Kernel Regression Utilizing External Information as Constraints

Department of Statistics, University of Wisconsin-Madison

School of Statistics, East China Normal University

Supplementary Material

S1 Proof of Theorem 1

Consider the following decomposition:

$$\sqrt{nb^p}\{\widehat{\mu}_{CK}(\boldsymbol{u}) - \mu(\boldsymbol{u})\} = T_1 + \cdots + T_6, \tag{S1.1}$$

where

$$T_1 = n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{B}_l^{-1}\boldsymbol{\Delta}_l\boldsymbol{\epsilon}/\widehat{f}_b(\boldsymbol{u}),$$

$$T_2 = n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top\{\widetilde{\boldsymbol{\mu}} - \mu(\boldsymbol{u})\mathbf{1}_n\}/\widehat{f}_b(\boldsymbol{u}),$$

$$T_3 = n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top(\boldsymbol{B}_l^{-1}\boldsymbol{\Delta}_l\widetilde{\boldsymbol{\mu}} - \widetilde{\boldsymbol{\mu}})/\widehat{f}_b(\boldsymbol{u}),$$

$$T_4 = -n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top\boldsymbol{P}(\boldsymbol{B}_l^{-1}\boldsymbol{\Delta}_l\widetilde{\boldsymbol{\mu}} - \widetilde{\boldsymbol{\mu}})/\widehat{f}_b(\boldsymbol{u}),$$

$$T_5 = n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{h})/\widehat{f}_b(\boldsymbol{u}),$$

$$T_6 = n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top\boldsymbol{P}(\boldsymbol{h} - \widetilde{\boldsymbol{\mu}})/\widehat{f}_b(\boldsymbol{u}),$$

$\widehat{f}_b(\boldsymbol{u}) = \sum_{i=1}^n \kappa_b(\boldsymbol{u} - \boldsymbol{U}_i)/n$, $\boldsymbol{\delta}_b(\boldsymbol{u}) = (\kappa_b(\boldsymbol{u} - \boldsymbol{U}_1), \ldots, \kappa_b(\boldsymbol{u} - \boldsymbol{U}_n))^\top$, $\boldsymbol{I}_n$ is the identity matrix of order n, $\boldsymbol{P} = \boldsymbol{G}(\boldsymbol{G}^\top\boldsymbol{G})^{-1}\boldsymbol{G}^\top$ with $\boldsymbol{G}$ defined in (2.6), $\boldsymbol{B}_l$ is the $n \times n$ diagonal matrix whose ith diagonal element is $\widehat{f}_l(\boldsymbol{U}_i)$, $\boldsymbol{\Delta}_l$ is the $n \times n$ matrix whose (i, j)th entry is $\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)/n$, $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^\top$ with $\epsilon_i = Y_i - \mu(\boldsymbol{U}_i)$, $\widetilde{\boldsymbol{\mu}} = (\mu(\boldsymbol{U}_1), ..., \mu(\boldsymbol{U}_n))^\top$, $\mathbf{1}_n$ is the n-vector with all components being 1, $\boldsymbol{h} = \boldsymbol{G}\boldsymbol{\beta}_g$ with $\boldsymbol{\beta}_g$ defined in (2.3), and $\widehat{\boldsymbol{h}} = \boldsymbol{G}\widehat{\boldsymbol{\beta}}_g$ with $\widehat{\boldsymbol{\beta}}_g$ defined in (2.5).

We first show that T_1 in (S1.1) is asymptotically normal with mean 0 and variance $V_{CK}(\boldsymbol{u})$ defined in (2.10). Define

$$S(\boldsymbol{U}_i, \epsilon_i, \boldsymbol{U}_j, \epsilon_j) = \frac{b^{p/2}}{2f_U(\boldsymbol{u})}\left\{\frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i-\boldsymbol{U}_j)\epsilon_j}{f_U(\boldsymbol{U}_i)} + \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_j)\kappa_l(\boldsymbol{U}_j-\boldsymbol{U}_i)\epsilon_i}{f_U(\boldsymbol{U}_j)}\right\}.$$

Consider a further decomposition of T_1:

$$T_1 = \sqrt{n}\,V + T_{11} + T_{12} + T_{13}, \tag{S1.2}$$

where

$$V = \frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n S(\boldsymbol{U}_i, \epsilon_i, \boldsymbol{U}_j, \epsilon_j),$$

$$T_{11} = \frac{b^{p/2}}{n^{3/2}}\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u} - \boldsymbol{U}_i)\kappa_l(0)\epsilon_i}{f_U(\boldsymbol{u})f_U(\boldsymbol{U}_i)},$$

$$T_{12} = \frac{b^{p/2}}{n^{3/2}}\sum_{j=1}^n\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u} - \boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)}{f_U(\boldsymbol{u})f_U(\boldsymbol{U}_i)}\left\{\frac{f_U(\boldsymbol{u})f_U(\boldsymbol{U}_i)}{\widehat{f}_b(\boldsymbol{u})\widehat{f}_l(\boldsymbol{U}_i)} - 1\right\}\epsilon_j,$$

$$T_{13} = -n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\boldsymbol{u})^\top\boldsymbol{P}\boldsymbol{B}_l^{-1}\boldsymbol{\Delta}_l\boldsymbol{\epsilon}/\widehat{f}_b(\boldsymbol{u}).$$

The decomposition (S1.2) follows from

$$\sqrt{n}V + T_{11} = \frac{b^{p/2}}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\kappa_b(\boldsymbol{u} - \boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)}{f_U(\boldsymbol{u}) f_U(\boldsymbol{U}_i)} \epsilon_j$$

and

$$\begin{aligned}\sqrt{n}V + T_{11} + T_{12} &= \frac{b^{p/2}}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\kappa_b(\boldsymbol{u} - \boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i - \boldsymbol{U}_j)}{\widehat{f}_l(\boldsymbol{u}) \widehat{f}_l(\boldsymbol{U}_i)} \epsilon_j \\ &= n^{-1/2} b^{p/2} \boldsymbol{\delta}_b(\boldsymbol{u})^\top \boldsymbol{B}_l^{-1} \boldsymbol{\Delta}_l \boldsymbol{\epsilon} / \widehat{f}_b(\boldsymbol{u}).\end{aligned}$$

Note that V is a V-statistic with

$$\begin{aligned}S_1(\boldsymbol{U}_1, \epsilon_1) &= \mathrm{E}\{S(\boldsymbol{U}_1, \epsilon_1, \boldsymbol{U}_2, \epsilon_2) \mid \boldsymbol{U}_1, \epsilon_1\} \\ &= \frac{b^{p/2}}{2 f_U(\boldsymbol{u})} \left\{ \int \kappa_l(\boldsymbol{u}_2 - \boldsymbol{U}_1) \kappa_b(\boldsymbol{u} - \boldsymbol{u}_2) d\boldsymbol{u}_2 \right\} \epsilon_1,\end{aligned}$$

which has variance

$$\begin{aligned}\mathrm{Var}\{S_1(\boldsymbol{U}_1, \epsilon_1)\} &= \frac{b^{p/2}}{4 f_U^2(\boldsymbol{u})} \int f_U(\boldsymbol{u}_1) \sigma^2(\boldsymbol{u}_1) \left\{ \int \kappa_l(\boldsymbol{u}_2 - \boldsymbol{u}_1) \kappa_b(\boldsymbol{u} - \boldsymbol{u}_2) d\boldsymbol{u}_2 \right\}^2 d\boldsymbol{u}_1 \\ &= \frac{b^{p/2}}{4 f_U^2(\boldsymbol{u})} \int f_U(\boldsymbol{u}_1) \sigma^2(\boldsymbol{u}_1) \left\{ \int \kappa_l(\boldsymbol{v}) \kappa_b(\boldsymbol{u} - \boldsymbol{u}_1 - l\boldsymbol{\nu}) d\boldsymbol{\nu} \right\}^2 d\boldsymbol{u}_1 \\ &= \frac{1}{4 f_U^2(\boldsymbol{u})} \int f_U(\boldsymbol{u} - b\boldsymbol{w}) \sigma^2(\boldsymbol{u} - b\boldsymbol{w}) \left\{ \int \kappa(\boldsymbol{v}) \kappa\left(\boldsymbol{w} - \boldsymbol{\nu}\frac{l}{b}\right) d\boldsymbol{\nu} \right\}^2 d\boldsymbol{w},\end{aligned}$$

where $\sigma^2(\cdot)$ is given in condition (A2), the second and third equalities follow from changing variables $\boldsymbol{u}_2 - \boldsymbol{u}_1 = l\boldsymbol{\nu}$ and $\boldsymbol{u} - \boldsymbol{u}_1 = b\boldsymbol{w}$, respectively. From the continuity of $f_U(\cdot)$ and $\sigma^2(\cdot)$, $\mathrm{Var}\{S_1(\boldsymbol{u}_1, \epsilon_1)\}$ converges to $V_{CK}(\boldsymbol{u})$. Therefore, by the theory for asymptotic normality of V-statistics (e.g., Theorem 3.16 in Shao (2003)),

