Supplement to "Smoothing Spline Semiparametric Density Models"

Jiahui Yu Jian Shi Anna Liu Yuedong Wang*

In this document, we provide technical proofs of some results in the main paper, a theory for local existence and uniqueness of the penalized likelihood estimators for our proposed semiparametric density models, and two additional simulation results. The verification of inner product $\langle \cdot, \cdot \rangle_{\mathcal{Q}}$ is in Section S.1. The proof of Proposition 2 for the verification of Assumptions 4 and 5 when $L_{\boldsymbol{\theta}} = D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}_0, h_0)$ is in Section S.2. The intermediate results Lemma 1 and Theorem 3 for the convergence $\tilde{\tau} - \tau_0$ are proved in Sections S.3 and S.4. Section S.5 shows local existence and uniqueness of the penalized likelihood estimators. Section S.6 presents two additional simulation results.

S.1 Verification of inner product $\langle \cdot, \cdot \rangle_{\mathcal{Q}}$

We now turn to the discussion of the validity of (3.2). For any $p_1 \times 1$ and $p_2 \times 1$ vectors of functions in $\mathcal{H}$, say $\boldsymbol{G}_1 = [G_1^k]_{k=1}^{p_1}$ and $\boldsymbol{G}_2 = [G_2^k]_{k=1}^{p_2}$, we use the vector form of the inner product $\langle \boldsymbol{G}_1, \boldsymbol{G}_2 \rangle_{\mathcal{H}}$ to denote a $p_1 \times p_2$ matrix in which the (i,j)th entry is $\langle G_1^i, G_2^j \rangle_{\mathcal{H}}$. For any $g \in \mathcal{H}$, let $\mathcal{F}^k g = \sum_{l=1}^m V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}^k, L_{l,h} g]$. For any fixed $\lambda > 0$, since

$$\left|\mathcal{F}^k g\right| \leq \left[\sum_{l=1}^m V_{l,\tau_0}(L_{l,\boldsymbol{\theta}}^k)\right]^{\frac{1}{2}} \left[\sum_{l=2}^m V_{l,\tau_0}(L_{l,h} g)\right]^{\frac{1}{2}} \leq C \left\|g\right\|_{\mathcal{H}}$$

*Jiahui Yu (email: jyu32@bu.edu) is Postdoctoral Research Associate, Department of Mathematics and Statistics, Boston University, Boston, MA 02215; Jian Shi (email: shijiannk@gmail.com) is Data Scientist, PayPal, San Jose, CA 95131; Anna Liu (email: anna@math.umass.edu) is Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003; and Yuedong Wang (email: yuedong@pstat.ucsb.edu) is Professor, Department Statistics and Applied Probability, University of California, Santa Barbara, CA 93106. This research is supported by National Science Foundation grants DMS-1507078 for Anna Liu, and DMS-1507620 for Yuedong Wang. The authors gratefully acknowledge support from the Center for Scientific Computing from the CNSI, MRL: an NSF MRSEC (DMR-1720256). The authors would also like to thank the associate editor and two referees for constructive comments that substantially improved an earlier draft.

for some positive constant C, $\mathcal{F}^k$ is a bounded linear functional on $\mathcal{H}$. By the Riesz representation theorem, there exists a $F^k \in \mathcal{H}$ such that for any $g \in \mathcal{H}$, $\mathcal{F}^k g = \langle F^k, g\rangle_{\mathcal{H}}$. Let $\boldsymbol{F} = \left[F^k\right]_{k=1}^p$. We define $V_{l,\tau_0}(L_{l,\boldsymbol{\theta}}, L_{l,h}g)$, $V_{l,\tau_0}(L_{l,h}\boldsymbol{F}, L_{l,h}g)$, and $J(\boldsymbol{F}, g)$ to be $p \times 1$ vectors whose kth entries are $V_{l,\tau_0}(L^k_{l,\boldsymbol{\theta}}, L_{l,h}g)$, $V_{l,\tau_0}(L_{l,h}F^k, L_{l,h}g)$, and $J(F^k, g)$, respectively. Therefore,

$$\sum_{l=1}^m V_{l,\tau_0}(L_{l,\boldsymbol{\theta}}, L_{l,h}g) = \sum_{l=1}^m V_{l,\tau_0}(L_{l,h}\boldsymbol{F}, L_{l,h}g) + \lambda J(\boldsymbol{F}, g) = \langle \boldsymbol{F}, g\rangle_{\mathcal{H}}.$$

We also define the $p \times p$ matrix $\boldsymbol{\Omega}_F = \sum_{l=1}^m V_{l,\tau_0}(L_{l,\boldsymbol{\theta}} - L_{l,h}\boldsymbol{F}, L_{l,\boldsymbol{\theta}} - L_{l,h}\boldsymbol{F})$, whose (i,j)th entry is $\sum_{l=1}^m V_{l,\tau_0}(L^i_{l,\theta} - L_{l,h}F^i, L^j_{l,\theta} - L_{l,h}F^j)$.

Lemma S.1. *Under Assumption 2, $\boldsymbol{\Omega}_F$ is positive definite and the eigenvalues of $\boldsymbol{\Omega}_F$ are greater than c_δ (see Assumption 2(ii)).*

Proof. Fix a non-zero vector $\boldsymbol{\zeta} \in \mathbb{R}^p$ and write $\boldsymbol{\zeta_*} = \boldsymbol{\zeta}/\left\|\boldsymbol{\zeta}\right\|_{l^2}$. We have

$$\begin{aligned}\boldsymbol{\zeta}^T\boldsymbol{\Omega}_F\boldsymbol{\zeta} &= \sum_{l=1}^m \boldsymbol{\zeta}^T V_{l,\tau_0}(L_{l,\boldsymbol{\theta}} - L_{l,h}\boldsymbol{F}, L_{l,\boldsymbol{\theta}} - L_{l,h}\boldsymbol{F})\boldsymbol{\zeta} \\ &= \|\boldsymbol{\zeta}\|^2_{l^2} \sum_{l=1}^m V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}\boldsymbol{\zeta_*} - L_{l,h}(\boldsymbol{\zeta_*}^T\boldsymbol{F}), L_{l,\boldsymbol{\theta}}\boldsymbol{\zeta_*} - L_{l,h}(\boldsymbol{\zeta_*}^T\boldsymbol{F})] \\ &> c_\delta \left\|\boldsymbol{\zeta}\right\|^2_{l^2},\end{aligned}$$

where the last inequality holds by Assumption 2(ii) because $\|\boldsymbol{\zeta_*}\|_{l^2} = 1$ and $\boldsymbol{\zeta_*}^T\boldsymbol{F} \in \mathcal{H}$. Therefore, $\boldsymbol{\Omega}_F$ is positive definite.

Let δ be any eigenvalue of $\boldsymbol{\Omega}_F$, and let $\boldsymbol{\zeta_\delta} \in \mathbb{R}^p$ be a unit eigenvector associated with δ. By definition, we have $\boldsymbol{\Omega}_F\boldsymbol{\zeta_\delta} = \delta\boldsymbol{\zeta_\delta}$. We have

$$\delta = \boldsymbol{\zeta_\delta}^T\delta\boldsymbol{\zeta_\delta} = \boldsymbol{\zeta_\delta}^T\boldsymbol{\Omega}_F\boldsymbol{\zeta_\delta} = \sum_{l=1}^m V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}\boldsymbol{\zeta_\delta} - L_{l,h}(\boldsymbol{\zeta_\delta}^T\boldsymbol{F}), L_{l,\boldsymbol{\theta}}\boldsymbol{\zeta_*} - L_{l,h}(\boldsymbol{\zeta_*}^T\boldsymbol{F})] > c_\delta.$$

□

Theorem S.1. *Suppose Assumption 2 holds. Then $\langle\cdot,\cdot\rangle_{\mathcal{Q}}$ given by* (3.2) *is a well-defined inner product on $\mathcal{Q}$, and $\mathcal{Q}$ is complete with respect to the norm $\left\|\cdot\right\|_{\mathcal{Q}}$ induced by this inner product. Hence, $\mathcal{Q}$ is a Hilbert space.*

Proof. It is easy to check that (3.2) satisfies symmetry, linearity and positive semi-definiteness for an inner product. If $(\boldsymbol{\zeta}, g) = 0$, $\langle(\boldsymbol{\zeta}, g), (\boldsymbol{\zeta}, g)\rangle_{\mathcal{Q}} = 0$ is obvious. We will now show that $\langle(\boldsymbol{\zeta}, g), (\boldsymbol{\zeta}, g)\rangle_{\mathcal{Q}} = 0$ implies $(\boldsymbol{\zeta}, g) = 0$. We see that

$$\begin{aligned}
\langle(\boldsymbol{\zeta}, g), (\boldsymbol{\zeta}, g)\rangle_{\mathcal{Q}} &= \sum_{l=1}^{m} V_{l,\tau_0}(L_{l,\boldsymbol{\theta}}\boldsymbol{\zeta} + L_{l,h}g, L_{l,\boldsymbol{\theta}}\boldsymbol{\zeta} + L_{l,h}g) + \lambda J(g, g) \\
&= \sum_{l=1}^{m} \boldsymbol{\zeta}^T V_{l,\tau_0}(L_{l,\boldsymbol{\theta}} - L_{l,h}\boldsymbol{F}, L_{l,\boldsymbol{\theta}} - L_{l,h}\boldsymbol{F})\boldsymbol{\zeta} \\
&\quad - \sum_{l=1}^{m} \boldsymbol{\zeta}^T \left[V_{l,\tau_0}(L_{l,h}\boldsymbol{F}, L_{l,h}\boldsymbol{F}) - 2V_{l,\tau_0}(L_{l,\boldsymbol{\theta}}, L_{l,h}\boldsymbol{F})\right]\boldsymbol{\zeta} \\
&\quad + 2\boldsymbol{\zeta}^T \sum_{l=1}^{m} V_{l,\tau_0}(L_{l,\boldsymbol{\theta}}, L_{l,h}g) + \langle g, g\rangle_{\mathcal{H}} \\
&= \boldsymbol{\zeta}^T \boldsymbol{\Omega}_F \boldsymbol{\zeta} + \langle \boldsymbol{\zeta}^T\boldsymbol{F} + g, \boldsymbol{\zeta}^T\boldsymbol{F} + g\rangle_{\mathcal{H}} + \lambda J(\boldsymbol{\zeta}^T\boldsymbol{F}, \boldsymbol{\zeta}^T\boldsymbol{F}),
\end{aligned} \tag{S.1}$$

and every term in (S.1) is non-negative. If $\langle(\boldsymbol{\zeta}, g), (\boldsymbol{\zeta}, g)\rangle = 0$, the first term in (S.1) implies $\boldsymbol{\zeta} = 0$ by Lemma S.1. This further implies that

$$\langle \boldsymbol{\zeta}^T\boldsymbol{F} + g, \boldsymbol{\zeta}^T\boldsymbol{F} + g\rangle_{\mathcal{H}} = \langle g, g\rangle_{\mathcal{H}} = 0.$$

Therefore, $g = 0$ because $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is an inner product on $\mathcal{H}$. Hence, $\langle\cdot,\cdot\rangle_{\mathcal{Q}}$ is a well-defined inner product on $\mathcal{Q}$.

