSTS

Let $\boldsymbol{Y}_{t,T} = (Y_{t,T}^1, \ldots, Y_{t,T}^q)^\top \in \mathbb{R}^q$ $(q \geq 2)$, $\boldsymbol{X}_{t,T} \in \mathbb{R}^d$, and consider the multivariate regression model:

$$\boldsymbol{Y}_{t,T} = \boldsymbol{m}^\star\big(\frac{t}{T}, \boldsymbol{X}_{t,T}\big) + \boldsymbol{\varepsilon}_{t,T},$$

where the conditional mean is given as $\boldsymbol{m}^\star\big(\frac{t}{T},\boldsymbol{X}_{t,T}\big) = \big(m^{\star 1}\big(\frac{t}{T},\boldsymbol{X}_{t,T}\big),\ldots,m^{\star q}\big(\frac{t}{T},\boldsymbol{X}_{t,T}\big)\big)^\top$ and the noise $\boldsymbol{\varepsilon}_{t,T} = \big(\varepsilon_{t,T}^1,\ldots,\varepsilon_{t,T}^q\big)^\top$. The variables $\{\varepsilon_{t,T}^l\}_{t\in\mathbb{Z}}$, for $l \in \{1,\ldots,q\}$, are i.i.d random variables independent of $\{\boldsymbol{X}_{t,T}\}_{t=1,\ldots,T}$. We denote by $\boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x}) \in \mathcal{P}(\mathbb{R}^q)$ the conditional distribution of $\boldsymbol{Y}_{t,T}|\boldsymbol{X}_{t,T} = \boldsymbol{x}$. An example that fits this framework is the time-varying vector autoregressive (tvVAR) model: $\boldsymbol{Y}_{t,T} = \boldsymbol{m}^\star\big(\frac{t}{T},\boldsymbol{Y}_{t-1,T},\ldots,\boldsymbol{Y}_{t-d,T}\big) + \boldsymbol{\varepsilon}_{t,T}$, where $\boldsymbol{X}_{t,T} = \big(\boldsymbol{Y}_{t-1,T},\ldots,\boldsymbol{Y}_{t-d,T}\big)^\top$ is the d-lag of the q-dimensional vector $\boldsymbol{Y}_{t,T}$ (Haslbeck et al., 2020; Li and Yuan, 2024).

Definition 4 *The NW estimator of $\boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x})$ is defined as $\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}) = \sum_{a=1}^T \omega_a(\frac{t}{T},\boldsymbol{x})\delta_{\boldsymbol{Y}_{a,T}}$, where $\omega_a(\frac{t}{T},\boldsymbol{x})$ is given in Definition 3 and $\delta_{\boldsymbol{Y}_{a,T}}$ represents a point mass at $\boldsymbol{Y}_{a,T} \in \mathbb{R}^q$. The associated conditional CDF to $\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x})$ writes as, for all $\boldsymbol{y} = (y^1,\ldots,y^q)^\top \in \mathbb{R}^q$,*

$$\hat{F}_t(\boldsymbol{y}|\boldsymbol{x}) = \sum_{a=1}^T \omega_a\big(\frac{t}{T},\boldsymbol{x}\big)\mathbb{1}_{Y_{a,T}^1\leq y^1,\ldots,Y_{a,T}^q\leq y^q}.$$

Remark 2 *The NW estimator of $\boldsymbol{m}^\star$ is given by $\hat{\boldsymbol{m}}(u,\boldsymbol{x}) = \sum_{a=1}^T \omega_a(u,\boldsymbol{x})\boldsymbol{Y}_{a,T}$.*

When $\boldsymbol{Y}_{t,T} \in \mathbb{R}^q$, estimating the Wasserstein distance is often affected by the curse of dimensionality due to high computational complexity (Bayraktar and Guo, 2021). To overcome this complexity, sliced Wasserstein distance was introduced (Bonnotte, 2013). It only requires estimating the distance of the projected unidimensional distributions.

Sliced Wasserstein distance. Let $\mathbb{S}^{q-1} = \{\boldsymbol{\theta} \in \mathbb{R}^q : \|\boldsymbol{\theta}\|_2 = 1\}$ be the unit sphere in $\mathbb{R}^q$. Let $\boldsymbol{\theta}_\# : \mathbb{R}^q \to \mathbb{R}$ be the map defined by $\boldsymbol{\theta}_\#(\boldsymbol{v}) = \langle\boldsymbol{\theta},\boldsymbol{v}\rangle = \boldsymbol{\theta}^\top\boldsymbol{v}$. For any $\boldsymbol{\mu} \in \mathcal{P}_1(\mathbb{R}^q)$ and $\boldsymbol{\theta} \in \mathbb{S}^{q-1}$, we define the push-forward measure $\boldsymbol{\theta}_\#\boldsymbol{\mu}(I) = \boldsymbol{\mu}(\{\boldsymbol{v} \in \mathbb{R}^q : \boldsymbol{\theta}^\top\boldsymbol{v} \in I\})$, for any I Borelian in $\mathbb{R}$. The sliced Wasserstein distance of order one between $\boldsymbol{\mu},\boldsymbol{\eta} \in \mathcal{P}_1(\mathbb{R}^q)$ is defined as follows.

Definition 5 *For $\boldsymbol{\mu},\boldsymbol{\eta} \in \mathcal{P}_1(\mathbb{R}^q)$, the sliced Wasserstein distance of order one is defined as*

$$SW_1(\boldsymbol{\mu},\boldsymbol{\eta}) = \int_{\mathbb{S}^{q-1}} W_1(\boldsymbol{\theta}_\#\boldsymbol{\mu},\boldsymbol{\theta}_\#\boldsymbol{\eta})\sigma_{q-1}(\mathrm{d}\boldsymbol{\theta}), \tag{9}$$

where σ_{q-1} stands for the uniform measure on $\mathbb{S}^{q-1}$.

Sliced Wasserstein distance can be determined by averaging the Wasserstein distance between random 1-dimensional projections of distributions. Generally, this metric is weaker than Wasserstein distance, but it still preserves similar properties, making it an alternative application computation (Bonnotte, 2013; Manole et al., 2022).

For $\boldsymbol{\theta} \in \mathbb{S}^{q-1}$, let $\boldsymbol{\theta}_{\#}\boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x})$ be the pushforward measure of $\boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x})$ in the direction $\boldsymbol{\theta}$ with conditional CDF $F_{t,\boldsymbol{\theta}}^\star(\cdot|\boldsymbol{x})$. We estimate this pushforward measure by $\boldsymbol{\theta}_{\#}\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x})$ with conditional CDF $\hat{F}_{t,\boldsymbol{\theta}}(\cdot|\boldsymbol{x})$ defined, for all $y \in \mathbb{R}$,

$$\hat{F}_{t,\boldsymbol{\theta}}(y|\boldsymbol{x}) = \sum_{a=1}^{T} \omega_a\big(\frac{t}{T}, \boldsymbol{x}\big)\mathbb{1}_{\boldsymbol{\theta}^\top \boldsymbol{Y}_{a,T} \le y}. \tag{10}$$

We assume that, for fixed direction $\boldsymbol{\theta}$, $\hat{F}_{t,\boldsymbol{\theta}}(\cdot|\boldsymbol{x})$ obeys the following condition.

Assumption 7 (Multivariate conditional CDF) *For any $\boldsymbol{\theta} \in \mathbb{S}^{q-1}$, the projected conditional CDF $F_{\cdot,\boldsymbol{\theta}}^\star(\cdot|\cdot)$ is Lipschitzian, i.e., $\big|F_{a,\boldsymbol{\theta}}^\star(\cdot|\boldsymbol{x}) - F_{t,\boldsymbol{\theta}}^\star(\cdot|\boldsymbol{x}')\big| \le L_{F_{\boldsymbol{\theta}}^\star}\big(\|\boldsymbol{x}-\boldsymbol{x}'\| + \big|\frac{a}{T} - \frac{t}{T}\big|\big)$, for some constant $L_{F_{\boldsymbol{\theta}}^\star} < \infty$, and for all $a, t \in \{1, \ldots, T\}$, $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$.*

Similar to the univariate case, the projected conditional CDF $F_{\cdot,\boldsymbol{\theta}}^\star(\cdot|\cdot)$ likewise exhibits smooth behavior, changing slowly as observations change.

Using the aforementioned setting and condition, we yield the result below.

Theorem 2 *Let Assumptions 1 - 3 and 5 - 7 hold. Then,*

$$\sup_{\boldsymbol{x}\in\mathcal{X}, \frac{t}{T}\in I_h} \mathbb{E}\big[SW_1\big(\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}), \boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x})\big)\big] = \mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}}h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\Big).$$

Theorem 2 is an extension of Theorem 1 to the multivariate LSTS response $\boldsymbol{Y}_{t,T} \in \mathbb{R}^q$. We use sliced Wasserstein distance that allows the convergence of measures on $\mathbb{R}^q$ to be reduced to the convergence of their unidimensional projections with respect to a direction $\boldsymbol{\theta} \in \mathbb{S}^{q-1}$. As a by-product, at a direction $\boldsymbol{\theta}$, the convergence of the multidimensional measure $\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x})$ is identical to that of the univariate case.

5 NUMERICAL EXPERIMENTS

We conduct numerical experiments on synthetic and real-world datasets to calculate the empirical Wasserstein distance between NW estimator and true conditional CDF. We have made the implementation code of the experiments in Python using Pytorch and Scikit-learn packages. The code that generates all figures is available from `https://github.com/mzalaya/wasslsp` in the form of annotated programs, together with notebook tutorials.

5.1 Synthetic data

We consider univariate response case $Y_{t,T} \in \mathbb{R}$ and illustrate the convergence of NW estimator w.r.t Wasserstein distance for each of the following processes:

Gaussian tvAR(1). The time-varying autoregressive model for $p = 1$, tvAR(1) (Richter and Dahlhaus, 2019), with Gaussian noise is defined by

$$Y_{t,T} = \alpha\big(\frac{t}{T}\big)Y_{t-1,T} + \varepsilon_t,$$

where $\alpha(u) = 0.9\sin(2\pi u)$ and $\varepsilon_t \sim \mathcal{N}(0, 1)$. Its strictly stationary approximation at rescaled time u (Dahlhaus, 2012), is $Y_t(u) = \alpha(u)Y_{t-1}(u) + \zeta_t$, where $\zeta_t \sim \mathcal{N}(0, 1)$. The topmost time plot of Figure A.1 shows the resulting process $Y_{t,T}$ for $T = 1000$.

Cauchy tvAR(2). The second synthetic process is time-varying autoregressive model for $p = 2$, tvAR(2) (Birr et al., 2017), with Cauchy noise:

$$Y_{t,T} = 1.8\cos\big(1.5 - \cos(2\pi\frac{t}{T})\big)Y_{t-1,T} - 0.81Y_{t-2,T} + \varepsilon_t,$$

with i.i.d. Cauchy noise ε_t. For a rescaled time u, the strictly stationary approximation reads as $Y_t(u) = 1.8\cos\big(1.5 - \cos(2\pi u)\big)Y_{t-1}(u) - 0.81Y_{t-2}(u) + \zeta_t$, with i.i.d. Cauchy noise ζ_t (Birr et al., 2017). The process $Y_{t,T}$, for $T = 1000$, in this example is depicted in the second time plot of Figure A.1.

Gaussian AR(1). Finally, we also give an example of a stationary autoregressive process of order $p = 1$, AR(1), with Gaussian noise: $Y_t = \alpha Y_{t-1} + \varepsilon_t$, where $\alpha = 0.9\sin(2\pi)$, and $\varepsilon_t \sim \mathcal{N}(0, 1)$. This process is plotted at the bottom portion of Figure A.1, which behaves stationarily compared to the plot of Gaussian tvAR(1).

We compute Wasserstein distance between the true and NW conditional distributions of these processes. However, we highlight a remark below.

Remark 3 *The true conditional probability distribution and NW estimator are calculated for a fixed time $t \in \{1, \ldots, T\}$. Hence, obtaining empirical versions of these quantities from a single one-shot sampling is impossible. Towards this end, we replicate each process $L \ge 1$ times. At specific time t, we compute NW conditional CDF, for each $l \in \{1, \ldots, L\}$, and calculate* *the empirical conditional CDF. We then measure the corresponding approximated Wasserstein distance between the average NW and the empirical conditional CDFs. This Wasserstein distance computation using a replication procedure is detailed in Algorithm 1.*

Algorithm 1: Wasserstein distance calculus using data replication

---

input : sample size T, time point $t \in \{1, \ldots, T\}$, number of replications $L \geq 1$, based kernels $K_1(\cdot), K_2(\cdot)$, bandwidth h;

for $l = 1, \ldots, L$ **do**

  `# Generate` l`-th replication process` $\{Y_{a,T}^{(l)}\}_{a=1,\ldots,T}$

  for $a = 1, \ldots, T$ **do**

    $Y_{a,T}^{(l)} \leftarrow m^\star\big(\frac{a}{T}, \boldsymbol{X}_{a,T}^{(l)}\big) + \varepsilon_{a,T}^{(l)}$;

  `# Calculate` l`-th NW conditional CDF estimator`

  $$\hat{F}_t^{(l)}(y|\boldsymbol{x}) \leftarrow \sum_{a=1}^{T} \omega_a\Big(\frac{t}{T}, \boldsymbol{x}\Big) \mathbb{1}_{Y_{a,T}^{(l)} \leq y};$$

`# Calculate average NW conditional CDF estimator`

$$\hat{F}_t^L(y|\boldsymbol{x}) \leftarrow \frac{1}{L}\sum_{l=1}^{L} \hat{F}_t^{(l)}(y|\boldsymbol{x});$$

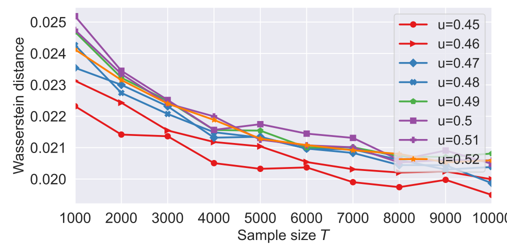
`# Calculate empirical conditional CDF`

$$F_t^L(y|\boldsymbol{x}) \leftarrow \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{Y_{t,T}^{(l)} \leq y};$$

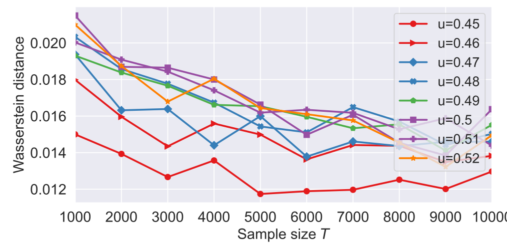
return : $W_1(\hat{F}_t^L(y|\boldsymbol{x}), F_t^L(y|\boldsymbol{x}))$;

Monte Carlo simulations. To illustrate theoretical results in Section 3, we provide 50 Monte Carlo runs of Algorithm 1 to get the expected W_1 distance between the underlying conditional distributions. We set $L = 1000$ and calculate the empirical conditional and NW CDFs, using (4). We consider different kernels $K_1(\cdot)$ and $K_2(\cdot)$ for each chosen process. We select $h = T^{-\xi}$, where $\xi = \frac{0.2}{d+1}$ for Gaussian tvAR(1) and Gaussian AR(1), and $\xi = \frac{0.3}{d+1}$ for Cauchy tvAR(2) and increase sample sizes $T \in [10^3, \ldots, 10^4]$. For based kernel K_1 belonging to Uniform, Rectangle, Triangle, and tricube, the constant $C_1 = 1$ and $I_h = [h, 1-h]$. Recall that our theoretical results are valid when $u = \frac{t}{T} \in I_h$, we then fix rescaled sample time $u = \frac{t}{T} \in [0.45, \ldots, 0.52]$.

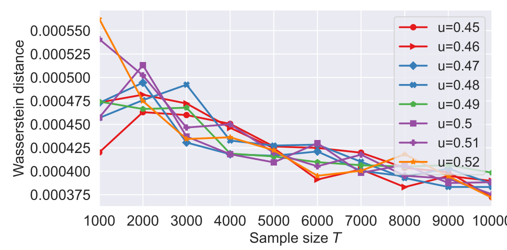
The expected Wasserstein distances are depicted in Figure 1. For each considered process, the supremum over u of the expected Wasserstein distances is approaching zero as T increases, which validates our theoretical result. It is also evident that Wasserstein distances of Cauchy tvAR(2) converge to zero more slowly than those of Gaussian tvAR(1) and Gaussian AR(1), since Cauchy tvAR(2) has higher dimension and extremely large fluctuations. This demonstrates how the theoretical convergence rate is affected by the covariate dimension d and local stationarity approximation. However,



(a) Gaussian tvAR(1), $K_1 = \text{Uniform}, K_2 = \text{Gaussian}$



(b) Cauchy tvAR(2) , $K_1 = \text{Triangle}, K_2 = \text{Gaussian}$



Gaussian AR(1), (c) $K_1 = \text{Uniform}, K_2 = \text{Gaussian}$

Figure 1: Expected Wasserstein distances between empirical conditional CDFs and estimated NW conditional CDFs, using (4), at different $u = \frac{t}{T}$ for increasing sample sizes $T \in [10^3, \ldots, 10^4]$ using specified kernels K_1 and K_2 for each process; bandwidths h are selected such that $h = T^{-\xi}$ with $0 < \xi < 0.5/(d+1)$; distances are calculated using $L = 1000$ replications and 50 Monte Carlo runs.

Wasserstein distances of Gaussian AR(1) are relatively smaller than Gaussian tvAR(1), indicating that the proposed estimation method is even more accurate when applied to stationary data. In general, the produced Wasserstein distances for all considered processes are small, signifying that NW conditional distribution estimator is robust in dealing with nonstationarity and extreme values. Moreover, the proposed estimator performs better compared to a classical method as discussed in Section A.1.1.

5.2 Real-world data

We utilize BabyECG ($T = 2048$) and HRV ($T = 17178$) datasets. BabyECG contains a record of the heart rate in beats per minute (bpm) of a 66-day-old infant, sampled every 16 seconds, and HRV records observations of instantaneous noninterpolated heart rate (niHR) frequency in bpm.

Remark 4 *Similar to our synthetic experiment, we want to calculate the NW and empirical conditional CDFs at fixed t. However, for this real data experiment, we could not get the empirical conditional CDF since we only have one data point value at specific t. In order to produce Wasserstein distances, we create copies of these datasets through replication, that relies on a Gaussian smoothed procedure explained below.*

Gaussian smoothing procedure. We replicate a given data $L \geq 1$ times by adding $Z_{t,T}^{(l)} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma > 0$ to each data observation $Y_{t,T}$: $Y_{t,T}^{(l)} = Y_{t,T} + Z_{t,T}^{(l)}$, for all replication $l \in \{1, \dots, L\}$ and $t \in \{1, \dots, T\}$. This is done since Corollary 1 in Nietert et al. (2021) ensures that $\lim_{\sigma \to 0} W(\mu + \mathcal{N}(0, \sigma^2), \nu + \mathcal{N}(0, \sigma^2)) = W(\mu, \nu)$, for $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. As an example, Figure A.3 presents time plots of Gaussian-smoothed datasets $Y_{t,T}^{(l)}$ with $Z_{t,T}^{(l)} \sim \mathcal{N}(0, 1)$ for $L = 3$ replications.

After replicating L times, we calculate the empirical and NW conditional CDFs at a specific time point t. We then measure the corresponding approximated Wasserstein distance between the empirical and average NW conditional CDFs. Algorithm 2 presents the calculus of Wasserstein distance approximation.

