Local Regression Distribution Estimators
Supplemental Appendix

Matias D. Cattaneo* Michael Jansson† Xinwei Ma‡

September 29, 2020

Abstract

This Supplemental Appendix contains general theoretical results encompassing those discussed in the main paper, includes proofs of those general results, and discusses additional methodological and technical results.

*Department of Operations Research and Financial Engineering, Princeton University.
†Department of Economics, UC Berkeley and CREATES.
‡Department of Economics, UC San Diego.

Contents

1 Setup

Suppose $x_1, x_2, \cdots, x_n$ is a random sample from a univariate distribution with cumulative distribution function $F(\cdot)$. Also assume the distribution function admits a (sufficiently accurate) linear-in-parameters local approximation near an evaluation point x:

$$\varrho(h, \mathsf{x}) := \sup_{|x-\mathsf{x}|\leq h} \left| F(x) - R(x-\mathsf{x})'\theta(\mathsf{x}) \right| \text{ is small for } h \text{ small},$$

where $R(\cdot)$ is a known basis function. The parameter $\theta(\mathsf{x})$ can be estimated by the following local L^2 method:

$$\hat{\theta}_G = \operatorname*{argmin}_{\theta} \int_{\mathcal{X}} \left(\hat{F}(u) - R(u-\mathsf{x})'\theta \right)^2 \frac{1}{h} K\left(\frac{u-\mathsf{x}}{h} \right) \mathrm{d}G(u), \qquad \hat{F}(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \leq u), \tag{1}$$

where $K(\cdot)$ is a kernel function, $\mathcal{X}$ is the support of $F(\cdot)$, and $G(\cdot)$ is a known weighting function to be specified later. The local projection estimator (1) is closely related to another estimator, which is constructed by local regression:

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} \left(\hat{F}(x_i) - R(x_i-\mathsf{x})'\theta \right)^2 \frac{1}{h} K\left(\frac{x_i-\mathsf{x}}{h} \right). \tag{2}$$

The local regression estimator can be equivalently expressed as $\hat{\theta}_{\hat{F}}$, meaning that it can be viewed as a special case of the local projection estimator, with $G(\cdot)$ in (1) replaced by the empirical measure (empirical distribution function) $\hat{F}(\cdot)$.

For future reference, we first discuss some of the notation we use in the main paper and this Supplemental Appendix (SA). For a function $g(\cdot)$, we denote its j-th derivative as $g^{(j)}(\cdot)$. For simplicity, we also use the "dot" notation to denote the first derivative: $\dot{g}(\cdot) = g^{(1)}(\cdot)$. Assume $g(\cdot)$ is suitably smooth on $[\mathsf{x}-\delta, \mathsf{x}) \cup (\mathsf{x}, \mathsf{x}+\delta]$ for some $\delta > 0$, but not necessarily continuous or differentiable at x. If $g(\cdot)$ and its one-sided derivatives can be continuously extended to x from the two sides, we adopt the following special notation:

$$g_u^{(j)} = \mathbb{1}(u < 0) g^{(j)}(\mathsf{x}-) + \mathbb{1}(u \geq 0) g^{(j)}(\mathsf{x}+).$$

With $j = 0$, the above is simply $g_u = \mathbb{1}(u < 0)g(\mathsf{x}-) + \mathbb{1}(u \geq 0)g(\mathsf{x}+)$. Also for $j = 1$, we use $\dot{g}_u = g_u^{(1)}$. Convergence in probability and in distribution are denoted by $\overset{\mathbb{P}}{\to}$ and $\rightsquigarrow$, respectively, and limits are taken with respect to the sample size n going to infinity unless otherwise specified. We use $|\cdot|$ to denote the Euclidean norm.

The following matrices will feature in asymptotic expansions of our estimators:

$$\begin{aligned}\Gamma_{h,\mathsf{x}} &= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(u)'K(u)\,g(\mathsf{x}+hv)\mathrm{d}u \\ &= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(u)'K(u)\,g_u\mathrm{d}u + O(h) = \Gamma_{1h,\mathsf{x}} + O(h)\end{aligned}$$

and

$$\begin{aligned}\Sigma_{h,\mathsf{x}} &= \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'\Big[F(\mathsf{x}+h(u\wedge v)) - F(\mathsf{x}+hu)F(\mathsf{x}+hv)\Big]K(u)\,K(v)\,g(\mathsf{x}+hu)g(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v \\ &= F(\mathsf{x})(1-F(\mathsf{x}))\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)K(u)g_u\mathrm{d}u\right)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)K(u)g_u\mathrm{d}u\right)' \\ &\quad + h\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'K(u)K(v)\Big[-F(\mathsf{x})(uf_u+vf_v)g_ug_v + F(\mathsf{x})(1-F(\mathsf{x}))(u\dot{g}_ug_v+v\dot{g}_vg_u)\Big]\mathrm{d}u\mathrm{d}v \\ &\quad + h\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'K(u)K(v)(u\wedge v)f_{u\wedge v}g_ug_v\mathrm{d}u\mathrm{d}v + O(h^2) \\ &= \Sigma_{1h,\mathsf{x}} + h\Sigma_{2h,\mathsf{x}} + O(h^2).\end{aligned}$$

Now we list the main assumptions.

Assumption 1. *$x_1,\ldots,x_n$ is a random sample from a distribution $F(\cdot)$ supported on $\mathcal{X}\subseteq\mathbb{R}$, and $\mathsf{x}\in\mathcal{X}$.*

(i) For some $\delta>0$, $F(\cdot)$ is absolutely continuous on $[\mathsf{x}-\delta,\mathsf{x}+\delta]$ with a density $f(\cdot)$ admitting constants $f(\mathsf{x}-)$, $f(\mathsf{x}+)$, $\dot{f}(\mathsf{x}-)$, and $\dot{f}(\mathsf{x}+)$, such that

$$\sup_{u\in[-\delta,0)}\frac{f(\mathsf{x}+u)-f(\mathsf{x}-)-u\dot{f}(\mathsf{x}-)}{u^2} + \sup_{u\in(0,\delta]}\frac{f(\mathsf{x}+u)-f(\mathsf{x}+)-u\dot{f}(\mathsf{x}+)}{u^2} < \infty.$$

(ii) $K(\cdot)$ is nonnegative, symmetric, and continuous on its support $[-1,1]$, and integrates to 1.

(iii) $R(\cdot)$ is locally bounded, and there exists a positive-definite diagonal matrix Υ_h for each $h>0$, such that $\Upsilon_h R(u) = R(u/h)$

(iv) For all h sufficiently small, the minimum eigenvalues of $\Gamma_{h,\mathsf{x}}$ and $h^{-1}\Sigma_{h,\mathsf{x}}$ are bounded away from zero. ■

Assumption 2. *For some $\delta>0$, $G(\cdot)$ is absolutely continuous on $[\mathsf{x}-\delta,\mathsf{x}+\delta]$ with a derivative $g(\cdot)\geq 0$ admitting constants $g(\mathsf{x}-)$, $g(\mathsf{x}+)$, $\dot{g}(\mathsf{x}-)$, and $\dot{g}(\mathsf{x}+)$, such that*

$$\sup_{u\in[-\delta,0)}\frac{g(\mathsf{x}+u)-g(\mathsf{x}-)-u\dot{g}(\mathsf{x}-)}{u^2} + \sup_{u\in(0,\delta]}\frac{g(\mathsf{x}+u)-g(\mathsf{x}+)-u\dot{g}(\mathsf{x}+)}{u^2} < \infty.$$ ■

Example 1 (Local Polynomial Estimator). Before closing this section, we briefly introduce the local polynomial estimator of Cattaneo, Jansson, and Ma (2020), which is a special case of our local regression distribution estimator. The local polynomial estimator employs the following polynomial basis:

$$R(u) = \Big(1,\ u,\ \frac{1}{2}u^2,\ \cdots,\ \frac{1}{p!}u^p\Big)',$$

for some $p \in \mathbb{N}$. As a result, it estimates the distribution function, the density function, and derivatives thereof. To be precise,

$$\theta(\mathsf{x}) = \Big(F(\mathsf{x}),\ f(\mathsf{x}),\ f^{(1)}(\mathsf{x}),\ \cdots,\ f^{(p-1)}(\mathsf{x})\Big)'.$$

With $R(\cdot)$ being a polynomial basis, it is straightforward to characterize the approximation bias $\varrho(h, \mathsf{x})$. Assuming the distribution function $F(\cdot)$ is at least $p+1$ times continuously differentiable in a neighborhood of x, one can employ a Taylor expansion argument and show that $\varrho(h, \mathsf{x}) = O(h^{p+1})$. We will revisit this local polynomial estimator below as a leading example when we discuss pointwise and uniform asymptotic properties of our local distribution estimator. ■

2 Pointwise Distribution Theory

We discuss pointwise (i.e., for a fixed evaluation point $\mathsf{x} \in \mathcal{X}$) large-sample properties of the local L^2 distribution estimator (1), and that of the local regression estimator (2). For ease of exposition, we suppress the dependence on the evaluation point x whenever possible.

2.1 Local L^2 Distribution Estimation

With simple algebra, the local $L^2(G)$ estimator in (1) takes the following form

$$\hat{\theta}_G = \left(\int_{\mathcal{X}} R(u-\mathsf{x})R(u-\mathsf{x})'\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}G(u)\right)^{-1}\left(\int_{\mathcal{X}} R(u-\mathsf{x})\hat{F}(u)\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}G(u)\right).$$

We can further simplify the above. First note that the "denominator" can be rewritten as

$$\begin{aligned}&\int_{\mathcal{X}} R(u-\mathsf{x})R(u-\mathsf{x})'\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}G(u)\\&=\Upsilon_h^{-1}\left(\int_{\mathcal{X}} \Upsilon_h R(u-\mathsf{x})R(u-\mathsf{x})'\Upsilon_h\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)g(u)\mathrm{d}u\right)\Upsilon_h^{-1} = \Upsilon_h^{-1}\Gamma_h\Upsilon_h^{-1}.\end{aligned}$$

The same technique can be applied to the "numerator", which leads to

$$\hat{\theta}_G - \theta = \Upsilon_h\Gamma_h^{-1}\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\hat{F}(\mathsf{x}+hu)K(u)\,g(\mathsf{x}+hu)\mathrm{d}u\right) - \theta$$

$$=\Upsilon_h\Gamma_h^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u \tag{3}$$

$$+\ \Upsilon_h\frac{1}{n}\sum_{i=1}^{n}\Gamma_h^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u. \tag{4}$$

The above provides further expansion of the local L^2 estimator into a term that contributes as bias, and another term that contributes asymptotically to the variance.

The large-sample properties of the local L^2 distribution estimator (1) are as follows.

Theorem 1 (Local L^2 Distribution Estimation: Asymptotic Normality). *Assume Assumptions 1 and 2 hold, and that $h \to 0$, $nh \to \infty$ and $n\varrho(h)^2/h \to 0$. Then*
(i) (3) *satisfies*

$$\left| \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big] K(u)\,\mathrm{d}u \right| = O(\varrho(h)).$$

(ii) (4) *satisfies*

$$\mathbb{V}\left[\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big] K(u)\, g(\mathsf{x}+hu)\mathrm{d}u\right] = \Sigma_h,$$

and

$$\Sigma_h^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big] K(u)\, g(\mathsf{x}+hu)\mathrm{d}u\right) \rightsquigarrow \mathcal{N}(0, I).$$

(iii) The local L^2 distribution estimator is asymptotically normally distributed

$$\sqrt{n}\left(\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}\right)^{-1/2}\Upsilon_h^{-1}(\hat\theta_G - \theta) \rightsquigarrow \mathcal{N}(0, I).$$

■

For valid inference, one needs to construct standard errors. To start, note that Γ_h is known, and hence we only need to estimate Σ_h. Consider the following:

$$\hat\Sigma_h = \frac{1}{n}\sum_{i=1}^{n}\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - \hat F(\mathsf{x}+hu)\Big]\Big[\mathbb{1}(x_i \le \mathsf{x}+hv) - \hat F(\mathsf{x}+hv)\Big]$$
$$K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v, \tag{5}$$

where $\hat F(\cdot)$ is the empirical distribution function. The following theorem shows that standard errors constructed using $\hat\Sigma_h$ are consistent.

Theorem 2 (Local L^2 Distribution Estimation: Standard Errors). *Assume Assumptions 1 and 2 hold, and that $h \to 0$ and $nh \to \infty$. Let c be a nonzero vector of suitable dimension, then*

$$\left|\frac{c'\hat\Sigma_h c}{c'\Sigma_h c} - 1\right| = O_{\mathbb{P}}\left(\sqrt{\frac{1}{nh}}\right).$$

If, in addition that $n\varrho(h)^2/h \to 0$, *then*

$$\frac{c'(\hat{\theta}_G - \theta)}{\sqrt{c'\Upsilon_h \Gamma_h^{-1}\hat{\Sigma}_h\Gamma_h^{-1}\Upsilon_h c/n}} \rightsquigarrow \mathcal{N}(0,1).$$

■

2.2 Local Regression Distribution Estimation

The local regression distribution estimator (2) can be understood as a special case of the local L^2 based estimator, by setting $G = \hat{F}$ (i.e., using the empirical distribution as the design). However, the empirical measure $\hat{F}$ is not smooth, so that large-sample properties of the local regression estimator cannot be deduced directly from Theorem 1. In this subsection, we will show that estimates obtained by the two approaches, (1) and (2), are asymptotically equivalent under suitable regularity conditions. To be precise, we establish the equivalence of the local regression distribution estimator, $\hat{\theta}$, and the (infeasible) local L^2 distribution estimator, $\hat{\theta}_F$ (i.e., using F as the design weighting in (1)). As before, we suppress the dependence on the evaluation point x.

First, the local regression estimator can be written as

$$\begin{aligned}\hat{\theta} - \theta &= \left(\frac{1}{n}\sum_{i=1}^n R(x_i - \mathsf{x})R(x_i - \mathsf{x})'\frac{1}{h}K\left(\frac{x_i - \mathsf{x}}{h}\right)\right)^{-1} \\ &\qquad \left(\frac{1}{n}\sum_{i=1}^n R(x_i - \mathsf{x})\Big[\hat{F}(x_i) - R(x_i - \mathsf{x})'\theta\Big]\frac{1}{h}K\left(\frac{x_i - \mathsf{x}}{h}\right)\right) \\ &= \Upsilon_h\hat{\Gamma}_h^{-1}\Gamma_h\Gamma_h^{-1}\left(\frac{1}{n}\sum_{i=1}^n \Upsilon_h R(x_i - \mathsf{x})\Big[\hat{F}(x_i) - R(x_i - \mathsf{x})'\theta\Big]\frac{1}{h}K\left(\frac{x_i - \mathsf{x}}{h}\right)\right),\end{aligned}$$

where

$$\hat{\Gamma}_h = \frac{1}{n}\sum_{i=1}^n \Upsilon_h R(x_i - \mathsf{x})R(x_i - \mathsf{x})'\Upsilon_h\frac{1}{h}K\left(\frac{x_i - \mathsf{x}}{h}\right),$$

and Γ_h is defined as before with $G = F$.

To proceed, we further expand as follows

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \Upsilon_h R(x_i - \mathsf{x})\Big[\hat{F}(x_i) - R(x_i - \mathsf{x})'\theta\Big]\frac{1}{h}K\left(\frac{x_i - \mathsf{x}}{h}\right) \\ &= \frac{1}{n^2}\sum_{i,j=1, i\neq j}^n \Upsilon_h R(x_j - \mathsf{x})\Big[\mathbb{1}(x_i \le x_j) - F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j - \mathsf{x}}{h}\right) \\ &\quad + \frac{1}{n^2}\sum_{j=1}^n \Upsilon_h R(x_j - \mathsf{x})\Big[1 - F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j - \mathsf{x}}{h}\right) \qquad (6) \\ &\quad + \frac{1}{n}\sum_{j=1}^n \Upsilon_h R(x_j - \mathsf{x})\Big[F(x_j) - R(x_j - \mathsf{x})'\theta\Big]\frac{1}{h}K\left(\frac{x_j - \mathsf{x}}{h}\right). \qquad (7)\end{aligned}$$

The last two terms correspond to the leave-in bias and the approximation bias, respectively. We further decompose the first term with conditional expectation:

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} \Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\\
&=\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} \mathbb{E}\left[\Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\middle| x_i\right]\\
&\quad+\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} \Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\\
&\qquad\qquad-\mathbb{E}\left[\Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\middle| x_i\right]\\
&=\frac{n-1}{n^2}\sum_{i=1}^{n}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i\le \mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]K\left(u\right)f(\mathsf{x}+hu)\mathrm{d}u \qquad (8)\\
&\quad+\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} \Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\\
&\qquad\qquad-\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i\le \mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]K\left(u\right)f(\mathsf{x}+hu)\mathrm{d}u. \qquad (9)
\end{aligned}
$$

The following theorem studies the large-sample properties of each term in the above decomposition, and shows that the local regression distribution estimator is asymptotically equivalent to the local L^2 based estimator by setting $G=F$, and hence it is asymptotically normally distributed.

Theorem 3 (Local Regression Distribution Estimation: Asymptotic Normality). *Assume Assumption 1 holds, and that $h\to 0$, $nh^2\to\infty$ and $n\varrho(h)^2/h\to 0$. Then*
(i) $\hat{\Gamma}_h$ satisfies

$$\left|\hat{\Gamma}_h-\Gamma_h\right| = O_{\mathbb{P}}\left(\sqrt{\frac{1}{nh}}\right).$$

(ii) (6) and (7) satisfy

$$(6)=O_{\mathbb{P}}\left(\frac{1}{n}\right),\qquad (7)=O_{\mathbb{P}}(\varrho(h)).$$

(iii) (9) satisfies

$$(9)=O_{\mathbb{P}}\left(\sqrt{\frac{1}{n^2h}}\right).$$

(iv) The local regression distribution estimator (2) *satisfies*

$$\sqrt{n}\left(\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}\right)^{-1/2}\Upsilon_h^{-1}(\hat\theta-\theta) = \sqrt{n}\left(\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}\right)^{-1/2}\Upsilon_h^{-1}(\hat\theta_F-\theta) + o_{\mathbb{P}}(1)$$
$$\rightsquigarrow \mathcal{N}(0, I).$$

■

We now discuss how to construct standard errors in the local regression framework. Note that Γ_h can be estimated by $\hat\Gamma_h$, whose properties have already been studied in Theorem 3. For Σ_h, we propose the following

$$\hat\Sigma_h = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}\Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-\hat F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\right]$$
$$\left[\frac{1}{n}\sum_{j=1}^{n}\Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-\hat F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\right]'.$$

where $\hat F(\cdot)$ is the empirical distribution function. The following theorem shows that standard errors constructed using $\hat\Sigma_h$ are consistent.

Theorem 4 (Local Regression Distribution Estimation: Standard Errors). *Assume Assumption 1 holds. In addition, assume $h\to 0$ and $nh^2\to\infty$. Let c be a nonzero vector of suitable dimension. Then*

$$\left|\frac{c'\hat\Gamma_h^{-1}\hat\Sigma_h\hat\Gamma_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c}-1\right| = O_{\mathbb{P}}\left(\sqrt{\frac{1}{nh^2}}\right).$$

If, in addition that $n\varrho(h)^2/h\to 0$, one has

$$\frac{c'(\hat\theta-\theta)}{\sqrt{c'\Upsilon_h\hat\Gamma_h^{-1}\hat\Sigma_h\hat\Gamma_h^{-1}\Upsilon_h c/n}} \rightsquigarrow \mathcal{N}(0,1).$$

■

3 Efficiency

We focus on the local L^2 distribution estimator $\hat\theta_G$, but all the results in this section are applicable to the local regression distribution estimator $\hat\theta$, as we showed earlier that it is asymptotically equivalent to $\hat\theta_F$. In addition, we consider a specific basis:

$$R(u) = \big(1,\ P(u)',\ Q(u)\big)', \tag{10}$$

where $P(u)$ is a polynomial basis of order p:

$$P(u) = \left(u,\ \frac{1}{2}u^2,\ \cdots,\ \frac{1}{p!}u^p\right)',$$

and $Q(u)$ is a scalar function, and hence is a "redundant regressor." Without $Q(\cdot)$, the above reduces to the local polynomial estimator of Cattaneo, Jansson, and Ma (2020). See Section 1 and Example 1 for an introduction.

