Supplementary Material for "Nonparametric Estimation of the Continuous Treatment Effect with Measurement Error"

Wei Huang
School of Mathematics and Statistics,
University of Melbourne, Australia
and
Zheng Zhang
Center for Applied Statistics, Institute of Statistics & Big Data, Renmin University of China

A. Side results

A.1. Some Discussions

A.1.1. Two-step Kernel Estimator of $\pi_0(t, \boldsymbol{x})$

An alternative approach to estimating $\pi_0(t, \boldsymbol{x})$ is to separately estimate both the marginal density $f_T(t)$ and the conditional density $f_{T|X}(t|\boldsymbol{x})$ based on deconvolutional kernels:

$$\widehat{f}_T(t) := (Nh_1)^{-1} \sum_{i=1}^{N} L_U\{(t - S_i)/h_1\},$$

$$\widehat{f}_{T|X}(t|\boldsymbol{x}) := \frac{\sum_{i=1}^{N} h_2^{-1} L_U\{(t - S_i)/h_2\} \prod_{k=3}^{r+2} L\{(X_{ik} - x_k)/h_k\}}{\sum_{i=1}^{N} \prod_{k=1}^{r} L\{(X_{ik} - x_k)/h_k\}},$$

and construct the two-step kernel estimator $\widehat{\pi}_{kernel}(t, \boldsymbol{x}) = \widehat{f}_T(t)/\widehat{f}_{T|X}(t|\boldsymbol{x})$, which is sensitive to low denominator values. Moreover, the deconvolution estimates of $\widehat{f}_T(t)$ and $\widehat{f}_{T|X}(t|\boldsymbol{x})$ may take negative values, causing instability of the estimated $\mu(t)$. By contrast, our proposed estimator $\widehat{\pi}(t, \boldsymbol{x})$ is always positive and satisfies the following empirical moment restriction:

$$\frac{\sum_{i=1}^{N} \widehat{\pi}(t, \boldsymbol{X}_i) L_U\{(t - S_i)/h_0\} u_K(\boldsymbol{X}_i)}{\sum_{i=1}^{N} L_U\{(t - S_i)/h_0\}} = \frac{1}{N} \sum_{i=1}^{N} u_K(\boldsymbol{X}_i),$$

which improves the robustness of the estimation results. Finally, the number of tuning parameters in $\widehat{\pi}$ (i.e. (h_0, K)) is lower than that in $\widehat{\pi}_{kernel}$ (i.e. $(h_1, h_2, ..., h_{r+2})$), which simplifies the computation in practice. Figure 5 depicts an example of the instability of $\widehat{\pi}_{kernel}(t, \boldsymbol{x})$. There, we generate a sample of size 250 from the following

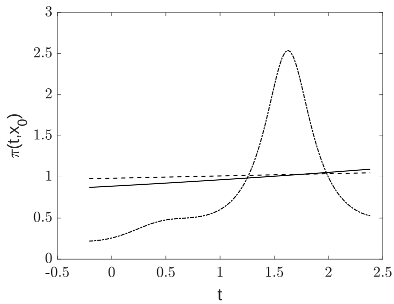

Fig. 5: Plots of the benchmark estimator $\widehat{\pi}_0(t, x_0)$ (solid line), our estimator $\widehat{\pi}(t, x_0)$ (dashed line), and the two-step kernel estimator $\widehat{\pi}_{kernel}(t, x_0)$ (dash-dotted line).

model:

$$X = \sum_{j=1}^{2} 0.3\xi_{x,j}\ , \ S = T + U \text{ and } Y^*(t) = \frac{\exp(-6+6t)}{1+\exp(-6+6t)} + X + \xi_y\ ,$$

where $T = 1 + X^2 + \xi_t$, U is a Laplace random variable with mean 0 and $\mathrm{var}(U)/\mathrm{var}(T) = 0.25$, ξ_1, ξ_2 are i.i.d. uniform random variables supported on $[0,1]$, and ξ_t and ξ_y are independent standard normal random variables. We estimate $\pi_0(t, x_0)$, where $x_0 = 0.0128$, using our proposed estimator and the two-step deconvolution kernel estimator. The tuning parameters are the theoretically optimal ones that minimise the integrated mean squared error $\int \{\widehat{\mu}(t) - \mu(t)\}^2\, dt$. The benchmark estimator

$$\widehat{\pi}_0(t, x_0) := \frac{\int f_N(t-u)\widehat{f}_{X^2}(u)\, du}{f_{N|x_0}(t)}\ ,$$

where f_N and $f_{N|x_0}$ are respectively the densities of the normal distribution $N(1,1)$ and $N(1+x_0^2, 1)$ and $\widehat{f}_{X^2}$ is the kernel density estimator of X^2, calculated from a sample of (T, X, Y) of size 10000.

A.1.2. Results of Ai et al. (2021 a)

Without measurement error, Ai et al. (2021) proposed estimating $\pi_0(\cdot,\cdot)$ based on the moment restriction (9) and maximum of entropy:

$$\begin{cases} \max\left\{-\sum_{i=1}^{N}\pi_i\log\pi_i\right\} \\ \text{s.t. } \frac{1}{N}\sum_{i=1}^{N}\pi_i v_{K_1}(T_i)u_{K_2}(\boldsymbol{X}_i)^\top = \left\{\frac{1}{N}\sum_{i=1}^{N}v_{K_1}(T_i)\right\}\left\{\frac{1}{N}\sum_{j=1}^{N}u_{K_2}(\boldsymbol{X}_j)^\top\right\}. \end{cases}$$

The dual solution to above maximisation problem is given by

$$\widetilde{\pi}(t,\boldsymbol{x}) = \tilde{\rho}'\left(v_{K_1}^\top(t)\widetilde{\Lambda}_{K_1\times K_2}u_{K_2}(\boldsymbol{x})\right) \text{ for } (t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}, \tag{A.1}$$

where $\tilde{\rho}(v) \equiv= -\exp(-v-1)$, $v_{K_1}(t)$ is another sieve basis function with dimension K_1 that can approximate any suitable function $v(t)$ and $\widetilde{\Lambda}_{K_1\times K_2}$ is a $K_1\times K_2$ matrix that maximises the following concave objective function $\widetilde{G}_{K_1\times K_2}(\Lambda)$, that is, $\widetilde{\Lambda}_{K_1\times K_2} = \arg\max_\Lambda \widetilde{G}_{K_1\times K_2}(\Lambda)$, where

$$\widetilde{G}_{K_1\times K_2}(\Lambda) := \frac{1}{N}\sum_{i=1}^{N}\tilde{\rho}\left\{v_{K_1}(T_i)^\top\Lambda u_{K_2}(\boldsymbol{X}_i)\right\} - \left\{\frac{1}{N}\sum_{i=1}^{N}v_{K_1}(T_i)\right\}^\top\Lambda\left\{\frac{1}{N}\sum_{j=1}^{N}u_{K_2}(\boldsymbol{X}_j)\right\}.$$

Ai et al. (2021 a) then estimate $\mu(t)$ by regressing the $\widetilde{\pi}(T_i,\boldsymbol{X}_i)Y_i$'s on the T_i's. The convergence rates of $\widetilde{\pi}(\cdot,\cdot)$ to $\pi_0(\cdot,\cdot)$ under the L_∞ and $L_2(dF_{T,X})$ distances are found. However, the appearance of the measurement error in the treatment observation makes the $\widetilde{\pi}(T_i,\boldsymbol{X}_i)$'s uncomputable. To the best of our knowledge, ours is the first study to address this problem. Moreover, the concave function $\widetilde{\rho}$ in Ai et al.'s (2021) estimator only corresponds to exponential tilting due to certain technical difficulties, namely, that the solution relies on the differential invariant of the exponential function (i.e. $\widetilde{\rho}''(v) = -\widetilde{\rho}'(v)$), while ours has a broader interpretation.

A.2. Proof of Theorem 3.1

We first prove that for every fixed $t\in\mathcal{T}$ and any integrable function $u(\boldsymbol{X})$, $\mathbb{E}\{\pi(t,\boldsymbol{X})u(\boldsymbol{X})|T=t\} = \mathbb{E}\{u(\boldsymbol{X})\}$ holds if and only if $\pi(t,\boldsymbol{X}) = \pi_0(t,\boldsymbol{X})$ a.s.. The sufficient part is obvious and we here show the necessary part. Since for all $t\in\mathcal{T}$ and any integrable function $u(\boldsymbol{X})$, we have $\mathbb{E}\{\pi(t,\boldsymbol{X})u(\boldsymbol{X})|T=t\} = \mathbb{E}\{u(\boldsymbol{X})\}$, comparing to (11), we see that

$$\mathbb{E}\left[\{\pi(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}u(\boldsymbol{X})|T=t\right] = 0$$

for all $t\in\mathcal{T}$ and any integrable function $u(\boldsymbol{X})$. Taking $u(\boldsymbol{X}) = \exp(a^\top\boldsymbol{X})$ for $a\in\mathbb{R}^r$, we have

$$\mathbb{E}\left[\{\pi(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}\exp(a^\top\boldsymbol{X})|T=t\right] = 0$$

for all $a \in \mathbb{R}^r$. Thus, according to the uniqueness of Laplace transform, we have that $\pi(t, \cdot) = \pi_0(t, \cdot)$ a.s..

Next, we show that

$$\lim_{h_0 \to 0} \frac{\mathbb{E}\left[\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) L_U\{(t-S)/h_0\}\right]}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]} = \mathbb{E}\{\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) | T = t\}\,.$$

The results shall then follows. Note that

$$\begin{aligned}
&\lim_{h_0 \to 0} \frac{\mathbb{E}\left[\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) L_U\{(t-S)/h_0\}\right]}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]} \\
=& \lim_{h_0 \to 0} \frac{\mathbb{E}\Big(\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) \mathbb{E}[L_U\{(t-S)/h_0\} | T, \boldsymbol{X}]\Big)}{\mathbb{E}\Big(\mathbb{E}[L_U\{(t-S)/h_0\} | T, \boldsymbol{X}]\Big)} \\
=& \lim_{h_0 \to 0} \frac{\mathbb{E}\Big(\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) \mathbb{E}[L_U\{(t-S)/h_0\} | T]\Big)}{\mathbb{E}\Big(\mathbb{E}[L_U\{(t-S)/h_0\} | T]\Big)} \quad (S \perp \boldsymbol{X} | T) \\
=& \lim_{h_0 \to 0} \frac{h_0^{-1} \mathbb{E}[\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) L\{(t-T)/h_0\}]}{h_0^{-1} \mathbb{E}[L\{(t-T)/h_0\}]} \quad (\text{by } (8))\,.
\end{aligned}$$

For the numerator, we have

$$\begin{aligned}
&\lim_{h_0 \to 0} h_0^{-1} \mathbb{E}[\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) L\{(t-T)/h_0\}] \\
=& \lim_{h_0 \to 0} h_0^{-1} \int \int \pi(t, \boldsymbol{x}) u(\boldsymbol{x}) L\{(t-t^{'})/h_0\} f_{T,\boldsymbol{X}}(t^{'}, \boldsymbol{x})\, dt^{'}\, d\boldsymbol{x} \\
=& - \lim_{h_0 \to 0} \int \int \pi(t, \boldsymbol{x}) u(\boldsymbol{x}) L(z) f_{T,\boldsymbol{X}}(t - z h_0, \boldsymbol{x})\, dz\, d\boldsymbol{x} \\
=& - \int \int \pi(t, \boldsymbol{x}) u(\boldsymbol{x}) L(z) f_{T,\boldsymbol{X}}(t, \boldsymbol{x})\, dz\, d\boldsymbol{x} \\
=& \mathbb{E}\{\pi(t, \boldsymbol{X}) u(\boldsymbol{X}) | T = t\} \cdot f_T(t)\,.
\end{aligned}$$

Similarly, we have

$$\lim_{h_0 \to 0} h_0^{-1} \mathbb{E}[L\{(t-T)/h_0\}] = f_T(t)\,.$$

The results then follows.

A.3. Dual solution of (16)

We derive the dual of the constraint maximization problem (16) using the methodology introduced in Tseng and Bertsekas (1991). Define

$$b_K := \frac{1}{N}\sum_{i=1}^{N} u_K(\boldsymbol{X}_i)\,, \quad L_{t,i} := \frac{L_U\{(t-S_i)/h_0\}}{\sum_{i=1}^{N} L_U\{(t-S_i)/h_0\}}\,,$$

$D_i(v) := D(v/L_{t,i})$ and $w_{t,i} = L_{t,i}\pi_i$ for $i = 1, \ldots, N$. Moreover, let $\boldsymbol{w}_t := (w_{t,1}, \ldots, w_{t,N})^\top$, $M_{K\times N} := (u_K(\boldsymbol{X}_1), \ldots, u_K(\boldsymbol{X}_N)) \in \mathbb{R}^{K\times N}$ and $F(\boldsymbol{w}_t) := \sum_{i=1}^{N} L_{t,i} D_i(w_{t,i})$. Then we can rewrite (16) as

$$\begin{cases} \min_{\boldsymbol{w}_t} F(\boldsymbol{w}_t) \\ \text{subject to } M_{K\times N}\cdot \boldsymbol{w}_t = b_K\,. \end{cases} \tag{A.2}$$

We define the conjugate convex function (Tseng and Bertsekas, 1991) of F to be

$$\begin{aligned} F^*(\boldsymbol{z}) &= \sup_{\boldsymbol{w}_t} \sum_{i=1}^{N} \{z_i w_{t,i} - L_{t,i} D_i(w_{t,i})\} \\ &= \sup_{\{\pi_i\}_{i=1}^N} \sum_{i=1}^{N} L_{t,i}\{z_i\pi_i - D(\pi_i)\} \\ &= \sum_{i=1}^{N} L_{t,i}\{z_i\pi_i^* - D(\pi_i^*)\}\,, \end{aligned}$$

where the π_i^*'s satisfy the first order conditions:

$$z_i = D'(\pi_i^*) \Rightarrow \pi_i^* = (D')^{-1}(z_i)\,, \quad i = 1, \ldots, N\,.$$

By defining $\rho(-z) := D\{(D')^{-1}(z)\} - z\cdot(D')^{-1}(z)$, we have

$$F^*(\boldsymbol{z}) = -\sum_{i=1}^{N} L_{t,i}\rho(-z_i)\,.$$

By Tseng and Bertsekas (1991), the dual problem of (A.2) is

$$\begin{aligned} \max_{\lambda\in\mathbb{R}^K}\{-F^*(\lambda^\top M_{K\times N}) + \lambda^\top b_K\} &= \max_{\lambda\in\mathbb{R}^K}\left[\sum_{i=1}^{N} L_{t,i}\rho\{-\lambda^\top u_K(\boldsymbol{X}_i)\} + \lambda^\top b_K\right] \\ &= \max_{\lambda\in\mathbb{R}^K}\left[\sum_{i=1}^{N} L_{t,i}\rho\{\lambda^\top u_K(\boldsymbol{X}_i)\} - \lambda^\top b_K\right] \\ &= \max_{\lambda\in\mathbb{R}^K} \widehat{G}_t(\lambda)\,. \end{aligned}$$

A.4. Undersmoothing pointwise confidence band

A.4.1. Methodology

In this section, we introduce how to construct an undersmoothing pointwise confidence band for our proposed estimator $\widehat{\mu}(t)$, $t \in \mathcal{T}$. From Theorem 4.2 and 4.4, we can see that if both h and h_0 are small enough such that the asymptotic bias is negligible compared to the asymptotic variance, then we have an undersmoothed estimator of $\mu(t)$, denoted by $\widehat{\mu}_{h,h_0}(t)$, such that for every fixed $t \in \mathcal{T}$,

$$\widehat{\mu}_{h,h_0}(t) - \mu(t) = \sum_{i=1}^{N} \frac{\eta_{h,h_0}(S_i, \boldsymbol{X}_i, Y_i; t)}{N \cdot f_T(t)} \cdot \{1 + o_P(1)\}, \tag{A.3}$$

and

$$\frac{\sqrt{N} f_T(t)\{\widehat{\mu}_{h,h_0}(t) - \mu(t)\}}{\{V_{\eta,h,h_0}(t)\}^{1/2}} \xrightarrow{D} N(0,1),$$

where $V_{\eta,h,h_0}(t)$ is the variance of $\eta_{h,h_0}(S, \boldsymbol{X}, Y; t)$.

Let h_{us} and $\tilde{h}_{\mathrm{us}}$ be respectively the largest bandwidths of h and h_0 such that the asymptotic bias is negligible compared to the asymptotic variance. We can construct an asymptotic pointwise confidence band with confidence level $1-\alpha$ by

$$\mathcal{B}_\alpha(\mathcal{T}) := \Bigg\{ (t,m) : t \in \mathcal{T}; \widehat{\mu}_{h_{\mathrm{us}},\tilde{h}_{\mathrm{us}}}(t) - \frac{\widehat{V}^{1/2}_{\eta,h_{\mathrm{us}},\tilde{h}_{\mathrm{us}}}(t)}{\sqrt{N}\widehat{f}_T(t)} \cdot z_{1-\alpha/2} \le m \le \widehat{\mu}_{h_{\mathrm{us}},\tilde{h}_{\mathrm{us}}}(t) + \frac{\widehat{V}^{1/2}_{\eta,h_{\mathrm{us}},\tilde{h}_{\mathrm{us}}}(t)}{\sqrt{N}\widehat{f}_T(t)} \cdot z_{1-\alpha/2} \Bigg\},$$

where $\widehat{f}_T$ and $\widehat{V}_{\eta,h,h_0}$ are some consistent estimators of f_T and V_{η,h,h_0}, respectively. For example, $\widehat{f}_T$ can be a deconvolution kernel density estimator with the plug-in bandwidth h_{PI} proposed by Delaigle and Gijbels (2002). To estimate $V_{\eta,h,h_0}(t)$, we can first estimate η_{h,h_0} by $\widehat{\eta}_{h,h_0}$, which is calculated as η_{h,h_0}'s definition above Theorem 4.2, replacing π_0, μ and $\mathbb{E}(Y|T=t,\boldsymbol{X})$ with $\widehat{\pi}$, $\widehat{\mu}_h$ and the partially linear estimator, $\widehat{m}(t,\boldsymbol{X})$, proposed by Liang (2000), respectively. Then $\widehat{V}_{\eta,h,h_0}(t)$ is the sample variance of $\widehat{\eta}_{h,h_0}(S_i, \boldsymbol{X}_i, Y_i; t)$ for $i = 1, \ldots, N$.

Now, the remaining problem is how to choose the undersmoothing bandwidths h_{us} and $\tilde{h}_{\mathrm{us}}$. Recall that our optimal bandwidths of h and h_0 for estimating $\mu(t)$ are respetively $\widehat{h}$ and h_{PI}. Our idea here is to estimate the asymptotic bias of $\widehat{\mu}_{h,h_0}(t)$, denoted by $\widehat{\mathrm{bias}}_{h,h_0}(t)$, for $(h,h_0) \in \{(h,h_0) : h = \widehat{h}/[\min(1.01^a, 2)], h_0 = h_{PI}/[\min(1.01^a, 1.1)], a = 1, 2, \ldots\}$, i.e. gradually reducing from $\widehat{h}$ and h_{PI} by dividing 1.01 each time, until $\int_{\mathcal{T}} \{\widehat{\mathrm{bias}}_{h,h_0}(t)\}^2\, dt < \int_{\mathcal{T}} \widehat{V}_{\eta,h,h_0}(t)\, dt/C$ for a large enough constant C, for example $C = 100$. Note that we set lower bounds for h_{us} and $\widetilde{h}_{\mathrm{us}}$ to be $\widehat{h}/2$ and $h_{PI}/1.1$, respectively. This is because the deconvolution kernel $L_U\{(t-S_i)/h\}$'s take more nagative values as the bandwidth gets smaller, making the computation unstable.

From Theorem 4.2 and 4.4, we can estimate the asymptotic bias by

$$\widehat{\mathrm{bias}}_{h,h_0}(t) := \frac{\kappa_{21}}{2}\left[\frac{\widehat{f}_T(t)\widehat{\Phi}_1(t) - \widehat{\mu}(t)\widehat{\partial_t^2 f}_T(t)}{\widehat{f}_T(t)}\right]\cdot h^2 + \frac{\kappa_{21}}{2}\left[\frac{\widehat{\mu}(t)\widehat{\partial_t^2 f}_T(t) - \widehat{f}_T(t)\widehat{\Phi}_2(t)}{\widehat{f}_T(t)}\right]\cdot h_0^2\,,$$

where

$$\widehat{\Phi}_1(t) := \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i\widehat{\partial_t^2 f}_{T|Y,\boldsymbol{X}}(t|Y_i,\boldsymbol{X}_i)}{\widehat{f}_{T|\boldsymbol{X}}(t|\boldsymbol{X}_i)}\,,$$

$$\widehat{\Phi}_2(t) := \frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{m}(t,\boldsymbol{X}_i)\widehat{\partial_t^2 f}_{T|\boldsymbol{X}}(t|\boldsymbol{X}_i)}{\widehat{f}_{T|\boldsymbol{X}}(t|\boldsymbol{X}_i)}\,,$$

and the conditional density $f_{T|\boldsymbol{Z}}(t|\boldsymbol{Z})$, for $\boldsymbol{Z} = \boldsymbol{X}$ or $\{Y, \boldsymbol{X}\}$, can be estimated by

$$\widehat{f}_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}) := \frac{\sum_{i=1}^{N} L_{U,h_{PI}}(t-S_i)L_{h_Z}(\|\boldsymbol{Z}-\boldsymbol{Z}_i\|)}{\sum_{i=1}^{N} L_{h_Z}(\|\boldsymbol{Z}-\boldsymbol{Z}_i\|)}\,,$$

the second partial derivatives, $\partial_t^2 f_T$, $\partial_t^2 f_{T|Y,\boldsymbol{X}}$ and $\partial_t^2 f_{T|\boldsymbol{X}}$, are estimated by replacing the $L_{U,h_{PI}}(t-S_i)$'s in the corresponding density estimators, $\widehat{f}_T$ and $\widehat{f}_{T|Y,\boldsymbol{X}}$, respectively, by the $\partial_t^2 L_{U,h_{PI}}(t-S_i)$'s (see Meister, 2009, Chap 2.7.2), and the bandwidth h_Z is chosen by a cross-validation method described below. Note that contructing a very good estimator of the asymptotic bias in measurement error context is difficult (see e.g. Delaigle et al., 2015). Our $\widehat{\mathrm{bias}}_{h,h_0}(t)$ does not estimate $\mathrm{bias}_{h,h_0}(t)$ very well either, but it is good enough for finding reliable undersmoothing bandwidths for h and h_0.