Algorithm 2: Wasserstein distance calculus using data replication and Gaussian smoothing

input : real dataset $\{Y_{a,T}\}_{a=1,\dots,T}$, $\sigma > 0$, time point $t \in \{1, \dots, T\}$, number of replications $L \geq 1$, based kernels $K_1(\cdot), K_2(\cdot)$, bandwidth h;

for $l = 1, \dots, L$ **do**

  `# Generate` l`-th replication` $\{Y_{a,T}^{(l)}\}_{a=1,\dots,T}$

  for $a = 1, \dots, T$ **do**

    $Y_{a,T}^{(l)} \leftarrow Y_{a,T} + Z_{a,T}^{(l)}$, where $Z_{a,T}^{(l)} \sim \mathcal{N}(0, \sigma^2)$;

  `# Calculate` l`-th NW conditional CDF estimator`

  $\hat{F}_t^{(l)}(y|\boldsymbol{x}) \leftarrow \sum_{a=1}^{T} \omega_a(\frac{t}{T}, \boldsymbol{x}) \mathbb{1}_{Y_{a,T}^{(l)} \leq y}$;

`# Calculate average NW conditional CDF estimator`

$\hat{F}_t^L(y|\boldsymbol{x}) \leftarrow \frac{1}{L} \sum_{l=1}^{L} \hat{F}_t^{(l)}(y|\boldsymbol{x})$;

`# Calculate empirical conditional CDF`

$F_t^L(y|\boldsymbol{x}) \leftarrow \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}_{Y_{t,T}^{(l)} \leq y}$;

return: $W_1(\hat{F}_t^L(y|\boldsymbol{x}), F_t^L(y|\boldsymbol{x}))$;

We next conduct an experiment to check the behavior of Wasserstein distance for increasing partitions of sample sizes T. Towards this end, we cut the observations at $S \in \{\frac{1}{3}T, \frac{2}{3}T, T\}$. We set $L = 1000$ and smoothing levels $\sigma \in \{1, 10^{-1}, 10^{-2}, 10^{-3}\}$. We quantify NW conditional CDF, through (4), using uniform and Gaussian kernels for K_1 and K_2, respectively. Similarly, we select $h = T^{-\xi}$ for $\xi = 0.2/(d + 1)$, and $d = 1$, and fix $u = \frac{t}{S} \in [h, 1 - h]$.

Figure A.5 shows the resulting Wasserstein distances. Plots in the first column represent Wasserstein distances for the partition $S = \frac{1}{3}T$, the second column for $S = \frac{2}{3}T$, and the last column for $S = T$. For each Gaussian-smoothed dataset, the supremum over $\frac{t}{T}$ of Wasserstein distances tends to be smaller for larger partitions $S = \frac{2}{3}T, T$, verifying our theoretical results. Moreover, it conveys the sensitivity of Wasserstein distance to the smoothing level σ. It is worth noting that Wasserstein distances do not change much for σ near to zero ($\sigma = 10^{-1}, 10^{-2}, 10^{-3}$) but vary slightly for $\sigma = 1$. This confirms that as $\sigma \to 0$, we preserve, to the maximum extent, the behavior of the given dataset, hence, Wasserstein distance for Gaussian-smoothed processes converges to the actual Wasserstein distance, as stated in Corollary 1 in Nietert et al. (2021).

CONCLUSION

We investigated Nadaraya-Watson (NW) conditional probability estimation for LSTS and established convergence rates w.r.t the Wasserstein distance in the univariate setting and the sliced Wasserstein distance in the multivariate case. These rates depend on the bandwidth in kernel smoothing, the degree of deviation from local stationarity approximation, and the weak dependence structure of the underlying time series. We then proposed a data-generating procedure for the synthetic data to compute the NW estimator, while for the real-world data, we used a Gaussian smoothing procedure. For future direction of this work, one may consider the Wasserstein-Fourier distance to measure similarity between power spectral densities involving LSTS as it was done in Cazelles et al. (2020) for stationary time series. In addition, the Gaussian smoothing procedure may be used to analyze privacy sensitive LSTS data, that usually arises in other research fields like medicine and psychology. Moreover, we view the development of lower bounds, such as those considered by Antos et al. (2000), and ultimately a complete optimality theory for conditional distribution estimation under Wasserstein metrics as an important direction for future research. This paper also lays the groundwork for future extensions involving nonparametric or machine-learning-based distributional estimators such as k-nearest neighbors, random forests, gradient boosting, or deep distributional models adapted to LSTS.

References

Ahmed, H., Salha, R, B., and EL-Sayed, H. (2020). Adaptive weighted nadaraya-watson estimation of the conditional quantiles by varying bandwidth. *Communications in Statistics - Simulation and Computation*, 49(5):1105–1117. (Cited on pages: 1, 4.)

Amato, F., Laib, M., Guignard, F., and Kanevski, M. (2020). Analysis of air pollution time series using complexity-invariant distance and information measures. *Physica A: Statistical Mechanics and Its Applications*, 547. (Cited on page: 1.)

Antos, A., Györfi, L., and Kohler, M. (2000). Lower bounds on the rate of convergence of nonparametric regression estimates. *Journal of Statistical Planning and Inference*, 83(1):91–100. (Cited on page: 9.)

Aue, A., Dubart Nourinho, D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.*, 110:378–392. (Cited on page: 1.)

Bayraktar, E. and Guo, G. (2021). Strong equivalence between metrics of Wasserstein type. *Electronic Communications in Probability*, 26:1–13. (Cited on pages: 2, 6.)

Benhenni, K., Ferraty, F., Rachdi, M., and Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.*, 22(3):353–369. (Cited on page: 16.)

Bernstein, S. N. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, 97:1–59. (Cited on pages: 4, 25.)

Birr, S., Volgushev, S., Kley, T., Dette, H., and Hallin, M. (2017). Quantile spectral analysis for locally stationary time series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79(5):1619–1643. (Cited on page: 7.)

Bonnotte, N. (2013). *Unidimensional and Evolution Methods for Optimal Transportation*. PhD thesis, Université paris sud. (Cited on pages: 2, 6, and 7.)

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*. Springer Science and Business Media. (Cited on page: 3.)

Bouanani, O. and Bouzebda, S. (2024). Limit theorems for local polynomial estimation of regression for functional dependent data. *AIMS Math.*, 9(9):23651–23691. (Cited on page: 1.)

Bouzebda, S. (2024). Weak convergence of the conditional single index U-statistics for locally stationary functional time series. *AIMS Math.*, 9(6):14807–14898. (Cited on pages: 3, 25.)

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory*, 18(1):169–192. (Cited on page: 16.)

Cazelles, E., Robert, A., and Tobar, F. (2020). The wasserstein-fourier distance for stationary time series. *IEEE Transactions on Signal Processing*, 69:709–721. (Cited on page: 9.)

Chen, S., Lei, L., and Tu, Y. (2016). Functional coefficient moving average model with applications to forecasting chinese cpi. *Statistica Sinica*, 26:1649–1672. (Cited on page: 1.)

Dadashova, B., Li, X., Turner, S., and Koeneman, P. (2021). Multivariate time series analysis of traffic congestion measures in urban areas as they relate to socioeconomic indicators. *Socio-Economic Planning Sciences*, 75. (Cited on page: 1.)

Dahlhaus, R. (1996). Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37. (Cited on page: 1.)

Dahlhaus, R. (2012). Locally stationary processes. In Subba Rao, T., Subba Rao, S., and Rao, C., editors, *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, pages 351–413. Elsevier. (Cited on pages: 1, 7.)

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for timevarying arch processes. *Ann. Statist.*, 34:1075–1114. (Cited on page: 1.)

Davydov, J. (1973). Mixing conditions for markov chains. *Teor. Verojatnost. i Primenen.*, pages 321–338. (Cited on page: 46.)

Dedecker, J., Doukhan, P., Lang, G., Leon, J., Louhichi, S., and Prieur, C. (2007). *Weak dependence: with examples and applications*. Springer Science and Business Media. (Cited on page: 3.)

Dedecker, J. and Merlevede, F. (2017). Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α-dependent sequences. *Bernoulli*, 23(3):2083–2127. (Cited on page: 3.)

Dombry, C., Modeste, T., and Pic, R. (2024). Stone's theorem for distributional regression in Wasserstein distance. *J. Nonparametr. Stat.*, 0(0):1–23. (Cited on page: 1.)

Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer-Verlag New York. (Cited on pages: 2, 3.)

Fan, J. and Masry, E. (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *Journal of Multivariate Analysis*, 43:237–271. (Cited on pages: 4, 25.)

Ferraty, F., Rabhi, A., and Vieu, P. (2005). Conditional quantiles for dependent functional data with application to the climatic “el niño” phenomenon. *Sankhyā: The Indian Journal of Statistics*, pages 378–398. (Cited on page: 16.)

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer, 233 Spring Street, New York,NY10013, USA. (Cited on pages: 1, 3, and 16.)

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94:154–163. (Cited on pages: 1, 4, 5, and 14.)

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748. (Cited on page: 4.)

Haslbeck, J., Bringmann, L., and Waldorp, L. (2020). A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivariate Behavioral Research*. (Cited on page: 6.)

Jaber, A., Rashid, T., Rasheed, M., Mahmood, R. S., and Maalej, O. (2024). Analysis of cauchy distribution and its applications. *Journal of Positive Sciences*, 4(4). (Cited on page: 13.)

Jiang, Y., Wu, X., and Guan, Y. (2020). Effect of ambient air pollutants and meteorological variables on covid-19 incidence. *Infection Control and Hospital Epidemiology*, 41(9):1011–1015. (Cited on page: 1.)

Kolluru, S., Patra, A., Nazneen, T., and Shiva Nagendra, S. (2021). Association of air pollution and meteorological variables with covid-19 incidence: Evidence from five megacities in india. *Infection Control and Hospital Epidemiology*, 195. (Cited on page: 1.)

Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory*, 25:1433–1445. (Cited on page: 1.)

Kurisu, D. (2022). Nonparametric regression for locally stationary functional time series. *Electron. J. Stat.*, 16(2):3973–3995. (Cited on pages: 1, 25.)

Lee, O. (2012). Exponential Ergodicity and β-Mixing Property for Generalized Ornstein-Uhlenbeck Processes. *Theoretical Economics Letters*, 2:21–25. (Cited on page: 3.)

Li, X. and Yuan, J. (2024). Deeptvar: Deep learning for a time-varying var model with extension to integrated var. *International Journal of Forecasting*, 40(3):1123–1133. (Cited on page: 6.)

Manole, T., Balakrishnan, S., and Wasserman, L. (2022). Minimax confidence intervals for the sliced Wasserstein distance. *Electron. J. Stat.*, 16(1):2252–2345. (Cited on pages: 2, 7.)

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, 115(1):155–177. (Cited on page: 25.)

Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56. (Cited on page: 3.)

McDonald, D., Shalizi, C., and Schervish, M. (2011). Estimating beta-mixing coefficients. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15, Fort Lauderdale, FL, USA. JMLR: W and CP. (Cited on page: 3.)

Nadaraya, E. A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.*, 9:157–159. (Cited on page: 1.)

Nietert, S., Goldfeld, Z., and Kato, K. (2021). Smooth p-Wasserstein Distance: Structure, Empirical Approximation, and Statistical Applications. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139. PMLR. (Cited on page: 9.)

Otneim, H. and Tjøstheim, D. (2018). Conditional density estimation using the local Gaussian correlation. *Stat. Comput.*, 28(2):303–321. (Cited on page: 4.)

Owen, A. (1986). *Nonparametric conditional estimation*. PhD thesis, Stanford University. (Cited on page: 4.)

Peyré, G. and Cuturi, M. (2020). Computational optimal transport. *Foundations and Trends in Machine Learning*, 11:355–607. (Cited on page: 2.)

Poinas, A. (2019). A bound of the β-mixing coefficient for point processes in terms of their intensity functions. *Statist. Probab. Lett.*, 148:88–93. (Cited on page: 3.)

Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, 137(9):2784–2801. (Cited on pages: 16, 17.)

Richter, S. and Dahlhaus, R. (2019). Cross validation for locally stationary processes. *Ann. Statist.*, 47(4):2145–2173. (Cited on pages: 2, 7, and 17.)

Rio, E. (2017). *Asymptotic theory of weakly dependent random processes*. Springer, Berlin Heidelberg. (Cited on pages: 2, 3.)

Rojo, J. (2013). Heavy-tailed densities. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(1):30–40. (Cited on page: 13.)

Silverman, B. (1998). *Density estimation for statistics and data analysis*. Chapman and Hall/CRC, FL. (Cited on page: 4.)

Truquet, L. (2019). Local stationarity and time-inhomogeneous Markov chains. *Ann. Statist.*, 47(4):2023–2050. (Cited on page: 1.)

Veraverbeke, N., Gijbels, I., and Omelka, M. (2014). Preadjusted non-parametric estimation of a conditional distribution function. *J. R. Statist. Soc. B*, 76(2):399–438. (Cited on page: 4.)

Vidyasagar, M. (1997). *A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems*. Springer Verlag, Berlin. (Cited on page: 3.)

Villani, C. (2009). *Optimal transport: old and new*. Springer, Verlag Berlin Heidelberg. (Cited on pages: 2, 3, and 43.)

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40(5):2601–2633. (Cited on pages: 1, 2, 3, 4, 6, 23, 24, and 46.)

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372. (Cited on page: 1.)

Weng, B., Martinez, W., Tsai, T., Li, C., Lu, L., Barth, J., and Megahed, F. (2018). Macroeconomic indicators alone can predict the monthly closing price of major u.s. indices: Insights from artificial intelligence, time-series analysis and hybrid models. *Applied Soft Computing*, 71:685–697. (Cited on page: 1.)

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Optimal Transport Guarantees to Nonparametric Regression for Locally Stationary Time Series: Supplementary Material

This supplementary material is organized as follows. Appendix A provides additional results of the conducted synthetic and real-world data experiments. It also includes an application of NW estimator to real-world data. We present additional theoretical results in Appendix B. Proofs of the main results and some helpful lemmas are given in Appendices C and D, respectively.

A ADDITIONAL EXPERIMENT RESULTS

To visualize and understand more the conducted numerical experiments, we also provide additional plots in this section.

A.1 Plots of synthetic data

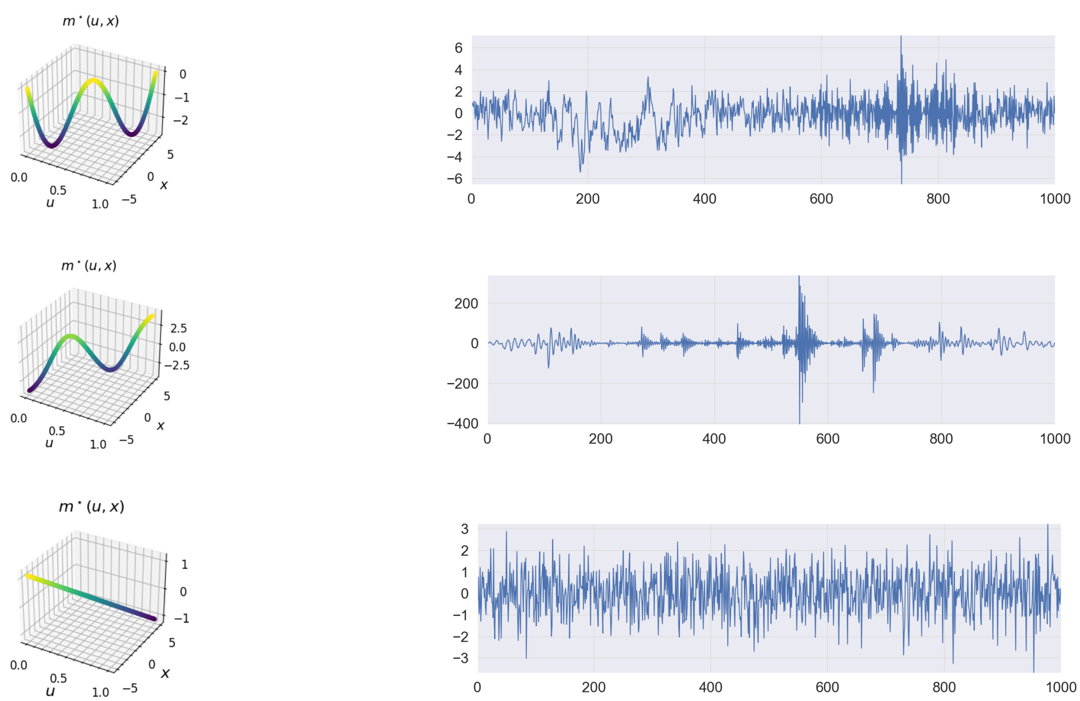


Figure A.1: Time plots of the simulated processes (right) with their corresponding true conditional mean function $m^\star(u, \boldsymbol{x})$ (left) for sample size $T = 1000$; from top to bottom: Gaussian tvAR(1), Cauchy tvAR(2), and Gaussian AR(1).

As shown in Figure A.1, Gaussian tvAR(1) (topmost) exhibits gradual downward and upward trends between time points $t = 100$ and $t = 400$. However, these trends are smooth over time, that is, the values remain tight at finer time intervals. For Cauchy tvAR(2) (second plot), most observations are centered around zero with relatively low-valued fluctuations. The process is affected by intermittent high-valued spikes at some time points of the series, which are due to the heavy-tailed property of Cauchy distributed error term ε_t (Rojo, 2013; Jaber et al., 2024). Lastly, Gaussian AR(1) (bottom) behaves stationarily compared to Gaussian tvAR(1) and Cauchy tvAR(2). The conditional mean functions $m^\star(u, x)$ of these example processes are also correspondingly placed beside each time plot. As shown, only the conditional mean function of the Gaussian AR(1) process is stationary for different values of $u \in [0, 1]$.

A.1.1 Comparison to a classical method

We also compare the performance of the proposed NW conditional CDF estimator for LSTS (NWLSTS) with the adjusted NW conditional CDF estimator (ANWSTS) introduced in Hall et al. (1999), given below:

$$\tilde{F}(y|x) = \frac{\sum_{t=1}^{T} \mathbb{1}_{Y_t \leq y} p_t(x)\, K_h(X_t - x)}{\sum_{t=1}^{T} p_t(x)\, K_h(X_t - x)},$$

where $p_t = p_t(x)$, for $1 \leq t \leq T$, denote weights depending on the data $X_1, \ldots, X_T$ and the point x, with the conditions that $p_t \geq 0$, $\sum_t p_t = 1$, and

$$\sum_{t=1}^{T} p_t(x)(X_t - x) K_h(X_t - x) = 0.$$

The weights p_t are chosen to maximize $\prod_t p_t$ while satisfying the above conditions. Using the same algorithm and parameter setting used in the Monte Carlo experiment in Section 5.1, Figure A.2 shows the Wasserstein distances obtained from using both NWLSTS and ANWSTS for Gaussian tvAR(1) with $T = 1000$. As shown by this result, the proposed NW estimator consistently has smaller Wasserstein distances compared to the adjusted NW estimator, suggesting that it captures better the conditional distribution of the given LSTS.