We consider additional regressors because they may help improve efficiency (i.e., reduce the asymptotic variance). Following Assumption 1, we assume there exists a scalar υ_h (depending on h) such that $\upsilon_h Q(u) = Q(u/h)$. Therefore, Υ_h is a diagonal matrix containing $1, h^{-1}, h^{-2}, \cdots, h^{-p}, \upsilon_h$. As we consider a (local) polynomial basis, it is natural to impose smoothness assumptions on $F(\cdot)$. In particular,

Assumption 3 (Smoothness). *For some $\delta > 0$, $F(\cdot)$ is $(p+1)$-times continuously differentiable in $\mathcal{X} \cap [\mathsf{x}-\delta, \mathsf{x}+\delta]$ for some $p \geq 1$, and $G(\cdot)$ is twice continuously differentiable in $\mathcal{X} \cap [\mathsf{x}-\delta, \mathsf{x}+\delta]$.* ■

Under the above assumption, the approximation error satisfies $\varrho(h) = O(h^{p+1})$, and the parameter θ can be partitioned into the following:

$$\theta = \Big(\theta_1,\ \theta_P',\ \theta_Q'\Big) = \Big(F(\mathsf{x}), \quad f(\mathsf{x}),\ \cdots,\ f^{(p-1)}(\mathsf{x}), \quad 0\Big)'.$$

We first state a corollary, which specializes Theorem 1 to the polynomial basis (10). That is, we study the (infeasible) estimator

$$\hat{\theta}_F = \underset{\theta}{\operatorname{argmin}} \int_{\mathcal{X}} \Big(\hat{F}(u) - R(u-\mathsf{x})'\theta\Big)^2 \frac{1}{h} K\left(\frac{u-\mathsf{x}}{h}\right) \mathrm{d}F(u). \tag{11}$$

Corollary 5 (Local Polynomial L^2 Distribution Estimation: Asymptotic Normality). *Assume Assumptions 1, 2 and 3 hold, and that $h \to 0$, $nh \to \infty$, and $n\varrho(h)^2/h \to 0$. Then the local polynomial L^2 distribution estimator in (11) satisfies*

$$\sqrt{n}\left(\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}\right)^{-1/2}\Upsilon_h^{-1}(\hat{\theta}_F - \theta) \rightsquigarrow \mathcal{N}(0, I).$$

■

3.1 Effect of Orthogonalization

For ease of presentation, consider the following (sequentially) orthogonalized basis:

$$R^{\perp}(u) = \Big(1,\ P^{\perp}(u)',\ Q^{\perp}(u)\Big)', \tag{12}$$

where

$$P^{\perp}(u) = P^{\perp}(u) - \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)\mathrm{d}u,$$

$$Q^{\perp}(u) = Q(u) - \Big(1, P(v)'\Big)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v)\Big(1, P(v)'\Big)'\Big(1, P(v)'\Big)\mathrm{d}v\right)^{-1}\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v)\Big(1, P(v)'\Big)'Q(v)\mathrm{d}v\right).$$

The above transformation can be represented by the following:

$$R^{\perp}(u) = \Lambda_h' R(u),$$

where Λ_h is a nonsingular upper triangular matrix. (Note that the matrix Λ_h depends on the bandwidth only because we would like to handle both interior and boundary evaluation points. If, for example, we fix the evaluation point to be in the interior of the support of the data, then Λ_h is a fixed matrix and no longer depends on h. Alternatively, one could also use the notation "Λ_{x}" to denote such dependence.) Now consider the following orthogonalized local polynomial L^2 distribution estimator

$$\hat{\theta}_F^{\perp} = \operatorname*{argmin}_{\theta} \int_{\mathcal{X}} \left(\hat{F}(u) - \Lambda_h' R(u-\mathsf{x})'\theta\right)^2 \frac{1}{h} K\left(\frac{u-\mathsf{x}}{h}\right) \mathrm{d}F(u). \tag{13}$$

To discuss its properties, we partition the estimator and the target parameter as

$$\hat{\theta}_F^{\perp} = \left(\hat{\theta}_{1,F}^{\perp},\ (\hat{\theta}_{P,F}^{\perp})',\ \hat{\theta}_{Q,F}^{\perp}\right)',$$

where $\hat{\theta}_{1,F}^{\perp}$ is the first element of $\hat{\theta}_F^{\perp}$ and $\hat{\theta}_{Q,F}^{\perp}$ is the last element of $\hat{\theta}_F^{\perp}$. Similarly, we can partition the target parameter,

$$\theta^{\perp} = \Lambda_h^{-1}\theta = \left(\theta_1^{\perp},\ (\theta_P^{\perp})',\ \theta_Q^{\perp}\right)',$$

so that $\theta_1^{\perp}$ is the first element of $\Lambda_h^{-1}\theta$ and $\theta_Q^{\perp}$ is the last element of $\Lambda_h^{-1}\theta$. As $\theta_Q = 0$, simple least squares algebra implies

$$\theta^{\perp} = \left(\theta_1^{\perp},\ \theta_P',\ 0\right)' = \left(\theta_1^{\perp},\ f(\mathsf{x}),\ f^{(1)}(\mathsf{x}),\ \cdots,\ f^{(p-1)}(\mathsf{x}),\ 0\right)'.$$

Note that, in general, $\theta_1^{\perp} \neq \theta_1$, meaning that after orthogonalization, the intercept of the local polynomial estimator no longer estimates the distribution function $F(\mathsf{x})$.

The following corollary gives the large-sample properties of the orthogonalized local polynomial estimator, excluding the intercept.

Corollary 6 (Orthogonalized Local Polynomial L^2 Distribution Estimation: Asymptotic Normality). *Assume Assumptions 1, 2 and 3 hold, and that $h \to 0$, $nh \to \infty$, and $n\varrho(h)^2/h \to 0$. Then the orthogonalized local polynomial L^2 distribution estimator in (13) satisfies*

$$\begin{bmatrix} (\Gamma_{P,h}^{\perp})^{-1}\Sigma_{PP,h}^{\perp}(\Gamma_{P,h}^{\perp})^{-1} & (\Gamma_{P,h}^{\perp})^{-1}\Sigma_{PQ,h}^{\perp}(\Gamma_{Q,h}^{\perp})^{-1} \\ (\Gamma_{Q,h}^{\perp})^{-1}\Sigma_{QP,h}^{\perp}(\Gamma_{P,h}^{\perp})^{-1} & (\Gamma_{Q,h}^{\perp})^{-1}\Sigma_{QQ,h}^{\perp}(\Gamma_{Q,h}^{\perp})^{-1} \end{bmatrix}^{-1/2} \sqrt{\frac{n}{hf(\mathsf{x})}}\Upsilon_{-1,h}^{-1} \begin{bmatrix} \hat{\theta}_{P,F}^{\perp} - \theta_P \\ \hat{\theta}_{Q,F}^{\perp} \end{bmatrix} \rightsquigarrow \mathcal{N}(0, I),$$

where

$$\Gamma^{\perp}_{P,h} = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} P^{\perp}(u)P^{\perp}(u)'K(u)\mathrm{d}u, \quad \Gamma^{\perp}_{Q,h} = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} Q^{\perp}(u)^2K(u)\mathrm{d}u,$$

$$\Sigma^{\perp}_{PP,h} = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)P^{\perp}(u)P^{\perp}(v)'(u\wedge v)\mathrm{d}u\mathrm{d}v,$$

$$\Sigma^{\perp}_{QQ,h} = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)Q^{\perp}(u)Q^{\perp}(v)(u\wedge v)\mathrm{d}u\mathrm{d}v,$$

$$\Sigma^{\perp}_{PQ,h} = (\Sigma^{\perp}_{QP,h})' = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)P^{\perp}(u)Q^{\perp}(v)(u\wedge v)\mathrm{d}u\mathrm{d}v,$$

and $\Upsilon_{-1,h}$ *is a diagonal matrix containing* $h^{-1}, h^{-2}, \cdots, h^{-p}, \upsilon_h$*.* ■

3.2 Optimal Q

Now we discuss the optimal choice of Q, which minimizes the asymptotic variance of the minimum distance estimator. Recall from the main paper that, with orthogonalized basis, the minimum distance estimator of $f^{(\ell)}(\mathsf{x})$, for $0 \le \ell \le p-1$, has an asymptotic variance

$$f(\mathsf{x})\Big[e_\ell'(\Gamma^{\perp}_{P,h})^{-1}\Sigma^{\perp}_{PP,h}(\Gamma^{\perp}_{P,h})^{-1}e_\ell - e_\ell'(\Gamma^{\perp}_{P,h})^{-1}\Sigma^{\perp}_{PQ,h}(\Sigma^{\perp}_{QQ,h})^{-1}\Sigma^{\perp}_{QP,h}(\Gamma^{\perp}_{P,h})^{-1}e_\ell\Big],$$

where e_ℓ is the $(\ell+1)$-th standard basis vector. In subsequent analysis, we drop the multiplicative factor $f(\mathsf{x})$.

Let $p_\ell(u)$ be defined as

$$p_\ell(u) = e_\ell'(\Gamma^{\perp}_{P,h})^{-1}P^{\perp}(u),$$

then the objective is to maximize

$$\left(\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)Q^{\perp}(u)Q^{\perp}(v)(u\wedge v)\mathrm{d}u\mathrm{d}v\right)^{-1}\left(\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)p_\ell(u)Q^{\perp}(v)(u\wedge v)\mathrm{d}u\mathrm{d}v\right)^{2}.$$

Alternatively, we would like to solve (recall that $Q(u)$ is a scaler function)

$$\text{maximize}\qquad \frac{\left(\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)p_\ell(u)q(v)(u\wedge v)\mathrm{d}u\mathrm{d}v\right)^2}{\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)q(u)q(v)(u\wedge v)\mathrm{d}u\mathrm{d}v},\quad \text{subject to}\quad \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)q(u)(1,P(u)')\mathrm{d}u = 0.$$

To proceed, define the following transformation for a function $g(\cdot)$:

$$\mathcal{H}(g)(u) = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \mathbb{1}(v\ge u)K(v)g(v)\mathrm{d}v.$$

This transformation satisfies two important properties, which are summarized in the following lemma.

Lemma 7 ($\mathcal{H}(\cdot)$ Transformation).

(i) If $g_1(\cdot)$ and $g_2(\cdot)$ are bounded, and that either $\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)g_1(u)\mathrm{d}u$ or $\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)g_2(u)\mathrm{d}u$ is zero, then

$$\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(g_1)(u)\mathcal{H}(g_2)(u)\mathrm{d}u = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)g_1(u)g_1(v)(u \wedge v)\mathrm{d}u\mathrm{d}v.$$

(ii) If $g_1(\cdot)$ and $g_2(\cdot)$ are bounded, $g_2(\cdot)$ is continuously differentiable with a bounded derivative, and that either $\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)g_1(u)\mathrm{d}u$ or $\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)g_2(u)\mathrm{d}u$ is zero, then

$$\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(g_1)(u)\dot{g}_2(u)\mathrm{d}u = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)g_1(u)g_2(u)\mathrm{d}u.$$

■

With the previous lemma, we can rewrite the maximization problem as

$$\text{maximize} \qquad \frac{\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(p_\ell)(u)\mathcal{H}(q)(u)\mathrm{d}u\right)^2}{\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(q)(u)^2\mathrm{d}u} \tag{14}$$

$$\text{subject to} \qquad \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\mathcal{H}(q)(u)\mathrm{d}u = 0, \qquad \mathcal{H}(q)\left(\frac{\inf\mathcal{X}-\mathsf{x}}{h} \vee (-1)\right) = 0. \tag{15}$$

Theorem 8 (Variance Bound of the Minimum Distance Estimator).
An upper bound of the maximization problem (14) and (15) is

$$e_\ell'(\Gamma_{P,h}^{\perp})^{-1}\Sigma_{PP,h}^{\perp}(\Gamma_{P,h}^{\perp})^{-1}e_\ell - e_\ell'\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1} e_\ell.$$

Therefore, the asymptotic variance of the minimum distance estimator is bounded below by

$$f(\mathsf{x})e_\ell'\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1} e_\ell,$$

where $\dot{P}(u) = (1,\ u,\ u^2/2,\ u^3/3!,\ \cdots,\ u^{p-1}/(p-1)!)'$. ■

Example 2 (Local Linear/Quadratic Minimum Distance Density Estimation). Consider a simple example where $\ell = 0$ and $P(u) = u$, which means we focus on the asymptotic variance of the estimated density in a local linear regression. Also assume we employ a uniform kernel: $K(u) = \frac{1}{2}\mathbb{1}(|u| \leq 1)$, and that the integration region is $\frac{\mathcal{X}-\mathsf{x}}{h} = \mathbb{R}$ (i.e., x is an interior evaluation point). Note that this example also applies to local quadratic regressions, as u and u^2 are orthogonal for interior evaluation points.

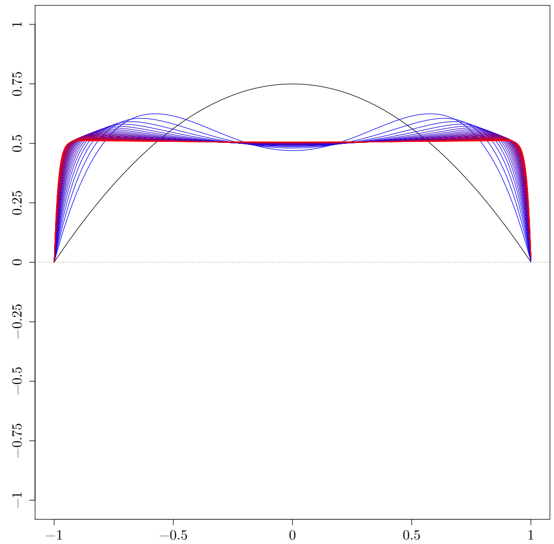

Figure 1. Equivalent Kernel of the Local Linear Minimum Distance Density Estimator.

Notes: The basis function $R(u)$ consists of an intercept, a linear term u (i.e., local linear regression), and an odd higher-order polynomial term u^{2j+1} for $j = 1, 2, \cdots, 30$. Without the higher-order polynomial regressor, the local linear density estimator using the uniform kernel is equivalent to the kernel density estimator using the Epanechnikov kernel (black line). Including a higher-order redundant regressor leads to an equivalent kernel that approaches the uniform kernel as j tends to infinity (red).

Taking $P(u) = u$, the variance bound in Theorem 8 is easily found to be

$$f(\mathsf{x})\left(\int_{-1}^{1}\dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1} = f(\mathsf{x})\frac{1}{2}.$$

We now calculate the asymptotic variance of the minimum distance estimator. To be specific, we choose $Q(u) = u^{2j+1}$, which is a higher-order polynomial function. With tedious calculation, one can show that the minimum distance estimator has the following asymptotic variance

$$\mathrm{Asy}\mathbb{V}[\hat{f}_{\mathtt{MD}}(\mathsf{x})] = f(\mathsf{x})\frac{11+4j}{20+8j},$$

which asymptotes to $f(\mathsf{x})/2$ as $j \to \infty$. As a result, it is possible to achieve the maximum amount of efficiency gain by including one higher-order polynomial and using our minimum distance estimator.

In Figure 1, we plot the equivalent kernel of the local linear minimum distance density estimator using a uniform kernel. Without the redundant regressor, it is equivalent to the kernel density estimator using the Epanechnikov kernel. As j gets larger, however, the equivalent kernel of the minimum distance estimator becomes closer to the uniform kernel, which is why, as $j \to \infty$, the minimum distance estimator has an asymptotic variance the same as the kernel density estimator using the uniform kernel. ■

Example 3 (Local Cubic Minimum Distance Estimation). We adopt the same setting in Example 2, i.e., local polynomial density estimation with the uniform kernel at an interior evaluation point. The difference is that we now consider a local cubic regression: $P(u) = (u, \frac{1}{2}u^2, \frac{1}{3!}u^3)'$.

As before, the variance bound in Theorem 8 is easily found to be

$$f(\mathsf{x})\left(\int_{-1}^{1}\dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1} = f(\mathsf{x})\begin{bmatrix} \frac{9}{8} & 0 & -\frac{15}{4} \\ 0 & \frac{3}{2} & 0 \\ -\frac{15}{4} & 0 & \frac{45}{2} \end{bmatrix}.$$

Again, we compute the asymptotic variance of our minimum distance estimator. Note, however, that now we have both odd and even order polynomials in our basis $P(u)$, therefore we include two higher-order polynomials. That is, we set $Q(u) = (u^{2j}, u^{2j+1})'$. The asymptotic variance of our minimum distance estimator is

$$\mathrm{Asy}\mathbb{V}\begin{bmatrix} \hat{f}_{\mathtt{MD}}(\mathsf{x}) \\ \hat{f}^{(1)}_{\mathtt{MD}}(\mathsf{x}) \\ \hat{f}^{(2)}_{\mathtt{MD}}(\mathsf{x}) \end{bmatrix} = f(\mathsf{x})\begin{bmatrix} \frac{9(4j+15)}{16(2j+7)} & 0 & -\frac{15(4j+17)}{8(2j+7)} \\ 0 & \frac{12j+39}{8j+20} & 0 \\ -\frac{15(4j+17)}{8(2j+7)} & 0 & \frac{45(4j+19)}{8j+28} \end{bmatrix},$$

which, again, asymptotes to the variance bound as $j \to \infty$. See also Table 1 for the efficiency gain of employing the minimum distance technique. ■

Example 4 (Local $p = 5$ Minimum Distance Estimation). We consider the same setting in Example 2 and 3, but with $p = 5$: $P(u) = (u, \frac{1}{2}u^2, \cdots, \frac{1}{5!}u^5)'$.

The variance bound in Theorem 8 is

$$f(\mathsf{x})\left(\int_{-1}^{1}\dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1} = f(\mathsf{x})\begin{bmatrix} \frac{225}{128} & 0 & -\frac{525}{32} & 0 & \frac{2835}{16} \\ 0 & \frac{75}{8} & 0 & -\frac{315}{4} & 0 \\ -\frac{525}{32} & 0 & \frac{2205}{8} & 0 & -\frac{14175}{4} \\ 0 & -\frac{315}{4} & 0 & \frac{1575}{2} & 0 \\ \frac{2835}{16} & 0 & -\frac{14175}{4} & 0 & \frac{99225}{2} \end{bmatrix}.$$

Again, we include two higher order polynomials: $Q(u) = (u^{2j}, u^{2j+1})'$. The asymptotic variance of our minimum distance estimator is

$$\mathrm{Asy}\mathbb{V}\begin{bmatrix} \hat{f}_{\mathtt{MD}}(\mathsf{x}) \\ \hat{f}^{(1)}_{\mathtt{MD}}(\mathsf{x}) \\ \hat{f}^{(2)}_{\mathtt{MD}}(\mathsf{x}) \\ \hat{f}^{(3)}_{\mathtt{MD}}(\mathsf{x}) \\ \hat{f}^{(4)}_{\mathtt{MD}}(\mathsf{x}) \end{bmatrix} = f(\mathsf{x})\begin{bmatrix} \frac{225(4j+19)}{256(2j+9)} & 0 & -\frac{525(4j+21)}{64(2j+9)} & 0 & \frac{2835(4j+23)}{32(2j+9)} \\ 0 & \frac{75(4j+17)}{16(2j+7)} & 0 & -\frac{315(4j+19)}{8(2j+7)} & 0 \\ -\frac{525(4j+21)}{64(2j+9)} & 0 & \frac{2205(4j+23)}{16(2j+9)} & 0 & -\frac{14175(4j+25)}{8(2j+9)} \\ 0 & -\frac{315(4j+19)}{8(2j+7)} & 0 & \frac{1575(4j+21)}{8j+28} & 0 \\ \frac{2835(4j+23)}{32(2j+9)} & 0 & -\frac{14175(4j+25)}{8(2j+9)} & 0 & \frac{99225(4j+27)}{8j+36} \end{bmatrix},$$

which converges to the variance bound as $j \to \infty$. See also Table 1 for the efficiency gain of employing the minimum distance technique. ■

Table 1. Variance Comparison.