Our cross-validation method is an extension of that in Meister (2009) (Chap 2.5.1) to the condtional density estimator case. Consider the ISE

$$\begin{aligned}
&\int\{\widehat{f}_{T|\boldsymbol{Z}}(t|\boldsymbol{z}) - f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})\}^2 f_{\boldsymbol{Z}}(\boldsymbol{z})\,dt\,d\boldsymbol{z}\\
&= \int \widehat{f}_{T|\boldsymbol{Z}}^2(t|\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})\,dt\,d\boldsymbol{z} - 2\int \widehat{f}_{T|\boldsymbol{Z}}(t|\boldsymbol{z})f_{T|\boldsymbol{Z}}(t|\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})\,dt\,d\boldsymbol{z}\\
&\quad + \int f_{T|\boldsymbol{Z}}^2(t|\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})\,dt\,d\boldsymbol{z}\,.
\end{aligned}$$

Observe that the third integral is independent of the bandwidth. The first two integrals can be estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\int \widehat{f}^2_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}_i)\,dt - \frac{2}{N}\sum_{i=1}^{N}\int \widehat{f}_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}_i) f_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}_i)\,dt\,.$$

Using Plancherel's isometry (Theorem A.4 of Meister, 2009),

$$\begin{aligned}
I(h_Z;\boldsymbol{Z}_i) &:= \int \widehat{f}_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}_i) f_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}_i)\,dt \\
&= \frac{1}{2\pi}\int \widehat{f}^{ft}_{T|\boldsymbol{Z}}(w;\boldsymbol{Z}_i) f^{ft}_{T|\boldsymbol{Z}}(-w;\boldsymbol{Z}_i)\,dw \\
&= \frac{1}{2\pi}\int \frac{\sum_{j\neq i}\exp(iwS_j)\phi_L(wh_{PI}) f^{ft}_{T|\boldsymbol{Z}}(-w;\boldsymbol{Z}_i) L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_j\|)}{\phi_U(w)\sum_{j\neq i} L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_j\|)}\,dw\,,
\end{aligned}$$

where f^{ft} denotes the Fourier transform of a function f. Then we can obtain an empirically accessible version of this integral by replacing the $f^{ft}_{T|\boldsymbol{Z}}(-w;\boldsymbol{Z}_i)$'s with

$$\frac{\sum_{k\neq i,j}\exp(-iwS_k) L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_k\|)}{\phi_U(w)\sum_{k\neq i,j} L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_k\|)}\,.$$

We thus define an empirical version of $I(h_Z;\boldsymbol{Z}_i)$, $\widehat{I}(h_Z;\boldsymbol{Z}_i)$, by

$$\frac{1}{2\pi}\int \frac{\sum_{j\neq i}\sum_{k\neq i,j}\exp\{-iw(S_k-S_j)\}\phi_L(wh_{PI})\cdot L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_j\|) L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_k\|)}{|\phi_U(w)|^2\sum_{j\neq i} L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_j\|)\sum_{k\neq i,j} L_{h_Z}(\|\boldsymbol{Z}_i-\boldsymbol{Z}_k\|)}\,dw\,.$$

Then, we choose h_Z that minimises

$$\frac{1}{N}\sum_{i=1}^{N}\int \widehat{f}^2_{T|\boldsymbol{Z}}(t|\boldsymbol{Z}_i)\,dt - \frac{2}{N}\sum_{i=1}^{N}\mathrm{Re}\{\widehat{I}(h_Z;\boldsymbol{Z}_i)\}\,.$$

A.4.2. Numerical Results

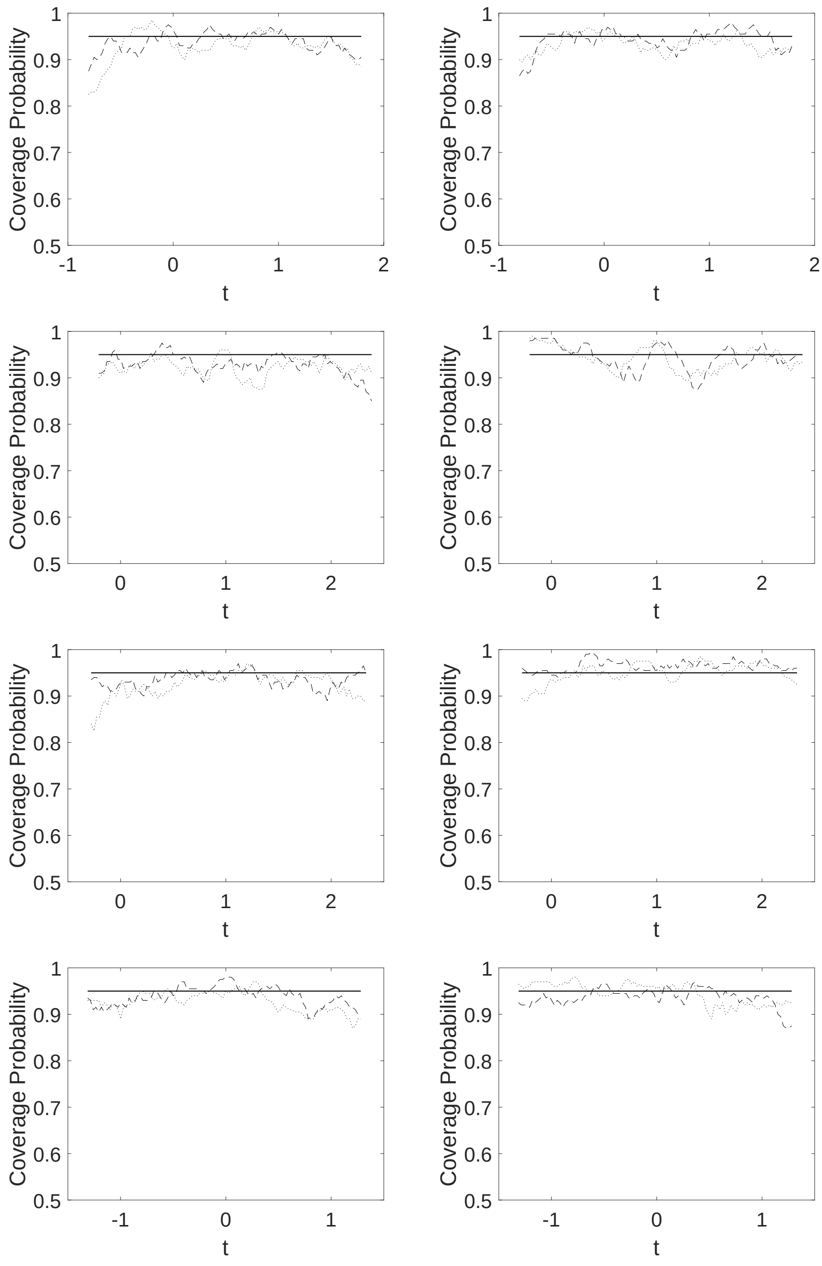

Fig. 6: The empirical coverage probabilities of our 95% undersmoothing poitwise confidence bands of models 1 to 4 (rows 1 to 4), Laplace (left) and Normal (right) measurement errors with $N = 250$ (dotted line) and $N = 500$ (dashed line).

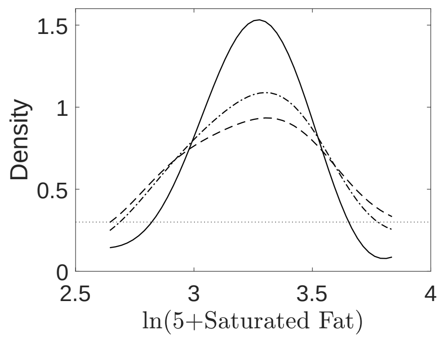

Fig. 7: Estimated denstiy functions of the treatment effect of the log-saturated fat intake for a Gaussian error of $\text{var}(U)/\text{var}(S) = 0.17$ (dashe-dotted), $\text{var}(U)/\text{var}(S) = 0.43$ (dashed) and $\text{var}(U)/\text{var}(S) = 0.75$ (solid), and the benchmark line at Density $= 0.3$ (dotted).

We applied our method to construct 95% confidence bands for the simulation models in section 6. Figure 6 shows the empirical coverage probabilities of our confidence bands. There, we can see that our method is reasonably good, with all the empirical coverage probabilities close to 95%.

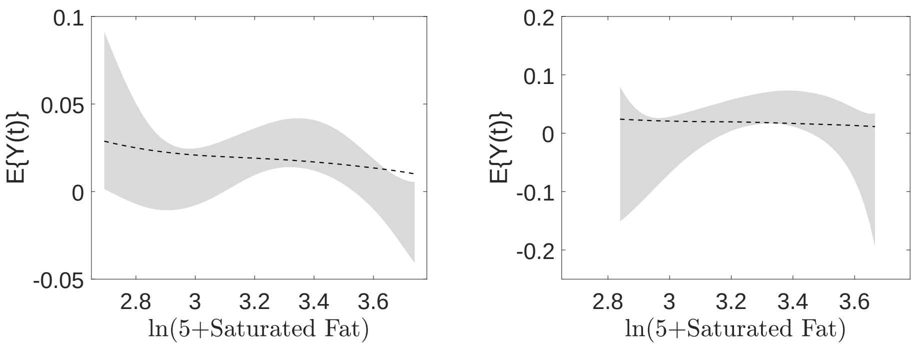

Fig. 8: Estimation of the treatment effect of the log-saturated fat intake on the risk of breast cancer and a 95% pointwise confidence band for a Gaussian error of $\text{var}(U)/\text{var}(S) = 0.43$ (left) and $\text{var}(U)/\text{var}(S) = 0.75$ (right).

We next show the 95% pointwise confidence bands for the Epidemiologic Study Cohort data from NHANES-I, when the variance of the measurement error is assumed to be $\text{var}(U)/\text{var}(S) = 0.43$ and 0.75. Note that when the variance of measurement error is large, the estimation is unstable, espeically in the tail parts. The reliable range of T is then different for different $\mathrm{var}(U)/\mathrm{var}(S)$. Based on the estimated density of the log-transformed saturated fat (see Figure 7), we plot in Figure 8 the confidence bands on $\ln(5+\text{Saturated Fat}) \in [2.69, 3.74]$ and $[2.84, 3.67]$ for $\mathrm{var}(U)/\mathrm{var}(S) = 0.43$ and 0.75, respectively, where the minimum estimated density values are at least 0.3.

For the case where $\mathrm{var}(U)/\mathrm{var}(S) = 0.43$, the reliable range of t is slightly shorter than that for $\mathrm{var}(U)/\mathrm{var}(S) = 0.17$. However, we can see the confidence bands show a very similar trend to that in Figure 4. For the case where $\mathrm{var}(U)/\mathrm{var}(S) = 0.75$, the reliable range of t is much shorter. But we can see from the confidence band that the trend between $t = 3$ to 3.6 is similar to those for $\mathrm{var}(U)/\mathrm{var}(S) = 0.17$ and 0.43, with a slight increase before $t = 3.4$ and a significant decrease from $t = 3.4$ to $t = 3.6$.

A.5. The rates of the tuning parameters

Note from section 5 that we set $h \asymp h_0 \asymp h_{PI}$ and $K \asymp h_{PI}^{-2}\log(h_{PI}+1)$. These K, h_0 and h give us the optimal convergence rate of our estimator $\widehat{\mu}$: for our final estimator $\widehat{\mu}$, the optimal rate is achieved when $h + h_0 \asymp [\{v_h(t) + v_{h_0}(t)\}/N]^{1/4}$. Since $h_{PI} \asymp \{v_{h_{PI}}(t)\}/N^{1/4}$, setting $h \asymp h_0 \asymp h_{PI}$ gives us the optimal convergence rate of our estimator $\widehat{\mu}$.

Regarding K, recalling that to obtain our Theorems 4.2 and 4.4, we require

$$\zeta(K)(K^{-\alpha} + h_0^2 + h^2) \to 0\,, \quad \frac{(K^{-\ell} + h_0^2)(K^{-\alpha} + h_0^2)}{h^2} \to 0\,, \text{ and}$$
$$\frac{v_{h_0}(t)}{\sqrt{v_h(t) \vee v_{h_0}(t)}} \frac{K}{\sqrt{N}} \to 0 \text{ ((D.28) on page 41 of Appendix D)}\,.$$

Thus, given $h \asymp h_0 \asymp \{v_h(t)/N\}^{1/4}$, we require

$$\zeta(K)(K^{-\alpha} + h^2) \to 0\,, \quad \frac{K^{-\ell-\alpha}}{h^2} \to 0\,, \text{ and } K \cdot h^2 \to 0\,,$$

which gives $\zeta(K) = o(K^{\alpha})$, $K = o(h^{-2})$ and $h^{-2} = o(K^{\alpha+\ell})$, because $\zeta(K) = O(\sqrt{K})$ for B-spline and $\zeta(K) = O(K)$ for polynomial sieve. Thus, we require $\alpha + \ell > 1$ and $\alpha > 1/2$ if B-spline basis is used and $\alpha > 1$ if the polynomial sieve is used. Therefore, we set $K = \lfloor \tilde{c} \cdot h_{PI}^{-2}\log(h_{PI}+1)\rfloor$ and select $\tilde{c}$.

Recalling from the definition of α and ℓ in Assumptions 5 and 7, if we use B-spline then no additional regularity condition is required. If the polynomial sieve is used, then we need to resctrict more on the smoothness of $\boldsymbol{x} \mapsto \pi_0(t, \boldsymbol{x})$.

A.6. Full results of simulation in section 6

In this section, we present the boxplots of our simulation results from Models 1 to 4.

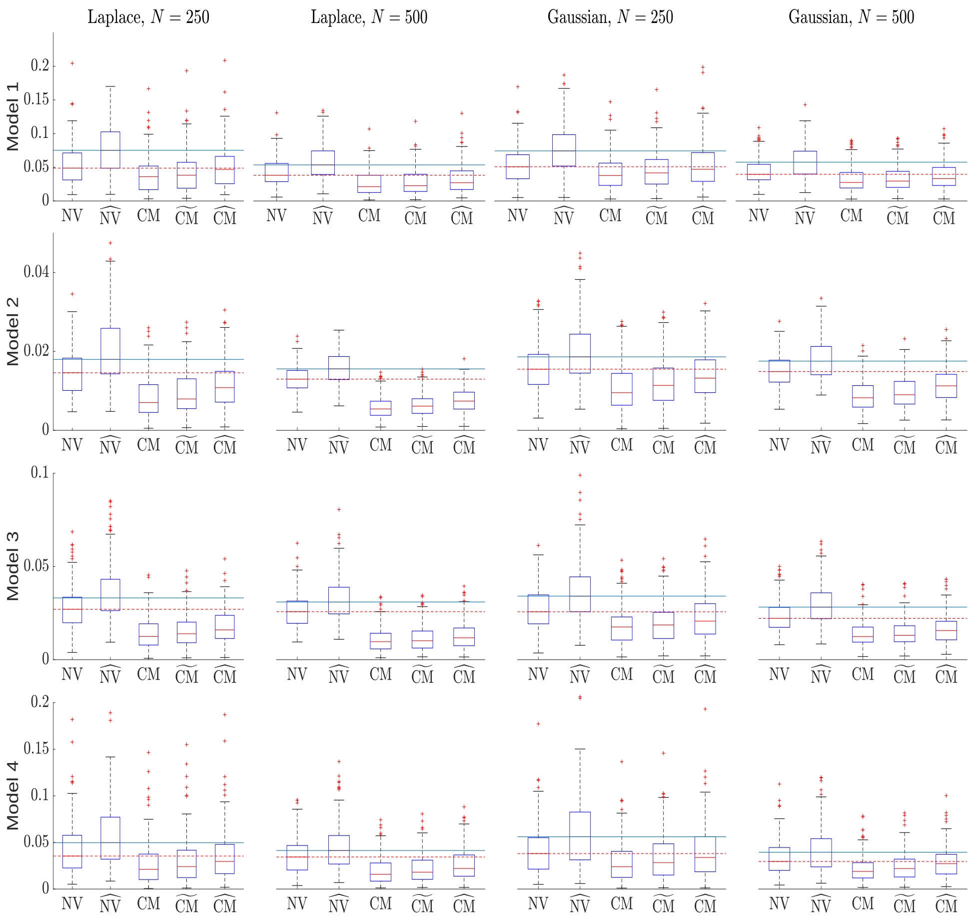

Fig. 9: Box plots of the ISEs of each estimator calculated from 200 samples generated from Model 1 to 4. The red dashed line indicates the median ISE value of NV estimators and the blue solid line indicates that of $\widehat{NV}$ estiamtors.

B. Asymptotic properties of deconvolution kernel

Note that the differences of the asymptotic behaviour of our estimators between the ordinary smooth error case and the suppersmooth error case come from the difference of the aysmptotic variance of the deconvolution kernel. In this section, we present a lemma regarding the asymptotic first and second moments, and the asymptotic variance of the deconvolution kernel, which shows the differences between the ordinary smooth error and the supersmooth error. For the simplicity of the notation, we define

$$v_h(t) := \mathbb{E}\left\{L_{U,h}^2(t-S)\right\}. \tag{B.1}$$

LEMMA 1. *Suppose that $\phi_U(t) \neq 0$ for all t and $h \to 0$ as $N \to \infty$. For any integrable function $R(\boldsymbol{X}, Y)$, let $\tau^2(t) := \mathbb{E}\{R^2(\boldsymbol{X}, Y)|T = t\}$ be a continuous function on $\mathcal{T}$. Then for any $t \in \mathcal{T}$,*

$$\mathbb{E}\left\{R(\boldsymbol{X}, Y)L_{U,h}(t-S)\right\} = \mathbb{E}\{R(\boldsymbol{X}, Y)|T = t\}f_T(t)\{1 + o(1)\},$$

and there exists a $t_h \in \mathcal{T}$ such that

$$\mathbb{E}\left\{R^2(\boldsymbol{X}, Y)L_{U,h}^2(t-S)\right\} = v_h(t) \cdot \tau^2(t_h).$$

Furthermore,

(a) if Assumption O and (18) are satisfied,

$$v_h(t) = C \cdot h^{-(1+2\beta)} \cdot f_T * f_U(t)\{1 + o(1)\},$$

where $C := \int_{-\infty}^{\infty} J^2(v)\,dv = (2\pi c^2)^{-1}\int_{-\infty}^{\infty}|w|^{2\beta}\phi_L^2(w)\,dw$ with $J(v) := (2\pi c)^{-1}\int_{-\infty}^{\infty}\exp(-iwv)w^{\beta}\phi_L(w)\,dw$ and c defined in (18). Suppose further that $(\tau^2 f_T) * f_U(t)$ is bounded for all $t \in \mathcal{T}$, we have

$$var\left\{R(\boldsymbol{X}_i, Y_i)L_{U,h}(t-S_i)\right\} = \frac{C}{h^{1+2\beta}} \cdot (\tau^2 f_T) * f_U(t)\{1 + o(1)\};$$

(b) if Assumption S and (19) are satisfied,

$$v_h(t) = O\{\exp(2h^{-\beta}/\gamma)/h\}.$$

Suppose further that $v_h(t) \to \infty$ as $N \to \infty$, we have

$$var\left\{R(\boldsymbol{X}_i, Y_i)L_{U,h}(t-S_i)\right\} = v_h(t) \cdot \tau^2(t_h)\{1 + o(1)\}.$$

Remark: If the measurement error U is super smooth, we can only obtain an upper bound for $v_h(t)$ based Assumption S and (19), i.e. $O\{\exp(2h^{-\beta}/\gamma)/h\}$. This is why we need to impose an additional lower bound condition $v_h(t) \to \infty$ in case (b). A discussion on the cases when this lower bound is satisfied can be found in the main text after Theorem 4.4.

Proof. Regarding the first moment, using (8), we have

$$\begin{aligned}
&\mathbb{E}\left\{R(\boldsymbol{X}_i, Y_i)L_{U,h}\left(t-S\right)\right\} \\
&= \frac{1}{h}\mathbb{E}\left[\mathbb{E}\{R(\boldsymbol{X}, Y)|T\} \cdot \mathbb{E}\left\{L_U\left(\frac{t-S}{h}\right)\middle| T\right\}\right] \\
&= \frac{1}{h}\mathbb{E}\left[\mathbb{E}\{R(\boldsymbol{X}, Y)|T\} \cdot L\left(\frac{t-T}{h}\right)\right] \\
&= \mathbb{E}\{R(\boldsymbol{X}, Y)|T=t\}f_T(t)\{1+o(1)\}\,.
\end{aligned}$$

For the second moment,

$$\begin{aligned}
&\mathbb{E}\left\{R^2(\boldsymbol{X}, Y)L^2_{U,h}\left(t-S\right)\right\} \\
&= \mathbb{E}\left\{\tau^2(T)L^2_{U,h}\left(t-T-U\right)\right\} \\
&= \int_{-\infty}^{\infty}\int_{\mathcal{T}} \tau^2(w)L^2_{U,h}(t-w-u)f_T(w)f_U(u)\,dw\,du\,.
\end{aligned}$$

Note that $\mathcal{T}$ is a compact interval, using the mean value theorem for definite integral, there exists a $t_h \in \mathcal{T}$ such that

$$\begin{aligned}
&\mathbb{E}\left\{R^2(\boldsymbol{X}, Y)L^2_{U,h}\left(t-S\right)\right\} \\
&= \tau^2(t_h)\int_{-\infty}^{\infty}\int_{\mathcal{T}} L^2_{U,h}(t-w-u)f_T(w)f_U(u)\,dw\,du = v_h(t)\cdot\tau^2(t_h).
\end{aligned}$$

Then if $v_h(t) \to \infty$ as $N \to \infty$, we have

$$\begin{aligned}
&\operatorname{var}\left\{R(\boldsymbol{X}_i, Y_i)L_{U,h}\left(t-S\right)\right\} \\
&= v_h(t)\tau^2(t_h) - f_T(t)\mathbb{E}\{R(\boldsymbol{X}, Y)|T=t\}\{1+o(1)\} \\
&= v_h(t)\tau^2(t_h)\{1+o(1)\}\,.
\end{aligned}$$

(a) Suppose that Assumption O and (18) are satisfied. We decompose $v_h(t)\tau^2(t_h)$ as follows, the arguments for the result of $v_h(t)$ is the same by taking $\tau^2(t) = 1$ for all t. Note that

$$v_h(t)\tau^2(t_h) = \frac{1}{h^2}\int_{-\infty}^{\infty}\int_{\mathcal{T}} \tau^2(w)L^2_U\{(t-w-u)/h\}f_T(w)f_U(u)\,dw\,du$$

$$= \frac{1}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tau^2(t-u-vh)L_U^2(v)f_T(t-u-vh)f_U(u)\,dw\,du\,.$$

Using (6), (18) and dominated convergence theorem, we have

$$h^\beta L_U(v) \to \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-iwv)\phi_L(w)\cdot\frac{w^\beta}{c}dw =: J(v) \quad \text{as} \quad h\to 0. \tag{B.2}$$

Under Assumption O (i), by Lemma 3 of Fan and Truong (1993), $|h^\beta L_U(v)| \le C_0/(1+|v|)$, for some positive constant C_0. Given that $(\tau^2 f_T)*f_U(t)$ is bounded for all $t\in\mathcal{T}$, we then have

$$\begin{aligned}
&\frac{1}{h}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tau^2(t-u-vh)L_U^2(v)f_T(t-u-vh)f_U(u)\,dw\,du\\
&= \frac{1}{h^{1+2\beta}}\int_{-\infty}^{\infty}J^2(v)\,dv\int_{-\infty}^{\infty}\tau^2(t-u)f_T(t-u)f_U(u)\,dw\,du\{1+o(1)\}\\
&= \frac{1}{h^{1+2\beta}}\int_{-\infty}^{\infty}J^2(v)\,dv\cdot(\tau^2 f_T)*f_U(t)\{1+o(1)\}\,. \qquad \text{(B.3)}
\end{aligned}$$

Now, by Parseval's identity,

$$\int_{-\infty}^{\infty}J^2(v)\,dv = \frac{1}{2\pi}\int_{-\infty}^{\infty}|w^\beta/c|^2\phi_L^2(w)\,dw\,.$$

The result then follows.