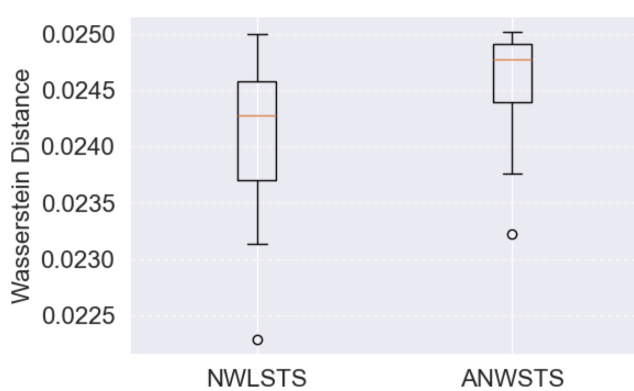
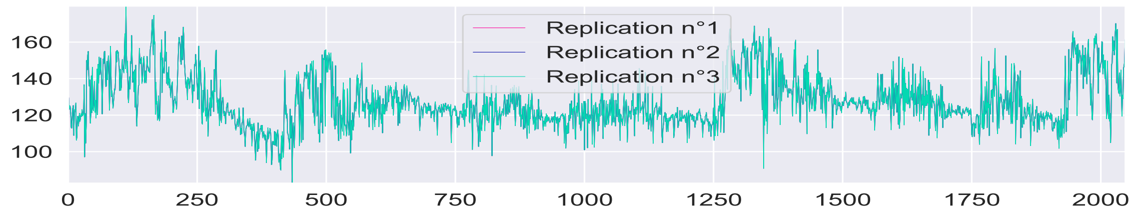
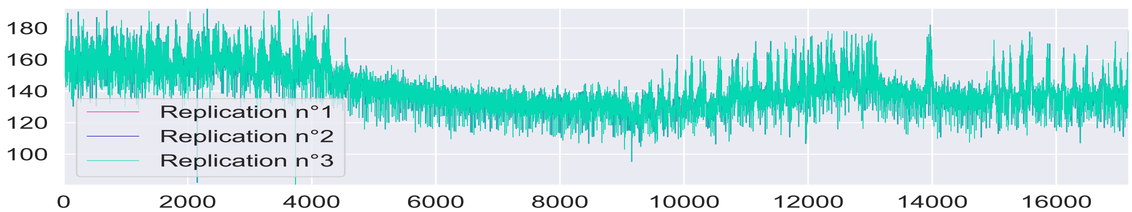


Figure A.2: Comparison of Wasserstein distances generated using NWLSTS and ANWSTS using Gaussian tvAR(1) for $T = 1000$ with $K_1 =$ Uniform and $K_2 =$ Gaussian; bandwidths h are selected such that $h = T^{-\xi}$ with $0 < \xi < 0.5/(d + 1)$; distances are calculated using $L = 1000$ replications and 50 Monte Carlo runs.

A.2 Plots of real-world data



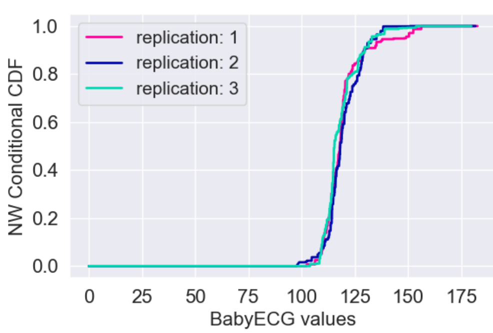
(a) Gaussian-smoothed BabyECG ($T = 2048$)



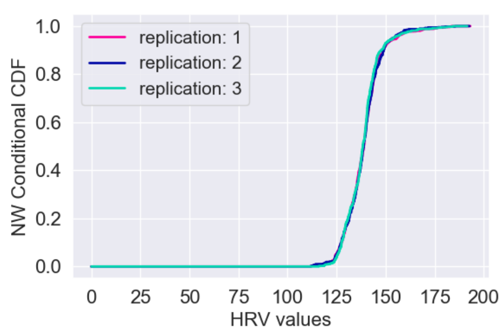
(b) Gaussian-smoothed HRV ($T = 17178$)

Figure A.3: Example plots of $Y^l_{t,T}$: real datasets smoothed with Gaussian noise $\mathcal{N}(0, 1)$ for replications $l = 1, 2, 3$ and $t = 1, \ldots, T$.

We plot $L = 3$ replicated Gaussian-smoothed datasets $Y^{(l)}_{t,T}$ with $Z^{(l)}_{t,T} \sim \mathcal{N}(0, 1)$ in Figure A.3. Looking at each replication, the mean of BabyECG and HRV changes gradually. Moreover, we computed NW conditional CDFs of the replicated Gaussian-smoothed BabyECG at $t = 970$ and Gaussian-smoothed HRV at $t = 7950$, see Figure A.4. We observe that NW conditional CDFs of Gaussian-smoothed HRV, having more data points, tend to be smoother.



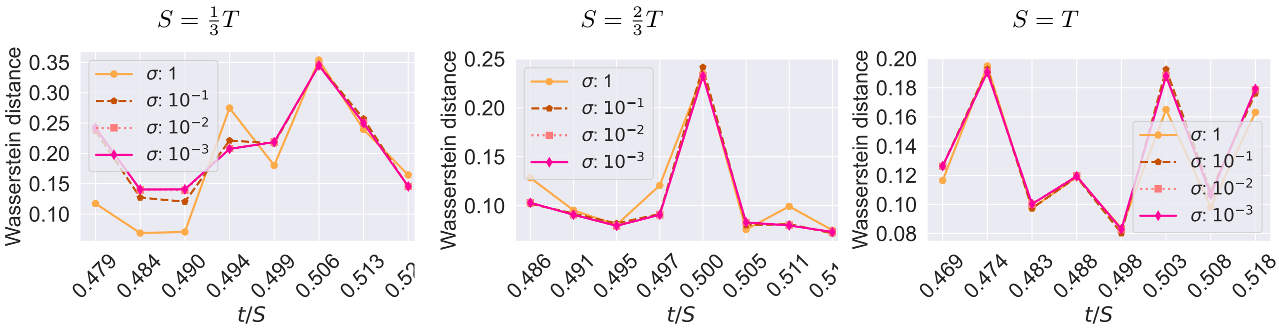
(a) Gaussian-smoothed BabyECG; $t = 970$



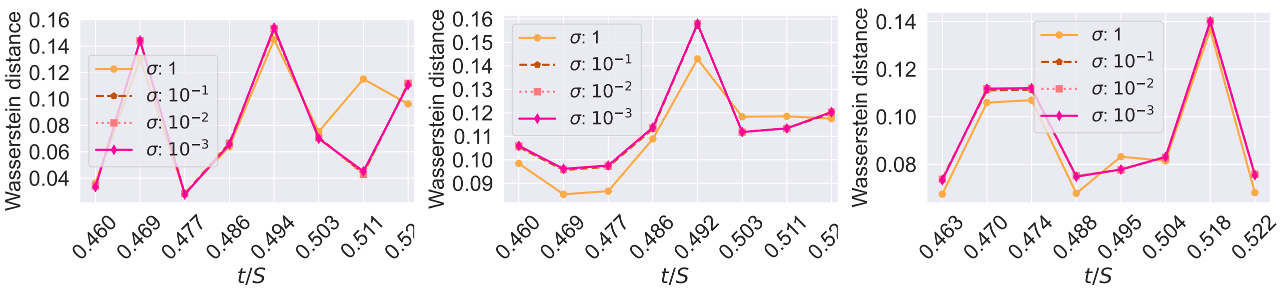
(b) Gaussian-smoothed HRV; $t = 7950$

Figure A.4: Plots of estimated NW conditional CDFs, using (4), for each Gaussian smoothed-dataset shown in Figure A.3 at specified t using $K_1 =$ Uniform and $K_2 =$ Gaussian, and bandwidths $h = T^{-\xi}$ for $\xi = 0.2/(d + 1)$, and $d = 1$.

We plot in Figure A.5 the Wasserstein distances obtained by the conducted real-world data experiments in Section 5.2. The figure shows the behavior of Wasserstein distance w.r.t. the sample partition S and the choice of the Gaussian smoothing parameter σ.



(a) Gaussian-smoothed BabyECG



(b) Gaussian-smoothed HRV

Figure A.5: Wasserstein distance between empirical conditional CDFs of Gaussian-smoothed datasets, for different smoothness level σ, and corresponding estimated NW conditional CDFs using K_1 = Uniform and K_2 = Gaussian and bandwidths $h = S^{-\xi}$ for $\xi = 0.2/(d+1)$, and $d = 1$; values are computed for sample size partitions $S = \frac{1}{3}T, \frac{2}{3}T, T$.

A.3 NW conditional median estimator for BabyECG and HRV data

In both synthetic and real-world data experiments, we have shown that the proposed approach is efficient in capturing the distributional behavior of LSTS. We further demonstrate its applicability by modeling both BabyECG and HRV datasets. We employ NW conditional median estimator, a by-product of NW CDF estimator (4). Since $F_t^\star(y|\boldsymbol{x})$ is strictly increasing, the conditional quantile of order $\tau \in [0,1]$ can be written as $F_{\tau,t}^{-1}(\boldsymbol{x}) = \inf\{y \in \mathbb{R} : F_t^\star(y|\boldsymbol{x}) \geq \tau\}$ with guaranteed uniqueness (Ferraty et al., 2005; Ferraty and Vieu, 2006). Hence, we can define NW conditional quantile estimator as (Cai, 2002):

$$\hat{F}_{\tau,t}^{-1}(\boldsymbol{x}) = \inf\{y \in \mathbb{R} : \hat{F}_t(y|\boldsymbol{x}) \geq \tau\}, \qquad \text{for } \tau \in [0,1]. \tag{11}$$

Setting $\tau = 0.5$, Equation (11) is denoted as the NW conditional median estimator.

In this application, we set $\boldsymbol{X}_{t,T} = Y_{t-1,T}$ and calculate NW conditional median $\hat{Y}_{t,T} = \hat{F}_{0.5,t}^{-1}(\boldsymbol{x})$ using (11) with $\tau = 0.5$. We fix uniform and Gaussian kernels for K_1 and K_2, respectively. The bandwidth h is selected using a leave-one-out cross-validation procedure used in (Benhenni et al., 2007; Rachdi and Vieu, 2007), describe as follows.

Bandwidth selection criteria. We define

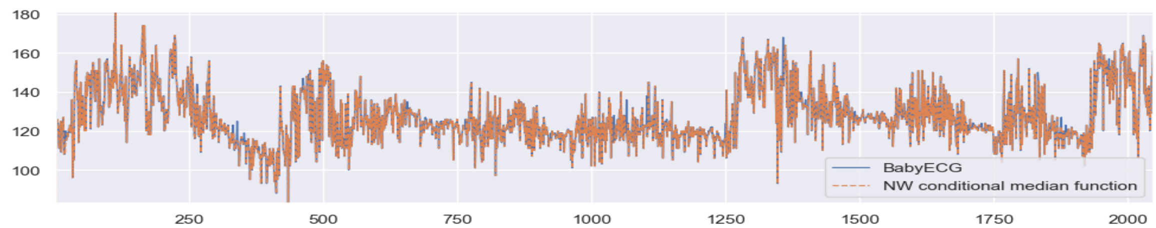
$$\hat{m}_i(\frac{t}{T}, \boldsymbol{x}) = \sum_{a=1; a \neq i}^{T} \omega_a(\frac{t}{T}, \boldsymbol{x}) Y_{a,T}, \tag{12}$$

for any fixed $i \in \{1, \ldots, T\}$, where $\omega_a(\frac{t}{T}, \boldsymbol{x})$ is given Definition 3. We denote (12) as the leave-out-$(\boldsymbol{X}_{i,T}, Y_{i,T})$ estimator of $m_i^\star(\frac{t}{T}, \boldsymbol{x})$. We then introduce the following leave-one-out cross-validation (LOOCV) criterion

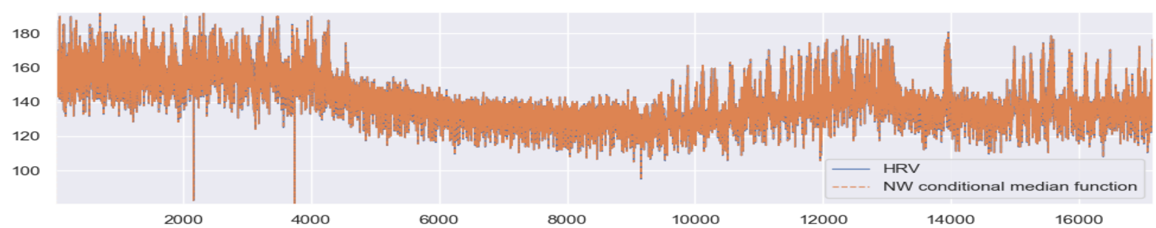
$$CV(y, \boldsymbol{x}, h) := \frac{1}{T} \sum_{i=1}^{T} \left(Y_{i,T} - \hat{m}_i(\frac{t}{T}, \boldsymbol{x})\right)^2. \tag{13}$$

We select a bandwidth $\hat{h}$ among $h \in [a_T, b_T]$ that minimizes (13) (Rachdi and Vieu, 2007). As highlighted in Theorem 3.8 in Richter and Dahlhaus (2019), $\hat{h}/h_0 \to 1$, for $h_0 \approx T^{-1/(d+4)}$, which satisfies Assumption 3. Hence, bandwidths selected using this cross-validation method are guaranteed to follow the established convergence rates in Section 3.

As it can be seen in Figure A.6, $\hat{Y}_{t,T}$ provides an accurate representation of the observed series $Y_{t,T}$. It effectively characterizes the central tendency of the conditional distribution of the given data, yielding very low mean absolute errors (MAEs). These results underscore the proposed NW estimation procedure's efficacy and precision in modeling LSTS.



(a) BabyECG; MAE $= 0.363001$.



(b) HRV; MAE $= 0.004863$.

Figure A.6: Actual values of real-world datasets with fitted values of the NW conditional median $\hat{Y}_{t,T}$ using $K_1 =$ uniform and $K_2 =$ Gaussian; h is selected using a leave-one-out cross-validation procedure.

B ADDITIONAL THEORETICAL RESULTS

Notation. For any real random variable X, we denote $\|X\|_{L_q}$ as the L_q-norm of X, for $q \geq 1$, i.e., $\|X\|_{L_q} = (\mathbb{E}[|X|^q])^{\frac{1}{q}}$. We say $a_T \lesssim b_T$ if there exists a constant C independent of T such that $a_T \leq C b_T$. For a given a_T and a sequence of random variables X_T, we write $X_T = \mathcal{O}_{\mathbb{P}}(a_T)$ if for any $\epsilon > 0$, there exists $C_\epsilon > 0$ and $T_\epsilon \in \mathbb{N}$ such that, for all $T \geq T_\epsilon$, $\mathbb{P}\big[\frac{|X_T|}{a_T} > C_\epsilon\big] < \epsilon$.

Let us examine the convergence rate of the second moment of W_1 and some lemmas involving the general r-Wasserstein distance between the considered NW estimator and true conditional distribution.

Corollary B.1 *Let Assumptions 1 - 6 hold. Then,*

$$\sup_{\boldsymbol{x} \in \mathcal{X}, \frac{t}{T} \in I_h} \|W_1(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x}))\|_{L_2} = \mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\Big).$$

Proof. We use the definition of W_1 given by (3) and Minkowski's integral inequality, that is, for any $r \geq 1$,

$$\Big\| \int \big|\hat{F}_t(y|\boldsymbol{x}) - F_t^\star(y|\boldsymbol{x})\big| \mathrm{d}y \Big\|_{L_r} \leq \int \big\|\hat{F}_t(y|\boldsymbol{x}) - F_t^\star(y|\boldsymbol{x})\big\|_{L_r} \mathrm{d}y.$$

By (3), $\|W_1(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x}))\|_{L_2} = \left\| \int_{\mathbb{R}} |\hat{F}_t(y|\boldsymbol{x}) - F_t^\star(y|\boldsymbol{x})| \mathrm{d}y \right\|_{L_2}$. For $r = 2$, we have

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

the last equality is obtained by (40) and (19). However, using Proposition C.2, $J_{t,T}^{-1}(\frac{t}{T},\boldsymbol{x}) = \mathcal{O}(1)$ and Proposition C.3, we get

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

Therefore,

$$
\|W_1(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x}))\|_{L_2} = \mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}}h^{(d+1)-\frac{1}{p}(1-\nu)}} + \frac{1}{T^{\nu}h^{d+\nu-1}} + h\Big),
$$

where $\nu = \rho \wedge 1$ and $p > 2$. □

Lemma B.1 (A moment-based control of W_r by W_1) *Let $r \geq 1$ and let $\mu, \nu \in \mathcal{P}_r(\mathbb{R})$ be probability measures with finite r-th moments*

$$
m_r(\mu) := \int_{\mathbb{R}} |x|^r \, \mu(\mathrm{d}x) < \infty, \qquad m_r(\nu) := \int_{\mathbb{R}} |y|^r \, \nu(\mathrm{d}y) < \infty.
$$

Then

$$
W_r^r(\mu,\nu) \;\leq\; W_1(\mu,\nu) \;+\; 2^{\,r-1}\big(m_r(\mu) + m_r(\nu)\big). \tag{14}
$$

Proof. By definition,

$$
W_r^r(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^2} |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y),
$$

where $\Pi(\mu,\nu)$ is the set of all couplings of μ and ν.

Fix any coupling $\pi \in \Pi(\mu,\nu)$ and consider two real numbers $x, y \in \mathbb{R}$. We use the elementary inequality

$$
|x-y|^r \;\leq\; |x-y| + 2^{\,r-1}\big(|x|^r + |y|^r\big), \tag{15}
$$

which holds for all $x, y \in \mathbb{R}$ and all $r \geq 1$. Indeed:

- If $|x-y| \leq 1$, then $|x-y|^r \leq |x-y|$, and (15) is immediate.
- If $|x-y| > 1$, then by the triangle inequality and $(a+b)^r \leq 2^{r-1}(a^r + b^r)$,

  $$
  |x-y| \leq |x| + |y| \quad \Longrightarrow \quad |x-y|^r \leq \big(|x| + |y|\big)^r \leq 2^{\,r-1}\big(|x|^r + |y|^r\big),
  $$

  so (15) also holds.

Integrating (15) with respect to π gives

$$
\iint |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) \;\leq\; \iint |x-y| \, \pi(\mathrm{d}x, \mathrm{d}y) \;+\; 2^{\,r-1} \iint \big(|x|^r + |y|^r\big) \, \pi(\mathrm{d}x, \mathrm{d}y).
$$

Since the marginals of π are μ and ν, we have

$$\iint |x|^r \, \pi(\mathrm{d}x, \mathrm{d}y) = \int |x|^r \, \mu(\mathrm{d}x) = m_r(\mu), \qquad \iint |y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) = m_r(\nu),$$

so

$$\iint |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) \;\le\; \iint |x-y| \, \pi(\mathrm{d}x, \mathrm{d}y) \;+\; 2^{\,r-1}\big(m_r(\mu) + m_r(\nu)\big).$$

Now take the infimum over all $\pi \in \Pi(\mu, \nu)$:

$$W_r^r(\mu, \nu) = \inf_\pi \iint |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) \;\le\; \inf_\pi \iint |x-y| \, \pi(\mathrm{d}x, \mathrm{d}y) \;+\; 2^{\,r-1}\big(m_r(\mu) + m_r(\nu)\big).$$

By definition of $W_1(\mu, \nu)$,

$$\inf_\pi \iint |x-y| \, \pi(\mathrm{d}x, \mathrm{d}y) = W_1(\mu, \nu),$$

which yields (14). □

Lemma B.2 (Truncation inequality for W_r) *Let $r \ge 1$ and $\mu, \nu \in \mathcal{P}_r(\mathbb{R})$. Then, for every $R > 0$,*

$$W_r^r(\mu, \nu) \;\le\; (2R)^{\,r-1} W_1(\mu, \nu) \;+\; 2^{\,r-1}\left(\int_{\{|x|>R\}} |x|^r \, \mu(\mathrm{d}x) + \int_{\{|y|>R\}} |y|^r \, \nu(\mathrm{d}y)\right). \tag{16}$$

Proof. Let $\pi \in \Pi(\mu, \nu)$ be any coupling of μ and ν. Write

$$\iint |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) = \iint_{\{|x|\vee|y|\le R\}} |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) + \iint_{\{|x|\vee|y|>R\}} |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y).$$

On the set $\{|x| \vee |y| \le R\}$, we have $|x-y| \le 2R$, and hence

$$|x-y|^r \le (2R)^{\,r-1} |x-y|.$$

Therefore,

$$\iint_{\{|x|\vee|y|\le R\}} |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) \le (2R)^{\,r-1} \iint |x-y| \, \pi(\mathrm{d}x, \mathrm{d}y).$$

On the complement $\{|x| \vee |y| > R\}$, we bound using

$$|x-y|^r \le 2^{\,r-1}\big(|x|^r + |y|^r\big),$$

so

$$\begin{aligned}
\iint_{\{|x|\vee|y|>R\}} |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) &\le 2^{\,r-1} \iint_{\{|x|\vee|y|>R\}} \big(|x|^r + |y|^r\big) \, \pi(\mathrm{d}x, \mathrm{d}y) \\
&\le 2^{\,r-1}\left(\iint_{\{|x|>R\}} |x|^r \, \pi(\mathrm{d}x, \mathrm{d}y) + \iint_{\{|y|>R\}} |y|^r \, \pi(\mathrm{d}x, \mathrm{d}y)\right) \\
&= 2^{\,r-1}\left(\int_{\{|x|>R\}} |x|^r \, \mu(\mathrm{d}x) + \int_{\{|y|>R\}} |y|^r \, \nu(\mathrm{d}y)\right),
\end{aligned}$$

using the marginal property of π in the last step.