Next, we want to show that $\mathcal{Q}$ is complete with respect to the norm $\|\cdot\|_{\mathcal{Q}}$. Let $\{(\boldsymbol{\zeta}_i, g_i)\}_{i=1}^{\infty} \subset \mathcal{Q}$ be a Cauchy sequence. For any $\epsilon > 0$, there exist a positive integer M such that for all $i, j > M$, we have

$$\|(\boldsymbol{\zeta}_i, g_i) - (\boldsymbol{\zeta}_j, g_j)\|_{\mathcal{Q}}^2 = \sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j) + L_{l,h}(g_i - g_j)] + \lambda J(g_i - g_j) \le \epsilon.$$

This implies that

$$\sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j) + L_{l,h}(g_i - g_j)] = \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{l^2}^2 \sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)^* + L_{l,h}(g_i - g_j)^*] \le \epsilon,$$

where $(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)^* = (\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)/\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{l^2}$, and $(g_i - g_j)^* = (g_i - g_j)/\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{l^2}$. By Assumption 2(*ii*), $\sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)^* + L_{l,h}(g_i - g_j)^*] > c_\delta$ for some positive constant c_δ as defined in (3.1) Therefore,

$$\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{l^2}^2 \le \frac{\epsilon}{c_\delta}, \tag{S.2}$$

and $\{\boldsymbol{\zeta}_i\}_{i=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}^p$ under the Euclidean norm, which therefore converges to some limit $\boldsymbol{\zeta}_\infty \in \mathbb{R}^p$.

To find a limit for the sequence $\{g_i\}_{i=1}^{\infty}$ in $\mathcal{H}$, we consider

$$\begin{aligned}
&\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2 + \|g_i - g_j\|_{\mathcal{H}}^2 - 2\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}\|g_i - g_j\|_{\mathcal{H}} \\
&\le \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2 + \|g_i - g_j\|_{\mathcal{H}}^2 - 2\left|\sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j), L_{l,h}(g_i - g_j)]\right| \\
&\le \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2 + \|g_i - g_j\|_{\mathcal{H}}^2 + 2\sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j), L_{l,h}(g_i - g_j)] \qquad \text{(S.3)} \\
&= \sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j) + L_{l,h}(g_i - g_j)] + \lambda J(g_i - g_j) \\
&= \|(\boldsymbol{\zeta}_i, g_i) - (\boldsymbol{\zeta}_j, g_j)\|_{\mathcal{Q}}^2 \le \epsilon
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from the triangle inequality. For $a > 0, b > 0$, we have $(1/4)a + b - a^{1/2}b^{1/2} = [(1/2)a^{1/2} - b^{1/2}]^2 \ge 0$, and it follows that $2a^{1/2}b^{1/2} \le (1/2)a + 2b$. For $a = \|g_i - g_j\|_{\mathcal{H}}^2$ and $b = \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2$, (S.3) becomes

$$\begin{aligned}
\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2 + \|g_i - g_j\|_{\mathcal{H}}^2 &\le \epsilon + 2\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}\|g_i - g_j\|_{\mathcal{H}} \\
&\le \epsilon + \frac{1}{2}\|g_i - g_j\|_{\mathcal{H}}^2 + 2\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2 .
\end{aligned} \qquad \text{(S.4)}$$

Since $\|\cdot\|_{\mathbb{R}^p}$ is equivalent to $\|\cdot\|_{l^2}$ on $\mathbb{R}^p$, $\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_{\mathbb{R}^p}^2 \le C\epsilon$ for some positive constant C by (S.2). Therefore, after rearranging (S.4), we get $\|g_i - g_j\|_{\mathcal{H}}^2 \le (2 + C)\epsilon$. Hence, $\{g_i\}_{i=1}^{\infty}$ is a Cauchy sequence in $\mathcal{H}$ under the norm $\|\cdot\|_{\mathcal{H}}$. By Assumption 2(i), this sequence converges to some limit $g_\infty \in \mathcal{H}$.

Lastly, we show that $(\boldsymbol{\zeta}_i, g_i)$ converges to $(\boldsymbol{\zeta}_\infty, g_\infty)$ in $\|\cdot\|_{\mathcal{Q}}$. By the Cauchy-Schwarz inequality and triangle inequality, as $i \to \infty$, we have

$$\begin{aligned}
&\|(\boldsymbol{\zeta}_i, g_i) - (\boldsymbol{\zeta}_\infty, g_\infty)\|_{\mathcal{Q}}^2 \\
&= \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_\infty\|_{\mathbb{R}^p}^2 + \|g_i - g_\infty\|_{\mathcal{H}}^2 + 2\sum_{l=1}^{m} V_{l,\tau_0}[L_{l,\boldsymbol{\theta}}(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_\infty), L_{l,h}(g_i - g_\infty)] \\
&\le \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_\infty\|_{\mathbb{R}^p}^2 + \|g_i - g_\infty\|_{\mathcal{H}}^2 + 2\|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_\infty\|_{\mathbb{R}^p}\|g_i - g_\infty\|_{\mathcal{H}} \to 0.
\end{aligned}$$

Therefore, we conclude that $\mathcal{Q}$ is a Hilbert space with respect to the inner product $\langle\cdot,\cdot\rangle_{\mathcal{Q}}$. $\square$

S.2 Proof of Proposition 2

Proof. Denote by $\mathscr{L}(\mathbb{R}^p, L_0^2(\mathcal{X}))$ the space of bounded linear operators from $\mathbb{R}^p$ to $L_0^2(\mathcal{X})$ equipped with the operator norm $\|\cdot\|_{op}$, and denote by $S_{\boldsymbol{\theta},\boldsymbol{\theta}_0}(R)$, $S_{h,h_0}(R)$ the balls of radius R in $\mathbb{R}^p$ and $\mathcal{H}$ centered at $\boldsymbol{\theta}_0$ and h_0, respectively. Since $D_{\boldsymbol{\theta}}\eta : U \to \mathscr{L}(\mathbb{R}^p, L_0^2(\mathcal{X}))$ is continuous, given any $\epsilon > 0$, there exist $R_1, R_2 > 0$ such that for any $\tau_1 \in S_{\boldsymbol{\theta},\boldsymbol{\theta}_0}(R_1) \times S_{h,h_0}(R_2)(= \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0})$, we have $\|D_{\boldsymbol{\theta}}\eta(\tau_1) - D_{\boldsymbol{\theta}}\eta(\tau_0)\|_{op} \leq \epsilon$. This implies that for any $\boldsymbol{\zeta} \in \mathbb{R}^p$ with $\|\boldsymbol{\zeta}\|_{l^2} = 1$, $V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_1)\boldsymbol{\zeta} - D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}] \leq \epsilon$, and

$$V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}] - \epsilon \leq V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_1)\boldsymbol{\zeta}] \leq V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}] + \epsilon. \tag{S.5}$$

Let $m = \inf_{\|\boldsymbol{\zeta}\|_{l^2}=1} V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}]$. Since $D_{\boldsymbol{\theta}}\eta(\tau_0)$ has zero null space and its domain $\mathbb{R}^p$ is of finite dimension, we have $m > 0$. For $0 < \epsilon < m$, let $\sqrt{C_1} = 1 - \epsilon/m > 0$ and $\sqrt{C_2} = 1 + \epsilon/m > 0$. Then (S.5) implies

$$C_1 V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}] \leq V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_1)\boldsymbol{\zeta}] \leq C_2 V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}].$$

Note that for any nonzero $\boldsymbol{\zeta} \in \mathbb{R}^p$, one can write $V_{\tau_0}[D_\theta\eta(\tau_j)\boldsymbol{\zeta}] = \|\boldsymbol{\zeta}\|_{l^2}^2 V_{\tau_0}[D_\theta\eta(\tau_j)(\boldsymbol{\zeta}/\|\boldsymbol{\zeta}\|_{l^2})]$ for $j = 0, 1$, part (i) holds by the above inequality.

For any $\tau_1, \tau_2 \in S_{\boldsymbol{\theta},\boldsymbol{\theta}_0}(R_1) \times S_{h,h_0}(R_2)$,

$$\|D_{\boldsymbol{\theta}}\eta(\tau_1) - D_{\boldsymbol{\theta}}\eta(\tau_2)\|_{op} \leq \|D_{\boldsymbol{\theta}}\eta(\tau_1) - D_{\boldsymbol{\theta}}\eta(\tau_0)\|_{op} + \|D_{\boldsymbol{\theta}}\eta(\tau_2) - D_{\boldsymbol{\theta}}\eta(\tau_0)\|_{op} \leq 2\epsilon.$$

Let $\sqrt{C_d} = 2\epsilon/m$. Since for any $\boldsymbol{\zeta} \in \mathbb{R}^p$, $V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}] \geq m\|\boldsymbol{\zeta}\|_{l^2}$, we have

$$V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_1)\boldsymbol{\zeta} - D_{\boldsymbol{\theta}}\eta(\tau_2)\boldsymbol{\zeta}] \leq 2\epsilon\|\boldsymbol{\zeta}\|_{l^2} \leq \sqrt{C_d} V_{\tau_0}^{\frac{1}{2}}[D_{\boldsymbol{\theta}}\eta(\tau_0)\boldsymbol{\zeta}].$$

Note that as $\epsilon \to 0^+$, $2C_1 \to 2^-$ and $C_d \to 0^+$. Thus, we can choose $0 < \epsilon < m$ such that $0 < C_d < 2C_1$ and (ii) holds. $\square$

S.3 Proof of Lemma 1

We now give detailed calculations and a proof of Lemma 1. Let $\mathcal{G}^k g = V_{\tau_0}[D_{\theta^k}\eta(\tau_0), D_h\eta(\tau_0)g]$ for any $g \in \mathcal{H}$. Since $|\mathcal{G}^k g| \leq V_{\tau_0}^{\frac{1}{2}}[D_{\theta^k}\eta(\tau_0)]V_{\tau_0}^{\frac{1}{2}}[D_h\eta(\tau_0)g] \leq C\|g\|_{\mathcal{H}}$ for some positive constant C, $\mathcal{G}^k$ is a bounded linear functional on $\mathcal{H}$. By the Riesz representation theorem, there exists $G^k \in \mathcal{H}$ such that for any $g \in \mathcal{H}$,

$$\mathcal{G}^k g = V_{\tau_0}[D_{\theta^k}\eta(\tau_0), D_h\eta(\tau_0)g] = V_{\tau_0}[D_h\eta(\tau_0)G^k, D_h\eta(\tau_0)g] + \lambda J(G^k, g).$$

Let $\boldsymbol{G} = \left[G^k\right]_{k=1}^p$. Define $V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)g]$, $V_{\tau_0}[D_h\eta(\tau_0)\boldsymbol{G}, D_h\eta(\tau_0)g]$, and $J(\boldsymbol{G}, g)$ to be $p \times 1$ vectors whose kth entries are $V_{\tau_0}[D_{\theta^k}\eta(\tau_0), D_h\eta(\tau_0)g]$, $V_{\tau_0}[D_h\eta(\tau_0)G^k, D_h\eta(\tau_0)g]$, and $J(G^k, g)$, respectively. Therefore,

$$V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)g] = V_{\tau_0}[D_h\eta(\tau_0)\boldsymbol{G}, D_h\eta(\tau_0)g] + \lambda J(\boldsymbol{G}, g).$$