(b) Suppose that Assumption S and (19) are satisfied. By a change of variable,

$$v_h(t) = \frac{1}{h}\int_{-\infty}^{\infty}L_U^2(v)f_S(t-vh)\,dv \le \frac{\sup_t f_S(t)}{h}\int_{-\infty}^{\infty}L_U^2(v)\,dv\,.$$

Then by Parseval's identity, we have

$$\int_{-\infty}^{\infty}L_U^2(v)\,dv = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\phi_L^2(w)}{\phi_U^2(w/h)}\,dw\,.$$

Using (19), there exists a constant M such that $(d_0/2)|t|^{\beta_0}\exp(-|t|^\beta/\gamma) < |\phi_U(t)|$ for $|t|>M$. Then by the bounded support of $\phi_L(t)$, we have

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\phi_L^2(w)|}{|\phi_U^2(w/h)|}\,dw\\
\le&\frac{1}{2\pi}\int_{|w|\le Mh}\frac{|\phi_L^2(w)|}{|\phi_U^2(w/h)|}\,dw + \frac{2}{\pi d_0^2}\int_{Mh\le|w|\le 1}|\phi_L^2(w)||w/h|^{-2\beta_0}\exp(2|w/h|^\beta/\gamma)\,dw
\end{aligned}$$

$$
\begin{aligned}
&\leq \inf_{|w|\leq M} |\phi_U(w)|^{-2} \sup_{|w|\leq M} |\phi_L^2(w)| \frac{1}{2\pi} \int_{|w|\leq Mh} dw \\
&\quad + \frac{2}{\pi d_0^2} \sup_{|w|\leq 1} |\phi_L^2(w)| \sup_{Mh\leq |w|\leq 1} \exp(2|w/h|^\beta/\gamma)|h|^{2\beta_0} \int_{Mh\leq|w|\leq 1} |w|^{-2\beta_0}\, dw \\
&=O\{\exp(2h^{-\beta}/\gamma)\}\,.
\end{aligned}
$$

Thus, we have $v_h(t) = O\{\exp(2h^{-\beta}/\gamma)/h\}$.

C. Proof of Theorems 4.1 and 4.3

Theorems 4.1 and 4.3 provide the convergence rates of $\widehat{\pi}(t,\boldsymbol{X}) \to \pi_0(t,\boldsymbol{X})$ for every fixed $t \in \mathcal{T}$ in the ordinary smooth and supersmooth cases, respectively. They directly follow from Lemmas 2 and 3 stated in the following subsections. We introduce some notations which will be used later. By Assumption 6 (i), without loss of generality, we assume the sieve basis $u_K(\boldsymbol{X})$ is orthonormalized, namely

$$\mathbb{E}[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})] = I_K. \tag{C.1}$$

Recall

$$\widehat{\lambda}_t = \arg\max_{\lambda\in\mathbb{R}^K} \widehat{G}_t(\lambda) \text{ and}$$
$$\widehat{G}_t(\lambda) = \frac{\sum_{i=1}^N \rho\{\lambda^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} - \lambda^\top\left\{\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\right\}.$$

Let $G_t^*(\lambda)$ and λ_t^* be the theoretical counterparts of $\widehat{G}_t(\lambda)$ and $\widehat{\lambda}_t$, i.e.,

$$\lambda_t^* := \arg\max_{\lambda\in\mathbb{R}^K} G_t^*(\lambda) \tag{C.2}$$

and

$$\begin{aligned} G_t^*(\lambda) :=& \frac{\mathbb{E}\left[\rho\{\lambda^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}\right]}{\mathbb{E}\left[L_U\{(t-S_i)/h_0\}\right]} - \lambda^\top\mathbb{E}\left[u_K(\boldsymbol{X}_i)\right] \\ =& \frac{\mathbb{E}\left[\rho\{\lambda^\top u_K(\boldsymbol{X}_i)\}L\{(t-T_i)/h_0\}\right]}{\mathbb{E}\left[L\{(t-T_i)/h_0\}\right]} - \lambda^\top\mathbb{E}\left[u_K(\boldsymbol{X}_i)\right], \end{aligned}$$

where the last equality comes from (8). Let

$$\pi^*(t,\boldsymbol{x}) := \rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\},$$

be the theoretical counterpart of $\widehat{\pi}(t,\boldsymbol{x})$. Theorem 4.1 holds by applying triangular inequality to the results of Lemmas 2 and 3, which are established in the following subsections.

C.1. Lemma 2

The first lemma gives the approximation rate of the intermediate quantity $\pi^*(t,\boldsymbol{x})$. Recall the notation $\zeta(K) = \sup_{\boldsymbol{x}\in\mathcal{X}} \|u_K(\boldsymbol{x})\|$.

LEMMA 2. *Under Assumptions 2-6, for every fixed $t \in \mathcal{T}$, we have*

$$\sup_{\boldsymbol{x}\in\mathcal{X}} |\pi^*(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})| = O\left\{\zeta(K)(K^{-\alpha} + h_0^2)\right\},$$

$$\int_{\mathcal{X}} |\pi^*(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})|^2 \, dF_X(\boldsymbol{x}) = O\left(K^{-2\alpha} + h_0^4\right),$$

$$\frac{1}{N}\sum_{i=1}^N |\pi^*(t,\boldsymbol{X}_i) - \pi_0(t,\boldsymbol{X}_i)|^2 = O_p\left(K^{-2\alpha} + h_0^4\right).$$

PROOF. By Assumption 3, for every fixed $t \in \mathcal{T}$, we can find two positive constants η_1 and η_2 such that $\pi_0(t,\boldsymbol{x}) \in [\eta_1, \eta_2]$, $\forall \boldsymbol{x} \in \mathcal{X}$ and the fact $(\rho')^{-1}$ is strictly decreasing, we have

$$(\rho')^{-1}(\eta_2) \le \inf_{\boldsymbol{x}\in\mathcal{X}} (\rho')^{-1}\left\{\pi_0(t,\boldsymbol{x})\right\} \le \sup_{\boldsymbol{x}\in\mathcal{X}} (\rho')^{-1}\left\{\pi_0(t,\boldsymbol{x})\right\} \le (\rho')^{-1}(\eta_1) \ .$$

By Assumption 5,

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \left|(\rho')^{-1}\left\{\pi_0(t,\boldsymbol{x})\right\} - \lambda_t^\top u_K(\boldsymbol{x})\right| < C_1 K^{-\alpha}, \tag{C.3}$$

where $C_1 > 0$ is a universal constant. Then we have

$$\begin{aligned} \lambda_t^\top u_K(\boldsymbol{x}) &\in \left((\rho')^{-1}\left\{\pi_0(t,\boldsymbol{x})\right\} - C_1K^{-\alpha}, (\rho')^{-1}\left\{\pi_0(t,x)\right\} + C_1K^{-\alpha}\right) \\ &\subset \left[(\rho')^{-1}(\eta_2) - C_1K^{-\alpha}, (\rho')^{-1}(\eta_1) + C_1K^{-\alpha}\right] \ , \ \forall \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{C.4}$$

and

$$\begin{aligned} &\rho'\{\lambda_t^\top u_K(\boldsymbol{x}) + C_1K^{-\alpha}\} - \rho'\{\lambda_K^\top u_K(\boldsymbol{x})\} < \pi_0(t,\boldsymbol{x}) - \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} \\ &< \rho'\{\lambda_t^\top u_K(\boldsymbol{x}) - C_1K^{-\alpha}\} - \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} \ , \ \forall \boldsymbol{x} \in \mathcal{X}. \end{aligned}$$

By the mean value theorem, for large enough K, there exists

$$\xi_1(\boldsymbol{x}) \in (\lambda_t^\top u_K(\boldsymbol{x}), \lambda_t^\top u_K(\boldsymbol{x}) + C_1K^{-\alpha}) \subset [(\rho')^{-1}(\eta_2) - C_1K^{-\alpha}, (\rho')^{-1}(\eta_1) + 2C_1K^{-\alpha}] \subset \Gamma_1$$

$$\xi_2(\boldsymbol{x}) \in (\lambda_t^\top u_K(\boldsymbol{x}) - C_1K^{-\alpha}, \lambda_t^\top u_K(\boldsymbol{x})) \subset [(\rho')^{-1}(\eta_2) - 2C_1K^{-\alpha}, (\rho')^{-1}(\eta_1) + C_1K^{-\alpha}] \subset \Gamma_1,$$

where

$$\Gamma_1 := \left[(\rho')^{-1}(\eta_2) - 1, (\rho')^{-1}(\eta_1) + 1\right],$$

such that

$$\rho'\{\lambda_t^\top u_K(\boldsymbol{x}) + C_1K^{-\alpha}\} - \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} = \rho''\{\xi_1(\boldsymbol{x})\}C_1K^{-\alpha} \ge \inf_{y\in\Gamma_1} \rho''(y) C_1 K^{-\alpha}$$

$$\rho'\{\lambda_t^\top u_K(\boldsymbol{x}) - C_1K^{-\alpha}\} - \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} = -\rho''\{\xi_2(\boldsymbol{x})\}C_1K^{-\alpha} \leq \sup_{y\in\Gamma_1} -\rho''(y)C_1K^{-\alpha} .$$

Let $a := \max\left\{-\inf_{y\in\Gamma_1}\rho''(y), \sup_{y\in\Gamma_1} -\rho''(y)\right\}$, which is a finite positive constant because the set Γ_1 is compact and the function $\rho''(y)$ is continuous. Therefore, for every fixed $t \in \mathcal{T}$,

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\pi_0(t,\boldsymbol{x}) - \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\}\right| < aC_1K^{-\alpha}. \tag{C.5}$$

For some fixed $C_2 > 0$ (to be chosen later), define the set

$$\Lambda_t := \left\{\lambda \in \mathbb{R}^K : \|\lambda - \lambda_t\| \leq C_2\left(K^{-\alpha} + h_0^2\right)\right\}.$$

For sufficiently large K, by (C.4), Assumption 6 (ii), we have that $\forall\lambda \in \Lambda_t$, $\forall \boldsymbol{x} \in \mathcal{X}$,

$$\begin{aligned}
&|\lambda^\top u_K(\boldsymbol{x}) - \lambda_t^\top u_K(\boldsymbol{x})| = |(\lambda - \lambda_t)^\top u_K(\boldsymbol{x})| \leq \|\lambda - \lambda_t\|\|u_K(\boldsymbol{x})\| \leq C_2(K^{-\alpha} + h_0^2)\zeta(K)\\
\Rightarrow &\lambda^\top u_K(\boldsymbol{x}) \in \left(\lambda_t^\top u_K(\boldsymbol{x}) - C_2(K^{-\alpha} + h_0^2)\zeta(K), \lambda_t^\top u_K(\boldsymbol{x}) + C_2(K^{-\alpha} + h_0^2)\zeta(K)\right)\\
&\subset \Big[(\rho')^{-1}(\eta_1) - C_1K^{-\alpha} - C_2(K^{-\alpha} + h_0^2)\zeta(K),\\
&\quad (\rho')^{-1}(\eta_2) + C_1K^{-\alpha} + C_2(K^{-\alpha} + h_0^2)\zeta(K)\Big]\\
&\subset \Gamma_1 .
\end{aligned} \tag{C.6}$$

By (C.1), (C.5), and (C.6), we can deduce that

$$\begin{aligned}
&\|\nabla G_t^*(\lambda_t)\| = \left\|\frac{\mathbb{E}\left[\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}L_U\{(t-S)/h_0\}u_K(\boldsymbol{X})\right]}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]} - \mathbb{E}\left[u_K(\boldsymbol{X})\right]\right\|\\
=&\left\|\mathbb{E}\left[\left\{\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}L_U\{(t-S)/h_0\}}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]} - 1\right\}u_K(\boldsymbol{X})\right]\right\|\\
\leq&\left\|\mathbb{E}\left[\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}u_K(\boldsymbol{X})\cdot\frac{L_U\{(t-S)/h_0\}}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]}\right] - \mathbb{E}\left[\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}u_K(\boldsymbol{X})\Big|T=t\right]\right\|\\
&+\left\|\mathbb{E}\left(\left[\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\} - \pi_0(t,\boldsymbol{X})\right]u_K(\boldsymbol{X})\Big|T=t\right)\right\|\\
\leq&C_0\cdot h_0^2 + a\cdot C_1\cdot K^{-\alpha}\\
\leq&(aC_1 + C_0)\cdot(K^{-\alpha} + h_0^2),
\end{aligned} \tag{C.7}$$

where the second inequality holds by using the following results: by (8),

$$\begin{aligned}
&\mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}L_U\{(t-S)/h_0\}}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]}u_K(\boldsymbol{X})\right]\\
=&\frac{\mathbb{E}\left[\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}L\{(t-T)/h_0\}u_K(\boldsymbol{X})\right]}{\mathbb{E}\left[L\{(t-T)/h_0\}\right]}
\end{aligned}$$

$$
\begin{aligned}
=& \frac{1}{f_T(t)+O(h_0^2)} \cdot \left[\frac{1}{h_0}\int \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} L\left\{\frac{t-s}{h_0}\right\} u_K(\boldsymbol{x}) f_{T,X}(s,\boldsymbol{x}) ds d\boldsymbol{x}\right] \\
=& \frac{1}{f_T(t)+O(h_0^2)} \cdot \left[\int \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} L\{v\} u_K(\boldsymbol{x}) f_{T,X}(t-h_0\cdot v,\boldsymbol{x}) dv d\boldsymbol{x}\right] \\
=& \frac{1}{f_T(t)+O(h_0^2)} \cdot \left[\int \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} u_K(\boldsymbol{x}) f_{T,X}(t,\boldsymbol{x}) d\boldsymbol{x}\right. \\
& \left. + \frac{h_0^2}{2}\int \rho'\{\lambda_t^\top u_K(\boldsymbol{x})\} u_K(\boldsymbol{x}) \partial_{tt} f_{T,X}(t,\boldsymbol{x}) d\boldsymbol{x} + o(h_0^2)\right] \\
=& \mathbb{E}\left[\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\} u_K(\boldsymbol{X}) | T=t\right] + \frac{h_0^2}{2}\cdot \mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K(\boldsymbol{X})\right] + o(h_0^2),
\end{aligned}
$$

and

$$
\begin{aligned}
& \left\|\mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K(\boldsymbol{X})\right]\right\|^2 \\
=& \mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K^\top(\boldsymbol{X})\right] \cdot \mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K(\boldsymbol{X})\right] \\
=& \mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K^\top(\boldsymbol{X})\right] \cdot \mathbb{E}\left[u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\right]^{-1} \\
& \cdot \mathbb{E}\left[u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\right] \\
& \cdot \mathbb{E}\left[u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\right]^{-1} \mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K(\boldsymbol{X})\right] \qquad \text{(by (C.1))} \\
=& \left\|\mathbb{E}\left[\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})} u_K^\top(\boldsymbol{X})\right] \cdot \mathbb{E}\left[u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\right]^{-1} u_K(\boldsymbol{X})\right\|^2_{L^2(dF_X)} \\
\leq & \left\|\frac{\rho'\{\lambda_t^\top u_K(\boldsymbol{X})\}\partial_{tt} f_{T,X}(t,\boldsymbol{X})}{f_T(t) f_X(\boldsymbol{X})}\right\|^2_{L^2(dF_X)} = O(1).
\end{aligned}
$$

Then for any $\lambda \in \partial\Lambda_t$, i.e. $\|\lambda - \lambda_t\| = C_2\{K^{-\alpha} + h_0^2\}$, by the mean value theorem we have

$$
\begin{aligned}
& G_t^*(\lambda) - G_t^*(\lambda_t) \\
&= (\lambda-\lambda_t)^\top \nabla G_t^*(\lambda_t) + \frac{1}{2}(\lambda-\lambda_t)^\top \nabla^2 G_t^*(\bar{\lambda}_t)(\lambda-\lambda_t) \\
&\leq \|\lambda-\lambda_t\|\|\nabla G_t^*(\lambda_t)\| \\
&\quad + \frac{1}{2}(\lambda-\lambda_t)^\top \mathbb{E}\left[\frac{\rho''\{\bar{\lambda}_t^\top u_K(\boldsymbol{X})\} L_U\{(t-S)/h_0\} u_K(\boldsymbol{X}) u_K(\boldsymbol{X})^\top}{\mathbb{E}\left[L_U\{(t-S)/h_0\}\right]}\right](\lambda-\lambda_K)
\end{aligned}
$$

$$
\begin{aligned}
&\le \|\lambda-\lambda_t\|\|\nabla G_t^*(\lambda_t)\| - \frac{a_1}{2}\cdot(\lambda-\lambda_t)^\top \mathbb{E}[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t]\cdot(\lambda-\lambda_t)\\
&\le \|\lambda-\lambda_t^*\|\|\nabla G_t^*(\lambda_t)\| - \frac{a_1}{2}\cdot\lambda_{\min}\left(\mathbb{E}[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t]\right)\cdot\|\lambda-\lambda_t\|^2\\
&= \|\lambda-\lambda_t\|\left(\|\nabla G_t^*(\lambda_t)\| - \frac{a_1}{2}\cdot\lambda_{\min}\left(\mathbb{E}[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t]\right)\cdot\|\lambda-\lambda_t\|\right)\\
&\le \|\lambda-\lambda_t\|\Big\{(aC_1+C_0)\cdot(K^{-\alpha}+h_0^2)\\
&\qquad -\frac{a_1}{2}\cdot\lambda_{\min}\left[\mathbb{E}\{u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\}\right]\cdot C_2\cdot(K^{-\alpha}+h_0^2)\Big\} \qquad (C.8)
\end{aligned}
$$

where $\bar{\lambda}_t$ lies between λ and λ_t on $\partial\Lambda_t$, $\lambda_{\min}(A)$ denotes the smallest eigenvalue of a matrix A, and $a_1=\inf_{y\in\Gamma_1}\{-\rho''(y)\}+o(h_0^2)\in(0,\infty)$ for sufficiently small h_0 under Assumption 6, and the last inequality follows from (C.7). By choosing

$$
C_2 > \frac{2\{aC_1+C_0\}}{a_1\cdot\lambda_{\min}\left(\mathbb{E}[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t]\right)}\ ,
$$

we can obtain that

$$
G_t^*(\lambda) < G_t^*(\lambda_t), \quad \lambda\in\partial\Lambda_t.
$$

In light of the continuity of G_t^*, there is a local maximum of G_t^* in the interior of Λ_t. On the other hand, G_t^* is a strictly concave function with a unique global maximum point λ_t^*, therefore we can claim

$$
\lambda_t^*\in\Lambda_t^\circ\ ,\quad i.e.\ \ \|\lambda_t^*-\lambda_t\|\le C_2\cdot\{K^{-\alpha}+h_0^2\}\ . \qquad (C.9)
$$

By the mean value theorem, for large enough K, there exists $\xi^*(\boldsymbol{x})$ lying between $(\lambda_t^*)^\top u_K(\boldsymbol{x})$ and $\lambda_t^\top u_K(\boldsymbol{x})$, which implies $\xi^*(\boldsymbol{x})\in\Gamma_1$, such that for any $\boldsymbol{x}\in\mathcal{X}$,

$$
\begin{aligned}
&|\rho'\{\lambda_t^\top u_K(\boldsymbol{x})\}-\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\}| \qquad (C.10)\\
&=|\rho''\{\xi^*(\boldsymbol{x})\}||\lambda_t^\top u_K(\boldsymbol{x})-(\lambda_t^*)^\top u_K(\boldsymbol{x})|\\
&\le -\rho''\{\xi^*(\boldsymbol{x})\}\|\lambda_t-\lambda_t^*\|\|u_K(\boldsymbol{x})\|\le a_2C_2\cdot\zeta(K)\{K^{-\alpha}+h_0^2\},
\end{aligned}
$$

where $a_2=\sup\limits_{\gamma\in\Gamma_1}-\rho''(\gamma)<\infty$. Therefore,

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\pi_0(t,\boldsymbol{x})-\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\}\right| \qquad (C.11)\\
&=\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\pi_0(t,\boldsymbol{x})-\rho'\{\lambda_t^\top u_K(\boldsymbol{x})\}+\rho'\{\lambda_t^\top u_K(\boldsymbol{x})\}-\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\}\right|\\
&\le \sup_{\boldsymbol{x}\in\mathcal{X}}\left|\pi_0(t,\boldsymbol{x})-\rho'\{\lambda_t^\top u_K(\boldsymbol{x})\}\right| + \sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho'\{\lambda_t^\top u_K(\boldsymbol{x})\}-\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\}\right|
\end{aligned}
$$

$$\leq aC_1K^{-\alpha} + a_2C_2 \cdot \zeta(K)\{K^{-\alpha} + h_0^2\}$$
$$\leq (aC_1 + a_2C_2)\{K^{-\alpha} + h_0^2\}\zeta(K) = O\left(\{K^{-\alpha} + h_0^2\}\zeta(K)\right),$$

where the second inequality follows from (C.5), (C.6) and (C.10).