(a) Density $f(\mathsf{x})$

	$p=1$	$p=2$	$p=3$	$p=4$
Kernel Function				
Uniform	0.600	0.600	1.250	1.250
Triangular	0.743	0.743	1.452	1.452
Epanechnikov	0.714	0.714	1.407	1.407
MD Variance Bound	0.500	0.500	1.125	1.125

(b) Density Derivative $f^{(1)}(\mathsf{x})$

	$p=2$	$p=3$	$p=4$	$p=5$
Kernel Function				
Uniform	2.143	2.143	11.932	11.932
Triangular	3.498	3.498	17.353	17.353
Epanechnikov	3.182	3.182	15.970	15.970
MD Variance Bound	1.500	1.500	9.375	9.375

Notes: Panel (a) compares asymptotic variance of the local polynomial density estimator of Cattaneo, Jansson, and Ma (2020) for different polynomial orders ($p = 1$, 2, 3, and 4) and different kernel functions (uniform, triangular and Epanechnikov). Also shown are the variance bound of the minimum distance estimator (MD Variance Bound), calculated according to Theorem 8. Panel(b) provides the same information for the estimated density derivative. All comparisons assume an interior evaluation point x.

Before closing this section, we make several remarks on the variance bound derived in Theorem 8, as well as to what extent it is achievable.

Remark 1 (Achievability of the Variance Bound). The previous two examples suggest that the variance bound derived in Theorem 8 can be achieved by employing a minimum distance estimator with two additional regressors, one higher-order even polynomial and one higher-order odd polynomial. With analytic calculation, we verify that this is indeed the case for $p \leq 10$ when a uniform kernel is used. ■

Remark 2 (Optimality of the Variance Bound). Granovsky and Müller (1991) discuss the problem of finding the optimal kernel function for kernel-type estimators. To be precise, consider the following

$$\frac{1}{nh^{\ell+1}}\sum_{i=1}^{n}\phi_{\ell,k}\left(\frac{x_i-\mathsf{x}}{h}\right),$$

where $\phi_{\ell,k}(u)$ is a function satisfying

$$\int_{-1}^{1} u^j \phi_{\ell,k}(u)\mathrm{d}u = \begin{cases} 0 & 0 \le j < k,\ j \ne \ell \\ \ell! & j = \ell \end{cases}, \qquad \int_{-1}^{1} u^k \phi_{\ell,k}(u)\mathrm{d}u \ne 0.$$

Then it is easy to see that, with a Taylor expansion argument,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{nh^{\ell+1}}\sum_{i=1}^{n}\phi_{\ell,k}\left(\frac{x_i - \mathsf{x}}{h}\right)\right] &= \frac{1}{h^{\ell+1}}\int_{-1}^{1}\phi_{\ell,k}\left(\frac{u-\mathsf{x}}{h}\right)f(u)\mathrm{d}u \\
&= \frac{1}{h^{\ell}}\int_{-1}^{1}\phi_{\ell,k}\left(u\right)f(\mathsf{x}+hu)\mathrm{d}u \\
&= \frac{1}{h^{\ell}}\int_{-1}^{1}\phi_{\ell,k}\left(u\right)\left[\sum_{j=0}^{k-1}\frac{(hu)^j}{j!}f^{(j)}(\mathsf{x}) + u^k O(h^k)\right]\mathrm{d}u \\
&= f^{(\ell)}(\mathsf{x}) + O(h^{k-\ell}).
\end{aligned}$$

That is, the kernel $\phi_{\ell,k}(u)$ facilitates estimating the ℓ-th derivative of the density function with a leading bias of order $h^{k-\ell}$. Asymptotic variance of this kernel-type estimator is easily found to be

$$\mathrm{Asy}\mathbb{V}\left[\frac{1}{nh^{\ell+1}}\sum_{i=1}^{n}\phi_{\ell,k}\left(\frac{x_i - \mathsf{x}}{h}\right)\right] = f(\mathsf{x})\int_{-1}^{1}\phi_{\ell,k}(u)^2\mathrm{d}u.$$

Granovsky and Müller (1991) provide the exact form of the kernel function $\phi_{\ell,k}(u)$ that minimizes the asymptotic variance subject to the order of the leading bias.

Take $\ell = 0$ and $k = 2$, $\phi_{\ell,k}(u)$ takes the following form:

$$\phi_{\ell,k}(u) = \frac{1}{2}\mathbb{1}(|u| \le 1),$$

which is the uniform kernel and minimizes variance among all second order kernels for density estimation. As illustrated in Example 2, our variance bound matches $f(\mathsf{x})\int_{-1}^{1}\phi_{\ell,k}(u)^2\mathrm{d}u$.

Now take $\ell = 1$ and $k = 3$. This will give an estimator for the density derivative $f^{(1)}(\mathsf{x})$ with a leading bias of order $O(h^2)$. The optimal choice of $\phi_{\ell,k}(u)$ is

$$\phi_{\ell,k}(u) = \frac{3}{2}u\mathbb{1}(|u| \le 1).$$

to match the order of bias, we consider the minimum distance estimator with $p = 3$. Again, the variance bound in Theorem 8 matches $f(\mathsf{x})\int_{-1}^{1}\phi_{\ell,k}(u)^2\mathrm{d}u$.

As a final illustration, take $\ell = 1$ and $k = 5$, which gives an estimator for the density derivative $f^{(1)}(\mathsf{x})$ with a leading bias of order $O(h^4)$. The optimal choice of $\phi_{\ell,k}(u)$ is

$$\phi_{\ell,k}(u) = \left(\frac{75}{8}u - \frac{105}{8}u^3\right)\mathbb{1}(|u| \le 1).$$

It is easy to see that $f(\mathsf{x})\int_{-1}^{1}\phi_{\ell,k}(u)^2\mathrm{d}u = 75f(\mathsf{x})/8$. To match the bias order, we take $p = 5$ for our minimum distance estimator. The variance bound is $75f(\mathsf{x})/8$, which is the same as $f(\mathsf{x})\int_{-1}^{1}\phi_{\ell,k}(u)^2\mathrm{d}u$.

With analytic calculations, we verify that the variance bound stated in Theorem 8 is the same as the minimum variance found in Granovsky and Müller (1991). Together with the previous remark, we reach a much stronger conclusion: including two higher-order polynomials in our minimum distance estimator can help achieve the variance bound in Theorem 8, which, in turn, is the smallest variance any kernel-type estimator can achieve (given a specific leading bias order). ■

Remark 3 (Another Density Estimator Which Achieves the Variance Bound). The following estimator achieves the bound of Theorem 8, although it does not belong to the class of estimators we consider in this paper.

$$\hat{\theta}_{\texttt{ND}} = \left(\int_{\mathcal{X}} \dot{P}(u-\mathsf{x})\dot{P}(u-\mathsf{x})'\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}u\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\dot{P}(x_i-\mathsf{x})\frac{1}{h}K\left(\frac{x_i-\mathsf{x}}{h}\right)\right),$$

where $\dot{P}(u) = (1, u, u^2/2, \cdots, u^{p-1}/(p-1)!)'$ is the $(p-1)$-th order polynomial basis. The subscript represents "numerical derivative," because the above estimator can be understood as

$$\begin{aligned}\hat{\theta}_{\texttt{ND}} &= \left(\int_{\mathcal{X}} \dot{P}(u-\mathsf{x})\dot{P}(u-\mathsf{x})'\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}u\right)^{-1}\left(\int_{\mathcal{X}}\dot{P}(u-\mathsf{x})\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\frac{\mathrm{d}\hat{F}(u)}{\mathrm{d}u}\mathrm{d}u\right)\\ &= \operatorname*{argmin}_{\theta}\int_{\mathcal{X}}\left(\frac{\mathrm{d}\hat{F}(u)}{\mathrm{d}u} - \dot{P}(u-\mathsf{x})'\theta\right)^2\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}u,\end{aligned}$$

where the derivative $\mathrm{d}\hat{F}(u)/\mathrm{d}u$ is interpreted in the sense of generalized functions. From the above, it is clear that this estimator requires the knowledge of the boundary position (that is, the knowledge of $\mathcal{X}$).

With straightforward calculations, this estimator has a leading bias

$$\begin{aligned}\mathbb{E}[\hat{\theta}_{\texttt{ND}}] &= \left(\int_{\mathcal{X}} \dot{P}(u-\mathsf{x})\dot{P}(u-\mathsf{x})'\frac{1}{h}K\left(\frac{u-\mathsf{x}}{h}\right)\mathrm{d}u\right)^{-1}\mathbb{E}\left[\dot{P}(x_i-\mathsf{x})\frac{1}{h}K\left(\frac{x_i-\mathsf{x}}{h}\right)\right]\\ &= \theta + h^p\Upsilon_h f^{(p)}(\mathsf{x})\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}\dot{P}(u)\dot{P}(u)'K(u)\mathrm{d}u\right)^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}\dot{P}(u)u^pK(u)\mathrm{d}u + o(h^p\Upsilon_h),\end{aligned}$$

where Υ_h is a diagonal matrix containing $1, h^{-1}, \cdots, h^{-(p-1)}$. Its leading variance is also easy to establish:

$$\mathbb{V}[\hat{\theta}_{\mathtt{ND}}] = \frac{1}{nh}\Upsilon_h f(\mathsf{x}) \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \dot{P}(u)\dot{P}(u)' K(u)\,\mathrm{d}u\right)^{-1} \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \dot{P}(u)\dot{P}(u)' K(u)^2\,\mathrm{d}u\right)$$
$$\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \dot{P}(u)\dot{P}(u)' K(u)\,\mathrm{d}u\right)^{-1} \Upsilon_h$$
$$+ o\left(\frac{1}{nh}\Upsilon_h^2\right).$$

To reach the efficiency bound in Theorem 8, it suffices to set $K(\cdot)$ to be the uniform kernel. Loader (2006), Section 5.1.1 also discussed this estimator, although it seems its efficiency property has not been realized in the literature. ■

4 Uniform Distribution Theory

We establish distribution approximation for $\{\hat{\theta}_G(\mathsf{x}), \mathsf{x} \in \mathcal{I}\}$ and $\{\hat{\theta}(\mathsf{x}), \mathsf{x} \in \mathcal{I}\}$, which can be viewed as processes indexed by the evaluation point x in some set $\mathcal{I} \subseteq \mathcal{X}$. Recall the definition of $\Gamma_{h,\mathsf{x}}$ and $\Sigma_{h,\mathsf{x}}$ from Section 1, and we define $\Omega_{h,\mathsf{x}} = \Gamma_{h,\mathsf{x}}^{-1}\Sigma_{h,\mathsf{x}}\Gamma_{h,\mathsf{x}}^{-1}$.

We first study the following (infeasible) centered and Studentized process:

$$\mathfrak{T}_G(\mathsf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big] K(u)\, g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}, \ \mathsf{x} \in \mathcal{I}, \tag{16}$$

where we consider linear combinations through a (known) vector $c_{h,\mathsf{x}}$, which can depend on the sample size through the bandwidth h, and can depend on the evaluation point. Again, we use the subscript G to denote the local L^2 approach with G being the design distribution. To economize notation, let

$$\mathcal{K}_{h,\mathsf{x}}(x) = \frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big] K(u)\, g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}},$$

then we can rewrite (16) as

$$\mathfrak{T}_G(\mathsf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{K}_{h,\mathsf{x}}(x_i),$$

and hence the centered and Studentized process $\mathfrak{T}_G(\cdot)$ takes a kernel form. The difference compared to standard kernel density estimators, however, is that the (equivalent) kernel in our case changes with the evaluation point, which is why our estimator is able to adapt to boundary points automatically. From the pointwise distribution theory developed in Section 2, the process $\mathfrak{T}_G(\mathsf{x})$ has variance

$$\mathbb{V}\left[\mathfrak{T}_G(\mathsf{x})\right] = \mathbb{E}\left[\mathcal{K}_{h,\mathsf{x}}(x_i)^2\right] = 1.$$

We can also compute the covariance as

$$\mathbb{C}\mathrm{ov}\left[\mathfrak{T}_G(\mathsf{x}), \mathfrak{T}_G(\mathsf{y})\right] = \mathbb{E}\left[\mathcal{K}_{h,\mathsf{x}}(x_i)\mathcal{K}_{h,\mathsf{y}}(x_i)\right] = \frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x},\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}} + O(h),$$

where $\Omega_{h,\mathsf{x},\mathsf{y}} = \Gamma_{h,\mathsf{x}}^{-1}\Sigma_{h,\mathsf{x},\mathsf{y}}\Gamma_{h,\mathsf{y}}^{-1}$, and

$$\Sigma_{h,\mathsf{x},\mathsf{y}} = \int_{\frac{\mathcal{X}-\mathsf{y}}{h}}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'\Big[F((\mathsf{x}+hu)\wedge(\mathsf{y}+hv)) - F(\mathsf{x}+hu)F(\mathsf{y}+hv)\Big]$$
$$K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{y}+hv)\mathrm{d}u\mathrm{d}v.$$

Of course we can further expand the above, but this is unnecessary for our purpose.

For future reference, let

$$r_1(\varepsilon, h) = \sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}, |\mathsf{x}-\mathsf{y}|\le\varepsilon} \left|c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h\right|, \qquad r_2(h) = \sup_{\mathsf{x}\in\mathcal{I}} \frac{1}{|c_{h,\mathsf{x}}'\Upsilon_h|}.$$

Remark 4 (On the Order of $r_1(\varepsilon, h)$, $r_2(h)$ and $\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x})$). In general, it is not possible to give precise orders of the quantities introduced above. In this remark, we consider the local polynomial estimator of Cattaneo, Jansson, and Ma (2020) (see Section 1 for an introduction). The local polynomial estimator employs a polynomial basis, and hence estimates the density function and higher-order derivatives by (it also estimates the distribution function)

$$\hat{F}^{(\ell)}(\mathsf{x}) = e_\ell'\hat{\theta}(\mathsf{x}),$$

where e_ℓ is the $(\ell+1)$-th standard basis vector. As a result, $c_{h,\mathsf{x}} = e_\ell$, which does not depend on the evaluation point. For the scaling matrix Υ_h, we note that it is diagonal with elements $1, h^{-1}, \cdots, h^{-p}$, and hence it does not depend on the evaluation point either. Therefore, we conclude that, for density (and higher-order) derivative estimation using the local polynomial estimator, $r_1(\varepsilon, h)$ is identically zero. Similarly, we have that $r_2(h) = h^\ell$. Finally, given the discussion in Section 1, the bias term generally has order $\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x}) = h^{p+1}$ for the local polynomial density estimator.

The above discussion restricts to the local polynomial density estimator, but more can be said about $r_2(h)$. We will argue that, in general, one should expect $r_2(h) = O(1)$. Recall that the leading variance of $c_{h,\mathsf{x}}'\hat{\theta}(\mathsf{x})$ and $c_{h,\mathsf{x}}'\hat{\theta}_G(\mathsf{x})$ is $\frac{1}{n}c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}$, and that the maximum eigenvalue of $\Omega_{h,\mathsf{x}}$ is bounded. Therefore, the variance has order $O(1/(nr_2(h)^2))$. In general, we do not expect the variance to shrink faster than $1/n$, which is why $r_2(h)$ is usually bounded. In fact, for most interesting cases, $c_{h,\mathsf{x}}'\hat{\theta}(\mathsf{x})$ and $c_{h,\mathsf{x}}'\hat{\theta}_G(\mathsf{x})$ will be "nonparametric" estimators in the sense that they estimate local features of the distribution function. If this is the case, we may even argue that $r_2(h)$ will be vanishing as the bandwidth shrinks. ■

We also make some additional assumptions.

Assumption 4. *Let $\mathcal{I}$ be a compact interval.*

(i) The density function is twice continuously differentiable and bounded away from zero in $\mathcal{I}$.

(ii) There exists some $\delta > 0$ and compactly supported kernel functions $K^\dagger(\cdot)$ and $\{K^{\ddagger,d}(\cdot)\}_{d\le\delta}$, such that (ii.1) $\sup_{u\in\mathbb{R}}|K^\dagger(u)|, \sup_{d\le\delta,u\in\mathbb{R}}|K^{\ddagger,d}(u)| < \infty$; (ii.2) the support of $K^{\ddagger,d}(\cdot)$ has Lebesgue measure bounded by Cd, where C is independent of d; and (ii.3) Further, for all u and v such that $|u-v|\le\delta$,

$$|K(u) - K(v)| \le |u-v|\cdot K^\dagger(u) + K^{\ddagger,|u-v|}(u).$$

(iii) The basis function $R(\cdot)$ is Lipschitz continuous in $[-1,1]$.

(iv) For all h sufficiently small, and the minimum eigenvalues of $\Gamma_{h,\mathsf{x}}$ and $h^{-1}\Sigma_{h,\mathsf{x}}$ are bounded away from zero uniformly for $\mathsf{x}\in\mathcal{I}$.

(v) $h\to 0$ and $nh/\log n\to\infty$ as $n\to\infty$.

(vi) For some $C_1, C_2 \ge 0$ and $C_3 \ge 0$,

$$r_1(\varepsilon, h) = O\left(\varepsilon^{C_1}h^{-C_2}\right), \qquad r_2(h) = O\left(h^{C_3}\right).$$

In addition,

$$\frac{\sup_{\mathsf{x}\in\mathcal{I}}|c_{h,\mathsf{x}}'\Upsilon_h|}{\inf_{\mathsf{x}\in\mathcal{I}}|c_{h,\mathsf{x}}'\Upsilon_h|} = O(1).$$

■

Assumption 5. *The design density function $g(\cdot)$ is twice continuously differentiable and bounded away from zero in $\mathcal{I}$.* ■

For any $h > 0$ (and fixed n), we can define a centered Gaussian process, $\{\mathfrak{B}_G(\mathsf{x}) : \mathsf{x}\in\mathcal{I}\}$, which has the same variance-covariance structure as the process $\mathfrak{T}_G(\cdot)$. The following lemma shows that it is possible to construct such a process, and that $\mathfrak{T}_G(\cdot)$ and $\mathfrak{B}_G(\cdot)$ are "close in distribution."

Theorem 9 (Strong Approximation). *Assume Assumptions 1, 2, 4 and 5 hold. Then on a possibly enlarged probability space there exist two processes, $\{\tilde{\mathfrak{T}}_G(\mathsf{x}) : \mathsf{x}\in\mathcal{I}\}$ and $\{\mathfrak{B}_G(\mathsf{x}) : \mathsf{x}\in\mathcal{I}\}$, such that (i) $\tilde{\mathfrak{T}}_G(\cdot)$ has the same distribution as $\mathfrak{T}_G(\cdot)$; (ii) $\mathfrak{B}_G(\cdot)$ is a Gaussian process with the same covariance structure as $\mathfrak{T}_G(\cdot)$; and (iii)*

$$\mathbb{P}\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\tilde{\mathfrak{T}}_G(\mathsf{x}) - \mathfrak{B}_G(\mathsf{x})\right| > \frac{C_4(u + C_1\log n)}{\sqrt{nh}}\right] \le C_2 e^{-C_3 u},$$

where C_1, C_2, C_3 and C_4 are constants that do not depend on h or n. ■

Next we consider the continuity property of the implied (equivalent) kernel of the process $\mathfrak{T}_G(\cdot)$, which will help control the complexity of the Gaussian process $\mathfrak{B}_G(\cdot)$. To be precise, define the pseudo-metric $\sigma_G(\mathsf{x},\mathsf{y})$ as

$$\sigma_G(\mathsf{x},\mathsf{y}) = \sqrt{\mathbb{V}\left[\mathfrak{T}_G(\mathsf{x}) - \mathfrak{T}_G(\mathsf{y})\right]} = \sqrt{\mathbb{E}\left[(\mathcal{K}_{h,\mathsf{x}}(x_i) - \mathcal{K}_{h,\mathsf{y}}(x_i))^2\right]},$$

we would like to provide an upper bound of $\sigma_G(\mathsf{x},\mathsf{y})$ in terms of $|\mathsf{x}-\mathsf{y}|$ (at least for all x and y such that $|\mathsf{x}-\mathsf{y}|$ is small enough).