$$\sqrt{n}\,V \xrightarrow{d} N\big(0, V_{CK}(\boldsymbol{u})\big).$$

Conditioned on $\boldsymbol{U}_1, ..., \boldsymbol{U}_n$, T_{11} has mean 0 and variance

$$\begin{aligned}\mathrm{Var}(T_{11}|\boldsymbol{U}_1, ..., \boldsymbol{U}_n) &= \frac{b^p}{4f_U^2(\boldsymbol{u})n^3}\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)^2\kappa_l(0)^2\sigma^2(\boldsymbol{U}_i)}{f_U(\boldsymbol{U}_i)} \\ &\le \frac{\sup_{\boldsymbol{u}\in\mathbb{U}}\kappa(\boldsymbol{u})^3}{4f_U^2(\boldsymbol{u})n^3l^{2p}}\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\sigma^2(\boldsymbol{U}_i)}{f_U(\boldsymbol{U}_i)} \\ &= n^{-2}l^{-2p}\, O_p(1) = o_p(1),\end{aligned}$$

where $O_p(a_n)$ denotes a term bounded by a_n in probability and $o_p(1)$ denotes a term $\xrightarrow{p} 0$. This proves that $T_{11} = o_p(1)$. Note that $\mathrm{E}(T_{12} \mid \boldsymbol{U}_1, ..., \boldsymbol{U}_n) = 0$ and, from

$$\begin{aligned}&\left|\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i-\boldsymbol{U}_j)}{f_U(\boldsymbol{u})f_U(\boldsymbol{U}_i)} - \sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i-\boldsymbol{U}_j)}{\widehat{f}_b(\boldsymbol{u})\widehat{f}_l(\boldsymbol{U}_i)}\right| \\ &\le \max\left\{\frac{1}{f_U(\boldsymbol{u})}, \frac{1}{\widehat{f}_b(\boldsymbol{u})}\right\}\max_{i=1,\ldots,n}\left|\frac{f_U(\boldsymbol{U}_i)}{\widehat{f}_l(\boldsymbol{U}_i)} - 1\right| \sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\kappa_l(\boldsymbol{U}_i-\boldsymbol{U}_j)}{f_U(\boldsymbol{u})f_U(\boldsymbol{U}_i)},\end{aligned}$$

$\mathrm{Var}(T_{12} \mid \boldsymbol{U}_1, ..., \boldsymbol{U}_n)$ is bounded by

$$\max\left\{\frac{1}{f_U^2(\boldsymbol{u})}, \frac{1}{\widehat{f}_b^2(\boldsymbol{u})}\right\}\max_{i=1,\ldots,n}\left|\frac{f_U(\boldsymbol{U}_i)}{\widehat{f}_l(\boldsymbol{U}_i)} - 1\right|^2 \mathrm{Var}(\sqrt{n}V + T_{11}|\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n),$$

Therefore, under the assumed condition that f_U is bounded away from zero, (S4.7) in Lemma 3 implies $T_{12} = o_p(1)$. Define

$$W_j(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)^\top}{\widehat{f}_b(\boldsymbol{u})}(\boldsymbol{G}^\top\boldsymbol{G})^{-1}\sum_{i=1}^n \frac{\kappa_l(\boldsymbol{U}_i-\boldsymbol{U}_j)\boldsymbol{g}(\boldsymbol{X}_i)}{\widehat{f}_l(\boldsymbol{U}_i)}.$$

Then

$$T_{13} = \frac{b^{p/2}}{n^{1/2}} \sum_{j=1}^{n} W_j(\boldsymbol{u})\epsilon_j.$$

Conditioned on $\boldsymbol{U}_1, ..., \boldsymbol{U}_n$, T_{13} has mean 0 and variance

$$\text{Var}(T_{13} \mid \boldsymbol{U}_1, ..., \boldsymbol{U}_n) = \frac{b^p}{n} \sum_{j=1}^{n} W_j^2(\boldsymbol{u})\sigma^2(\boldsymbol{U}_j).$$

Under the assumed condition that f_U is bounded away from zero, (S4.7) and (S4.9) in Lemma 3 imply

$$\max_{j=1,...,n} \left| W_j(\boldsymbol{u}) - \boldsymbol{g}(\boldsymbol{u})^\top \boldsymbol{\Sigma}_g^{-1} \boldsymbol{g}(\boldsymbol{X}_j) \right| = o_p(1).$$

As a result, $\text{Var}(T_{13} \mid \boldsymbol{U}_1, ..., \boldsymbol{U}_n) = O_p(b^p) = o_p(1)$. This concludes that $T_{13} = o_p(1)$. Consequently, by (S1.2), T_1 has the same asymptotic distribution as $\sqrt{n}V$, the claimed result.

It remains to show that $T_2 + \cdots + T_6$ in (S1.1) converges to $B_{CK}(\boldsymbol{u})$ in probability. From Lemma 4 and (A4),

$$T_2 = \sqrt{nb^p} b^2 A(\boldsymbol{u})\{1 + o_p(1)\} = \sqrt{c} A(\boldsymbol{u})\{1 + o_p(1)\}.$$

Note that

$$\begin{aligned} T_3 &= \frac{\sqrt{nb^p} l^2}{n\widehat{f}_b(\boldsymbol{u})} \sum_{j=1}^{n} \kappa_b(\boldsymbol{u} - \boldsymbol{U}_j) \left[\frac{1}{nl^2 \widehat{f}(\boldsymbol{U}_j)} \sum_{i=1}^{n} \kappa_l(\boldsymbol{u} - \boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i) - \mu(\boldsymbol{U}_j)\} \right] \\ &= \left\{ \frac{\sqrt{c} r^2}{n f_b(\boldsymbol{u})} \sum_{j=1}^{n} \kappa_b(\boldsymbol{u} - \boldsymbol{U}_j) A(\boldsymbol{U}_j) \right\} \{1 + o_p(1)\} \\ &= \sqrt{c} r^2 A(\boldsymbol{u})\{1 + o_p(1)\}, \end{aligned}$$

where the second equality follows from (S4.7) in Lemma 3, Lemma 4 and (A4), and the last equality follows from Lemma 2 and continuity of $A(\cdot)$.

The term T_4 divided by $-n^{-1/2}b^{p/2}$ is equal to

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)^\top}{\widehat{f}_b(\boldsymbol{u})}(\boldsymbol{G}^\top\boldsymbol{G})^{-1}\sum_{j=1}^{n}\frac{\boldsymbol{g}(\boldsymbol{X}_j)}{n\widehat{f}(\boldsymbol{U}_j)}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{U}_j)\}\\
&=\left\{\boldsymbol{g}(\boldsymbol{x})^\top\boldsymbol{\Sigma}_g^{-1}\frac{1}{n}\sum_{j=1}^{n}\frac{\boldsymbol{g}(\boldsymbol{X}_j)}{n\widehat{f}(\boldsymbol{U}_j)}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{U}_j)\}\right\}\{1+o_p(1)\}\\
&=\left\{\boldsymbol{g}(\boldsymbol{x})^\top\boldsymbol{\Sigma}_g^{-1}\frac{l^{2/p}}{n}\sum_{j=1}^{n}\boldsymbol{g}(\boldsymbol{X}_j)A(\boldsymbol{U}_j)\right\}\{1+o_p(1)\}\\
&=l^{2/p}\boldsymbol{g}(\boldsymbol{x})^\top\boldsymbol{\Sigma}_g^{-1}\mathrm{E}\{\boldsymbol{g}(\boldsymbol{X})A(\boldsymbol{U})\}\{1+o_p(1)\},
\end{aligned}$$

where the first equality follows from (S4.7)-(S4.8) in the Appendix and the law of large numbers, the second equality follows from Lemma 4, and the last equality follows from the law of large numbers. Hence,

$$T_4=-\sqrt{c}r^2g(\boldsymbol{x})^\top\boldsymbol{\Sigma}_g^{-1}\mathrm{E}\{\boldsymbol{g}(\boldsymbol{X})A(\boldsymbol{U})\}\{1+o_p(1)\}.$$