Similarly, by (C.5), (C.6), (C.9), we can deduce that

$$\begin{aligned}
&\int_{\mathcal{X}} |\pi_0(t, \boldsymbol{x}) - \pi^*(t, \boldsymbol{x})|^2 \, dF_X(\boldsymbol{x}) \\
\leq& 2\int_{\mathcal{X}} \left|\pi_0(t, \boldsymbol{x}) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\}\right|^2 dF_X(\boldsymbol{x}) \\
&+ 2\int_{\mathcal{X}} \left|\rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\} - \rho'\left\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\right\}\right|^2 dF_X(\boldsymbol{x}) \\
\leq& 2\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\pi_0(t, \boldsymbol{x}) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\}\right|^2 + 2\cdot\int_{\mathcal{X}} |\rho''\{\xi^*(\boldsymbol{x})\}|^2 \cdot |(\lambda_t - \lambda_t^*)^\top u_K(\boldsymbol{x})|^2 dF_X(\boldsymbol{x}) \\
\leq& 2\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\pi_0(t, \boldsymbol{x}) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\}\right|^2 \\
&+ 2\sup_{\boldsymbol{x}\in\mathcal{X}} |\rho''\{\xi^*(\boldsymbol{x})\}|^2 \cdot (\lambda_t - \lambda_t^*)^\top \int_{\mathcal{X}} u_K(\boldsymbol{x})u_K(\boldsymbol{x})^\top dF_X(\boldsymbol{x}) \cdot (\lambda_t - \lambda_t^*) \\
=& 2\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\pi_0(t, \boldsymbol{x}) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\}\right|^2 + 2\sup_{\boldsymbol{x}\in\mathcal{X}} |\rho''\{\xi^*(\boldsymbol{x})\}|^2 \cdot \|\lambda_t - \lambda_t^*\|^2 \\
\leq& 2\cdot a^2 \cdot C_1^2 \cdot K^{-2\alpha} + 2\cdot a_2^2 \cdot C_2^2 \cdot \{K^{-\alpha} + h_0^2\}^2 = O(K^{-2\alpha} + h_0^4).
\end{aligned}$$

We can also obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N |\pi_0(t, \boldsymbol{X}_i) - \pi^*(t, \boldsymbol{X}_i)|^2 \\
\leq& \frac{2}{N}\sum_{i=1}^N \left|\pi_0(t, \boldsymbol{X}_i) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{X}_i)\right\}\right|^2 + \frac{2}{N}\sum_{i=1}^N \left|\rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\} - \rho'\left\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\right\}\right|^2 \\
=& \frac{2}{N}\sum_{i=1}^N \left|\pi_0(t, \boldsymbol{X}_i) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{X}_i)\right\}\right|^2 + \frac{2}{N}\sum_{i=1}^N \left|\rho''\left\{\xi^*(\boldsymbol{x})\right\}^\top (\lambda_t - \lambda_t^*) u_K(\boldsymbol{X}_i)\right|^2 \\
\leq& 2\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\pi_0(t, \boldsymbol{x}) - \rho'\left\{\lambda_K^\top u_K(\boldsymbol{x})\right\}\right|^2 \\
&+ 2\sup_{\boldsymbol{x}\in\mathcal{X}} |\rho''\{\xi^*(\boldsymbol{x})\}|^2 \cdot (\lambda_t - \lambda_t^*)^\top \left\{\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top\right\}(\lambda_t - \lambda_t^*) \\
\leq& 2\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\pi_0(t, \boldsymbol{x}) - \rho'\left\{\lambda_t^\top u_K(\boldsymbol{x})\right\}\right|^2
\end{aligned}$$

$$
\begin{aligned}
&+2\sup_{\boldsymbol{x}\in\mathcal{X}}|\rho''\{\xi^*(\boldsymbol{x})\}|^2\cdot\lambda_{\max}\left\{\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top\right\}\|\lambda_t-\lambda_t^*\|^2\\
=&O(K^{-2\alpha})+O(1)\cdot O_p(1)\cdot O(\{K^{-2\alpha}+h_0^4\})=O_p(K^{-2\alpha}+h_0^4),
\end{aligned}
$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of a matrix A; the second equality follows from Chebyshev's inequality and the following facts

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top-\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right\|^2\right] \qquad \text{(C.12)}\\
=&\frac{1}{N}\mathbb{E}\left[\left\|u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top-\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right\|^2\right]\\
=&\frac{1}{N}\mathbb{E}\left[\mathrm{tr}\left\{u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right\}\right]-\frac{1}{N}\mathrm{tr}\left\{\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\cdot\mathbb{E}[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top]\right\}\\
\leq&\frac{1}{N}\cdot\zeta(K)^2\mathbb{E}\left[\|u_K(\boldsymbol{X})\|^2\right]=\zeta(K)^2\frac{K}{N}\to 0\ ,
\end{aligned}
$$

and $\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]=I_K$.

C.2. Lemma 3

LEMMA 3. *Under Assumptions 2-6, if for any $t\in\mathcal{T}$, $\zeta(K)\sqrt{Kv_{h_0}(t)/N}\to 0$ as $N\to\infty$, we have*

$$
\begin{aligned}
&\left\|\widehat{\lambda}_t-\lambda_t^*\right\|=O_p\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right),\\
&\sup_{\boldsymbol{x}\in\mathcal{X}}|\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})|=O_p\left(\zeta(K)\sqrt{\frac{Kv_{h_0}(t)}{N}}\right),\\
&\int_{\mathcal{X}}|\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})|^2dF_X(\boldsymbol{x})=O_p\left(\frac{Kv_{h_0}(t)}{N}\right),\\
&\frac{1}{N}\sum_{i=1}^N|\widehat{\pi}(t,\boldsymbol{X}_i)-\pi^*(t,\boldsymbol{X}_i)|^2=O_p\left(\frac{Kv_{h_0}(t)}{N}\right)\ ,
\end{aligned}
$$

where $v_h(t)$ is defined in (B.1) *and its rate is derived in Lemma 1.*

Note that by Lemma 1, we have

(a) if Assumption O and (18) are satisfied, $Kv_{h_0}(t)/N\asymp K/(Nh_0^{1+2\beta})$;

(b) if Assumption S and (19) are satisfied, then $Kv_{h_0}(t)/N = O\{\exp(2h_0^{-\beta}/\gamma)\cdot K/(Nh_0)\}$.

Proof. Define

$$\widehat{S}_N := \frac{\sum_{i=1}^N L_{U,h_0}(t-S_i)u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top}{\sum_{i=1}^N L_{U,h_0}(t-S_i)},$$

where $L_{U,h_0}(v) := h_0^{-1}L_U(v/h_0)$, and $L_{h_0}(v) := h_0^{-1}L(v/h_0)$ for $v \in \mathbb{R}$. We have

$$\begin{aligned}
\widehat{S}_N =& \frac{1}{\mathbb{E}[L_{U,h_0}\{t-S_i\}]} \times \{1+o_P(1)\} \\
&\times \left[\frac{1}{N}\sum_{i=1}^N L_{U,h_0}(t-S_i)u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top - \mathbb{E}[L_{U,h_0}\{t-S_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]\right] \\
&+ \frac{\mathbb{E}[L_{U,h_0}\{t-S_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]}{\mathbb{E}[L_{U,h_0}\{t-S_i\}]} \times \{1+o_P(1)\} \\
=& O_P\left(\zeta(K)\sqrt{\frac{v_{h_0}(t)K}{N}}\right) + \frac{\mathbb{E}[L_{U,h_0}\{t-S_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]}{\mathbb{E}[L_{U,h_0}\{t-S_i\}]} \times \{1+o_P(1)\} \\
=& \frac{\mathbb{E}[L_{h_0}\{t-T_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]}{\mathbb{E}[L_{h_0}\{t-T_i\}]} + o_P(1),
\end{aligned} \tag{C.13}$$

where the last equality holds by (8) and the fact that, under Assumption 6 (i), for sufficiently large N, there exist two positive constants s_1 and s_2 such that

$$\begin{aligned}
0 < s_1 \le& \lambda_{\min}\left(\frac{\mathbb{E}[L_{h_0}\{t-T_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]}{\mathbb{E}[L_{h_0}\{t-T_i\}]}\right) \\
&\le \lambda_{\max}\left(\frac{\mathbb{E}[L_{h_0}\{t-T_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]}{\mathbb{E}[L_{h_0}\{t-T_i\}]}\right) \le s_2 < \infty,
\end{aligned} \tag{C.14}$$

and the second equality holds by Chebyshev's inequality and the following result:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^N L_{U,h_0}(t-S_i)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top - \mathbb{E}\left[L_{U,h_0}(t-S_i)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\right\|^2\right] \\
\le& \frac{1}{N}\cdot\mathbb{E}\left[\left\|L_{U,h_0}(t-S_i)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right\|^2\right] \\
\le& \frac{\zeta(K)^2}{N}\cdot\mathbb{E}\left[L_{U,h_0}^2(t-S_i)R^2(\boldsymbol{X})\right] \quad (\text{where } R^2(\boldsymbol{X}) := u_K^\top(\boldsymbol{X})u_K(\boldsymbol{X})) \\
=& O\left(\zeta(K)^2\frac{v_{h_0}(t)K}{N}\right),
\end{aligned}$$

where the last equality holds by Lemma 1 and Assumption 6. Thus, all the eigenvalues of

$$\frac{1}{N}\sum_{i=1}^{N} L_{U,h_0}(t-S_i)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top - \mathbb{E}\left[L_{U,h_0}(t-S_i)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]$$

is of the rate

$$O_P\left(\zeta(K)\sqrt{\frac{v_{h_0}(t)K}{N}}\right) . \tag{C.15}$$

Consider the event set

$$\begin{aligned} E_N :=& \Bigg\{(\lambda-\lambda_t^*)^\top \widehat{S}_N(\lambda-\lambda_t^*) \\ &> (\lambda-\lambda_t^*)^\top\left(\frac{\mathbb{E}[L_{h_0}\{t-T_i\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top]}{\mathbb{E}[L_{h_0}\{t-T_i\}]} - \frac{s_1}{2}I_K\right)(\lambda-\lambda_t^*), \lambda\neq\lambda_t^*\Bigg\}, \end{aligned}$$

then (C.13) and (C.14) imply that for any $\epsilon>0$, there exists $N_0(\epsilon)\in\mathbb{N}$ such that $N>N_0(\epsilon)$ large enough

$$\mathbb{P}\left\{(E_N)^c\right\} < \frac{\epsilon}{2} . \tag{C.16}$$

Note that λ_t^* is the unique maximizer of $G_t^*(\lambda)$, which implies

$$\nabla G_t^*(\lambda_t^*) = \frac{\mathbb{E}\left[\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L\{(t-T_i)/h_0\}u_K(\boldsymbol{X}_i)\right]}{\mathbb{E}\left[L\{(t-T_i)/h_0\}\right]} - \mathbb{E}\left[u_K(\boldsymbol{X}_i)\right] = 0.$$

Note that

$$\begin{aligned} \nabla\widehat{G}_t(\lambda_t^*) =& \frac{\sum_{i=1}^N \rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} - \frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i) \\ =& \frac{1}{N}\sum_{i=1}^N\left\{\frac{\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}}{\mathbb{E}[L_U\{(t-S_i)/h_0\}]} - 1\right\}u_K(\boldsymbol{X}_i) \\ &- \frac{\sum_{i=1}^N \rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} \\ &\times\left\{\frac{1}{N}\sum_{i=1}^N \frac{L_U\{(t-S_i)/h_0\}}{\mathbb{E}[L_U\{(t-S_i)/h_0\}]} - 1\right\}. \end{aligned}$$

From Assumption 6, we have that $\mathbb{E}(\|u_K(\boldsymbol{X})\|^2|T=t) \asymp K$ for all $t \in \mathcal{T}$. Then by $\mathbb{E}[\nabla G_t(\lambda_t^*)] = 0$, Lemmas 1 and 2, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}}{\mathbb{E}[L_U\{(t-S_i)/h_0\}]}-1\right\}u_K(\boldsymbol{X}_i)\right\|^2\right]\\
=&\frac{1}{N}\cdot\mathbb{E}\left[\left(\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}\cdot\frac{L_U\{(t-S_i)/h_0\}}{\mathbb{E}[L\{(t-T_i)/h_0\}]}-1\right)^2\cdot\|u_K(\boldsymbol{X}_i)\|^2\right]\\
\leq&\frac{2}{N}\cdot\frac{\mathbb{E}\left[\left\{\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}\right\}^2\cdot L_U^2\{(t-S_i)/h_0\}\cdot\|u_K(\boldsymbol{X}_i)\|^2\right]}{\mathbb{E}[L\{(t-T_i)/h_0\}]^2}+\frac{2}{N}\cdot\mathbb{E}\left[\|u_K(\boldsymbol{X}_i)\|^2\right]\\
=&O\left(\frac{Kv_{h_0}(t)}{N}\right)+O\left(\frac{K}{N}\right)=O\left(\frac{Kv_{h_0}(t)}{N}\right),
\end{aligned}\tag{C.17}$$

and

$$\begin{aligned}
&\frac{\sum_{i=1}^N\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}\cdot\left\{\frac{1}{N}\sum_{i=1}^N\frac{L_U\{(t-S_i)/h_0\}}{\mathbb{E}[L_U\{(t-S_i)/h_0\}]}-1\right\}\\
=&\{\mathbb{E}[u_K(\boldsymbol{X})]+o_P(1)\}\cdot O_P\left(\sqrt{\frac{v_{h_0}(t)}{N}}\right),
\end{aligned}$$

so that

$$\begin{aligned}
&\left\|\frac{\sum_{i=1}^N\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}\cdot\left\{\frac{1}{N}\sum_{i=1}^N\frac{L_U\{(t-S_i)/h_0\}}{\mathbb{E}[L_U\{(t-S_i)/h_0\}]}-1\right\}\right\|\\
=&O_P\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right).
\end{aligned}$$

Therefore, we have

$$\left\|\nabla\widehat{G}_t(\lambda_t^*)\right\|=O_P\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right),\tag{C.18}$$

then for any $t \in \mathcal{T}$ and every $\epsilon > 0$, there exists a constant $C_4 > 0$ such that

$$\mathbb{P}\left(\left\|\nabla\widehat{G}_t(\lambda_t^*)\right\|\geq C_4\sqrt{\frac{Kv_{h_0}(t)}{N}}\right)<\frac{\epsilon}{2}.$$

Let $\epsilon > 0$, fix some $C_5(\epsilon) > 0$ (to be chosen later) and define

$$\widehat{\Lambda}_t(\epsilon):=\left\{\lambda\in\mathbb{R}^K:\|\lambda-\lambda_t^*\|\leq C_5(\epsilon)C_4\sqrt{\frac{Kv_{h_0}(t)}{N}}\right\}.\tag{C.19}$$

For $\forall\lambda \in \widehat{\Lambda}_t(\epsilon)$, $\boldsymbol{x} \in \mathcal{X}$, and sufficiently large N, by Assumption 6 (ii), $\zeta(K)\sqrt{Kv_{h_0}(t)/N} \to 0$ for all $t \in \mathcal{T}$ as $N \to \infty$, (C.6) and (C.9), we have

$$
\begin{aligned}
&|\lambda^\top u_K(\boldsymbol{x}) - (\lambda_t^*)^\top u_K(\boldsymbol{x})| \le \|\lambda - \lambda_t^*\|\|u_K(\boldsymbol{x})\| \le C_5(\epsilon)C_4\sqrt{\frac{Kv_{h_0}(t)}{N}}\zeta(K) \\
\Rightarrow\ & \lambda^\top u_K(\boldsymbol{x}) \\
&\in \left[(\lambda_t^*)^\top u_K(\boldsymbol{x}) - C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{Kv_{h_0}(t)}{N}},\ (\lambda_t^*)^\top u_K(\boldsymbol{x}) + C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{Kv_{h_0}(t)}{N}}\right] \\
&\subset \Bigg[(\rho')^{-1}(\eta_1) - C_1K^{-\alpha} - C_2\{K^{-\alpha} + h_0^2\}\zeta(K) - C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{Kv_{h_0}(t)}{N}}, \\
&\qquad (\rho')^{-1}(\eta_2) + C_1K^{-\alpha} + C_2\{K^{-\alpha} + h_0^2\}\zeta(K) + C_5(\epsilon)C_4\zeta(K)\sqrt{\frac{Kv_{h_0}(t)}{N}}\Bigg] \\
&\subset \Gamma_2(\epsilon)\ ,
\end{aligned}
\tag{C.20}
$$

where

$$\Gamma_2(\epsilon) := [\underline{\gamma} - 1 - C_5(\epsilon), \overline{\gamma} + 1 + C_5(\epsilon)]$$

with $\underline{\gamma} := (\rho')^{-1}(\eta_1)$ and $\overline{\gamma} := (\rho')^{-1}(\eta_2)$, is a compact set independent of $\boldsymbol{x}$.

By the mean value theorem, for any $\lambda \in \partial\widehat{\Lambda}_t(\epsilon)$, there exists $\bar{\lambda}$ on the line joining λ and λ_t^* such that

$$\widehat{G}_t(\lambda) = \widehat{G}_t(\lambda_t^*) + (\lambda - \lambda_t^*)^\top \nabla\widehat{G}_t(\lambda_t^*) + \frac{1}{2}\cdot(\lambda - \lambda_t^*)^\top \nabla^2\widehat{G}_t(\bar{\lambda})(\lambda - \lambda_t^*)\ .$$

For the second order term in above expression, when N is large enough, we have

$$
\begin{aligned}
&(\lambda - \lambda_t^*)^\top \nabla^2\widehat{G}_t(\bar{\lambda})(\lambda - \lambda_t^*) \\
&= (\lambda - \lambda_t^*)^\top \frac{\sum_{i=1}^N \rho''\{\bar{\lambda}^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}(\lambda - \lambda_t^*) \\
&\le -\bar{b}(\epsilon)\cdot(\lambda - \lambda_t^*)^\top \frac{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}(\lambda - \lambda_t^*) \\
&= -\bar{b}(\epsilon)\cdot(\lambda - \lambda_t^*)^\top \widehat{S}_N(\lambda - \lambda_t^*),
\end{aligned}
\tag{C.21}
$$

where $-\bar{b}(\epsilon) := \sup_{\gamma\in\Gamma_2(\epsilon)} \rho''(\gamma) < \infty$ because $\Gamma_2(\epsilon)$ is compact and ρ'' is a continuous function. Then on the event E_N with large enough N, we have that for any $\lambda \in \partial\widehat{\Lambda}_t(\epsilon)$,

$$\widehat{G}_t(\lambda) - \widehat{G}_t(\lambda_t^*)$$

$$
\begin{aligned}
&= (\lambda - \lambda_t^*)^\top \nabla \widehat{G}_t(\lambda_t^*) + \frac{1}{2}(\lambda - \lambda_t^*)^\top \nabla^2 \widehat{G}_t(\bar{\lambda})(\lambda - \lambda_t^*) \\
&\leq \|\lambda - \lambda_t^*\| \|\nabla \widehat{G}_t(\lambda_t^*)\| - \frac{\bar{b}(\epsilon)}{2}(\lambda - \lambda_t^*)^\top \widehat{S}_N(\lambda - \lambda_t^*) \quad \text{(using (C.21))} \\
&\leq \|\lambda - \lambda_t^*\| \|\nabla \widehat{G}_t(\lambda_t^*)\| - \frac{\bar{b}(\epsilon)}{2}(\lambda - \lambda_t^*)^\top \left(\frac{\mathbb{E}[L_{U,h_0}(t - S_i) u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)]}{\mathbb{E}[L_{U,h_0}(t - S_i)]} - \frac{s_1}{2} I_K \right)(\lambda - \lambda_t^*) \\
&\leq \|\lambda - \lambda_t^*\| \|\nabla \widehat{G}_t(\lambda_t^*)\| - \frac{\bar{b}(\epsilon)}{2}(\lambda - \lambda_t^*)^\top \left(s_1 \cdot I_K - \frac{s_1}{2} I_K \right)(\lambda - \lambda_t^*) \quad \text{(using (C.14))} \\
&< \|\lambda - \lambda_t^*\| \left(\|\nabla \widehat{G}_t(\lambda_t^*)\| - \frac{\bar{b}(\epsilon)}{4} \cdot s_1 \cdot \|\lambda - \lambda_t^*\| \right). \qquad \text{(C.22)}
\end{aligned}
$$

By Chebyshev's inequality and (C.17), for sufficiently large N,

$$
\begin{aligned}
&\mathbb{P}\left\{ \|\nabla \widehat{G}_t(\lambda_t^*)\| \geq \frac{\bar{b}(\epsilon)}{4} \cdot s_1 \cdot \|\lambda - \lambda_t^*\| \right\} \\
\leq &\mathbb{P}\left\{ \|\nabla \widehat{G}_t(\lambda_t^*)\| \geq \frac{\bar{b}(\epsilon)}{4} \cdot s_1 \cdot \|\lambda - \lambda_t^*\|, \left\| \nabla \widehat{G}_t(\lambda_t^*) \right\| \leq C_4 \sqrt{\frac{K v_{h_0}(t)}{N}} \right\} \\
&+ \mathbb{P}\left(\left\| \nabla \widehat{G}_t(\lambda_t^*) \right\| > C_4 \sqrt{\frac{K v_{h_0}(t)}{N}} \right) \\
\leq &\mathbb{P}\left\{ C_4 \sqrt{\frac{K v_{h_0}(t)}{N}} \geq \frac{\bar{b}(\epsilon)}{4} \cdot s_1 \cdot C_5(\epsilon) \cdot C_4 \sqrt{\frac{K v_{h_0}(t)}{N}} \right\} + \frac{\epsilon}{2} \leq \frac{\epsilon}{2}, \qquad \text{(C.23)}
\end{aligned}
$$

where the last inequality holds by choosing $C_5(\epsilon) > \dfrac{4}{\bar{b}(\epsilon) s_1}$. Therefore, for sufficiently large N, by (C.16) and (C.23) we have that

$$
\begin{aligned}
&\mathbb{P}\left((E_N)^c \text{ or } \|\nabla \widehat{G}_t(\lambda_t^*)\| \geq \frac{\bar{b}(\epsilon)}{4} \cdot s_1 \cdot \|\lambda - \lambda_t^*\| \right) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \\
\Rightarrow &\mathbb{P}\left(E_N \text{ and } \|\nabla \widehat{G}_t(\lambda_t^*)\| < \frac{\bar{b}(\epsilon)}{4} \cdot s_1 \cdot \|\lambda - \lambda_t^*\| \right) > 1 - \epsilon \ . \qquad \text{(C.24)}
\end{aligned}
$$

Then by (C.22) and (C.24), we get

$$
\mathbb{P}\left\{ \widehat{G}_t(\lambda) - \widehat{G}_t(\lambda_t^*) < 0, \ \forall \lambda \in \partial \widehat{\Lambda}_t(\epsilon) \right\} \geq 1 - \epsilon,
$$

for sufficiently large N. Note that the event set $\{\widehat{G}_t(\lambda_t^*) > \widehat{G}_t(\lambda), \ \forall \lambda \in \partial \hat{\Lambda}_t(\epsilon)\}$ implies that there exists a local maximum point in the interior of $\widehat{\Lambda}_K(\epsilon)$. On the other hand, with probability approaching to one, $\widehat{G}_t$ is strictly concave function and $\widehat{\lambda}_t$ is the unique global maximum point of $\widehat{G}_t$, then we get

$$\mathbb{P}\left(\widehat{\lambda}_t \in \widehat{\Lambda}_t(\epsilon)\right) > 1-\epsilon, \tag{C.25}$$

i.e.

$$\left\|\widehat{\lambda}_t - \lambda_t^*\right\| = O_p\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right).$$

We next show that $\sup_{\boldsymbol{x}\in\mathcal{X}}|\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x})| = O_p\left(\zeta(K)\sqrt{Kv_{h_0}(t)/N}\right)$. By the mean value theorem, we have

$$\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x}) = \rho'\left(\widehat{\lambda}_t^\top u_K(\boldsymbol{x})\right) - \rho'\left((\lambda_t^*)^\top u_K(\boldsymbol{x})\right) = \rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)(\widehat{\lambda}_t - \lambda_t^*)^\top u_K(\boldsymbol{x}),$$

where $\tilde{\lambda}_t$ lies on the line joining $\widehat{\lambda}_t$ and λ_t^*. From (C.25) and (C.20), we have

$$\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\right| = O_p(1)\ , \tag{C.26}$$

therefore, we can obtain that

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in\mathcal{X}}|\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x})| \qquad\qquad (\text{C.27})\\
\leq &\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\right| \cdot \|\widehat{\lambda}_t - \lambda_t^*\| \cdot \sup_{x\in\mathcal{X}}\|u_K(\boldsymbol{x})\| = O_p\left(\zeta(K)\sqrt{\frac{Kv_{h_0}(t)}{N}}\right).
\end{aligned}$$

By the mean value theorem, (C.26), the fact $\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right] = I_K$, and $\left\|\widehat{\lambda}_t - \lambda_t^*\right\| = O_p\left(\sqrt{Kv_{h_0}(t)/N}\right)$ we have

$$\begin{aligned}
&\int_{\mathcal{X}}|\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x})|^2 dF_X(\boldsymbol{x})\\
=&\int_{\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\cdot(\widehat{\lambda}_t - \lambda_t^*)^\top u_K(\boldsymbol{x})\right|^2 dF_X(\boldsymbol{x})\\
\leq&\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\right|^2\cdot(\widehat{\lambda}_t - \lambda_t^*)^\top\cdot\int_{\mathcal{X}}u_K(\boldsymbol{x})u_K(\boldsymbol{x})^\top dF_X(x)\cdot(\widehat{\lambda}_t - \lambda_t^*)\\
=&\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\right|^2\cdot\|\widehat{\lambda}_t - \lambda_t^*\|^2\\
=&O_p(1)\cdot O_p\left(\frac{Kv_{h_0}(t)}{N}\right) = O_p\left(\frac{Kv_{h_0}(t)}{N}\right).
\end{aligned} \tag{C.28}$$

Similarly, we obtain that

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}|\widehat{\pi}(t,\boldsymbol{X}_i)-\pi^*(t,\boldsymbol{X}_i)|^2\\
=&\frac{1}{N}\sum_{i=1}^{N}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{X}_i)\right)\cdot(\widehat{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\right|^2\\
\leq&\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\right|^2\cdot(\widehat{\lambda}_t-\lambda_t^*)^\top\cdot\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top\right\}(\widehat{\lambda}_t-\lambda_t^*)\\
\leq&\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\rho''\left(\tilde{\lambda}_t^\top u_K(\boldsymbol{x})\right)\right|^2\cdot\|\widehat{\lambda}_t-\lambda_t^*\|^2\cdot\lambda_{\max}\left\{\frac{1}{N}\sum_{i=1}^{N}u_K(\boldsymbol{X}_i)u_K(\boldsymbol{X}_i)^\top\right\}\\
\leq&O_p(1)\cdot O_p\left(\frac{Kv_{h_0}(t)}{N}\right)\cdot O_p(1)=O_p\left(\frac{Kv_{h_0}(t)}{N}\right).
\end{aligned}
\tag{C.29}
$$

D. Proof of Theorems 4.2 and 4.4

In this section, we prove Theorem 4.2, the asymptotic behaviour of $\widehat{\mu}(t)$ when U is ordinary smooth, and Theorem 4.4, the asymptotic behaviour of $\widehat{\mu}(t)$ when U is supersmooth. The difference in these two theorems comes from the different asymptotic variance of the deconvolution kernel, reflected in Lemma 1.

