Metric Learning via Cross-Validation

Linlin Dai, Kani Chen, Gang Li and Yuanyuan Lin

Southwestern University of Finance and Economics,

Hong Kong University of Science and Technology,

University of California, Los Angeles and

The Chinese University of Hong Kong

Supplementary Material

S1. Technical Proofs

S1.1 Proof of Theorem 1

It suffices to show $\|\mathbf{L}_0^\top \hat{\mathbf{L}}_1^\perp\| \to 0$ in probability, where $\hat{\mathbf{L}}_1^\perp$ is the basis of the subspace orthogonal to that spanned by the columns of $\hat{\mathbf{L}}_1$, which is obtained via the minimization in problem (2.2). First, we show that $\hat{f}^{(-i)}(\mathbf{X}_i)$ is not a consistent estimator of $f(\mathbf{X}_i)$ when $\|\mathbf{L}_0^\top \hat{\mathbf{L}}_1^\perp\| \nrightarrow 0$. On the one hand,

$$\begin{aligned}
\mathrm{CM}_n(\mathbf{M}) &= \frac{1}{n}\sum_{i=1}^{n}\{\hat{f}^{(-i)}(\mathbf{X}_i) - f(\mathbf{X}_i) - \epsilon_i\}^2 w(\mathbf{X}_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{X}_i)\epsilon_i^2 + \frac{1}{n}\sum_{i=1}^{n}\{\hat{f}^{(-i)}(\mathbf{X}_i) - f(\mathbf{X}_i)\}^2 w(\mathbf{X}_i) \\
&\quad -\frac{2}{n}\sum_{i=1}^{n}\{\hat{f}^{(-i)}(\mathbf{X}_i) - f(\mathbf{X}_i)\} w(\mathbf{X}_i)\epsilon_i \\
&\equiv \eta_0 + \Pi_1 + \Pi_2,
\end{aligned}$$

where η_0 is irrelevant to the minimization over $\mathbf{M}$. Since Π_2 is mean 0 and

$$E(\Pi_1|\mathbf{X}_1,\ldots,\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{X}_i)\Big\{B(\mathbf{X}_i)^2 + \sum_{j\neq i} K_{j,i}^{*2}\sigma_j^2\Big\},$$

$$E(\Pi_2^2|\mathbf{X}_1,\ldots,\mathbf{X}_n) = \frac{4}{n^2}\sum_{i=1}^{n} \sigma_i^2 w(\mathbf{X}_i)\Big\{B(\mathbf{X}_i)^2 + \sum_{j\neq i} K_{j,i}^{*2}\sigma_j^2\Big\},$$

where $B(\mathbf{X}_i) = \tilde{f}(\mathbf{X}_i) - f(\mathbf{X}_i)$ and $\tilde{f}(\mathbf{X}_i) = \sum_{j=1}^{n} f(\mathbf{X}_j)K_{j,i}^*$. We have $\Pi_2 = O_p(n^{-1}\sqrt{\sum_{i=1}^{n}\{B(\mathbf{X}_i)^2 + \sum_{j\neq i} K_{j,i}^{*2}\sigma_j^2\}})$. Hence, Π_1 is the dominant term compared with Π_2 in $\mathrm{CM}_n(\mathbf{M}) - \eta_0$. On the other hand,

$$\hat{f}^{(-i)}(\mathbf{X}_i) - f(\mathbf{X}_i) = \sum_{j=1}^{n} f(\mathbf{X}_j)K_{j,i}^* - f(\mathbf{X}_i) + \sum_{j=1}^{n}\epsilon_j K_{j,i}^*.$$

Suppose that there exists a subsequence of $n = 1, 2, \cdots$, such that $\hat{\mathbf{L}}_1 \to \mathbf{L}_1^{\dagger}$ but $\|\mathbf{L}_0^{\top}\mathbf{L}_1^{\dagger\perp}\| \neq 0$. For notational simplicity, we still denote the subsequence as the original n. With $\|\hat{\mathbf{h}}\| \to 0$,

$$\sum_{j=1}^{n} f(\mathbf{X}_j)K_{j,i}^* \to \mathrm{E}\{f(\mathbf{X})|\mathbf{X}\in\mathbf{x}+\mathbf{L}_1^{\dagger\perp}\}|_{\mathbf{x}=\mathbf{X}_i} \equiv f^{\dagger}(\mathbf{X}_i)$$

in probability, as $n \to \infty$. We intend to show

$$P\{f^{\dagger}(\mathbf{X}) = f(\mathbf{X})\} < 1.$$

Indeed, suppose that $P\{f^{\dagger}(\mathbf{X}) = f(\mathbf{X})\} = 1$ and then $f^{\dagger}(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$. Since $f^{\dagger}(\mathbf{x}) = f^{\dagger}(\mathbf{t})$ if $\mathbf{x} - \mathbf{t} \in \mathbf{L}_1^{\dagger\perp}$, we have $f(\mathbf{x}) = f(\mathbf{t})$ if $\mathbf{x} - \mathbf{t} \in \mathbf{L}_1^{\dagger\perp}$. It follows that $f\{\mathbf{t} + c(\mathbf{x} - \mathbf{t})\} = f(\mathbf{t})$ for all $c \in \mathbb{R}$. By the identifiability condition (C4), we have $\mathbf{x} - \mathbf{t} \in \mathcal{F}$ and thus $\mathbf{L}_1^{\dagger\perp} \subseteq \mathcal{F} = \mathbf{L}_0^{\perp}$ or equivalently $\mathcal{S}(\mathbf{L}_0) \subseteq \mathcal{S}(\mathbf{L}_1^{\dagger})$. Since $\mathbf{L}_0$ and $\mathbf{L}_1^{\dagger}$ are both column orthogonal matrices of size $p \times r_0$, we have $\mathcal{S}(\mathbf{L}_0) = \mathcal{S}(\mathbf{L}_1^{\dagger})$. This is in contradiction with the assumption that $\|\mathbf{L}_0^{\top}\mathbf{L}_1^{\dagger\perp}\| \neq 0$. Hence, $P\{f^{\dagger}(\mathbf{X}) = f(\mathbf{X})\} < 1$.

Since $f^{\dagger}(\cdot)$ and $f(\cdot)$ are smooth functions, we write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\{\hat{f}^{(-i)}(\mathbf{X}_i) - f(\mathbf{X}_i)\}^2 w(\mathbf{X}_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\{\hat{f}^{(-i)}(\mathbf{X}_i) - f^{\dagger}(\mathbf{X}_i) + f^{\dagger}(\mathbf{X}_i) - f(\mathbf{X}_i)\}^2 w(\mathbf{X}_i) \\
&\geq \frac{1}{n}\sum_{i=1}^{n}\{f^{\dagger}(\mathbf{X}_i) - f(\mathbf{X}_i)\}^2 w(\mathbf{X}_i) + \frac{1}{n}\sum_{i=1}^{n}\{\hat{f}^{(-i)}(\mathbf{X}_i) - f^{\dagger}(\mathbf{X}_i)\}^2 w(\mathbf{X}_i) \\
&\quad - \frac{2}{n}\sum_{i=1}^{n}|\{\hat{f}^{(-i)}(\mathbf{X}_i) - f^{\dagger}(\mathbf{X}_i)\}\{f^{\dagger}(\mathbf{X}_i) - f(\mathbf{X}_i)\}| w(\mathbf{X}_i) \\
&\geq c_0 + o_p(1). \qquad \text{(S1.1)}
\end{aligned}$$

The last inequality is followed by the Cauchy-Schwarz inequality, the strong law of large number of $W_i \equiv \{f^{\dagger}(\mathbf{X}_i) - f(\mathbf{X}_i)\}^2 w(\mathbf{X}_i)$ and the consistency of $\hat{f}^{(-i)}(\mathbf{X}_i)$ with repect to $f^{\dagger}(\mathbf{X}_i)$. As a result, $\mathrm{CM}_n(\hat{\mathbf{M}}) - \eta_0$ is at the order of $O_p(1)$ with some positive lower bound $c_0 > 0$. Nevertheless, we now show that with $\|\mathbf{L}_0^\top \hat{\mathbf{L}}_1^\perp\| \to 0$ and $\|\mathbf{h}\| \to 0$,

$$\sup_{1\le i\le n} \left|\hat{f}^{(-i)}(\mathbf{X}_i) - f(\mathbf{X}_i)\right| \to 0 \quad \text{in probability}, \tag{S1.2}$$

as $n \to \infty$. As a consequence, $\text{CM}_n(\mathbf{M}) - \eta_0$ is at the order of $o_p(1)$.