Similarly, T_5 divided by $n^{-1/2}b^{p/2}$ is equal to

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)^\top}{\widehat{f}_b(\boldsymbol{u})}(\boldsymbol{G}^\top\boldsymbol{G})^{-1}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{X}_i)\{\widehat{h}(\boldsymbol{X}_i)-h(\boldsymbol{X}_i)\}\\
&=\left[\boldsymbol{g}(\boldsymbol{x})^\top\boldsymbol{\Sigma}_g^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{X}_i)\{\widehat{h}(\boldsymbol{X}_i)-h(\boldsymbol{X}_i)\}\right]\{1+o_p(1)\}\\
&\leq\{1+o_p(1)\}M\max_{j=1,\ldots,n}|\widehat{h}(\boldsymbol{X}_j)-h(\boldsymbol{X}_j)|,
\end{aligned}$$

where $\widehat{h}(\boldsymbol{X}_i)$ and $h(\boldsymbol{X}_i)$ are the ith components of $\widehat{\boldsymbol{h}}$ and $\boldsymbol{h}$, respectively, the first equality follows from (S4.7)-(S4.8), and the inequality holds for a non-random constant M. Since $\widehat{h}(\boldsymbol{X}_j) - h(\boldsymbol{X}_j) = \boldsymbol{g}(\boldsymbol{X}_j)^\top(\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g)$,

$$\max_{j=1,\ldots,n} |\widehat{h}(\boldsymbol{X}_j) - h(\boldsymbol{X}_j)| \le \max_{j=1,\ldots,n} \|\boldsymbol{g}(\boldsymbol{X}_j)\| \|\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g\| = O_p(1/\sqrt{m}),$$

following from $\|\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g\| = O_p(1/\sqrt{m})$ and the boundedness of $\boldsymbol{g}(\boldsymbol{X})$. By assumption (A5), $O_p(1/\sqrt{m}) = o_p(n^{-2/(p+4)})$, where $o_p(a_n) = a_n o_p(1)$. Consequently, by (A4),

$$T_5 = o_p(1).$$

From (S4.7)-(S4.8) in Lemma 3 and the Central Limit Theorem,

$$T_6 = n^{-1/2} b^{p/2} \sum_{i=1}^{n} \frac{\kappa_b(\boldsymbol{u} - \boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)^\top}{\widehat{f}_b(\boldsymbol{u})} (\boldsymbol{G}^\top \boldsymbol{G})^{-1} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{X}_i)\{h(\boldsymbol{X}_i) - \mu(\boldsymbol{U}_i)\}$$
$$= O_p(b^{p/2}) = o_p(1).$$

Combining these results, we obtain that $T_2 + \cdots + T_6 = B_{CK}(\boldsymbol{u}) + o_p(1)$. This establishes the result for $\widehat{\mu}_{CK}$.

For $\widehat{\mu}_{DK}$, the result follows from the fact that $\widehat{\mu}_{DK}$ equals to $\widehat{\mu}_{CK}$ with $\boldsymbol{g} = 0$.

As a technical note, we show that a large enough s ensures that Lemma 2 and (A4) hold. The condition $s > 2 + p/2$ as stated in (A1) is sufficient for

$$n^{1-2/s-\theta} b^p \to \infty, \qquad \text{where } \theta > 0 \text{ and } b = O_p(n^{-1/(p+4)}).$$

S2 Proof of Proposition 1

Using the notation in Section 2.1 and in the proof of Theorem 1 and defining $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$ and $\widehat{\boldsymbol{h}} = \boldsymbol{G}\widehat{\boldsymbol{\beta}}_g$, we obtain that

$$\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = \|(\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{Y} - \boldsymbol{\mu})\|^2 + \|\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{\mu})\|^2,$$

$$\|\widehat{\boldsymbol{\mu}}_K - \boldsymbol{\mu}\|^2 = \|(\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{Y} - \boldsymbol{\mu})\|^2 + \|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{Y} - \boldsymbol{\mu})\|^2$$

and, hence,

$$\|\widehat{\boldsymbol{\mu}}_K - \boldsymbol{\mu}\|^2 - \|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = \|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{Y} - \boldsymbol{\mu})\|^2 - \|\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{\mu})\|^2.$$

Let $\boldsymbol{h} = \boldsymbol{G}\boldsymbol{\beta}_g$, $\widehat{\boldsymbol{\Sigma}}_g = \boldsymbol{G}^\top\boldsymbol{G}/n$, and $\boldsymbol{q}(\boldsymbol{U}) = \{\boldsymbol{\beta}_g^\top\boldsymbol{g}(\boldsymbol{X}) - \mu(\boldsymbol{U})\}\boldsymbol{g}(\boldsymbol{X})$. Then

$$\|\boldsymbol{P}(\boldsymbol{h} - \boldsymbol{\mu})\|^2 = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{q}(\boldsymbol{U}_i)\right\}^\top \widehat{\boldsymbol{\Sigma}}_g^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{q}(\boldsymbol{U}_i)\right\}$$

From constraint (4), $\mathrm{E}\{\boldsymbol{q}(\boldsymbol{U})\} = 0$. By the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{q}(\boldsymbol{U}_i) \ \xrightarrow{d}\ N\left(0,\ \mathrm{E}\{\boldsymbol{q}(\boldsymbol{U})^\top\boldsymbol{q}(\boldsymbol{U})\}\right).$$

By the law of large numbers, $\widehat{\boldsymbol{\Sigma}}_g \xrightarrow{p} \boldsymbol{\Sigma}_g$. Hence,

$$\|\boldsymbol{P}(\boldsymbol{h} - \boldsymbol{\mu})\|^2 = O_p(1).$$

Since $\boldsymbol{h} = \boldsymbol{G}\boldsymbol{\beta}_g$ and $\widehat{\boldsymbol{h}} = \boldsymbol{G}\widehat{\boldsymbol{\beta}}_g$,

$$\|\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{h})\|^2 = (\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g)^\top\boldsymbol{G}^\top\boldsymbol{G}(\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g) = n(\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g)^\top\widehat{\boldsymbol{\Sigma}}_g(\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g)$$

As $\widehat{\boldsymbol{\beta}}_g$ is the least squares estimator, $\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g = O_p(m^{-1/2})$, which together with (A5) imply that $\|\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{h})\|^2 = O_p(1)$. Thus,

$$\|\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{\mu})\|^2 \leq 2\|\boldsymbol{P}(\boldsymbol{h} - \boldsymbol{\mu})\|^2 + 2\|\boldsymbol{P}(\widehat{\boldsymbol{h}} - \boldsymbol{h})\|^2 = O_p(1).$$

Let $\boldsymbol{A} = (A(\boldsymbol{U}_1), \ldots, A(\boldsymbol{U}_n))^\top$, where $A(\boldsymbol{u})$ is defined in (11). From Lemma 4 and (A4),

$$\frac{\|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\mu} - \boldsymbol{\mu})\|^2}{nb^4} - \frac{\boldsymbol{A}^\top\boldsymbol{P}\boldsymbol{A}}{n} \xrightarrow{p} 0.$$

By the law of large numbers,

$$\begin{aligned}\frac{\boldsymbol{A}^\top\boldsymbol{P}\boldsymbol{A}}{n} &= \left\{\frac{1}{n}\sum_{i=1}^n A(\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)^\top\right\}\left\{\frac{1}{n}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{X}_i)\boldsymbol{g}(\boldsymbol{X}_i)^\top\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^n A(\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)\right\} \\ &\xrightarrow{p} \mathrm{E}\{A(\boldsymbol{U})\boldsymbol{g}(\boldsymbol{X})^\top\}\boldsymbol{\Sigma}_g^{-1}\mathrm{E}\{A(\boldsymbol{U})\boldsymbol{g}(\boldsymbol{X})\}.\end{aligned}$$

As a result,

$$\frac{\|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\mu} - \boldsymbol{\mu})\|^2}{nb^4} \xrightarrow{p} \mathrm{E}\{A(\boldsymbol{U})\boldsymbol{g}(\boldsymbol{X})^\top\}\boldsymbol{\Sigma}_g^{-1}\mathrm{E}\{A(\boldsymbol{U})\boldsymbol{g}(\boldsymbol{X})\}.$$

From the identity

$$\begin{aligned}\|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{Y} - \boldsymbol{\mu})\|^2 &= \|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\mu} - \boldsymbol{\mu})\|^2 + \|\boldsymbol{P}\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\epsilon}\|^2 \\ &\quad + 2\{\boldsymbol{\epsilon}^\top\boldsymbol{\Delta}_b\boldsymbol{B}_b^{-1}\}\{\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\mu} - \boldsymbol{\mu})\}\end{aligned}$$

and the fact that

$$|\{\boldsymbol{\epsilon}^\top\boldsymbol{\Delta}_b\boldsymbol{B}_b^{-1}\}\{\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\mu} - \boldsymbol{\mu})\}| \leq \|\boldsymbol{P}(\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\mu} - \boldsymbol{\mu})\| \times \|\boldsymbol{P}\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\epsilon}\|,$$

the desired result follows if we can show that

$$\frac{\|\boldsymbol{P}\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\epsilon}\|^2}{nb^4} \xrightarrow{p} 0.$$