Proof. Define

$$\mu^*(t) := \frac{\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_U\{(t-S)/h\}]}{\mathbb{E}[L_U\{(t-S)/h\}]}.$$

Note from (8) that

$$\mathbb{E}[L_U\{(t-S)/h\}] = \mathbb{E}\big(\mathbb{E}[L_U\{(t-S)/h\}|T]\big) = \mathbb{E}[L\{(t-T)/h\}]\,,$$

and

$$\begin{aligned}\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_U\{(t-S)/h\}] =&\mathbb{E}\big(\pi_0(t,\boldsymbol{X})Y\cdot\mathbb{E}[L_U\{(t-S)/h\}|\boldsymbol{X},Y,T]\big)\\ =&\mathbb{E}[\pi_0(t,\boldsymbol{X})YL\{(t-T)/h\}]\,. \end{aligned}\tag{D.1}$$

Then

$$\mu^*(t) = \frac{\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_U\{(t-S)/h\}]}{\mathbb{E}[L_U\{(t-S)/h\}]} = \frac{\mathbb{E}[\pi_0(t,\boldsymbol{X})YL\{(t-T)/h\}]}{\mathbb{E}[L\{(t-T)/h\}]}.$$

We have the following *bias-variance* decomposition:

$$\widehat{\mu}(t)-\mu(t) = \{\widehat{\mu}(t)-\mu^*(t)\} + \{\mu^*(t)-\mu(t)\} =: A_N(t) + B_{N,1}(t),$$

where $B_{N,1}(t)$ contributes to a part of the asymptotic bias. The most difficult part of the proof is dealing $A_N(t)$ as it involves both the bias and the variance arisen from the nonparametric estimation of $\pi_0(t,\cdot)$.

We first consider $B_{N,1}(t)$:

$$\begin{aligned}B_{N,1}(t) =&\mu^*(t)-\mu(t) = \frac{\mathbb{E}[\pi_0(t,\boldsymbol{X})YL\{(t-T)/h\}]/h}{\mathbb{E}[L\{(t-T)/h\}]/h} - \mu(t)\\ =&\big\{\mathbb{E}[\pi_0(t,\boldsymbol{X})YL\{(t-T)/h\}]/h - \mu(t)\cdot f_T(t)\big\}\big(\mathbb{E}[L\{(t-T)/h\}]/h\big)^{-1}\\ &-\mu(t)\cdot\big\{\mathbb{E}[L\{(t-T)/h\}]/h - f_T(t)\big\}\big(\mathbb{E}[L\{(t-T)/h\}]/h\big)^{-1}.\end{aligned}$$

By Taylor's expansion and Assumption 4 (i), we have

$$\mathbb{E}[L\{(t-T)/h\}]/h = f_T(t) + \frac{h^2}{2}\partial_t^2 f_T(t)\kappa_{21} + O(h^3)\,,\tag{D.2}$$

where $\kappa_{21} = \int u^2 L(u)du$. Under Assumptions 4 (i) and (ii), we have

$$
\begin{aligned}
&\mathbb{E}[\pi_0(t,\boldsymbol{X})YL\{(t-T)/h\}]/h - \mu(t)\cdot f_T(t) \qquad\qquad \text{(D.3)}\\
=&\frac{1}{h}\int \pi_0(t,\boldsymbol{x})yL\left(\frac{t-t'}{h}\right)f_{T|Y,\boldsymbol{X}}(t'|y,\boldsymbol{x})f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})d\boldsymbol{x}dydt' - \mu(t)\cdot f_T(t)\\
=&\int \pi_0(t,\boldsymbol{x})yL(u)f_{T|Y,\boldsymbol{X}}(t-uh|y,\boldsymbol{x})f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})d\boldsymbol{x}dydu - \mu(t)\cdot f_T(t)\\
=&\int \pi_0(t,\boldsymbol{x})yf_{T|Y,\boldsymbol{X}}(t|y,\boldsymbol{x})f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})\,d\boldsymbol{x}\,dy - \mu(t)\cdot f_T(t)\\
&+\frac{h^2}{2}f_T(t)\kappa_{21}\mathbb{E}\left\{\frac{Y\partial_t^2 f_{T|Y,\boldsymbol{X}}(t|Y,\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})}\right\}\\
&+\frac{h^3}{6}f_T(t)\mathbb{E}\left\{\frac{Y}{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})}\int \partial_t^3 f_{T|Y,\boldsymbol{X}}(t+\xi uh|Y,\boldsymbol{X})u^3L(u)du\right\}\\
=&\frac{h^2}{2}f_T(t)\kappa_{21}\mathbb{E}\left\{\frac{Y\partial_t^2 f_{T|Y,\boldsymbol{X}}(t|Y,\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})}\right\}+O(h^3),
\end{aligned}
$$

for some $\xi \in (0,1)$. Thus, we have

$$
\begin{aligned}
B_{N,1}(t) =&\left[\frac{h^2}{2}f_T(t)\kappa_{21}\mathbb{E}\left\{\frac{Y\partial_t^2 f_{T|Y,\boldsymbol{X}}(t|Y,\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})}\right\}+O(h^3)\right]\{f_T(t)+O(h^2)\}^{-1}\\
&-\mu(t)\left\{\frac{h^2}{2}\partial_t^2 f_T(t)\kappa_{21}+O(h^3)\right\}\{f_T(t)+O(h^2)\}^{-1}\\
=&\frac{\kappa_{21}}{2}\left[\frac{f_T(t)\Phi(t)-\mu(t)\partial_t^2 f_T(t)}{f_T(t)}\right]h^2+O(h^3)\,, \qquad\qquad \text{(D.4)}
\end{aligned}
$$

where $\Phi(t) := \mathbb{E}\big[\{Y\partial_t^2 f_{T|Y,\boldsymbol{X}}(t|Y,\boldsymbol{X})\}/\{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})\}\big]$.

Regarding $A_N(t)$, we need the following decomposition and notations. The first order condition for maximising (15) implies

$$
\nabla\widehat{G}_t(\widehat{\lambda}_t) = \frac{\sum_{i=1}^N \rho'\{\widehat{\lambda}_t^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} - \left\{\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\right\} = 0,
$$

with probability approching to one. Then using the mean value theorem,

$$
\frac{\sum_{i=1}^N \rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} - \left\{\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\right\}
$$

$$= -\left[\frac{\sum_{i=1}^N \rho''\{(\widetilde{\lambda}_t)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}\right]\left\{\widehat{\lambda}_t - \lambda_t^*\right\},$$

where $\widetilde{\lambda}_t$ lies between $\widehat{\lambda}_t$ and λ_t^*, and λ_t^* is defined at (C.2), which gives

$$\begin{aligned}
\widehat{\lambda}_t - \lambda_t^* =& -\left[\frac{\sum_{i=1}^N \rho''\{(\widetilde{\lambda}_t)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}\right]^{-1} \\
&\times\left[\frac{\sum_{i=1}^N \rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} - \left\{\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\right\}\right] \\
=& -\widetilde{\Sigma}_t^{-1}\cdot\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\left[\frac{\pi^*(t,\boldsymbol{X}_i)L_U\{(t-S_i)/h_0\}}{N^{-1}\sum_{i=1}^N L_U\{(t-S_i)/h_0\}} - 1\right], \qquad \text{(D.5)}
\end{aligned}$$

with

$$\widetilde{\Sigma}_t := \frac{\sum_{i=1}^N \rho''\{(\widetilde{\lambda}_t)^\top u_K(\boldsymbol{X}_i)\}L_U\{(t-S_i)/h_0\}u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}.$$

Define

$$\Sigma_t := \frac{\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_{h_0}(t-T_i)u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)\right]}{\mathbb{E}\left[L_{h_0}(t-T_i)\right]},$$

where $L_{h_0}(t-t') := h_0^{-1}L\{(t-t')/h_0\}$.

Note that by Lemmas 2 and 4 of Fan and Truong (1993), we have, under Assumptions 2 and 4 (i), and Assumption O for ordinary smooth U or Assumption S for supersmooth U,

$$\frac{1}{Nh}\sum_{i=1}^N L_U\{(t-S_i)/h\} = f_T(t)\{1+o_P(1)\}. \qquad \text{(D.6)}$$

Then, combining this with (D.1) and (D.3), we have

$$\begin{aligned}
&A_N(t) = \widehat{\mu}(t) - \mu^*(t) \\
=&\frac{\sum_{i=1}^N \widehat{\pi}(t,\boldsymbol{X}_i)Y_iL_U\{(t-S_i)/h\}}{\sum_{i=1}^N L_U\{(t-S_i)/h\}} - \frac{\mathbb{E}[\pi_0(t,\boldsymbol{X}_i)Y_iL_U\{(t-S_i)/h\}]}{\mathbb{E}[L_U\{(t-S_i)/h\}]} \\
=&\left\{\frac{1}{Nh}\sum_{i=1}^N L_U\{(t-S_i)/h\}\right\}^{-1}
\end{aligned}$$

$$
\begin{aligned}
&\times\left\{\frac{1}{Nh}\sum_{i=1}^{N}\widehat{\pi}(t,\boldsymbol{X}_i)Y_iL_U\{(t-S_i)/h\}-\frac{1}{h}\cdot\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_U\{(t-S)/h\}]\right\}\\
&-\frac{1}{h}\cdot\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_U\{(t-S)/h\}]\cdot\left\{\frac{1}{Nh}\sum_{i=1}^{N}L_U\{(t-S_i)/h\}\right\}^{-1}\\
&\quad\times\left\{\frac{1}{h}\cdot\mathbb{E}[L_U\{(t-S)/h\}]\right\}^{-1}\\
&\quad\times\left\{\frac{1}{Nh}\sum_{i=1}^{N}L_U\{(t-S_i)/h\}-\frac{1}{h}\cdot\mathbb{E}[L_U\{(t-S)/h\}]\right\}\\
=&\,[f_T(t)\{1+o_p(1)\}]^{-1}\\
&\quad\times\left\{\frac{1}{Nh}\sum_{i=1}^{N}\widehat{\pi}(t,\boldsymbol{X}_i)Y_iL_U\{(t-S_i)/h\}-\frac{1}{h}\cdot\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_U\{(t-S)/h\}]\right\}\\
&-\{\mu(t)+o_P(1)\}\cdot\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{L_U\{(t-S_i)/h\}}{\mathbb{E}[L_U\{(t-S)/h\}]}-1\right\}\\
=&\frac{1}{N}\sum_{i=1}^{N}\left\{\widehat{\pi}(t,\boldsymbol{X}_i)Y_i\cdot\frac{L_{U,h}(t-S_i)}{f_T(t)}-\frac{\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_{U,h}(t-S)]}{f_T(t)}\right\}\cdot\{1+o_p(1)\}
\end{aligned}
\tag{D.7}
$$

$$
-\mu(t)\cdot\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{L_{U,h}(t-S_i)-\mathbb{E}[L_{U,h}(t-S)]}{f_T(t)}\right\}\cdot\{1+o_P(1)\}. \tag{D.8}
$$

We decompose the numerator of (D.7) as follows

$$
\frac{1}{N}\sum_{i=1}^{N}\{\widehat{\pi}(t,\boldsymbol{X}_i)Y_iL_{U,h}(t-S_i)-\mathbb{E}[\pi_0(t,\boldsymbol{X})YL_{U,h}(t-S)]\} \tag{D.9}
$$

$$
\begin{aligned}
=&\frac{1}{N}\sum_{i=1}^{N}\pi_0(t,\boldsymbol{X}_i)Y_i\cdot L_{U,h}(t-S_i)-\mathbb{E}\left[\pi_0(t,\boldsymbol{X}_i)Y_i\cdot L_{U,h}(t-S_i)\right]\\
&+\frac{1}{Nh}\sum_{i=1}^{N}\left\{\{\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\}Y_iL_U\left(\frac{t-S_i}{h}\right)\right.\\
&\qquad\qquad\left.-\mathbb{E}\left[\{\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\}Y_iL_U\left(\frac{t-S_i}{h}\right)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{h}\cdot\mathbb{E}\left[\{\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\}Y_iL_U\left(\frac{t-S_i}{h}\right)\right]\\
&+\frac{1}{Nh}\sum_{i=1}^N\left\{(\widehat{\pi}(t,\boldsymbol{X}_i)-\pi^*(t,\boldsymbol{X}_i))\,Y_i\cdot L_U\left(\frac{t-S_i}{h}\right)\right.\\
&\qquad\left.-\int_{\mathcal{T}}\int_{\mathcal{X}}(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x}))\,m(t',\boldsymbol{x})\cdot L\left(\frac{t-t'}{h}\right)dF_{X,T}(\boldsymbol{x},t')\right\}\\
&+\int_{\mathcal{T}}\int_{\mathcal{X}}(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x}))\,\{m(t',\boldsymbol{x})L_h\left(t-t'\right)-m(t,\boldsymbol{x})L_{h_0}(t-t')\}dF_{X,T}(\boldsymbol{x},t')\\
&+\int_{\mathcal{T}}\int_{\mathcal{X}}(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x}))\,m(t,\boldsymbol{x})L_{h_0}(t-t')dF_{X,T}(\boldsymbol{x},t') \qquad \text{(D.10)}
\end{aligned}
$$

Now, noting that $\widehat{\pi}(t,\boldsymbol{x})=\rho'\{\widehat{\lambda}_t^\top u_K(\boldsymbol{x})\}$ and $\pi^*(t,\boldsymbol{x})=\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\}$, applying the mean value theorem, we have

$$
\text{(D.10)}=\breve{\Psi}_t^\top(\widehat{\lambda}_t-\lambda_t^*)\,,
$$

where $\breve{\Psi}_t:=\int_{\mathcal{T}}\int_{\mathcal{X}}m(t,\boldsymbol{x})\rho''\left\{(\breve{\lambda}_t)^\top u_K(\boldsymbol{x})\right\}u_K(\boldsymbol{x})L_{h_0}(t-t')dF_{X,T}(\boldsymbol{x},t')$, and $\breve{\lambda}_t$ lies between $\widehat{\lambda}_t$ and λ_t^*.

Letting $\Psi_t:=\int_{\mathcal{T}}\int_{\mathcal{X}}m(t,\boldsymbol{x})\rho''\left\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\right\}u_K(\boldsymbol{x})L_{h_0}(t-t')dF_{X,T}(\boldsymbol{x},t')$, and recalling (D.5), we have

$$
\begin{aligned}
\text{(D.10)}&=-\breve{\Psi}_t^\top\widetilde{\Sigma}_t^{-1}\cdot\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\left[\frac{\pi^*(t,\boldsymbol{X}_i)L_U\{(t-S_i)/h_0\}}{N^{-1}\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}-1\right]\\
&=-\left\{\breve{\Psi}_t^\top\widetilde{\Sigma}_t^{-1}-\Psi_t^\top\Sigma_t^{-1}\right\}\cdot\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\left[\frac{\pi^*(t,\boldsymbol{X}_i)L_U\{(t-S_i)/h_0\}}{N^{-1}\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}-1\right]\\
&\quad-\left\{\Psi_t^\top\Sigma_t^{-1}\cdot\frac{1}{N}\sum_{i=1}^N u_K(\boldsymbol{X}_i)\frac{\pi^*(t,\boldsymbol{X}_i)L_U\{(t-S_i)/h_0\}}{N^{-1}\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}\right. \qquad \text{(D.11)}\\
&\qquad\left.-\frac{1}{N}\sum_{i=1}^N m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i)\right\}\\
&\quad+\Psi_t^\top\Sigma_t^{-1}\cdot\frac{1}{N}\sum_{i=1}^N\Big[u_K(\boldsymbol{X}_i)-\mathbb{E}\{u_K(\boldsymbol{X})\}\Big]\\
&\quad+\Psi_t^\top\Sigma_t^{-1}\mathbb{E}\{u_K(\boldsymbol{X})\}-\mathbb{E}\{m(t,\boldsymbol{X})\}\mathbb{E}\{L_{U,h_0}(t-S)\}\\
&\quad+\mu(t)\mathbb{E}\{L_{U,h_0}(t-S)\}-\mathbb{E}\Big\{m(t,\boldsymbol{X})\pi_0(t,\boldsymbol{X})L_{U,h_0}(t-S)\Big\}
\end{aligned}
$$

$$
-\frac{1}{N}\sum_{i=1}^{N}\Big[m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i)-\mathbb{E}\big\{m(t,\boldsymbol{X})\pi_0(t,\boldsymbol{X})L_{U,h_0}(t-S)\big\}\Big]\ ,
$$

Let

$$
J(\boldsymbol{X};t):=\Psi_t^{\top}\Sigma_t^{-1}u_K(\boldsymbol{X})\pi^*(t,\boldsymbol{X})-m(t,\boldsymbol{X})\mathbb{E}[L_{h_0}(t-T)]\pi_0(t,\boldsymbol{X})
$$

and using (D.6), we further decompose (D.11) as

$$
\begin{aligned}
\text{(D.11)} =& -\frac{1}{N}\sum_{i=1}^{N}\left(\frac{J(\boldsymbol{X}_i;t)L_{U,h_0}(t-S_i)}{N^{-1}\sum_{i=1}^{N}L_{U,h_0}(t-S_i)}-\frac{\mathbb{E}\Big[J(\boldsymbol{X};t)L_{U,h_0}(t-S)\Big]}{\mathbb{E}\{L_{h_0}(t-T)\}}\right)\\
&-\frac{1}{N}\sum_{i=1}^{N}\Bigg[\frac{m(t,\boldsymbol{X}_i)\mathbb{E}[L_{h_0}(t-T_i)]\pi_0(t,\boldsymbol{X}_i)\cdot L_{U,h_0}(t-S_i)}{N^{-1}\sum_{i=1}^{N}L_{U,h_0}\{t-S_i\}}\\
&\qquad\qquad -m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i)\Bigg]\\
&-\frac{\mathbb{E}\Big[J(\boldsymbol{X};t)L_{U,h_0}(t-S)\Big]}{\mathbb{E}\{L_{h_0}(t-T)\}}\\
=& -\frac{N^{-1}\sum_{i=1}^{N}J(\boldsymbol{X}_i;t)L_{U,h_0}(t-S_i)-\mathbb{E}\{J(\boldsymbol{X};t)L_{U,h_0}(t-S)\}}{N^{-1}\sum_{i=1}^{N}L_{U,h_0}(t-S_i)}\\
&-\frac{\mathbb{E}[J(\boldsymbol{X};t)L_{U,h_0}(t-S)]}{N^{-1}\sum_{i=1}^{N}L_{U,h_0}(t-S_i)}\left[\frac{\mathbb{E}\{L_{U,h_0}(t-S)\}-N^{-1}\sum_{i=1}^{N}L_{U,h_0}(t-S_i)}{\mathbb{E}\{L_{U,h_0}(t-S)\}}\right]\\
&-\frac{1}{N}\sum_{i=1}^{N}\Bigg[\frac{m(t,\boldsymbol{X}_i)\mathbb{E}[L_{h_0}(t-T_i)]\pi_0(t,\boldsymbol{X}_i)\cdot L_{U,h_0}(t-S_i)}{N^{-1}\sum_{i=1}^{N}L_{U,h_0}\{t-S_i\}}\\
&\qquad\qquad -m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i)\Bigg]\\
&-\frac{\mathbb{E}\Big[J(\boldsymbol{X};t)L_{U,h_0}(t-S)\Big]}{\mathbb{E}\{L_{h_0}(t-T)\}}\\
=& -\frac{N^{-1}\sum_{i=1}^{N}J(\boldsymbol{X}_i;t)L_{U,h_0}(t-S_i)-\mathbb{E}\{J(\boldsymbol{X};t)L_{U,h_0}(t-S)\}}{f_T(t)}\cdot\{1+o_P(1)\}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Bigg[\frac{m(t,\boldsymbol{X}_i)\mathbb{E}[L_{h_0}(t-T_i)]\pi_0(t,\boldsymbol{X}_i)\cdot L_{U,h_0}(t-S_i)}{N^{-1}\sum_{i=1}^{N}L_{U,h_0}\{t-S_i\}}\\
&\qquad\qquad -m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i)\Bigg]
\end{aligned}
$$

$$- \frac{\mathbb{E}[J(\boldsymbol{X};t)L_{U,h_0}(t-S)]}{\mathbb{E}\{L_{U,h_0}(t-S)\}} \cdot \{1+o_P(1)\}\,.$$