Combining the two parts, we obtain

$$\iint |x-y|^r \, \pi(\mathrm{d}x, \mathrm{d}y) \le (2R)^{\,r-1} \iint |x-y| \, \pi(\mathrm{d}x, \mathrm{d}y) + 2^{\,r-1}\left(\int_{\{|x|>R\}} |x|^r \, \mu(\mathrm{d}x) + \int_{\{|y|>R\}} |y|^r \, \nu(\mathrm{d}y)\right).$$

Taking the infimum over all couplings $\pi \in \Pi(\mu, \nu)$ yields (16). □

C PROOFS OF MAIN RESULTS

Before providing the proofs of the main results, we begin with the following propositions that will be useful in the succeeding proofs. Proposition C.1 controls the expectation of terms involving $K_{h,2}$. Proposition C.2 bounds the general kernel density estimator for LSTS. While Proposition C.3 controls the squares of the weighted conditional CDF around the true conditional CDF.

Proposition C.1 *Let Assumptions 1 to 4 hold. Then, for $a, t \in \{1, \ldots, T\}$, the following inequalities hold:*

(C.1.1) $\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^d K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\Big|\Big] \le \frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^\nu}.$

(C.1.2) $\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j)\Big|\Big] \le \frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^\nu} + h^d f\big(\frac{t}{T}, \boldsymbol{x}\big) + h^{d+2}\frac{M}{2}\kappa d.$

(C.1.3) $K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j)[\mathbb{1}_{Y_{a,T} \le y} - F_t^\star(\cdot|\boldsymbol{x})]\Big]$

$$\le (\sqrt{d}C_2 + C_1)L_{F^\star}K_{h,1}\big(\tfrac{t}{T} - \tfrac{a}{T}\big)\Big\{\frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^{\nu-1}} + h^{d+1} f\big(\tfrac{t}{T}, \boldsymbol{x}\big) + h^{d+3}\frac{M}{2}\kappa d\Big\},$$

where $\nu = \rho \wedge 1$, $\kappa = \int z^2 K_2(z)\mathrm{d}z$, and $\sum_{j=1}^d \big|\partial_j f\big(\frac{t}{T}, \boldsymbol{x}\big)\big| \le M$.

Proof.

Proof of (C.1.1). Using *(D.1.1)* and K_2 is bounded by C_2, we have

$$\Big|\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^d K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\Big| \le C_2^{d-1}\sqrt{d}\sum_{j=1}^d \big|K_{h,2}(x^j - X_{a,T}^j) - K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\big|.$$

Note that for any bounded function $|f(x)| \le \iota$, we have $|f(x)|^{1-\nu} \le \iota^{1-\nu}$, which implies that $|f(x)| \le \iota^{1-\nu}|f(x)|^\nu$, for $1 - \nu \ge 0$. This yields,

$$\big|K_{h,2}(x^j - X_{a,T}^j) - K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\big| \le C_2^{1-\nu}\big|K_{h,2}(x^j - X_{a,T}^j) - K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\big|^\nu.$$

Set,

$$\begin{aligned}\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^d K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\Big|\Big] &\le C_2^{d-1}\sqrt{d}\,\mathbb{E}\Big[\sum_{j=1}^d \big|K_{h,2}(x^j - X_{a,T}^j) - K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\big|\Big] \\ &\le C_2^{d-\nu}\sqrt{d}\,\mathbb{E}\Big[\sum_{j=1}^d \big|K_{h,2}(x^j - X_{a,T}^j) - K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\big|^\nu\Big].\end{aligned}$$

Recall that K_2 is Lipschitz, we get

$$\begin{aligned}\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^d K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\Big|\Big] &\le \mathbb{E}\Big[L_2 C_2^{d-\nu}\sqrt{d}\sum_{j=1}^d \Big|\Big(\frac{x^j - X_{a,T}^j}{h}\Big) - \Big(\frac{x^j - X_a^j(\frac{a}{T})}{h}\Big)\Big|^\nu\Big] \\ &\le \mathbb{E}\Big[L_2 C_2^{d-\nu}\sqrt{d}\sum_{j=1}^d \Big|\frac{1}{h}\big(X_{a,T}^j - X_a^j\big(\frac{a}{T}\big)\big)\Big|^\nu\Big] \\ &= \frac{L_2 C_2^{d-\nu}\sqrt{d}}{h^\nu}\sum_{j=1}^d \mathbb{E}\Big[\big|\big(X_{a,T}^j - X_a^j\big(\frac{a}{T}\big)\big)\big|^\nu\Big].\end{aligned}$$

Clearly $\big|X_{a,T}^j - X_{a,T}^j\big(\frac{a}{T}\big)\big| \le \|X_{a,T}^j - X_{a,T}^j\big(\frac{a}{T}\big)\|_1$ and by Assumption 1, $\|X_{a,T}^j - X_{a,T}^j\big(\frac{a}{T}\big)\|_1 \le \frac{1}{T}U_{a,T}\big(\frac{a}{T}\big)$, where $\mathbb{E}\big[\big(U_{a,T}\big(\frac{a}{T}\big)\big)^\nu\big] < C_U$, we arrive at

$$\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^d K_{h,2}\big(x^j - X_a^j\big(\frac{a}{T}\big)\big)\Big|\Big] \le \frac{L_2 C_2^{d-\nu}\sqrt{d}}{T^\nu h^\nu}\sum_{j=1}^d \mathbb{E}\Big[\big|U_{a,T}\big(\frac{a}{T}\big)\big|^\nu\Big] \le \frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^\nu}.$$

Proof of (C.1.2). Using Assumption 1, $\boldsymbol{X}_{a,T}$ is locally stationary, and *(C.1.1)*,

$$\begin{aligned}\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X^j_{a,T})\Big|\Big] &\leq \mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j - X^j_{a,T}) - \prod_{j=1}^d K_{h,2}\big(x^j - X^j_a(\frac{a}{T})\big)\Big|\Big] + \mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}\big(x^j - X^j_a(\frac{a}{T})\big)\Big|\Big]\\ &\leq \frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^\nu} + \mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}\big(x^j - X^j_a(\frac{a}{T})\big)\Big|\Big].\end{aligned}$$

For the second term in the previous inequality, we have

$$\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}\big(x^j - X^j_a(\frac{a}{T})\big)\Big|\Big] = \int\cdots\int K_{h,2}(x^1-y^1)\cdots K_{h,2}(x^d-y^d) f(\frac{t}{T}, y^1,\ldots,y^d)\mathrm{d}y^1\cdots\mathrm{d}y^d.$$

Let the change of variable $z^j = \frac{x^j - y^j}{h}$, then

$$\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}\big(x^j - X^j_a(\frac{a}{T})\big)\Big|\Big] = \int\cdots\int K_2(z^1)\cdots K_2(z^d) f(\frac{t}{T}, x^1-hz^1,\ldots,x^d-hz^d)(-h)\mathrm{d}z^1\cdots(-h)\mathrm{d}z^d.$$

By Assumption 1, we use the first order Taylor expansion of $f(\frac{t}{T}, x^1-hz^1,\ldots,x^d-hz^d)$ wrt all x^j. Setting $f(\frac{t}{T}, x^1,\ldots,x^d) = f(\frac{t}{T},\boldsymbol{x})$, we have

$$\begin{aligned}f(\frac{t}{T}, x^1-hz^1,\ldots,x^d-hz^d) &= f(\frac{t}{T}, x^1,\ldots,x^d) + \sum_{j=1}^d \partial_j f(\frac{t}{T}, x^1,\ldots,x^d)(-h)z^j + R_1(h\boldsymbol{z})\\ &= f(\frac{t}{T},\boldsymbol{x}) + \sum_{j=1}^d \partial_j f(\frac{t}{T},\boldsymbol{x})(-h)z^j + R_1(h\boldsymbol{z}).\end{aligned}$$

The remainder part of this expansion verifies $R_1(h\boldsymbol{z}) \leq \frac{M}{2}h^2\|\boldsymbol{z}\|^2$, since $\partial_j f(\frac{t}{T},\boldsymbol{x})$ are continuous for $\boldsymbol{x}\in\mathcal{X}$. Hence, $\sum_{j=1}^d \big|\partial_j f_{X_t(\frac{t}{T})}(\boldsymbol{x})\big| \leq M < \infty$ for $\|\boldsymbol{x}-\boldsymbol{y}\| \leq h\|\boldsymbol{z}\|$, where $\boldsymbol{y} = (x^1-hz^1,\ldots,x^d-hz^d)$. That is, $R_1(h\boldsymbol{z})$ goes to zero as $h\to 0$. Now by Assumption 2, $\int K_2(z^j)\mathrm{d}z^j = 1$, $\int z^j K_2(z^j)\mathrm{d}z^j = 0$, and $\int (z^j)^2 K_2(z^j)\mathrm{d}z^j = \kappa$, so we have

$$\begin{aligned}&\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}\big(x^j - X^j_a(\frac{a}{T})\big)\Big|\Big]\\ &= (-h)^d\int\cdots\int K_2(z^1)\cdots K_2(z^d)\Big\{f(\frac{t}{T},\boldsymbol{x}) + \sum_{j=1}^d \partial_j f(\frac{t}{T},\boldsymbol{x})(-h)z^j + R_1(h\boldsymbol{z})\Big\}\mathrm{d}z^1\cdots\mathrm{d}z^d\\ &\leq (-h)^d\int\cdots\int K_2(z^1)\cdots K_2(z^d) f(\frac{t}{T},\boldsymbol{x})\mathrm{d}z^1\cdots\mathrm{d}z^d\\ &\quad - (-1)^{d+1}h^{d+1}\int\cdots\int K_2(z^1)\cdots K_2(z^d)\sum_{j=1}^d \partial_j f(\frac{t}{T},\boldsymbol{x})z^j\mathrm{d}z^1\cdots\mathrm{d}z^d\\ &\quad + (-1)^d h^d\frac{M}{2}\int\cdots\int K_2(z^1)\cdots K_2(z^d)h^2\|\boldsymbol{z}\|^2\mathrm{d}z^1\cdots\mathrm{d}z^d\\ &\leq (-h)^d f(\frac{t}{T},\boldsymbol{x}) - (-1)^{d+1}h^{d+1}\Big\{\partial_1 f(\frac{t}{T},\boldsymbol{x})\int\cdots\int K_2(z^2)\cdots K_2(z^d)\Big(\int z^1 K_2(z^1)\mathrm{d}z^1\Big)\mathrm{d}z^2\cdots\mathrm{d}z^d\\ &\quad + \cdots + \partial_d f(\frac{t}{T},\boldsymbol{x})\int\cdots\int K_2(z^1)\cdots K_2(z^{d-1})\Big(\int z^d K_2(z^d)\mathrm{d}z^d\Big)\mathrm{d}z^1\cdots\mathrm{d}z^{d-1}\Big\}\\ &\quad + (-1)^d h^{d+2}\frac{M}{2}\Big\{\int\cdots\int K_2(z^2)\cdots K_2(z^d)\Big(\int (z^1)^2 K_2(z^1)\mathrm{d}z^1\Big)\mathrm{d}z^2\cdots\mathrm{d}z^d\\ &\quad + \cdots + \int\cdots\int K_2(z^1)\cdots K_2(z^{d-1})\Big(\int (z^d)^2 K_2(z^d)\mathrm{d}z^d\Big)\mathrm{d}z^1\cdots\mathrm{d}z^{d-1}\Big\}.\end{aligned}$$

It entails

$$\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big)\Big|\Big] \leq (-h)^d f(\frac{t}{T}, \boldsymbol{x}) + (-1)^d h^{d+2}\frac{M}{2}\kappa d \leq h^d f(\frac{t}{T}, \boldsymbol{x}) + h^{d+2}\frac{M}{2}\kappa d. \tag{17}$$

Therefore,

$$\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big|\Big] \leq \frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^\nu} + h^d f(\frac{t}{T}, \boldsymbol{x}) + h^{d+2}\frac{M}{2}\kappa d.$$

Proof of (C.1.3). By Assumption 4, $\big|F_a^\star(y|\boldsymbol{X}_{a,T}) - F_t^\star(y|\boldsymbol{x})\big| \leq L_{F^\star}\big(\|\boldsymbol{X}_{a,T} - \boldsymbol{x}\| + \big|\frac{a}{T} - \frac{t}{T}\big|\big)$,

$$\begin{aligned}
&K_{h,1}(\frac{t}{T} - \frac{a}{T})\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)[\mathbb{1}_{Y_{a,T}\leq y} - F_t^\star(y|\boldsymbol{x})]\Big] \\
&\leq K_{h,1}(\frac{t}{T} - \frac{a}{T})\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\mathbb{E}\Big[\big(\mathbb{1}_{Y_{a,T}\leq y} - F_t^\star(y|\boldsymbol{x})\big)\Big|\boldsymbol{X}_{a,T}\Big]\Big] \\
&\leq K_{h,1}(\frac{t}{T} - \frac{a}{T})\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\big|F_a^\star(y|\boldsymbol{X}_{a,T}) - F_t^\star(y|\boldsymbol{x})\big|\Big] \\
&\leq L_{F^\star} K_{h,1}(\frac{t}{T} - \frac{a}{T})\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\big(\|\boldsymbol{X}_{a,T} - \boldsymbol{x}\| + \big|\frac{a}{T} - \frac{t}{T}\big|\big)\Big] \\
&\leq L_{F^\star} K_{h,1}(\frac{t}{T} - \frac{a}{T})\Big\{\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\|\boldsymbol{X}_{a,T} - \boldsymbol{x}\|\Big] + \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\big|\frac{a}{T} - \frac{t}{T}\big|\Big]\Big\}.
\end{aligned}$$

However, using Assumption 2, $|x^j - X_{a,T}^j| \leq C_2 h$ otherwise, $K_{h,2}(x^j - X_{a,T}^j) = 0$,

$$\begin{aligned}
\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\|\boldsymbol{X}_{a,T} - \boldsymbol{x}\|_2 &= \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\sqrt{\sum_{j=1}^{d}|x^j - X_{a,T}^j|^2} \\
&\leq \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\sqrt{d\max_j |x^j - X_{a,T}^j|^2} \\
&\leq \sqrt{d} C_2 h \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j).
\end{aligned}$$

Simultaneously, for the kernel $K_{h,1}$, $\big|\frac{a}{T} - \frac{t}{T}\big| \leq C_1 h$, otherwise $K_{h,1}\big(\big|\frac{a}{T} - \frac{t}{T}\big|\big) = 0$, then latter combined with *C.1.2*, we get

$$\begin{aligned}
&K_{h,1}(\frac{t}{T} - \frac{a}{T})\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)[\mathbb{1}_{Y_{a,T}\leq y} - F_t^\star(y|\boldsymbol{x})]\Big] \\
&\leq L_{F^\star} K_{h,1}(\frac{t}{T} - \frac{a}{T})\Big\{\sqrt{d} C_2 h \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big] + C_1 h \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big]\Big\} \\
&\leq (\sqrt{d} C_2 + C_1) L_{F^\star} h K_{h,1}(\frac{t}{T} - \frac{a}{T})\mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big] \\
&\leq (\sqrt{d} C_2 + C_1) L_{F^\star} K_{h,1}(\frac{t}{T} - \frac{a}{T})\Big\{\frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^{\nu-1}} + h^{d+1} f(\frac{t}{T}, \boldsymbol{x}) + h^{d+3}\frac{M}{2}\kappa d\Big\},
\end{aligned}$$

as claimed.

□

Proposition C.2 *Let Assumptions 1 - 3 hold, then for $I_h = [C_1 h, 1 - C_1 h]$,*

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \Big| J_{t,T}^{-1}(\frac{t}{T}, \boldsymbol{x}) \Big| = \sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in I_h} \Big| \Big(\frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j) \Big)^{-1} \Big| = \mathcal{O}(1).$$

Proof.