Lemma 10 (VC-type Property). *Assume Assumptions 1, 2, 4 and 5 hold. Then for all* $\mathsf{x},\mathsf{y}\in\mathcal{I}$ *such that* $|\mathsf{x}-\mathsf{y}| = \varepsilon \le h$,

$$\sigma_G(\mathsf{x},\mathsf{y}) = O\left(\frac{1}{\sqrt{h}}\frac{\varepsilon}{h} + \frac{1}{\sqrt{h}} r_1(\varepsilon,h) r_2(h) + \frac{1}{h} r_1(\varepsilon,h)^2 r_2(h)^2\right).$$

Therefore,

$$\mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}} |\mathfrak{B}_G(\mathsf{x})|\right] = O\left(\sqrt{\log n}\right), \qquad \text{and} \qquad \mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}} |\mathfrak{T}_G(\mathsf{x})|\right] = O\left(\sqrt{\log n}\right).$$

■

4.1 Local L^2 Distribution Estimation

We first discuss the covariance estimator. For the local L^2 distribution estimator, let $\hat{\Omega}_{h,\mathsf{x},\mathsf{y}} = \Gamma_{h,\mathsf{x}}^{-1}\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}\Gamma_{h,\mathsf{y}}^{-1}$ with $\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}$ given by

$$\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}} = \frac{1}{n}\sum_{i=1}^{n} \int_{\frac{\mathcal{X}-\mathsf{y}}{h}} \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)' \Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - \hat{F}(\mathsf{x}+hu)\Big]\Big[\mathbb{1}(x_i \le \mathsf{y}+hv) - \hat{F}(\mathsf{y}+hv)\Big]$$
$$K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{y}+hv)\mathrm{d}u\mathrm{d}v.$$

The next lemma characterizes the convergence rate of $\hat{\Omega}_{h,\mathsf{x},\mathsf{y}}$.

Lemma 11 (Local L^2 Distribution Estimation: Covariance Estimation). *Assume Assumptions 1, 2, 4 and 5 hold, and that* $nh^2/\log n \to \infty$. *Then*

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}} \left| \frac{c_{h,\mathsf{x}}'\Upsilon_h(\hat{\Omega}_{h,\mathsf{x},\mathsf{y}} - \Omega_{h,\mathsf{x},\mathsf{y}})\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}} \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^2}}\right)$$

■

We now consider the estimator $c_{h,\mathsf{x}}'\hat{\theta}_G(\mathsf{x})$. From (3) and (4), one has

$$
\begin{aligned}
T_G(\mathsf{x}) &= \frac{\sqrt{n}c_{h,\mathsf{x}}'\left(\hat{\theta}_G(\mathsf{x}) - \theta(\mathsf{x})\right)}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} \\
&= \sqrt{n}\frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_h^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big]K\left(u\right)g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} && (17) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K\left(u\right)g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}. && (18)
\end{aligned}
$$

In the following two lemmas, we analyze the two terms in the above decomposition separately.

Lemma 12. *Assume Assumptions 1, 2, 4 and 5 hold, and that $nh^2/\log n \to \infty$. Then*

$$
\sup_{\mathsf{x}\in\mathcal{I}}\Big|(17)\Big| = O_{\mathbb{P}}\left(\sqrt{\frac{n}{h}}\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x})\right).
$$

■

Lemma 13. *Assume Assumptions 1, 2, 4 and 5 hold, and that $nh^2/\log n \to \infty$. Then*

$$
\sup_{\mathsf{x}\in\mathcal{I}}\Big|(18) - \mathfrak{T}_G(\mathsf{x})\Big| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{nh^2}}\right).
$$

■

Now we state the main result on uniform distributional approximation.

Theorem 14 (Local L^2 Distribution Estimation: Uniform Distributional Approximation). *Assume Assumptions 1, 2, 4 and 5 hold, and that $nh^2/\log n \to \infty$. Then on a possibly enlarged probability space there exist two processes, $\{\tilde{\mathfrak{T}}_G(\mathsf{x}) : \mathsf{x}\in\mathcal{I}\}$ and $\{\mathfrak{B}_G(\mathsf{x}) : \mathsf{x}\in\mathcal{I}\}$, such that (i) $\tilde{\mathfrak{T}}_G(\cdot)$ has the same distribution as $\mathfrak{T}_G(\cdot)$; (ii) $\mathfrak{B}_G(\cdot)$ is a Gaussian process with the same covariance structure as $\mathfrak{T}_G(\cdot)$; and (iii)*

$$
\sup_{\mathsf{x}\in\mathcal{I}}\Big|T_G(\mathsf{x}) - \mathfrak{T}_G(\mathsf{x})\Big| + \sup_{\mathsf{x}\in\mathcal{I}}\Big|\tilde{\mathfrak{T}}_G(\mathsf{x}) - \mathfrak{B}_G(\mathsf{x})\Big| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{nh^2}} + \sqrt{\frac{n}{h}}\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x})\right).
$$

■

The following theorem shows that a feasible approximation to the process $\mathfrak{B}_G(\cdot)$ can be achieved by simulating a Gaussian process with covariance estimated from the data. In the following, we use $\mathbb{P}^\star$, $\mathbb{E}^\star$ and $\mathbb{C}\mathrm{ov}^\star$ to denote the probability, expectation and covariance operator conditioning on the data.

Theorem 15 (Local L^2 Distribution Estimation: Feasible Distributional Approximation). *Assume Assumptions 1, 2, 4 and 5 hold, and that $nh^2/\log n \to \infty$. Then on a possibly enlarged probability space there exist two centered Gaussian processes, $\{\tilde{\mathfrak{B}}_G(\mathsf{x}), \mathsf{x}\in\mathcal{I}\}$ and $\{\hat{\mathfrak{B}}_G(\mathsf{x}), \mathsf{x}\in\mathcal{I}\}$, satisfying (i) $\tilde{\mathfrak{B}}_G(\mathsf{x})$ is independent of the data, and has the same distribution as*

$\mathfrak{B}_G(\cdot)$; *(ii)* $\hat{\mathfrak{B}}_G(\cdot)$ *has covariance*

$$\mathbb{C}\mathrm{ov}^\star\left[\hat{\mathfrak{B}}_G(\mathsf{x}), \hat{\mathfrak{B}}_G(\mathsf{y})\right] = \frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x},\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}};$$

and (iii)

$$\begin{aligned}&\mathbb{E}^\star\left[\sup_{\mathsf{x}\in\mathcal{I}}|\tilde{\mathfrak{B}}_G(\mathsf{x}) - \hat{\mathfrak{B}}_G(\mathsf{x})|\right]\\ =&O_{\mathbb{P}}\left(\sqrt{\log n}\inf_{\varepsilon\le h}\left[\frac{1}{\sqrt{h}}\frac{\varepsilon}{h} + \frac{1}{\sqrt{h}}r_1(\varepsilon,h)r_2(h) + \frac{1}{h}r_1(\varepsilon,h)^2 r_2(h)^2 + \sqrt{\frac{1}{\varepsilon}\sqrt{\frac{\log n}{n}}}\right]\right).\end{aligned}$$

■

Remark 5 (On the Remainders in Theorems 14 and 15). Both errors involve numerous quantities which are difficult to further simplify. To understand the magnitude of these two error bounds, we again consider the local polynomial estimator of Cattaneo, Jansson, and Ma (2020) (see Section 1 for a brief introduction, and also Remark 4).

Recall that the local polynomial density estimator employs a polynomial basis, which implies that $\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x}) = h^{p+1}$, where p is the highest polynomial order. Then the error in Theorem 14 reduces to

$$\sqrt{nh^{2p+1}} + \frac{\log n}{\sqrt{nh^2}}.$$

As discussed in Remark 4, $r_1(\varepsilon,h) = 0$ for density (or higher-order derivative) estimation, which implies that the error in Theorem 15 becomes

$$\sqrt{\log n}\inf_{\varepsilon\le h}\left(\frac{1}{\sqrt{h}}\frac{\varepsilon}{h} + \sqrt{\frac{1}{\varepsilon}\sqrt{\frac{\log n}{n}}}\right) = O\left(\left(\frac{\log^2 n}{\sqrt{nh^3}}\right)^{1/3}\right).$$

A sufficient set of conditions for both errors to be negligible is $nh^{2p+1}\to 0$ and $nh^3/\log^4 n\to\infty$. ■

4.2 Local Regression Distribution Estimation

Now we consider the local regression estimator $\{\hat{\theta}(\mathsf{x}), \mathsf{x}\in\mathcal{I}\}$. As before, we first discuss the construction of the covariance $\Omega_{h,\mathsf{x},\mathsf{y}}$. Let $\hat{\Omega}_{h,\mathsf{x},\mathsf{y}} = \hat{\Gamma}_{h,\mathsf{x}}^{-1}\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}\hat{\Gamma}_{h,\mathsf{y}}^{-1}$. Construction of $\hat{\Gamma}_{h,\mathsf{x}}$ is given in Section 2.2. The following lemma shows that $\hat{\Gamma}_{h,\mathsf{x}}$ is uniformly consistent.

Lemma 16 (Uniform Consistency of $\hat{\Gamma}_{h,\mathsf{x}}$). *Assume Assumptions 1 and 4 hold. Then*

$$\sup_{\mathsf{x}\in\mathcal{I}}\left|\hat{\Gamma}_{h,\mathsf{x}} - \Gamma_{h,\mathsf{x}}\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh}}\right).$$

■

Construction of $\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}$ also mimics that in Section 2.2. To be precise, we let

$$\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}\Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-\hat{F}(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\right]$$
$$\left[\frac{1}{n}\sum_{j=1}^{n}\Upsilon_h R(x_j-\mathsf{y})\Big[\mathbb{1}(x_i\le x_j)-\hat{F}(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{y}}{h}\right)\right]'.$$

where $\hat{F}(\cdot)$ remains to be the empirical distribution function. The following result justifies consistency of $\hat{\Omega}_{h,\mathsf{x},\mathsf{y}}$.

Lemma 17 (Local Regression Distribution Estimation: Covariance Estimation). *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n\to\infty$. Then*

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|\frac{c_{h,\mathsf{x}}'\Upsilon_h(\hat{\Omega}_{h,\mathsf{x},\mathsf{y}}-\Omega_{h,\mathsf{x},\mathsf{y}})\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^2}}\right)$$

■

The following is an expansion of $T(\cdot)$.

$$T(\mathsf{x}) = \frac{\sqrt{n}c_{h,\mathsf{x}}'\Big(\hat{\theta}(\mathsf{x})-\theta(\mathsf{x})\Big)}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}$$
$$= \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}\Upsilon_h R(x_i-\mathsf{x})[1-F(x_i)]\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} \tag{19}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}\Upsilon_h R(x_i-\mathsf{x})[F(x_i)-\theta(\mathsf{x})'R(x_i-\mathsf{x})]\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} \tag{20}$$
$$+\frac{1}{n\sqrt{n}}\sum_{i,j=1,i\neq j}^{n}\frac{1}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\left\{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}\Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\le x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\right.$$
$$\left.-\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}R(u)\Big[\mathbb{1}(x_i\le\mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]K(u)f(\mathsf{x}+hu)\mathrm{d}u\right\} \tag{21}$$
$$+\frac{n-1}{n\sqrt{n}}\sum_{i=1}^{n}\frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}R(u)\Big[\mathbb{1}(x_i\le\mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]K(u)f(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}. \tag{22}$$

Lemma 18. *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n\to\infty$. Then*

$$\sup_{\mathsf{x}\in\mathcal{I}}\Big|(19)\Big| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh}}\right).$$

■

Lemma 19. *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n \to \infty$. Then*

$$\sup_{\mathsf{x}\in\mathcal{I}} \Big|(20)\Big| = O_{\mathbb{P}}\Big(\sqrt{\frac{n}{h}} \sup_{\mathsf{x}\in\mathcal{I}} \varrho(h,\mathsf{x})\Big).$$

■

Lemma 20. *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n \to \infty$. Then*

$$\sup_{\mathsf{x}\in\mathcal{I}} \Big|(21)\Big| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{nh^2}}\right).$$

■

Lemma 21. *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n \to \infty$. Then*

$$\sup_{\mathsf{x}\in\mathcal{I}} \Big|(22) - \mathfrak{T}_F(\mathsf{x})\Big| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{nh^2}}\right).$$

■

Finally we have the following result on uniform distributional approximation for the local regression distribution estimator, as well as a feasible approximation by simulating from a Gaussian process with estimated covariance.

Theorem 22 (Local Regression Distribution Estimation: Uniform Distributional Approximation). *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n \to \infty$. Then on a possibly enlarged probability space there exist two processes, $\{\tilde{\mathfrak{T}}_F(\mathsf{x}) : \mathsf{x} \in \mathcal{I}\}$ and $\{\mathfrak{B}_F(\mathsf{x}) : \mathsf{x} \in \mathcal{I}\}$, such that (i) $\tilde{\mathfrak{T}}_F(\cdot)$ has the same distribution as $\mathfrak{T}_F(\cdot)$; (ii) $\mathfrak{B}_F(\cdot)$ is a Gaussian process with the same covariance structure as $\mathfrak{T}_F(\cdot)$; and (iii)*

$$\sup_{\mathsf{x}\in\mathcal{I}} \Big|T(\mathsf{x}) - \mathfrak{T}_F(\mathsf{x})\Big| + \sup_{\mathsf{x}\in\mathcal{I}} \Big|\tilde{\mathfrak{T}}_F(\mathsf{x}) - \mathfrak{B}_F(\mathsf{x})\Big| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{nh^2}} + \sqrt{\frac{n}{h}} \sup_{\mathsf{x}\in\mathcal{I}} \varrho(h,\mathsf{x})\right).$$

■

Theorem 23 (Local Regression Distribution Estimation: Feasible Distributional Approximation). *Assume Assumptions 1 and 4 hold, and that $nh^2/\log n \to \infty$. Then on a possibly enlarged probability space there exist two centered Gaussian processes, $\{\tilde{\mathfrak{B}}_F(\mathsf{x}), \mathsf{x} \in \mathcal{I}\}$ and $\{\hat{\mathfrak{B}}_F(\mathsf{x}), \mathsf{x} \in \mathcal{I}\}$, satisfying (i) $\tilde{\mathfrak{B}}_F(\mathsf{x})$ is independent of the data, and has the same distribution as $\mathfrak{B}_F(\cdot)$; (ii) $\hat{\mathfrak{B}}_F(\cdot)$ has covariance*

$$\mathbb{C}\mathrm{ov}^\star\left[\hat{\mathfrak{B}}_F(\mathsf{x}), \hat{\mathfrak{B}}_F(\mathsf{y})\right] = \frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x},\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}};$$

and (iii)

$$\begin{aligned}&\mathbb{E}^\star\left[\sup_{\mathsf{x}\in\mathcal{I}} |\tilde{\mathfrak{B}}_F(\mathsf{x}) - \hat{\mathfrak{B}}_F(\mathsf{x})|\right]\\ =&O_{\mathbb{P}}\left(\sqrt{\log n}\inf_{\varepsilon\le h}\left[\frac{1}{\sqrt{h}}\frac{\varepsilon}{h} + \frac{1}{\sqrt{h}}r_1(\varepsilon,h)r_2(h) + \frac{1}{h}r_1(\varepsilon,h)^2r_2(h)^2 + \sqrt{\frac{1}{\varepsilon}\sqrt{\frac{\log n}{n}}}\,\right]\right).\end{aligned}$$

■

5 Proofs

5.1 Additional Preliminary Lemmas

Lemma 24. *Assume* $\{u_{i,h}(a) : \ a \in A \subset \mathbb{R}^d\}$ *are independent across* i, *and* $\mathbb{E}[u_{i,h}(a)] = 0$ *for all* $a \in A$ *and all* $h > 0$. *In addition, assume for each* $\varepsilon > 0$ *there exists* $\{u_{i,h,\varepsilon}(a) : a \in A\}$, *such that*

$$|a - b| \leq \varepsilon \quad \Rightarrow \quad |u_{i,h}(a) - u_{i,h}(b)| \leq u_{i,h,\varepsilon}(a).$$

Define

$$C_1 = \sup_{a\in A} \max_{1\leq i\leq n} \mathbb{V}[u_{i,h}(a)], \qquad C_2 = \sup_{a\in A} \max_{1\leq i\leq n} |u_{i,h}(a)|$$

$$C_{1,\varepsilon} = \sup_{a\in A} \max_{1\leq i\leq n} \mathbb{V}[u_{i,h,\varepsilon}(a)], \quad C_{2,\varepsilon} = \sup_{a\in A} \max_{1\leq i\leq n} |u_{i,h,\varepsilon}(a) - \mathbb{E}[u_{i,h,\varepsilon}(a)]|, \quad C_{3,\varepsilon} = \sup_{a\in A} \max_{1\leq i\leq n} \mathbb{E}[|u_{i,h,\varepsilon}(a)|].$$

Then

$$\sup_{a\in A} \left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) \right| = O_{\mathbb{P}}\left(\gamma + \gamma_\varepsilon + C_{3,\varepsilon}\right),$$

where γ *and* γ_ε *are any sequences satisfying*

$$\frac{\gamma^2 n}{(C_1 + \frac{1}{3}\gamma C_2)\log N(\varepsilon, A, |\cdot|)} \quad \text{and} \quad \frac{\gamma_\varepsilon^2 n}{(C_{1,\varepsilon} + \frac{1}{3}\gamma_\varepsilon C_{2,\varepsilon})\log N(\varepsilon, A, |\cdot|)} \quad \text{are bounded from below,}$$

and $N(\varepsilon, A, |\cdot|)$ *is the covering number of* A. ■

Remark 6. Provided that $u_{i,h}(\cdot)$ is reasonably smooth, one can always choose ε (as a function of n and h) small enough, and the leading order will be given by γ (and hence is determined by C_1 and C_2). ■

Proof. Let A_ε be an ε-covering of A, then

$$\sup_{a\in A} \left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) \right| \leq \sup_{a\in A_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) \right| + \sup_{a\in A_\varepsilon, b\in A, |a-b|\leq\varepsilon} \left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) - u_{i,h}(b) \right|.$$

Next we apply the union bound and Bernstein's inequality:

$$\begin{aligned}
\mathbb{P}\left[\sup_{a\in A_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) \right| \geq \gamma u \right] &\leq N(\varepsilon, A, |\cdot|) \sup_{a\in A} \mathbb{P}\left[\left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) \right| \geq \gamma u \right] \\
&\leq 2N(\varepsilon, A, |\cdot|) \exp\left\{-\frac{1}{2} \frac{\gamma^2 n u^2}{C_1 + \frac{1}{3}\gamma C_2 u}\right\} \\
&= 2\exp\left\{-\frac{1}{2} \frac{\gamma^2 n u^2}{C_1 + \frac{1}{3}\gamma C_2 u} + \log N(\varepsilon, A, |\cdot|)\right\}.
\end{aligned}$$

Now take u sufficiently large, then the above is further bounded by:

$$\mathbb{P}\left[\sup_{a\in A_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^{n} u_{i,h}(a) \right| \geq \gamma u \right] \leq 2\exp\left\{-\log N(\varepsilon, A, |\cdot|)\left[\frac{1}{2} \frac{1}{\log N(\varepsilon, A, |\cdot|)} \frac{\gamma^2 n}{C_1 + \frac{1}{3}\gamma C_2} u - 1\right]\right\},$$

which tends to zero if $\log N(\varepsilon, A, |\cdot|) \to \infty$ and

$$\frac{\gamma^2 n}{(C_1 + \frac{1}{3}\gamma C_2)\log N(\varepsilon, A, |\cdot|)} \text{ is bounded from below,}$$

in which case we have

$$\sup_{a\in A_\varepsilon}\left|\frac{1}{n}\sum_{i=1}^n u_{i,h}(a)\right| = O_{\mathbb{P}}\left(\gamma\right).$$

We can apply the same technique to the other term, and obtain

$$\sup_{a\in A_\varepsilon, b\in A, |a-b|\le\varepsilon}\left|\frac{1}{n}\sum_{i=1}^n u_{i,h}(a) - u_{i,h}(b)\right| = O_{\mathbb{P}}\left(\gamma_\varepsilon\right),$$

where γ_ε is any sequence satisfying

$$\frac{\gamma_\varepsilon^2 n}{(C_{1,\varepsilon}+\frac{1}{3}\gamma_\varepsilon C_{2,\varepsilon})\log N(\varepsilon, A, |\cdot|)} \text{ is bounded from below.}$$