In fact, recall that Ω° is the support of $w(\cdot)$ and $f_{r_0}(\mathbf{u}) = f(u_1, \ldots, u_{r_0})$ is the density function of $\mathbf{U} = \mathbf{L}_0^\top \mathbf{X}$. Now, define $\Omega^{\bar{\delta}} = \{\mathbf{y} \in \mathbb{R}^p : \inf_{\mathbf{x}\in\Omega^\circ} \|\mathbf{y}-\mathbf{x}\| < \bar{\delta}\}$, where $\bar{\delta} > 0$ is a small constant such that $\min_{\mathbf{x}\in\Omega^{\bar{\delta}}} f_{\mathbf{X}}(\mathbf{x}) > 0$. Hence, there exists $\tau > 0$ such that $\min_{\mathbf{x}\in\Omega^{\bar{\delta}}} f_{r_0}(\mathbf{L}_0^\top \mathbf{x}) \ge \tau$. To show (S1.2), it is sufficient to prove that for any $\epsilon > 0$,

$$P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\hat{f}_n(\mathbf{x}) - f(\mathbf{x})| > \epsilon\Big\} \to 0, \quad \text{as } n \to \infty. \tag{S1.3}$$

For simplicity, denote $\nu(\mathbf{x}) = f_{r_0}(\mathbf{L}_0^\top \mathbf{x})$. Let $\phi(\mathbf{x}) = f(\mathbf{x})\nu(\mathbf{x}) = g(\mathbf{L}_0^\top \mathbf{x})\nu(\mathbf{x})$,

$$\phi_n(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^n Y_j K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{x}), \quad \nu_n(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^n K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{x}).$$

And thus $\hat{f}_n(\mathbf{x}) = \phi_n(\mathbf{x})/\nu_n(\mathbf{x})$. It is not hard to verify that

$$\begin{aligned}
&P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\hat{f}_n(\mathbf{x}) - f(\mathbf{x})| > \epsilon\Big\} \\
\le\ & P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\phi_n(\mathbf{x}) - f(\mathbf{x})\nu_n(\mathbf{x})| \ge \epsilon(\tau-\epsilon)\Big\} + P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\nu_n(\mathbf{x}) - \nu(\mathbf{x})| > \epsilon\Big\} \\
\le\ & P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\nu_n(\mathbf{x}) - \nu(\mathbf{x})| > \frac{\epsilon(\tau-\epsilon)}{2b}\Big\} + P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\nu_n(\mathbf{x}) - \nu(\mathbf{x})| > \epsilon\Big\} \\
&+ P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\phi_n(\mathbf{x}) - f(\mathbf{x})\nu(\mathbf{x})| \ge \frac{\epsilon(\tau-\epsilon)}{2}\Big\},
\end{aligned} \tag{S1.4}$$

where $b = \sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |f(\mathbf{x})| < \infty$. Recall that

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_1 & \mathbf{L}_2 \end{pmatrix}$$

is a $p \times p$ orthonormal matrix, where $\mathbf{L}_1 \in \mathbb{R}^{p\times r_0}$ and $\mathbf{L}_2$ is the augmented orthonormal basis in $\mathbb{R}^p$. Define $f_{\mathbf{L}_2}(\mathbf{x}) = \int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} f_{\mathbf{X}}(\mathbf{x} + \mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2$.

To proceed, we first show that as $n \to \infty$, $\|\mathbf{h}\| \to 0$ and $\|\mathbf{L}_0^\top\mathbf{L}_2\| \to 0$,

$$\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} \left|\mathrm{E}\{\phi_n(\mathbf{x})\} - \phi(\mathbf{x})\right| \to 0. \tag{S1.5}$$

Define $\tilde{\phi}(\mathbf{x}) = f(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})$, $\phi_{\mathbf{L}_2}(\mathbf{x}) = \int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} \tilde{\phi}(\mathbf{x} + \mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2$ and $\mathbf{I}_{r_0}$ be the $r_0 \times r_0$ identity matrix. We have

$$\begin{aligned}
&\mathrm{E}\{f(\mathbf{X})K_{\mathbf{M}}(\mathbf{X}-\mathbf{x})\} \qquad\qquad \text{(S1.6)}\\
= \;& \frac{1}{h_1\cdots h_{r_0}} \int_{\mathbf{t}\in\mathbb{R}^p} K\big\{(\mathbf{t}-\mathbf{x})^\top\mathbf{M}(\mathbf{t}-\mathbf{x})\big\}\tilde{\phi}(\mathbf{t})d\mathbf{t}\\
= \;& \frac{1}{h_1\cdots h_{r_0}} \int_{\mathbf{t}\in\mathbb{R}^p} K(\mathbf{t}^\top\mathbf{L}_1\mathbf{H}^{-2}\mathbf{L}_1^\top\mathbf{t})\tilde{\phi}(\mathbf{t}+\mathbf{x})d\mathbf{t}\\
= \;& \frac{1}{h_1\cdots h_{r_0}} \int_{\mathbf{s}\in\mathbb{R}^p} K(\mathbf{s}^\top\mathbf{L}^\top\mathbf{L}_1\mathbf{H}^{-2}\mathbf{L}_1^\top\mathbf{L}\mathbf{s})\tilde{\phi}(\mathbf{L}\mathbf{s}+\mathbf{x})d\mathbf{s}\\
= \;& \frac{1}{h_1\cdots h_{r_0}} \int_{\mathbf{s}_1\in\mathbb{R}^{r_0},\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} K\left\{\begin{pmatrix}\mathbf{s}_1^\top & \mathbf{s}_2^\top\end{pmatrix}\begin{pmatrix}\mathbf{I}_{r_0}\\ \mathbf{0}\end{pmatrix}\mathbf{H}^{-2}\begin{pmatrix}\mathbf{I}_{r_0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{pmatrix}\right\}\\
&\qquad\times\tilde{\phi}(\mathbf{L}_1\mathbf{s}_1+\mathbf{L}_2\mathbf{s}_2+\mathbf{x})d\mathbf{s}_1d\mathbf{s}_2\\
= \;& \frac{1}{h_1\cdots h_{r_0}} \int_{\mathbf{s}_1\in\mathbb{R}^{r_0},\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} K(\mathbf{s}_1^\top\mathbf{H}^{-2}\mathbf{s}_1)\tilde{\phi}(\mathbf{L}_1\mathbf{s}_1+\mathbf{L}_2\mathbf{s}_2+\mathbf{x})d\mathbf{s}_1d\mathbf{s}_2\\
= \;& \int_{\mathbf{s}_1\in\mathbb{R}^{r_0}} K(\|\mathbf{s}_1\|^2)\phi_{\mathbf{L}_2}(\mathbf{x}+\mathbf{L}_1\mathbf{H}\mathbf{s}_1)d\mathbf{s}_1,\\
= \;& \phi_{\mathbf{L}_2}(\mathbf{x}) + \frac{R_1(K)}{2}\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top\ddot{\phi}_{\mathbf{L}_2}(\mathbf{x})\mathbf{L}_1\mathbf{H}\} + o(\|\mathbf{h}\|^2),
\end{aligned}$$

where the last equality is due to the Taylor expansion of $\phi_{\mathbf{L}_2}(\mathbf{x}+\mathbf{L}_1\mathbf{H}\mathbf{s}_1)$ and the condition $\int_{\mathbf{s}_1\in\mathbb{R}^{r_0}}\mathbf{s}_1 K(\|\mathbf{s}_1\|^2)d\mathbf{s}_1=0$. Therefore, we have

$$\begin{aligned}\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}|\mathrm{E}\{\phi_n(\mathbf{x})\}-\phi_{\mathbf{L}_2}(\mathbf{x})| &\leq \sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}|R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top\ddot{\phi}_{\mathbf{L}_2}(\mathbf{x})\mathbf{L}_1\mathbf{H}\}|\{1+o(1)\}\\ &= O(\|\mathbf{h}\|^2). \qquad (S1.7)\end{aligned}$$