Letting $\widehat{\boldsymbol{g}}(\boldsymbol{x}) = \sum_{j=1}^n \kappa_b(\boldsymbol{U}_i - \boldsymbol{x})\boldsymbol{g}(\boldsymbol{X}_j)/\{n\widehat{f}_b(\boldsymbol{U}_j)\}$, we obtain that

$$\begin{aligned}\|\boldsymbol{P}\boldsymbol{B}_b^{-1}\boldsymbol{\Delta}_b\boldsymbol{\epsilon}\|^2 &= \left\{\sum_{i=1}^n \widehat{\boldsymbol{g}}(\boldsymbol{X}_i)\epsilon_i\right\}^\top (\boldsymbol{G}^\top\boldsymbol{G})^{-1}\left\{\sum_{i=1}^n \widehat{\boldsymbol{g}}(\boldsymbol{X}_i)\epsilon_i\right\} \\ &\le \frac{\tau_n}{n}\sum_{j=1}^n\sum_{i=1}^n \widehat{\boldsymbol{g}}(\boldsymbol{X}_i)^\top\widehat{\boldsymbol{g}}(\boldsymbol{X}_j)\epsilon_i\epsilon_j,\end{aligned}$$

where τ_n is the largest eigenvalue of the matrix $(\boldsymbol{G}^\top\boldsymbol{G}/n)^{-1}$. Since τ_n has order $O_p(1)$ from the assumed conditions and $nb^4 \to \infty$ under (A4), the result follows if

$$D = \frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n \widehat{\boldsymbol{g}}(\boldsymbol{X}_i)^\top\widehat{\boldsymbol{g}}(\boldsymbol{X}_j)\epsilon_i\epsilon_j = O_p(1).$$

Since $\widehat{\boldsymbol{g}}(\boldsymbol{X}_i)$'s are constants conditioned on $\boldsymbol{U}_1, ..., \boldsymbol{U}_n$,

$$\mathrm{E}(D|\boldsymbol{U}_1,\ldots,\boldsymbol{U}_n) = \frac{1}{n}\sum_{i=1}^n \widehat{\boldsymbol{g}}(\boldsymbol{X}_i)^\top\widehat{\boldsymbol{g}}(\boldsymbol{X}_i)\sigma^2(\boldsymbol{U}_i).$$

Under the assumed condition that f_U is bounded away from zero, (S4.7) and (S4.8) in Lemma 3 imply that $D = O_p(1)$.

S3 Proof of Proposition 2

Let $c \geq 0$ and $r > 0$ be the constants in (A4) and $\varphi_c(r)$ denote the AMISE of $\widehat{\mu}_{CK}$ for particular c and r. By (2.8),

$$\varphi_c(r) = c\tau_1 + c\tau_2 r^2(2 + r^2) + \tau_3 \int \left\{ \int \kappa(\boldsymbol{w} - r\boldsymbol{v})\kappa(\boldsymbol{v})d\boldsymbol{v} \right\}^2 d\boldsymbol{w},$$

where $\tau_1 = \mathrm{E}\{A(\boldsymbol{U})\}^2$, $\tau_2 = \mathrm{E}\{A(\boldsymbol{U}) - \boldsymbol{g}(\boldsymbol{X})^\top \boldsymbol{\Sigma}_g^{-1} \mathrm{E}\{\boldsymbol{g}(\boldsymbol{X})A(\boldsymbol{U})\}\}^2$, and $\tau_3 = \mathrm{E}\{\sigma^2(\boldsymbol{U})/f_U(\boldsymbol{U})\}$. The first and second order derivatives of this function with respect to r are

$$\varphi_c'(r) = 4c\tau_2(r+r^3)+\tau_3 \int \left\{ \int \kappa(\boldsymbol{w} - \boldsymbol{v}r)\kappa(\boldsymbol{v})d\boldsymbol{v} \right\} \left\{ \int -\nabla\kappa(\boldsymbol{w} - \boldsymbol{v}r)^\top \boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v} \right\} d\boldsymbol{w},$$

and

$$\begin{aligned}\varphi_c''(r) = 4c\tau_2(1 + 3r^2) &+ \tau_3 \int \left\{ \int \nabla\kappa(\boldsymbol{w} - \boldsymbol{v}r)^\top \boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v} \right\}^2 d\boldsymbol{w} \\ &+ \tau_3 \int \left\{ \int \kappa(\boldsymbol{w} - \boldsymbol{v}r)\kappa(\boldsymbol{v})d\boldsymbol{v} \right\} \left\{ \int \boldsymbol{v}^\top \nabla^2\kappa(\boldsymbol{w} - \boldsymbol{v}r)\boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v} \right\} d\boldsymbol{w}.\end{aligned}$$

Since the mean of $\kappa(\cdot)$ is 0,

$$\varphi_c'(0) = 0 \quad \text{and} \quad \varphi_c''(0) = 4c\tau_2 + \tau_3 \int \boldsymbol{v}^\top \left\{ \int \nabla^2\kappa(\boldsymbol{w})\kappa(\boldsymbol{w})d\boldsymbol{w} \right\} \boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v}.$$

Under the assumed condition that $\int \nabla^2\kappa(\boldsymbol{w})\kappa(\boldsymbol{w})d\boldsymbol{w}$ is negative definite, there exists a constant $c^* > 0$ such that for all $c < c^*$, $\varphi_c''(0) < 0$. This means that for $c < c^*$, $r = 0$ is a local maximum of $\varphi_c(r)$. Consequently, for all c and r in a neighborhood of 0,

$$\text{AMISE}(\widehat{\mu}_{CK})(r) < \text{AMISE}(\widehat{\mu}_K) = \varphi_c(0).$$

Let $\rho(r) = \int \left\{\int \kappa(\boldsymbol{w} - r\boldsymbol{v})\kappa(\boldsymbol{v})d\boldsymbol{v}\right\}^2 d\boldsymbol{w}$, then The first and second order derivatives of this function with respect to r are

$$\rho_c'(r) = \int \left\{\int \kappa(\boldsymbol{w} - \boldsymbol{v}r)\kappa(\boldsymbol{v})d\boldsymbol{v}\right\}\left\{\int -\nabla\kappa(\boldsymbol{w} - \boldsymbol{v}r)^\top \boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v}\right\}d\boldsymbol{w},$$

and

$$\begin{aligned}\rho_c''(r) = &\int \left\{\int \nabla\kappa(\boldsymbol{w} - \boldsymbol{v}r)^\top \boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v}\right\}^2 d\boldsymbol{w} \\ &+ \int \left\{\int \kappa(\boldsymbol{w} - \boldsymbol{v}r)\kappa(\boldsymbol{v})d\boldsymbol{v}\right\}\left\{\int \boldsymbol{v}^\top \nabla^2\kappa(\boldsymbol{w} - \boldsymbol{v}r)\boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v}\right\}d\boldsymbol{w}.\end{aligned}$$

Since the mean of $\kappa(\cdot)$ is 0,

$$\rho_c'(0) = 0 \quad \text{and} \quad \rho_c''(0) = \int \boldsymbol{v}^\top \left\{\int \nabla^2\kappa(\boldsymbol{w})\kappa(\boldsymbol{w})d\boldsymbol{w}\right\}\boldsymbol{v}\kappa(\boldsymbol{v})d\boldsymbol{v}.$$

Under the assumed condition that $\int \nabla^2\kappa(\boldsymbol{w})\kappa(\boldsymbol{w})d\boldsymbol{w}$ is negative definite, $\rho_c''(0)$ is negative. Let $c \geq 0$ and $r > 0$ be the constants in (A4) and $\varphi_c(r)$ denote the AMISE of $\widehat{\mu}_{CK}$ for particular c and r. By (2.8),

$$\text{AMISE}(\widehat{\mu}_{CK})(r) = \varphi_c(r) = c\tau_1 + c\tau_2 r^2(2 + r^2) + \tau_3\rho(r),$$

where $\tau_1 = \text{E}\{A(\boldsymbol{U})\}^2$, $\tau_2 = \text{E}\{A(\boldsymbol{U}) - \boldsymbol{g}(\boldsymbol{X})^\top \boldsymbol{\Sigma}_g^{-1}\text{E}\{\boldsymbol{g}(\boldsymbol{X})A(\boldsymbol{U})\}\}^2$, and $\tau_3 \qquad =$

$\mathrm{E}\{\sigma^2(\boldsymbol{U})/f_U(\boldsymbol{U})\}$.

$$\varphi_c(0) = c\tau_1 + \tau_3\rho(0)$$

$$\varphi_c(0) - \varphi_c(r) = \tau_3\{\rho(0) - \rho(r)\} - c\tau_2 r^2(2 + r^2)$$

Then, $\rho_c(0) - \rho_c(r) > 0$ if and only if

$$c < \frac{\tau_3\{\rho_c(0) - \rho_c(r)\}}{\tau_2 r^2(2 + r^2)}$$

From Taylor's expansion, we have

$$\rho_c(r) - \rho_c(0) = \rho'(0)r + \frac{r^2}{2}\int_0^1 (1-t)^2\rho''(tr)dt = \frac{r^2}{2}\int_0^1 (1-t)^2\rho''(tr)dt,$$

the second equality comes from $\rho'(0) = 0$.

And hence, $\rho_c(0) - \rho_c(r) > 0$ if and only if

$$c < -\frac{\tau_3\int_0^1 (1-t)^2\rho''(tr)dt}{2\tau_2(2 + r^2)}. \tag{S3.3}$$

Since $\rho''(0) < 0$ and the continuity, here $\rho''(r) < 0$ in a neighborhood of 0. As a result, for all r in such neighborhood of 0, and c satisfying (S3.3), we have

$$\mathrm{AMISE}(\widehat{\mu}_{CK})(r) < \mathrm{AMISE}(\widehat{\mu}_K) = \varphi_c(0).$$

Remark 1. Our proposed method CK has a smaller τ_2. And hence, in the right hand side of (S3.3), it has a larger upper bound for c. CK provide a more flexible choice of the bandwidthes.