The Fourier expansion of G^k with respect to the eigensystem discussed in the appendix following Assumptions 7 and 8 is $G^k = \sum_\nu V_{\tau_0}[D_h\eta(\tau_0)G^k, D_h\eta(\tau_0)\phi_{0,\nu}]\phi_{0,\nu}$. A simple calculation shows that

$$V_{\tau_0}[D_h\eta(\tau_0)G^k, D_h\eta(\tau_0)\phi_{0,\nu}] = (1+\lambda\rho_{0,\nu})^{-1}V_{\tau_0}[D_{\theta^k}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}], \tag{S.6}$$

and hence,

$$G^k = \sum_\nu \frac{1}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\theta^k}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\phi_{0,\nu}. \tag{S.7}$$

Recall from Section A.4.1 that

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \boldsymbol{\Omega}_\lambda^{-1}\left\{\bar{\boldsymbol{\alpha}}_n - \sum_\nu \frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu}h_{0,\nu}}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right\},$$

where

$$\boldsymbol{\Omega}_\lambda = V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0)] - \sum_\nu \frac{V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]^{\otimes 2}}{1+\lambda\rho_{0,\nu}}.$$

The (i,j)th entry of this $p \times p$ matrix $\boldsymbol{\Omega}_\lambda$ can be written as

$$\begin{aligned}\boldsymbol{\Omega}_\lambda^{i,j} &= V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_{\theta^j}\eta(\tau_0)] - V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)G^j] \\ &= V_{\tau_0}[D_{\theta^i}\eta(\tau_0) - D_h\eta(\tau_0)G^i, D_{\theta^j}\eta(\tau_0) - D_h\eta(\tau_0)G^j] + \lambda J(G^i, G^j).\end{aligned}$$

Let $\boldsymbol{\Omega} = V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0) - D_h\eta(\tau_0)G]$ and $\boldsymbol{\Sigma}_\lambda = \lambda J(\boldsymbol{G})$ be the matrices such that

$$\Omega^{i,j} = V_{\tau_0}[D_{\theta^i}\eta(\tau_0) - D_h\eta(\tau_0)G^i, D_{\theta^j}\eta(\tau_0) - D_h\eta(\tau_0)G^j],$$

and $\Sigma_\lambda^{i,j} = \lambda J(G^i, G^j)$. Thus, $\boldsymbol{\Omega}_\lambda = \boldsymbol{\Omega} + \boldsymbol{\Sigma}_\lambda$.

We now prove some properties of $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_\lambda$, which will be used to establish the bound for $\mathrm{E}[(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)]$.

Lemma S.2. $\mathbf{\Sigma}_\lambda \to 0$ *as* $\lambda \to 0$.

Proof. By (S.7), we get that the (i,j)th entry of $\mathbf{\Sigma}_\lambda$ is

$$\sum_\nu a_\nu^\lambda = \sum_\nu \frac{\lambda \rho_{0,\nu}}{(1+\lambda\rho_{0,\nu})^2} V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}] V_{\tau_0}[D_{\theta^j}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]. \tag{S.8}$$

We have $\sum_\nu a_\nu^\lambda \le \sum_\nu |a_\nu^\lambda| \le \sum_\nu b_\nu$, where

$$\begin{aligned}\sum_\nu b_\nu &= \sum_\nu |V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}] V_{\tau_0}[D_{\theta^j}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]| \\ &\le \left\{\sum_\nu V_{\tau_0}^2[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right\}^{\frac{1}{2}} \left\{\sum_\nu V_{\tau_0}^2[D_{\theta^j}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right\}^{\frac{1}{2}}.\end{aligned}$$

Since $V_{\tau_0}^{\frac{1}{2}}(\cdot)$ is a norm on $L^2(\mathcal{X}) \ominus \{1\}$, by the definition of $\{\phi_{0,\nu}\}_{\nu\in\mathbb{N}}$, it follows that $\{D_h\eta(\tau_0)\phi_{0,\nu}\}_{\nu\in\mathbb{N}}$ is an orthonormal basis on $L^2(\mathcal{X}) \ominus \{1\}$ with respect to $V_{\tau_0}^{\frac{1}{2}}(\cdot)$. Since $D_{\theta^i}\eta(x;\tau_0) \in L^2(\mathcal{X}) \ominus \{1\}$, for any $i = 1, \ldots, p$, we have

$$V_{\tau_0}[D_{\theta^i}\eta(\tau_0)] = \sum_\nu V_{\tau_0}^2[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}] < \infty.$$

Hence, $\{V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\}_{\nu\in\mathbb{N}}$ is square summable and $\sum_\nu b_\nu < \infty$. By the dominated convergence theorem, as $\lambda \to 0$, $\lambda J(G^i, G^j) = \sum_\nu a_\nu^\lambda \to 0$, because $a_\nu^\lambda \to 0$ for any fixed $\nu \in \mathbb{N}$.

□

Lemma S.3. *Under Assumption 2,* $\mathbf{\Omega}$ *is positive definite. Let* c_δ *and* C_1 *be constants defined in Assumptions 2 and 4 respectively, and let* δ *be any eigenvalue of* $\mathbf{\Omega}$. *Then* $\delta > C_1 c_\delta = \tilde{c}_\delta$.

Proof. The proof is similar to the one for Lemma S.1. □

Note that by Lemma S.3, the eigenvalues of $\mathbf{\Omega}$ have a uniform lower bound independent of λ. Then as $\lambda \to 0$, we have

$$\begin{aligned}\mathrm{E}\left[(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right] &= \mathrm{E}\left[\boldsymbol{a}_n^T(\mathbf{\Omega} + \mathbf{\Sigma}_\lambda)^{-2}\boldsymbol{a}_n\right] \\ &\to \mathrm{E}\left[\boldsymbol{a}_n^T\mathbf{\Omega}^{-2}\boldsymbol{a}_n\right] \le \tilde{c}_\delta^{-2}\,\mathrm{E}\left[\boldsymbol{a}_n^T\boldsymbol{a}_n\right] = \tilde{c}_\delta^{-2}\sum_{i=1}^p \mathrm{E}\left[(a_n^i)^2\right],\end{aligned}$$

where

$$\boldsymbol{a}_n = \bar{\boldsymbol{\alpha}}_n - \sum_\nu \frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu} h_{0,\nu}}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}],$$

and a_n^i is the ith entry of $\boldsymbol{a}_n$.

Before we proceed to derive the bound for $\mathrm{E}\,[(a_n^i)^2]$, we also need the following lemma, whose proof is given by Lemma 5.2 in Gu and Qiu (1993).

Lemma S.4. *Under Assumption 8, as $\lambda \to 0$,*

$$\sum_\nu \frac{\lambda\rho_{0,\nu}}{(1+\lambda\rho_{0,\nu})^2} = O(\lambda^{-\frac{1}{r}}), \quad \sum_\nu \frac{1}{(1+\lambda\rho_{0,\nu})^2} = O(\lambda^{-\frac{1}{r}}), \quad \sum_\nu \frac{1}{1+\lambda\rho_{0,\nu}} = O(\lambda^{-\frac{1}{r}}).$$

We now ready to establish the upper bound for $\mathrm{E}\,[(a_n^i)^2]$. We have

$$\begin{aligned}\mathrm{E}\left[(a_n^i)^2\right] &= \mathrm{E}\left\{\left[\bar{\alpha}_n^i - \sum_\nu \frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu} h_{0,\nu}}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right]^2\right\} \\ &\le 2\,\mathrm{E}\left[(\bar{\alpha}_n^i)^2\right] + 2\,\mathrm{E}\left[\left(\sum_\nu \frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu} h_{0,\nu}}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right)^2\right].\end{aligned}$$

Note that $\mathrm{E}\,[(\bar{\alpha}_n^i)^2] = O(n^{-1})$, and by square summability of $\{V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\}_{\nu\in\mathbb{N}}$ (See proof of Lemma S.2) and $\{h_{0,\nu}\}_{\nu\in\mathbb{N}}$, the dominated convergence theorem, the Cauchy-Schwarz inequality, $\mathrm{E}(\bar{\beta}_{\nu,n}^2) = n^{-1}$, and Lemma S.4, we have

$$\begin{aligned}&\mathrm{E}\left[\left(\sum_\nu \frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu} h_{0,\nu}}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right)^2\right] \\ &\quad \to \mathrm{E}\left[\left(\sum_\nu \frac{\bar{\beta}_{\nu,n}}{1+\lambda\rho_{0,\nu}} V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]\right)^2\right] \\ &\quad \le C\left[\sum_\nu \frac{\mathrm{E}(\bar{\beta}_{\nu,n}^2)}{(1+\lambda\rho_{0,\nu})^2}\right] = O(n^{-1}\lambda^{-\frac{1}{r}}).\end{aligned}$$

Therefore, we conclude that as $\lambda \to 0$, $\mathrm{E}\,[(a_n^i)^2] = O(n^{-1}\lambda^{-\frac{1}{r}})$, which implies that

$$\mathrm{E}\left\{V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)]\right\} \le c\,\mathrm{E}\left[(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right] = O(n^{-1}\lambda^{-\frac{1}{r}}).$$

This concludes the proof for the first bound in Lemma 1.

Now for the second bound in Lemma 1, we see that

$$V_{\tau_0}[D_h\eta(\tau_0)(\tilde{h} - h_0)] = \sum_\nu \left(\tilde{h}_\nu - h_{0,\nu}\right)^2, \lambda J(\tilde{h} - h_0) = \sum_\nu \lambda\rho_\nu^0 \left(\tilde{h}_\nu - h_{0,\nu}\right)^2.$$

Plugging in the formula of $\tilde{h}_\nu$ given in Section A.4.1, we get

$$\begin{aligned}\sum_\nu \left(\tilde{h}_\nu - h_{0,\nu}\right)^2 &= \sum_\nu \left\{ \frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu}h_{0,\nu}}{1+\lambda\rho_{0,\nu}} - (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]}{1+\lambda\rho_{0,\nu}} \right\}^2 \\ &\leq C\left[(I_h) + (II_h)\right],\end{aligned}$$

where

$$\begin{aligned}(I_h) &= \sum_\nu \left(\frac{\bar{\beta}_{\nu,n} - \lambda\rho_{0,\nu}h_{0,\nu}}{1+\lambda\rho_{0,\nu}}\right)^2, \\ (II_h) &= \sum_\nu \left[(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{V_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]}{1+\lambda\rho_{0,\nu}}\right]^2.\end{aligned}$$

Since $\mathrm{E}(\bar{\beta}_{\nu,n}) = 0$, $\mathrm{E}(\bar{\beta}^2_{\nu,n}) = \frac{1}{n}$, and $\sum_\nu \rho_{0,\nu}^{1+\epsilon_0} h_{0,\nu}^2 < \infty$ for $\epsilon_0 \in [0,1]$, by Lemma S.4, we have

$$\mathrm{E}[(I_h)] = \frac{1}{n}\sum_\nu \frac{1}{(1+\lambda\rho_{0,\nu})^2} + \lambda^{1+\epsilon_0}\sum_\nu \frac{(\lambda\rho_{0,\nu})^{1-\epsilon_0}}{(1+\lambda\rho_{0,\nu})^2}\rho_{0,\nu}^{1+\epsilon_0}h_{0,\nu}^2 = O(n^{-1}\lambda^{-\frac{1}{r}} + \lambda^{1+\epsilon_0}).$$