By application of Theorem 4.1 in Vogt (2012) with $W_{t,T} = 1$,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\, \frac{t}{T}\in[0,1]} \Big| J_{t,T}(\frac{t}{T}, \boldsymbol{x}) - \mathbb{E}\Big[J_{t,T}(\frac{t}{T}, \boldsymbol{x}) \Big] \Big| = \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th^{d+1}}} \Big).$$

Additionally, using Assumption 1, $J_{t,T}(\frac{t}{T}, \boldsymbol{x})$ can be decomposed as $J_{t,T}(\frac{t}{T}, \boldsymbol{x}) = \widetilde{J}_{t,T}(\frac{t}{T}, \boldsymbol{x}) + \bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})$, where

$$\widetilde{J}_{t,T}(\frac{t}{T}, \boldsymbol{x}) = \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) \prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big),$$

and

$$\bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x}) = \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) \Big\{ \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big) \Big\}.$$

Then,

$$\begin{aligned}
\Big| J_{t,T}(\frac{t}{T}, \boldsymbol{x}) \Big| &= \Big| J_{t,T}(\frac{t}{T}, \boldsymbol{x}) - \mathbb{E}[J_{t,T}(\frac{t}{T}, \boldsymbol{x})] + \mathbb{E}[J_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| \\
&\leq \Big| J_{t,T}(\frac{t}{T}, \boldsymbol{x}) - \mathbb{E}[J_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| + \Big| \mathbb{E}[J_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| \\
&\leq \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th^{d+1}}} \Big) + \Big| \mathbb{E}[J_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| \\
&\leq \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th^{d+1}}} \Big) + \Big| \mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T}, \boldsymbol{x}) + \bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| \\
&\leq \mathcal{O}_{\mathbb{P}}\Big(\sqrt{\frac{\log T}{Th^{d+1}}} \Big) + \Big| \mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| + \Big| \mathbb{E}[\bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big|.
\end{aligned}$$

Let us first observe $\Big| \mathbb{E}[\bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big|$. Using Assumptions 1 and 2 together with *(C.1.1)*, we have

$$\begin{aligned}
\Big| \mathbb{E}[\bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})] \Big| &\leq \mathbb{E}\Big[\Big| \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) \Big\{ \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big) \Big\} \Big| \Big] \\
&\leq \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) \mathbb{E}\Big[\Big\{ \Big| \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j) - \prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big) \Big\} \Big| \Big] \\
&\leq \Big(\frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^\nu} \Big) \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big).
\end{aligned}$$

For $I_h = [C_1 h, 1 - C_1 h]$, Lemma D.4 implies

$$
\begin{aligned}
\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) &\leq \sup_{u \in I_h} \Big| \frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(u - \frac{a}{T}\big)\Big| \\
&\leq \sup_{u \in I_h} \Big| \frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(u - \frac{a}{T}\big) - 1\Big| + 1 \\
&= \mathcal{O}\Big(\frac{1}{Th^2}\Big) + o(h) + 1 = \mathcal{O}(1). \qquad (18)
\end{aligned}
$$

It gives,

$$
\Big|\mathbb{E}[\bar{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})]\Big| \leq \big(\frac{L_2 C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^{d+\nu}}\big) \underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} \leq \frac{L_2 C C_U C_2^{d-\nu} d^{\frac{3}{2}}}{T^\nu h^{d+\nu}} \lesssim \frac{1}{T^\nu h^{d+\nu}}.
$$

On the other hand, using (17) and (18), we get

$$
\begin{aligned}
\Big|\mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})]\Big| &\leq \mathbb{E}\Big[\Big|\frac{1}{Th^{d+1}}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)\prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big)\Big|\Big] \\
&\leq \frac{1}{Th^{d+1}}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}\big(x^j - X_a^j(\frac{a}{T})\big)\Big|\Big] \\
&\leq \frac{1}{Th^{d+1}}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)\Big(h^d f(\frac{t}{T}, \boldsymbol{x}) + h^{d+2}\frac{M}{2}\kappa d\Big) \\
&\leq \Big(f(\frac{t}{T}, \boldsymbol{x}) + h^2 \frac{M}{2}\kappa d\Big) \underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} \\
&\lesssim f(\frac{t}{T}, \boldsymbol{x}) + h^2.
\end{aligned}
$$

Observe that $\big|\mathbb{E}[\widetilde{J}_{t,T}(\frac{t}{T}, \boldsymbol{x})]\big| > 0$, since $f(\dfrac{t}{T}, \boldsymbol{x}) \geq \inf\limits_{\boldsymbol{x} \in \mathcal{X}, u \in I_h} f(\dfrac{t}{T}, \boldsymbol{x})$. Moreover, using Theorem 4.1 in Vogt (2012),

$$
\begin{aligned}
J_{t,T}(\frac{t}{T}, \boldsymbol{x}) &\leq \big|J_{t,T}(\frac{t}{T}, \boldsymbol{x}) - f(\frac{t}{T}, \boldsymbol{x})\big| + f(\frac{t}{T}, \boldsymbol{x}) \\
&\leq \sup_{\boldsymbol{x} \in \mathcal{X}, u \in [0,1]} \big|J_{t,T}(u, \boldsymbol{x}) - f(\frac{t}{T}, \boldsymbol{x})\big| + f(\frac{t}{T}, \boldsymbol{x}) \\
&\leq o(1) + f(\frac{t}{T}, \boldsymbol{x}).
\end{aligned}
$$

Hence $\inf_{\boldsymbol{x} \in \mathcal{X}, u \in I_h} J_{t,T}(u, \boldsymbol{x}) \leq o(1) + \inf_{\boldsymbol{x} \in \mathcal{X}, u \in I_h} f(\frac{t}{T}, \boldsymbol{x})$. We finally get

$$
\frac{1}{J_{t,T}(\frac{t}{T}, \boldsymbol{x})} \leq \sup_{\boldsymbol{x} \in \mathcal{X}, u \in I_h} \frac{1}{J_{t,T}(u, \boldsymbol{x})} = \frac{1}{\inf_{\boldsymbol{x} \in \mathcal{X}, u \in I_h} J_{t,T}(u, \boldsymbol{x})} = \mathcal{O}(1).
$$

□

Proposition C.3 *Let Assumptions 1 - 6 be satisfied. For $\boldsymbol{x}, y \in \mathbb{R}^{d+1}$, define*

$$Z_{t,T}(y,\boldsymbol{x}) = \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\big[\mathbb{1}_{Y_{a,T} \leq y} - F_t^\star(y|\boldsymbol{x})\big].$$

Then

$$\mathbb{E}\big[Z_{t,T}^2(y,\boldsymbol{x})\big] = \mathcal{O}\Big(\frac{1}{Th^{2(d+1)+\frac{2}{p}(\nu-1)}} + \frac{1}{T^{2\nu}h^{2(d+\nu-1)}} + h^2\Big),$$

where $\nu = \rho \wedge 1$ and $p > 2$.

Sketch of proof. We employ Bernstein's blocking procedure (Bernstein, 1927): we first decompose $Z_{t,T}(y,\boldsymbol{x})$ as a sum of independent blocks: big blocks, small blocks, and a remainder block. The proof then continues in three steps. Step one controls the expectation of the square of sums of the big blocks, step two controls the expectation of the square of sums of the small blocks, and step three controls the expectation of the square of sum of the remainder block.

Proof.

To begin, rewrite

$$Z_{t,T}(y,\boldsymbol{x}) := \frac{1}{Th^{d+1}} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) Z_{a,t,T}(y,\boldsymbol{x}), \tag{19}$$

where

$$Z_{a,t,T}(y,\boldsymbol{x}) = \prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\big[\mathbb{1}_{Y_{a,T} \leq y} - F_t^\star(y|\boldsymbol{x})\big].$$

Applying Bernstein's big-block and small-block procedure on $Z_{t,T}(y,\boldsymbol{x})$, we partition the set $\{1,\ldots,T\}$ 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 (Fan and Masry, 1992; Masry, 2005; Kurisu, 2022; Bouzebda, 2024). So, we decompose $Z_{t,T}(y,\boldsymbol{x})$ as

$$Z_{t,T}(y,\boldsymbol{x}) = \Lambda_{t,T}(y,\boldsymbol{x}) + \Pi_{t,T}(y,\boldsymbol{x}) + \Xi_{t,T}(y,\boldsymbol{x}) := \sum_{l=0}^{v_T-1} \Lambda_{l,t,T}(y,\boldsymbol{x}) + \sum_{l=0}^{v_T-1} \Pi_{l,t,T}(y,\boldsymbol{x}) + \Xi_{t,T}(y,\boldsymbol{x}), \tag{20}$$

where

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

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

and

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

Let us define the size of the big blocks as $r_T = \lfloor \sqrt{Th^{d+1}}/q_T \rfloor$, where q_T satisfies Assumption 6, i.e., $q_T = o(\sqrt{Th^{d+1}})$. This further implies that there exists a sequence of positive integers $\{q_T\}$, $q_T \to \infty$, such that $q_T s_T = o\big(\sqrt{Th^{d+1}}\big)$. As $T \to \infty$, $\frac{s_T}{r_T} \to 0$ and $\frac{r_T}{T} \to 0$. Note that by defining $r_T = \lfloor \sqrt{Th^{d+1}}/q_T \rfloor$ immediately implies that $r_T = o\big(\sqrt{Th^{d+1}}\big)$, $s_T = o(r_T)$ and $v_T = o(q_T\sqrt{Th^{d+1}})$. Now,

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

However, the defined size of big blocks ensures 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

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

For convenience of notation, in the succeeding steps, the dependency on y and $\boldsymbol{x}$ is implicit.

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

$$\begin{aligned}\mathbb{E}\big[\Lambda_{t,T}^2\big] &= \sum_{l=0}^{v_T-1}\mathbb{E}\big[\Lambda_{l,t,T}^2\big] + \sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\mathbb{E}[\Lambda_{l,t,T}]\mathbb{E}[\Lambda_{l',t,T}]\\ &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\mathbb{E}\Big[\Big(\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)Z_{a,t,T}\Big)^2\Big]\\ &\quad + \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}Z_{b,t,T}\big]\\ &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\big[Z_{a,t,T}^2\big]\\ &\quad + \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{a=l(r_T+s_T)+1\\ |a-b|>0}}^{l(r_T+s_T)+r_T}\sum_{b=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}Z_{b,t,T}\big]\\ &\quad + \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}Z_{b,t,T}\big]\\ &=: \mathsf{S}_1^{\Lambda} + \mathsf{S}_2^{\Lambda} + \mathsf{S}_3^{\Lambda}.\end{aligned}$$

Step 1.1. Control of S_1^{Λ}. Considering S_1^{Λ}, we have,

$$\begin{aligned}\mathsf{S}_1^{\Lambda} &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\big[Z_{a,t,T}^2\big]\\ &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}^2(x^j-X_{a,T}^j)(\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x}))^2\Big].\end{aligned}$$

Observe that,

$$\begin{aligned}&K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}^2(x^j-X_{a,T}^j)(\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x}))^2\Big]\\ &\qquad\qquad\leq 2C_2^dK_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j-X_{a,T}^j)\big|\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x})\big|\Big].\end{aligned}$$

By *(C.1.3)*,

$$
\begin{aligned}
&K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}^{2}(x^{j}-X_{a,T}^{j})(\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x}))^{2}\Big]\\
&\quad\leq 2C_2^{d}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^{j}-X_{a,T}^{j})\big|\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x})\big|\Big]\\
&\quad\leq 2C_2^{d}(\sqrt{d}C_2+C_1)L_{F^{\star}}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\Big(\frac{L_2C_UC_2^{d-\nu}d^{\frac{3}{2}}}{T^{\nu}h^{\nu-1}}+h^{d+1}f\big(\frac{t}{T},\boldsymbol{x}\big)+h^{d+3}\frac{M}{2}\kappa d\Big)\\
&\quad\lesssim K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathsf{S}_1^{\Lambda}&\lesssim\frac{1}{T^{2}h^{2d+2}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}K_{h,1}^{2}\big(\frac{t}{T}-\frac{a}{T}\big)\\
&\leq\frac{C_1}{Th^{2d+1}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)\underbrace{\frac{1}{Th}\sum_{a=1}^{T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)}_{\mathcal{O}(1)}\\
&\lesssim\frac{1}{Th^{2d+1}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big) &(21)\\
&\lesssim\frac{1}{Th^{2d+\nu}}. &(22)
\end{aligned}
$$

Step 1.2. Control of S_2^{Λ}*.* On the other hand,

$$
\begin{aligned}
\mathsf{S}_2^{\Lambda}&=\frac{1}{(Th^{d+1})^{2}}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{\substack{b=l(r_T+s_T)+1\\|a-b|>0}}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}Z_{b,t,T}\big]\\
&=\frac{1}{(Th^{d+1})^{2}}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{\substack{b=l(r_T+s_T)+1\\|a-b|>0}}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{a,t,T},Z_{b,t,T}\big)\\
&\quad+\frac{1}{(Th^{d+1})^{2}}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{\substack{b=l(r_T+s_T)+1\\|a-b|>0}}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}\big]\mathbb{E}\big[Z_{b,t,T}\big]\\
&:=\mathsf{S}_{21}^{\Lambda}+\mathsf{S}_{22}^{\Lambda}.
\end{aligned}
$$

Step 1.2.1. Control of $\mathsf{S}_{21}^{\Lambda}$*.* We have

$$
\begin{aligned}
\mathsf{S}_{21}^{\Lambda}&=\frac{1}{(Th^{d+1})^{2}}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{\substack{b=l(r_T+s_T)+1\\|a-b|>0}}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{a,t,T},Z_{b,t,T}\big)\\
&=\frac{1}{(Th^{d+1})^{2}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{r_T}K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T},Z_{\lambda+n_2,t,T}\big)\\
&\leq\frac{1}{(Th^{d+1})^{2}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{r_T}K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\big|\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T},Z_{\lambda+n_2,t,T}\big)\big|,
\end{aligned}
$$

where $\lambda = l(r_T + s_T)$. Note that $\{\boldsymbol{X}_{t,T}, \varepsilon_{t,T}\}$ is β-mixing (Assumption 5), using Davydov's inequality (see Lemma D.3), for $p > 2$ and by Lemma D.2, we have

$$\begin{aligned}
&K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\Big|\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\big)\Big| \\
&\quad \le 8K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\big\|Z_{\lambda+n_1,t,T}\big\|_{L_p}\big\|Z_{\lambda+n_2,t,T}\big\|_{L_p}\beta(\sigma(\boldsymbol{X}_{\lambda+n_1,t,T}), \sigma(\boldsymbol{X}_{\lambda+n_2,t,T}))^{1-\frac{2}{p}} \\
&\quad \le 8K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\Big(\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X^j_{\lambda+n_1,T})(\mathbb{1}_{Y_{\lambda+n_1,T}\le y} - F_t^\star(y|\boldsymbol{x}))\Big|^p\Big]\Big)^{\frac{1}{p}} \\
&\qquad\qquad \times \Big(\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X^j_{\lambda+n_2,T})(\mathbb{1}_{Y_{\lambda+n_2,T}\le y} - F_t^\star(y|\boldsymbol{x}))\Big|^p\Big]\Big)^{\frac{1}{p}}\beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned}$$

Using *(C.1.3)*,

$$\begin{aligned}
&K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\Big|\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\big)\Big| \\
&\lesssim K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{1}{p}}K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{1}{p}}\beta(|n_1-n_2|)^{1-\frac{2}{p}} \\
&\lesssim K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned} \tag{23}$$

In consequence, by Assumption 5, $\sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} < \infty$, which can be expressed as $\sum_{k=1}^{r_T} k^\zeta \beta(k)^{1-\frac{2}{p}} + \sum_{k=r_T+1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}}$.

$$\begin{aligned}
\mathsf{S}_{21}^{\Lambda} &\lesssim \frac{1}{T^2h^{2d+2}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{r_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big)\beta(|n_1-n_2|)^{1-\frac{2}{p}} \\
&\le \frac{C_1^2}{T^2h^{2d+2}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{r_T}\beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned}$$

Letting $k = |n_1 - n_2|$ yields

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

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

$$\begin{aligned}
\mathsf{S}_{21}^{\Lambda} &\leq \frac{C_1^2 r_T}{T^2 h^{2d+2}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3} \Big)^{\frac{2}{p}} \sum_{l=0}^{v_T-1} \sum_{k=1}^{\infty} k^{\zeta} \beta(k)^{1-\frac{2}{p}} \\
&\lesssim \frac{v_T r_T}{T^2 h^{2d+2}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3} \Big)^{\frac{2}{p}} \sum_{k=1}^{\infty} k^{\zeta} \beta(k)^{1-\frac{2}{p}} \\
&\lesssim \frac{1}{T h^{2d+2}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3} \Big)^{\frac{2}{p}}, \quad \text{since } v_T r_T \leq \frac{T}{r_T} r_T = T, \\
&\lesssim \Big(\frac{1}{T^{p+2\nu} h^{2(d+1)p+2(\nu-1)}} + \frac{1}{T^{p} h^{2(d+1)p-2(d+1)}} \Big)^{\frac{1}{p}} \\
&\lesssim \frac{1}{T h^{2(d+1)-\frac{2}{p}(1-\nu)}}. \qquad (24)
\end{aligned}$$

Step 1.2.2. Control of $\mathsf{S}_{22}^{\Lambda}$. We have

$$\begin{aligned}
\mathsf{S}_{22}^{\Lambda} &= \frac{1}{(Th^{d+1})^2} \sum_{l=0}^{v_T-1} \sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T} \sum_{\substack{b=l(r_T+s_T)+1 \\ |a-b|>0}}^{l(r_T+s_T)+r_T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{b}{T}\big) \mathbb{E}\big[Z_{a,t,T}\big] \mathbb{E}\big[Z_{b,t,T}\big] \\
&= \frac{1}{(Th^{d+1})^2} \sum_{l=0}^{k_T-1} \sum_{n_1=1}^{r_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{r_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \mathbb{E}\big[Z_{\lambda+n_1,t,T}\big] \mathbb{E}\big[Z_{\lambda+n_2,t,T}\big] \\
&= \frac{1}{(Th^{d+1})^2} \sum_{l=0}^{v_T-1} \sum_{n_1=1}^{r_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{r_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \\
&\quad \times \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{\lambda+n_1,T}^j)(\mathbb{1}_{Y_{\lambda+n_1,T} \leq y} - F_t^{\star}(y|\boldsymbol{x})) \Big] \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{\lambda+n_2,T}^j)(\mathbb{1}_{Y_{\lambda+n_2,T} \leq y} - F_t^{\star}(y|\boldsymbol{x})) \Big].
\end{aligned}$$

By *(C.1.3)*, for $i = 1, 2$,

$$\begin{aligned}
K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_i}{T}\big) \mathbb{E}\big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{\lambda+n_i,T}^j)(\mathbb{1}_{Y_{\lambda+n_i,T} \leq y} - F_t^{\star}(y|\boldsymbol{x})) \big] \\
\lesssim K_{h,1}\big(\tfrac{t}{T} - \tfrac{\lambda+n_i}{T}\big) \big(\tfrac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3} \big),
\end{aligned}$$

then,

$$\begin{aligned}
\mathsf{S}_{22}^{\Lambda} &\lesssim \frac{1}{T^2 h^{2d+2}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3} \Big)^2 \sum_{l=0}^{v_T-1} \sum_{n_1=1}^{r_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{r_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \\
&\leq \frac{C_1}{T h^{2d+1}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3} \Big)^2 \underbrace{\frac{1}{Th} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} \\
&\lesssim \frac{1}{T h^{2d+1}} \Big(\frac{1}{T^{2\nu} h^{2(\nu-1)}} + h^{2(d+1)} \Big) \\
&\lesssim \frac{1}{T h^{2(d+\nu)-1}}. \qquad (25)
\end{aligned}$$

Step 1.3 Control of S_3^Λ. Observe that

$$\begin{aligned}
\mathsf{S}_3^\Lambda &= \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}[Z_{a,t,T}Z_{b,t,T}]\\
&= \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{a,t,T},Z_{b,t,T}\big)\\
&\quad + \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}[Z_{a,t,T}]\mathbb{E}[Z_{b,t,T}]\\
&=: \mathsf{S}_{31}^\Lambda + \mathsf{S}_{32}^\Lambda.
\end{aligned}$$

Step 1.3.1 Control of S_{31}^Λ. We have

$$\begin{aligned}
\mathsf{S}_{31}^\Lambda &= \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{a,t,T},Z_{b,t,T}\big)\\
&= \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\ l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{n_2=1}^{r_T} K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda'+n_2}{T}\big)\times\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T},Z_{\lambda'+n_2,t,T}\big),
\end{aligned}$$

where $\lambda = l(r_T+s_T)$ and $\lambda' = l'(r_T+s_T)$, however, for $l\neq l'$, see that for $n_1,n_2\in\{1,\dots,r_T\}$,

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

Define $m=\lambda+n_1$ and $m'=\lambda'+n_2$, we have

$$\begin{aligned}
\mathsf{S}_{31}^\Lambda &= \frac{1}{(Th^{d+1})^2}\sum_{\substack{m=1\\ |m-m'|>s_T}}^{v_T(r_T+s_T)-s_T}\sum_{m'=1}^{v_T(r_T+s_T)-s_T} K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{m,t,T},Z_{m',t,T}\big)\\
&\leq \frac{1}{(Th^{d+1})^2}\sum_{\substack{m=1\\ |m-m'|>s_T}}^{T}\sum_{m'=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)\big|\mathbb{C}\mathrm{ov}\big(Z_{m,t,T},Z_{m',t,T}\big)\big|.
\end{aligned}$$

Using (23),

$$\begin{aligned}
&K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)\big|\mathbb{C}\mathrm{ov}\big(Z_{m,t,T},Z_{m',t,T}\big)\big|\\
&\qquad\lesssim K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)\Big(\frac{1}{T^\nu h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\beta(|m-m'|)^{1-\frac{2}{p}}.
\end{aligned}$$