Remark 2. Let $\kappa(\boldsymbol{z}) = (2\pi)^{-p/2} e^{-\|\boldsymbol{z}\|^2/2}$, then

$$\begin{aligned}
\rho(r) = & \int \left[\int (2\pi)^{-p} e^{-\frac{1}{2}\left(\|\boldsymbol{w}-r\boldsymbol{v}\|^2 - \|\boldsymbol{v}\|^2\right)} d\boldsymbol{v} \right]^2 d\boldsymbol{w} \\
= & \int \left[\int (2\pi)^{-p} e^{-\frac{1}{2}\left\{(1+r^2)\left\|\boldsymbol{v}-\frac{r}{1+r^2}\boldsymbol{w}\right\|^2 + \|\boldsymbol{w}\|^2/(1+r^2)\right\}} d\boldsymbol{v} \right]^2 d\boldsymbol{w} \\
= & \int e^{-\frac{1}{1+r^2}\|\boldsymbol{w}\|^2} \{(1+r^2)2\pi\}^{-p} \left[\int \left(\frac{1+r^2}{2\pi}\right)^{-p/2} e^{-\frac{1+r^2}{2}\left\|\boldsymbol{v}-\frac{r}{1+r^2}\boldsymbol{w}\right\|^2} d\boldsymbol{v} \right]^2 d\boldsymbol{w} \\
= & \, 2^{-p}\{\pi(1+r^2)\}^{-p/2} \int \{(1+r^2)\pi\}^{-p/2} e^{-\frac{1}{1+r^2}\|\boldsymbol{w}\|^2} d\boldsymbol{w} = 2^{-p}\{\pi(1+r^2)\}^{-p/2},
\end{aligned}$$

which is a decreasing function on r. Hence, $\rho(r) < \rho(0)$. So, the upper bound of c is

$$c < \frac{\tau_3\{1-(1+r^2)^{-n/2}\}}{\tau_2 2^p \pi^{p/2} r^2 (2+r^2)}.$$

S4 Some Lemmas

Lemma 1 (Bernstein's inequalities)**.** *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables such that $|X_i| \leq C$ (a constant) and $\bar{X}$ be the sample mean of X_i's. Then, for any $t > 0$,*

$$\mathrm{P}(|\bar{X} - \mathrm{E}(X_1)| > t) \leq 2\exp\left\{-\frac{nt^2/2}{\mathrm{Var}(X_1) + Ct/3}\right\}.$$

Lemma 2. *Suppose that the kernel κ is a bounded Lipschitz continuous density with mean zero and $\int \kappa(\boldsymbol{v})\|\boldsymbol{v}\|^2 d\boldsymbol{v} < \infty$, the bandwidth b satisfies*

$n^{1-2/s-\theta}b^p \to \infty$ *for some constants* $s > 0$ *and* $\theta > 0$, $\mathrm{E}|Y|^s < \infty$, *the covariate* $\boldsymbol{U}$ *has a compact support* $\mathbb{U} \subset \mathbb{R}^p$, *and* $\mathrm{E}(|Y|^s|\boldsymbol{U})f_U(\boldsymbol{U})$ *is bounded. Define*

$$\widehat{\nu}_n(\boldsymbol{u}) = \frac{1}{nb^p}\sum_{i=1}^{n}\kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right)Y_i$$

and

$$a_n = \sqrt{\frac{\log n}{nb^p}} \to 0.$$

Then

$$\sup_{\boldsymbol{u}\in\mathbb{U}} |\widehat{\nu}_n(\boldsymbol{u}) - \mathrm{E}\{\widehat{\nu}_n(\boldsymbol{u})\}| = O_p(a_n). \tag{S4.4}$$

If, in addition, $\nu(\boldsymbol{u}) = \mu(\boldsymbol{u})f_U(\boldsymbol{u})$ *has bounded second-order derivatives, then*

$$\sup_{\boldsymbol{u}\in\mathbb{U}} |\widehat{\nu}_n(\boldsymbol{u}) - \nu(\boldsymbol{u})| = O_p(a_n) + O_p(b^2). \tag{S4.5}$$

Proof. We modify the proof of Newey (1994) and Hansen (2008). First, we reduce the supremum in (S4.4) to a maximization over a finite J points grid by creating a grid using regions of the form $A_\ell = \{\boldsymbol{u} : \|\boldsymbol{u} - \boldsymbol{u}_j\| \le a_n b^{p+1}\}$. By putting $\boldsymbol{u}_j$ on a grid, the domain $\mathbb{U}$ can be covered with $J = O(b^{-p(p+1)}a_n^{-p})$ such regions. Note that

$$\sup_{\boldsymbol{u}\in\mathbb{U}} |\widehat{\nu}_n(\boldsymbol{u}) - \mathrm{E}\{\widehat{\nu}_n(\boldsymbol{u})\}| \le S_1 + S_2 + S_3,$$

where

$$S_1 = \sup_{\|\boldsymbol{u}-\boldsymbol{u}'\|\le a_n b^{p+1}} |\widehat{\nu}_n(\boldsymbol{u}) - \widehat{\nu}_n(\boldsymbol{u}')\}|$$

$$S_2 = \sup_{\|\boldsymbol{u}-\boldsymbol{u}'\|\le a_n b^{p+1}} |\mathrm{E}\{\widehat{\nu}_n(\boldsymbol{u}) - \widehat{\nu}_n(\boldsymbol{u}')\}|$$

$$S_3 = \sup_{j=1,\dots,J} |\widehat{\nu}_n(\boldsymbol{u}_j) - \mathrm{E}\{\widehat{\nu}_n(\boldsymbol{u}_j)\}|.$$

To establish (S4.4), we show that each $S_j = O_p(a_n)$. First,

$$\begin{aligned} S_1 &\le \sup_{\|\boldsymbol{u}-\boldsymbol{u}'\|\le a_n b^{p+1}} \left| \frac{1}{nb^p} \sum_{i=1}^n \left\{ \kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right) - \kappa\left(\frac{\boldsymbol{u}'-\boldsymbol{U}_i}{b}\right)\right\} Y_i \right| \\ &\le \sup_{\|\boldsymbol{u}-\boldsymbol{u}'\|\le a_n b^{p+1}} \frac{1}{nb^{p+1}} \sum_{i=1}^n L\|\boldsymbol{u}-\boldsymbol{u}'\||Y_i| \\ &\le a_n \frac{L}{n} \sum_{i=1}^n |Y_i| = O_p(a_n) \end{aligned}$$

for some constant $L > 0$, as κ is Lipschitz continuous. Similarly,

$$S_2 \le a_n L \mathrm{E}|Y_i| = O(a_n).$$

Let $1(\cdot)$ be the indicator function. Define

$$\widetilde{\nu}_n(\boldsymbol{u}) = \frac{1}{nb^p} \sum_{i=1}^n \kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right) Y_i 1(|Y_i| \le a_n^{-1}),$$

$$R_n(\boldsymbol{u}) = \widehat{\nu}_n(\boldsymbol{u}) - \widetilde{\nu}_n(\boldsymbol{u}) = \frac{1}{nb^p} \sum_{i=1}^n \kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right) Y_i 1(|Y_i| > a_n^{-1}).$$

Then

$$S_3 \le \max_{j=1,\dots,J} |R_n(\boldsymbol{u}_j)| + \max_{j=1,\dots,J} \mathrm{E}|R_n(\boldsymbol{u}_j)| + \max_{j=1,\dots,J} |\widetilde{\nu}_n(\boldsymbol{u}_j) - \mathrm{E}\{\widetilde{\nu}_n(\boldsymbol{u}_j)\}|. \tag{S4.6}$$

From the definition of $R_n(\boldsymbol{u})$,

$$\begin{aligned}
\mathrm{P}\left(\max_{j=1,\dots,J}|R_n(\boldsymbol{u}_j)|>0\right) &= \mathrm{P}\left(\max_{i=1,\dots,n}|Y_i|>a_n^{-1}\right) \\
&\le nP(|Y|>a_n^{-1}) \\
&= \mathrm{E}|Y|^s\left(\frac{\log n}{n^{1-2/s}b^p}\right)^{s/2} \\
&\to 0
\end{aligned}$$