If $\boldsymbol{a}, \boldsymbol{b}$ are p dimensional vectors, then $\boldsymbol{a}^T\boldsymbol{b}$ is 1×1, which implies that $\boldsymbol{a}^T\boldsymbol{b} = \boldsymbol{b}^T\boldsymbol{a}$ and $(\boldsymbol{a}^T\boldsymbol{b})^2 = \boldsymbol{b}^T\boldsymbol{a}\boldsymbol{a}^T\boldsymbol{b}$. Using this fact and the bound for $\mathrm{E}[(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)]$, we have

$$\begin{aligned}\mathrm{E}[(II_h)] = &\sum_{i,j=1}^p \Bigg\{ \mathrm{E}\left[(\tilde{\theta} - \theta_0)^i(\tilde{\theta} - \theta_0)^j\right] \\ &\quad \cdot \sum_\nu \frac{1}{(1+\lambda\rho_{0,\nu})^2} V_{\tau_0}[D_{\theta^i}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}]V_{\tau_0}[D_{\theta^j}\eta(\tau_0), D_h\eta(\tau_0)\phi_{0,\nu}] \Bigg\} \\ &\leq C\sum_{i,j=1}^p \mathrm{E}\left[(\tilde{\theta} - \theta_0)^i(\tilde{\theta} - \theta_0)^j\right] \\ &\leq C\sum_{i,j=1}^p \left\{\mathrm{E}\left[(\tilde{\theta}^i - \theta_0^i)^2\right]\right\}^{\frac{1}{2}} \left\{\mathrm{E}\left[(\tilde{\theta}^j - \theta_0^j)^2\right]\right\}^{\frac{1}{2}} = O(n^{-1}\lambda^{-\frac{1}{r}}).\end{aligned}$$

Therefore, $\mathrm{E}\{V_{\tau_0}[D_h\eta(\tau_0)(\tilde{h} - h_0)]\} = O(n^{-1}\lambda^{-\frac{1}{r}} + \lambda^{1+\epsilon_0})$. Similar analysis shows that

$$\mathrm{E}\left[\lambda J(\tilde{h} - h_0)\right] = \mathrm{E}\left[\sum_\nu \lambda\rho_{0,\nu}\left(\tilde{h}_\nu - h_{0,\nu}\right)^2\right] = O(n^{-1}\lambda^{-\frac{1}{r}} + \lambda^{1+\epsilon_0}).$$

Hence, the second bound in Lemma 1 is established.

S.4 Proof of Theorem 3

By Assumption 4, Lemma 1 implies that

$$\mathrm{E}\left\{V_{\tau_0}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)]\right\} = O(n^{-1}\lambda^{-\frac{1}{r}}),$$
$$\mathrm{E}\left\{V_{\tau_0}[L_h(\tilde{h}-h_0)] + \lambda J(\tilde{h}-h_0)\right\} = O(n^{-1}\lambda^{-\frac{1}{r}} + \lambda^{1+\epsilon_0}).$$

By the Cauchy-Schwarz inequality and completing the square, we have

$$\begin{aligned}
&V_{\tau_0}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + L_h(\tilde{h}-h_0)]\\
&\quad= V_{\tau_0}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)] + V_{\tau_0}[L_h(\tilde{h}-h_0)] + 2V_{\tau_0}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0), L_h(\tilde{h}-h_0)]\\
&\quad\le V_{\tau_0}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)] + V_{\tau_0}[L_h(\tilde{h}-h_0)] + 2V_{\tau_0}^{\frac{1}{2}}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)]V_{\tau_0}^{\frac{1}{2}}[L_h(\tilde{h}-h_0)]\\
&\quad= \left[V_{\tau_0}^{\frac{1}{2}}[L_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)] + V_{\tau_0}^{\frac{1}{2}}[L_h(\tilde{h}-h_0)]\right]^2\\
&\quad= O_p(n^{-1}\lambda^{-\frac{1}{r}} + \lambda^{1+\epsilon_0}).
\end{aligned}$$

Together with $\lambda J(\tilde{h}-h_0) = O_p(n^{-1}\lambda^{-\frac{1}{r}} + \lambda^{1+\epsilon_0})$, we have the desired result.

S.5 Local existence and uniqueness

we establish a theory for local existence and uniqueness of the semiparametric estimator of the penalized likelihood given by (2.1) for the case when $m = 1$. The proof for the case when $m > 1$ can be carried out in a similar manner. We follow a framework that was often used to study nonparametric models (Cox, 1988; Cox and O'Sullivan, 1990; O'Sullivan, 1990; Ke and Wang, 2004). Note that we have the same assumptions as given in Section 3 and the Appendix for the proof of consistency in the main text, with the exception of assuming the existence of $(\hat{\boldsymbol{\theta}}, \hat{h})$ in $\mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$. For convenience, we drop the subscript $l = 1$. Recall the penalized likelihood

$$\mathfrak{L}_{n,\lambda}(\boldsymbol{\theta}, h) = \mathfrak{L}_n(\boldsymbol{\theta}, h) + \frac{\lambda}{2}J(h) = -\frac{1}{n}\sum_{i=1}^{n}\eta(X_i;\boldsymbol{\theta}, h) + \log\int_{\mathcal{X}} e^{\eta(x;\boldsymbol{\theta},h)}dx + \frac{\lambda}{2}J(h),$$

where $\mathfrak{L}_n(\boldsymbol{\theta}, h)$ is the negative log likelihood of the density function

$$f(x;\boldsymbol{\theta}, h) = \frac{\exp\{\eta(x;\boldsymbol{\theta}, h)\}}{\int_{\mathcal{X}}\exp\{\eta(x;\boldsymbol{\theta}, h)\}dx},$$

$J(h)$ is the roughness penalty term, which is assumed to be a quadratic functional, and λ is the smoothing parameter. We consider $\tau = (\boldsymbol{\theta}, h) \in \mathcal{Q} = \mathbb{R}^p \times \mathcal{H}$ and a known function $\eta : \mathcal{Q} \to L_0^2(\mathcal{X})$, where $\mathcal{H}$ is a real reproducing kernel Hilbert space (RKHS), $\mathcal{Q} = \mathbb{R}^p \times \mathcal{H}$, $L_0^2(\mathcal{X}) = L_{\tau_0}^2(\mathcal{X}) \ominus \{1\}$, $L_{\tau_0}^2(\mathcal{X})$ is the space of functions with finite second moments with respect to the measure given by the true density, and $\tau_0 = (\boldsymbol{\theta}_0, h_0)$ is the true parameter. Denote the kth order partial Fréchet derivative operators by $D^k_{a_1 \ldots a_k} = D_{a_1} \ldots D_{a_k}$, where $a_i \in \{\boldsymbol{\theta}, h\}$ for $i = 1, \ldots, k$. The penalized likelihood estimator

$$(\hat{\boldsymbol{\theta}}, \hat{h}) = \underset{(\boldsymbol{\theta}, h) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}}{\arg\min} \quad \mathfrak{L}_{n,\lambda}(\boldsymbol{\theta}, h)$$

satisfies

$$\begin{cases} D_{\boldsymbol{\theta}} \mathfrak{L}_{n,\lambda}(\hat{\boldsymbol{\theta}}, \hat{h}) = 0 \\ D_h \mathfrak{L}_{n,\lambda}(\hat{\boldsymbol{\theta}}, \hat{h}) = 0. \end{cases}$$

We also have $\mathfrak{L}_\lambda(\boldsymbol{\theta}, h) = \mathrm{E}[\mathfrak{L}_{n,\lambda}(\boldsymbol{\theta}, h)]$, and it is easy to verify that the true parameter $\tau_0 = (\boldsymbol{\theta}_0, h_0)$ is the solution for

$$\begin{cases} D_{\boldsymbol{\theta}} \mathfrak{L}_0(\boldsymbol{\theta}, h) = 0 \\ D_h \mathfrak{L}_0(\boldsymbol{\theta}, h) = 0. \end{cases}$$

Note that by Assumption 3, τ_0 is the unique solution for the above system in $\mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$. Lastly, throughout this document, we use $\langle \cdot, \cdot \rangle_{\mathbb{R}^p}, \langle \cdot, \cdot \rangle_{\mathcal{H},1}$ as defined in the main text to be the inner products on $\mathbb{R}^p$ and $\mathcal{H}$, respectively, and $\|\cdot\|_{\mathbb{R}^p}$, $\|\cdot\|_{\mathcal{H},1}$ denote the corresponding induced norms. Recall that $\langle \cdot, \cdot \rangle_{\mathcal{H},1}$ is the inner product given by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ when $\lambda = 1$.

S.5.1 Linearization

In the next four sections, we extend the linearization technique used to approximate the systematic and stochastic components of the estimation error as in Cox and O'Sullivan (1990); O'Sullivan (1990) to our semiparametric setting by using the bivariate Taylor series expansions for nonlinear operators. We first state the following proposition, whose proof is provided in Ke and Wang (2004).

Proposition S.1. *Let $f : D(f) \subset X \times Y \to Z$, where X, Y and Z are Banach spaces. If f'' exists at (x, y), then the partial Fréchet derivatives f_{xx}, f_{xy}, f_{yx} and f_{yy} exist at (x, y).*

For any $h, a \in X,\ k, b \in Y$,

$$f''(x,y)(h,k)(a,b) = f_{xx}(x,y)ha + f_{xy}(x,y)ka + f_{yx}(x,y)hb + f_{yy}(x,y)kb.$$

By the above theorem and the Taylor formula given in Chapter 1 Section 4 in Ambrosetti and Prodi (1995), we can write the first order Taylor series expansion of $f(x,y)$ as

$$f(x+h, y+k) = f(x,y) + f_x(x,y)h + f_y(x,y)k + R,$$

where R is the remainder, given by

$$\begin{aligned} R &= \int_0^1 (1-t) f''(x+th, y+tk)(h,k)(h,k)dt \\ &= \int_0^1 (1-t)\left[f_{xx}(x+th, y+tk)hh + f_{xy}(x+th, y+tk)kh \right. \\ &\qquad \left. + \ f_{yx}(x+th, y+tk)hk + f_{yy}(x+th, y+tk)kk\right]\ dt. \end{aligned}$$

S.5.2 Linear expansions

Since $D_{\boldsymbol{\theta}}\mathfrak{L}_\lambda(\boldsymbol{\theta}, h)$ is a bounded linear functional on $\mathbb{R}^p$, by the Riesz representation theorem, there exists $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h) \in \mathbb{R}^p$ such that for any $\boldsymbol{a} \in \mathbb{R}^p$,

$$D_{\boldsymbol{\theta}}\mathfrak{L}_\lambda(\boldsymbol{\theta}, h)\boldsymbol{a} = \langle Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h), \boldsymbol{a}\rangle_{l^2}.$$