■

Lemma 25 (Equation (3.5) of Giné, Latała, and Zinn (2000)). *For a degenerate second order U-statistic, $\sum_{i,j=1,i\neq j}^n h_{ij}(x_i,x_j)$, the following holds:*

$$\mathbb{P}\left[\left|\sum_{i,j,i\neq j}^n u_{ij}(x_i,x_j)\right| > t\right] \le C\exp\left\{-\frac{1}{C}\min\left[\frac{t}{D},\ \left(\frac{t}{B}\right)^{\frac{2}{3}},\ \left(\frac{t}{A}\right)^{\frac{1}{2}}\right]\right\},$$

where C is some universal constant, and A, B and D are any constants satisfying

$$\begin{aligned}
A &\ge \max_{1\le i,j\le n}\sup_{u,v}|u_{ij}(u,v)| \\
B^2 &\ge \max_{1\le i,j\le n}\left[\sup_v\left|\sum_{i=1}^n \mathbb{E}u_{ij}(x_i,v)^2\right|,\ \sup_u\left|\sum_{j=1}^n \mathbb{E}u_{ij}(u,x_j)^2\right|\right] \\
D^2 &\ge \sum_{i,j=1,i\neq j}^n \mathbb{E}u_{ij}(x_i,x_j)^2.
\end{aligned}$$

■

5.2 Proof of Theorem 1

Part (i)

The bias term can be bounded by

$$\begin{aligned}
\left|\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big]K(u)\,\mathrm{d}u\right| &\le \sup_{u\in[-1,1]}\Big|F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big|\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}|R(u)|K(u)\,\mathrm{d}u \\
&= \varrho(h)\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}|R(u)|K(u)\,\mathrm{d}u.
\end{aligned}$$

Part (ii)

The variance can be found by

$$\begin{aligned}
&\mathbb{V}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i\le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u\right] \\
&= \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'K(u)K(v)\Big[F(\mathsf{x}+h(u\wedge v)) - F(\mathsf{x}+hu)F(\mathsf{x}+hv)\Big]g(\mathsf{x}+hu)g(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v.
\end{aligned}$$

To establish asymptotic normality, we verify a Lyapunov-type condition with a fourth moment calculation. Take c to be a nonzero vector of conformable dimension, and we employ the Cramer-Wold device:

$$\frac{1}{n}\left(c'\Sigma_h c\right)^{-2}\mathbb{E}\left[\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} c'R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K\left(u\right)g(\mathsf{x}+hu)\mathrm{d}u\right)^4\right].$$

If $c'\Sigma_h c$ is bounded away from zero as the bandwidth decreases, the above will have order n^{-1}, as $K(\cdot)$ is bounded and compactly supported and $R(\cdot)$ is locally bounded. Therefore, the Lyapunov condition holds in this case. The more challenging case is when $c'\Sigma_h c$ is of order h. In this case, it implies

$$F(\mathsf{x})(1-F(\mathsf{x}))\left|\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} c'R(u)K(u)g_u\mathrm{d}u\right|^2 = O(h).$$

Now consider the fourth moment. The leading term is

$$F(\mathsf{x})(1-F(\mathsf{x}))(3F(\mathsf{x})^2-3F(\mathsf{x})+1)\left|\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} c'R(u)K(u)g(\mathsf{x}+hu)\mathrm{d}u\right|^4 = O(h),$$

meaning that for the Lyapunov condition to hold, we need the requirement that $nh\to\infty$.

Part (iii)

This follows immediately from Part (i) and (ii).

5.3 Proof of Theorem 2

To study the property of $\hat{\Sigma}_h$, we make the following decomposition:

$$\hat{\Sigma}_h = \frac{1}{n}\sum_{i=1}^n \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'\Big[\mathbb{1}(x_i\le \mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]\Big[\mathbb{1}(x_i\le \mathsf{x}+hv)-F(\mathsf{x}+hv)\Big]K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v \tag{I}$$

$$-\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)R(v)'\Big[\hat{F}(\mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]\Big[\hat{F}(\mathsf{x}+hv)-F(\mathsf{x}+hv)\Big]K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v. \tag{II}$$

First, it is obvious that term (II) is of order $O_{\mathbb{P}}(1/n)$. Term (I) requires more delicate analysis. Let c be a vector of unit length and suitable dimension, and define

$$c_i = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} c'R(u)R(v)'c\Big[\mathbb{1}(x_i\le \mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]\Big[\mathbb{1}(x_i\le \mathsf{x}+hv)-F(\mathsf{x}+hv)\Big]K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v.$$

Then

$$c'(\mathrm{I})c = \mathbb{E}[c'(\mathrm{I})c] + O_{\mathbb{P}}\left(\sqrt{\mathbb{V}[c'(\mathrm{I})c]}\right) = \mathbb{E}[c_i] + O_{\mathbb{P}}\left(\sqrt{\frac{1}{n}\left(\mathbb{E}[c_i^2]-(\mathbb{E}[c_i])^2\right)}\right),$$

which implies that

$$\frac{c'(\mathrm{I})c}{\mathbb{E}[c'(\mathrm{I})c]} - 1 = O_{\mathbb{P}}\left(\sqrt{\frac{1}{n}\left(\frac{\mathbb{E}[c_i^2]}{(\mathbb{E}[c_i])^2}-1\right)}\right).$$

With the same argument used in the proof of Theorem 1, one can show that

$$\frac{\mathbb{E}[c_i^2]}{(\mathbb{E}[c_i])^2} = O\left(\frac{1}{h}\right),$$

which implies

$$\frac{c'(\mathrm{I})c}{c'\Sigma_h c} - 1 = O_{\mathbb{P}}\left(\sqrt{\frac{1}{nh}}\right).$$

5.4 Proof of Theorem 3

Part (i)

For the "denominator," its variance is bounded by

$$\begin{aligned}
&\left|\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}\Upsilon_h R(x_i-\mathsf{x})R(x_i-\mathsf{x})'\Upsilon_h\frac{1}{h}K\left(\frac{x_i-\mathsf{x}}{h}\right)\right]\right| \le \frac{1}{n}\mathbb{E}\left[\left|\Upsilon_h R(x_i-\mathsf{x})R(x_i-\mathsf{x})'\Upsilon_h\right|^2\frac{1}{h^2}K\left(\frac{x_i-\mathsf{x}}{h}\right)^2\right]\\
&= \frac{1}{n}\int_{\mathcal{X}}\left|\Upsilon_h R(u-\mathsf{x})R(u-\mathsf{x})'\Upsilon_h\right|^2\frac{1}{h^2}K\left(\frac{u-\mathsf{x}}{h}\right)^2 f(u)\mathrm{d}u = \frac{1}{nh}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}\left|R(u)R(u)'\right|^2 K(u)^2 f(\mathsf{x}+hu)\mathrm{d}u\\
&= O\left(\frac{1}{nh}\right).
\end{aligned}$$

Therefore, under the assumption that $h\to 0$ and $nh\to\infty$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}\Upsilon_h R(x_i-\mathsf{x})R(x_i-\mathsf{x})'\Upsilon_h\frac{1}{h}K\left(\frac{x_i-\mathsf{x}}{h}\right) - \Gamma_h\right| = O_{\mathbb{P}}\left(\sqrt{\frac{1}{nh}}\right).$$

As a result, we have

$$\hat{\theta}-\theta = \Upsilon_h\Gamma_h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\Upsilon_h R(x_i-\mathsf{x})\left[\hat{F}(x_i) - R(x_i-\mathsf{x})'\theta_0\right]\frac{1}{h}K\left(\frac{x_i-\mathsf{x}}{h}\right)\right)(1+o_{\mathbb{P}}(1)).$$

Part (ii)

The order of the leave-in bias is clearly $1/n$. For the approximation bias (7), we obtained its mean in the proof of Theorem 1 by setting $G=F$, which has an order of $\varrho(h)$. The approximation bias has a variance of order

$$\begin{aligned}
&\left|\mathbb{V}\left[\frac{1}{n}\sum_{j=1}^{n}\Upsilon_h R(x_j-\mathsf{x})\left[F(x_j)-R(x_j-\mathsf{x})'\theta_0\right]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\right]\right|\\
&\le \frac{1}{n}\mathbb{E}\left[\left|\Upsilon_h R(x_j-\mathsf{x})\left[F(x_j)-R(x_j-\mathsf{x})'\theta_0\right]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right)\right|^2\right]\\
&= \frac{1}{n}\int_{\mathcal{X}}\left|\Upsilon_h R(u-\mathsf{x})\left[F(u)-R(u-\mathsf{x})'\theta_0\right]\right|^2\frac{1}{h^2}K\left(\frac{u-\mathsf{x}}{h}\right)^2 f(u)\mathrm{d}u\\
&= \frac{1}{nh}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}\left|R(u)\left[F(\mathsf{x}+hu)-R(u-\mathsf{x})'\theta_0\right]\right|^2 K(u)^2 f(\mathsf{x}+hu)\mathrm{d}u\\
&\le \frac{1}{nh}\varrho(h)^2\int_{\frac{\mathcal{X}-\mathsf{x}}{h}}|R(u)|^2K(u)^2f(\mathsf{x}+hu)\mathrm{d}u\\
&= O\left(\frac{\rho(h)^2}{nh}\right).
\end{aligned}$$

Therefore,

$$(7) = O_{\mathbb{P}}\left(\varrho(h)+\varrho(h)\sqrt{\frac{1}{nh}}\right) = O_{\mathbb{P}}(\varrho(h)),$$

under the assumption that $nh\to\infty$.

Part (iii)

We compute the variance of the U-statistic (9). For simplicity, define

$$u_{ij} = \Upsilon_h R(x_j - \mathsf{x})\Big[\mathbb{1}(x_i \le x_j) - F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j - \mathsf{x}}{h}\right) - \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x} + hu) - F(\mathsf{x} + hu)\Big]K\left(u\right)f(\mathsf{x} + hu)\mathrm{d}u,$$

which satisfies $\mathbb{E}[u_{ij}] = \mathbb{E}[u_{ij}|x_i] = \mathbb{E}[u_{ij}|x_j] = 0$. Therefore

$$\mathbb{V}\left[(9)\right] = \frac{1}{n^4}\sum_{i,j=1,i\neq j}^{n}\sum_{i',j'=1,i\neq j}^{n}\mathbb{E}\left[u_{ij}u'_{i'j'}\right] = \frac{1}{n^4}\sum_{i,j=1,i\neq j}^{n}\mathbb{E}\left[u_{ij}u'_{ij}\right] + \mathbb{E}\left[u_{ij}u'_{ji}\right],$$

meaning that

$$(9) = O_{\mathbb{P}}\left(\sqrt{\frac{1}{n^2 h}}\right).$$

Part (iv)

This follows immediately from Part (i)–(iii) and Theorem 1.

5.5 Proof of Theorem 4

We first split $\hat{\Sigma}_h$ into two terms,

$$\begin{aligned}
\text{(I)} &= \frac{1}{n^3}\sum_{i,j,k=1}^{n}\Upsilon_h R_j R_k' \Upsilon_h W_j W_k \Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big)\Big(\mathbb{1}(x_i \le x_k) - F(x_k)\Big)\\
\text{(II)} &= -\frac{1}{n^2}\sum_{j,k=1}^{n}\Upsilon_h R_j R_k' \Upsilon_h W_j W_k \Big(\hat{F}(x_j) - F(x_j)\Big)\Big(\hat{F}(x_k) - F(x_k)\Big),
\end{aligned}$$

where $R_i = R(x_i - \mathsf{x})$ and $W_i = K((x_i - \mathsf{x})/h)/h$.

(II) satisfies

$$|\text{(II)}| \le \sup_x |\hat{F}(x) - F(x)|^2 \frac{1}{n^2}\sum_{j,k=1}^{n}\left|\Upsilon_h R_j R_k' \Upsilon_h W_j W_k\right|.$$

It is obvious that

$$\sup_x |\hat{F}(x) - F(x)|^2 = O_{\mathbb{P}}\left(\frac{1}{n}\right).$$

As for the second part, we have

$$\begin{aligned}
\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}\left[\left|\Upsilon_h R_j R_k' \Upsilon_h W_j W_k\right|\right] =& \frac{n-1}{n}\mathbb{E}\left[\left|\Upsilon_h R_j R_k' \Upsilon_h W_j W_k\right| \,\middle|\, j \neq k\right] + \frac{1}{n}\mathbb{E}\left[\left|\Upsilon_h R_k R_k' \Upsilon_h W_k W_k\right|\right]\\
=& O_{\mathbb{P}}\left(1 + \frac{1}{nh}\right) = O_{\mathbb{P}}\left(1\right),
\end{aligned}$$

implying that

$$\text{(II)} = O_{\mathbb{P}}\left(\frac{1}{n}\right),$$

provided that $nh \to \infty$.

For (I), we further expand into "diagonal" and "non-diagonal" sums:

$$
\begin{aligned}
\text{(I)} = {} & \frac{1}{n^3} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \Upsilon_h R_j R_k' \Upsilon_h W_j W_k \Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big)\Big(\mathbb{1}(x_i \le x_k) - F(x_k)\Big) && \text{(I.1)} \\
& + \frac{1}{n^3} \sum_{\substack{i,k=1 \\ \text{distinct}}}^{n} \Upsilon_h R_i R_k' \Upsilon_h W_i W_k \Big(\mathbb{1}(x_i \le x_i) - F(x_i)\Big)\Big(\mathbb{1}(x_i \le x_k) - F(x_k)\Big) && \text{(I.2)} \\
& + \frac{1}{n^3} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} \Upsilon_h R_j R_i' \Upsilon_h W_j W_i \Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big)\Big(\mathbb{1}(x_i \le x_i) - F(x_i)\Big) && \text{(I.3)} \\
& + \frac{1}{n^3} \sum_{\substack{i,j=1 \\ \text{distinct}}}^{n} \Upsilon_h R_j R_j' \Upsilon_h W_j W_j \Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big)\Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big) && \text{(I.4)} \\
& + \frac{1}{n^3} \sum_{i} \Upsilon_h R_i R_i' \Upsilon_h W_i W_i \Big(\mathbb{1}(x_i \le x_i) - F(x_i)\Big)\Big(\mathbb{1}(x_i \le x_i) - F(x_i)\Big). && \text{(I.5)}
\end{aligned}
$$

By calculating the expectation of the absolute value of the summands above, it is straightforward to show

$$
\text{(I.2)} = O_{\mathbb{P}}\left(\frac{1}{n}\right), \quad \text{(I.3)} = O_{\mathbb{P}}\left(\frac{1}{n}\right), \quad \text{(I.4)} = O_{\mathbb{P}}\left(\frac{1}{nh}\right), \quad \text{(I.5)} = O_{\mathbb{P}}\left(\frac{1}{n^2 h}\right).
$$

Therefore, we have

$$
\begin{aligned}
\hat{\Sigma}_h &= \text{(I.1)} + O_{\mathbb{P}}\left(\frac{1}{nh}\right) \\
&= \frac{1}{n^3} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \Upsilon_h R_j R_k' \Upsilon_h W_j W_k \Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big)\Big(\mathbb{1}(x_i \le x_k) - F(x_k)\Big) + O_{\mathbb{P}}\left(\frac{1}{nh}\right).
\end{aligned}
$$

To proceed, define

$$
u_{ij} = \Upsilon_h R_j W_j \Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big) \quad \text{and} \quad \bar{u}_i = \mathbb{E}[u_{ij} | x_i; i \ne j].
$$

Then we can further decompose (I.1) into

$$
\begin{aligned}
\text{(I.1)} = \frac{1}{n^3} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} u_{ij} u_{ik}' &= \frac{1}{n^3} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \mathbb{E}[u_{ij} u_{ik}' | x_i] \quad + \frac{1}{n^3} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \Big(u_{ij} u_{ik}' - \mathbb{E}[u_{ij} u_{ik}' | x_i]\Big) \\
&= \frac{(n-1)(n-2)}{n^3} \sum_{i=1}^{n} \bar{u}_i \bar{u}_i' && \text{(I.1.1)} \\
&+ \frac{1}{n^3} \sum_{\substack{i,j,k=1 \\ \text{distinct}}}^{n} \Big(u_{ij} u_{ik}' - \bar{u}_i \bar{u}_i'\Big). && \text{(I.1.2)}
\end{aligned}
$$

We have already analyzed (I.1.1) in Theorem 2, which satisfies

$$
\text{(I.1.1)} = \Sigma_h + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).
$$

Now we study (I.1.2), which satisfies

$$(\text{I.1.2}) = \frac{n-2}{n^3}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\Big(u_{ij}-\bar{u}_i\Big)\bar{u}_i' \tag{I.1.2.1}$$

$$+\frac{n-2}{n^3}\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\bar{u}_i\Big(u_{ij}-\bar{u}_i\Big)' \tag{I.1.2.2}$$

$$+\frac{1}{n^3}\sum_{\substack{i,j,k=1\\ \text{distinct}}}^{n}\Big(u_{ij}-\bar{u}_i\Big)\Big(u_{ik}-\bar{u}_i\Big)'. \tag{I.1.2.3}$$

By variance calculation, it is easy to see that

$$(\text{I.1.2.3}) = O_{\mathbb{P}}\left(\frac{1}{nh}\right).$$

Therefore we have

$$\frac{c'(\hat{\Sigma}_h-\Sigma_h)c}{c'\Sigma_h c} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^2}}\right) + 2\frac{c'(\text{I.1.2.1})c}{c'\Sigma_h c},$$

since (I.1.2.1) and (I.1.2.2) are transpose of each other. To close the proof, we calculate the variance of the last term in the above.

$$\begin{aligned}
\mathbb{V}\left[\frac{c'(\text{I.1.2.1})c}{c'\Sigma_h c}\right] &= \frac{1}{(c'\Sigma_h c)^2}\frac{(n-2)^2}{n^6}\mathbb{E}\left[\sum_{\substack{i,j=1\\ \text{distinct}}}^{n}\sum_{\substack{i',j'=1\\ \text{distinct}}}^{n} c'\Big(u_{ij}-\bar{u}_i\Big)\bar{u}_i'cc'\Big(u_{i'j'}-\bar{u}_{i'}\Big)\bar{u}_{i'}'c\right]\\
&= \frac{1}{(c'\Sigma_h c)^2}\frac{(n-2)^2}{n^6}\mathbb{E}\left[\sum_{\substack{i,j,i'=1\\ \text{distinct}}}^{n} c'u_{ij}\bar{u}_i'cc'u_{i'j}\bar{u}_{i'}'c\right] + \text{higher order terms.}
\end{aligned}$$

The expectation is further given by (note that i, j and i' are assumed to be distinct indices)

$$\begin{aligned}
&\mathbb{E}\left[c'u_{ij}\bar{u}_i'cc'u_{i'j}\bar{u}_{i'}'c\right]\\
&= \mathbb{E}\iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} W_j^2\left[c'\Upsilon_h R_j R(u)cc'\Upsilon_h R_j R(v)c\right]K(u)K(v)\\
&\qquad [F(x_j\wedge(\mathsf{x}+hu))-F(x_j)F(\mathsf{x}+hu)]\,[F(x_j\wedge(\mathsf{x}+hv))-F(x_j)F(\mathsf{x}+hv)]\,f(\mathsf{x}+hu)f(\mathsf{x}+hv)\mathrm{d}u\mathrm{d}v\\
&= \frac{1}{h}\iiint_{\frac{\mathcal{X}-\mathsf{x}}{h}}\left[c'R(w)R(u)cc'R(w)R(v)c\right]K(u)K(v)K(w)^2\\
&\qquad [F(\mathsf{x}+h(w\wedge u))-F(\mathsf{x}+hw)F(\mathsf{x}+hu)]\,[F(\mathsf{x}+h(w\wedge v))-F(\mathsf{x}+hw)F(\mathsf{x}+hv)]\,f(\mathsf{x}+hw)f(\mathsf{x}+hu)f(\mathsf{x}+hv)\mathrm{d}w\mathrm{d}u\mathrm{d}v\\
&= \frac{1}{h}F(\mathsf{x})^2(1-F(\mathsf{x}))^2\iiint_{\frac{\mathcal{X}-\mathsf{x}}{h}}\left[c'R(w)R(u)cc'R(w)R(v)c\right]K(u)K(v)K(w)^2 f_w f_u f_v\mathrm{d}w\mathrm{d}u\mathrm{d}v + \text{higher-order terms.}
\end{aligned}$$

If $c'\Sigma_h c = O(1)$, then the above will have order h, which means

$$\mathbb{V}\left[\frac{c'(\text{I.1.2.1})c}{c'\Sigma_h c}\right] = O\left(\frac{1}{nh}\right).$$

If $c'\Sigma_h c = O(h)$, however, $\mathbb{E}\left[c'u_{ij}\bar{u}_i'cc'u_{i'j}\bar{u}_{i'}'c\right]$ will be $O(1)$, which will imply that

$$\mathbb{V}\left[\frac{c'(\text{I.1.2.1})c}{c'\Sigma_h c}\right] = O\left(\frac{1}{nh^2}\right).$$

As a result, we have

$$\frac{c'(\hat{\Sigma}_h - \Sigma_h)c}{c'\Sigma_h c} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^2}}\right).$$

Now consider

$$\begin{aligned}\frac{c'\hat{\Gamma}_h^{-1}\hat{\Sigma}_h\hat{\Gamma}_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} - 1 &= \frac{c'\hat{\Gamma}_h^{-1}(\hat{\Sigma}_h - \Sigma_h)\hat{\Gamma}_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} + \frac{c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h\hat{\Gamma}_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} + \frac{c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h\Gamma_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c}\\ &= \frac{c'\hat{\Gamma}_h^{-1}(\hat{\Sigma}_h - \Sigma_h)\hat{\Gamma}_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} + 2\frac{c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h\Gamma_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} + \frac{c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c}.\end{aligned}$$

From the analysis of $\hat{\Sigma}_h$, we have

$$\frac{c'\hat{\Gamma}_h^{-1}(\hat{\Sigma}_h - \Sigma_h)\hat{\Gamma}_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^2}}\right).$$

For the second term, we have

$$\begin{aligned}\left|\frac{c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h\hat{\Gamma}_h^{-1}c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c}\right| &\leq \frac{|c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h^{1/2}| \cdot |c'\Gamma_h^{-1}\Sigma_h^{1/2}|}{|c'\Gamma_h^{-1}\Sigma_h^{1/2}|^2}\\ &= \frac{|c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h^{1/2}|}{|c'\Gamma_h^{-1}\Sigma_h^{1/2}|} = O_{\mathbb{P}}\left(\sqrt{\frac{1}{nh^2}}\right).\end{aligned}$$

The third term has order

$$\frac{c'(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})\Sigma_h(\hat{\Gamma}_h^{-1} - \Gamma_h^{-1})c}{c'\Gamma_h^{-1}\Sigma_h\Gamma_h^{-1}c} = O_{\mathbb{P}}\left(\frac{1}{nh^2}\right).$$

5.6 Proof of Corollary 5

This follows directly from Theorem 1.