Now, combining (D.7) to the decomposition of (D.11), we have

$$A_N(t) = \{B_{N,2}(t) + V_N(t)\} \cdot \{1+o_P(1)\}\,,$$

where $B_{N,2}(t)$ gethers the non-random terms that contribute to the bias arisen from the nonparametric estimation of $\pi_0(t,\cdot)$ and $V_N(t)$ contributes to the variance of our estimator as follows: note from (8) that

$$\begin{aligned}
&B_{N,2}(t) \\
&:= f_T(t)^{-1}\mathbb{E}\Big[\{\pi^*(t,\boldsymbol{X}) - \pi_0(t,\boldsymbol{X})\}YL_h(t-T)\Big] && \text{(D.12)}\\
&+ f_T(t)^{-1}\Big[\Psi_t^\top\Sigma_t^{-1}\mathbb{E}\{u_K(\boldsymbol{X})\} - \mathbb{E}\{m(t,\boldsymbol{X})\}\mathbb{E}\{L_{h_0}(t-T)\}\Big] && \text{(D.13)}\\
&+ f_T(t)^{-1}\Big[\mu(t)\mathbb{E}\{L_{h_0}(t-T)\} - \mathbb{E}\{m(t,\boldsymbol{X})\pi_0(t,\boldsymbol{X})L_{h_0}(t-T)\}\Big] && \text{(D.14)}\\
&- \frac{\mathbb{E}\Big[J(\boldsymbol{X};t)L_{h_0}(t-T)\Big]}{f_T(t)\cdot\mathbb{E}\{L_{h_0}(t-T)\}}\,, && \text{(D.15)}
\end{aligned}$$

and

$$V_N(t) := f_T(t)^{-1}\left[\sqrt{\frac{v_h(t)}{N}}\{V_{N,1}(t)+V_{N,2}(t)\}+\sqrt{\frac{v_{h_0}(t)}{N}}V_{N,3}(t)+\sqrt{\frac{\{v_h(t)\vee v_{h_0}(t)\}}{N}}V_{N,4}(t)\right],$$

with

$$V_{N,1}(t) := -\sqrt{N/v_h(t)}\mu(t)\cdot\frac{1}{N}\sum_{i=1}^N\{L_{U,h}\{t-S_i\} - \mathbb{E}[L_{U,h}\{t-S\}]\}\,,$$

$$V_{N,2}(t) := \sqrt{N/v_h(t)} \times \frac{1}{N}\sum_{i=1}^N\Big\{\pi_0(t,\boldsymbol{X}_i)Y_i\cdot L_{U,h}(t-S_i) - \mathbb{E}\left[\pi_0(t,\boldsymbol{X})Y\cdot L_{U,h}(t-S)\right]\Big\},$$

$$\begin{aligned}
V_{N,3}(t) := -\sqrt{\frac{1}{Nv_{h_0}(t)}}\sum_{i=1}^N\Bigg[&\frac{m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)\mathbb{E}[L_{h_0}(t-T)]\cdot L_{U,h_0}(t-S_i)}{N^{-1}\sum_{i=1}^N L_{U,h_0}\{t-S_i\}}\\
&- m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i)\Bigg] \qquad \text{(D.16)}
\end{aligned}$$

$$-\sqrt{\frac{1}{Nv_{h_0}(t)}}\sum_{i=1}^N\Big\{m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)L_{U,h_0}(t-S_i) - \mathbb{E}\big[m(t,\boldsymbol{X})\pi_0(t,\boldsymbol{X})L_{U,h_0}(t-S)\big]\Big\}, \qquad \text{(D.17)}$$

and

$$V_{N,4}(t) := \sqrt{N/(v_h(t) \vee v_{h_0}(t))} \times \frac{1}{Nh} \sum_{i=1}^{N} \Bigg\{ (\widehat{\pi}(t, \boldsymbol{X}_i) - \pi^*(t, \boldsymbol{X}_i))\, Y_i \cdot L_U \left(\frac{t - S_i}{h} \right)$$

$$- \int_{\mathcal{T}} \int_{\mathcal{X}} (\widehat{\pi}(t, \boldsymbol{x}) - \pi^*(t, \boldsymbol{x}))\, m(t', \boldsymbol{x}) \cdot L \left(\frac{t - t'}{h} \right) dF_{X,T}(\boldsymbol{x}, t') \Bigg\} \quad \text{(D.18)}$$

$$+ \sqrt{N/(v_h(t) \vee v_{h_0}(t))} \times \frac{1}{Nh} \sum_{i=1}^{N} \Bigg\{ \{\pi^*(t, \boldsymbol{X}_i) - \pi_0(t, \boldsymbol{X}_i)\} Y_i L_U \left(\frac{t - S_i}{h} \right)$$

$$- \mathbb{E} \left[\{\pi^*(t, \boldsymbol{X}_i) - \pi_0(t, \boldsymbol{X}_i)\} Y_i L_U \left(\frac{t - S_i}{h} \right) \right] \Bigg\} \quad \text{(D.19)}$$

$$+ \sqrt{N/(v_h(t) \vee v_{h_0}(t))} \times \Bigg\{ \int_{\mathcal{T}} \int_{\mathcal{X}} (\widehat{\pi}(t, \boldsymbol{x}) - \pi^*(t, \boldsymbol{x}))$$

$$\times \{m(t', \boldsymbol{x}) L_h(t - t') - m(t, \boldsymbol{x}) L_{h_0}(t - t')\} dF_{X,T}(\boldsymbol{x}, t') \Bigg\} \quad \text{(D.20)}$$

$$- \frac{1}{\sqrt{N(v_h(t) \vee v_{h_0}(t))}} \sum_{i=1}^{N} \left\{ \breve{\Psi}_t^\top \widetilde{\Sigma}_t^{-1} - \Psi_t^\top \Sigma_t^{-1} \right\} u_K(\boldsymbol{X}_i) \left[\frac{\pi^*(t, \boldsymbol{X}_i) L_U\{(t - S_i)/h_0\}}{N^{-1} \sum_{i=1}^{N} L_U\{(t - S_i)/h_0\}} - 1 \right] \quad \text{(D.21)}$$

$$- \sqrt{N/(v_h(t) \vee v_{h_0}(t))} \left(\frac{N^{-1} \sum_{i=1}^{N} J(\boldsymbol{X}_i; t) L_{U,h_0}(t - S_i) - \mathbb{E}\Big[J(\boldsymbol{X}; t) L_{U,h_0}(t - S) \Big]}{f_T(t)} \right) \quad \text{(D.22)}$$

$$+ \frac{1}{\sqrt{N(v_h(t) \vee v_{h_0}(t))}} \sum_{i=1}^{N} \Big[\Psi_t^\top \Sigma_t^{-1} \{u_K(\boldsymbol{X}_i) - \mathbb{E}[u_K(\boldsymbol{X})]\} \Big]. \quad \text{(D.23)}$$

For the bias term $B_{N,2}(t)$**:** We first decompose (D.14). As in (D.2) and (D.3), we have

$$f_T(t)^{-1} \mu(t) \mathbb{E}\{L_{h_0}(t - T)\} = \mu(t) + \frac{h_0^2 \kappa_{21}}{2} \frac{\partial_t^2 f_T(t) \cdot \mu(t)}{f_T(t)} + O(h_0^3)$$

and

$$f_T(t)^{-1} \mathbb{E}\{m(t, \boldsymbol{X}) \pi_0(t, \boldsymbol{X}) L_{h_0}(t - T)\} = \mu(t) + \frac{h_0^2 \kappa_{21}}{2} \mathbb{E}\left\{ \frac{m(t, \boldsymbol{X}) \partial_t^2 f_{T|\boldsymbol{X}}(t|\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})} \right\} + O(h_0^3).$$

Thus,

$$\text{(D.14)} = \frac{\kappa_{21}}{2} \left[\frac{\mu(t) \partial_t^2 f_T(t)}{f_T(t)} - \mathbb{E}\left\{ \frac{m(t, \boldsymbol{X}) \partial_t^2 f_{T|\boldsymbol{X}}(t|\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t|\boldsymbol{X})} \right\} \right] \cdot h_0^2 + O(h_0^3).$$

We next decompose (D.15) as

$$
\begin{aligned}
&-\frac{\mathbb{E}\Big[J(\boldsymbol{X};t)L_{h_0}(t-T)\Big]}{f_T(t)\cdot\mathbb{E}\{L_{h_0}(t-T)\}}\\
&=-\frac{\mathbb{E}\Big[m(t,\boldsymbol{X})\mathbb{E}[L_{h_0}(t-T)]\{\pi^*(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}L_{h_0}(t-T)\Big]}{f_T(t)\cdot\mathbb{E}\{L_{h_0}(t-T)\}} &\text{(D.24)}\\
&-\frac{\mathbb{E}\Big[\big\{\Psi_t^\top\Sigma_t^{-1}u_K(\boldsymbol{X})-m(t,\boldsymbol{X})\mathbb{E}[L_{h_0}(t-T)]\big\}\cdot\{\pi^*(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}L_{h_0}(t-T)\Big]}{f_T(t)\cdot\mathbb{E}\{L_{h_0}(t-T)\}} &\text{(D.25)}\\
&-\frac{\mathbb{E}\Big[\big\{\Psi_t^\top\Sigma_t^{-1}u_K(\boldsymbol{X})-m(t,\boldsymbol{X})\mathbb{E}[L_{h_0}(t-T)]\big\}\pi_0(t,\boldsymbol{X})L_{h_0}(t-T)\Big]}{f_T(t)\cdot\mathbb{E}\{L_{h_0}(t-T)\}} &\text{(D.26)}
\end{aligned}
$$

Using (8) and Taylor's expansion, we have

$$
\begin{aligned}
\text{(D.12)} =&f_T(t)^{-1}\mathbb{E}\left[m(t,\boldsymbol{X})\cdot\left\{\frac{1}{h}L\left(\frac{t-T}{h}\right)\right\}\cdot(\pi^*(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X}))\right]\\
=&\mathbb{E}[m(t,\boldsymbol{X})\{\pi^*(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}|T=t]\cdot\{1+O(h^2)\}\,,
\end{aligned}
$$

and

$$
\text{(D.24)}=-\mathbb{E}[m(t,\boldsymbol{X})\{\pi^*(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}|T=t]\cdot\{1+O(h_0^2)\}\,.
$$

Thus, using Cauchy-Schwarz inequality and Lemma 2,

$$
\text{(D.12)+(D.24)}=\mathbb{E}[m(t,\boldsymbol{X})\{\pi^*(t,\boldsymbol{X})-\pi_0(t,\boldsymbol{X})\}|T=t]\cdot\{O(h_0^2+h^2)\}=O\{(K^{-\alpha}+h_0^2)(h_0^2+h^2)\}\,.
$$

Note that $\rho''(v)<0$ for all $v\in\mathbb{R}$, then $-L_{h_0}(t-T)\rho''[(\lambda_t^*)^\top u_K(\boldsymbol{X})]=L_{h_0}(t-T)|\rho''[(\lambda_t^*)^\top u_K(\boldsymbol{X})]|>0$. We consider the following weighted least square projection of $m(t,\boldsymbol{X})$ on the space linearly spanned by $u_K(\boldsymbol{X})$:

$$
\gamma_t^*:=\arg\min_{\gamma\in\mathbb{R}^K}\mathbb{E}\left[\big\{-L_{h_0}(t-T)\rho''[(\lambda_t^*)^\top u_K(\boldsymbol{X})]\big\}\big\{m(t,\boldsymbol{X})-\gamma^\top u_K(\boldsymbol{X})\big\}^2\right], \tag{D.27}
$$

which gives

$$
\begin{aligned}
\gamma_t^*=&\mathbb{E}\left[L_{h_0}(t-T)\rho''[(\lambda_t^*)^\top u_K(\boldsymbol{X})]u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})\right]^{-1}\\
&\cdot\mathbb{E}\left[L_{h_0}(t-T)\rho''[(\lambda_t^*)^\top u_K(\boldsymbol{X})]m(t,\boldsymbol{X})u_K(\boldsymbol{X})\right]\\
=&\Psi_t^\top\Sigma_t^{-1}/\mathbb{E}[L_{h_0}(t-T)],
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi_t&=\int_{\mathcal{T}}\int_{\mathcal{X}}m(t,\boldsymbol{x})\rho''\left\{(\lambda_t^*)^\top u_K(\boldsymbol{x})\right\}u_K(\boldsymbol{x})L_{h_0}(t-t')dF_{X,T}(\boldsymbol{x},t'),\\
\Sigma_t&=\frac{\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}L_{h_0}(t-T_i)u_K(\boldsymbol{X}_i)u_K^\top(\boldsymbol{X}_i)\right]}{\mathbb{E}\left[L_{h_0}(t-T_i)\right]}.
\end{aligned}
$$

Therefore, we have that $(\gamma_t^*)^\top u_K(\boldsymbol{X}) = \Psi_t^\top \Sigma_t^{-1} u_K(\boldsymbol{X})/\mathbb{E}[L_{h_0}(t-T)]$ is the $L^2\{L_{h_0}(t-t')\big|\rho''[(\lambda_t^*)^\top u_K(x)]\big|dF_{T,X}(t',x)\}$-projection of $m(t,\boldsymbol{X})$ on the space linearly spanned by $u_K(\boldsymbol{X})$ and under Assumption 7, the approximation rate is $O(K^{-\ell})$. Then, using (8), Cauchy-Schwarz inequality and Lemma 2, we have

$$\text{(D.25)} = O\{K^{-\ell}\cdot(K^{-\alpha}+h_0^2)\}\,.$$

Finally, using Taylor's expansion and equation (10), we have

$$\begin{aligned}\text{(D.26)} =& -\mathbb{E}\Big[\{(\gamma_t^*)^\top u_K(\boldsymbol{X}) - m(t,\boldsymbol{X})\}\pi_0(t,\boldsymbol{X})|T=t\Big]\cdot\{1+O(h_0^2)\}\\ =& -\mathbb{E}\Big[(\gamma_t^*)^\top u_K(\boldsymbol{X}) - m(t,\boldsymbol{X})\Big]\cdot\{1+O(h_0^2)\}\,,\end{aligned}$$

and

$$\text{(D.13)} = \mathbb{E}\Big[(\gamma_t^*)^\top u_K(\boldsymbol{X}) - m(t,\boldsymbol{X})\Big]\cdot\{1+O(h_0^2)\}\,.$$

Then, (D.26)+(D.13) $= \mathbb{E}\Big[(\gamma_t^*)^\top u_K(\boldsymbol{X}) - m(t,\boldsymbol{X})\Big]\cdot O(h_0^2) = O(K^{-\ell}\cdot h_0^2)$. Therefore, we have

$$B_{N,2}(t) = O\{(K^{-\ell}+h_0^2+h^2)\cdot(K^{-\alpha}+h_0^2)\} + \text{(D.14)} = \text{(D.14)} + o(h^2),$$

given that $(K^{-\ell}+h_0^2)\cdot(K^{-\alpha}+h_0^2) = o(h^2)$.

For the variance term $V_N(t)$**:** By the definition of $v_h(t)$ in (B.1), using Lemma 1 and Assumption 8, we have $\mathrm{var}\{V_{N,1}(t)+V_{N,2}(t)\}\asymp 1$. Since $\mathbb{E}\{V_{N,1}(t)+V_{N,2}(t)\}=0$, we have $V_{N,1}(t)+V_{N,2}(t)\asymp 1$ with probability approaching 1. For $V_{N,3}(t)$, note that

$$\begin{aligned}\text{(D.16)} =& \frac{\sum_{i=1}^N m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)\cdot L_{U,h_0}(t-S_i)}{\sum_{i=1}^N L_{U,h_0}\{t-S_i\}}\\ &\times\sqrt{\frac{1}{Nv_{h_0}(t)}}\sum_{i=1}^N\{L_{U,h_0}(t-S_i)-\mathbb{E}[L_{h_0}(t-T_i)]\}\\ =&\mu(t)\cdot\sqrt{\frac{1}{Nv_{h_0}(t)}}\sum_{i=1}^N\{L_{U,h_0}\{t-S_i\}-\mathbb{E}[L_{h_0}(t-T_i)]\}\cdot\{1+o_P(1)\},\end{aligned}$$

where the last equality comes from the consistency result of standard deconvolution kernel regression. Now, $\mathbb{E}\{L_{U,h_0}(t-S_i)-\mathbb{E}[L_{h_0}(t-T_i)]\}=0$ by (8). Under Assumption 8 and using Lemma 1, we have

$$\mathbb{E}\Bigg(\Bigg[\mu(t)\cdot\{L_{U,h_0}\{t-S_i\}-\mathbb{E}[L_{h_0}(t-T)]\}$$

$$- \Big\{ m(t, \boldsymbol{X}_i)\pi_0(t, \boldsymbol{X}_i)L_{U,h_0}(t - S_i) - \mathbb{E}\big[m(t, \boldsymbol{X})\pi_0(t, \boldsymbol{X}_i)L_{U,h_0}(t - S_i)\,\big]\Big\}\Big]^2\Big)$$

$$= \mathrm{var}[\{\mu(t) - \pi_0(t, \boldsymbol{X})m(t, \boldsymbol{X})\}L_{U,h_0}(t - S)] \asymp v_{h_0}(t).$$

Thus, $V_{N,3}(t) \asymp 1$ with probability approaching 1. We shall show that $V_{N,4}(t) = o_P(1)$ in Subsection D.3 under the following assumption

$$\frac{v_{h_0}(t)}{\sqrt{v_h(t) \vee v_{h_0}(t)}} \frac{K}{\sqrt{N}} \to 0\,. \tag{D.28}$$

Then, we have

$$V_N(t) = \frac{1}{N \cdot f_T(t)} \sum_{i=1}^{N} \eta_{h,h_0}(S_i, \boldsymbol{X}_i, Y_i; t) \cdot \{1 + o_P(1)\}\,,$$

where for $i = 1, \ldots, N$, $\eta_{h,h_0}(S_i, \boldsymbol{X}_i, Y_i; t) = \phi_h(S_i, \boldsymbol{X}_i, Y_i; t) + \psi_{h_0}(S_i, \boldsymbol{X}_i, Y_i; t)$, with

$$\begin{aligned}\phi_h(S_i, \boldsymbol{X}_i, Y_i; t) = &[\pi_0(t, \boldsymbol{X}_i)Y_iL_{U,h}(t - S_i) - \mathbb{E}\{\pi_0(t, \boldsymbol{X})YL_{U,h}(t - S)\}] \\ &- \mu(t)\,[L_{U,h}(t - S_i) - \mathbb{E}\{L_{U,h}(t - S)\}]\,,\end{aligned}$$

and

$$\begin{aligned}&\psi_{h_0}(S_i, \boldsymbol{X}_i, Y_i; t) = \mu(t)\,[L_{U,h_0}(t - S_i) - \mathbb{E}\{L_{U,h_0}(t - S)\}] \\ &\quad - \Big[m(t, \boldsymbol{X}_i)\pi_0(t, \boldsymbol{X}_i) \cdot L_{U,h_0}(t - S_i) - \mathbb{E}\{m(t, \boldsymbol{X})\pi_0(t, \boldsymbol{X})L_{U,h_0}(t - S)\}\Big].\end{aligned}$$

D.1. Asymptotic distribution of $V_N(t)$ in ordinary smooth case

Using Lemma 1, we have, if U is ordinary smooth and Assumption O holds, (D.28) is equivalent to

$$\frac{(h \wedge h_0)^{1/2+\beta}}{h_0^{1+2\beta}} \frac{K}{\sqrt{N}} \to 0$$

and

$$\mathrm{var}\left\{\frac{1}{N}\sum_{i=1}^{N} f_T^{-1}(t)\phi_h(S_i, \boldsymbol{X}_i, Y_i; t)\right\} = \frac{1}{Nh^{1+2\beta}} \cdot f_T^{-2}(t)(R_1^2 f_T) * f_U(t) \cdot C\{1 + o(1)\}\,,$$

and

$$\mathrm{var}\left\{\frac{1}{N}\sum_{i=1}^{N} f_T^{-1}(t)\psi_{h_0}(S_i, \boldsymbol{X}_i, Y_i; t)\right\} = \frac{1}{Nh_0^{1+2\beta}} \cdot f_T^{-2}(t)(R_2^2 f_T) * f_U(t) \cdot C\{1 + o(1)\}\,,$$

where $R_1^2(t) = \mathbb{E}[\{\pi_0(t,\boldsymbol{X})Y - \mu(t)\}^2|T=t]$, $R_2^2(t) = \mathbb{E}[\{\pi_0(t,\boldsymbol{X})m(t,\boldsymbol{X}) - \mu(t)\}^2|T=t]$ and $C = \int_{-\infty}^{\infty} J^2(v)\,dv = (2\pi c^2)^{-1}\int_{-\infty}^{\infty}|w|^{2\beta}\phi_L^2(w)dw$.