Similarly, we can denote the Riesz representer of $D_h\mathfrak{L}_\lambda(\boldsymbol{\theta}, h)$ in $\mathcal{H}$ by $Z_h(\boldsymbol{\theta}, h)$. For convenience, we use

$$Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h) = D_{\boldsymbol{\theta}}\mathfrak{L}_\lambda(\boldsymbol{\theta}, h) \quad \text{and} \quad Z_h(\boldsymbol{\theta}, h) = D_h\mathfrak{L}_\lambda(\boldsymbol{\theta}, h)$$

to represent either the functionals or their Riesz representers in $\mathbb{R}^p$ and $\mathcal{H}$, respectively. For any $\boldsymbol{\theta}_0 + \boldsymbol{a} \in \mathcal{N}_{\boldsymbol{\theta}_0}$ and $h_0 + g \in \mathcal{N}_{h_0}$, the first order Taylor series expansions of $Z_h, Z_{\boldsymbol{\theta}}$ at the true parameter $(\boldsymbol{\theta}_0, h_0)$ are

$$\begin{aligned} Z_h(\boldsymbol{\theta}_0 + \boldsymbol{a}, h_0 + g) &= Z_h(\boldsymbol{\theta}_0, h_0) + D_{\boldsymbol{\theta}} Z_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{a} + D_h Z_h(\boldsymbol{\theta}_0, h_0)g + R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g, \\ Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{a}, h_0 + g) &= Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) + D_{\boldsymbol{\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a} + D_h Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)g + R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g, \end{aligned}$$

where

$$
\begin{aligned}
R_h(\boldsymbol{\theta}, h)\boldsymbol{a}g = \int_0^1 (1-t)\, \big[& D^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Z_h(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)\boldsymbol{a}\boldsymbol{a} + D^2_{h\boldsymbol{\theta}} Z_h(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)ag \\
& + D^2_{\boldsymbol{\theta}h} Z_h(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)g\boldsymbol{a} + D^2_{hh} Z_h(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)gg \big]\ dt, \\
R_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}g = \int_0^1 (1-t)\, \big[& D^2_{\boldsymbol{\theta}\boldsymbol{\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)\boldsymbol{a}\boldsymbol{a} + D^2_{h\boldsymbol{\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)\boldsymbol{a}g \\
& + D^2_{\boldsymbol{\theta}h} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)g\boldsymbol{a} + D^2_{hh} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta} + t\boldsymbol{a}, h + tg)gg \big]\ dt.
\end{aligned}
$$

For $Z_h(\boldsymbol{\theta}, h)$, $u, v \in \mathcal{H}$, $\boldsymbol{a} \in \mathbb{R}^p$, we have

$$
\begin{aligned}
Z_h(\boldsymbol{\theta}, h)u &= -\mu_{\tau_0}[D_h\eta(\boldsymbol{\theta}, h)u] + \mu_\tau[D_h\eta(\boldsymbol{\theta}, h)u] + \lambda J(h, u), \\
D_h Z_h(\boldsymbol{\theta}, h)uv &= -\left\{\mu_{\tau_0}[D^2_{hh}\eta(\boldsymbol{\theta}, h)uv] - \mu_\tau[D^2_{hh}\eta(\boldsymbol{\theta}, h)uv]\right\} \\
&\quad + V_\tau[D_h\eta(\boldsymbol{\theta}, h)v, D_h\eta(\boldsymbol{\theta}, h)u] + \lambda J(v, u), \\
D_{\boldsymbol{\theta}} Z_h(\boldsymbol{\theta}, h)u\boldsymbol{a} &= -\left\{\mu_{\tau_0}[D^2_{\boldsymbol{\theta}h}\eta(\boldsymbol{\theta}, h)u\boldsymbol{a}] - \mu_\tau[D^2_{\boldsymbol{\theta}h}\eta(\boldsymbol{\theta}, h)ua]\right\} \\
&\quad + V_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)u].
\end{aligned}
$$

For $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)$, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$, $u \in \mathcal{H}$, we have

$$
\begin{aligned}
Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a} &= -\mu_{\tau_0}[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}] + \mu_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}], \\
D_h Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}u &= -\left\{\mu_{\tau_0}[D^2_{h\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}u] - \mu_\tau[D^2_{h\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}u]\right\} + V_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)u], \\
D_{\boldsymbol{\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}\boldsymbol{b} &= -\left\{\mu_{\tau_0}[D^2_{\boldsymbol{\theta}\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\boldsymbol{b}] - \mu_\tau[D^2_{\boldsymbol{\theta}\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\boldsymbol{b}]\right\} + V_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}].
\end{aligned}
$$

For any $u, v \in \mathcal{H}$, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$, and $\tau = (\boldsymbol{\theta}, h) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$, define the operators $U_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)$ and $U_h(\boldsymbol{\theta}, h)$ on $\mathbb{R}^p$ and $\mathcal{H}$, respectively, such that

$$
\begin{aligned}
\langle u, U_h(\boldsymbol{\theta}, h)v\rangle &= V_\tau[D_h\eta(\boldsymbol{\theta}, h)u, D_h\eta(\boldsymbol{\theta}, h)v], \\
\langle \boldsymbol{a}, U_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{b}\rangle_{l^2} &= V_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}],
\end{aligned}
$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $\mathcal{H}$ as defined in Assumption 2 and $\langle \cdot, \cdot \rangle_{l^2}$ is the l^2 inner product on $\mathbb{R}^p$. Note that these operators are well-defined by the Riesz representation theorem applied to the linear functionals

$$
v \mapsto V_\tau[D_h\eta(\boldsymbol{\theta}, h)u, D_h\eta(\boldsymbol{\theta}, h)v], \qquad \boldsymbol{b} \mapsto V_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}],
$$

which are bounded in the corresponding norms on $\mathcal{H}$ and $\mathbb{R}^p$, respectively. Similarly, we also define $U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}, h) : \mathcal{H} \to \mathbb{R}^p$ and $U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}, h) : \mathbb{R}^p \to \mathcal{H}$ by

$$\langle \boldsymbol{a}, U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}, h)u\rangle_{l^2} = V_\tau[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)u] = \langle u, U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}, h)\boldsymbol{a}\rangle .$$

By Lemma S.2 in the supplement of Cheng and Shang (2015), there exists a bounded linear operator W_λ on $\mathcal{H}$ such that $\langle u, W_\lambda v\rangle = \lambda J(u, v)$. Therefore,

$$\begin{aligned} Z_h(\boldsymbol{\theta}_0 + \boldsymbol{a}, h_0 + g) &= Z_h(\boldsymbol{\theta}_0, h_0) + G_h(\boldsymbol{\theta}_0, h_0)g + U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a} + R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g, \\ Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{a}, h_0 + g) &= Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) + U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a} + U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)g + R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g, \end{aligned} \tag{S.9}$$

where $G_h(\boldsymbol{\theta}, h) = U_h(\boldsymbol{\theta}, h) + W_\lambda$. We provide the presentations of the remainder terms $R_h(\boldsymbol{\theta}, h)\boldsymbol{a}g$ and $R_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}g$ in Section S.5.3.

Suppose $(\boldsymbol{\theta}_\lambda, h_\lambda)$ is a solution for $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h) = Z_h(\boldsymbol{\theta}, h) = 0$. We define the systematic error as $(\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}_0, h_\lambda - h_0)$. Ignoring the remainder terms, we get an approximation to the systematic error by setting the system of equations (S.9) to 0 and solving for $\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0$, and $\bar{h}_\lambda - h_0$, i.e.,

$$\begin{aligned} Z_h(\boldsymbol{\theta}_0, h_0) + G_h(\boldsymbol{\theta}_0, h_0)(\bar{h}_\lambda - h_0) + U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0)(\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0) = 0, \\ Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) + U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)(\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0) + U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)(\bar{h}_\lambda - h_0) = 0. \end{aligned}$$

By the Lax-Milgram theorem (Section 3.6 of Aubin (1979)) and Assumptions 2 and 4 in the main text, for any $(\boldsymbol{\theta}, h) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$, the operators $G_h(\boldsymbol{\theta}, h)$, $U_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)$ have bounded inverses on $\mathcal{H}$ and $\mathbb{R}^p$, respectively. Let

$$\begin{aligned} G_{hh}(\boldsymbol{\theta}, h) &= (G_h - U_{\boldsymbol{\theta}h}U_{\boldsymbol{\theta}}^{-1}U_{h\boldsymbol{\theta}})(\boldsymbol{\theta}, h) : \mathcal{H} \to \mathcal{H}, \\ G_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, h) &= (U_{\boldsymbol{\theta}} - U_{h\boldsymbol{\theta}}G_h^{-1}U_{\boldsymbol{\theta}h})(\boldsymbol{\theta}, h) : \mathbb{R}^p \to \mathbb{R}^p. \end{aligned}$$

Assuming both operators above have bounded inverses for any $(\boldsymbol{\theta}, h) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$, we get

$$\begin{aligned} \bar{h}_\lambda - h_0 &= -G_{hh}^{-1}(\boldsymbol{\theta}_0, h_0)\left[Z_h(\boldsymbol{\theta}_0, h_0) - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0)U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\right], \\ \bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0 &= -G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\left[Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)G_h^{-1}(\boldsymbol{\theta}_0, h_0)Z_h(\boldsymbol{\theta}_0, h_0)\right]. \end{aligned}$$

Next, we define the stochastic error as $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\lambda, \hat{h} - h_\lambda)$. Similar to the definition of $Z_{\boldsymbol{\theta}}$ and Z_h, we let

$$Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h) = D_{\boldsymbol{\theta}}\mathfrak{L}_{n,\lambda}(\boldsymbol{\theta}, h) \quad \text{and} \quad Z_{nh}(\boldsymbol{\theta}, h) = D_h\mathfrak{L}_{n,\lambda}(\boldsymbol{\theta}, h).$$

The approximation of the stochastic errors can be obtained by the linearizations of $Z_{n\boldsymbol{\theta}}$ and Z_{nh}. For $u, v \in \mathcal{H}$ and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$, we have

$$
\begin{aligned}
Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a} &= -\frac{1}{n}\sum_{i=1}^{n} D_{\boldsymbol{\theta}}\eta(x_i;\boldsymbol{\theta}, h)\boldsymbol{a} + \mu_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right], \\
D_{\boldsymbol{\theta}} Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{ab} &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{\boldsymbol{\theta\theta}}\eta(x_i;\boldsymbol{\theta}, h)\boldsymbol{ab} + \mu_\tau\left[D^2_{\boldsymbol{\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{ab}\right] + V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right], \\
D_h Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}u &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{h\boldsymbol{\theta}}\eta(x_i;\boldsymbol{\theta}, h)\boldsymbol{a}u + \mu_\tau\left[D^2_{h\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}u\right] + V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)u\right], \\
Z_{nh}(\boldsymbol{\theta}, h)u &= -\frac{1}{n}\sum_{i=1}^{n} D_h\eta(x_i;\boldsymbol{\theta}, h)u + \mu_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u\right] + \lambda J(h, u), \\
D_{\boldsymbol{\theta}} Z_{nh}(\boldsymbol{\theta}, h)u\boldsymbol{a} &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{\boldsymbol{\theta}h}\eta(x_i;\boldsymbol{\theta}, h)u\boldsymbol{a} + \mu_\tau\left[D^2_{\boldsymbol{\theta}h}\eta(\boldsymbol{\theta}, h)u\boldsymbol{a}\right] + V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right], \\
D_h Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h)uv &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{hh}\eta(x_i;\boldsymbol{\theta}, h)uv + \mu_\tau\left[D^2_{hh}\eta(\boldsymbol{\theta}, h)uv\right] + V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u, D_h\eta(\boldsymbol{\theta}, h)v\right] \\
&\quad + \lambda J(v, u).
\end{aligned}
$$

