

## 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:

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

where

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

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

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

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

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

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

$\widehat{f}_b(\mathbf{u}) = \sum_{i=1}^n \kappa_b(\mathbf{u} - \mathbf{U}_i)/n$ ,  $\boldsymbol{\delta}_b(\mathbf{u}) = (\kappa_b(\mathbf{u} - \mathbf{U}_1), \dots, \kappa_b(\mathbf{u} - \mathbf{U}_n))^\top$ ,  $\mathbf{I}_n$  is the identity matrix of order  $n$ ,  $\mathbf{P} = \mathbf{G}(\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top$  with  $\mathbf{G}$  defined in (2.6),  $\mathbf{B}_l$  is the  $n \times n$  diagonal matrix whose  $i$ th diagonal element is  $\widehat{f}_l(\mathbf{U}_i)$ ,  $\boldsymbol{\Delta}_l$  is the  $n \times n$  matrix whose  $(i, j)$ th entry is  $\kappa_l(\mathbf{U}_i - \mathbf{U}_j)/n$ ,  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$  with  $\epsilon_i = Y_i - \mu(\mathbf{U}_i)$ ,  $\tilde{\boldsymbol{\mu}} = (\mu(\mathbf{U}_1), \dots, \mu(\mathbf{U}_n))^\top$ ,  $\mathbf{1}_n$  is the  $n$ -vector with all components being 1,  $\mathbf{h} = \mathbf{G}\boldsymbol{\beta}_g$  with  $\boldsymbol{\beta}_g$  defined in (2.3), and  $\widehat{\mathbf{h}} = \mathbf{G}\widehat{\boldsymbol{\beta}}_g$  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  $V_{CK}(\mathbf{u})$  defined in (2.10). Define

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

Consider a further decomposition of  $T_1$ :

$$T_1 = \sqrt{n}V + T_{11} + T_{12} + T_{13}, \quad (\text{S1.2})$$

where

$$\begin{aligned} V &= \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n S(\mathbf{U}_i, \epsilon_i, \mathbf{U}_j, \epsilon_j), \\ T_{11} &= \frac{b^{p/2}}{n^{3/2}} \sum_{i=1}^n \frac{\kappa_b(\mathbf{u} - \mathbf{U}_i)\kappa_l(0)\epsilon_i}{f_U(\mathbf{u})f_U(\mathbf{U}_i)}, \\ T_{12} &= \frac{b^{p/2}}{n^{3/2}} \sum_{j=1}^n \sum_{i=1}^n \frac{\kappa_b(\mathbf{u} - \mathbf{U}_i)\kappa_l(\mathbf{U}_i - \mathbf{U}_j)}{f_U(\mathbf{u})f_U(\mathbf{U}_i)} \left\{ \frac{f_U(\mathbf{u})f_U(\mathbf{U}_i)}{\widehat{f}_b(\mathbf{u})\widehat{f}_l(\mathbf{U}_i)} - 1 \right\} \epsilon_j, \\ T_{13} &= -n^{-1/2}b^{p/2}\boldsymbol{\delta}_b(\mathbf{u})^\top \mathbf{P}\mathbf{B}_l^{-1}\boldsymbol{\Delta}_l\boldsymbol{\epsilon}/\widehat{f}_b(\mathbf{u}). \end{aligned}$$

The decomposition (S1.2) follows from

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

and

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

Note that  $V$  is a V-statistic with

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

which has variance

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

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

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

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

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

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  $E(T_{12} | \mathbf{U}_1, \dots, \mathbf{U}_n) = 0$  and, from

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

$\text{Var}(T_{12} | \mathbf{U}_1, \dots, \mathbf{U}_n)$  is bounded by

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

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

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

Then

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

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

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

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

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

As a result,  $\text{Var}(T_{13} \mid \mathbf{U}_1, \dots, \mathbf{U}_n) = O_p(b^p) = o_p(1)$ . 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  $T_2 + \dots + T_6$  in (S1.1) converges to  $B_{CK}(\mathbf{u})$  in probability. From Lemma 4 and (A4),

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

Note that

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

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

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

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,

$$T_4 = -\sqrt{cr^2} \mathbf{g}(\mathbf{x})^\top \Sigma_g^{-1} \mathbb{E}\{\mathbf{g}(\mathbf{X}) A(\mathbf{U})\} \{1 + o_p(1)\}.$$

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

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

where  $\widehat{h}(\mathbf{X}_i)$  and  $h(\mathbf{X}_i)$  are the  $i$ th components of  $\widehat{\mathbf{h}}$  and  $\mathbf{h}$ , respectively, the first equality follows from (S4.7)-(S4.8), and the inequality holds for a non-random constant  $M$ . Since

$$\widehat{h}(\mathbf{X}_j) - h(\mathbf{X}_j) = \mathbf{g}(\mathbf{X}_j)^\top (\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g),$$

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

following from  $\|\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}_g\| = O_p(1/\sqrt{m})$  and the boundedness of  $\mathbf{g}(\mathbf{X})$ . 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,

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

Combining these results, we obtain that  $T_2 + \dots + T_6 = B_{CK}(\mathbf{u}) + o_p(1)$ . 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  $\mathbf{g} = 0$ .