Hence, the first term on the right side of (S4.6) equals to zero with probability going to one. By the conditions on κ and the boundedness of f_U,

$$\begin{aligned}
\mathrm{E}|R_n(\boldsymbol{u}^*)| =& \frac{1}{b^p}\int \kappa\left(\frac{\boldsymbol{u}^*-\boldsymbol{u}}{b}\right)|y|1(|y|>a_n^{-1})f(y|\boldsymbol{u})f_U(\boldsymbol{u})dyd\boldsymbol{u} \\
=& \int \kappa\left(\boldsymbol{v}\right)|y|1(|y|>a_n^{-1})f(y|\boldsymbol{u}^*-b\boldsymbol{v})f_U(\boldsymbol{u}^*-b\boldsymbol{v})dyd\boldsymbol{v} \\
\le& \int \kappa(\boldsymbol{v})d\boldsymbol{v}a_n^{s-1}\sup_{\boldsymbol{u}}\mathrm{E}(|Y|^s|\boldsymbol{U}=\boldsymbol{u})f_U(\boldsymbol{u}) = O(a_n^{s-1}).
\end{aligned}$$

Since this is an upper bound for all $\boldsymbol{u}^* \in \mathbb{U}$, the second term on the right side of (S4.6) is $O(a_n)$. By the conditions on κ, there are constants C_1 and C_2 such that

$$\sup_{\boldsymbol{u}\in\mathbb{U}}\frac{1}{b^p}\kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}}{b}\right)Y1(|Y|\le a_n^{-1}) \le \frac{C_1}{a_n b^p}$$

and

$$\begin{aligned}
\mathrm{Var}\left\{\frac{1}{b^p}\kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}}{b}\right)Y1(|Y|\le a_n^{-1})\right\} &\le \mathrm{E}\left\{\frac{1}{b^p}\kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}}{b}\right)Y1(|Y|\le a_n^{-1})\right\}^2 \\
&\le \frac{1}{b^{2p}}\int \kappa\left(\frac{\boldsymbol{u}-\boldsymbol{v}}{b}\right)^2 y^2 f(y|\boldsymbol{v})f_U(\boldsymbol{v})dyd\boldsymbol{v} \\
&\le \frac{1}{b^p}\sup_{\boldsymbol{u}\in\mathbb{U}}\mathrm{E}(|Y|^s+1|\boldsymbol{U}=\boldsymbol{u})f_U(\boldsymbol{u})\int \kappa(\boldsymbol{v})^2 d\boldsymbol{v} \\
&\le \frac{C_2}{b^p}.
\end{aligned}$$

Apply Lemma 1 with $t = ca_n$, we obtain that

$$\begin{aligned}
\mathrm{P}\left(\max_{j=1,\ldots,J}|\widetilde{\nu}_n(\boldsymbol{u}_j)-\mathrm{E}\{\widetilde{\nu}_n(\boldsymbol{u}_j)\}|>ca_n\right) &\le 2J\exp\left\{\frac{-nc^2a_n^2/2}{C_2/b^p+ca_nC_1\tau_n/3b^p}\right\} \\
&\le 2J\exp\left\{\frac{-c^2\log n}{2C_2+2cC_1/3}\right\} \\
&\le C^*\exp\left[\log n\left\{p+1-\frac{c^2}{2C_2+2cC_1/3}\right\}\right]
\end{aligned}$$

for some constant C^*. If c is large enough, the last term in the previous expression tends to 0. This proves that the third term on the right side of (S4.6) is $O(a_n)$. Combining the results, we obtain (S4.4).

From Taylor's expansion, there exists $\xi_v \in (0,1)$ such that

$$\begin{aligned}
|\mathrm{E}[\widehat{\nu}_n(\boldsymbol{u}^*)]-\nu(\boldsymbol{u}^*)| &= \left|\int \kappa(\boldsymbol{v})\{\nu(\boldsymbol{u}^*-b\boldsymbol{v})-\nu(\boldsymbol{u}^*)\}d\boldsymbol{v}\right| \\
&= \left|\int b\kappa(\boldsymbol{v})\nabla\nu(\boldsymbol{u}^*)^\top\boldsymbol{v}+\frac{b^2}{2}\kappa(\boldsymbol{v})\boldsymbol{v}^\top\nabla^2\nu(\boldsymbol{u}^*-b\xi_v\boldsymbol{v})^\top\boldsymbol{v}d\boldsymbol{v}\right| \\
&\le Lb^2\int \kappa(\boldsymbol{v})\|\boldsymbol{v}\|_2^2 d\boldsymbol{v} = O(b^2)
\end{aligned}$$

where the inequality holds since $\int \kappa(\boldsymbol{v})d\boldsymbol{v} = 0$ and $\nabla^2\nu(\boldsymbol{u})$ is bounded.

As this is a bound for all $\boldsymbol{u}$, (S4.5) holds. This completes the proof of this Lemma. $\square$

Lemma 3. *Under (A1)-(A4), each of the following terms is* $o_p(1)$ *:*

$$\sup_{\boldsymbol{u}\in\mathbb{U}} |\widehat{f}_U(\boldsymbol{u}) - f_U(\boldsymbol{u})|, \tag{S4.7}$$

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \left| \frac{1}{n}\sum_{i=1}^{n} \kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i) - f_U(\boldsymbol{u})\boldsymbol{g}(\boldsymbol{x}) \right|, \tag{S4.8}$$

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\boldsymbol{g}(\boldsymbol{X}_i)}{f_U(\boldsymbol{U}_i)} - \boldsymbol{g}(\boldsymbol{x}) \right|, \tag{S4.9}$$

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)(\boldsymbol{u}-\boldsymbol{U}_i)^\top \nabla^2\mu(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{U}_i)}{b^2} - f_U(\boldsymbol{u})\int \kappa(\boldsymbol{v})\boldsymbol{v}^\top \nabla^2 \mu(\boldsymbol{u})\boldsymbol{v}d\boldsymbol{v} \right|, \tag{S4.10}$$

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\nabla\mu(\boldsymbol{u})^\top(\boldsymbol{u}-\boldsymbol{U}_i)}{b^2} - \int \kappa(\boldsymbol{v})\nabla^\top \boldsymbol{v}\boldsymbol{v}^\top \nabla f_U(\boldsymbol{u})d\boldsymbol{v} \right|, \tag{S4.11}$$

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \left| \frac{1}{nl^2}\sum_{i=1}^{n} \kappa_l(u-\boldsymbol{U}_i)\|\boldsymbol{u}-\boldsymbol{U}_i\|^3 \right|. \tag{S4.12}$$

Proof. Note that (A1)-(A4) ensure the conditions on Lemma 2. With Y being 1 or $g(\boldsymbol{X})$, (S4.5) in Lemma 2 implies that the terms in (S4.7) and (S4.8) are $o_p(1)$. Since f_U is bounded away from zero, this also implies that the term in (S4.9) is $o_p(1)$.

For terms in (S4.10)-(S4.12), we cannot apply directly Lemma 2, but an argument similar to that in the proof of Lemma 2 leads to the desired results.

By the boundedness of κ and $\boldsymbol{U}$, (A1), and (A3), there are constants C_1 and C_2 such that

$$\left|\frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)(\boldsymbol{u}-\boldsymbol{U}_i)^\top\nabla^2\mu(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{U}_i)}{b^2}\right| \le \frac{C_1}{b^{p+2}}$$

and

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)(\boldsymbol{u}-\boldsymbol{U}_i)^\top\nabla^2\mu(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{U}_i)}{b^2}\right\}\\
&\le \frac{1}{b^p}\mathrm{E}\left[\frac{1}{b^p}\kappa\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right)^2\left\{\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right)^\top\nabla^2\mu(\boldsymbol{u})\left(\frac{\boldsymbol{u}-\boldsymbol{U}_i}{b}\right)\right\}^2\right]\\
&= \frac{1}{b^p}\int \kappa(\boldsymbol{v})^2\{\boldsymbol{v}^\top\nabla^2\mu(\boldsymbol{u})\boldsymbol{v}\}^2 f_U(\boldsymbol{u}-b\boldsymbol{v})d\boldsymbol{v}\\
&\le \frac{C_2}{b^p}
\end{aligned}$$

Then, applying Lemma 1 with the above two bounds and defining

$$D_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^n \frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\,(\boldsymbol{u}-\boldsymbol{U}_i)^\top\,\nabla^2\mu(\boldsymbol{u})\,(\boldsymbol{u}-\boldsymbol{U}_i)}{b^2},$$

we obtain that

$$\mathrm{P}\left(\sup_{j\in 1\ldots,J}|D_n(\boldsymbol{u}_j)-\mathrm{E}\{D_n(\boldsymbol{u}_j)\}|>t\right)\le 2J\exp\left\{-\frac{nt^2/2}{C_2/b^p+tC_1/(3b^{p+2})}\right\},$$

which goes to 0 under (A4). From the Lipschitz continuity of f_U, the boundedness of κ's third moment, and the boundedness of second gradient of μ,

$$\begin{aligned}
&\sup_{\boldsymbol{u}\in\mathbb{U}}\left|f_U(\boldsymbol{u})\int\kappa(\boldsymbol{v})\boldsymbol{v}^\top\nabla^2\mu(\boldsymbol{u})\boldsymbol{v}d\boldsymbol{v}-\mathrm{E}\left\{D_n(\boldsymbol{u})\right\}\right|\\
&=\sup_{\boldsymbol{u}\in\mathbb{U}}\left|f_U(\boldsymbol{u})\int\kappa(\boldsymbol{v})\boldsymbol{v}^\top\nabla^2\mu(\boldsymbol{u})\boldsymbol{v}d\boldsymbol{v}-\int\kappa(\boldsymbol{v})\boldsymbol{v}^\top\nabla^2\mu(\boldsymbol{u})\boldsymbol{v}f_U(\boldsymbol{u}-b\boldsymbol{v})d\boldsymbol{v}\right|\\
&\le bL\sup_{\boldsymbol{u}\in\mathbb{U}}\|\nabla^2\mu(\boldsymbol{u})\|\int\kappa(\boldsymbol{v})\|\boldsymbol{v}\|^3d\boldsymbol{v}\\
&=o(1)
\end{aligned}$$

since $b\to 0$. Considering grid of regions A_ℓ described in the proof of Lemma 2, these results show that the term in (S4.10) is $o_p(1)$.