Recall that $f(\mathbf{x})=g(\mathbf{L}_0^\top\mathbf{x})$. According to the Taylor's expansion,

$$\begin{aligned}&\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}|\phi_{\mathbf{L}_2}(\mathbf{x})-\phi(\mathbf{x})|\\ &= \sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}\left|\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}g(\mathbf{L}_0^\top\mathbf{x}+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2)f_{\mathbf{X}}(\mathbf{x}+\mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2-g(\mathbf{L}_0^\top\mathbf{x})\nu(\mathbf{x})\right|\\ &= \sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}\Big|g(\mathbf{L}_0^\top\mathbf{x})f_{\mathbf{L}_2}(\mathbf{x})+\dot{g}(\mathbf{L}_0^\top\mathbf{x})^\top\mathbf{L}_0^\top\mathbf{L}_2\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}\mathbf{s}_2 f(\mathbf{x}+\mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2\\ &\qquad -g(\mathbf{L}_0^\top\mathbf{x})\nu(\mathbf{x})+o(||\mathbf{L}_0^\top\mathbf{L}_2||)\Big|\\ &= \sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}\left|g(\mathbf{L}_0^\top\mathbf{x})f_{\mathbf{L}_2}(\mathbf{x})-g(\mathbf{L}_0^\top\mathbf{x})\nu(\mathbf{x})+O(||\mathbf{L}_0^\top\mathbf{L}_2||)\right|\\ &\leq \sup_{\mathbf{x}\in\Omega^{\bar{\delta}}}\left|g(\mathbf{L}_0^\top\mathbf{x})\right|o(1)+O(||\mathbf{L}_0^\top\mathbf{L}_2||), \qquad (S1.8)\end{aligned}$$

where the last inequality holds by the fact that $f_{\mathbf{L}_2}(\mathbf{x})=f_{r_0}(\mathbf{L}_0^\top\mathbf{x})\{1+o(1)\}$, as $n\to\infty$ and $\|\mathbf{L}_0^\top\mathbf{L}_2\|\to 0$. This result can be derived using the condition (C1) and the Taylor expansion of $f_{\mathbf{L}_2}(\mathbf{x})$. In fact, recall that $f_{\mathbf{X}}(\mathbf{x})$ is the density of $\mathbf{X}$ and thus the density function $f_Q(\mathbf{u})$ of $\mathbf{U}=\mathbf{Q}\mathbf{X}$ satisfies $f_Q(\mathbf{u})=f_{\mathbf{X}}(\mathbf{x})$ for any $p\times p$ rotation matrix $\mathbf{Q}$. By taking $\mathbf{Q}^\top=(\mathbf{L}_0\ \mathbf{L}_0^\perp)$, we have that as $\|\mathbf{L}_0^\top \mathbf{L}_2\| \to 0$,

$$\begin{aligned}
f_{\mathbf{L}_2}(\mathbf{x}) &= \int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} f_{\mathbf{X}}(\mathbf{x}+\mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2 \\
&= \int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} f_{\mathbf{Q}}\begin{pmatrix} \mathbf{L}_0^\top\mathbf{x}+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2 \\ \mathbf{L}_0^{\perp\top}\mathbf{x}+\mathbf{L}_0^{\perp\top}\mathbf{L}_2\mathbf{s}_2 \end{pmatrix} d\mathbf{s}_2 \\
&= \int_{\tilde{\mathbf{s}}_2\in\mathbb{R}^{(p-r_0)}} f_{\mathbf{Q}}\begin{pmatrix} \mathbf{L}_0^\top\mathbf{x} \\ \tilde{\mathbf{s}}_2 \end{pmatrix} d\tilde{\mathbf{s}}_2\{1+o(1)\} \\
&= f_{r_0}(\mathbf{L}_0^\top\mathbf{x})\{1+o(1)\}.
\end{aligned}$$

Hence, as $n \to 0$, if $\|\mathbf{h}\| \to 0$ and $\|\mathbf{L}_0^\top \mathbf{L}_2\| \to 0$, (S1.7) in conjunction with (S1.8) yields (S1.5). On the other hand, following a similar proof of Lemma B.1 in Newey (1994) and applying condition (C5) and $h_1 \cdots h_{r_0} > n^{-\delta}$ for some $0 < \delta < 1$, we have

$$\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} \left|\phi_n(\mathbf{x}) - \mathrm{E}\{\phi_n(\mathbf{x})\}\right| \to 0, \quad \text{in probability.} \tag{S1.9}$$

Therefore, (S1.7), (S1.8) and (S1.9) yield

$$P\Big\{\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\phi_n(\mathbf{x}) - f(\mathbf{x})\nu(\mathbf{x})| \geq \frac{\epsilon(\tau-\epsilon)}{2}\Big\} \to 0.$$

Likewise, by replacing Y_i with 1, it can be shown that $\sup_{\mathbf{x}\in\Omega^{\bar{\delta}}} |\nu_n(\mathbf{x}) - \nu(\mathbf{x})| \to 0$ in probability. As a result, combining inequalities (S1.4), we have proved (S1.3).

In conclusion, in case of $||\mathbf{L}_0^\top \mathbf{L}_2|| \to 0$ and $\|\mathbf{h}\| \to 0$, $\mathrm{CM}_n(\mathbf{M}) - \eta_0$ is at the order of $o_p(1)$. This violates the definition that $\mathrm{CM}_n(\hat{\mathbf{M}}) \le \mathrm{CM}_n(\mathbf{M})$ for $\mathbf{M} \in \mathbb{S}_+^p$. The proof is complete.

S1.2 Proof of Theorem 2

For brevity, we write $w(\mathbf{X}_i)$ by w_i. The following lemmas are needed to prove Theorem 2. The proofs of Lemmas 1–4 is given latter.

Lemma 1. *Suppose conditions (C1)–(C5) hold. Then,* $\mathrm{E}\{K_{\mathbf{M}}(\mathbf{X}-\mathbf{x})\} = f_{\mathbf{L}_2}(\mathbf{x}) + O(\|\mathbf{h}\|^2)$*. Moreover, for any* $i = 1, \ldots, n$, $\sum_{j\neq i}^n K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{X}_i) = n f_{\mathbf{L}_2}(\mathbf{X}_i)\{1 + o_p(1)\}$.

Lemma 2. *Define* $\sigma^2_{\mathbf{L}_2}(\mathbf{x}) = \int_{s\in\mathbb{R}^{(p-r_0)}} f_{\mathbf{X}}(\mathbf{x}+\mathbf{L}_2\mathbf{s})\sigma^2(\mathbf{x}+\mathbf{L}_2\mathbf{s})d\mathbf{s}$*. Under conditions (C1)–(C5), we have*

$$\mathrm{E}\{(K^*_{j,i})^2\sigma_j^2 w_i\} = \frac{R_2(K)V_0}{n^2 h_1\cdots h_{r_0}}\{1+o(1)\},$$

where $R_2(K) = \int_{\mathbf{s}\in\mathbb{R}^{r_0}} K^2(\|\mathbf{s}\|^2)d\mathbf{s}$ *and*

$$V_0 = \int_{\mathbf{x}\in\mathbb{R}^p} \sigma^2(\mathbf{L}_0^\top\mathbf{x})\frac{f_{\mathbf{X}}(\mathbf{x})w(\mathbf{x})}{f_{r_0}(\mathbf{L}_0^\top\mathbf{x})}d\mathbf{x}.$$

Lemma 3. *Under regularity conditions (C1)–(C5), suppose* $\|\mathbf{L}_0^\top\mathbf{L}_2\| \to 0$ *and* $\|\mathbf{h}\| \to 0$*. Then, for any* $\mathbf{t} \in \Omega$,

$$\mathrm{E}[\{f(\mathbf{X}) - f(\mathbf{t})\}K_{\mathbf{M}}(\mathbf{X}-\mathbf{t})] = \psi(\mathbf{t},\mathbf{h},\mathbf{L}_1) + o(\|\mathbf{h}\|^2 + \|\mathbf{L}_0^\top\mathbf{L}_2\|),$$

where the definition of $\psi(\cdot)$ *is given in Theorem 2.*

Lemma 4. *Under regularity conditions (C1)–(C5), for any* $\mathbf{t} \in \Omega$, $\mathrm{Var}[\{f(\mathbf{X}) - f(\mathbf{t})\}K_{\mathbf{M}}(\mathbf{X}-\mathbf{t})] = O\{\|\mathbf{h}\|^2/(h_1 \cdots h_{r_0})\}$. *Consequently,*

$$\mathrm{Var}\left[\frac{1}{n}\sum_{j=1}^{n}\{f(\mathbf{X}_j) - f(\mathbf{t})\}K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{t})\right] = O\left(\frac{1}{nh_1 \cdots h_{r_0}}\right).$$

Proof of Theorem 2. Write

$$\begin{aligned}
\mathrm{CM}_n(\mathbf{M}) &= \frac{1}{n}\sum_{i=1}^{n} w_i\epsilon_i^2 + \frac{1}{n}\sum_{i=1}^{n}\{B(\mathbf{X}_i)\}^2 w_i + \frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}\epsilon_j K_{j,i}^*\Big)^2 w_i \\
&\quad -\frac{2}{n}\sum_{i=1}^{n} B(\mathbf{X}_i) w_i \epsilon_i - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\epsilon_i\epsilon_j K_{j,i}^* w_i \\
&\quad +\frac{2}{n}\sum_{i=1}^{n} B(\mathbf{X}_i)\sum_{j=1}^{n}\epsilon_j K_{j,i}^* w_i \\
&\equiv \eta_0 + \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5,
\end{aligned}$$

where $\tilde{f}(\mathbf{X}_i) = \sum_{j=1}^{n} f(\mathbf{X}_j)K_{j,i}^*$ and $B(\mathbf{X}_i) = \tilde{f}(\mathbf{X}_i) - f(\mathbf{X}_i)$. Here $B(\cdot)$ stands for the bias. Observe the facts that

(a) $\eta_0 \equiv n^{-1}\sum_{i=1}^{n} w_i\epsilon_i^2$ is free of $\mathbf{M}$ and thus it is irrelevant to the minimization over $\mathbf{M}$.