Thus, for any $\boldsymbol{\theta}_\lambda + \boldsymbol{a} \in \mathcal{N}_{\boldsymbol{\theta}_0}, h_\lambda + g \in \mathcal{N}_{h_0}$, the first order Taylor series expansions of $Z_{n\boldsymbol{\theta}}$ and Z_{nh} at $(\boldsymbol{\theta}_\lambda, h_\lambda) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$ can be written as

$$
\begin{aligned}
Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{a}, h_\lambda + g) &= Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) + U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a} + U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)g + e_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a}g + R_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a}g, \\
Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{a}, h_\lambda + g) &= Z_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda) + G_h(\boldsymbol{\theta}_\lambda, h_\lambda)g + U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a} + e_h(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a}g + R_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a}g,
\end{aligned}
\tag{S.10}
$$

where the error terms are given by

$$
\begin{aligned}
e_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a}g &= e_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a} + e_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda)g, \\
e_h(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a}g &= e_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\boldsymbol{a} + e_{hh}(\boldsymbol{\theta}_\lambda, h_\lambda)g, \\
e_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a} &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{\boldsymbol{\theta\theta}}\eta(X_i;\boldsymbol{\theta}, h)\boldsymbol{a} + \mu_\tau\left[D^2_{\boldsymbol{\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right], \\
e_{\boldsymbol{\theta}h}(\boldsymbol{\theta}, h)g &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{h\boldsymbol{\theta}}\eta(X_i;\boldsymbol{\theta}, h)g + \mu_\tau\left[D^2_{\boldsymbol{\theta\theta}}\eta(\boldsymbol{\theta}, h)g\right], \\
e_{h\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a} &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{\boldsymbol{\theta}h}\eta(X_i;\boldsymbol{\theta}, h)\boldsymbol{a} + \mu_\tau\left[D^2_{\boldsymbol{\theta}h}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right], \\
e_{hh}(\boldsymbol{\theta}, h)g &= -\frac{1}{n}\sum_{i=1}^{n} D^2_{hh}\eta(X_i;\boldsymbol{\theta}, h)g + \mu_\tau\left[D^2_{hh}\eta(\boldsymbol{\theta}, h)g\right].
\end{aligned}
$$

$R_{nh}, R_{n\boldsymbol{\theta}}$ are defined similarly to $R_h, R_{\boldsymbol{\theta}}$ by replacing $Z_h, Z_{\boldsymbol{\theta}}$ with Z_{nh}, $Z_{n\boldsymbol{\theta}}$, respectively. Recall that $(\hat{\boldsymbol{\theta}}, \hat{h})$ is the solution for $Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h) = Z_{nh}(\boldsymbol{\theta}, h) = 0$. Dropping the error terms and the remainder terms, we get an approximation to the stochastic error $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\lambda, \hat{h} - h_\lambda)$ by setting the linearizations (S.10) to 0 and solving for $\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda$ and $\bar{h}_{n\lambda} - h_\lambda$. We get

$$\bar{h}_{n\lambda} - h_\lambda = -G_{hh}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)[Z_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda) - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda)U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)],$$
$$\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda = -G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)[Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)G_{h}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)Z_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda)].$$

S.5.3 Remainder terms

To find the representation of R_h and $R_{\boldsymbol{\theta}}$ given in Section S.5.2, we need to find the second partial Fréchet derivatives of $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)$ and $Z_h(\boldsymbol{\theta}, h)$. For $u, v, w \in \mathcal{H}$, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^p$, we have

$$\begin{aligned}
&D_{hh}^2 Z_h(\boldsymbol{\theta}, h)uvw \\
&= -\left\{\mu_{\tau_0}\left[D_{hhh}^2\eta(\boldsymbol{\theta}, h)uvw\right] - \mu_\tau\left[D_{hhh}^2\eta(\boldsymbol{\theta}, h)uvw\right]\right\} \\
&\quad + V_\tau\left[D_{hh}^2\eta(\boldsymbol{\theta}, h)uv, D_h\eta(\boldsymbol{\theta}, h)w\right] + V_\tau\left[D_{hh}^2\eta(\boldsymbol{\theta}, h)vw, D_h\eta(\boldsymbol{\theta}, h)u\right] \\
&\quad + V_\tau\left[D_{hh}^2\eta(\boldsymbol{\theta}, h)uw, D_h\eta(\boldsymbol{\theta}, h)v\right] + V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)v, D_h\eta(\boldsymbol{\theta}, h)u \cdot D_h\eta(\boldsymbol{\theta}, h)w\right] \\
&\quad - \mu_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u\right]V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)v, D_h\eta(\boldsymbol{\theta}, h)w\right] - \mu_\tau\left[D_h\eta(\boldsymbol{\theta}, h)w\right]V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)v, D_h\eta(\boldsymbol{\theta}, h)u\right],
\end{aligned}$$

$$\begin{aligned}
&D_{\boldsymbol{\theta}h}^2 Z_h(\boldsymbol{\theta}, h)uv\boldsymbol{a} = D_{h\boldsymbol{\theta}}^2 Z_h(\boldsymbol{\theta}, h)u\boldsymbol{a}v = D_{hh}^2 Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}uv \\
&= -\left\{\mu_{\tau_0}\left[D_{\boldsymbol{\theta}}D_{hh}^2\eta(\boldsymbol{\theta}, h)uv\boldsymbol{a}\right] - \mu_\tau\left[D_{\boldsymbol{\theta}}D_{hh}^2\eta(\boldsymbol{\theta}, h)uv\boldsymbol{a}\right]\right\} \\
&\quad + V_\tau\left[D_{hh}^2\eta(\boldsymbol{\theta}, h)uv, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right] + V_\tau\left[D_{\boldsymbol{\theta}h}^2\eta(\boldsymbol{\theta}, h)v\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)u\right] \\
&\quad + V_\tau\left[D_{\boldsymbol{\theta}h}^2\eta(\boldsymbol{\theta}, h)u\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)v\right] + V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u \cdot D_h\eta(\boldsymbol{\theta}, h)v, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right] \\
&\quad - \mu_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u\right]V_\tau\left[D_h\eta(\boldsymbol{\theta}, h)v, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right] - \mu_\tau\left[D_h\eta(\boldsymbol{\theta}, h)v\right]V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)a, D_h\eta(\boldsymbol{\theta}, h)u\right],
\end{aligned}$$

$$\begin{aligned}
&D_{\boldsymbol{\theta\theta}}^2 Z_h(\boldsymbol{\theta}, h)u\boldsymbol{ab} = D_{\boldsymbol{\theta}h}^2 Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{a}u\boldsymbol{b} = D_{h\boldsymbol{\theta}}^2 Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{ab}u \\
&= -\left\{\mu_{\tau_0}\left[D_{\boldsymbol{\theta\theta}}^2 D_h\eta(\boldsymbol{\theta}, h)u\boldsymbol{ab}\right] - \mu_\tau\left[D_{\boldsymbol{\theta\theta}}^2 D_h\eta(\boldsymbol{\theta}, h)u\boldsymbol{ab}\right]\right\} \\
&\quad + V_\tau\left[D_{\boldsymbol{\theta}h}^2\eta(\boldsymbol{\theta}, h)u\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right] + V_\tau\left[D_{\boldsymbol{\theta\theta}}^2\eta(\boldsymbol{\theta}, h)\boldsymbol{ab}, D_h\eta(\boldsymbol{\theta}, h)u\right] \\
&\quad + V_\tau\left[D_{\boldsymbol{\theta}h}^2\eta(\boldsymbol{\theta}, h)u\boldsymbol{b}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right] + V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b} \cdot D_h\eta(\boldsymbol{\theta}, h)u\right] \\
&\quad - \mu_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right]V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_h\eta(\boldsymbol{\theta}, h)u\right] - \mu_\tau\left[D_h\eta(\boldsymbol{\theta}, h)u\right]V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right].
\end{aligned}$$

$$
\begin{aligned}
&D^2_{\boldsymbol{\theta\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}, h)\boldsymbol{abc} \\
&= -\left\{\mu_{\tau_0}\left[D^3_{\boldsymbol{\theta\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{abc}\right] - \mu_\tau\left[D^3_{\boldsymbol{\theta\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{abc}\right]\right\} \\
&\quad + V_\tau\left[D^2_{\boldsymbol{\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{ab}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{c}\right] + V_\tau\left[D^2_{\boldsymbol{\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{bc}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}\right] \\
&\quad + V_\tau\left[D^2_{\boldsymbol{\theta\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{ac}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right] + V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b} \cdot D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{c}\right] \\
&\quad - \mu_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right] V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{c}\right] - \mu_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{c}\right] V_\tau\left[D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{a}, D_{\boldsymbol{\theta}}\eta(\boldsymbol{\theta}, h)\boldsymbol{b}\right],
\end{aligned}
$$

By replacing the terms $\mu_{\tau_0}[\cdot(x)]$ with $\frac{1}{n}\sum_{i=1}^n \cdot(x_i)$ in each term above, we have the second partial Fréchet derivatives of $Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}, h)$ and $Z_{nh}(\boldsymbol{\theta}, h)$ for the remainder terms R_{nh} and $R_{n\boldsymbol{\theta}}$.