Consider (S4.11). Following the previous proof with $\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)(\boldsymbol{u}-\boldsymbol{U}_i)^\top\nabla^2\mu(\boldsymbol{u})(\boldsymbol{u}-\boldsymbol{U}_i)$ replaced by $\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\nabla\mu(\boldsymbol{u})^\top(\boldsymbol{u}-\boldsymbol{U}_i)$, we can establish that

$$\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\kappa_b(\boldsymbol{u}_j-\boldsymbol{U}_i)\nabla\mu(\boldsymbol{u}_j)^\top(\boldsymbol{u}_j-\boldsymbol{U}_i)}{b^2}-\mathrm{E}\left\{\frac{\kappa_b(\boldsymbol{u}_j-\boldsymbol{U}_i)\nabla\mu(\boldsymbol{u}_j)^\top(\boldsymbol{u}_j-\boldsymbol{U}_i)}{b^2}\right\}\right|\xrightarrow{p}0$$

From Taylor's expansion and the boundedness of the second derivative of density f_U and κ's third moment, we obtain that

$$\begin{aligned}
&\left|\mathrm{E}\left\{\frac{\kappa_b(\boldsymbol{u}-\boldsymbol{U}_i)\nabla\mu(\boldsymbol{u})^\top(\boldsymbol{u}-\boldsymbol{U}_i)}{b^2}\right\}-\int\kappa(\boldsymbol{v})\nabla\mu(\boldsymbol{u})^\top\boldsymbol{v}\boldsymbol{v}^\top\nabla f_U(\boldsymbol{u})d\boldsymbol{v}\right| \\
&=\left|\frac{1}{b}\int\kappa(\boldsymbol{v})\nabla\mu(\boldsymbol{u})^\top\boldsymbol{v}f_U(\boldsymbol{u}-b\boldsymbol{v})d\boldsymbol{v}-\int\kappa(\boldsymbol{v})\nabla\mu(\boldsymbol{u})^\top\boldsymbol{v}\boldsymbol{v}^\top\nabla f_U(\boldsymbol{u})d\boldsymbol{v}\right| \\
&=\left|\frac{1}{b}f_U(\boldsymbol{u})\int\kappa(\boldsymbol{v})\nabla\mu(\boldsymbol{u})^\top\boldsymbol{v}d\boldsymbol{v}+b\int\kappa(\boldsymbol{v})\nabla\mu(\boldsymbol{u})^\top\boldsymbol{v}\boldsymbol{v}^\top\nabla^2 f_U(\boldsymbol{\xi})\boldsymbol{v}d\boldsymbol{v}\right| \\
&=\left|b\int\kappa(\boldsymbol{v})\nabla\mu(\boldsymbol{u})^\top\boldsymbol{v}\boldsymbol{v}^\top\nabla^2 f_U(\boldsymbol{\xi})\boldsymbol{v}d\boldsymbol{v}\right| \\
&\leq b\sup_{\boldsymbol{u}\in\mathbb{U}}\|\nabla^2 f_U(\boldsymbol{u})\|\sup_{\boldsymbol{u}\in\mathbb{U}}\|\nabla\mu(\boldsymbol{u})\|\int\kappa(\boldsymbol{v})\|\boldsymbol{v}\|^3d\boldsymbol{v} \\
&=o(1),
\end{aligned}$$

where the second equality follows from Taylor's expansion of f_U at $\boldsymbol{u}$ and the third equality follows from the fact that κ has mean zero. This proves that the term in (S4.11) is $o_p(1)$.

For (S4.12), with the same argument we can show that

$$\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\frac{1}{nl^3}\sum_{i=1}^n\kappa_l(\boldsymbol{u}_j-\boldsymbol{U}_i)\|\boldsymbol{u}_j-\boldsymbol{U}_i\|^3-\mathrm{E}\left\{\frac{1}{l^3}\kappa_l(\boldsymbol{u}_j-\boldsymbol{U}_i)\|\boldsymbol{u}_j-\boldsymbol{U}_i\|^3\right\}\right|\xrightarrow{p}0.$$

Note that

$$\begin{aligned}
\sup_{\boldsymbol{u}\in\mathbb{U}}\mathrm{E}\left\{\frac{1}{l^3}\kappa_l(u-\boldsymbol{U}_i)\|\boldsymbol{u}-\boldsymbol{U}_i\|^3\right\}&=\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\int\kappa(\boldsymbol{v})\|\boldsymbol{v}\|_2^3f_U(\boldsymbol{u}-b\boldsymbol{v})d\boldsymbol{v}\right| \\
&\leq\sup_{\boldsymbol{u}\in\mathbb{U}}f_U(\boldsymbol{u})\int\kappa(\boldsymbol{v})\|\boldsymbol{v}\|^3d\boldsymbol{v}.
\end{aligned}$$

Since $l\to0$, this proves that the term in (S4.12) is $o_p(1)$.

□

Lemma 4. *Under (A1)-(A4),*

$$\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\frac{1}{nl^2\widehat{f}_U(\boldsymbol{u})}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{u})\}-A(\boldsymbol{u})\right|=o_p(1)$$

Proof. Applying Taylor's expansion at $\boldsymbol{u}$, we obtain that

$$\frac{1}{nl^2 f_U(\boldsymbol{u})}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{u})\}=S_4(\boldsymbol{u})+S_5(\boldsymbol{u})+S_6(\boldsymbol{u}),$$

where

$$\begin{aligned}
S_4(\boldsymbol{u}) &= \frac{1}{nl^2 f_U(\boldsymbol{u})}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\nabla\mu(\boldsymbol{u})^\top(\boldsymbol{U}_i-\boldsymbol{u}),\\
S_5(\boldsymbol{u}) &= \frac{1}{2nl^2 f_U(\boldsymbol{u})}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)(\boldsymbol{U}_i-\boldsymbol{u})^\top\nabla^2\mu(\boldsymbol{u})(\boldsymbol{U}_i-\boldsymbol{u}),\\
S_6(\boldsymbol{u}) &= \frac{1}{nl^2 f_U(\boldsymbol{u})}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)R(\boldsymbol{U}_i-\boldsymbol{u}),
\end{aligned}$$

and $R(\boldsymbol{U}_i-\boldsymbol{u})$ is the residual term in Taylor's expansion, which is bounded in absolute value by $M\|\boldsymbol{U}_i-\boldsymbol{u}\|^3$ for some constant M under (A2). Since f_U is bounded away from zero, the results in Lemma 3 about the terms in (S4.10)-(S4.12) imply that $S_4(\boldsymbol{u})+S_5(\boldsymbol{u})-A(\boldsymbol{u})\xrightarrow{p}0$ and $S_6(\boldsymbol{u})\xrightarrow{p}0$ uniformly for $\boldsymbol{u}\in\mathbb{U}$. Hence,

$$\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\frac{1}{nl^2 f_U(\boldsymbol{u})}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{u})\}-A(\boldsymbol{u})\right|=o_p(1).$$

Then, the result follows from

$$
\begin{aligned}
&\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\left[\frac{1}{nl^2}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{u})\}\right]\left\{\frac{1}{\widehat{f}_U(\boldsymbol{u})}-\frac{1}{f_U(\boldsymbol{u})}\right\}\right|\\
&\leq \sup_{\boldsymbol{u}\in\mathbb{U}}\left|\frac{1}{nl^2}\sum_{i=1}^{n}\kappa_l(\boldsymbol{u}-\boldsymbol{U}_i)\{\mu(\boldsymbol{U}_i)-\mu(\boldsymbol{u})\}\right|\left|\frac{1}{\widehat{f}_U(\boldsymbol{u})}-\frac{1}{f_U(\boldsymbol{u})}\right|\\
&= O_p(1)\sup_{\boldsymbol{u}\in\mathbb{U}}\left|\frac{1}{\widehat{f}_U(\boldsymbol{u})}-\frac{1}{f_U(\boldsymbol{u})}\right|\\
&= O_p\left(\sqrt{\frac{\log n}{nl^p}}+l^2\right)\\
&= o_p(1),
\end{aligned}
$$

where the second equality follows from the result in Lemma 2 about the term in (S4.7) and the fact that f_U is bounded away from zero, and the last equality follows from (A4). □

Lemma 5. *Under the conditions in Theorem 3,*

$$
\sup_{\boldsymbol{x}\in\mathbb{X}}|\widehat{g}^*(\boldsymbol{x})-g^*(\boldsymbol{x})| = o_p(n^{-2/(p+4)}).
$$

Proof. Using the notation in Section 3, we define

$$
\widehat{\boldsymbol{\theta}} = (\widehat{f}_U, \widehat{\nu}_0, \widehat{\nu}_1 \ldots \widehat{\nu}_p, \nabla^2_{11}\widehat{f}_U, \ldots, \nabla^2_{pp}\widehat{f}_U)
$$

and

$$
\boldsymbol{\theta} = (f_U, \nu_0, \nu_1 \ldots, \nu_p, \nabla_{11} f_U, \ldots, \nabla_{pp} f_U).
$$

Then, there is a function η such that $A(\boldsymbol{u}) = \eta(\boldsymbol{\theta}(\boldsymbol{u}))$ and $\widehat{A} = \eta(\widehat{\boldsymbol{\theta}}(\boldsymbol{u}))$.