(b) $\eta_1 \equiv n^{-1}\sum_{i=1}^{n}\{B(\mathbf{X}_i)\}^2 w_i$ stands for the bias term and $\eta_1 \geq 0$.

(c) η_2 is viewed as the variance term and $\eta_2 \geq 0$. $E(\eta_2|\mathbf{X}_1, \ldots, \mathbf{X}_n) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(K_{j,i}^*)^2\sigma_j^2 w_i$.

(d) $E(\eta_3|\mathbf{X}_1, \ldots, \mathbf{X}_n) = 0$ and $E(\eta_3^2|\mathbf{X}_1, \ldots, \mathbf{X}_n) = 4n^{-2}\sum_{i=1}^{n}\{B(\mathbf{X}_i)\}^2\sigma_i^2 w_i^2$. Hence, $\eta_3 = O_p(n^{-1}\sqrt{\sum_{i=1}^{n}\{B(\mathbf{X}_i)\}^2})$.

(e) $E(\eta_4|\mathbf{X}_1,\ldots,\mathbf{X}_n)=0$ and $E(\eta_4^2|\mathbf{X}_1,\ldots,\mathbf{X}_n)=4n^{-2}\sum_{i=1}^n\sum_{j=1}^n\big\{(K_{j,i}^*)^2w_i^2$ $+K_{j,i}^*K_{i,j}^*w_iw_j\big\}\sigma_i^2\sigma_j^2$. Hence, $\eta_4=O_p(n^{-1}\sqrt{\sum_{i=1}^n\sum_{j=1}^n\{K_{j,i}^*\}^2})$.

(f) $\mathrm{E}(\eta_5|\mathbf{X}_1,\ldots,\mathbf{X}_n)=0$ and $\eta_5=O_p\big(n^{-1/2}\sup_{\mathbf{X}\in\Omega^{\bar\delta}}|B(\mathbf{X})|\big)$.

The above statements (a)–(e) are trivial and we only give one justification for statement (f).

Since $\|\mathbf{h}\|\to 0$ as $n\to\infty$, $K_{j,i}^*w_i=0$ for all $(\mathbf{X}_i,\mathbf{X}_j)$ if $\mathbf{X}_i$ or $\mathbf{X}_j$ is outside $\Omega^{\bar\delta}$ for all large n. Set $a_n(\mathbf{x})=n^{-1}\sum_{l\neq i}K_{\mathbf{M}}(\mathbf{X}_l-\mathbf{x})$. By the Lemma 1, with probability one, for all large n, there exists some constant $C>0$ such that

$$1/C\le\inf_{\mathbf{x}\in\Omega^{\bar\delta}}a_n(\mathbf{x})\le\sup_{\mathbf{x}\in\Omega^{\bar\delta}}a_n(\mathbf{x})\le C.$$

It follows that with probability one, for all large n,

$$\begin{aligned}\sup_{1\le j\le n}\Big|\sum_{i=1}^n B(\mathbf{X}_i)K_{j,i}^*w_i\Big| &= \sup_{\mathbf{X}_j\in\Omega^\circ}\Big|\sum_{i=1}^n B(\mathbf{X}_i)K_{j,i}^*w_i\Big|\\ &\le\Big\{\sup_{\mathbf{X}_i\in\Omega^{\bar\delta}}\big|B(\mathbf{X}_i)w_i\big|\Big\}\Big\{\sup_{\mathbf{X}_j\in\Omega^{\bar\delta}}\sum_{j=1}^n K_{j,i}^*\Big\}\\ &=\Big\{\sup_{\mathbf{X}_i\in\Omega^{\bar\delta}}\big|B(\mathbf{X}_i)w_i\big|\Big\}\Bigg\{\sup_{\mathbf{X}_j\in\Omega^{\bar\delta}}\frac{1}{n}\sum_{j=1}^n\frac{K_{\mathbf{M}}(\mathbf{X}_j-\mathbf{X}_i)}{a_n(\mathbf{X}_i)}\Bigg\}\\ &\le C^2\sup_{\mathbf{X}_i\in\Omega^{\bar\delta}}\big|B(\mathbf{X}_i)w_i\big|.\end{aligned}$$

Hence,

$$\mathrm{E}(\eta_5^2|\mathbf{X}_1,\ldots,\mathbf{X}_n)=\frac{4}{n^2}\sum_{j=1}^n\sigma_j^2\Big\{\sum_{i=1}^n B(\mathbf{X}_i)K_{j,i}^*w_i\Big\}^2=O_p\big(n^{-1}\sup_{\mathbf{X}_i\in\Omega^{\bar\delta}}\{B(\mathbf{X}_i)\}^2w_i^2\big).$$

In the following, we intend to show that η_1 and η_2 are the dominating terms, compared with η_3, η_4 and η_5.

Write

$$\begin{aligned}\{B(\mathbf{X}_i)\}^2 &= \{\tilde{f}(\mathbf{X}_i) - f(\mathbf{X}_i)\}^2 \\ &= \frac{\Big(n^{-1}\Big[\sum_{j\neq i}\{f(\mathbf{X}_j) - f(\mathbf{X}_i)\}K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{X}_i)\Big]\Big)^2}{\Big\{n^{-1}\sum_{j\neq i}K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{X}_i)\Big\}^2}.\end{aligned}$$

By Lemma 1, the above denominator is $f^2_{\mathbf{L}_2}(\mathbf{X}_i)\{1 + o_p(1)\}$. And it follows from Lemmas 3–4 that the above numerator is

$$\psi^2(\mathbf{X}_i, \mathbf{h}, \mathbf{L}_1) + o_p\left(\|\mathbf{h}\|^4 + \|\mathbf{L}_0^\top \mathbf{L}_2\|^2 + \frac{1}{nh_1\cdots h_{r_0}}\right).$$

Hence, by the law of large number and the continuous mapping theorem, we have

$$\begin{aligned}\eta_1 = \frac{1}{n}\sum_{i=1}^n\{B(\mathbf{X}_i)\}^2 w_i &= \int_{\mathbf{x}\in\mathbb{R}^p}\frac{\psi^2(\mathbf{x}, \mathbf{h}, \mathbf{L}_1)}{f^2_{r_0}(\mathbf{L}_0^\top\mathbf{x})}f_{\mathbf{X}}(\mathbf{x})w(\mathbf{x})d\mathbf{x} \\ &\quad + o_p\left(\|\mathbf{L}_0^\top\mathbf{L}_2\|^2 + \|\mathbf{h}\|^4 + \frac{1}{nh_1\cdots h_{r_0}}\right).\end{aligned}$$

Write

$$\eta_2 = \frac{1}{n}\sum_{j=1}^n \epsilon_j^2\tilde{a}_j + \frac{1}{n}\sum_{j_1=1}^n\sum_{j_2\neq j_1}^n \epsilon_{j_1}\epsilon_{j_2}\tilde{a}_{j_1,j_2},$$

where $\tilde{a}_j = \sum_{i=1}^n (K^*_{j,i})^2 w_i$ for all i and $\tilde{a}_{j_1,j_2} = \sum_{i=1}^n K^*_{j_1,i}K^*_{j_2,i}w_i$ for all $j_2 \neq j_1$. We now show that $\sum_{j=1}^n \epsilon_j^2\tilde{a}_j$ is the dominant term in η_2.