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

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

## S2 Proof of Proposition 1

Using the notation in Section 2.1 and in the proof of Theorem 1 and defining  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$

and  $\hat{\mathbf{h}} = \mathbf{G}\hat{\boldsymbol{\beta}}_g$ , we obtain that

$$\begin{aligned}\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 &= \|(I - \mathbf{P})(\mathbf{B}_b^{-1}\Delta_b\mathbf{Y} - \boldsymbol{\mu})\|^2 + \|\mathbf{P}(\hat{\mathbf{h}} - \boldsymbol{\mu})\|^2, \\ \|\hat{\boldsymbol{\mu}}_K - \boldsymbol{\mu}\|^2 &= \|(I - \mathbf{P})(\mathbf{B}_b^{-1}\Delta_b\mathbf{Y} - \boldsymbol{\mu})\|^2 + \|\mathbf{P}(\mathbf{B}_b^{-1}\Delta_b\mathbf{Y} - \boldsymbol{\mu})\|^2\end{aligned}$$

and, hence,

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

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

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

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

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

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

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

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

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



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

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

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

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

By the law of large numbers,

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

As a result,

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

From the identity

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

and the fact that

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

the desired result follows if we can show that

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

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

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

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

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

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

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

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),

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

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

are

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

and

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

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

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

Under the assumed condition that  $\int \nabla^2\kappa(\mathbf{w})\kappa(\mathbf{w})d\mathbf{w}$  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,

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

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

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

and

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

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

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

Under the assumed condition that  $\int \nabla^2 \kappa(\mathbf{w}) \kappa(\mathbf{w}) d\mathbf{w}$  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  $\hat{\mu}_{CK}$  for particular  $c$  and  $r$ . By (2.8),

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

where  $\tau_1 = \text{E}\{A(\mathbf{U})\}^2$ ,  $\tau_2 = \text{E}\{A(\mathbf{U}) - \mathbf{g}(\mathbf{X})^\top \Sigma_g^{-1} \text{E}\{\mathbf{g}(\mathbf{X})A(\mathbf{U})\}\}^2$ , and  $\tau_3 = \text{E}\{\sigma^2(\mathbf{U})/f_U(\mathbf{U})\}$ .

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

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

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

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

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

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

**Remark 1.** Our proposed method CK has a smaller  $\tau_2$ . 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 bandwidths.

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

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

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  $X_1, \dots, X_n$  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  $t > 0$ ,*

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

**Lemma 2.** *Suppose that the kernel  $\kappa$  is a bounded Lipschitz continuous density with mean zero and  $\int \kappa(\mathbf{v}) \|\mathbf{v}\|^2 d\mathbf{v} < \infty$ , the bandwidth  $b$  satisfies  $n^{1-2/s-\theta} b^p \rightarrow \infty$  for some constants  $s > 0$  and  $\theta > 0$ ,  $E|Y|^s < \infty$ , the covariate  $\mathbf{U}$  has a compact support  $\mathbb{U} \subset \mathbb{R}^p$ , and  $E(|Y|^s | \mathbf{U}) f_U(\mathbf{U})$  is bounded. Define*

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

and

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

Then

$$\sup_{\mathbf{u} \in \mathbb{U}} |\hat{\nu}_n(\mathbf{u}) - E\{\hat{\nu}_n(\mathbf{u})\}| = O_p(a_n). \quad (\text{S4.4})$$

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

$$\sup_{\mathbf{u} \in \mathbb{U}} |\hat{\nu}_n(\mathbf{u}) - \nu(\mathbf{u})| = O_p(a_n) + O_p(b^2). \quad (\text{S4.5})$$

*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  $A_\ell = \{\mathbf{u} : \|\mathbf{u} - \mathbf{u}_j\| \leq a_n b^{p+1}\}$ . By putting  $\mathbf{u}_j$  on a grid, the domain  $\mathbb{U}$  can be covered with  $J = O(b^{-p(p+1)} a_n^{-p})$  such regions. Note that

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

where

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

$$S_2 = \sup_{\|\mathbf{u} - \mathbf{u}'\| \leq a_n b^{p+1}} |\mathbb{E}\{\widehat{\nu}_n(\mathbf{u}) - \widehat{\nu}_n(\mathbf{u}')\}|$$

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

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

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

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

$$S_2 \leq a_n L \mathbb{E}|Y_i| = O(a_n).$$

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

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

Then

$$S_3 \leq \max_{j=1,\dots,J} |R_n(\mathbf{u}_j)| + \max_{j=1,\dots,J} \mathbb{E}|R_n(\mathbf{u}_j)| + \max_{j=1,\dots,J} |\tilde{\nu}_n(\mathbf{u}_j) - \mathbb{E}\{\tilde{\nu}_n(\mathbf{u}_j)\}|. \quad (\text{S4.6})$$

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

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

Hence, the first term on the right side of (S4.6) equals to zero with probability going to one.