Under (B1)-(B3), Lemma 8.10 in Newey and McFadden (1994) shows that

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \|\widehat{\boldsymbol{\theta}}(\boldsymbol{u}) - \boldsymbol{\theta}(\boldsymbol{u})\| = O_p(\sqrt{\log n}\, n^{-\varsigma/(p+4+2\varsigma)}).$$

Since $\varsigma > 2 + 8/p$, this implies that

$$\sup_{\boldsymbol{u}\in\mathbb{U}} \|\widehat{\boldsymbol{\theta}}(\boldsymbol{u}) - \boldsymbol{\theta}(\boldsymbol{u})\| = o_p(n^{-2/(p+4)}). \tag{S4.13}$$

From the mean value theorem, we have

$$\widehat{A}(\boldsymbol{u}) - A(\boldsymbol{u}) = \eta(\widehat{\boldsymbol{\theta}}(\boldsymbol{u})) - \eta(\boldsymbol{\theta}(\boldsymbol{u})) = \nabla\eta(\boldsymbol{\xi})^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\boldsymbol{u}),$$

for some $\boldsymbol{\xi}$ in the segment between $\widehat{\boldsymbol{\theta}}(\boldsymbol{u})$ and $\boldsymbol{\theta}(\boldsymbol{u})$. Since f_U is bounded away from zero, $\nabla\eta$ is bounded on a neighborhood of $\boldsymbol{\theta}$. This, together with (S4.13) imply that

$$\sup_{\boldsymbol{u}\in\mathbb{U}} |\widehat{A}(\boldsymbol{u}) - A(\boldsymbol{u})| = o_p(n^{-2/(p+4)}). \tag{S4.14}$$

Note that

$$\begin{aligned}\sup_{\boldsymbol{x}\in\mathbb{X}} |\widehat{g}^*(\boldsymbol{x}) - g^*(\boldsymbol{x})| \leq \sup_{\boldsymbol{x}\in\mathbb{X}} &\left| \frac{\sum_{i=1}^n \kappa_{\lambda_2}(\boldsymbol{x} - \boldsymbol{X}_i)\{\widehat{A}(\boldsymbol{U}_i) - A(\boldsymbol{U}_i)\}}{\sum_{i=1}^n \kappa_{\lambda_2}(\boldsymbol{x} - \boldsymbol{X}_i)} \right| \\ &+ \sup_{\boldsymbol{x}\in\mathbb{X}} \left| \frac{\sum_{i=1}^n \kappa_{\lambda_2}(\boldsymbol{x} - \boldsymbol{X}_i)\{A(\boldsymbol{U}_i) - g^*(\boldsymbol{x})\}}{\sum_{i=1}^n \kappa_{\lambda_2}(\boldsymbol{x} - \boldsymbol{X}_i)} \right|\end{aligned}$$

The first term on the right side of the previous inequality is bounded by the quantity in (S4.14), which is $o_p(n^{-2/(p+4)})$. Applying the results in Lemma 2 with $Y_i = A(\boldsymbol{U}_i) - g^*(\boldsymbol{x})$ shows that the second term on the right side of the previous inequality is $O_p(\sqrt{\log n}\, n^{-2/(q+4)})$, which is$o_p(n^{-2/(p+4)})$ as $q < p$. This proves the result. $\square$

S5 Bias tables

Table A1: Simulated Bias (3.2) and IMP (3.3) with $S = 500$ under setting S1

					$\widehat{\mu}_{CK}$ in (2.7) with constraint (2.5) $\boldsymbol{g} =$			
Covariate	Model	Test data	b, l	$\widehat{\mu}_K$ (2.2)	1	$(1, X)$	$(1, \widehat{h})$	$(1, X, \widehat{h})$
Normal	M1	Sample	Best	0.018	0.014	0.014	0.016	0.020
			CV	0.022	0.014	0.008	0.012	0.017
		Grid	Best	-0.003	-0.024	-0.074	-0.058	-0.035
			CV	-0.017	-0.023	-0.041	-0.030	-0.002
	M2	Sample	Best	0.016	-0.001	0.016	0.018	0.019
			CV	0.020	0.012	0.012	0.012	0.013
		Grid	Best	0.020	0.021	0.019	0.038	0.046
			CV	0.022	0.014	0.012	0.035	0.054
	M3	Sample	Best	0.012	0.013	0.013	0.017	0.020
			CV	0.021	0.007	0.008	0.011	0.013
		Grid	Best	0.007	0.008	0.007	0.026	0.050
			CV	0.009	-0.005	-0.004	0.013	0.040
	M4	Sample	Best	0.026	0.018	0.017	0.020	0.024
			CV	0.035	0.018	0.016	0.021	0.026
		Grid	Best	-0.002	-0.010	-0.038	-0.069	-0.042
			CV	-0.001	-0.020	-0.031	-0.015	0.015
Bounded	M1	Sample	Best	0.004	-0.000	0.001	0.002	0.004
			CV	0.014	0.005	0.001	0.002	0.004
		Grid	Best	0.039	0.043	0.044	0.037	0.036
			CV	0.046	0.037	0.038	0.034	0.029
	M2	Sample	Best	0.010	0.010	0.010	0.011	0.014
			CV	0.019	0.010	0.009	0.011	0.011
		Grid	Best	0.117	0.082	0.081	0.069	0.055
			CV	0.083	0.076	0.076	0.071	0.053
	M3	Sample	Best	0.014	0.006	0.005	0.005	0.009
			CV	0.009	0.003	0.002	0.003	0.007
		Grid	Best	0.071	0.053	0.052	0.049	0.033
			CV	0.060	0.055	0.054	0.052	0.035
	M4	Sample	Best	0.009	-0.001	-0.001	0.001	0.002
			CV	0.008	0.005	0.002	0.004	0.006
		Grid	Best	0.053	0.053	0.053	0.043	0.040
			CV	0.050	0.048	0.050	0.046	0.042

Table A2: Simulated Bias (3.2) and IMP (3.3) with $S = 500$ under setting S2

					$\widehat{\mu}_{CK}$ in (2.7) with constraint	
Covariate	Model	Test data	b, l	$\widehat{\mu}_K$ in (2.2)	(2.5)	(2.17)
Normal	M1	Sample	Best	0.020	0.015	0.014
			CV	0.020	0.011	0.011
		Grid	Best	-0.002	-0.074	-0.075
			CV	-0.015	-0.034	-0.034
	M2	Sample	Best	0.021	0.022	0.022
			CV	0.019	0.013	0.013
		Grid	Best	0.021	0.022	0.022
			CV	0.020	0.014	0.014
	M3	Sample	Best	0.022	0.023	0.016
			CV	0.016	0.015	0.015
		Grid	Best	0.016	0.016	0.007
			CV	0.005	0.004	0.003
	M4	Sample	Best	0.023	0.018	0.014
			CV	0.037	0.018	0.017
		Grid	Best	-0.005	-0.035	-0.039
			CV	-0.001	-0.029	-0.032
Bounded	M1	Sample	Best	0.009	0.000	-0.001
			CV	0.009	0.003	0.002
		Grid	Best	0.045	0.044	0.043
			CV	0.043	0.042	0.041
	M2	Sample	Best	0.017	0.014	0.011
			CV	0.018	0.013	0.009
		Grid	Best	0.122	0.083	0.080
			CV	0.087	0.084	0.079
	M3	Sample	Best	0.013	0.003	0.008
			CV	0.013	0.003	0.000
		Grid	Best	0.070	0.050	0.055
			CV	0.060	0.052	0.049
	M4	Sample	Best	0.013	0.002	0.000
			CV	0.007	0.001	0.000
		Grid	Best	0.054	0.056	0.055
			CV	0.048	0.049	0.048

For CK estimator under all constraints, $\boldsymbol{g}(X) = (1, X)$.

Bibliography

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

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory 10*, 233–253.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Shao, J. (2003). *Mathematical Statistics* (2nd ed.). Springer, New York.