Thus, by Assumption 5, $\sum_{k=1}^\infty k^\zeta\beta(k)^{1-\frac{2}{p}}<\infty$,

$$\begin{aligned}
\mathsf{S}_{31}^\Lambda &\lesssim \frac{1}{T^2h^{2d+2}}\Big(\frac{1}{T^\nu h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\sum_{\substack{m=1\\ |m-m'|>s_T}}^{T}\sum_{m'=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)\beta(|m-m'|)^{1-\frac{2}{p}}\\
&\leq \frac{C_1^2}{T^2h^{2d+2}}\Big(\frac{1}{T^\nu h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\sum_{\substack{m=1\\ |m-m'|>s_T}}^{T}\sum_{m'=1}^{T}\beta(|m-m'|)^{1-\frac{2}{p}}.
\end{aligned}$$

Set $k = |m - m'|$, this yields

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

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

$$\begin{aligned}\mathsf{S}_{31}^{\Lambda} &\le \frac{C_1^2}{s_T^{\zeta}T^2h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\sum_{k=s_T+1}^{\infty}k^{\zeta}\beta(k)^{1-\frac{2}{p}}\\ &\lesssim \frac{1}{T^2h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}, \quad \text{since } \frac{1}{s_T^{\zeta}} \le 1,\\ &\lesssim \Big(\frac{1}{T^{2p}h^{2(d+1)p}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2\Big)^{\frac{1}{p}}\\ &\lesssim \Big(\frac{1}{T^{2p}h^{2(d+1)p}}\Big(\frac{1}{T^{2\nu}h^{2(\nu-1)}} + h^{2(d+1)}\Big)\Big)^{\frac{1}{p}}\\ &\lesssim \frac{1}{T^2h^{2(d+1)-\frac{2}{p}(1-\nu)}}.\end{aligned} \tag{26}$$

Step 1.3.2 Control of $\mathsf{S}_{32}^{\Lambda}$. Observe that

$$\begin{aligned}\mathsf{S}_{32}^{\Lambda} &= \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{b=l'(r_T+s_T)+1}^{l'(r_T+s_T)+r_T}K_{h,1}\Big(\frac{t}{T}-\frac{a}{T}\Big)K_{h,1}\Big(\frac{t}{T}-\frac{b}{T}\Big)\mathbb{E}[Z_{a,t,T}]\mathbb{E}[Z_{b,t,T}]\\ &= \frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{n_1=1}^{r_T}\sum_{n_2=1}^{r_T}K_{h,1}\Big(\frac{t}{T}-\frac{\lambda+n_1}{T}\Big)K_{h,1}\Big(\frac{t}{T}-\frac{\lambda'+n_2}{T}\Big)\mathbb{E}[Z_{\lambda+n_1,t,T}]\mathbb{E}[Z_{\lambda'+n_2,t,T}].\end{aligned}$$

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

$$\begin{aligned}\mathsf{S}_{32}^{\Lambda} &\le \frac{1}{(Th^{d+1})^2}\sum_{m=1}^{T}\sum_{\substack{m'=1\\|m-m'|>s_T}}^{T}K_{h,1}\Big(\frac{t}{T}-\frac{m}{T}\Big)K_{h,1}\Big(\frac{t}{T}-\frac{m'}{T}\Big)\mathbb{E}[Z_{m,t,T}]\mathbb{E}[Z_{m',t,T}]\\ &= \frac{1}{(Th^{d+1})^2}\sum_{m=1}^{T}\sum_{\substack{m'=1\\|m-m'|>s_T}}^{T}K_{h,1}\Big(\frac{t}{T}-\frac{m}{T}\Big)K_{h,1}\Big(\frac{t}{T}-\frac{m'}{T}\Big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j - X_{m,T}^j)(\mathbb{1}_{Y_{m,T}\le y} - F_t^{\star}(y|\boldsymbol{x}))\Big]\\ &\qquad \times \mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j - X_{m',T}^j)(\mathbb{1}_{Y_{m',T}\le y} - F_t^{\star}(y|\boldsymbol{x}))\Big].\end{aligned}$$

Using *(C.1.3)*, $K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)\mathbb{E}\big[\prod_{j=1}^{d}K_{h,2}(x^j - X_{m,T}^j)(\mathbb{1}_{Y_{m,T}\le y} - F_t^{\star}(y|\boldsymbol{x}))\big] \lesssim K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)\big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\big)$, then,

$$\begin{aligned}
\mathsf{S}_{32}^{\Lambda} &\lesssim \frac{1}{(Th^{d+1})^2}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^2 \sum_{m=1}^{T}\sum_{\substack{m'=1\\|m-m'|>s_T}}^{T} K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)\\
&\leq \frac{1}{h^{2d}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^2 \underbrace{\frac{1}{Th}\sum_{m=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{m}{T}\big)}_{\mathcal{O}(1)}\underbrace{\frac{1}{Th}\sum_{m'=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{m'}{T}\big)}_{\mathcal{O}(1)}\\
&\lesssim \frac{1}{h^{2d}}\Big(\frac{1}{T^{2\nu}h^{2(\nu-1)}}+h^{2(d+1)}\Big)\\
&\lesssim \frac{1}{T^{2\nu}h^{2(d+\nu-1)}}+h^2.
\end{aligned} \tag{27}$$

Hence, comparing (21), (24), (25), (26), and (27), we have

$$\mathbb{E}\big[\Lambda_{t,T}^2\big] \lesssim \frac{1}{Th^{2(d+1)-\frac{2}{p}(1-\nu)}}+\frac{1}{T^{2\nu}h^{2(d+\nu-1)}}+h^2. \tag{28}$$

Next, we deal with the small blocks.

Step 2. Control of the small blocks. We have

$$\begin{aligned}
\mathbb{E}\big[\Pi_{t,T}^2\big] &= \mathbb{E}\Big[\sum_{l=0}^{v_T-1}\Pi_{l,t,T}^2+\sum_{\substack{l=0\\l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\Pi_{l,t,T}\Pi_{l',t,T}\Big]\\
&= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\big[Z_{a,t,T}^2\big]\\
&\quad+\frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{a=l(r_T+s_T)+r_T+1\\a\neq b}}^{(l+1)(r_T+s_T)}\sum_{b=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}Z_{b,t,T}\big]\\
&\quad+\frac{1}{(Th^{d+1})^2}\sum_{\substack{l=0\\l\neq l'}}^{v_T-1}\sum_{l'=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)}\sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}Z_{b,t,T}\big]\\
&=: \mathsf{S}_1^{\Pi}+\mathsf{S}_2^{\Pi}+\mathsf{S}_3^{\Pi}.
\end{aligned}$$

Step 2.1. Control of S_1^{Π}. We have

$$\begin{aligned}
\mathsf{S}_1^{\Pi} &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}^2(x^j-X_{a,T}^j)(\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x}))^2\Big]\\
&\leq \frac{2C_2^d}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j-X_{a,T}^j)\big|\mathbb{1}_{Y_{a,T}\leq y}-F_t^{\star}(y|\boldsymbol{x})\big|\Big].
\end{aligned}$$

By *(C.1.3)*, we get

$$\begin{aligned}
\mathsf{S}_1^{\Pi} &\lesssim \frac{1}{T^2 h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}^2\big(\frac{t}{T}-\frac{a}{T}\big)\\
&\leq \frac{C_1}{T^2 h^{2d+2}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\\
&\leq \frac{C_1}{T h^{2d+1}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)\underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)}_{\mathcal{O}(1)}\\
&\lesssim \frac{1}{T h^{2d+1}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)\\
&\lesssim \frac{1}{T h^{2d+\nu}}. \qquad (29)
\end{aligned}$$

Step 2.2. Control of S_2^{Π}. On the other hand,

$$\begin{aligned}
\mathsf{S}_2^{\Pi} &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)}\sum_{\substack{b=l(r_T+s_T)+r_T+1\\ a\neq b}}^{(l+1)(r_T+s_T)} K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}[Z_{a,t,T}Z_{b,t,T}]\\
&= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{n_1=1\\|n_1-n_2|>0}}^{s_T}\sum_{n_2=1}^{s_T} K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\\
&\qquad\times\big\{\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\big) + \mathbb{E}[Z_{\lambda+n_1,t,T}]\mathbb{E}[Z_{\lambda+n_2,t,T}]\big\},
\end{aligned}$$

where $\lambda = l(r_T+s_T)+r_T$. We arrive at,

$$\begin{aligned}
\mathsf{S}_2^{\Pi} &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{n_1=1\\|n_1-n_2|>0}}^{s_T}\sum_{n_2=1}^{s_T} K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\mathbb{C}\mathrm{ov}\Big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\Big)\\
&\quad+\frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{n_1=1\\|n_1-n_2|>0}}^{s_T}\sum_{n_2=1}^{s_T} K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\mathbb{E}\Big[Z_{\lambda+n_1,t,T}\Big]\mathbb{E}\Big[Z_{\lambda+n_2,t,T}\Big]\\
&=: \mathsf{S}_{21}^{\Pi} + \mathsf{S}_{22}^{\Pi}.
\end{aligned}$$

Step 2.2.1. Control of S_{21}^{Π}. We have

$$\mathsf{S}_{21}^{\Pi} = \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{n_1=1\\|n_1-n_2|>0}}^{s_T}\sum_{n_2=1}^{s_T} K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\mathbb{C}\mathrm{ov}\Big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\Big).$$

Using (23),

$$\begin{aligned}
&K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\big|\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\big)\big|\\
&\qquad\lesssim K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned}$$

Thus, by Assumption 5,

$$\begin{aligned}
\mathsf{S}_{21}^{\Pi} &\lesssim \frac{1}{T^2h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{s_T}K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\beta(|n_1-n_2|)^{1-\frac{2}{p}}\\
&\le \frac{C_1^2}{T^2h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{s_T}\beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned}$$

In addition, letting $k=|n_1-n_2|$ yields

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

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

$$\begin{aligned}
\mathsf{S}_{21}^{\Pi} &\le \frac{C_1^2 s_T}{T^2h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\sum_{l=0}^{v_T-1}\sum_{k=1}^{\infty}k^{\zeta}\beta(k)^{1-\frac{2}{p}}\\
&\lesssim \frac{v_T s_T}{T^2h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}}\sum_{k=1}^{\infty}k^{\zeta}\beta(k)^{1-\frac{2}{p}}\\
&\lesssim \frac{1}{Th^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}}+h^{d+1}+h^{d+3}\Big)^{\frac{2}{p}},\quad \text{since } v_Ts_T\le\frac{T}{s_T}s_T=T,\\
&\lesssim \Big(\frac{1}{T^ph^{2(d+1)p}}\Big(\frac{1}{T^{2\nu}h^{2(\nu-1)}}+h^{2(d+1)}\Big)\Big)^{\frac{1}{p}}\\
&\lesssim \frac{1}{Th^{2(d+1)-\frac{2}{p}(1-\nu)}}.
\end{aligned}\tag{30}$$

Step 2.2.2. Control of S_{22}^{Π}. We have

$$\begin{aligned}
\mathsf{S}_{22}^{\Pi} &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{a=l(r_T+s_T)+1}^{l(r_T+s_T)+r_T}\sum_{\substack{b=l(r_T+s_T)+1\\|a-b|>0}}^{l(r_T+s_T)+r_T}K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}\big]\mathbb{E}\big[Z_{b,t,T}\big]\\
&= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{s_T}K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T}-\frac{\lambda+n_2}{T}\big)\\
&\quad\times\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j-X_{\lambda+n_1,T}^j)(\mathbb{1}_{Y_{\lambda+n_1,T}\le y}-F_t^{\star}(y|\boldsymbol{x}))\Big]\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j-X_{\lambda+n_2,T}^j)(\mathbb{1}_{Y_{\lambda+n_2,T}\le y}-F_t^{\star}(y|\boldsymbol{x}))\Big].
\end{aligned}$$

By *(C.1.3)*, for $i = 1, 2$,

$$K_{h,1}\big(\frac{t}{T} - \frac{\lambda + n_i}{T}\big)\mathbb{E}\big[\prod_{j=1}^{d} K_{h,2}(x^j - X^j_{\lambda+n_i,T})(\mathbb{1}_{Y_{\lambda+n_i,T}\leq y} - F^\star_t(y|\boldsymbol{x}))\big]$$
$$\lesssim K_{h,1}\big(\frac{t}{T} - \frac{\lambda + n_i}{T}\big)\big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\big).$$

Then

$$\begin{aligned}
\mathsf{S}^{\Pi}_{22} &\lesssim \frac{1}{T^2h^{2d+2}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \sum_{l=0}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{\substack{n_2=1\\|n_1-n_2|>0}}^{s_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \\
&\leq \frac{C_1}{Th^{2d+1}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} \\
&\lesssim \frac{1}{T^{1+2\nu}h^{2(d+\nu)-1}} + \frac{h}{T} \qquad (31)\\
&\lesssim \frac{1}{Th^{2(d+\nu)-1}}. \qquad (32)
\end{aligned}$$

Step 2.3. Control of S^{Π}_3. S^{Π}_3 can be decomposed as

$$\begin{aligned}
\mathsf{S}^{\Pi}_3 &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{l'=0\\l\neq l'}}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)}\sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{b}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{a,t,T}, Z_{b,t,T}\big) \\
&\quad + \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{l'=0\\l\neq l'}}^{v_T-1}\sum_{a=l(r_T+s_T)+r_T+1}^{(l+1)(r_T+s_T)}\sum_{b=l'(r_T+s_T)+r_T+1}^{(l'+1)(r_T+s_T)} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{b}{T}\big)\mathbb{E}\big[Z_{a,t,T}\big]\mathbb{E}\big[Z_{b,t,T}\big] \\
&= \mathsf{S}^{\Pi}_{31} + \mathsf{S}^{\Pi}_{32}.
\end{aligned}$$

Step 2.3.1 Control of S^{Π}_{31}. Setting

$$\mathsf{S}^{\Pi}_{31} = \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{l'=0\\l\neq l'}}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{n_2=1}^{s_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda'+n_2}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda'+n_2,t,T}\big),$$

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

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

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

$$\begin{aligned}
\mathsf{S}^{\Pi}_{31} &= \frac{1}{(Th^{d+1})^2}\sum_{q=r_T+1}^{v_T(r_T+s_T)}\sum_{\substack{q'=r_T+1\\|q-q'|>r_T}}^{v_T(r_T+s_T)} K_{h,1}\big(\frac{t}{T} - \frac{q}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{q'}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{q,t,T}, Z_{q',t,T}\big) \\
&= \frac{1}{(Th^{d+1})^2}\sum_{m=1}^{v_T(r_T+s_T)-r_T}\sum_{\substack{m'=1\\|m-m'|>r_T}}^{v_T(r_T+s_T)-r_T} K_{h,1}\big(\frac{t}{T} - \frac{m}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{m'}{T}\big)\mathbb{C}\mathrm{ov}\big(Z_{m,t,T}, Z_{m',t,T}\big) \\
&\leq \frac{1}{(Th^{d+1})^2}\sum_{m=1}^{T}\sum_{\substack{m'=1\\|m-m'|>r_T}}^{T} K_{h,1}\big(\frac{t}{T} - \frac{m}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{m'}{T}\big)\big|\mathbb{C}\mathrm{ov}\big(Z_{m,t,T}, Z_{m',t,T}\big)\big|,
\end{aligned}$$

where $m = q - r_T$ and $m' = q' - r_T$. Now, using (23), we have

$$K_{h,1}\big(\frac{t}{T} - \frac{m}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{m'}{T}\big)\big|\mathbb{C}\mathrm{ov}\big(Z_{m,t,T}, Z_{m',t,T}\big)\big|$$
$$\lesssim K_{h,1}\big(\frac{t}{T} - \frac{m}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{m'}{T}\big)\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\beta(|m - m'|)^{1-\frac{2}{p}}.$$

Thus, by Assumption 5,

$$\begin{aligned}
\mathsf{S}^{\Pi}_{31} &\lesssim \frac{1}{(Th^{d+1})^2}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}} \sum_{m=1}^{T}\sum_{\substack{m'=1\\|m-m'|>r_T}}^{T} K_{h,1}\big(\frac{t}{T} - \frac{m}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{m'}{T}\big)\beta(|m - m'|)^{1-\frac{2}{p}}\\
&\le \frac{C_1^2}{T^2h^{2d+2}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}} \sum_{m=1}^{T}\sum_{\substack{m'=1\\|m-m'|>r_T}}^{T}\beta(|m - m'|)^{1-\frac{2}{p}}.
\end{aligned}$$

Letting $k = |m - m'|$ entails

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

Hence,

$$\begin{aligned}
\mathsf{S}^{\Pi}_{31} &\lesssim \frac{1}{r_T^\zeta T^2h^{2(d+1)}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}\sum_{k=r_T+1}^{\infty}k^\zeta\beta(k)^{1-\frac{2}{p}}\\
&\lesssim \frac{1}{T^2h^{2(d+1)}}\Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}, \quad \text{since } \frac{1}{r_T^\zeta} \le 1,\\
&\lesssim \Big(\frac{1}{T^{2p}h^{2(d+1)p}}\Big(\frac{1}{T^{2\nu}h^{2(\nu-1)}} + h^{2(d+1)}\Big)\Big)^{\frac{1}{p}}\\
&\lesssim \frac{1}{T^2h^{2(d+1)-\frac{2}{p}(1-\nu)}}. \qquad (33)
\end{aligned}$$

Step 2.3.2 Control of S^{Π}_{32}. Observe that

$$\begin{aligned}
\mathsf{S}^{\Pi}_{32} &= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{l'=0\\l\neq l'}}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{n_2=1}^{s_T}K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda'+n_2}{T}\big)\mathbb{E}\big[Z_{\lambda+n_1,t,T}\big]\mathbb{E}\big[Z_{\lambda'+n_2,t,T}\big]\\
&= \frac{1}{(Th^{d+1})^2}\sum_{l=0}^{v_T-1}\sum_{\substack{l'=0\\l\neq l'}}^{v_T-1}\sum_{n_1=1}^{s_T}\sum_{n_2=1}^{s_T}K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)K_{h,1}\big(\frac{t}{T} - \frac{\lambda'+n_2}{T}\big)\\
&\quad\times\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j - X^j_{\lambda+n_1,T})(\mathbb{1}_{Y_{\lambda+n_1,T}\le y} - F_t^\star(y|\boldsymbol{x}))\Big]\mathbb{E}\Big[\prod_{j=1}^{d}K_{h,2}(x^j - X^j_{\lambda'+n_2,T})(\mathbb{1}_{Y_{\lambda'+n_2,T}\le y} - F_t^\star(y|\boldsymbol{x}))\Big].
\end{aligned}$$

Using *(C.1.3)*,

$$K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)\mathbb{E}\big[\prod_{j=1}^{d}K_{h,2}(x^j - X^j_{\lambda+n_1,T})(\mathbb{1}_{Y_{\lambda+n_1,T}\le y} - F_t^\star(y|\boldsymbol{x}))\big]$$
$$\lesssim K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big)\big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\big),$$

and

$$\mathsf{S}_{32}^{\Pi} \lesssim \frac{1}{(Th^{d+1})^2} \sum_{l=0}^{v_T-1} \sum_{\substack{l'=0 \\ l \neq l'}}^{v_T-1} \sum_{n_1=1}^{s_T} \sum_{n_2=1}^{s_T} K_{h,1}\Big(\frac{t}{T} - \frac{\lambda + n_1}{T}\Big) K_{h,1}\Big(\frac{t}{T} - \frac{\lambda' + n_2}{T}\Big) \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2.$$