5.7 Proof of Corollary 6

To understand (13), note that

$$\begin{aligned}\hat{\theta}_F^{\perp} &= \left(\int_{\mathcal{X}} \Lambda_h' R(u - \mathsf{x})R(u - \mathsf{x})'\Lambda_h \frac{1}{h}K\left(\frac{u - \mathsf{x}}{h}\right)\mathrm{d}F(u)\right)^{-1}\left(\int_{\mathcal{X}} \Lambda_h' R(u - \mathsf{x})\hat{F}(u)\frac{1}{h}K\left(\frac{u - \mathsf{x}}{h}\right)\mathrm{d}F(u)\right)\\ &= \Lambda_h^{-1}\left(\int_{\mathcal{X}} R(u - \mathsf{x})R(u - \mathsf{x})'\frac{1}{h}K\left(\frac{u - \mathsf{x}}{h}\right)\mathrm{d}F(u)\right)^{-1}\left(\int_{\mathcal{X}} R(u - \mathsf{x})\hat{F}(u)\frac{1}{h}K\left(\frac{u - \mathsf{x}}{h}\right)\mathrm{d}F(u)\right),\end{aligned}$$

which means $\hat{\theta}_F^{\perp} = \Lambda_h^{-1}\hat{\theta}_F$. Then we have (up to an approximation bias term)

$$\begin{aligned}&\hat{\theta}_F^{\perp} - \Lambda_h^{-1}\theta_0 = \Lambda_h^{-1}(\hat{\theta}_F - \theta_0)\\ &= \Lambda_h^{-1}\left(\int_{\mathcal{X}} R(u - \mathsf{x})R(u - \mathsf{x})'\frac{1}{h}K\left(\frac{u - \mathsf{x}}{h}\right)\mathrm{d}F(u)\right)^{-1}\left(\int_{\mathcal{X}} R(u - \mathsf{x})\Big(\hat{F}(u) - F(u)\Big)\frac{1}{h}K\left(\frac{u - \mathsf{x}}{h}\right)\mathrm{d}F(u)\right)\\ &= \Lambda_h^{-1}\Upsilon_h\left(\int_{\frac{\mathcal{X} - \mathsf{x}}{h}} R(u)R(u)'K(u)f(\mathsf{x} + hu)\mathrm{d}u\right)^{-1}\left(\int_{\frac{\mathcal{X} - \mathsf{x}}{h}} R(u)\Big(\hat{F}(\mathsf{x} + hu) - F(\mathsf{x} + hu)\Big)K(u)f(\mathsf{x} + hu)\mathrm{d}u\right)\\ &= \Lambda_h^{-1}\Upsilon_h\Lambda_h\left(\int_{\frac{\mathcal{X} - \mathsf{x}}{h}} R^{\perp}(u)R^{\perp}(u)'K(u)f(\mathsf{x} + hu)\mathrm{d}u\right)^{-1}\left(\int_{\frac{\mathcal{X} - \mathsf{x}}{h}} R^{\perp}(u)\Big(\hat{F}(\mathsf{x} + hu) - F(\mathsf{x} + hu)\Big)K(u)f(\mathsf{x} + hu)\mathrm{d}u\right).\end{aligned}$$

As a result, we have the following

We first discuss the transformed parameter vector $\Lambda_h^{-1}\theta_0$. By construction, the matrix Λ_h takes the following form:

$$\Lambda_h = \begin{bmatrix} 1 & c_{1,2} & c_{1,3} & \cdots & c_{1,p+2} \\ 0 & 1 & 0 & \cdots & c_{2,p+2} \\ 0 & 0 & 1 & \cdots & c_{2,p+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where $c_{i,j}$ are some constants (possibly depending on h). Therefore, the above matrix differs from the identity matrix only in its first row and last column. This observation also holds for Λ_h^{-1}. Since the last component of θ_0 is zero (because the extra regressor $Q_h(\cdot)$ is redundant), we conclude that, except for the first element, $\Lambda_h\theta$ and θ are identical. More specifically, let I_{-1} be the identity matrix excluding the first row:

$$I_{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

which is used to extract all elements of a vector except for the first one, then by Theorem 1,

$$\sqrt{n}\left(I_{-1}(\Lambda_h^{-1}\Upsilon_h\Lambda_h)(\Gamma_h^{\perp})^{-1}\Sigma_h^{\perp}(\Gamma_h^{\perp})^{-1}(\Lambda_h^{-1}\Upsilon_h\Lambda_h)'I_{-1}'\right)^{-1/2}\begin{bmatrix} \hat{\theta}_{P,F}^{\perp} - \theta_P \\ \hat{\theta}_{Q,F}^{\perp} \end{bmatrix} \rightsquigarrow \mathcal{N}(0, I),$$

where $\theta_{P,F}^{\perp}$ contains the second to the $p+1$-th element of $\theta_F^{\perp}$, and $\theta_{Q,F}^{\perp}$ is the last element.

Now we discuss the covariance matrix in the above display. Due to orthogonalization, $\Gamma_h^{\perp}$ is block diagonal. To be precise,

$$\Gamma_h^{\perp} = f(\mathsf{x})\begin{bmatrix} \Gamma_{1,h}^{\perp} & 0 & 0 \\ & \Gamma_{P,h}^{\perp} & 0 \\ 0 & 0 & \Gamma_{Q,h}^{\perp} \end{bmatrix},\ \Gamma_{1,h}^{\perp} = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)\mathrm{d}u,\ \Gamma_{P,h}^{\perp} = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} P^{\perp}(u)P^{\perp}(u)'K(u)\mathrm{d}u,\ \Gamma_{Q,h}^{\perp} = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} Q^{\perp}(u)^2K(u)\mathrm{d}u.$$

Finally, using the structure of Λ_h and Υ_h, we have

$$I_{-1}(\Lambda_h^{-1}\Upsilon_h\Lambda_h)(\Gamma_h^{\perp})^{-1} = I_{-1}\Upsilon_h(\Gamma_h^{\perp})^{-1}.$$

The form of $\Sigma_h^{\perp}$ is quite involved, but with some algebra, and using the fact that the basis $R(\cdot)$ (or $R^{\perp}(\cdot)$) includes a constant and polynomials, one can show the following:

$$(\Lambda_h^{-1}\Upsilon_h\Lambda_h)(\Gamma_h^{\perp})^{-1}\Sigma_h^{\perp}(\Gamma_h^{\perp})^{-1}(\Lambda_h^{-1}\Upsilon_h\Lambda_h)' = hf(\mathsf{x})\Upsilon_{-1,h}(\Gamma_{-1,h}^{\perp})^{-1}\Sigma_{-1,h}^{\perp}(\Gamma_{-1,h}^{\perp})^{-1}\Upsilon_{-1,h},$$

where $\Upsilon_{-1,h}$, $\Gamma_{-1,h}^{\perp}$ and $\Sigma_{-1,h}^{\perp}$ are obtained by excluding the first row and the first column of Υ_h, $\Gamma_h^{\perp}$ and $\Sigma_h^{\perp}$, respectively:

$$\Upsilon_{-1,h} = \begin{bmatrix} h^{-1} & 0 & 0 & \cdots & 0 \\ 0 & h^{-2} & 0 & \cdots & 0 \\ 0 & 0 & h^{-3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \upsilon_h \end{bmatrix},\ \Gamma_{-1,h}^{\perp} = f(\mathsf{x})\begin{bmatrix} \Gamma_{P,h}^{\perp} & 0 \\ 0 & \Gamma_{Q,h}^{\perp} \end{bmatrix},\ \Sigma_{-1,h}^{\perp} = f(\mathsf{x})^3\begin{bmatrix} \Sigma_{PP,h}^{\perp} & \Sigma_{PQ,h}^{\perp} \\ \Sigma_{QP,h}^{\perp} & \Sigma_{QQ,h}^{\perp} \end{bmatrix},$$

and

$$\Sigma_{PP,h}^{\perp} = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)P^{\perp}(u)P^{\perp}(v)'(u \wedge v)\mathrm{d}u\mathrm{d}v, \qquad \Sigma_{QQ,h}^{\perp} = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)Q^{\perp}(u)Q^{\perp}(v)(u \wedge v)\mathrm{d}u\mathrm{d}v$$

$$\Sigma_{PQ,h}^{\perp} = (\Sigma_{QP,h}^{\perp})' = \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)K(v)P^{\perp}(u)Q^{\perp}(v)(u \wedge v)\mathrm{d}u\mathrm{d}v.$$

With the above discussion, we have the desired result.

5.8 Proof of Lemma 7

Part (i)

To start,

$$\begin{aligned}
\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(g_1)(u)\mathcal{H}(g_2)(u)\mathrm{d}u &= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \mathbb{1}(v_1 \geq u)K(v_1)g(v_1)\mathrm{d}v_1\right)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \mathbb{1}(v_2 \geq u)K(v_2)g(v_2)\mathrm{d}v_2\right)\mathrm{d}u \\
&= \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v_1)K(v_2)g(v_1)g(v_2)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathbb{1}(v_1 \geq u)\mathbb{1}(v_2 \geq u)\mathrm{d}u\right)\mathrm{d}v_1\mathrm{d}v_2 \\
&= \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v_1)K(v_2)g(v_1)g(v_2)\left[(v_1 \wedge v_2) \wedge \left(\frac{\overline{x}-\mathsf{x}}{h} \wedge 1\right) - \left(\frac{\underline{x}-\mathsf{x}}{h} \vee (-1)\right)\right]\mathrm{d}v_1\mathrm{d}v_2 \\
&= \iint_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v_1)K(v_2)g(v_1)g(v_2)(v_1 \wedge v_2)\mathrm{d}v_1\mathrm{d}v_2,
\end{aligned}$$

where for the last equality, we used the fact that $v_1 \leq \frac{\overline{x}-\mathsf{x}}{h} \wedge 1$ and $v_2 \leq \frac{\overline{x}-\mathsf{x}}{h} \wedge 1$ for the double integral.

Part (ii)

For this part,

$$\begin{aligned}
\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(g_1)(u)\dot{g}_2(u)\mathrm{d}u &= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \mathbb{1}(v \geq u)K(v)g_1(v)\mathrm{d}v\right)\dot{g}_2(u)\mathrm{d}u \\
&= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v)g_1(v)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathbb{1}(v \geq u)\dot{g}_2(u)\mathrm{d}u\right)\mathrm{d}v \\
&= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v)g_1(v)\left[g_2\left(v \wedge \frac{\overline{x}-\mathsf{x}}{h} \wedge 1\right) - g_2\left(\frac{\underline{x}-\mathsf{x}}{h} \vee (-1)\right)\right]\mathrm{d}v \\
&= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(v)g_1(v)g_2(v)\mathrm{d}v.
\end{aligned}$$

5.9 Proof of Theorem 8

To find a bound of the maximization problem, we note that for any $c \in \mathbb{R}^{p-1}$, one has

$$\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(p_\ell)(u)\mathcal{H}(q)(u)\mathrm{d}u = \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \left[\mathcal{H}(p_\ell)(u) + c'\dot{P}(u)\right]\mathcal{H}(q)(u)\mathrm{d}u,$$

due to the constraint (15). Hence subject to this constraint, an upper bound of the objective function is (due to the Cauchy-Schwarz inequality)

$$
\begin{aligned}
&\inf_{c} \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \Big[\mathcal{H}(p_\ell)(u) + c'\dot{P}(u)\Big]^2 \mathrm{d}u \\
&= \inf_{c} \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \Big[\mathcal{H}(p_\ell)(u)^2 + 2c'\dot{P}(u)\mathcal{H}(p_\ell)(u) + c'\dot{P}(u)\dot{P}(u)'c\Big]\mathrm{d}u \\
&= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(p_\ell)(u)^2\mathrm{d}u + \inf_{c} \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \Big[2c'\dot{P}(u)\mathcal{H}(p_\ell)(u) + c'\dot{P}(u)\dot{P}(u)'c\Big]\mathrm{d}u \\
&= \int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(p_\ell)(u)^2\mathrm{d}u + \inf_{c} \left[2c'\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)p_\ell(u)\mathrm{d}u\right) + c'\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)c\right],
\end{aligned}
$$

which is minimized by setting

$$
c = -\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1}\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)p_\ell(u)\mathrm{d}u\right).
$$

As a result, an upper bound of (14) is

$$
\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(p_\ell)(u)^2\mathrm{d}u - \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)p_\ell(u)\mathrm{d}u\right)'\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1}\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)p_\ell(u)\mathrm{d}u\right).
$$

We may further simplify the above. First,

$$
\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \mathcal{H}(p_\ell)(u)^2\mathrm{d}u = e_\ell'(\Gamma^{\perp}_{P,h})^{-1}\Sigma^{\perp}_{PP,h}(\Gamma^{\perp}_{P,h})^{-1}e_\ell.
$$

Second, note that

$$
\begin{aligned}
\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)p_\ell(u)\mathrm{d}u &= \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P(u)P^{\perp}(u)'\mathrm{d}u\right)(\Gamma^{\perp}_{P,h})^{-1}e_\ell \\
&= \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} K(u)P^{\perp}(u)P^{\perp}(u)'\mathrm{d}u\right)(\Gamma^{\perp}_{P,h})^{-1}e_\ell \\
&= e_\ell.
\end{aligned}
$$

As a result, an upper bound of (14) is

$$
\begin{aligned}
&e_\ell'(\Gamma^{\perp}_{P,h})^{-1}\Sigma^{\perp}_{PP,h}(\Gamma^{\perp}_{P,h})^{-1}e_\ell - e_\ell'\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1}e_\ell \\
&= e_\ell'\left[(\Gamma^{\perp}_{P,h})^{-1}\Sigma^{\perp}_{PP,h}(\Gamma^{\perp}_{P,h})^{-1} - \left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}\cap[-1,1]} \dot{P}(u)\dot{P}(u)'\mathrm{d}u\right)^{-1}\right]e_\ell.
\end{aligned}
$$

5.10 Proof of Theorem 9

To show that the two processes, $\mathfrak{T}_G(\cdot)$ and $\mathfrak{B}_G(\cdot)$ are "close in distribution," we employ the proof strategy of Giné, Koltchinskii, and Sakhanenko (2004). Recall that F denotes the distribution of x_i, and we define

$$
\mathcal{K}_{h,\mathsf{x}} \circ F^{-1}(x) = \mathcal{K}_{h,\mathsf{x}}(F^{-1}(x)).
$$

Take $v < v'$ in $[0,1]$, we have

$$
\begin{aligned}
&\left|\mathcal{K}_{h,\mathsf{x}} \circ F^{-1}(v) - \mathcal{K}_{h,\mathsf{x}} \circ F^{-1}(v')\right| \\
&= \left| \frac{\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} c_{h,\mathsf{x}}' \Upsilon_h \Gamma_{h,\mathsf{x}}^{-1} R(u) \Big[\mathbb{1}(F^{-1}(v) \leq \mathsf{x} + hu) - \mathbb{1}(F^{-1}(v') \leq \mathsf{x} + hu)\Big] K\left(u\right) g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}' \Upsilon_h \Omega_{h,\mathsf{x}} \Upsilon_h c_{h,\mathsf{x}}}} \right| \\
&\leq \frac{\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \left|c_{h,\mathsf{x}}' \Upsilon_h \Gamma_{h,\mathsf{x}}^{-1} R(u)\right| \Big[\mathbb{1}(F^{-1}(v) \leq \mathsf{x} + hu) - \mathbb{1}(F^{-1}(v') \leq \mathsf{x} + hu)\Big] K\left(u\right) g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}' \Upsilon_h \Omega_{h,\mathsf{x}} \Upsilon_h c_{h,\mathsf{x}}}}.
\end{aligned}
$$

Therefore, the function $\mathcal{K}_{h,\mathsf{x}} \circ F^{-1}(\cdot)$ has total variation bounded by

$$
\begin{aligned}
&\frac{\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} \left|c_{h,\mathsf{x}}' \Upsilon_h \Gamma_{h,\mathsf{x}}^{-1} R(u)\right| \Big[\mathbb{1}(F^{-1}(0) \leq \mathsf{x} + hu) - \mathbb{1}(F^{-1}(1) \leq \mathsf{x} + hu)\Big] K\left(u\right) g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}' \Upsilon_h \Omega_{h,\mathsf{x}} \Upsilon_h c_{h,\mathsf{x}}}} \\
&= \frac{\int_{-1}^{1} \left|c_{h,\mathsf{x}}' \Upsilon_h \Gamma_{h,\mathsf{x}}^{-1} R(u)\right| K\left(u\right) g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}' \Upsilon_h \Omega_{h,\mathsf{x}} \Upsilon_h c_{h,\mathsf{x}}}} \leq C_4 \frac{1}{\sqrt{h}}.
\end{aligned}
$$

It is well-known that functions of bounded variation can be approximated (pointwise) by convex combination of indicator functions of half intervals. To be more precise,

$$
\Big\{\mathcal{K}_{h,\mathsf{x}} \circ F^{-1}(\cdot):\ \mathsf{x} \in \mathcal{I}\Big\} \subset C_4 \frac{1}{\sqrt{h}} \overline{\mathrm{conv}} \Big\{ \pm \mathbb{1}(\cdot \leq t), \pm \mathbb{1}(\cdot \geq t)\Big\}.
$$

Following (2.3) and (2.4) of Giné, Koltchinskii, and Sakhanenko (2004), we have

$$
\mathbb{P}\left[\sup_{\mathsf{x}\in\mathcal{I}} \left|\tilde{\mathfrak{T}}_G(\mathsf{x}) - \mathfrak{B}_G(\mathsf{x})\right| > \frac{C_4(u + C_1 \log n)}{\sqrt{nh}}\right] \leq C_2 e^{-C_3 u},
$$

where C_1, C_2 and C_3 are some universal constants.