S.5.4 Bounds for the remainders

We see that the magnitude of the remainder terms $R_{\boldsymbol{\theta}}$, R_h, $R_{n\boldsymbol{\theta}}$, R_{nh}, $e_{\boldsymbol{\theta}}$, and e_h determine how accurate $(\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0, \bar{h}_\lambda - h_0)$ and $(\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda, \bar{h}_{n\lambda} - h_\lambda)$ are as approximations of the systematic error and the stochastic error, respectively. To obtain bounds of these terms, we first define for $\lambda > 0$, $\tau_1 = (\boldsymbol{\theta}_1, h_1), \tau_2 = (\boldsymbol{\theta}_2, h_2) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$, and unit elements $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^p$ and $v_1, v_2 \in \mathcal{H}$,

$$
\begin{aligned}
K_h^1 &= \sup_{\tau_1,\tau_2}\sup_{v_1,v_2}\left\|G_{hh}^{-1}(\tau_1)[D^2_{hh}Z_h(\tau_2)v_1v_2 - U_{\boldsymbol{\theta}h}(\tau_1)U_{\boldsymbol{\theta}}^{-1}(\tau_1)D^2_{hh}Z_{\boldsymbol{\theta}}(\tau_2)v_1v_2]\right\|_{\mathcal{H},1}, \\
K_h^2 &= \sup_{\tau_1,\tau_2}\sup_{v_1,\boldsymbol{u}_1}\left\|G_{hh}^{-1}(\tau_1)[D^2_{\boldsymbol{\theta}h}Z_h(\tau_2)v_1\boldsymbol{u}_1 - U_{\boldsymbol{\theta}h}(\tau_1)U_{\boldsymbol{\theta}}^{-1}(\tau_1)D^2_{\boldsymbol{\theta}h}Z_{\boldsymbol{\theta}}(\tau_2)v_1\boldsymbol{u}_1]\right\|_{\mathcal{H},1}, \\
K_h^3 &= \sup_{\tau_1,\tau_2}\sup_{\boldsymbol{u}_1,v_1}\left\|G_{hh}^{-1}(\tau_1)[D^2_{h\boldsymbol{\theta}}Z_h(\tau_2)\boldsymbol{u}_1v_1 - U_{\boldsymbol{\theta}h}(\tau_1)U_{\boldsymbol{\theta}}^{-1}(\tau_1)D^2_{h\boldsymbol{\theta}}Z_{\boldsymbol{\theta}}(\tau_2)\boldsymbol{u}_1v_1]\right\|_{\mathcal{H},1}, \\
K_h^4 &= \sup_{\tau_1,\tau_2}\sup_{\boldsymbol{u}_1,\boldsymbol{u}_2}\left\|G_{hh}^{-1}(\tau_1)[D^2_{\boldsymbol{\theta\theta}}Z_h(\tau_2)\boldsymbol{u}_1\boldsymbol{u}_2 - U_{\boldsymbol{\theta}h}(\tau_1)U_{\boldsymbol{\theta}}^{-1}(\tau_1)D^2_{\boldsymbol{\theta\theta}}Z_{\boldsymbol{\theta}}(\tau_2)\boldsymbol{u}_1\boldsymbol{u}_2]\right\|_{\mathcal{H},1}, \\
K_{\boldsymbol{\theta}}^1 &= \sup_{\tau_1,\tau_2}\sup_{v_1,v_2}\left\|G_{\boldsymbol{\theta\theta}}^{-1}(\tau_1)[D^2_{hh}Z_{\boldsymbol{\theta}}(\tau_2)v_1v_2 - U_{h\boldsymbol{\theta}}(\tau_1)G_h^{-1}(\tau_1)D^2_{hh}Z_h(\tau_2)v_1v_2]\right\|_{\mathbb{R}^p}, \\
K_{\boldsymbol{\theta}}^2 &= \sup_{\tau_1,\tau_2}\sup_{v_1,\boldsymbol{u}_1}\left\|G_{\boldsymbol{\theta\theta}}^{-1}(\tau_1)[D^2_{\boldsymbol{\theta}h}Z_{\boldsymbol{\theta}}(\tau_2)v_1\boldsymbol{u}_1 - U_{h\boldsymbol{\theta}}(\tau_1)G_h^{-1}(\tau_1)D^2_{\boldsymbol{\theta}h}Z_h(\tau_2)v_1\boldsymbol{u}_1]\right\|_{\mathbb{R}^p}, \\
K_{\boldsymbol{\theta}}^3 &= \sup_{\tau_1,\tau_2}\sup_{\boldsymbol{u}_1,v_1}\left\|G_{\boldsymbol{\theta\theta}}^{-1}(\tau_1)[D^2_{h\boldsymbol{\theta}}Z_{\boldsymbol{\theta}}(\tau_2)\boldsymbol{u}_1v_1 - U_{h\boldsymbol{\theta}}(\tau_1)G_h^{-1}(\tau_1)D^2_{h\boldsymbol{\theta}}Z_h(\tau_2)\boldsymbol{u}_1v_1]\right\|_{\mathbb{R}^p}, \\
K_{\boldsymbol{\theta}}^4 &= \sup_{\tau_1,\tau_2}\sup_{\boldsymbol{u}_1,\boldsymbol{u}_2}\left\|G_{\boldsymbol{\theta\theta}}^{-1}(\tau_1)[D^2_{\boldsymbol{\theta\theta}}Z_{\boldsymbol{\theta}}(\tau_2)\boldsymbol{u}_1\boldsymbol{u}_2 - U_{h\boldsymbol{\theta}}(\tau_1)G_h^{-1}(\tau_1)D^2_{\boldsymbol{\theta\theta}}Z_h(\tau_2)\boldsymbol{u}_1\boldsymbol{u}_2]\right\|_{\mathbb{R}^p}.
\end{aligned}
$$

For $i = 1, 2, 3, 4$, we also define $K_{nh}^i, K_{n\boldsymbol{\theta}}^i$ by replacing $Z_{\boldsymbol{\theta}}, Z_h$ with $Z_{n\boldsymbol{\theta}}, Z_{nh}$ in K_h^i and $K_{\boldsymbol{\theta}}^i$, respectively. In addition, we define

$$
\begin{aligned}
E_{nh}^{12} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{\boldsymbol{u}_1} \left\| G_{hh}^{-1}(\boldsymbol{\theta}_1, h_1) e_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_1, h_1) \boldsymbol{u}_1 \right\|_{\mathcal{H},1}, \\
E_{nh}^{11} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{v_1} \left\| G_{hh}^{-1}(\boldsymbol{\theta}_1, h_1) e_{hh}(\boldsymbol{\theta}_1, h_1) v_1 \right\|_{\mathcal{H},1}, \\
E_{nh}^{22} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{\boldsymbol{u}_1} \left\| G_{hh}^{-1}(\boldsymbol{\theta}_1, h_1) U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_1, h_1) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_1, h_1) e_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_1, h_1) \boldsymbol{u}_1 \right\|_{\mathcal{H},1}, \\
E_{nh}^{21} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{v_1} \left\| G_{hh}^{-1}(\boldsymbol{\theta}_1, h_1) U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_1, h_1) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_1, h_1) e_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_1, h_1) v_1 \right\|_{\mathcal{H},1}, \\
E_{n\boldsymbol{\theta}}^{22} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{\boldsymbol{u}_1} \left\| G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_1, h_1) e_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_1, h_1) \boldsymbol{u}_1 \right\|_{\mathbb{R}^p}, \\
E_{n\boldsymbol{\theta}}^{21} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{v_1} \left\| G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_1, h_1) e_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_1, h_1) v_1 \right\|_{\mathbb{R}^p}, \\
E_{n\boldsymbol{\theta}}^{12} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{\boldsymbol{u}_1} \left\| G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_1, h_1) U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_1, h_1) G_h^{-1}(\boldsymbol{\theta}_1, h_1) e_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_1, h_1) \boldsymbol{u}_1 \right\|_{\mathbb{R}^p}, \\
E_{n\boldsymbol{\theta}}^{11} &= \sup_{\boldsymbol{\theta}_1, h_1} \sup_{v_1} \left\| G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_1, h_1) U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_1, h_1) G_h^{-1}(\boldsymbol{\theta}_1, h_1) e_{hh}(\boldsymbol{\theta}_1, h_1) v_1 \right\|_{\mathbb{R}^p}.
\end{aligned}
$$

Therefore, for any $\boldsymbol{a} \in \mathbb{R}^p$ and $g \in \mathcal{H}$, standard analysis yields the following bounds for the remainder terms for the systematic error and the stochastic error,

$$
\begin{aligned}
&\left\| G_{hh}^{-1}(\boldsymbol{\theta}_0, h_0) \left[R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g \right] \right\|_{\mathcal{H},1} \\
&\quad \le \frac{1}{2} \left[\left(K_h^1 \|g\|_{\mathcal{H},1} + K_h^2 \|\boldsymbol{a}\|_{\mathbb{R}^p} \right) \|g\|_{\mathcal{H},1} + \left(K_h^3 \|g\|_{\mathcal{H},1} + K_h^4 \|\boldsymbol{a}\|_{\mathbb{R}^p} \right) \|\boldsymbol{a}\|_{\mathbb{R}^p} \right],
\end{aligned} \tag{S.11}
$$

$$
\begin{aligned}
&\left\| G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) \left[R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{a}g \right] \right\|_{\mathbb{R}^p} \\
&\quad \le \frac{1}{2} \left[\left(K_{\boldsymbol{\theta}}^1 \|g\|_{\mathcal{H},1} + K_{\boldsymbol{\theta}}^2 \|\boldsymbol{a}\|_{\mathbb{R}^p} \right) \|g\|_{\mathcal{H},1} + \left(K_{\boldsymbol{\theta}}^3 \|g\|_{\mathcal{H},1} + K_{\boldsymbol{\theta}}^4 \|\boldsymbol{a}\|_{\mathbb{R}^p} \right) \|\boldsymbol{a}\|_{\mathbb{R}^p} \right],
\end{aligned} \tag{S.12}
$$

$$
\begin{aligned}
&\Big\| G_{hh}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) \left\{ \left[e_h(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda) \right] \boldsymbol{a}g \right. \\
&\quad \left. - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) \left[e_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) \right] \boldsymbol{a}g \right\} \Big\|_{\mathcal{H},1} \\
&\quad \le \frac{1}{2} \left[(K_{nh}^1 \|g\|_{\mathcal{H},1} + K_{nh}^2 \|\boldsymbol{a}\|_{\mathbb{R}^p}) \|g\|_{\mathcal{H},1} + (K_{nh}^3 \|g\|_{\mathcal{H},1} + K_{nh}^4 \|\boldsymbol{a}\|_{\mathbb{R}^p}) \|\boldsymbol{a}\|_{\mathbb{R}^p} \right] \\
&\qquad + E_{nh}^1 \|g\|_{\mathcal{H},1} + E_{nh}^2 \|\boldsymbol{a}\|_{\mathbb{R}^p},
\end{aligned} \tag{S.13}
$$

$$
\begin{aligned}
&\Big\| G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) \left\{ \left[e_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) \right] \boldsymbol{a} g \right. \\
&\quad \left. - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) \left[e_h(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda) \right] \boldsymbol{a} g \right\} \Big\|_{\mathbb{R}^p} \\
&\quad \leq \frac{1}{2} \left[(K_{n\boldsymbol{\theta}}^1 \|g\|_{\mathcal{H},1} + K_{n\boldsymbol{\theta}}^2 \|\boldsymbol{a}\|_{\mathbb{R}^p}) \|g\|_{\mathcal{H},1} + (K_{n\boldsymbol{\theta}}^3 \|g\|_{\mathcal{H},1} + K_{n\boldsymbol{\theta}}^4 \|\boldsymbol{a}\|_{\mathbb{R}^p}) \|\boldsymbol{a}\|_{\mathbb{R}^p} \right] \\
&\qquad + E_{n\boldsymbol{\theta}}^1 \|g\|_{\mathcal{H},1} + E_{n\boldsymbol{\theta}}^2 \|\boldsymbol{a}\|_{\mathbb{R}^p},
\end{aligned}
\tag{S.14}
$$

where $E_{nh}^1 = E_{nh}^{11} + E_{nh}^{21}$, $E_{nh}^2 = E_{nh}^{12} + E_{nh}^{22}$, $E_{n\boldsymbol{\theta}}^1 = E_{n\boldsymbol{\theta}}^{21} + E_{n\boldsymbol{\theta}}^{11}$ and $E_{n\boldsymbol{\theta}}^2 = E_{n\boldsymbol{\theta}}^{22} + E_{n\boldsymbol{\theta}}^{12}$.