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

$$\begin{aligned}
\mathsf{S}_{32}^{\Pi} &\lesssim \frac{1}{T^2 h^{2d+2}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \sum_{m=1}^{T} \sum_{\substack{m'=1 \\ |m-m'| > r_T}}^{T} K_{h,1}\Big(\frac{t}{T} - \frac{m}{T}\Big) K_{h,1}\Big(\frac{t}{T} - \frac{m'}{T}\Big) \\
&\leq \frac{1}{h^{2d}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \underbrace{\frac{1}{Th} \sum_{m=1}^{T} K_{h,1}\Big(\frac{t}{T} - \frac{m}{T}\Big)}_{\mathcal{O}(1)} \underbrace{\frac{1}{Th} \sum_{m'=1}^{T} K_{h,1}\Big(\frac{t}{T} - \frac{m'}{T}\Big)}_{\mathcal{O}(1)} \\
&\lesssim \frac{1}{h^{2d}} \Big(\frac{1}{T^{\nu} h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \\
&\lesssim \frac{1}{T^{2\nu} h^{2(d+\nu-1)}} + h^2, \qquad (34)
\end{aligned}$$

Now, comparing (29), (30), (31), (33), and (34), we get

$$\mathbb{E}[\Pi_{t,T}^2] \lesssim \frac{1}{Th^{2(d+1) - \frac{2}{p}(1-\nu)}} + \frac{1}{T^{2\nu} h^{2(d+\nu-1)}} + h^2. \qquad (35)$$

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

$$\begin{aligned}
\mathbb{E}\big[\Xi_{t,T}^2\big] = &\frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} K_{h,1}^2\Big(\frac{t}{T} - \frac{a}{T}\Big) \mathbb{E}\big[Z_{a,t,T}^2\big] \\
&+ \frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^{T} K_{h,1}\Big(\frac{t}{T} - \frac{a}{T}\Big) K_{h,1}\Big(\frac{t}{T} - \frac{b}{T}\Big) \mathbb{E}\big[Z_{a,t,T} Z_{b,t,T}\big].
\end{aligned}$$

We can further expand this as

$$\begin{aligned}
\mathbb{E}\big[\Xi_{t,T}^2\big] = &\frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} K_{h,1}^2\Big(\frac{t}{T} - \frac{a}{T}\Big) \mathbb{E}\big[Z_{a,t,T}^2\big] \\
&+ \frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^{T} K_{h,1}\Big(\frac{t}{T} - \frac{a}{T}\Big) K_{h,1}\Big(\frac{t}{T} - \frac{b}{T}\Big) \mathbb{C}\mathrm{ov}\big(Z_{a,t,T}, Z_{b,t,T}\big) \\
&+ \frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a \neq b}}^{T} K_{h,1}\Big(\frac{t}{T} - \frac{a}{T}\Big) K_{h,1}\Big(\frac{t}{T} - \frac{b}{T}\Big) \mathbb{E}\big[Z_{a,t,T}\big] \mathbb{E}\big[Z_{b,t,T}\big] \\
=: &\ \mathsf{S}_1^{\Xi} + \mathsf{S}_2^{\Xi} + \mathsf{S}_3^{\Xi}.
\end{aligned}$$

Step 3.1. Control of S_1^{Ξ}. We have

$$\begin{aligned}
\mathsf{S}_1^{\Xi} &= \frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} K_{h,1}^2\Big(\frac{t}{T} - \frac{a}{T}\Big) \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}^2(x^j - X_{a,T}^j)(\mathbb{1}_{Y_{a,T} \leq y} - F_t^{\star}(y|\boldsymbol{x}))^2\Big] \\
&\leq \frac{2C_2^d}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} K_{h,1}^2\Big(\frac{t}{T} - \frac{a}{T}\Big) \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\big|\mathbb{1}_{Y_{a,T} \leq y} - F_t^{\star}(y|\boldsymbol{x})\big|\Big].
\end{aligned}$$

Using *(C.1.3)*, we have

$$
\begin{aligned}
\mathsf{S}_1^{\Xi} &\lesssim \frac{1}{T^2 h^{2d+2}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big) \sum_{a=v_T(r_T+s_T)+1}^{T} K_{h,1}^2\big(\frac{t}{T} - \frac{a}{T}\big) \\
&\leq \frac{C_1}{Th^{2d+1}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big) \underbrace{\frac{1}{Th}\sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} \\
&\lesssim \frac{1}{Th^{2d+1}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big) \\
&\lesssim \frac{1}{Th^{2d+\nu}}.
\end{aligned} \tag{36}
$$

Step 3.2. Control of S_2^{Ξ}. Observe that

$$
\begin{aligned}
\mathsf{S}_2^{\Xi} &= \frac{1}{(Th^{d+1})^2} \sum_{a=v_T(r_T+s_T)+1}^{T} \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a\neq b}}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{b}{T}\big) \mathbb{C}\mathrm{ov}\big(Z_{a,t,T}, Z_{b,t,T}\big) \\
&= \frac{1}{(Th^{d+1})^2} \sum_{n_1=1}^{T-v_T(r_T+s_T)} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{T-v_T(r_T+s_T)} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\big),
\end{aligned}
$$

where $\lambda = v_T(r_T+s_T)$. Now, using (23), we have

$$
\begin{aligned}
&K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \big|\mathbb{C}\mathrm{ov}\big(Z_{\lambda+n_1,t,T}, Z_{\lambda+n_2,t,T}\big)\big| \\
&\qquad \lesssim K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}} \beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned}
$$

By Assumption 5,

$$
\begin{aligned}
\mathsf{S}_2^{\Xi} &\lesssim \frac{1}{T^2 h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}} \sum_{n_1=1}^{T-v_T(r_T+s_T)} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{T-v_T(r_T+s_T)} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) \\
&\qquad \times K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \beta(|n_1-n_2|)^{1-\frac{2}{p}} \\
&\leq \frac{C_1^2}{T^2 h^{2(d+1)}}\Big(\frac{1}{T^{\nu}h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}} \sum_{n_1=1}^{T-v_T(r_T+s_T)} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{T-v_T(r_T+s_T)} \beta(|n_1-n_2|)^{1-\frac{2}{p}}.
\end{aligned}
$$

Moreover, letting $k = |n_1 - n_2|$ and $w_T = T - v_T(r_T+s_T)$ yields

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

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

$$\begin{aligned}
\mathsf{S}_2^{\Xi} &\leq \frac{C_1^2 w_T}{T^2 h^{2(d+1)}} \Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}} \sum_{k=1}^{\infty} k^\zeta \beta(k)^{1-\frac{2}{p}} \\
&\lesssim \frac{1}{T h^{2(d+1)}} \Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^{\frac{2}{p}}, \quad \text{since } w_T \ll T, \\
&\lesssim \Big(\frac{1}{T^p h^{2(d+1)p}} \Big(\frac{1}{T^{2\nu} h^{2(\nu-1)}} + h^{2(d+1)}\Big)\Big)^{\frac{1}{p}} \\
&\lesssim \frac{1}{T h^{2(d+1) - \frac{2}{p}(1-\nu)}}. \qquad (37)
\end{aligned}$$

Step 3.3. Control of S_3^{Ξ}. Lastly, let us look at S_3^{Ξ}.

$$\begin{aligned}
\mathsf{S}_3^{\Xi} = \frac{1}{(Th^{d+1})^2} &\sum_{a=v_T(r_T+s_T)+1}^{T} \sum_{\substack{b=v_T(r_T+s_T)+1 \\ a\neq b}}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{b}{T}\big) \mathbb{E}\big[Z_{a,t,T}\big]\mathbb{E}\big[Z_{b,t,T}\big] \\
= \frac{1}{(Th^{d+1})^2} &\sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \mathbb{E}\big[Z_{\lambda+n_1,t,T}\big]\mathbb{E}\big[Z_{\lambda+n_2,t,T}\big] \\
= \frac{1}{(Th^{d+1})^2} &\sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \\
&\times \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X^j_{\lambda+n_1,T})(\mathbb{1}_{Y_{\lambda+n_1,T}\leq y} - F_t^\star(y|\boldsymbol{x}))\Big] \mathbb{E}\Big[\prod_{j=1}^{d} K_{h,2}(x^j - X^j_{\lambda+n_2,T})(\mathbb{1}_{Y_{\lambda+n_2,T}\leq y} - F_t^\star(y|\boldsymbol{x}))\Big].
\end{aligned}$$

Using *(C.1.3)*, for $i = 1, 2$,

$$\begin{aligned}
K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_i}{T}\big) &\mathbb{E}\big[\prod_{j=1}^{d} K_{h,2}(x^j - X^j_{\lambda+n_i,T})(\mathbb{1}_{Y_{\lambda+n_i,T}\leq y} - F_t^\star(y|\boldsymbol{x}))\big] \\
&\lesssim K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_i}{T}\big)\big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\big),
\end{aligned}$$

then,

$$\begin{aligned}
\mathsf{S}_3^{\Xi} &\lesssim \frac{1}{T^2 h^{2d+2}} \Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \sum_{n_1=1}^{w_T} \sum_{\substack{n_2=1 \\ |n_1-n_2|>0}}^{w_T} K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_1}{T}\big) K_{h,1}\big(\frac{t}{T} - \frac{\lambda+n_2}{T}\big) \\
&\leq \frac{C_1}{T h^{2d+1}} \Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \underbrace{\frac{1}{Th} \sum_{a=1}^{T} K_{h,1}\big(\frac{t}{T} - \frac{a}{T}\big)}_{\mathcal{O}(1)} \\
&\lesssim \frac{1}{T h^{2d+1}} \Big(\frac{1}{T^\nu h^{\nu-1}} + h^{d+1} + h^{d+3}\Big)^2 \\
&\lesssim \frac{1}{T h^{2(d+\nu)-1}}. \qquad (38)
\end{aligned}$$

Now, comparing (36), (37), and (38), we have

$$\mathbb{E}[\Xi_{t,T}^2] \lesssim \frac{1}{T h^{2(d+1) - \frac{2}{p}(1-\nu)}}. \qquad (39)$$

Therefore, following (28), (35), and (39), we get

$$\mathbb{E}\big[Z_{t,T}^2\big] = \mathcal{O}\Big(\frac{1}{T h^{2(d+1) - \frac{2}{p}(1-\nu)}} + \frac{1}{T^{2\nu} h^{2(d+\nu-1)}} + h^2\Big).$$

□

C.1 Proof of Theorem 1

Recall that $\pi_t^\star(\cdot|\boldsymbol{x})$ is the probability measure of the random variable $Y_{t,T}|\boldsymbol{X}_{t,T}=\boldsymbol{x}$ with conditional CDF $F_t^\star(y|\boldsymbol{x})=\mathbb{P}(Y_{t,T}\leq y|\boldsymbol{X}_{t,T}=\boldsymbol{x})$. Observe that, by the definition of W_1 given in (3),

$$\mathbb{E}[W_1(\hat{\pi}_t(\cdot|\boldsymbol{x}),\pi_t^\star(\cdot|\boldsymbol{x}))]=\int_{\mathbb{R}}\mathbb{E}\big[\big|\hat{F}_t(y|\boldsymbol{x})-F_t^\star(y|\boldsymbol{x})\big|\big]\mathrm{d}y,$$

using Fubini's theorem. Now, using Definition 4,

$$\hat{F}_t(y|\boldsymbol{x})-F_t^\star(y|\boldsymbol{x})=\frac{\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)\mathbb{1}_{Y_{a,T}\leq y}}{\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)}-F_t^\star(y|\boldsymbol{x}).$$

Then observe that

$$\hat{F}_t(y|\boldsymbol{x})-F_t^\star(y|\boldsymbol{x})=\frac{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)\big[\mathbb{1}_{Y_{a,T}\leq y}-F_t^\star(y|\boldsymbol{x})\big]}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)}. \tag{40}$$

Further, by applying Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\mathbb{E}[W_1(\hat{\pi}_t(\cdot|\boldsymbol{x}),\pi_t^\star(\cdot|\boldsymbol{x}))]&=\int\mathbb{E}\big[\big|\hat{F}_t(y|\boldsymbol{x})-F_t^\star(y|\boldsymbol{x})\big|\big]\mathrm{d}y\\
&=\int\mathbb{E}\Big[\Big|\frac{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)\big[\mathbb{1}_{Y_{a,T}\leq y}-F_t^\star(y|\boldsymbol{x})\big]}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)}\Big|\Big]\mathrm{d}y\\
&\leq\int\Big(\mathbb{E}\Big[\Big(\frac{1}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)}\Big)^2\Big]\Big)^{\frac{1}{2}}\\
&\quad\times\Big(\mathbb{E}\Big[\Big(\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)\big[\mathbb{1}_{Y_{a,T}\leq y}-F_t^\star(y|\boldsymbol{x})\big]\Big)^2\Big]\Big)^{\frac{1}{2}}\mathrm{d}y.
\end{aligned}\tag{41}$$

Let $J_{t,T}(\frac{t}{T},\boldsymbol{x})=\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)$. Using Proposition C.2, the first term in (41) becomes

$$\Big\{\mathbb{E}\Big[\Big(\frac{1}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X_{a,T}^j)}\Big)^2\Big]\Big\}^{\frac{1}{2}}=\mathcal{O}(1). \tag{42}$$

Proposition C.3 implies the second term is of order $\mathcal{O}\big(\frac{1}{T^{\frac{1}{2}}h^{d+1-\frac{1}{p}(1-\nu)}}+\frac{1}{T^\nu h^{d+\nu-1}}+h\big)$. Therefore, from (41), we have

$$\sup_{\boldsymbol{x}\in\mathcal{X},\frac{t}{T}\in I_h}\mathbb{E}\big[W_1\big(\hat{\pi}_t(\cdot|\boldsymbol{x}),\pi_t^\star(\cdot|\boldsymbol{x})\big)\big]=\mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}}h^{d+1-\frac{1}{p}(1-\nu)}}+\frac{1}{T^\nu h^{d+\nu-1}}+h\Big),$$

where $\nu=\rho\wedge 1$ and $p>2$. □

C.2 Proof of Corollary 1

Proof. **Step 1: Bounded support of the conditional laws.** Since $|Y_{t,T}|\leq M$ a.s. for all t, both $\pi_t^\star(\cdot\mid\boldsymbol{x})$ and $\hat{\pi}_t(\cdot\mid\boldsymbol{x})=\sum_{a=1}^T\omega_a\big(\frac{t}{T},\boldsymbol{x}\big)\,\delta_{Y_{a,T}}$ are supported on $[-M,M]$. Hence their quantile functions also take values in $[-M,M]$.

Step 2: One-dimensional representation of W_r and comparison with W_1. In one dimension,

$$W_r^r(\mu,\nu)=\int_0^1\big|F_\mu^{-1}(z)-F_\nu^{-1}(z)\big|^r\,\mathrm{d}z.$$

Let $\mu = \hat{\pi}_t(\cdot \mid \boldsymbol{x})$ and $\nu = \pi_t^\star(\cdot \mid \boldsymbol{x})$. For every $z \in [0,1]$,

$$\left|F_\mu^{-1}(z) - F_\nu^{-1}(z)\right| \le 2M.$$

Hence, for all $r \ge 1$,

$$\left|F_\mu^{-1}(z) - F_\nu^{-1}(z)\right|^r \le (2M)^{r-1} \left|F_\mu^{-1}(z) - F_\nu^{-1}(z)\right|.$$

Integrating over z yields

$$W_r^r(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})) \le (2M)^{r-1}\, W_1(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x}))\,. \tag{43}$$

Step 3: Take expectations and apply Theorem 1. Taking expectation of (43) and using linearity,

$$\mathbb{E}[W_r^r(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x}))] \le (2M)^{r-1}\, \mathbb{E}[W_1(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x}))]\,.$$

Taking the supremum over $\boldsymbol{x} \in \mathcal{X}$ and $\frac{t}{T} \in I_h$, and using Theorem 1,

$$\sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}[W_r^r(\hat{\pi}_t, \pi_t^\star)] = \mathcal{O}\left(\frac{1}{T^{1/2} h^{d+1-\frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h\right),$$

since $(2M)^{r-1}$ is a finite constant independent of T, h, t, and $\boldsymbol{x}$. The case $r = 2$ follows immediately. □

C.3 Proof of Corollary 2

Proof. By Lemma B.1, for any t and $\boldsymbol{x}$,

$$W_r^r(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})) \le W_1(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})) + 2^{r-1}(m_r(\hat{\pi}_t) + m_r(\pi_t^\star))\,,$$

where $m_r(\mu) := \int |y|^r\, \mu(\mathrm{d}y)$. Under the given uniform moment condition,

$$\sup_{\boldsymbol{x},\, t, T} (m_r(\hat{\pi}_t) + m_r(\pi_t^\star)) \le C_r < \infty.$$

Taking expectations and suprema over $(\boldsymbol{x}, t/T)$ yields

$$\sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}[W_r^r(\hat{\pi}_t, \pi_t^\star)] \le \sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}[W_1(\hat{\pi}_t, \pi_t^\star)] + 2^{r-1} C_r.$$

Finally, applying Theorem 1 completes the proof. □

C.4 Proof of Corollary 3

Proof. Fix $\varepsilon > 0$. By the given uniform integrability condition, there exists $R = R(\varepsilon) > 0$ such that

$$\sup_{\boldsymbol{x}, t, T} \int_{\{|y| > R\}} |y|^r\, \pi_t^\star(\mathrm{d}y \mid \boldsymbol{x}) \le \varepsilon, \qquad \sup_{\boldsymbol{x}, t, T} \int_{\{|y| > R\}} |y|^r\, \hat{\pi}_t(\mathrm{d}y \mid \boldsymbol{x}) \le \varepsilon.$$

Apply Lemma B.2 with this R to $\mu = \hat{\pi}_t(\cdot \mid \boldsymbol{x})$ and $\nu = \pi_t^\star(\cdot \mid \boldsymbol{x})$:

$$W_r^r\big(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})\big) \le (2R)^{r-1} W_1\big(\hat{\pi}_t(\cdot \mid \boldsymbol{x}), \pi_t^\star(\cdot \mid \boldsymbol{x})\big) + 2^r \varepsilon,$$

uniformly in $\boldsymbol{x}, t, T$.

Taking expectations and then the supremum over $\boldsymbol{x} \in \mathcal{X}$ and $\frac{t}{T} \in I_h$ yields

$$\sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}\Big[W_r^r(\hat{\pi}_t, \pi_t^\star)\Big] \le (2R)^{r-1} \sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}\Big[W_1(\hat{\pi}_t, \pi_t^\star)\Big] + 2^r \varepsilon.$$

By Theorem 1, the first term on the right-hand side goes to zero as $T \to \infty$, for any fixed R. Hence

$$\limsup_{T \to \infty} \sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}\Big[W_r^r(\hat{\pi}_t, \pi_t^\star)\Big] \le 2^r \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the conclusion follows. □

C.5 Proof of Proposition 1

Proof. Fix $\boldsymbol{x} \in \mathcal{X}$ and t, and denote

$$\mu = \hat{\pi}_t(\cdot \mid \boldsymbol{x}), \qquad \nu = \pi^\star_t(\cdot \mid \boldsymbol{x}).$$

By Lemma B.2, for any $R > 0$,

$$W_r^r(\mu,\nu) \le (2R)^{r-1} W_1(\mu,\nu) + 2^{r-1}\left(\int_{\{|y|>R\}} |y|^r\,\mu(\mathrm{d}y) + \int_{\{|y|>R\}} |y|^r\,\nu(\mathrm{d}y)\right). \tag{44}$$

Step 1: Tail control via the s-th moment. By Markov's inequality and the assumption (5), for $R > 0$,

$$\int_{\{|y|>R\}} |y|^r\,\mu(\mathrm{d}y) \le \frac{1}{R^{s-r}} \int_{\mathbb{R}} |y|^s\,\mu(\mathrm{d}y) \le \frac{C_s}{R^{s-r}},$$

and the same bound holds for ν, i.e.

$$\int_{\{|y|>R\}} |y|^r\,\nu(\mathrm{d}y) \le \frac{C_s}{R^{s-r}},$$

for some constant C_s independent of $(\boldsymbol{x}, t, T)$. Hence, from (44),

$$W_r^r(\mu,\nu) \le (2R)^{r-1} W_1(\mu,\nu) + \frac{C}{R^{s-r}}, \tag{45}$$

for a suitable constant $C > 0$.