5.11 Proof of Lemma 10

Take $|\mathsf{x}-\mathsf{y}| \le \varepsilon$ to be some small number, then

$$
\begin{aligned}
&\mathcal{K}_{h,\mathsf{x}}(x) - \mathcal{K}_{h,\mathsf{y}}(x)\\
&= \frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x\le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\\
&-\frac{c_{h,\mathsf{y}}'\Upsilon_h\Gamma_{h,\mathsf{y}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{y}}{h}} R(u)\Big[\mathbb{1}(x\le \mathsf{y}+hu) - F(\mathsf{y}+hu)\Big]K(u)\,g(\mathsf{y}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\\
&= \left(\frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} - \frac{c_{h,\mathsf{y}}'\Upsilon_h\Gamma_{h,\mathsf{y}}^{-1}}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x\le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u\right)\\
&+\left(\frac{c_{h,\mathsf{y}}'\Upsilon_h\Gamma_{h,\mathsf{y}}^{-1}}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right)\left(\frac{1}{h}\int_{\mathcal{X}}\left[R\left(\frac{u-\mathsf{x}}{h}\right)K\left(\frac{u-\mathsf{x}}{h}\right) - R\left(\frac{u-\mathsf{y}}{h}\right)K\left(\frac{u-\mathsf{y}}{h}\right)\right]\Big[\mathbb{1}(x\le u) - F(u)\Big]g(u)\mathrm{d}u\right)\\
&= \frac{1}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\left(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h\right)\Gamma_{h,\mathsf{x}}^{-1}\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x\le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u\right) && \text{(I)}\\
&+ \frac{1}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}c_{h,\mathsf{y}}'\Upsilon_h\left(\Gamma_{h,\mathsf{x}}^{-1} - \Gamma_{h,\mathsf{y}}^{-1}\right)\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x\le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u\right) && \text{(II)}\\
&+\left(\frac{1}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} - \frac{1}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right)c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\left(\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x\le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\,g(\mathsf{x}+hu)\mathrm{d}u\right) && \text{(III)}\\
&+\left(\frac{c_{h,\mathsf{y}}'\Upsilon_h\Gamma_{h,\mathsf{y}}^{-1}}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right)\left(\frac{1}{h}\int_{\mathcal{X}}\left[R\left(\frac{u-\mathsf{x}}{h}\right)K\left(\frac{u-\mathsf{x}}{h}\right) - R\left(\frac{u-\mathsf{y}}{h}\right)K\left(\frac{u-\mathsf{y}}{h}\right)\right]\Big[\mathbb{1}(x\le u) - F(u)\Big]g(u)\mathrm{d}u\right). && \text{(IV)}
\end{aligned}
$$

For term (I), its variance (replace the placeholder x by x_i) is

$$\mathbb{V}[\text{(I)}] = \frac{1}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}\left(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h\right)\Omega_{h,\mathsf{x}}\left(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h\right)' = O\left(\frac{1}{h}r_1(\varepsilon,h)^2 r_2(h)^2\right).$$

Term (II) has variance

$$\mathbb{V}[\text{(II)}] = \frac{1}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}c_{h,\mathsf{y}}'\Upsilon_h\left(\Gamma_{h,\mathsf{x}}^{-1} - \Gamma_{h,\mathsf{y}}^{-1}\right)\Sigma_{h,\mathsf{x}}\left(\Gamma_{h,\mathsf{x}}^{-1} - \Gamma_{h,\mathsf{y}}^{-1}\right)'\left(c_{h,\mathsf{y}}'\Upsilon_h\right)' = O\left(\frac{1}{h}\left(\frac{\varepsilon}{h}\wedge 1\right)^2\right),$$

where the order $\frac{\varepsilon}{h}\wedge 1$ comes from the difference $\Gamma_{h,\mathsf{x}}^{-1} - \Gamma_{h,\mathsf{y}}^{-1}$.

Next for term (III), we have

$$
\begin{aligned}
\mathbb{V}[(\mathrm{III})] &= \left(\frac{1}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} - \frac{1}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right)^2 c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}} \\
&= \left(1-\sqrt{1+\frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}} - c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right)^2 \\
&\asymp \left(\frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}} - c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}\right)^2 \\
&= \left(\frac{c_{h,\mathsf{x}}'\Upsilon_h(\Omega_{h,\mathsf{x}}-\Omega_{h,\mathsf{y}})\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}} + \frac{(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{x}} + (c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}\right)^2.
\end{aligned}
$$

The first term has bound

$$
\frac{c_{h,\mathsf{x}}'\Upsilon_h(\Omega_{h,\mathsf{x}}-\Omega_{h,\mathsf{y}})\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}} = O\left(\frac{\varepsilon}{h}\right).
$$

The third term has bound

$$
\frac{(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}} \precsim \frac{|(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}^{1/2}|}{\sqrt{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}} = O\left(\frac{1}{\sqrt{h}} r_1(\varepsilon,h) r_2(h)\right).
$$

Finally, the second term can be bounded as

$$
\begin{aligned}
\frac{(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}} &= \frac{(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}} + (c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)\Omega_{h,\mathsf{y}}(c_{h,\mathsf{x}}'\Upsilon_h - c_{h,\mathsf{y}}'\Upsilon_h)'}{c_{h,\mathsf{y}}'\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}} \\
&= O\left(\frac{1}{\sqrt{h}} r_1(\varepsilon,h) r_2(h) + \frac{1}{h} r_1(\varepsilon,h)^2 r_2(h)^2\right).
\end{aligned}
$$

Overall, we have that

$$
\mathbb{V}[(\mathrm{III})] = O\left(\frac{\varepsilon^2}{h^2} + \frac{1}{h} r_1(\varepsilon,h)^2 r_2(h)^2 + \frac{1}{h^2} r_1(\varepsilon,h)^4 r_2(h)^4\right).
$$

Given our assumptions on the basis function and on the kernel function, it is obvious that term (IV) has variance

$$
\mathbb{V}[(\mathrm{IV})] = O\left(\frac{1}{h}\left(\frac{\varepsilon}{h}\wedge 1\right)^2\right).
$$

The bound on $\mathbb{E}[\sup_{\mathsf{x}\in\mathcal{I}}|\mathfrak{B}_G(\mathsf{x})|]$ can be found by standard entropy calculation, and the bound on $\mathbb{E}[\sup_{\mathsf{x}\in\mathcal{I}}|\mathfrak{T}_G(\mathsf{x})|]$ is obtained by the following fact

$$
\mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\mathfrak{T}_G(\mathsf{x})|\right] \le \mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\mathfrak{B}_G(\mathsf{x})|\right] + \mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\tilde{\mathfrak{T}}_G(\mathsf{x}) - B_G(\mathsf{x})|\right],
$$

and that

$$
\mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\tilde{\mathfrak{T}}_G(\mathsf{x}) - \mathfrak{B}_G(\mathsf{x})|\right] = \int_0^\infty \mathbb{P}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\tilde{\mathfrak{T}}_G(\mathsf{x}) - \mathfrak{B}_G(\mathsf{x})| > u\right]\mathrm{d}u = O\left(\frac{\log n}{\sqrt{nh}}\right) = o(\sqrt{\log n}),
$$

which follows from Theorem 9 and our assumption that $\log n/(nh) \to 0$.

5.12 Proof of Lemma 11

We adopt the following decomposition (the integration is always on $\frac{\mathcal{X}-\mathsf{y}}{h} \times \frac{\mathcal{X}-\mathsf{x}}{h}$, unless otherwise specified):

$$\frac{1}{n}\sum_{i=1}^{n}\iint R(u)R(v)'\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]\Big[\mathbb{1}(x_i \le \mathsf{y}+hv) - F(\mathsf{y}+hv)\Big]K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{y}+hv)\mathrm{d}u\mathrm{d}v \tag{I}$$

$$-\iint R(u)R(v)'\Big[\hat{F}(\mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]\Big[\hat{F}(\mathsf{y}+hv) - F(\mathsf{y}+hv)\Big]K(u)K(v)g(\mathsf{x}+hu)g(\mathsf{y}+hv)\mathrm{d}u\mathrm{d}v. \tag{II}$$

By the uniform convergence of the empirical distribution function, we have that

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}} |(\mathrm{II})| = O_{\mathbb{P}}\left(\frac{1}{n}\right).$$

From the definition of $\Sigma_{h,\mathsf{x},\mathsf{y}}$, we know that

$$\mathbb{E}[(\mathrm{I})] = \Sigma_{h,\mathsf{x},\mathsf{y}}.$$

As (I) is a sum of bounded terms, we can apply Lemma 24 and easily show that

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}} |(\mathrm{I}) - \Sigma_{h,\mathsf{x},\mathsf{y}}| + O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right).$$

5.13 Proof of Lemma 12

We rewrite (17) as

$$\begin{aligned}
|(17)| &= \left|\sqrt{n}\sqrt{\frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_h^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big]K(u)\, g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right| \\
&\le \sqrt{\frac{n}{h}}\left[\sup_{\mathsf{x}\in\mathcal{I}}\sqrt{\frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right]\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[F(\mathsf{x}+hu) - \theta' R(u)\Upsilon_h^{-1}\Big]K(u)\, g(\mathsf{x}+hu)\mathrm{d}u\right|\right] \\
&= O_{\mathbb{P}}\left(\sqrt{\frac{n}{h}}\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x})\right),
\end{aligned}$$

where the final bound holds uniformly for $\mathsf{x}\in\mathcal{I}$.

5.14 Proof of Lemma 13

To start, we expand term (18) as

$$\begin{aligned}
(18) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\, g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[1 - \sqrt{\frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right]\frac{c_{h,\mathsf{x}}'\Upsilon_h\Gamma_{h,\mathsf{x}}^{-1}\int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i \le \mathsf{x}+hu) - F(\mathsf{x}+hu)\Big]K(u)\, g(\mathsf{x}+hu)\mathrm{d}u}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}} \\
&= \mathfrak{T}_G(\mathsf{x}) \\
&\quad + \left[1 - \sqrt{\frac{c_{h,\mathsf{x}}'\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right]\mathfrak{T}_G(\mathsf{x}). \qquad\qquad \text{(I)}
\end{aligned}$$

Term (I) can be easily bounded by

$$\sup_{\mathsf{x}\in\mathcal{I}}|(\mathrm{I})| = O_{\mathbb{P}}\left(\left(\sqrt{\frac{\log n}{nh^2}}\right)\mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\mathfrak{T}_G(\mathsf{x})|\right]\right) = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{nh^2}}\right).$$

5.15 Proof of Theorem 14

The claim follows from Theorem 9 and previous lemmas.

5.16 Proof of Theorem 15

Let $\mathcal{I}_\varepsilon$ be an ε-covering (with respect to the Euclidean metric) of $\mathcal{I}$, and assume $\varepsilon \le h$. Then the process $\tilde{\mathfrak{B}}_G(\cdot)$ can be decomposed into:

$$\tilde{\mathfrak{B}}_G(\mathsf{x}) = \tilde{\mathfrak{B}}_G(\Pi_{\mathcal{I}_\varepsilon}(\mathsf{x})) + \tilde{\mathfrak{B}}_G(\mathsf{x}) - \tilde{\mathfrak{B}}_G(\Pi_{\mathcal{I}_\varepsilon}(\mathsf{x})),$$

where $\Pi_{\mathcal{I}_\varepsilon}: \mathcal{I} \to \mathcal{I}_\varepsilon$ is a mapping satisfying:

$$\Pi_{\mathcal{I}_\varepsilon}(\mathsf{x}) = \operatorname*{argmin}_{\mathsf{y}\in\mathcal{I}_\varepsilon}|\mathsf{y}-\mathsf{x}|.$$

We first study the properties of $\tilde{\mathfrak{B}}_G(\mathsf{x}) - \tilde{\mathfrak{B}}_G(\Pi_{\mathcal{I}_\varepsilon}(\mathsf{x}))$. With standard entropy calculation, one has:

$$\begin{aligned}
\mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\tilde{\mathfrak{B}}_G(\mathsf{x}) - \tilde{\mathfrak{B}}_G(\Pi_{\mathcal{I}_\varepsilon}(\mathsf{x}))|\right] &\le \mathbb{E}\left[\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I},|\mathsf{x}-\mathsf{y}|\le\varepsilon}|\tilde{\mathfrak{B}}_G(\mathsf{x}) - \tilde{\mathfrak{B}}_G(\mathsf{y})|\right] \\
&\le \mathbb{E}\left[\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I},\sigma(\mathsf{x},\mathsf{y})\le\delta(\varepsilon)}|\tilde{\mathfrak{B}}_G(\mathsf{x}) - \tilde{\mathfrak{B}}_G(\mathsf{y})|\right] \\
&\precsim \int_0^{\delta(\varepsilon)}\sqrt{\log N(\lambda,\mathcal{I},\sigma_G)}\mathrm{d}\lambda,
\end{aligned}$$

where

$$\delta(\varepsilon) = C\left(\frac{1}{\sqrt{h}}\frac{\varepsilon}{h} + \frac{1}{\sqrt{h}}r_1(\varepsilon,h)r_2(h) + \frac{1}{h}r_1(\varepsilon,h)^2 r_2(h)^2\right),$$

for some $C > 0$ that does not depend on ε and h, and $N(\lambda,\mathcal{I},\sigma_G)$ is the covering number of $\mathcal{I}$ measured by the pseudo metric $\sigma_G(\cdot,\cdot)$, which satisfies

$$N(\lambda,\mathcal{I},\sigma_G) \precsim \frac{1}{\delta^{-1}(\lambda)}.$$

Therefore, we have

$$\mathbb{E}\left[\sup_{\mathsf{x}\in\mathcal{I}}|\tilde{\mathfrak{B}}_G(\mathsf{x}) - \tilde{\mathfrak{B}}_G(\Pi_{\mathcal{I}_\varepsilon}(\mathsf{x}))|\right] \precsim \left(\frac{1}{\sqrt{h}}\frac{\varepsilon}{h} + \frac{1}{\sqrt{h}}r_1(\varepsilon,h)r_2(h) + \frac{1}{h}r_1(\varepsilon,h)^2 r_2(h)^2\right)\sqrt{\log n}.$$

Now consider the discretized version of $\tilde{\mathfrak{B}}_G(\mathsf{x})$. To be precise, we denote $\mathcal{I}_\varepsilon$ by $\{\mathsf{x}_1,\mathsf{x}_2,\cdots,\mathsf{x}_{M(\varepsilon)}\}$, where $M(\varepsilon)$ depends on ε, and we have $M(\varepsilon) = O(\varepsilon^{-1})$. Also let $\Xi_{\mathcal{I}_\varepsilon}$ be the covariance matrix of $(\tilde{\mathfrak{B}}_G(\mathsf{x}_1),\tilde{\mathfrak{B}}_G(\mathsf{x}_2),\cdots,\tilde{\mathfrak{B}}_G(\mathsf{x}_{M(\varepsilon)}))'$, then the following representation holds:

$$(\tilde{\mathfrak{B}}_G(\mathsf{x}_1),\tilde{\mathfrak{B}}_G(\mathsf{x}_2),\cdots,\tilde{\mathfrak{B}}_G(\mathsf{x}_{M(\varepsilon)}))' = \Xi_{\mathcal{I}_\varepsilon}^{1/2}\mathfrak{N}_{M(\varepsilon)},$$

and $\mathfrak{N}_{M(\varepsilon)}$ is a standard Gaussian random vector (i.e., with zero mean, unit variance and zero correlation) of length $M(\varepsilon)$. In addition, $\mathfrak{N}_{M(\varepsilon)}$ is independent of the data by construction.

Similarly, we can define the Gaussian vector

$$(\hat{\mathfrak{B}}_G(\mathsf{x}_1), \hat{\mathfrak{B}}_G(\mathsf{x}_2), \cdots, \hat{\mathfrak{B}}_G(\mathsf{x}_{M(\varepsilon)}))' = \hat{\Xi}^{1/2}_{\mathcal{I}_\varepsilon} \mathfrak{N}_{M(\varepsilon)},$$

where $\hat{\Xi}^{1/2}_{\mathcal{I}_\varepsilon}$ is the estimated covariance matrix. From Lemma 11, we have

$$\left\| \hat{\Xi}_{\mathcal{I}_\varepsilon} - \Xi_{\mathcal{I}_\varepsilon} \right\|_\infty = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^2}} \right),$$

where $\|\cdot\|_\infty$ denote supremum norm (i.e., the maximum discrepancy between entries of $\Xi_{\mathcal{I}_\varepsilon}$ and $\hat{\Xi}_{\mathcal{I}_\varepsilon}$). Also we note that, in any row/column of $\hat{\Xi}_{\mathcal{I}_\varepsilon}$ or $\Xi_{\mathcal{I}_\varepsilon}$, at most $O(hM(\varepsilon)) = O(h/\varepsilon)$ entries are nonzero. Therefore we have

$$\left\| \hat{\Xi}_{\mathcal{I}_\varepsilon} - \Xi_{\mathcal{I}_\varepsilon} \right\|_{\mathrm{op}} = O_{\mathbb{P}}\left(\frac{h}{\varepsilon}\sqrt{\frac{\log n}{nh^2}} \right) = O_{\mathbb{P}}\left(\frac{1}{\varepsilon}\sqrt{\frac{\log n}{n}} \right),$$

with $\|\cdot\|_{\mathrm{op}}$ being the operator norm. We also have

$$\left\| \hat{\Xi}^{1/2}_{\mathcal{I}_\varepsilon} - \Xi^{1/2}_{\mathcal{I}_\varepsilon} \right\|^2_{\mathrm{op}} \le \left\| \hat{\Xi}_{\mathcal{I}_\varepsilon} - \Xi_{\mathcal{I}_\varepsilon} \right\|_{\mathrm{op}}.$$

Now we are ready to bound the difference of the two discretized processes:

$$\mathbb{E}^\star\left[\sup_{\mathsf{x}\in\mathcal{I}_\varepsilon} |\tilde{\mathfrak{B}}_G(\mathsf{x}) - \hat{\mathfrak{B}}_G(\mathsf{x})| \right] = O\left(\sqrt{\log\frac{1}{\varepsilon}} \right) \sqrt{\sup_{\mathsf{x}\in\mathcal{I}_\varepsilon} \mathbb{V}^\star\left[\tilde{\mathfrak{B}}_G(\mathsf{x}) - \hat{\mathfrak{B}}_G(\mathsf{x}) \right]} = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{\varepsilon}\sqrt{\frac{\log n}{n}}} \right).$$

5.17 Proof of Lemma 16

We apply Lemma 24. For simplicity, assume $R(\cdot)$ is scalar, and let

$$u_{i,h}(\mathsf{x}) = R\left(\frac{x_i - \mathsf{x}}{h}\right)^2 \frac{1}{h} K\left(\frac{x_i - \mathsf{x}}{h}\right) - \Gamma_{h,\mathsf{x}}.$$

Then it is easy to see that

$$\sup_{\mathsf{x}\in\mathcal{I}} \max_{1\le i\le n} \mathbb{V}[u_{i,h}(\mathsf{x})] = O(h^{-1}), \qquad \sup_{\mathsf{x}\in\mathcal{I}} \max_{1\le i\le n} |u_{i,h}(\mathsf{x})| = O(h^{-1}).$$