First, from Lemma 2, we have

$$E\Big(\frac{1}{n}\sum_{j=1}^n \epsilon_j^2\tilde{a}_j\Big) = \frac{R_2(K)V_0}{nh_1\cdots h_{r_0}}\{1 + o(1)\}.$$

Second, it can be easily verified that $E(\sum_{j_1=1}^n \sum_{j_2 \neq j_1}^n \epsilon_{j_1}\epsilon_{j_2}\tilde{a}_{j_1,j_2}) = 0$. By Lemmas 1–2, we have $\tilde{a}_{j_1,j_2} = O_p(n^{-1})$. And recall condition (C5), it follows that

$$\mathrm{E}\Big(\sum_{j_1=1}^n \sum_{j_2 \neq j_1}^n \epsilon_{j_1}\epsilon_{j_2}\tilde{a}_{j_1,j_2}\Big)^2 = \mathrm{E}\Big(\sum_{j_1=1}^n \sum_{j_2 \neq j_1}^n \sigma_{j_1}^2\sigma_{j_2}^2\tilde{a}_{j_1,j_2}^2\Big) = O(1).$$

As a result,

$$\eta_2 = \frac{1}{n}\sum_{j=1}^n \epsilon_j^2 \tilde{a}_j\{1 + o_p(1)\} = \frac{R_2(K)V_0}{nh_1 \cdots h_{r_0}}\{1 + o_p(1)\}.$$

Since η_3, η_4 and η_5 are of smaller order than $\|\mathbf{L}_0^\top \mathbf{L}_2\|^2 + \|\mathbf{h}\|^4 + 1/(nh_1 \cdots h_{r_0})$ and η_0 is free of $\mathbf{M}$, then

$$\begin{aligned}
\mathrm{CM}_n(\mathbf{M}) - \eta_0 &= \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 \\
&= \int_{\mathbf{x} \in \mathbb{R}^p} \frac{\psi^2(\mathbf{x}, \mathbf{h}, \mathbf{L}_1)}{f_{r_0}^2(\mathbf{L}_0^\top \mathbf{x})} f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} + \frac{R_2(K)V_0}{nh_1 \cdots h_{r_0}} \\
&\quad + o_p\left(\|\mathbf{L}_0^\top \mathbf{L}_2\|^2 + \|\mathbf{h}\|^4 + \frac{1}{nh_1 \cdots h_{r_0}}\right).
\end{aligned}$$

The proof is complete. □

S1.3 Proof of Corollary 1

It is seen that $\mathbf{L}_0^\top \mathbf{L}_2$ is only contained in the bias term of the asymptotic expansion shown in Theorem 2. And we can easily verify that

$$\psi(\mathbf{x}, \mathbf{h}, \mathbf{L}_1) = -\mathrm{vec}(\mathbf{T}^\top)^\top \mathrm{vec}\{\mathbf{b}(\mathbf{L}_0^\top \mathbf{t})\dot{g}(\mathbf{L}_0^\top \mathbf{t})^\top\} + R_1(K)\mathrm{tr}(\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \mathbf{A}(\mathbf{L}_0^\top \mathbf{t})\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}),$$

where $\mathbf{T} = \mathbf{L}_0^\top \mathbf{L}_2$. Let $\tilde{\mathbf{b}}(\mathbf{t}) = \text{vec}\{\mathbf{b}(\mathbf{L}_0^\top \mathbf{t})\dot{g}(\mathbf{L}_0^\top \mathbf{t})^\top\}$ and

$$\tilde{c}_1(\mathbf{t}, \mathbf{h}) = R_1(K)\text{tr}(\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \mathbf{A}(\mathbf{L}_0^\top \mathbf{t})\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}).$$

Since the objective function is quadratic, the optimization procedure over $\text{vec}(\mathbf{T}^\top)$ yields the solution

$$\left\{ -\int_{\mathbf{t}\in\mathbb{R}^p} \tilde{\mathbf{b}}(\mathbf{t})\tilde{\mathbf{b}}(\mathbf{t})^\top f_{\mathbf{X}}(\mathbf{t}) \frac{w(\mathbf{t})}{f_{r_0}^2(\mathbf{L}_0^\top \mathbf{t})} d\mathbf{t} \right\}^+ \int_{\mathbf{t}\in\mathbb{R}^p} \tilde{c}_1(\mathbf{t}, \mathbf{h})\tilde{\mathbf{b}}(\mathbf{t}) f_{\mathbf{X}}(\mathbf{t}) \frac{w(\mathbf{t})}{f_{r_0}^2(\mathbf{L}_0^\top \mathbf{t})} d\mathbf{t}, \tag{S1.10}$$

where the $\mathbf{A}^+$ denotes the generalized inverse of a matrix $\mathbf{A}$. By some simple calculations, it can be shown that the order of (S1.10) is $O(\|\mathbf{h}\|^2)$. Since $\mathbf{L}_0^\top \mathbf{L}_1$ is asymptotically orthonormal, we obtain that the $\|\mathbf{L}_0^\top \mathbf{L}_2\| = O(\|\mathbf{h}\|^2)$.

Further, to find the optimal rate of the bandwidth $\mathbf{h}$, we also consider optimizing the asymptotic expansion. Let $\mathbf{L}_0^\top \mathbf{L}_1 = (\tilde{\boldsymbol{\ell}}_1, \cdots, \tilde{\boldsymbol{\ell}}_{r_0})$ and then $\tilde{c}_1(\mathbf{t}, \mathbf{h}) = R_1(K) \sum_{j=1}^{r_0} h_j^2 \tilde{\boldsymbol{\ell}}_j^\top \mathbf{A}(\mathbf{L}_0^\top \mathbf{t})\tilde{\boldsymbol{\ell}}_j$. Taking derivative over h_k, $k = 1, \ldots, r_0$, we have

$$\frac{\partial\{\text{CM}_n(\mathbf{M}) - \eta_0\}}{\partial h_k} = 4h_k \tilde{C}_k(\mathbf{L}_1) + 4h_k \sum_{j=1}^{r_0} C_j h_j^2 - \frac{R_2(K)V_0}{nh_k^2(\prod_{j\neq k} h_j)} = 0,$$

where

$$C_j = \{R_1(K)\}^2 \int_{\mathbf{t}\in\mathbb{R}^p} \{\tilde{\boldsymbol{\ell}}_k^\top \mathbf{A}(\mathbf{L}_0^\top \mathbf{t})\tilde{\boldsymbol{\ell}}_k\}\{\tilde{\boldsymbol{\ell}}_j^\top \mathbf{A}(\mathbf{L}_0^\top \mathbf{t})\tilde{\boldsymbol{\ell}}_j\} f_{\mathbf{X}}(\mathbf{t}) \frac{w(\mathbf{t})}{f_{r_0}^2(\mathbf{L}_0^\top \mathbf{t})} d\mathbf{t} = O(1)$$

and

$$\begin{aligned}\tilde{C}_k(\mathbf{L}_1) &= R_1(K)\int_{\mathbf{t}\in\mathbb{R}^p}\dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{T}\mathbf{b}(\mathbf{L}_0^\top\mathbf{t})\tilde{\boldsymbol{\ell}}_k^\top\mathbf{A}(\mathbf{L}_0^\top\mathbf{t})\tilde{\boldsymbol{\ell}}_k f_{\mathbf{X}}(\mathbf{t})\frac{w(\mathbf{t})}{f_{r_0}^2(\mathbf{L}_0^\top\mathbf{t})}d\mathbf{t}\\
&= R_1(K)\text{vec}(\mathbf{T}^\top)^\top\\
&\quad\times\int_{\mathbf{t}\in\mathbb{R}^p}\text{vec}\{\mathbf{b}(\mathbf{L}_0^\top\mathbf{t})\dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\}\tilde{\boldsymbol{\ell}}_k^\top\mathbf{A}(\mathbf{L}_0^\top\mathbf{t})\tilde{\boldsymbol{\ell}}_k f_{\mathbf{X}}(\mathbf{t})\frac{w(\mathbf{t})}{f_{r_0}^2(\mathbf{L}_0^\top\mathbf{t})}d\mathbf{t}\\
&= O(\|\mathbf{h}\|^2).\end{aligned}$$

As a result, we obtain that $\hat{\mathbf{h}} = O\{n^{-1/(r_0+4)}\}$. This completes the proof.