For the case $h \asymp h_0$, suppose that $h_0 = \tilde{c}h$ for some constant $\tilde{c} > 0$. We have

$$\begin{aligned}\mathbb{E}\{\eta_{h,h_0}^2(S,\boldsymbol{X},Y;t)\} =& \mathbb{E}\Big\{\phi_h^2(S,\boldsymbol{X},Y;t) + \psi_{h_0}^2(S,\boldsymbol{X},Y;t)\\ &+ 2\phi_h(S,\boldsymbol{X},Y;t)\psi_{h_0}(S,\boldsymbol{X},Y;t)\Big\},\end{aligned}$$

where using Lemma 1 (a),

$$\mathbb{E}\{\phi_h^2(S,\boldsymbol{X},Y;t)\} = \frac{1}{h^{1+2\beta}}\cdot(R_1^2 f_T) * f_U(t)\cdot\int_{-\infty}^{\infty} J^2(v)\,dv\{1+o(1)\} \tag{D.29}$$

and

$$\mathbb{E}\{\psi_{h_0}^2(S,\boldsymbol{X},Y;t)\} = \frac{1}{\tilde{c}^{1+2\beta}h^{1+2\beta}}\cdot(R_2^2 f_T) * f_U(t)\cdot\int_{-\infty}^{\infty} J^2(v)\,dv\{1+o(1)\}\,. \tag{D.30}$$

For the cross term, note that

$$\begin{aligned}&\mathbb{E}\Big\{\phi_h(S,\boldsymbol{X},Y;t)\psi_{h_0}(S,\boldsymbol{X},Y;t)\Big\}\\ &= \mathbb{E}[\{\pi_0(t,\boldsymbol{X})Y - \mu(t)\}\{\mu(t) - \pi_0(t,\boldsymbol{X})m(t,\boldsymbol{X})\}L_{U,h}(t-S)L_{U,h_0}(t-S)]\\ &\quad - \mathbb{E}[\{\pi_0(t,\boldsymbol{X})Y - \mu(t)\}L_{U,h}(t-S)]\cdot\mathbb{E}[\{\mu(t) - \pi_0(t,\boldsymbol{X})m(t,\boldsymbol{X})\}L_{U,h_0}(t-S)]\,.\end{aligned}$$

For the second term, using Lemma 1, we have

$$\mathbb{E}[\{\pi_0(t,\boldsymbol{X})Y - \mu(t)\}L_{U,h}(t-S)]\cdot\mathbb{E}[\{\mu(t) - \pi_0(t,\boldsymbol{X})m(t,\boldsymbol{X})\}L_{U,h_0}(t-S)] = o(1)\,.$$

For the first term, letting $L_U'(v)$ be the same as (6) except that h is replaced by h_0, we have

$$\begin{aligned}&\mathbb{E}[\{\pi_0(t,\boldsymbol{X})Y - \mu(t)\}\{\mu(t) - \pi_0(t,\boldsymbol{X})m(t,\boldsymbol{X})\}L_{U,h}(t-S)L_{U,h_0}(t-S)]\\ &= \mathbb{E}[(R_1R_2)(T)L_{U,h}(t-S)L_{U,h_0}(t-S)]\\ &= \frac{1}{\tilde{c}h^2}\int(R_1R_2)(t')L_U\left(\frac{t-t'-u}{h}\right)L_U'\left(\frac{t-t'-u}{\tilde{c}h}\right)f_T(t')f_U(u)\,dt'\,du\\ &= \frac{1}{\tilde{c}h}\int(R_1R_2)(t-u-zh)L_U(z)\,L_U'(z/\tilde{c})\,f_T(t-u-zh)f_U(u)\,dz\,du\,.\end{aligned}$$

Note that similar to the arguments in the proof of Lemma 1 (a), we have

$$h^\beta L_U(z) \to \frac{1}{2\pi}\int\exp(-iwz)\phi_L(w)\frac{w^\beta}{c}\,dw =: J(z)\,,$$

$$h_0^{\beta} L_U'(z/\tilde{c}) \to \frac{1}{2\pi} \int \exp(-iwz/\tilde{c}) \phi_L(w) \frac{w^{\beta}}{c}\, dw =: J(z/\tilde{c})\,,$$

and

$$\begin{aligned}
&\frac{1}{\tilde{c}h} \int (R_1R_2)(t-u-zh) L_U(z)\, L_U'(z/\tilde{c})\, f_T(t-u-zh) f_U(u)\, dz\, du \\
&= \frac{1}{\tilde{c}^{1+\beta} h^{1+2\beta}} \{(R_1R_2) f_T\} * f_U(t) \cdot \int_{-\infty}^{\infty} J(v) J(v/\tilde{c})\, dv \cdot \{1+o(1)\}\,.
\end{aligned}$$

Note that by a change of variable,

$$\int J^2(v/\tilde{c})\, dv = \tilde{c} \int J^2(v)\, dv\,.$$

Combining with (D.29) and (D.30), we have

$$\begin{aligned}
&\mathbb{E}\{\eta^2_{h,h_0}(S,\boldsymbol{X},Y;t)\} \qquad\qquad \text{(D.31)}\\
&= \frac{1}{h^{1+2\beta}} \cdot \Bigg\{ (R_1^2 f_T) * f_U(t) \cdot \int_{-\infty}^{\infty} J^2(v)\, dv \\
&\quad + \frac{1}{\tilde{c}^{2+2\beta}} (R_2^2 f_T) * f_U(t) \cdot \int_{-\infty}^{\infty} J^2(v/\tilde{c})\, dv \\
&\quad + \frac{2}{\tilde{c}^{1+\beta}} \{(R_1R_2) f_T\} * f_U(t) \cdot \int_{-\infty}^{\infty} J(v) J(v/\tilde{c})\, dv \Bigg\} + o(1)\,.
\end{aligned}$$

Using Cauchy-Schwarz inequality, we have

$$\left| \int_{-\infty}^{\infty} J(v) J(v/\tilde{c})\, dv \right| \le \sqrt{\tilde{c}} \int_{-\infty}^{\infty} J^2(v)\, dv\,.$$

Note that for all $t \in \mathcal{T}$,

$$\begin{aligned}
&R_1^2(t) + \tilde{c}^{-(1+2\beta)} R_2^2(t) - 2\tilde{c}^{-(1/2+\beta)} |(R_1R_2)(t)| \\
=&\mathbb{E}\left[|\pi_0(t,\boldsymbol{X})Y - \mu(t)|^2 | T=t \right] + \tilde{c}^{-(1+2\beta)} \mathbb{E}\left[|\pi_0(t,\boldsymbol{X})\mathbb{E}[Y|T=t,\boldsymbol{X}] - \mu(t)|^2 | T=t \right] \\
&- 2\tilde{c}^{-(1/2+\beta)} \cdot \left| \mathbb{E}\left[\{\pi_0(t,\boldsymbol{X})Y - \mu(t)\}\{\mu(t) - \pi_0(t,\boldsymbol{X})\mathbb{E}[Y|T=t,\boldsymbol{X}]\} \right] \right| \\
\ge&\mathbb{E}\left[|\pi_0(t,\boldsymbol{X})Y - \mu(t)|^2 | T=t \right] + \tilde{c}^{-(1+2\beta)} \mathbb{E}\left[|\pi_0(t,\boldsymbol{X})\mathbb{E}[Y|T=t,\boldsymbol{X}] - \mu(t)|^2 | T=t \right] \\
&- 2\tilde{c}^{-(1/2+\beta)} \cdot \mathbb{E}\left[|\pi_0(t,\boldsymbol{X})Y - \mu(t)| \cdot |\mu(t) - \pi_0(t,\boldsymbol{X})\mathbb{E}[Y|T=t,\boldsymbol{X}]| \right] \\
=&\mathbb{E}\left[\left\{ |\pi_0(t,\boldsymbol{X})Y - \mu(t)| - \tilde{c}^{-(1/2+\beta)} \cdot |\pi_0(t,\boldsymbol{X})\mathbb{E}[Y|T=t,\boldsymbol{X}] - \mu(t)| \right\}^2 \middle| T=t \right] > 0,
\end{aligned}$$

we have that $\mathbb{E}\{\eta^2_{h,h_0}(S, \boldsymbol{X}, Y; t)\}$is strictly larger than 0. Then, we have

$$\begin{aligned}
&\operatorname{var}\left[\frac{1}{N}\sum_{i=1}^{N} f_T^{-1}(t)\{\phi_h(S_i, \boldsymbol{X}_i, Y_i; t) + \psi_{\tilde{c}h}(S_i, \boldsymbol{X}_i, Y_i; t)\}\right] \\
&= \frac{1}{Nh^{1+2\beta}} \cdot V_3 \cdot \{1 + o(1)\}\,.
\end{aligned}$$

Using the same arguments as in Lemma 1 (a), we have $\mathbb{E}\{L^{2+\delta}_{U,h}(t - S)\} = O(h^{-(1+2\beta)(2+\delta)/2-\delta/2})$. Then, under Assumption 8, we have

$$\frac{\mathbb{E}\{|\eta_{h,h_0}(S, \boldsymbol{X}, Y; t)|^{2+\delta}\}}{N^{\delta/2}[\operatorname{var}\{\eta_{h,h_0}(S, \boldsymbol{X}, Y; t)\}]^{(2+\delta)/2}} \asymp \frac{O\{(h \wedge h_0)^{-(1+2\beta)(2+\delta)/2-\delta/2}\}}{N^{\delta/2}(h \wedge h_0)^{-(1+2\beta)(2+\delta)/2}} \to 0\,,$$

as $N \to \infty$. Thus, by Lyapunov central limit theorem, the results in Theorem 4.2 follow.

D.2. Asymptotic distribution of $V_N(t)$ in supersmooth case

Note that $v_{h_0}(t)/\sqrt{v_h(t) \vee v_{h_0}(t)} \le \sqrt{v_h(t) \vee v_{h_0}(t)}$. Using Lemma 1 and recalling the definition of $e(h)$ in Theorem 4.4, we have, if U is supersmooth and Assumption S holds, then as long as

$$\frac{K}{\{e(h) \wedge e(h_0)\}\sqrt{N}} \to 0$$

holds, we have (D.28) holds. Moreover, using Lemma 1, if $v_h(t) \to \infty$ as $h \to 0$, then

$$\operatorname{var}\left\{\frac{1}{N}\sum_{i=1}^{N} f_T^{-1}(t)\phi_h(S_i, \boldsymbol{X}_i, Y_i; t)\right\} = O\{N^{-1}e(h)^{-2}\}\,,$$

and

$$\operatorname{var}\left\{\frac{1}{N}\sum_{i=1}^{N} f_T^{-1}(t)\psi_{h_0}(S_i, \boldsymbol{X}_i, Y_i; t)\right\} = O\{N^{-1}e(h_0)^{-2}\}\,.$$

Thus,

$$\frac{1}{N}\sum_{i=1}^{N} f_T^{-1}(t)\eta_{h,h_0}(S_i, \boldsymbol{X}_i, Y_i; t) = O\{N^{-1/2}\{e(h) \wedge e(h_0)\}^{-1}\}\,.$$

To conclude the proof of Theorem 4.4, we check the Lyapunov condition. That is, to show that

$$\frac{\mathbb{E}\{|\eta_{h,h_0}(S, \boldsymbol{X}, Y; t)|^{2+\delta}\}}{N^{\delta/2}[\operatorname{var}\{\eta_{h,h_0}(S, \boldsymbol{X}, Y; t)\}]^{(2+\delta)/2}} \to 0\,,$$

as $N \to \infty$. Under Assumption 8, we see that

$$\mathbb{E}\{|\eta_{h,h_0}(S,\boldsymbol{X},Y;t)|^{2+\delta}\} = O\big[\mathbb{E}\{|L_{U,h}|^{2+\delta}\} + \mathbb{E}\{|L_{U,h_0}|^{2+\delta}\}\big]\,.$$

By the definition of $L_{U,h}$, under Assumptions S and (19), we can see that

$$|L_{U,h}(v)| = O\{h^{-c_U}\exp(h^{-\beta}/\gamma)\}\,,$$

where c_U is a constant depending on β_0:

$$c_U = \begin{cases} 1, & \text{if } \beta_0 \geq 0 \\ 1-\beta_0, & \text{if } \beta_0 < 0\,. \end{cases}$$

Thus, we have

$$\mathbb{E}\{|\eta_{h,h_0}(S,\boldsymbol{X},Y;t)|^{2+\delta}\} = O\left[h^{-c_U(2+\delta)}\exp\left(\frac{2+\delta}{\gamma h^{\beta}}\right) + h_0^{-c_U(2+\delta)}\exp\left(\frac{2+\delta}{\gamma h_0^{\beta}}\right)\right]\,.$$

For the denominator, we first note that using Lemma 1 and Assumption 8,

$$\operatorname{var}\{\phi_h(S,\boldsymbol{X},Y;t)\} \asymp v_h(t)\,,$$
$$\operatorname{var}\{\psi_{h_0}(S,\boldsymbol{X},Y;t)\} \asymp v_{h_0}(t)\,.$$

If either one of them is a dominating term, then we have in the denominator that

$$\operatorname{var}\{\eta_{h,h_0}(S,\boldsymbol{X},Y;t)\} \asymp v_h(t) \vee v_{h_0}(t)\,.$$

However, when $v_h(t) \asymp v_{h_0}(t)$, we need to take the covariance between $\phi_h(S,\boldsymbol{X},Y;t)$ and $\psi_{h_0}(S,\boldsymbol{X},Y;t)$ into account.

To see when $v_h(t) \asymp v_{h_0}(t)$, we let $h_0 = c_N h$ for some $c_N > 0$ that may depend on N. Then under the assumption in Theorem 4.4 and Lemma 1, we have

$$h^{d_3}\exp(2h^{-\beta}/\gamma - d_2h^{-d_4\beta}) \preceq v_h(t) \preceq h^{-1}\exp(2h^{-\beta}/\gamma),$$
$$c_N^{d_3}h^{d_3}\exp(2c_N^{-\beta}h^{-\beta}/\gamma - d_2c_N^{-d_4}h^{-d_4\beta}) \preceq v_{h_0}(t) \preceq c_N^{-1}h^{-1}\exp(2c_N^{-\beta}h^{-\beta}/\gamma),$$

where d_1, d_2, d_3 and d_4 are defined in the statement of Theorem 4.4.

Noting that $d_1, d_2 > 0$ and $1 > d_4 > 0$, we have,

(a) if $c_N > 1$, noting that $0 < d_4 < 1$, we have $0 < 1 - c_N^{-\beta} \asymp 1$ and

$$\frac{v_h(t)}{v_{h_0}(t)} \succeq c_N h^{1+d_3}\exp\{2(1-c_N^{-\beta})h^{-\beta}/\gamma - d_2h^{-d_4\beta}\} \to \infty \text{ as } h \to 0\,;$$

(b) if $c_N < 1$, then $0 > 1 - c_N^{-\beta} \asymp -c_N^{-\beta}$ and

$$\frac{v_h(t)}{v_{h_0}(t)} \preceq c_N^{-d_3} h^{-(1+d_3)} \exp\{2(1-c_N^{-\beta})h^{-\beta}/\gamma + d_2 c_N^{-d_4\beta} h^{-d_4\beta}\} \to 0 \text{ as } h \to 0\,;$$

(c) if $c_N = 1$, then $v_h(t)/v_{h_0}(t) = 1$.

Therefore, $v_h(t) \asymp v_{h_0}(t)$ only $h = h_0$. In that case,

$$\mathrm{var}\{\eta_{h,h_0}(S,\boldsymbol{X},Y;t)\} = \mathrm{var}[\pi_0(t,\boldsymbol{X})\{Y - m(t,\boldsymbol{X})\}L_{U,h}(t)]\,,$$

which, under Assumption 8 and using Lemma 1, is of the rate of $v_h(t)$. Then, under any condition of (a), (b) and (c) above, we have

$$\mathrm{var}\{\eta_{h,h_0}(S,\boldsymbol{X},Y;t)\} \asymp v_{h\wedge h_0}(t)\,.$$

Thus, we have

$$\begin{aligned} &\frac{\mathbb{E}\{|\eta_{h,h_0}(S,\boldsymbol{X},Y;t)|^{2+\delta}\}}{N^{\delta/2}[\mathrm{var}\{\eta_{h,h_0}(S,\boldsymbol{X},Y;t)\}]^{(2+\delta)/2}} \\ &= O\left[\frac{(h\wedge h_0)^{-(c_U+d_3/2)(2+\delta)}\exp\{(h\wedge h_0)^{-d_4\beta}(1+\delta/2)d_2\}}{N^{\delta/2}}\right]. \end{aligned}$$

Note that the condition

$$\frac{K}{\{e(h)\wedge e(h_0)\}\sqrt{N}} \to 0$$

implies that $(h\wedge h_0)^{-\delta/2}\exp\{\delta\cdot(h\wedge h_0)^{-\beta}/\gamma\} = o(N^{\delta/2})$ for any positive δ. Since $0 < d_4 < 1$, we have

$$(h\wedge h_0)^{-(c_U+d_3/2)(2+\delta)}\exp\{(h\wedge h_0)^{-d_4\beta}(1+\delta/2)d_2\} = o[(h\wedge h_0)^{-\delta/2}\exp\{\delta\cdot(h\wedge h_0)^{-\beta}/\gamma\}]\,.$$

Then,

$$\frac{\mathbb{E}\{|\eta_{h,h_0}(S,\boldsymbol{X},Y;t)|^{2+\delta}\}}{N^{\delta/2}[\mathrm{var}\{\eta_{h,h_0}(S,\boldsymbol{X},Y;t)\}]^{(2+\delta)/2}} \to 0,$$

so the Lyapunov condition is satisfied.

D.3. The negligible term $V_{N,4}(t)$

Now, we show that

$$V_{N,4}(t) = o_P(1)\,. \tag{D.32}$$

For the term (D.18). Let

$$\mu_N\{g(S_i, \boldsymbol{X}_i, Y_i)\} := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{g(S_i, \boldsymbol{X}_i, Y_i) - \mathbb{E}[g(S_i, \boldsymbol{X}_i, Y_i)]\}$$

be the empirical process indexed by the function $g(\cdot)$. Then (D.18) $= \mu_N\left[M_{\widehat{\pi}}(S_i, \boldsymbol{X}_i, Y_i)\right]$, where

$$M_\pi(S_i, \boldsymbol{X}_i, Y_i) := \{\pi(t, \boldsymbol{X}_i) - \pi^*(t, \boldsymbol{X}_i)\}\, Y_i \cdot \frac{1}{\sqrt{v_h(t) \vee v_{h_0}(t)}} L_{U,h}\left(t - S_i\right).$$

We apply Theorem 3 of Chen et al. (2003) to show (D.18) is of $o_P(1)$ by verifying their conditions hold. Note that

$$\begin{aligned}
&\mathbb{E}\left(\sup_{\|\pi_1 - \pi_2\|_\infty < \delta} \left| M_{\pi_1}(S_i, \boldsymbol{X}_i, Y_i) - M_{\pi_2}(S_i, \boldsymbol{X}_i, Y_i) \right|^2\right) \\
=&\frac{1}{\{v_h(t) \vee v_{h_0}(t)\}} \mathbb{E}\left(\sup_{\|\pi_1 - \pi_2\|_\infty < \delta} \{\pi_1(t, \boldsymbol{X}_i) - \pi_2(t, \boldsymbol{X}_i)\}^2\, Y_i^2 \cdot L_{U,h}^2\left(t - S_i\right)\right) \\
\leq&\delta^2 \cdot \frac{O\{v_h(t)\}}{\{v_h(t) \vee v_{h_0}(t)\}} \leq O(\delta^2) \quad \text{(by Lemma 1 and Assumption 8).}
\end{aligned}$$

Thus, condition (3.2) of Theorem 3 of Chen et al. (2003) is satisified. Under Assumption 5, $\pi_0(t, \cdot) \in C_M^s(\mathcal{X})$, the set of continuous function defined in Van Der Vaart et al. (1996, Chapter 2.7). Then, by Van Der Vaart et al. (1996, Theorem 2.7.1), the $\|\cdot\|_\infty$-covering number of $C_M^s(\mathcal{X})$ using ϵ-balls, $N(\epsilon, C_M^s(\mathcal{X}), \|\cdot\|_\infty)$, satisfies that

$$\log N(\epsilon, C_M^s(\mathcal{X}), \|\cdot\|_\infty) \leq \text{const.} \times \epsilon^{-\frac{r}{s}}.$$

Under Assumption 5 that $s > r/2$, we have

$$\int_0^\infty \sqrt{\log N(\epsilon, C_M^s(\mathcal{X}), \|\cdot\|_\infty)}\, d\epsilon < \infty.$$

Then the last condition (3.3) of Chen et al. (2003, Theorem 3) holds, which gives that (D.18) is of $o_P(1)$.

For term (D.19)**:** Using Lemmas 1 and 2, we compute the second moment of (D.19) to get

$$\mathbb{E}[|\text{(D.19)}|^2]$$

$$
\begin{aligned}
\leq & \frac{1}{hv_h(t)} \cdot \mathbb{E}\Bigg[\frac{1}{h}\Bigg\{\{\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\}Y_i L_U\left(\frac{t-S_i}{h}\right) \\
& \qquad - \mathbb{E}\left[\{\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\}Y_i L_U\left(\frac{t-S_i}{h}\right)\right]\Bigg\}^2\Bigg] \\
\leq & \frac{1}{hv_h(t)} \cdot \mathbb{E}\left[\frac{1}{h}\{\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\}^2 Y_i^2 L_U^2\left(\frac{t-S_i}{h}\right)\right] \\
\leq & \frac{1}{hv_h(t)} \cdot O\left\{\zeta(K)^2(K^{-2\alpha}+h_0^4)\right\} \cdot O\left\{hv_h(t)\right\} = O\left\{\zeta(K)^2(K^{-2\alpha}+h_0^4)\right\}.
\end{aligned}
$$

Hence, (D.19) $= O_P\left\{\zeta(K)(K^{-\alpha}+h_0^2)\right\} = o_P(1)$ by Chebyshev's inequality.
For term (D.20)**:** Note that

$$
\begin{aligned}
& \int_{\mathcal{T}}\int_{\mathcal{X}} \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) m(t',\boldsymbol{x}) L_h\left(t-t'\right) dF_{X,T}(\boldsymbol{x},t') \\
= & \int_{\mathcal{T}}\int_{\mathcal{X}} \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) \int y f_{Y|T,X}(y|t',\boldsymbol{x}) L_h\left(t-t'\right) dF_{X,T}(\boldsymbol{x},t') \\
= & \int \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) y L_h\left(t-t'\right) f_{T|Y,\boldsymbol{X}}(t'|y,\boldsymbol{x}) f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})\, d\boldsymbol{x} dy dt'.
\end{aligned}
$$

Using the Taylor's expansion as in (D.3), we have

$$
\begin{aligned}
& \int_{\mathcal{T}}\int_{\mathcal{X}} \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) m(t',\boldsymbol{x}) L_h\left(t-t'\right) dF_{X,T}(\boldsymbol{x},t') \\
= & \int \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) y f_{T|Y,\boldsymbol{X}}(t|y,\boldsymbol{x}) f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})\, d\boldsymbol{x} dy \\
& + \frac{h^2}{2}\kappa_{21}\int_{\mathcal{X}} \mathbb{E}\{Y\partial_t^2 f_{T|Y,\boldsymbol{X}}(t|Y,\boldsymbol{X})|\boldsymbol{X}=\boldsymbol{x}\}\left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \\
& + \frac{h^3}{6}\int_{\mathcal{X}} \mathbb{E}\Bigg\{Y\int \partial_t^3 f_{T|Y,\boldsymbol{X}}(t+\xi uh|Y,\boldsymbol{X}) u^3 L(u)\, du|\boldsymbol{X}=\boldsymbol{x}\Bigg\}\left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x},
\end{aligned}
$$

and

$$
\begin{aligned}
& \int \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) y f_{T|Y,\boldsymbol{X}}(t|y,\boldsymbol{x}) f_{Y,\boldsymbol{X}}(y,\boldsymbol{x})\, d\boldsymbol{x} dy \\
= & \int_{\mathcal{X}} \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) m(t,\boldsymbol{x}) f_{X,T}(\boldsymbol{x},t)\, d\boldsymbol{x}.
\end{aligned}
$$