Let $|\mathsf{x} - \mathsf{y}| \le \varepsilon \le h$, we also have

$$\begin{aligned}
|u_{i,h}(\mathsf{x}) - u_{i,h}(\mathsf{y})| &\le \left| R\left(\frac{x_i - \mathsf{x}}{h}\right)^2 \frac{1}{h} K\left(\frac{x_i - \mathsf{x}}{h}\right) - R\left(\frac{x_i - \mathsf{y}}{h}\right)^2 \frac{1}{h} K\left(\frac{x_i - \mathsf{y}}{h}\right) \right| + |\Gamma_{h,\mathsf{x}} - \Gamma_{h,\mathsf{y}}| \\
&\le \left| R\left(\frac{x_i - \mathsf{x}}{h}\right)^2 - R\left(\frac{x_i - \mathsf{y}}{h}\right)^2 \right| \frac{1}{h} K\left(\frac{x_i - \mathsf{x}}{h}\right) + R\left(\frac{x_i - \mathsf{y}}{h}\right)^2 \frac{1}{h} \left| K\left(\frac{x_i - \mathsf{x}}{h}\right) - K\left(\frac{x_i - \mathsf{y}}{h}\right) \right| + |\Gamma_{h,\mathsf{x}} - \Gamma_{h,\mathsf{y}}| \\
&\le M\left[\frac{\varepsilon}{h}\frac{1}{h} K\left(\frac{x_i - \mathsf{x}}{h}\right) + \frac{\varepsilon}{h}\frac{1}{h} K^\dagger\left(\frac{x_i - \mathsf{x}}{h}\right) + \frac{1}{h} K^\ddagger\left(\frac{x_i - \mathsf{x}}{h}\right) + \frac{\varepsilon}{h} \right].
\end{aligned}$$

where M is some constant that does not depend on n, h or ε. Then it is easy to see that

$$\sup_{\mathsf{x}\in\mathcal{I}} \max_{1\le i\le n} \mathbb{V}[u_{i,h,\varepsilon}(\mathsf{x})] = O\left(\frac{\varepsilon}{h^2}\right), \qquad \sup_{\mathsf{x}\in\mathcal{I}} \max_{1\le i\le n} |u_{i,h,\varepsilon}(\mathsf{x}) - \mathbb{E}[u_{i,h,\varepsilon}(\mathsf{x})]| = O(h^{-1}), \quad \sup_{\mathsf{x}\in\mathcal{I}} \max_{1\le i\le n} \mathbb{E}[|u_{i,h,\varepsilon}(\mathsf{x})|] = O\left(\frac{\varepsilon}{h}\right).$$

Now take $\varepsilon = \sqrt{h\log n/n}$, then $\log N(\varepsilon, \mathcal{I}, |\cdot|) = O(\log n)$. Lemma 24 implies that

$$\sup_{\mathsf{x}\in\mathcal{I}} \left| \frac{1}{n}\sum_{i=1}^n R\left(\frac{x_i - \mathsf{x}}{h}\right)^2 \frac{1}{h} K\left(\frac{x_i - \mathsf{x}}{h}\right) - \Gamma_{h,\mathsf{x}} \right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh}} \right).$$

5.18 Proof of Lemma 17

Let $R_i(\mathsf{x}) = R(x_i - \mathsf{x})$ and $W_i(\mathsf{x}) = K((x_i - \mathsf{x})/h)/h$, then we split $\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}$ into two terms,

$$\text{(I)} = \frac{1}{n^3}\sum_{i,j,k} \Upsilon_h R_j(\mathsf{x})R_k(\mathsf{y})'\Upsilon_h W_j(\mathsf{x})W_k(\mathsf{y})\Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big)\Big(\mathbb{1}(x_i \le x_k) - F(x_k)\Big)$$

$$\text{(II)} = -\frac{1}{n^2}\sum_{j,k} \Upsilon_h R_j(\mathsf{x})R_k(\mathsf{y})'\Upsilon_h W_j(\mathsf{x})W_k(\mathsf{y})\Big(\hat{F}(x_j) - F(x_j)\Big)\Big(\hat{F}(x_k) - F(x_k)\Big).$$

(II) satisfies

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}} |\text{(II)}| \le \sup_x |\hat{F}(x) - F(x)|^2 \left(\sup_{\mathsf{x}\in\mathcal{I}} \frac{1}{n}\sum_j |\Upsilon_h R_j(\mathsf{x})W_j(\mathsf{x})|\right)^2.$$

It is obvious that

$$\sup_x |\hat{F}(x) - F(x)|^2 = O_{\mathbb{P}}\left(\frac{1}{n}\right).$$

As for the second part, one can employ the same technique used to prove Lemma 16 and show that

$$\sup_{\mathsf{x}\in\mathcal{I}} \frac{1}{n}\sum_j |\Upsilon_h R_j(\mathsf{x})W_j(\mathsf{x})| = O_{\mathbb{P}}(1),$$

implying that

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}} |\text{(II)}| = O_{\mathbb{P}}\left(\frac{1}{n}\right).$$

For (I), we first define

$$u_{ij}(\mathsf{x}) = \Upsilon_h R_j(\mathsf{x})W_j(\mathsf{x})\Big(\mathbb{1}(x_i \le x_j) - F(x_j)\Big),$$

and

$$\bar{u}_i(\mathsf{x}) = \mathbb{E}[u_{ij}(\mathsf{x})|x_i; i \ne j], \qquad \hat{u}_i(\mathsf{x}) = \frac{1}{n}\sum_j u_{ij}(\mathsf{x}).$$

Then

$$\begin{aligned}
\text{(I)} &= \frac{1}{n}\sum_i \left(\frac{1}{n}\sum_j u_{ij}(\mathsf{x})\right)\left(\frac{1}{n}\sum_j u_{ij}(\mathsf{y})\right)' = \frac{1}{n}\sum_i \hat{u}_i(\mathsf{x})\hat{u}_i(\mathsf{y})' \\
&= \frac{1}{n}\sum_i \bar{u}_i(\mathsf{x})\bar{u}_i(\mathsf{y})' + \frac{1}{n}\sum_i \left(\hat{u}_i(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\hat{u}_i(\mathsf{y})' + \frac{1}{n}\sum_i \bar{u}_i(\mathsf{x})\left(\hat{u}_i(\mathsf{y}) - \bar{u}_i(\mathsf{y})\right)' \\
&= \underbrace{\frac{1}{n}\sum_i \bar{u}_i(\mathsf{x})\bar{u}_i(\mathsf{y})'}_{\text{(I.1)}} + \underbrace{\frac{1}{n}\sum_i \left(\hat{u}_i(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})'}_{\text{(I.2)}} + \underbrace{\frac{1}{n}\sum_i \bar{u}_i(\mathsf{x})\left(\hat{u}_i(\mathsf{y}) - \bar{u}_i(\mathsf{y})\right)'}_{\text{(I.3)}} \\
&\qquad + \underbrace{\frac{1}{n}\sum_i \left(\hat{u}_i(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\left(\hat{u}_i(\mathsf{y}) - \bar{u}_i(\mathsf{y})\right)'}_{\text{(I.4)}}.
\end{aligned}$$

Term (I.1) has been analyzed in Lemma 11, which satisfies

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}} |\text{(I.1)} - \Sigma_{h,\mathsf{x},\mathsf{y}}| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right).$$

Term (I.2) has expansion:

$$(\text{I}.2) = \frac{1}{n^2}\sum_{i,j}\left(u_{ij}(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})' = \underbrace{\frac{1}{n^2}\sum_{\substack{i,j\\ \text{distinct}}}\left(u_{ij}(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})'}_{(\text{I}.2.1)} + \underbrace{\frac{1}{n^2}\sum_{i}\left(u_{ii}(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})'}_{(\text{I}.2.2)}.$$

By the same technique of Lemma 16, one can show that

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|(\text{I}.2.2)\right| = O_{\mathbb{P}}\left(\frac{1}{n}\right).$$

We need a further decomposition to make (I.2.1) a degenerate U-statistic:

$$\begin{aligned}(\text{I}.2.1) = &\underbrace{\frac{n-1}{n^2}\sum_{j}\mathbb{E}\left[\left(u_{ij}(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})'\middle| x_j\right]}_{(\text{I}.2.1.1)}\\ &+ \underbrace{\frac{1}{n^2}\sum_{\substack{i,j\\ \text{distinct}}}\left\{\left(u_{ij}(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})' - \mathbb{E}\left[\left(u_{ij}(\mathsf{x}) - \bar{u}_i(\mathsf{x})\right)\bar{u}_i(\mathsf{y})'\middle| x_j\right]\right\}}_{(\text{I}.2.1.2)}.\end{aligned}$$

(I.2.1) has zero mean. By discretizing $\mathcal{I}$ and apply Bernstein's inequality, one can show that the (I.2.1.1) has order $O_{\mathbb{P}}\left(\sqrt{\log n/n}\right)$.

For (I.2.1.2), we first discretize $\mathcal{I}$ and then apply a Bernstein-type inequality (Lemma 25) for degenerate U-statistics, which gives an order

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|(\text{I}.2.1.2)\right| = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{n^2h}}\right).$$

Overall, we have

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|(\text{I}.2)\right| = O_{\mathbb{P}}\left(\frac{1}{n} + \sqrt{\frac{\log n}{n}} + \frac{\log n}{\sqrt{n^2h}}\right) = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right),$$

and the same bound applies to (I.3).

For (I.4), one can show that

$$\sup_{\mathsf{x}\in\mathcal{I}}\sup_{x\in\mathcal{X}}\left|\frac{1}{n}\sum_{j}\Upsilon_h R_j(\mathsf{x})W_j(\mathsf{x})\Big(\mathbb{1}(x\le x_j) - F(x_j)\Big) - \mathbb{E}\left[\Upsilon_h R_j(\mathsf{x})W_j(\mathsf{x})\Big(\mathbb{1}(x\le x_j) - F(x_j)\Big)\right]\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh}}\right),$$

which means

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|(\text{I}.4)\right| = O_{\mathbb{P}}\left(\frac{\log n}{nh}\right) = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right),$$

under our assumption that $\log n/(nh^2)\to 0$.

As a result, we have

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}} - \Sigma_{h,\mathsf{x},\mathsf{y}}\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n}}\right).$$

Now take c to be a generic vector. Then we have

$$\frac{c'_{h,\mathsf{x}}\Upsilon_h(\hat{\Omega}_{h,\mathsf{x},\mathsf{y}}-\Omega_{h,\mathsf{x},\mathsf{y}})\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}} = \frac{c'_{h,\mathsf{x}}\Upsilon_h\hat{\Gamma}^{-1}_{h,\mathsf{x}}(\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}-\Sigma_{h,\mathsf{x},\mathsf{y}})\hat{\Gamma}^{-1}_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}$$
$$+\frac{c'_{h,\mathsf{x}}\Upsilon_h(\hat{\Gamma}^{-1}_{h,\mathsf{x}}-\Gamma^{-1}_{h,\mathsf{x}})\Sigma_{h,\mathsf{x},\mathsf{y}}\hat{\Gamma}^{-1}_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}$$
$$+\frac{c'_{h,\mathsf{x}}\Upsilon_h\Gamma^{-1}_{h,\mathsf{x}}\Sigma_{h,\mathsf{x},\mathsf{y}}(\hat{\Gamma}^{-1}_{h,\mathsf{y}}-\Gamma^{-1}_{h,\mathsf{y}})\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}.$$

From the analysis of $\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}$, we have

$$\sup_{\mathsf{x},\mathsf{y}\in\mathcal{I}}\left|\frac{c'_{h,\mathsf{x}}\Upsilon_h\hat{\Gamma}^{-1}_{h,\mathsf{x}}(\hat{\Sigma}_{h,\mathsf{x},\mathsf{y}}-\Sigma_{h,\mathsf{x},\mathsf{y}})\hat{\Gamma}^{-1}_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^2}}\right).$$

For the second term, we have

$$\left|\frac{c'_{h,\mathsf{x}}\Upsilon_h(\hat{\Gamma}^{-1}_{h,\mathsf{x}}-\Gamma^{-1}_{h,\mathsf{x}})\Sigma_{h,\mathsf{x},\mathsf{y}}\hat{\Gamma}^{-1}_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}\right| \le \frac{|c'_{h,\mathsf{x}}\Upsilon_h(\hat{\Gamma}^{-1}_{h,\mathsf{x}}-\Gamma^{-1}_{h,\mathsf{x}})\Sigma^{1/2}_{h,\mathsf{x}}|\cdot|c'_{h,\mathsf{y}}\Upsilon_h\hat{\Gamma}^{-1}_{h,\mathsf{y}}\Sigma^{1/2}_{h,\mathsf{y}}|}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\Omega_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}\sqrt{c'_{h,\mathsf{y}}\Upsilon_h\Omega_{h,\mathsf{y}}\Upsilon_h c_{h,\mathsf{y}}}}$$
$$= O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh^2}}\right).$$

The same bound holds for the third term.

5.19 Proof of Lemma 18

We decompose (19) as

$$\sup_{\mathsf{x}\in\mathcal{I}}|(19)| \le \frac{1}{\sqrt{n}}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{c'_{h,\mathsf{x}}\Upsilon_h\hat{\Gamma}^{-1}_{h,\mathsf{x}}}{\sqrt{c'_{h,\mathsf{x}}\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right|\right]}_{\text{(I)}}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^{n}R((x_i-\mathsf{x})/h)[1-F(x_i)]\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})\right|\right]}_{\text{(II)}}.$$

As both $\hat{\Gamma}_{h,\mathsf{x}}$ and $c'_{h,\mathsf{x}}\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}$ are uniformly consistent, term (I) has order

$$\text{(I)} = O_{\mathbb{P}}\left(\sqrt{\frac{1}{h}}\right).$$

For (II), we can employ the same technique used to prove Lemma 16 and show that

$$\text{(II)} = O_{\mathbb{P}}\left(1+\sqrt{\frac{\log n}{nh}}\right) = O_{\mathbb{P}}(1),$$

where the leading order in the above represents the mean of $R((x_i-\mathsf{x})/h)[1-F(x_i)]\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})$.

5.20 Proof of Lemma 19

To start, term (20) is bounded by

$$\sup_{\mathsf{x}\in\mathcal{I}}|(20)| \leq \sqrt{n}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right|\right]}_{\text{(I)}}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^{n}R((x_i-\mathsf{x})/h)[F(x_i)-\theta(\mathsf{x})'R(x_i-\mathsf{x})]\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})\right|\right]}_{\text{(II)}}.$$

Employing the same argument used to prove Lemma 18, we have

$$\text{(I)} = O_{\mathbb{P}}\left(\sqrt{\frac{1}{h}}\right).$$

To bound term (II), recall that $K(\cdot)$ is supported on $[-1,1]$, meaning that

$$\begin{aligned}
&\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^{n}R((x_i-\mathsf{x})/h)[F(x_i)-\theta(\mathsf{x})'R(x_i-\mathsf{x})]\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})\right| \\
&=\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^{n}R((x_i-\mathsf{x})/h)[F(x_i)-\theta(\mathsf{x})'R(x_i-\mathsf{x})]\mathbb{1}(|x_i-\mathsf{x}|\leq h)\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})\right| \\
&\leq \underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\frac{1}{n}\sum_{i=1}^{n}\left|R((x_i-\mathsf{x})/h)\frac{1}{h}K(\frac{x_i-\mathsf{x}}{h})\right|\right]}_{\text{(II.1)}}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\sup_{u\in[\mathsf{x}-h,\mathsf{x}+h]}\left|\Big[F(u)-\theta(\mathsf{x})'R(u-\mathsf{x})\Big]\right|\right]}_{\text{(II.2)}}.
\end{aligned}$$

Term (II.2) has the bound $\sup_{\mathsf{x}\in\mathcal{I}}\varrho(h,\mathsf{x})$. Term (II.1) can be bounded by mean and variance calculations and adopting the proof of Lemma 16, which leads to

$$\text{(II.1)} = O_{\mathbb{P}}\left(1+\sqrt{\frac{\log n}{nh}}\right) = O_{\mathbb{P}}(1).$$

5.21 Proof of Lemma 20

For ease of exposition, define the following:

$$u_{ij}(\mathsf{x}) = \Upsilon_h R(x_j-\mathsf{x})\Big[\mathbb{1}(x_i\leq x_j)-F(x_j)\Big]\frac{1}{h}K\left(\frac{x_j-\mathsf{x}}{h}\right) - \int_{\frac{\mathcal{X}-\mathsf{x}}{h}} R(u)\Big[\mathbb{1}(x_i\leq \mathsf{x}+hu)-F(\mathsf{x}+hu)\Big]K\left(u\right)f(\mathsf{x}+hu)\mathrm{d}u,$$

then $n^{-2}\sum_{i,j=1,i\neq j}^{n}u_{ij}(\mathsf{x})$ is a degenerate U-statistic. We rewrite (21) as

$$\sup_{\mathsf{x}\in\mathcal{I}}|(21)| \leq \sqrt{n}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Gamma}_{h,\mathsf{x}}^{-1}}{\sqrt{c_{h,\mathsf{x}}'\Upsilon_h\hat{\Omega}_{h,\mathsf{x}}\Upsilon_h c_{h,\mathsf{x}}}}\right|\right]}_{\text{(I)}}\underbrace{\left[\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n}u_{ij}\right|\right]}_{\text{(II)}}.$$

As before, we have

$$\text{(I)} = O_{\mathbb{P}}\left(\sqrt{\frac{1}{h}}\right).$$

Now we consider (II). Let $\mathcal{I}_\varepsilon$ be an $\frac{\varepsilon}{2}$-covering of $\mathcal{I}$, we have

$$\sup_{\mathsf{x}\in\mathcal{I}}\left|\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n}u_{ij}(\mathsf{x})\right| \le \underbrace{\max_{\mathsf{x}\in\mathcal{I}_\varepsilon}\left|\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n}u_{ij}(\mathsf{x})\right|}_{\text{(II.1)}} + \underbrace{\max_{\mathsf{x}\in\mathcal{I}_\varepsilon,\mathsf{y}\in\mathcal{I},|\mathsf{x}-\mathsf{y}|\le\varepsilon}\left|\frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n}\Big(u_{ij}(\mathsf{x})-u_{ij}(\mathsf{y})\Big)\right|}_{\text{(II.2)}}.$$

We rely on the concentration inequality in Lemma 25 for degenerate second order U-statistics. By our assumptions, A can be chosen to be C_1h^{-1} where C_1 is some constant that is independent of x. Similarly, B can be chosen to be $C_2\sqrt{n}h^{-1}$ for some constant C_2 which is independent of x, and D can be chosen as $C_3nh^{-1/2}$ for some C_3 independent of x. Therefore, by setting $\eta = K\log n/\sqrt{n^2h}$ for some large constant K, we have

$$\begin{aligned}\mathbb{P}\left[\text{(II.1)}\ge\eta\right] &\le C\frac{1}{\varepsilon}\max_{\mathsf{x}\in\mathcal{I}_\varepsilon}\mathbb{P}\left[\left|\sum_{i,j=1,i\neq j}^{n}u_{ij}(\mathsf{x})\right|\ge n^2\eta\right]\\ &\le C\frac{1}{\varepsilon}\exp\left\{-\frac{1}{C}\min\left[\frac{n^2h^{1/2}\eta}{nc_3},\ \left(\frac{n^2h\eta}{n^{1/2}c_2}\right)^{\frac{2}{3}},\ \left(\frac{n^2h\eta}{c_1}\right)^{\frac{1}{2}}\right]\right\}\\ &= C\frac{1}{\varepsilon}\exp\left\{-\frac{1}{C}\min\left[\frac{K\log n}{c_3},\ \left(\frac{K\sqrt{nh}\log n}{c_2}\right)^{\frac{2}{3}},\ \left(\frac{K\sqrt{n^2h}\log n}{c_1}\right)^{\frac{1}{2}}\right]\right\}.\end{aligned}$$

As ε is at most polynomial in n, the above tends to zero for all K large enough, which implies

$$\text{(II.1)} = O_{\mathbb{P}}\left(\frac{\log n}{\sqrt{n^2h}}\right).$$

With tedious but still straightforward calculations, it can be shown that

$$\text{(II.2)} = O_{\mathbb{P}}\left(\frac{\varepsilon}{h}+\frac{\log n}{\sqrt{n^2h}}+\frac{\varepsilon}{h}\frac{\log n}{\sqrt{n^2h}}\right),$$

and to match the rates, let $\varepsilon = h\log n/\sqrt{n^2h}$.

5.22 Proof of Lemma 21

The proof resembles that of of Lemma 13.

5.23 Proof of Theorem 22

The proof resembles that of Theorem 14.

5.24 Proof of Theorem 23

The proof resembles that of Theorem 15.

References

Cattaneo, M. D., M. Jansson, and X. Ma (2020): "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 115(531), 1449–1455.

Giné, E., V. Koltchinskii, and L. Sakhanenko (2004): "Kernel Density Estimators: Convergence in Distribution for Weighted Sup-Norms," *Probability Theory and Related Fields*, 130(2), 167–198.

GINÉ, E., R. LATAŁA, AND J. ZINN (2000): "Exponential and Moment Inequalities for U-statistics," in *High Dimensional Probability II*. Springer.

GRANOVSKY, B. L., AND H.-G. MÜLLER (1991): "Optimizing Kernel Methods: A Unifying Variational Principle," *International Statistical Review/Revue Internationale de Statistique*, 59(3), 373–388.

LOADER, C. (2006): *Local Regression and Likelihood*. Springer, New York.