S1.4 Proof of Proposition 1

Recall that from Theorem 1, we have $\text{CM}_n(\hat{\mathbf{M}}_{r_0}) - \tilde{\eta}_0 = o_p(1)$, where $\tilde{\eta}_0 = E(w(\mathbf{X})\sigma^2(\mathbf{L}_0^\top\mathbf{X}))$ is irrelevant to r_0. When $1 \le r < r_0$, we can show that $\text{CM}_n(\hat{\mathbf{M}}_r) - \tilde{\eta}_0 \ge c_1 + o_p(1)$ for some constant $c_1 > 0$. Let $\hat{\mathbf{L}}_1(r) \in \mathbb{R}^{p\times r}$ be the CVML estimator when the dimension of CMS is set to be r and $\hat{\mathbf{L}}_1^\perp(r)$ be the augmented orthonormal basis in $\mathbb{R}^p$. Since the column vectors of $\mathbf{L}_0$ and $\mathbf{L}_0^\perp$ form a set of basis in $\mathbb{R}^p$, there exists a unique decomposition of $\hat{\mathbf{L}}_1^\perp(r)$ such that

$$\hat{\mathbf{L}}_1^\perp(r) = \mathbf{L}_0\mathbf{A}(r) + \mathbf{L}_0^\perp\mathbf{B}(r), \tag{S1.11}$$

where $\mathbf{A}(r)$ is a $r_0 \times (p-r)$ matrix and $\mathbf{B}(r)$ is a $(p-r_0)\times(p-r)$ matrix. We now show that $\|\mathbf{L}_0^\top\hat{\mathbf{L}}_1^\perp(r)\|$ does not converge to zero. Suppose that $\|\mathbf{L}_0^\top\hat{\mathbf{L}}_1^\perp(r)\|$ converges to zero. Then by the decomposition (S1.11), we have that $\|\mathbf{L}_0^\top \hat{\mathbf{L}}_1^\perp(r)\| = \|\mathbf{A}(r)\| \to 0$. Since the column vectors of $\hat{\mathbf{L}}_1^\perp(r)$ are orthogonal, we have that $\hat{\mathbf{L}}_1^\perp(r)^\top \hat{\mathbf{L}}_1^\perp(r) = \mathbf{A}(r)^\top \mathbf{A}(r) + \mathbf{B}(r)^\top \mathbf{B}(r) = \mathbf{I}_{(p-r)}$, where $\mathbf{I}_{(p-r)}$ is a $(p-r)\times(p-r)$ identity matrix. It follows from $\|\mathbf{A}(r)\| \to 0$ that $\mathbf{B}(r)^\top \mathbf{B}(r) \to \mathbf{I}_{(p-r)}$. However, due to $r < r_0$, the rank of $\mathbf{B}(r)^\top \mathbf{B}(r)$ shall not exceed $(p - r_0)$ and is not able to attain $(p - r)$, which is in contradiction with $\mathbf{B}(r)^\top \mathbf{B}(r) \to \mathbf{I}_{(p-r)}$. Therefore, we have $\|\mathbf{L}_0^\top \hat{\mathbf{L}}_1^\perp(r)\| \nrightarrow 0$. Then it follows from simular proofs in Theorem 1 that $\hat{f}^{(-i)}(\mathbf{X}_i)$ with $\mathbf{M}$ set to be $\hat{\mathbf{M}}_r = \hat{\mathbf{L}}_1(r)\hat{\mathbf{H}}^{-2}\hat{\mathbf{L}}_1(r)^\top$ is not a consistent estimator of $f(\mathbf{X}_i)$ when $\|\mathbf{L}_0^\top \hat{\mathbf{L}}_1^\perp(r)\| \nrightarrow 0$. Moreover, by similar derivation of (S1.1), we obtain that for any $1 \le r < r_0$, there exists a positive constant c_1 such that $\mathrm{CM}_n(\hat{\mathbf{M}}_r) - \tilde{\eta}_0 \ge c_1 + o_p(1)$. As a result, $\mathrm{CM}_n(\hat{\mathbf{M}}_r) > \mathrm{CM}_n(\hat{\mathbf{M}}_{r_0})$ for all $1 \le r < r_0$ because of the lack of fit.

One the other hand, when $r > r_0$, let $\mathbf{L}_1(r)$ represent the column orthogonal matrix $\mathbf{L}_1$ of order $p \times r$ and $\mathbf{L}_2(r)$ be the augmented orthonormal basis in $\mathbb{R}^p$. Then, we have

$$\hat{f}^{(-i)}(\mathbf{X}_i) = \frac{\sum_{j\neq i} Y_j K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{X}_i)}{\sum_{j\neq i} K_{\mathbf{M}}(\mathbf{X}_j - \mathbf{X}_i)},$$

where $\mathbf{M} = \mathbf{L}_1(r)\mathbf{H}^{-2}\mathbf{L}_1(r)^\top$. Following a similar derivation as Theorem 1, we have that as $n \to \infty$, $\|\mathbf{h}\| \to 0$ and $\|\mathbf{L}_0^\top \mathbf{L}_2(r)\| \to 0$, $\hat{f}^{(-i)}(\mathbf{X}_i)$ is also a consistent estimate for $f(\mathbf{X}_i)$. As a result, $\mathrm{CM}_n(\hat{\mathbf{M}}_r) = \tilde{\eta}_0 + o_p(1)$ for all $r_0 \le r \le p$. Therefore, we have $\mathrm{CM}_n(\hat{\mathbf{M}}_r)/\mathrm{CM}_n(\hat{\mathbf{M}}_{r_0}) \to_p 1$, for all

$r_0 \le r \le p$.

S1.5 Proofs of Lemmas

Proof of Lemma 1. Recall that $\mathbf{L} = (\mathbf{L}_1 \ \ \mathbf{L}_2)$ is a $p \times p$ orthogonal matrix. Analogue to (S1.6), we have

$$\mathrm{E}\{K_{\mathbf{M}}(\mathbf{X}-\mathbf{x})\} = \int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}} K(\|\mathbf{s}_1\|^2) f_{\mathbf{L}_2}(\mathbf{x}+\mathbf{L}_1\mathbf{H}\mathbf{s}_1)d\mathbf{s}_1. \tag{S1.12}$$

According to condition (C3) and the Taylor expansion of $f_{\mathbf{L}_2}(\mathbf{x}+\mathbf{L}_1\mathbf{H}\mathbf{s}_1)$, (S1.12) equals

$$\begin{aligned}
& f_{\mathbf{L}_2}(\mathbf{x}) + \left\{\dot{f}_{\mathbf{L}_2}(\mathbf{x})\right\}^\top \mathbf{L}_1\mathbf{H}\int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}} \mathbf{s}_1 K(\|\mathbf{s}_1\|^2)d\mathbf{s}_1 \qquad \text{(S1.13)}\\
& \quad + \frac{1}{2}\int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}} \mathbf{s}_1^\top \mathbf{H}^\top \mathbf{L}_1^\top \ddot{f}_{\mathbf{L}_2}(\mathbf{x})\mathbf{L}_1\mathbf{H}\mathbf{s}_1 K(\|\mathbf{s}_1\|^2)d\mathbf{s}_1 + o(\|\mathbf{h}\|^2)\\
= \;& f_{\mathbf{L}_2}(\mathbf{x}) + \frac{R_1(K)}{2}\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top \ddot{f}_{\mathbf{L}_2}(\mathbf{x})\mathbf{L}_1\mathbf{H}\} + o(\|\mathbf{h}\|^2)\\
= \;& f_{\mathbf{L}_2}(\mathbf{x}) + O(||\mathbf{h}||^2),
\end{aligned}$$

where $\dot{f}_{\mathbf{L}_2}(\mathbf{x}) = \partial f_{\mathbf{L}_2}(\mathbf{x})/\partial \mathbf{x}$ and $\ddot{f}_{\mathbf{L}_2}(\mathbf{x}) = (\partial^2/\partial \mathbf{x}^2) f_{\mathbf{L}_2}(\mathbf{x})$.

On the other hand, a similar calculation to (S1.12) and (S1.13) yields

$$\mathrm{E}\{K_{\mathbf{M}}^2(\mathbf{X}-\mathbf{x})\} = \frac{R_2(K)}{h_1\cdots h_{r_0}} f_{\mathbf{L}_2}(\mathbf{x})\{1+o(1)\}.$$

Consequently, $\mathrm{Var}\{n^{-1}\sum_{i=1}^n K_{\mathbf{M}}(\mathbf{X}_i - \mathbf{x})\} = O\{(nh_1\cdots h_{r_0})^{-1}\} = o(1)$. □

Proof of Lemma 2. By Lemma 1 and the continuous mapping theorem, for

$j \neq i$, we write

$$
\begin{aligned}
& \mathrm{E}\{(K^*_{j,i})^2 w_i \sigma_j^2\} \\
= \;& \int_{\mathbf{t}\in\mathbb{R}^p,\ \mathbf{x}\in\mathbb{R}^p} \frac{K^2_{\mathbf{M}}(\mathbf{t}-\mathbf{x}) w(\mathbf{x}) \sigma^2(\mathbf{t}) f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{t})}{n^2 f^2_{\mathbf{L}_2}(\mathbf{x})} d\mathbf{t} d\mathbf{x}\{1+o(1)\} \\
= \;& \frac{1}{n^2} \int_{\mathbf{x}\in\mathbb{R}^p} \left\{ \int_{\mathbf{t}\in\mathbb{R}^p} K^2_{\mathbf{M}}(\mathbf{t}-\mathbf{x}) \sigma^2(\mathbf{t}) f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t} \right\} \frac{f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x})}{f^2_{\mathbf{L}_2}(\mathbf{x})} d\mathbf{x}\{1+o(1)\} \\
= \;& \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{\mathbf{x}\in\mathbb{R}^p} \int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} \sigma^2(\mathbf{x}+\mathbf{L}_2\mathbf{s}_2) f_{\mathbf{X}}(\mathbf{x}+\mathbf{L}_2\mathbf{s}_2) d\mathbf{s}_2 \\
& \quad \times \frac{f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x})}{f^2_{\mathbf{L}_2}(\mathbf{x})} d\mathbf{x}\{1+o(1)\} \\
\equiv \;& \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{\mathbf{x}\in\mathbb{R}^p} \sigma^2_{\mathbf{L}_2}(\mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x})}{f^2_{\mathbf{L}_2}(\mathbf{x})} d\mathbf{x}\{1+o(1)\}, \qquad \text{(S1.14)}
\end{aligned}
$$

where the last two equalities hold by invoking $R_2(K) = \int_{\mathbf{s}\in\mathbb{R}^{r_0}} \{K(\|\mathbf{s}\|^2)\}^2 d\mathbf{s}$.