Similarly, using Taylor's expansion, we have

$$
\int_{\mathcal{T}}\int_{\mathcal{X}} \left(\widehat{\pi}(t,\boldsymbol{x})-\pi^*(t,\boldsymbol{x})\right) m(t,\boldsymbol{x}) L_{h_0}\left(t-t'\right) dF_{X,T}(\boldsymbol{x},t')
$$

$$
\begin{aligned}
&= \int \left(\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x})\right) m(t,\boldsymbol{x}) f_{X,T}(\boldsymbol{x},t)\, d\boldsymbol{x} \\
&\quad + \frac{h_0^2}{2}\kappa_{21} \int_{\mathcal{X}} \partial_t^2 f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) \left(\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x})\right) m(t,\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \\
&\quad + \frac{h_0^3}{6} \int_{\mathcal{X}} \left\{ \int \partial_t^3 f_{T|\boldsymbol{X}}(t+\xi uh|\boldsymbol{x}) u^3 L(u)\, du \right\} \left(\widehat{\pi}(t,\boldsymbol{x}) - \pi^*(t,\boldsymbol{x})\right) m(t,\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \,.
\end{aligned}
$$

Under Assumptions 4 and 6, and Lemma 3, we have

$$
\begin{aligned}
\text{(D.20)} =& \sqrt{\frac{N}{v_h(t) \vee v_{h_0}(t)}} \times O_p\left\{ \sqrt{\frac{K v_{h_0}(t)}{N}} \right\} \times O(h^2 + h_0^2) \\
=& O_p\{\sqrt{K}(h^2 + h_0^2)\} = O_p\{\zeta(K)(h^2 + h_0^2)\} = o_P(1).
\end{aligned}
$$

For term (D.21)**:** Note that

$$
\begin{aligned}
\text{(D.21)} =& -\left\{ \breve{\Psi}_t^\top \widetilde{\Sigma}_t^{-1} - \Psi_t^\top \Sigma_t^{-1} \right\} \\
&\times \sqrt{\frac{1}{N\{v_h(t) \vee v_{h_0}(t)\}}} \sum_{i=1}^N u_K(\boldsymbol{X}_i) \left[\frac{\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\} L_U\{(t-S_i)/h\}}{N^{-1}\sum_{i=1}^N L_U\{(t-S_i)/h\}} - 1 \right] \\
&= (\Psi_t^\top \Sigma_t^{-1} - \breve{\Psi}_t^\top \widetilde{\Sigma}_t^{-1}) \sqrt{N/\{v_h(t) \vee v_{h_0}(t)\}} \cdot \nabla \widehat{G}_t(\lambda_t^*) \\
&= (\Psi_t - \breve{\Psi}_t)^\top \widetilde{\Sigma}_t^{-1} \sqrt{N/\{v_h(t) \vee v_{h_0}(t)\}} \cdot \nabla \widehat{G}_t(\lambda_t^*) \\
&\quad + \Psi_t^\top (\Sigma_t^{-1} - \widetilde{\Sigma}_t^{-1}) \sqrt{N/\{v_h(t) \vee v_{h_0}(t)\}} \cdot \nabla \widehat{G}_t(\lambda_t^*) \ . \qquad \text{(D.33)}
\end{aligned}
$$

Consider the first term in (D.33). By the mean value theorem and Lemma 3 we have

$$
\begin{aligned}
\|\Psi_t - \breve{\Psi}_t\| =& \left\| -\int m(t',\boldsymbol{x}) L_{h_0}(t-t') \left[\rho''(\breve{\lambda}_t^\top u_K(\boldsymbol{x})) - \rho''((\lambda_t^*)^\top u_K(\boldsymbol{x})) \right] u_K(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t') \right\| \\
=& \left\| \int m(t',\boldsymbol{x}) L_{h_0}(t-t') \rho'''(\xi_{t,3}(\boldsymbol{x})) u_K(\boldsymbol{x}) u_K^\top(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t') \cdot \{\breve{\lambda}_t - \lambda_t^*\} \right\| \\
\le& \|\breve{\lambda}_t - \lambda_t^*\| \cdot \lambda_{\max}^{1/2} \Bigg(\int m(t',\boldsymbol{x}) L_{h_0}(t-t') \rho'''(\xi_{t,3}(\boldsymbol{x})) u_K(\boldsymbol{x}) u_K^\top(x) dF_{X,T}(\boldsymbol{x},t') \\
&\cdot \int m(t',\boldsymbol{x}) L_{h_0}(t-t') \rho'''(\xi_{t,3}(\boldsymbol{x})) u_K(\boldsymbol{x}) u_K^\top(\boldsymbol{x}) dF_{X,T}(\boldsymbol{x},t') \Bigg) \\
=& O_p\left(\sqrt{\frac{K v_{h_0}(t)}{N}} \right)
\end{aligned}
$$

Using (C.14), (C.15), (C.26) and (D.6), we have, $\lambda_{\max}(\widetilde{\Sigma}_t)$ is negative and bounded away from zero with probability approaching to 1, we have $\lambda_{\min}\left(\widetilde{\Sigma}_t^{-1}\right) = \lambda_{\max}^{-1}(\widetilde{\Sigma}_t) < 0$ and $|\lambda_{\min}(\widetilde{\Sigma}_t^{-1})| = O_p(1)$. Therefore, together with (C.18), we have

$$
\begin{aligned}
&|(\Psi_t - \breve{\Psi}_t)^\top \widetilde{\Sigma}_t^{-1} \sqrt{N/\{v_h(t) \vee v_{h_0}(t)\}} \cdot \nabla \widehat{G}_t(\lambda_t^*)| \qquad \text{(D.34)}\\
\le\ & \sqrt{N/(v_h \vee v_{h_0})} \cdot \|(\breve{\Psi}_t - \Psi_t)^\top \tilde{\Sigma}_t^{-1}\| \|\nabla \widehat{G}_t(\lambda_t^*)\| \\
=\ & \sqrt{N/(v_h \vee v_{h_0})} \cdot \sqrt{(\breve{\Psi}_t - \Psi_t)^\top \left(\widetilde{\Sigma}_t^{-1}\right)^2 (\breve{\Psi}_t - \Psi_t)} \cdot \|\nabla \widehat{G}_t(\lambda_t^*)\| \\
\le\ & \sqrt{N/(v_h \vee v_{h_0})} \cdot \sqrt{\lambda_{\min}^2(\tilde{\Sigma}_t^{-1}) \|\breve{\Psi}_t - \Psi_t\|^2} \cdot \|\nabla \widehat{G}_t(\lambda_t^*)\| \\
\le\ & \sqrt{N/\{v_h(t) \vee v_{h_0}(t)\}} \cdot O_p(1) \cdot O_p\left(\sqrt{\frac{K v_{h_0}(t)}{N}}\right) O_p\left(\sqrt{\frac{K v_{h_0}(t)}{N}}\right) \\
=\ & O_p\left(\frac{K}{\sqrt{N}} \cdot \frac{v_{h_0}(t)}{\sqrt{v_h(t) \vee v_{h_0}(t)}}\right).
\end{aligned}
$$

Similarly, for the second term in (D.33), we can deduce that

$$
\begin{aligned}
&|\Psi_t^\top (\tilde{\Sigma}_t^{-1} - \Sigma_t^{-1}) \sqrt{N/(v_h(t) \vee v_{h_0}(t))} \cdot \nabla \widehat{G}_t(\lambda_t^*)| \qquad \text{(D.35)}\\
=&\sqrt{N/(v_h(t) \vee v_{h_0}(t))} \cdot |\nabla \widehat{G}_t(\lambda_t^*)^\top \widetilde{\Sigma}_t^{-1} (\Sigma_t - \tilde{\Sigma}_t) \Sigma_t^{-1} \Psi_t| \\
\le&\sqrt{N/(v_h(t) \vee v_{h_0}(t))} \cdot \|\nabla \widehat{G}_t(\lambda_t^*)\| \cdot \|\tilde{\Sigma}_t^{-1} (\Sigma_t - \tilde{\Sigma}_t) \Sigma_t^{-1} \Psi_t\| \\
=&\sqrt{N/(v_h(t) \vee v_{h_0}(t))} \cdot \|\nabla \widehat{G}_t(\lambda_t^*)\| \cdot \operatorname{tr}\left(\tilde{\Sigma}_t^{-1} \tilde{\Sigma}_t^{-1} (\Sigma_t - \tilde{\Sigma}_t) \Sigma_t^{-1} \Psi_t \Psi_t^\top \Sigma_t^{-1} (\Sigma_t - \tilde{\Sigma}_t)\right)^{\frac{1}{2}} \\
\le&\sqrt{N/(v_h(t) \vee v_{h_0}(t))} \cdot \|\nabla \widehat{G}_t(\lambda_t^*)\| \cdot \left|\lambda_{\max}(\tilde{\Sigma}_t^{-1} \tilde{\Sigma}_t^{-1})\right|^{\frac{1}{2}} \cdot \operatorname{tr}\left((\Sigma_t - \tilde{\Sigma}_t) \Sigma_t^{-1} \Psi_t \Psi_t^\top \Sigma_t^{-1} (\Sigma_t - \tilde{\Sigma}_t)\right)^{\frac{1}{2}} \\
=&\sqrt{N/(v_h(t) \vee v_{h_0}(t))} \cdot \|\nabla \widehat{G}_t(\lambda_t^*)\| \cdot \left|\lambda_{\min}(\tilde{\Sigma}_t^{-1})\right| \cdot \left\|(\Sigma_t - \tilde{\Sigma}_t) \gamma_t^*\right\| \cdot \mathbb{E}[L_{h_0}(t - T)]
\end{aligned}
$$

where the second inequality follow from the fact that $\operatorname{tr}(AB) \le \lambda_{\max}(B) \operatorname{tr}(A)$ for any symmetric B and positive semidefinite matrix A, and the last equality follows from the project property of $\Sigma_t^{-1} \Psi_t$ in (D.27).

Consider $\left\|(\Sigma_t - \tilde{\Sigma}_t) \gamma_t^*\right\|$. Using the mean value theorem, triangle inequality, and the arguments similar to those under (C.13), we have

$$
\begin{aligned}
&\left\|(\Sigma_t - \tilde{\Sigma}_t) \gamma_t^*\right\| \qquad \text{(D.36)}\\
\le&\left\|\frac{\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t - T) u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X}) \gamma_t^*\right]}{\mathbb{E}\left[L_{h_0}(t - T)\right]}\right.
\end{aligned}
$$

$$
- \frac{\sum_{i=1}^N \rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\} L_{U,h_0}(t-S_i) u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)\gamma_t^*}{\sum_{i=1}^N L_{U,h_0}(t-S_i)} \Bigg\|
$$

$$
+ \left\| \frac{\sum_{i=1}^N \rho'''\{\xi_{t,3}(\boldsymbol{X}_i)\} L_{U,h_0}(t-S_i) u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)\gamma_t^* \cdot (\widetilde{\lambda}_t - \lambda_t^*)^\top u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_{U,h_0}(t-S_i)} \right\| .
$$

For the first term in the above expression, using (D.6),

$$
\begin{aligned}
&\frac{\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t-T) u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\gamma_t^*\right]}{\mathbb{E}\left[L_{h_0}(t-T)\right]} \\
&\quad - \frac{\sum_{i=1}^N \rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\} L_{U,h_0}(t-S_i) u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)\gamma_t^*}{\sum_{i=1}^N L_{U,h_0}(t-S_i)} \\
&= \frac{\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t-T) u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\gamma_t^*\right]}{f_T^2(t)} \cdot \left[\frac{1}{N}\sum_{i=1}^N L_{U,h_0}(t-S_i) - \mathbb{E}\{L_{U,h_0}(t-S)\}\right] \\
&\quad + f_T^{-1}(t)\{1+o_P(1)\} \cdot \Bigg\{ \mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t-T) u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\gamma_t^*\right] \\
&\quad - \frac{1}{N}\sum_{i=1}^N \rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\} L_{U,h_0}(t-S_i) u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)\gamma_t^* \Bigg\} .
\end{aligned}
$$

Then using the projection property of $u_K^\top(\boldsymbol{X})\gamma_t^*$ in (D.27), Assumption 6 and Lemma 1, we have the first term in the above expression satisfies:

$$
\begin{aligned}
&\left\| \frac{\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t-T) u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\gamma_t^*\right]}{f_T^2(t)} \cdot \left[\frac{1}{N}\sum_{i=1}^N L_{U,h_0}(t-S_i) - \mathbb{E}\{L_{U,h_0}(t-S)\}\right] \right\| \\
&= \{1+o(1)\} \cdot O_P\left\{\sqrt{\frac{v_{h_0}(t)}{N}}\right\} \cdot \left\| \mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t-T) m(t,\boldsymbol{X}) u_K(\boldsymbol{X})\right] \right\| \\
&\leq \{1+o(1)\} \cdot O_P\left\{\sqrt{\frac{v_{h_0}(t)}{N}}\right\} \cdot \sqrt{\mathbb{E}\left[\left(\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\}\right)^2 m^2(t,\boldsymbol{X}) u_K^\top(\boldsymbol{X}) u_K(\boldsymbol{X}) | T=t\right]} \\
&= O_P\left\{\sqrt{\frac{v_{h_0}(t)K}{N}}\right\} .
\end{aligned}
$$

For the second term, note that

$$
\begin{aligned}
\mathbb{E}\Bigg[\Bigg\| &\mathbb{E}\left[\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} L_{h_0}(t-T) u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\gamma_t^*\right] \\
&- \frac{1}{N}\sum_{i=1}^N \rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\} L_{U,h_0}(t-S_i) u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)\gamma_t^* \Bigg\|^2\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N}\mathbb{E}\Big[\big(\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\}\big)^2 m^2(t,\boldsymbol{X}) L^2_{U,h_0}(t-S) u_K^\top(\boldsymbol{X}) u_K(\boldsymbol{X})\Big]\cdot\{1+o(1)\}\\
&= O\left\{\frac{v_{h_0}(t)K}{N}\right\}.
\end{aligned}
$$

Thus, we have the first term in (D.36) is of the rate $O_P\{\sqrt{v_{h_0}(t)K/N}\}$

For the last item in (D.36), by Assumption 6 (i), we have

$$
\begin{aligned}
&\left\|\frac{\sum_{i=1}^N \rho'''\{\xi_{t,3}(\boldsymbol{X}_i)\} L_U\{(t-S_i)/h_0\} u_K(\boldsymbol{X}_i) u_K^\top(\boldsymbol{X}_i)\gamma_t^*\cdot(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}_i)}{\sum_{i=1}^N L_U\{(t-S_i)/h_0\}}\right\|^2\\
=&\{1+o_P(1)\}\cdot\left\|\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\big|T=t\right]\right\|^2\\
=&\{1+o_P(1)\}\cdot\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\big|T=t\right]\\
&\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1}\cdot\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]\right)^2\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1}\\
&\cdot\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X}_i)(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K(\boldsymbol{X})\big|T=t\right]\\
\leq&\{1+o_P(1)\}\cdot\lambda^2_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]\right)\\
&\cdot\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\big|T=t\right]\\
&\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1}\\
&\cdot\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K(\boldsymbol{X})\big|T=t\right]\\
\leq&\{1+o_P(1)\}\cdot\lambda^2_{\max}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]\right)\cdot\lambda^{-1}_{\min}\left(\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]\right)\\
&\cdot\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\big|T=t\right]\\
&\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1}\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1}\\
&\cdot\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K(\boldsymbol{X})\big|T=t\right]\\
=&O_P(1)\cdot\mathbb{E}\left[\{L_t(\boldsymbol{X})\}^2|T=t\right], \qquad \text{(D.37)}
\end{aligned}
$$

where

$$
\begin{aligned}
L_t(\boldsymbol{X}):=&\mathbb{E}\left[\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\} m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X}) u_K^\top(\boldsymbol{X})\big|T=t\right]\\
&\cdot\mathbb{E}\left[u_K(\boldsymbol{X})u_K^\top(\boldsymbol{X})|T=t\right]^{-1} u_K(\boldsymbol{X}),
\end{aligned}
$$

is the least square projection (w.r.t. the metric induced by the $L^2(dF_{X|T}(\cdot|t))$ norm) of $\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\}m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X})$ on the space linearly spanned by $u_K(\boldsymbol{X})$. With

Lemma 3, it follows that

$$\begin{aligned}\mathbb{E}\left[\{L_t(\boldsymbol{X})\}^2|T=t\right] \leq& \mathbb{E}\left[\left\{\rho'''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\}m(t,\boldsymbol{X})(\widetilde{\lambda}_t-\lambda_t^*)^\top u_K(\boldsymbol{X})\right\}^2|T=t\right]\\ =&O_P\left(\frac{Kv_{h_0}}{N}\right). \qquad \text{(D.38)}\end{aligned}$$

Therefore, by combining (D.36), (D.37) and (D.38), we obtain that

$$\left\|(\Sigma_t-\tilde{\Sigma}_t)\gamma_t^*\right\| = O_p\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right). \qquad \text{(D.39)}$$

We next compute the order of $\Psi_t = -\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\}u_K(\boldsymbol{X})\right]$. By using the similar argument of obtaining (C.7), we can deduce that

$$\begin{aligned}&\|\Psi_t\|^2 \qquad \text{(D.40)}\\ =&\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))u_K(\boldsymbol{X})^\top\right]\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''((\lambda_t^*)^\top u_K(X))u_K(X)\right]\\ =&\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''((\lambda_t^*)^\top u_K(X))u_K(X)^\top\right]\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\\ &\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\\ &\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[m(T,\boldsymbol{X})L_h(t-T)\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))u_K(\boldsymbol{X})\right]\\ \leq&\lambda_{\max}\left(\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\right)\\ &\cdot\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))u_K(\boldsymbol{X})^\top\right]\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\\ &\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]\\ &\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\mathbb{E}\left[m(T,\boldsymbol{X})L_h(t-T)\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))u_K(\boldsymbol{X})\right]\\ =&O(1)\cdot\mathbb{E}\left[\left\{\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))u_K(\boldsymbol{X})^\top\right]\cdot\mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\right.\right.\\ &\left.\left.\cdot\sqrt{L_{h_0}(t-T)}u_K(\boldsymbol{X})\right\}^2\right]\\ \leq&O(1)\cdot\mathbb{E}\left[\left\{m(T,\boldsymbol{X})\sqrt{L_{h_0}(t-T)}\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))\right\}^2\right]\\ \leq&O(1)\cdot\sup_{x\in\mathcal{X}}|\rho''((\lambda_t^*)^\top u_K(\boldsymbol{x}))|^2\cdot\mathbb{E}\left[m(T,\boldsymbol{X})^2L_{h_0}(t-T)\right]=O(1), \qquad \text{(D.41)}\end{aligned}$$

where the first inequality follows from the fact that

$$\mathbb{E}\left[m(T,\boldsymbol{X})L_{h_0}(t-T)\rho''((\lambda_t^*)^\top u_K(\boldsymbol{X}))u_K(\boldsymbol{X})^\top\right]$$

$$\cdot \mathbb{E}\left[L_{h_0}(t-T)u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top\right]^{-1}\sqrt{L_{h_0}(t-T)}u_K(\boldsymbol{X})$$

is the $L^2(dF_X)$-projection of $m(T,\boldsymbol{X})\sqrt{L_{h_0}(t-T)}\rho''\{(\lambda_t^*)^\top u_K(\boldsymbol{X})\}$ on the space spanned by $\sqrt{L_{h_0}(t-T)}u_K(\boldsymbol{X})$. Combining (D.35), (D.39), and (D.41) we can obtain

$$\begin{aligned}
&\|\Psi_t^\top(\widetilde{\Sigma}_t^{-1}-\Sigma_t^{-1})\sqrt{N/(v_h(t)\vee v_{h_0}(t))}\nabla\widehat{G}_t(\lambda_t^*)\|\\
=&\sqrt{N/(v_h(t)\vee v_{h_0}(t))}\cdot O_p\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right)O(1)O_p(1)O_p\left(\sqrt{\frac{Kv_{h_0}(t)}{N}}\right)\\
=&O_p\left(\frac{v_{h_0}(t)}{\sqrt{(v_h(t)\vee v_{h_0}(t))}}\frac{K}{\sqrt{N}}\right),
\end{aligned}$$

then together with (D.34) and Assumption 6, we have

$$\begin{aligned}
\text{(D.21)} =&O_p\left(\frac{K}{\sqrt{N}}\cdot\frac{v_{h_0}(t)}{\sqrt{(v_h(t)\vee v_{h_0}(t))}}\right)+O_p\left(\frac{v_{h_0}(t)}{\sqrt{(v_h(t)\vee v_{h_0}(t))}}\frac{K}{\sqrt{N}}\right)\\
=&O_p\left(\frac{v_{h_0}(t)}{\sqrt{(v_h(t)\vee v_{h_0}(t))}}\frac{K}{\sqrt{N}}\right).
\end{aligned}$$

For term (D.22)**:** Note that $\mathbb{E}\{\text{(D.22)}\}=0$.

$$\begin{aligned}
&\operatorname{var}\{\text{(D.22)}\}\le[f_T^2(t)\{v_h(t)\vee v_{h_0}(t)\}]^{-1}\cdot\mathbb{E}\Bigg[\Bigg\{\Psi_t^\top\Sigma_t^{-1}u_K(\boldsymbol{X}_i)\rho'\{(\lambda_t^*)^\top u_K(\boldsymbol{X}_i)\}\\
&\qquad-m(t,\boldsymbol{X}_i)\pi_0(t,\boldsymbol{X}_i)\cdot\mathbb{E}[L_{h_0}(t-T_i)]\Bigg\}^2\cdot L_{U,h_0}^2\{t-S_i\}\Bigg]\\
&=[f_T^2(t)\{v_h(t)\vee v_{h_0}(t)\}]^{-1}\cdot\mathbb{E}\Bigg[\Bigg\{\Big[\Psi_t^\top\Sigma_t^{-1}u_K(\boldsymbol{X}_i)-m(t,\boldsymbol{X}_i)\mathbb{E}\{L_{h_0}(t-T)\}\Big]\pi^*(t,\boldsymbol{X}_i)\\
&+m(t,\boldsymbol{X}_i)\mathbb{E}[L_{h_0}(t-T)]\Big[\pi^*(t,\boldsymbol{X}_i)-\pi_0(t,\boldsymbol{X}_i)\Big]\Bigg\}^2\cdot L_{U,h_0}^2\{t-S_i\}\Bigg]\\
=&O\left(\{v_{h_0}(t)/(v_h(t)\vee v_{h_0}(t))\}\cdot\{K^{-2\ell}+(h_0^4+K^{-2\alpha})\}\right)=o(1),
\end{aligned}$$

where the second equality holds by using Lemmas 1, 2, and the projection approximation (D.27).

Finally, we can see from law of large numbers that (D.23) is of rate $(v_h(t) \vee v_{h_0}(t))^{-1/2} = o(1)$ given that $v_h(t), v_{h_0}(t) \to \infty$ as $N \to \infty$. Thus, we have (D.32) holds under the assumption

$$\frac{v_{h_0}(t)}{\sqrt{v_h(t) \vee v_{h_0}(t)}} \frac{K}{\sqrt{N}} \to 0.$$

The result follows.