Step 2: Take expectations and suprema. Taking expectations in (45) and the supremum over $\boldsymbol{x} \in \mathcal{X}$ and $\frac{t}{T} \in I_h$ yields

$$\sup_{\boldsymbol{x},\, \frac{t}{T}\in I_h} \mathbb{E}\Big[W_r^r(\hat{\pi}_t, \pi^\star_t)\Big] \le (2R)^{r-1} A_T + \frac{C}{R^{s-r}},$$

where A_T is defined in (6).

Thus, for each fixed T, we are led to control

$$f_T(R) := a_T R^{r-1} + b R^{-(s-r)},$$

with $a_T := 2^{r-1} A_T$ and $b := C$.

Step 3: Optimizing the choice of R. We choose $R = R_T$ to minimize the order of $f_T(R)$. Ignoring constants and differentiating formally, set

$$\frac{\mathrm{d}}{\mathrm{d}R}\big(a_T R^{r-1} + bR^{-(s-r)}\big) = a_T(r-1)R^{r-2} - b(s-r)R^{-(s-r)-1} = 0.$$

Solving for R gives

$$a_T(r-1)R^{r-2+s-r+1} = b(s-r) \quad\Longrightarrow\quad R^{s-1} = \frac{b(s-r)}{a_T(r-1)} \asymp \frac{1}{A_T},$$

so that

$$R_T \asymp A_T^{-1/(s-1)}.$$

Substituting R_T back into the two terms,

$$a_T R_T^{r-1} \;\asymp\; A_T \cdot A_T^{-(r-1)/(s-1)} = A_T^{(s-r)/(s-1)},$$

and

$$b R_T^{-(s-r)} \;\asymp\; A_T^{(s-r)/(s-1)}.$$

Hence, with this choice of R_T,

$$f_T(R_T) \lesssim A_T^{(s-r)/(s-1)}.$$

Therefore, there exists a constant $C > 0$ such that, for all sufficiently large T,

$$\sup_{\boldsymbol{x},\, \frac{t}{T} \in I_h} \mathbb{E}\Big[W_r^r(\hat{\pi}_t, \pi_t^\star)\Big] \leq C\, A_T^{(s-r)/(s-1)},$$

which proves (7). The specialization to $r = 2$ is immediate. □

C.6 Proof of Proposition 2

Observe that

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

In the last equality, we use duality formula of Kantorovich-Rubinstein distance (see Remark 6.5 in Villani (2009)), where $\mathcal{F}$ is the set of all continuous functions satisfying Lipschitz condition $\|f\|_{Lip} \leq 1$, i.e., $\sup_{y \neq y'} \frac{|f(y)-f(y')|}{|y-y'|} \leq 1$. Hence,

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

This finishes the proof. □

C.7 Proof of Proposition 3

If $h \asymp T^{-\xi}$, then directly from Theorem 1, for $\nu = \rho \wedge 1$, we get

$$\begin{aligned}
\mathbb{E}\big[W_1\big(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x})\big)\big] &\lesssim \frac{1}{T^{\frac{1}{2}} h^{(d+1) - \frac{1}{p}(1-\nu)}} + \frac{1}{T^\nu h^{d+\nu-1}} + h \\
&\lesssim \frac{1}{T^{\frac{1}{2}} T^{-\xi((d+1) - \frac{1}{p}(1-\nu))}} + \frac{1}{T^\nu T^{-\xi(d+\nu-1)}} + \frac{1}{T^\xi}.
\end{aligned}$$

Note that, as $T \to \infty$, the third component goes to zero for any $\xi > 0$. The second component converges to zero when $\xi < \frac{\nu}{d+\nu-1}$, which suggests that $\xi < \frac{\nu}{d+1}$. Lastly, the first component approaches zero if $\xi < \frac{1}{2(d+1-\frac{1}{p}(1-\nu))}$, which further implies that $\xi < \frac{1}{2(d+1)}$ since $\nu = \rho \wedge 1$ and $p > 2$. Therefore, for $h \asymp T^{-\xi}$, $\mathbb{E}\big[W_1\big(\hat{\pi}_t(\cdot|\boldsymbol{x}), \pi_t^\star(\cdot|\boldsymbol{x})\big)\big]$ converges to zero if,

$$\xi < \begin{cases} \frac{\nu}{d+1} & \text{if } \nu < \frac{1}{2}, \\ \frac{1}{2(d+1)} & \text{otherwise.} \end{cases}$$

As a consequence, $\xi < \frac{\frac{1}{2} \wedge \nu}{d+1}$. □

C.8 Proof of Theorem 2

Observe that using (9) and by Fubini's theorem, we have

$$\mathbb{E}[SW_1(\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}), \boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x}))] = \int_{\mathbb{S}^{q-1}} \mathbb{E}\big[W_1(\boldsymbol{\theta}_\# \hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}), \boldsymbol{\theta}_\# \boldsymbol{\pi}_t^\star(\cdot|\boldsymbol{x}))\big] \sigma_{q-1}(\mathrm{d}\boldsymbol{\theta}).$$

On the other hand,

$$\mathbb{E}\big[W_1(\boldsymbol{\theta}_{\#}\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}),\boldsymbol{\theta}_{\#}\boldsymbol{\pi}^\star_t(\cdot|\boldsymbol{x}))\big]=\mathbb{E}\Big[\int_{\mathbb{R}}\big|\hat{F}_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})\big|\mathrm{d}y\Big]=\int_{\mathbb{R}}\mathbb{E}\big[\big|\hat{F}_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})\big|\big]\mathrm{d}y.$$

Using the weights defined in Definition 3 and (10),

$$\hat{F}_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})=\frac{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})\big[\mathbb{1}_{\boldsymbol{\theta}^\top \boldsymbol{Y}_{a,T}\le y}-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})\big]}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})}.$$

Further, by applying Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
&\mathbb{E}\big[W_1(\boldsymbol{\theta}_{\#}\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}),\boldsymbol{\theta}_{\#}\boldsymbol{\pi}^\star_t(\cdot|\boldsymbol{x}))\big]\\
&=\int_{\mathbb{R}}\mathbb{E}\Big[\Big|\frac{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})\big[\mathbb{1}_{\boldsymbol{\theta}^\top \boldsymbol{Y}_{a,T}\le y}-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})\big]}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})}\Big|\Big]\mathrm{d}y\\
&\le\int_{\mathbb{R}}\Big(\mathbb{E}\Big[\Big(\frac{1}{\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})}\Big)^2\Big]\Big)^{\frac{1}{2}}\\
&\qquad\times\Big(\mathbb{E}\Big[\Big(\frac{1}{Th^{d+1}}\sum_{a=1}^T K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})\big[\mathbb{1}_{\boldsymbol{\theta}^\top \boldsymbol{Y}_{a,T}\le y}-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})\big]\Big)^2\Big]\Big)^{\frac{1}{2}}\mathrm{d}y.
\end{aligned}\tag{46}$$

Note that from Proposition C.2, the first term in (46) is $\mathcal{O}(1)$. Moreover, it can be observed that inequality (46) is similar to inequality (41), hence, we again use Bernstein's big-block and small-block procedure and consider (20) with $Z_{a,t,T}=\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})\big[\mathbb{1}_{\boldsymbol{\theta}^\top \boldsymbol{Y}_{a,T}\le y}-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x})\big]$. Furthermore, Assumption 7 and *(C.1.3)* give

$$K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\mathbb{E}\Big[\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})(\mathbb{1}_{\boldsymbol{\theta}^\top \boldsymbol{Y}_{a,T}\le y}-F^\star_{t,\boldsymbol{\theta}}(y|\boldsymbol{x}))\Big]\lesssim K_{h,1}\big(\frac{t}{T}-\frac{a}{T}\big)\big(\frac{1}{T^\nu h^{\nu-1}}+h^{d+1}+h^{d+3}\big).$$

The rest of the proof follows directly from the proof of Theorem 1. By Proposition C.3, we have

$$\mathbb{E}\big[\big(Z_{t,T}\big)^2\big]\lesssim\frac{1}{Th^{2(d+1)-\frac{2}{p}(1-\nu)}}+\frac{1}{T^{2\nu}h^{2(d+\nu-1)}}+h^2.\tag{47}$$

From (46), and incorporating (42) and (47), we have

$$\sup_{\boldsymbol{x}\in\mathcal{X},\,\frac{t}{T}\in I_h}\mathbb{E}\big[W_1(\boldsymbol{\theta}_{\#}\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}),\boldsymbol{\theta}_{\#}\boldsymbol{\pi}^\star_t(\cdot|\boldsymbol{x}))\big]=\mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}}h^{d+1-\frac{1}{p}(1-\nu)}}+\frac{1}{T^\nu h^{d+\nu-1}}+h\Big).$$

Therefore,

$$\sup_{\boldsymbol{x}\in\mathcal{X},\,\frac{t}{T}\in I_h}\mathbb{E}[SW_1(\hat{\boldsymbol{\pi}}_t(\cdot|\boldsymbol{x}),\boldsymbol{\pi}^\star_t(\cdot|\boldsymbol{x}))]=\mathcal{O}\Big(\frac{1}{T^{\frac{1}{2}}h^{d+1-\frac{1}{p}(1-\nu)}}+\frac{1}{T^\nu h^{d+\nu-1}}+h\Big),$$

where $\nu=\rho\wedge 1$. □

D TECHNICAL LEMMAS

Lemma D.1 *Let Assumption 2 hold, then*

(D.1.1) $$\Big|\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})-\prod_{j=1}^d K_{h,2}\big(x^j-X^j_a\big(\tfrac{a}{T}\big)\big)\Big|\le C_2^{d-1}\sqrt{d}\sum_{j=1}^d\big|K_{h,2}(x^j-X^j_{a,T})-K_{h,2}\big(x^j-X^j_a\big(\tfrac{a}{T}\big)\big)\big|.$$

(D.1.2) $$\Big|\prod_{j=1}^d K^2_{h,2}(x^j-X^j_{a,T})-\prod_{j=1}^d K^2_{h,2}\big(x^j-X^j_a\big(\tfrac{a}{T}\big)\big)\Big|\le C_2^{2d-2}\sqrt{d}\sum_{j=1}^d\big|K_{h,2}(x^j-X^j_{a,T})-K_{h,2}\big(x^j-X^j_a\big(\tfrac{a}{T}\big)\big)\big|.$$

(D.1.3) for $p\ge 2$, $\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})\Big|^p\Big]\le C_2^{d(p-1)}\mathbb{E}\Big[\Big|\prod_{j=1}^d K_{h,2}(x^j-X^j_{a,T})\Big|\Big]$.

Proof. *Proof of (D.1.1).* Let $g^j = K_{h,2}(x^j - X^j_{a,T})$, $\widetilde{g}^j = K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)$, and $G(g^1,\ldots,g^d) = \prod_{j=1}^d g^j$. The gradient of $G(g^1,\ldots,g^d)$ can be written as

$$\nabla G(g^1,\ldots,g^d) = \begin{bmatrix} \frac{\partial G(g^1,\ldots,g^d)}{\partial g^1} \\ \frac{\partial G(g^1,\ldots,g^d)}{\partial g^2} \\ \vdots \\ \frac{\partial G(g^1,\ldots,g^d)}{\partial g^d} \end{bmatrix} = \begin{bmatrix} \prod_{j=2}^d g^j \\ \prod_{j=1;j\neq 2}^d g^j \\ \vdots \\ \prod_{j=1}^{d-1} g^j \end{bmatrix}.$$

By Assumption 2, K_2 is bounded by C_2,

$$\begin{aligned}\|\nabla G(g^1,\ldots,g^d)\| &= \sqrt{\Big(\prod_{j=2}^d g^j\Big)^2 + \Big(\prod_{j=1;j\neq 2}^d g^j\Big)^2 + \cdots + \Big(\prod_{j=1}^{d-1} g^j\Big)^2} \\ &\le \sqrt{(C_2^{d-1})^2 + \cdots + (C_2^{d-1})^2} = \sqrt{d(C_2^{d-1})^2} = C_2^{d-1}\sqrt{d}.\end{aligned}$$

Now,

$$\begin{aligned}|G(g^1,\ldots,g^d) - G(\widetilde{g}^1,\ldots,\widetilde{g}^j)| &\le C_2^{d-1}\sqrt{d}\|(g^1,\ldots,g^d) - (\widetilde{g}^1,\ldots,\widetilde{g}^j)\|_2 \\ &= C_2^{d-1}\sqrt{d}\sqrt{\sum_{j=1}^d \big(K_{h,2}(x^j - X^j_{a,T}) - K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\big)^2} \\ &\le C_2^{d-1}\sqrt{d}\sum_{j=1}^d \big|K_{h,2}(x^j - X^j_{a,T}) - K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\big|,\end{aligned}$$

since for d-dimensional vector $\boldsymbol{z}$, $\|\boldsymbol{z}\|_2 \le \|\boldsymbol{z}\|_1$. Then,

$$\Big|\prod_{j=1}^d K_{h,2}(x^j - X^j_{a,T}) - \prod_{j=1}^d K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\Big| \le C_2^{d-1}\sqrt{d}\sum_{j=1}^d \big|K_{h,2}(x^j - X^j_{a,T}) - K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\big|.$$

Proof of (D.1.2). Let $g^j = K^2_{h,2}(x^j - X^j_{a,T})$, $\widetilde{g}^j = K^2_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)$, and $G(g^1,\ldots,g^d) = \prod_{j=1}^d g^j$. Using the gradient of $G(g^1,\ldots,g^d)$ given in (D.1.1) and noting that $K_2^2(\cdot)$ is bounded by C_2^2,

$$\begin{aligned}\|\nabla G(g^1,\ldots,g^d)\| &= \sqrt{\Big(\prod_{j=2}^d g^j\Big)^2 + \Big(\prod_{j=1;j\neq 2}^d g^j\Big)^2 + \cdots + \Big(\prod_{j=1}^{d-1} g^j\Big)^2} \\ &\le \sqrt{(C_2^{2d-2})^2 + \cdots + (C_2^{2d-2})^2} = \sqrt{d(C_2^{2d-2})^2} = C_2^{2d-2}\sqrt{d}.\end{aligned}$$

Observe that,

$$\begin{aligned}|G(g^1,\ldots,g^d) - G(\widetilde{g}^1,\ldots,\widetilde{g}^j)| &\le C_2^{2d-2}d^{\frac{1}{2}}\|(g^1,\ldots,g^d) - (\widetilde{g}^1,\ldots,\widetilde{g}^j)\|_2 \\ &= C_2^{2d-2}\sqrt{d}\sqrt{\sum_{j=1}^d \big(K_{h,2}(x^j - X^j_{a,T}) - K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\big)^2} \\ &\le C_2^{2d-2}\sqrt{d}\sum_{j=1}^d \big|K_{h,2}(x^j - X^j_{a,T}) - K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\big|,\end{aligned}$$

Finally,

$$\Big|\prod_{j=1}^d K_{h,2}(x^j - X^j_{a,T}) - \prod_{j=1}^d K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\Big| \le C_2^{2d-2}\sqrt{d}\sum_{j=1}^d \big|K_{h,2}(x^j - X^j_{a,T}) - K_{h,2}\big(x^j - X^j_a\big(\frac{a}{T}\big)\big)\big|.$$

Proof of (D.1.3). We use the boundedness of K_2, so

$$\begin{aligned}\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big|^p\Big] &\leq \mathbb{E}\Big[\max_{j=1,\ldots,d}\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big|^{p-1}\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big|\Big] \\ &\leq C_2^{d(p-1)}\mathbb{E}\Big[\Big|\prod_{j=1}^{d} K_{h,2}(x^j - X_{a,T}^j)\Big|\Big].\end{aligned}$$

□

Lemma D.2 *For $l \neq l'$, $\beta(\sigma(\boldsymbol{X}_{l,T}), \sigma(\boldsymbol{X}_{l',T})) \leq \beta(|l-l'|)$, where $\sigma(X)$ denotes the σ-algebra generated by X.*

Proof. Let us start the proof by first considering the case $l > l'$, that is

$$\begin{aligned}\beta(\sigma(\boldsymbol{X}_{l,T}), \sigma(\boldsymbol{X}_{l',T})) &\leq \beta(\sigma(\boldsymbol{X}_{s,T}, s \geq l), \sigma(\boldsymbol{X}_{s,T}, s \leq l')) \\ &= \beta(\sigma(\boldsymbol{X}_{s,T}, s \leq l'), \sigma(\boldsymbol{X}_{s,T}, l \leq s)) \\ &\leq \sup_t \beta(\sigma(\boldsymbol{X}_{s,T}, s \leq t), \sigma(\boldsymbol{X}_{s,T}, t + l - l' \leq s \leq T)), \quad (t = l') \\ &\leq \sup_{t,T: t \leq T - |l-l'|} \beta(\sigma(\boldsymbol{X}_{s,T}, s \leq t), \sigma(\boldsymbol{X}_{s,T}, t + |l - l'| \leq s \leq T)) \\ &= \beta(|l-l'|).\end{aligned}$$

The last inequality holds since $t \leq T - |l - l'|$. For $l' > l$, observe that

$$\begin{aligned}\beta(\sigma(\boldsymbol{X}_{l,T}), \sigma(\boldsymbol{X}_{l',T})) &\leq \beta(\sigma(\boldsymbol{X}_{s,T}, s \geq l'), \sigma(\boldsymbol{X}_{s,T}, s \leq l)) \\ &= \beta(\sigma(\boldsymbol{X}_{s,T}, s \leq l), \sigma(\boldsymbol{X}_{s,T}, l' \leq s)) \\ &\leq \sup_t \beta(\sigma(\boldsymbol{X}_{s,T}, s \leq t), \sigma(\boldsymbol{X}_{s,T}, t + l' - l \leq s \leq T)), \quad (t = l) \\ &\leq \sup_{t,T: t \leq T - |l'-l|} \beta(\sigma(\boldsymbol{X}_{s,T}, s \leq t), \sigma(\boldsymbol{X}_{s,T}, t + |l' - l| \leq s \leq T)) \\ &= \beta(|l-l'|).\end{aligned}$$

□

Lemma D.3 (Davydov (1973)) *Suppose that X and Y are random variables which are $\mathscr{G}$ and $\mathscr{H}$-measurable, respectively, and that $\mathbb{E}[|X|^p] < \infty$, $\mathbb{E}[|Y|^{p'}] < \infty$, where $p, p' > 1$, $p^{-1} + p'^{-1} < 1$. Then*

$$|\mathbb{C}ov(X,Y)| \leq 8\|X\|_{L_p}\|Y\|_{L_p'}[\beta(\mathscr{G}, \mathscr{H})]^{1-p^{-1}-p'^{-1}}.$$

Lemma D.4 (Vogt (2012)) *Suppose K fulfills Assumption 2 and let $g : [0,1] \times \mathbb{R}^d \to \mathbb{R}$, $(u,x) \mapsto g(u,x)$ be continuously differentiable wrt u. Then for any compact set $S \subset \mathbb{R}^d$,*

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