Noting that

$$
\begin{aligned}
\sigma^2(\mathbf{x}+\mathbf{L}_2\mathbf{s}_2) &= \sigma^2(\mathbf{L}_0^\top \mathbf{x} + \mathbf{L}_0^\top \mathbf{L}_2 \mathbf{s}_2) \\
&= \sigma^2(\mathbf{L}_0^\top \mathbf{x}) + \dot{\sigma}^2(\mathbf{L}_0^\top \mathbf{x})^\top \mathbf{L}_0^\top \mathbf{L}_2 \mathbf{s}_2 + o(||\mathbf{L}_0^\top \mathbf{L}_2||).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \mathrm{E}\{(K^*_{j,i})^2 w_i \sigma_j^2\} \\
= \;& \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{\mathbf{x}\in\mathbb{R}^p} \sigma^2_{\mathbf{L}_2}(\mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x})}{f^2_{\mathbf{L}_2}(\mathbf{x})} d\mathbf{x}\{1+o(1)\} \\
= \;& \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{\mathbf{x}\in\mathbb{R}^p} \sigma^2(\mathbf{L}_0^\top \mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x})}{f_{\mathbf{L}_2}(\mathbf{x})} d\mathbf{x}\{1+o(1)\} \\
= \;& \frac{R_2(K)}{n^2 h_1 \cdots h_{r_0}} \int_{\mathbf{x}\in\mathbb{R}^p} \sigma^2(\mathbf{L}_0^\top \mathbf{x}) \frac{f_{\mathbf{X}}(\mathbf{x}) w(\mathbf{x})}{f_{r_0}(\mathbf{L}_0^\top \mathbf{x})} d\mathbf{x}\{1+o(1)\} \\
= \;& \frac{R_2(K) V_0}{n^2 h_1 \cdots h_{r_0}}
\end{aligned}
$$

□

Proof of Lemma 3. It follows from $f(\mathbf{x}) = g(\mathbf{L}_0^\top \mathbf{x})$ that

$$\begin{aligned}
& \mathrm{E}[\{f(\mathbf{X}) - f(\mathbf{t})\} K_{\mathbf{M}}(\mathbf{X}-\mathbf{t})] \\
= & \frac{1}{h_1 \cdots h_{r_0}} \int_{\mathbf{s}\in\mathbb{R}^p} \{f(\mathbf{L}\mathbf{s}+\mathbf{t}) - f(\mathbf{t})\} K(\mathbf{s}_1^\top \mathbf{H}^{-2}\mathbf{s}_1) f_{\mathbf{X}}(\mathbf{L}\mathbf{s}+\mathbf{t}) d\mathbf{s} \\
= & \int_{\mathbf{s}_1\in\mathbb{R}^{r_0}, \mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} \{f(\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_2\mathbf{s}_2 + \mathbf{t}) - f(\mathbf{t})\} \\
& \times K(\|\mathbf{s}_1\|^2) f_{\mathbf{X}}(\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_2\mathbf{s}_2 + \mathbf{t}) d\mathbf{s}_1 d\mathbf{s}_2. \\
= & \int_{\mathbf{s}_1\in\mathbb{R}^{r_0}, \mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} \{g(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2 + \mathbf{L}_0^\top\mathbf{t}) - g(\mathbf{L}_0^\top\mathbf{t})\} \\
& \times K(\|\mathbf{s}_1\|^2) f_{\mathbf{X}}(\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_2\mathbf{s}_2 + \mathbf{t}) d\mathbf{s}_1 d\mathbf{s}_{21} d\mathbf{s}_{22}. \qquad \text{(S1.15)}
\end{aligned}$$

Now expanding both $g(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2+\mathbf{L}_0^\top\mathbf{t})$ and $f_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_1\mathbf{H}\mathbf{s}_1+\mathbf{L}_2\mathbf{s}_2)$ in Taylor expansions yield

$$\begin{aligned}
& g(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2 + \mathbf{L}_0^\top\mathbf{t}) - g(\mathbf{L}_0^\top\mathbf{t}) \\
= & \dot{g}(\mathbf{L}_0^\top\mathbf{t}_1)^\top(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2) \\
& + \frac{1}{2}(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2)^\top \ddot{g}(\mathbf{L}_0^\top\mathbf{t})(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + \mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2) + o(\|\mathbf{h}\|^2)
\end{aligned}$$

and

$$f_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_1\mathbf{H}\mathbf{s}_1+\mathbf{L}_2\mathbf{s}_2) = f_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2) + \dot{f}_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2)^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1 + o(\|\mathbf{h}\|).$$

Therefore, (S1.15) equals

$$
\begin{aligned}
& \int_{\mathbf{s}_1\in\mathbb{R}^{r_0},\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}} \big\{\dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2)+ \\
& \quad \frac{1}{2}(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2)^\top\ddot{g}(\mathbf{L}_0^\top\mathbf{t}_1)(\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2) \\
& \quad +o(\|\mathbf{h}\|^2)\big\}K(\|\mathbf{s}_1\|^2)\{f_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2)+\dot{f}_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2)^\top\mathbf{L}_1\mathbf{H}\mathbf{s}_1+o(\|\mathbf{h}\|)\}d\mathbf{s}_1d\mathbf{s}_2. \\
= & \quad \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\int_{\mathbf{s}_1\in\mathbb{R}^{r_0},\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}\mathbf{L}_2\mathbf{s}_2 f_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2 \\
& \quad +R_1(K)\dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}^2\mathbf{L}_1^\top\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}\dot{f}_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2 \\
& \quad +\frac{1}{2}R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top\mathbf{L}_0\ddot{g}(\mathbf{L}_0^\top\mathbf{t})\mathbf{L}_0^\top\mathbf{L}_1\mathbf{H}\}f_{\mathbf{L}_2}(\mathbf{t})+o(\|\mathbf{h}\|^2+\|\mathbf{L}_0^\top\mathbf{L}_2\|) \\
\equiv & \quad \Delta_1+\Delta_2+\Delta_3+o(\|\mathbf{h}\|^2+\|\mathbf{L}_0^\top\mathbf{L}_2\|)
\end{aligned}
$$

As $\|\mathbf{L}_0^\top\mathbf{L}_2\|\to 0$ and $\|\mathbf{h}\|\to 0$, taking $\mathbf{Q}^\top=(\mathbf{L}_0\ \mathbf{L}_0^\perp)$, we have