By the conditions on  $\kappa$  and the boundedness of  $f_U$ ,

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

Since this is an upper bound for all  $\mathbf{u}^* \in \mathbb{U}$ , the second term on the right side of (S4.6) is

$O(a_n)$ . By the conditions on  $\kappa$ , there are constants  $C_1$  and  $C_2$  such that

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



and

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

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

$$\begin{aligned}
\mathbb{P} \left( \max_{j=1, \dots, J} |\tilde{\nu}_n(\mathbf{u}_j) - \mathbb{E}\{\tilde{\nu}_n(\mathbf{u}_j)\}| > ca_n \right) &\leq 2J \exp \left\{ \frac{-nc^2 a_n^2/2}{C_2/b^p + ca_n C_1 \tau_n / 3b^p} \right\} \\
&\leq 2J \exp \left\{ \frac{-c^2 \log n}{2C_2 + 2cC_1/3} \right\} \\
&\leq C^* \exp \left[ \log n \left\{ p + 1 - \frac{c^2}{2C_2 + 2cC_1/3} \right\} \right]
\end{aligned}$$

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

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

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

for all  $\mathbf{u}$ , (S4.5) holds. This completes the proof of this Lemma.  $\square$

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

$$\sup_{\mathbf{u} \in \mathbb{U}} |\hat{f}_U(\mathbf{u}) - f_U(\mathbf{u})|, \quad (\text{S4.7})$$

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

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

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

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

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

*Proof.* Note that (A1)-(A4) ensure the conditions on Lemma 2. With  $Y$  being 1 or  $\mathbf{g}(\mathbf{X})$ , (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  $\kappa$  and  $\mathbf{U}$ , (A1), and (A3), there are constants  $C_1$  and  $C_2$  such that

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

and

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

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

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

we obtain that

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

which goes to 0 under (A4). From the Lipschitz continuity of  $f_U$ , the boundedness of  $\kappa$ 's

third moment, and the boundedness of second gradient of  $\mu$ ,

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

since  $b \rightarrow 0$ . Considering grid of regions  $A_\ell$  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  $\kappa_b(\mathbf{u} - \mathbf{U}_i)(\mathbf{u} - \mathbf{U}_i)^\top \nabla^2 \mu(\mathbf{u})(\mathbf{u} - \mathbf{U}_i)$  replaced by  $\kappa_b(\mathbf{u} - \mathbf{U}_i) \nabla \mu(\mathbf{u})^\top (\mathbf{u} - \mathbf{U}_i)$ , we can establish that

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

From Taylor's expansion and the boundedness of the second derivative of density  $f_U$  and  $\kappa$ 's

third moment, we obtain that

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

where the second equality follows from Taylor's expansion of  $f_U$  at  $\mathbf{u}$  and the third equality follows from the fact that  $\kappa$  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

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

Note that

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

Since  $l \rightarrow 0$ , this proves that the term in (S4.12) is  $o_p(1)$ .

□

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

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

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

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

where

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

and  $R(\mathbf{U}_i - \mathbf{u})$  is the residual term in Taylor's expansion, which is bounded in absolute value by  $M \|\mathbf{U}_i - \mathbf{u}\|^3$  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  $S_4(\mathbf{u}) + S_5(\mathbf{u}) - A(\mathbf{u}) \xrightarrow{p} 0$  and  $S_6(\mathbf{u}) \xrightarrow{p} 0$  uniformly for  $\mathbf{u} \in \mathbb{U}$ . Hence,

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

Then, the result follows from

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

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).  $\square$

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

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

*Proof.* Using the notation in Section 3, we define

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

and

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

Then, there is a function  $\eta$  such that  $A(\mathbf{u}) = \eta(\boldsymbol{\theta}(\mathbf{u}))$  and  $\widehat{A} = \eta(\widehat{\boldsymbol{\theta}}(\mathbf{u}))$ . Under (B1)-(B3),

Lemma 8.10 in Newey and McFadden (1994) shows that

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

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

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

From the mean value theorem, we have

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

for some  $\boldsymbol{\xi}$  in the segment between  $\widehat{\boldsymbol{\theta}}(\mathbf{u})$  and  $\boldsymbol{\theta}(\mathbf{u})$ . 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

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

Note that

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

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  $Y_i = A(\mathbf{U}_i) - g^*(\mathbf{x})$  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  $iso_p(n^{-2/(p+4)})$  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

Covariate	Model	Test data	$b, l$	$\hat{\mu}_K$ (2.2)	$\hat{\mu}_{CK}$ in (2.7) with constraint (2.5) $\mathbf{g} =$			
					1	(1, $X$ )	(1, $\hat{h}$ )	(1, $X, \hat{h}$ )
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

Covariate	Model	Test data	$b, l$	$\hat{\mu}_K$ in (2.2)	$\hat{\mu}_{CK}$ in (2.7) with constraint	
					(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,  $\mathbf{g}(X) = (1, X)$ .

## 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.