$$
\begin{aligned}
\Delta_1 &= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}\mathbf{s}_2 f_{\mathbf{X}}(\mathbf{t}+\mathbf{L}_2\mathbf{s}_2)d\mathbf{s}_2 \\
&= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}\mathbf{s}_2 f_{\mathbf{Q}}\begin{pmatrix}\mathbf{L}_0^\top\mathbf{t}+\mathbf{L}_0^\top\mathbf{L}_2\mathbf{s}_2\\ \mathbf{L}_0^{\perp\top}\mathbf{t}+\mathbf{L}_0^{\perp\top}\mathbf{L}_2\mathbf{s}_2\end{pmatrix}d\mathbf{s}_2 \\
&= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}\mathbf{s}_2 f_{\mathbf{Q}}\begin{pmatrix}\mathbf{L}_0^\top\mathbf{t}\\ \mathbf{L}_0^{\perp\top}\mathbf{t}+\mathbf{s}_2\end{pmatrix}d\mathbf{s}_2\{1+o(1)\} \\
&= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}(\mathbf{s}_2-\mathbf{L}_0^{\perp\top}\mathbf{t})f_{\mathbf{Q}}(\mathbf{L}_0^\top\mathbf{t},\mathbf{s}_2)d\mathbf{s}_2\{1+o(1)\} \\
&= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2\int_{\mathbf{s}_2\in\mathbb{R}^{(p-r_0)}}(\mathbf{s}_2-\mathbf{L}_0^{\perp\top}\mathbf{t})f_{\mathbf{u}_2|\mathbf{u}_1}(\mathbf{s}_2|\mathbf{L}_0^\top\mathbf{t})d\mathbf{s}_2 f_{r_0}(\mathbf{L}_0^\top\mathbf{t})\{1+o(1)\} \\
&= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2 E_{\mathbf{u}_2|\mathbf{u}_1}(\mathbf{U}_2-\mathbf{L}_0^{\perp\top}\mathbf{t}|\mathbf{U}_1=\mathbf{L}_0^\top\mathbf{t})f_{r_0}(\mathbf{L}_0^\top\mathbf{t})\{1+o(1)\} \\
&= \dot{g}(\mathbf{L}_0^\top\mathbf{t})^\top\mathbf{L}_0^\top\mathbf{L}_2\mathbf{b}(\mathbf{L}_0^\top\mathbf{t})\{1+o(1)\}, \qquad \text{(S1.16)}
\end{aligned}
$$

where $\mathbf{u}_1 \in \mathbb{R}^{r_0}$, $\mathbf{u}_2 \in \mathbb{R}^{(p-r_0)}$, $\mathbf{U}_1 \in \mathbb{R}^{r_0}$, $\mathbf{U}_2 \in \mathbb{R}^{(p-r_0)}$ and

$$\mathbf{b}(\mathbf{L}_0^\top \mathbf{t}) = E_{\mathbf{u}_2|\mathbf{u}_1}(\mathbf{U}_2 - \mathbf{L}_0^{\perp\top}\mathbf{t}|\mathbf{U}_1 = \mathbf{L}_0^\top \mathbf{t}) f_{r_0}(\mathbf{L}_0^\top \mathbf{t}).$$

For Δ_2, it is straightforward to show that

$$\Delta_2 = R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \dot{g}(\mathbf{L}_0^\top \mathbf{t}) \dot{f}_{r_0}(\mathbf{L}_0^\top \mathbf{t})^\top \mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}\}\{1 + o(1)\}, \quad \text{and}$$

$$\Delta_3 = \frac{1}{2} R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \ddot{g}(\mathbf{L}_0^\top \mathbf{t})\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}\} f_{r_0}(\mathbf{L}_0^\top \mathbf{t})\{1 + o(1)\}.$$

As a result,

$$\begin{aligned}
& E[\{f(\mathbf{X}) - f(\mathbf{t})\} K_{\mathbf{M}}(\mathbf{X} - \mathbf{t})] \\
= \; & \dot{g}(\mathbf{L}_0^\top \mathbf{t})^\top \mathbf{L}_0^\top \mathbf{L}_2 \mathbf{b}(\mathbf{L}_0^\top \mathbf{t}) + R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \dot{g}(\mathbf{L}_0^\top \mathbf{t}) \dot{f}_{r_0}(\mathbf{L}_0^\top \mathbf{t})^\top \mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}\} \\
& + \frac{1}{2} R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \ddot{g}(\mathbf{L}_0^\top \mathbf{t})\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}\} f_{r_0}(\mathbf{L}_0^\top \mathbf{t}) + o(\|\mathbf{h}\|^2 + \|\mathbf{L}_0^\top \mathbf{L}_2\|) \\
= \; & g(\mathbf{L}_0^\top \mathbf{t})^\top \mathbf{L}_0^\top \mathbf{L}_2 \mathbf{b}(\mathbf{L}_0^\top \mathbf{t}) + R_1(K)\mathrm{tr}\{\mathbf{H}\mathbf{L}_1^\top \mathbf{L}_0 \mathbf{A}(\mathbf{L}_0^\top \mathbf{t})\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}\} \\
& + o(\|\mathbf{h}\|^2 + \|\mathbf{L}_0^\top \mathbf{L}_2\|) \\
= \; & \psi(\mathbf{t}, \mathbf{h}, \mathbf{L}_1) + o(\|\mathbf{h}\|^2 + \|\mathbf{L}_0^\top \mathbf{L}_2\|),
\end{aligned}$$

where

$$\mathbf{A}(\mathbf{L}_0^\top \mathbf{t}) = \frac{1}{2}\ddot{g}(\mathbf{L}_0^\top \mathbf{t}) f_{r_0}(\mathbf{L}_0^\top \mathbf{t}) + \dot{g}(\mathbf{L}_0^\top \mathbf{t}) \dot{f}_{r_0}(\mathbf{L}_0^\top \mathbf{t})^\top.$$

□

Proof of Lemma 4. With an analogue calculation to Lemma 3, we have

$$
\begin{aligned}
&E[\{f(\mathbf{X}) - f(\mathbf{t})\} K_{\mathbf{M}}(\mathbf{X} - \mathbf{t})]^2 \\
&= \frac{1}{h_1 \cdots h_{r_0}} \int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}, \mathbf{s}_2 \in \mathbb{R}^{(p-r_0)}} \{g(\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H} \mathbf{s}_1 + \mathbf{L}_0^\top \mathbf{L}_2 \mathbf{s}_2 + \mathbf{L}_0^\top \mathbf{t}) - g(\mathbf{L}_0^\top \mathbf{t})\}^2 \\
&\quad \times K^2(\|\mathbf{s}_1\|^2) f_{\mathbf{X}}(\mathbf{L}_1 \mathbf{H} \mathbf{s}_1 + \mathbf{L}_2 \mathbf{s}_2 + \mathbf{t}) d\mathbf{s}_1 d\mathbf{s}_2 \\
&= \frac{1}{h_1 \cdots h_{r_0}} \int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}, \mathbf{s}_2 \in \mathbb{R}^{(p-r_0)}} \{\dot{g}(\mathbf{L}_0^\top \mathbf{t})^\top (\mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H} \mathbf{s}_1 + \mathbf{L}_0^\top \mathbf{L}_2 \mathbf{s}_2) + o(\|\mathbf{h}\| + \|\mathbf{L}_0^\top \mathbf{L}_2\|)\}^2 \\
&\quad \times K^2(\|\mathbf{s}_1\|^2) \{f_{\mathbf{X}}(\mathbf{t} + \mathbf{L}_2 \mathbf{s}_2) + O(\|\mathbf{h}\|)\} d\mathbf{s}_1 d\mathbf{s}_2.
\end{aligned}
$$

Recall that $\int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}} \mathbf{s}_1 K^2(\|\mathbf{s}_1\|^2) d\mathbf{s}_1 = \mathbf{0}$ and $\int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}} \mathbf{s}_1 \mathbf{s}_1^\top K^2(\|\mathbf{s}_1\|^2) d\mathbf{s}_1$ exists.

Let

$$
\int_{\mathbf{s}_1 \in \mathbb{R}^{r_0}} \mathbf{s}_1 \mathbf{s}_1^\top K^2(\|\mathbf{s}_1\|^2) d\mathbf{s}_1 = c_2 \mathbf{I}_{r_0 \times r_0}
$$

for some $c_2 \geq 0$. It follows that

$$
\begin{aligned}
& \mathrm{Var}[\{f(\mathbf{X}) - f(\mathbf{t})\} K_{\mathbf{M}}(\mathbf{X} - \mathbf{t})] \\
\leq\ & \mathrm{E}[\{f(\mathbf{X}) - f(\mathbf{t})\} K_{\mathbf{M}}(\mathbf{X} - \mathbf{t})]^2 \\
=\ & \frac{c_2}{h_1 \cdots h_{r_0}} \mathrm{tr}\Big\{\mathbf{H} \mathbf{L}_1^\top \mathbf{L}_0 \dot{g}(\mathbf{L}_0^\top \mathbf{t}) \dot{g}(\mathbf{L}_0^\top \mathbf{t})^\top \mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}\Big\} f_{\mathbf{L}_2}(\mathbf{t}) \{1 + o(1)\} \\
=\ & \frac{c_2}{h_1 \cdots h_{r_0}} \dot{g}(\mathbf{L}_0^\top \mathbf{t})^\top \mathbf{L}_0^\top \mathbf{L}_1 \mathbf{H}^2 \mathbf{L}_1^\top \mathbf{L}_0 \dot{g}(\mathbf{L}_0^\top \mathbf{t}) f_{\mathbf{L}_2}(\mathbf{t}) \{1 + o(1)\} \\
=\ & O\left(\frac{\|\mathbf{h}\|^2}{h_1 \cdots h_{r_0}}\right).
\end{aligned}
$$

□

Reference

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