S.5.5 Proof of existence and uniqueness

We are now ready to show the local existence and uniqueness of $(\boldsymbol{\theta}_\lambda, h_\lambda)$ and $(\hat{\boldsymbol{\theta}}, \hat{h})$ in the neighborhood $\mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$. Let

$$
\begin{aligned}
d_{\boldsymbol{\theta}}(\lambda) &= \left\| \bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0 \right\|_{\mathbb{R}^p}, \\
d_h(\lambda) &= \left\| \bar{h}_\lambda - h_0 \right\|_{\mathcal{H},1}, \\
r_{\boldsymbol{\theta}}(\lambda) &= (K_h^3 + K_{\boldsymbol{\theta}}^3) d_h(\lambda) + (K_h^4 + K_{\boldsymbol{\theta}}^4) d_{\boldsymbol{\theta}}(\lambda), \\
r_h(\lambda) &= (K_h^1 + K_{\boldsymbol{\theta}}^1) d_h(\lambda) + (K_h^2 + K_{\boldsymbol{\theta}}^2) d_{\boldsymbol{\theta}}(\lambda), \\
S_{\boldsymbol{\theta},\boldsymbol{\theta}_1}(\gamma) &= \{ \boldsymbol{a} \in \mathbb{R}^p : \|\boldsymbol{a} - \boldsymbol{\theta}_1\|_{\mathbb{R}^p} \leq \gamma \} \text{ for } \boldsymbol{\theta}_1 \in \mathbb{R}^p, \\
S_{h,h_1}(\gamma) &= \{ u \in \mathcal{H} : \|g - h_1\|_{\mathcal{H},1} \leq \gamma \} \text{ for } h_1 \in \mathcal{H}, \\
S_{\boldsymbol{\theta}}(\gamma) &= S_{\boldsymbol{\theta},\mathbf{0}}(\gamma), \\
S_h(\gamma) &= S_{h,0}(\gamma).
\end{aligned}
$$

One can get the following theorem for the existence and uniqueness of $(\boldsymbol{\theta}_\lambda, h_\lambda)$ via a contraction mapping argument.

Theorem S.2. *If $d_{\boldsymbol{\theta}}(\lambda) \to 0, d_h(\lambda) \to 0, r_{\boldsymbol{\theta}}(\lambda) \to 0, r_h(\lambda) \to 0$ as $\lambda \to 0$, there exists $\lambda_0 > 0$ such that for $\lambda \in [0, \lambda_0]$, there are unique $\boldsymbol{\theta}_\lambda \in S_{\boldsymbol{\theta},\boldsymbol{\theta}_0}(2d_{\boldsymbol{\theta}}(\lambda))$ and $h_\lambda \in S_{h,h_0}(2d_h(\lambda))$ satisfying $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) = 0, Z_h(\boldsymbol{\theta}_\lambda, h_\lambda) = 0$, and $(\boldsymbol{\theta}_\lambda, h_\lambda) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$. In addition, as $\lambda \to 0$,*

$$
\left\| \bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_\lambda \right\|_{\mathbb{R}^p} + \left\| \bar{h}_\lambda - h_\lambda \right\|_{\mathcal{H},1} \leq 4 \left[r_h(\lambda) d_h(\lambda) + r_{\boldsymbol{\theta}}(\lambda) d_{\boldsymbol{\theta}}(\lambda) \right].
$$

Proof. Let $t_{\boldsymbol{\theta}\lambda} = 2d_{\boldsymbol{\theta}}(\lambda)$, $t_{h\lambda} = 2d_h(\lambda)$. Define

$$
\begin{aligned}
F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}, g) &= \boldsymbol{\zeta} - G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) \left[Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) \right], \\
F_h(\boldsymbol{\zeta}, g) &= g - G_{hh}^{-1}(\boldsymbol{\theta}_0, h_0) \left[Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) \right].
\end{aligned}
$$

Let $\vec{F}(\boldsymbol{\zeta}, g) = (F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}, g), F_h(\boldsymbol{\zeta}, g))$ be a function on $\mathcal{Q} = \mathbb{R}^p \times \mathcal{H}$, and for any subset $\mathcal{Q}_1 \subset \mathcal{Q}$, denote by $\vec{F}(\mathcal{Q}_1)$ the image of $\mathcal{Q}_1$ under $\vec{F}$. The proof has three steps:

1. $\vec{F}(S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})) \subset S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})$.

2. $\vec{F}$ is a contraction map on $S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})$.

3. Obtaining the bound for $\|\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_\lambda\|_{\mathbb{R}^p} + \|\bar{h}_\lambda - h_\lambda\|_{\mathcal{H},1}$.

For step 1, by our assumption, we can choose λ_0 small enough that $S_{\boldsymbol{\theta},\boldsymbol{\theta}_0}(t_{\boldsymbol{\theta}\lambda}) \subset \mathcal{N}_{\boldsymbol{\theta}_0}$, $S_{h,h_0}(t_{h\lambda}) \subset \mathcal{N}_{h_0}$, and $r_{\boldsymbol{\theta}}(\lambda) < 1/2$ for all $\lambda \in (0, \lambda_0]$. For every $(\boldsymbol{\theta}, h) \in \mathcal{Q}$, we denote $\|(\boldsymbol{\theta}, h)\|_{\mathcal{Q},*1} = \|\boldsymbol{\theta}\|_{\mathbb{R}^p} + \|h\|_{\mathcal{H},1}$. For $(\boldsymbol{\zeta}, g) \in S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})$, we have

$$\left\|\vec{F}(\boldsymbol{\zeta}, g)\right\|_{\mathcal{Q},*1} = \|F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}, g)\|_{\mathbb{R}^p} + \|F_h(\boldsymbol{\zeta}, g)\|_{\mathcal{H},1}.$$

For $\|F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}, g)\|_{\mathbb{R}^p}$, by the triangle inequality, we have

$$\begin{aligned}\|F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}, g)\|_{\mathbb{R}^p} \leq \big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\,&[Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) \\ &-U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)\big] - (\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0)\big\|_{\mathbb{R}^p} + \big\|\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0\big\|_{\mathbb{R}^p}.\end{aligned}$$

By the definition of $\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0$ and $G_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, h)$, the Taylor series expansions of $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)$ and $Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)$, and the remainder bound (S.12), we get

$$\begin{aligned}&\big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\,[Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) \\ &\quad -U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)\big] - (\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0)\big\|_{\mathbb{R}^p} \\ &= \big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\,\{[Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) - Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)] \\ &\quad\quad -U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0)\,[Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g) - Z_h(\boldsymbol{\theta}_0, h_0)]\}\big\|_{\mathbb{R}^p} \\ &= \big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\,\{[U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} + R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} g] \\ &\quad\quad -U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0)\,[U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} + R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} g]\}\big\|_{\mathbb{R}^p} \\ &= \big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\,\big\{\big[U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0)\big]\,\boldsymbol{\zeta} \\ &\quad\quad + \big[R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) R_h(\boldsymbol{\theta}_0, h_0)\big]\,\boldsymbol{\zeta} g\big\}\big\|_{\mathbb{R}^p} \\ &= \big\|G_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0)\,\big[R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} g - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} g\big]\big\|_{\mathbb{R}^p} \\ &\leq \frac{1}{2}\left(K_{\boldsymbol{\theta}}^1 \|g\|_{\mathcal{H},1} + K_{\boldsymbol{\theta}}^2 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right)\|g\|_{\mathcal{H},1} + \frac{1}{2}\left(K_{\boldsymbol{\theta}}^3 \|g\|_{\mathcal{H},1} + K_{\boldsymbol{\theta}}^4 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right)\|\boldsymbol{\zeta}\|_{\mathbb{R}^p}.\end{aligned}$$

Similarly, by the definition of $\bar{h}_\lambda - h_0$ and $G_{hh}(\boldsymbol{\theta}, h)$, the Taylor series expansions of $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)$ and $Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)$, and the remainder bound (S.11), we also have

$$\begin{aligned}\|F_h(\boldsymbol{\zeta}, g)\|_{\mathcal{H},1} \leq & \left\| g - G_{hh}^{-1}(\boldsymbol{\theta}_0, h_0)\left[Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)\right.\right. \\ & \left.\left. - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)\right] - (\bar{h}_\lambda - h_0)\right\|_{\mathcal{H},1} \\ & + \left\|\bar{h}_\lambda - h_0\right\|_{\mathcal{H},1},\end{aligned}$$

and

$$\begin{aligned}&\left\| g - G_{hh}^{-1}(\boldsymbol{\theta}_0, h_0)\left[Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)\right.\right. \\ &\quad \left.\left. - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}, h_0 + g)\right] - (\bar{h}_\lambda - h_0)\right\|_{\mathcal{H},1} \\ &= \left\| G_{hh}^{-1}(\boldsymbol{\theta}_0, h_0)\left[R_h(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} g - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_0, h_0) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_0, h_0) R_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)\boldsymbol{\zeta} g\right]\right\|_{\mathcal{H},1} \\ &\leq \frac{1}{2}\left(K_h^1 \|g\|_{\mathcal{H},1} + K_h^2 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right)\|g\|_{\mathcal{H},1} + \frac{1}{2}\left(K_h^3 \|g\|_{\mathcal{H},1} + K_h^4 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right)\|\boldsymbol{\zeta}\|_{\mathbb{R}^p}.\end{aligned}$$

Since $t_{\boldsymbol{\theta}\lambda} = 2\left\|\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0\right\|_{\mathbb{R}^p}$, $t_{h\lambda} = 2\left\|\bar{h}_\lambda - h_0\right\|_{\mathcal{H},1}$, $r_h(\lambda) < 1/2$, and $r_{\boldsymbol{\theta}}(\lambda) < 1/2$, for $(\boldsymbol{\zeta}, g) \in S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})$, we have

$$\begin{aligned}\left\|\vec{F}(\boldsymbol{\zeta}, g)\right\|_{\mathcal{Q},*1} \leq & \frac{1}{2}\left[\left(K_h^1 + K_{\boldsymbol{\theta}}^1\right)\|g\|_{\mathcal{H},1} + \left(K_h^2 + K_{\boldsymbol{\theta}}^2\right)\|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right]\|g\|_{\mathcal{H},1} \\ & + \frac{1}{2}\left[\left(K_h^3 + K_{\boldsymbol{\theta}}^3\right)\|g\|_{\mathcal{H},1} + \left(K_h^4 + K_{\boldsymbol{\theta}}^4\right)\|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right]\|\boldsymbol{\zeta}\|_{\mathbb{R}^p} \\ & + \left\|\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0\right\|_{\mathbb{R}^p} + \left\|\bar{h}_\lambda - h_0\right\|_{\mathcal{H},1} \\ \leq & \frac{1}{2}\left[\left(K_h^1 + K_{\boldsymbol{\theta}}^1\right)t_{h\lambda} + \left(K_h^2 + K_{\boldsymbol{\theta}}^2\right)t_{\boldsymbol{\theta}\lambda}\right]t_{h\lambda} \\ & + \frac{1}{2}\left[\left(K_h^3 + K_{\boldsymbol{\theta}}^3\right)t_{h\lambda} + \left(K_h^4 + K_{\boldsymbol{\theta}}^4\right)t_{\boldsymbol{\theta}\lambda}\right]t_{\boldsymbol{\theta}\lambda} + \frac{1}{2}t_{\boldsymbol{\theta}\lambda} + \frac{1}{2}t_{h\lambda} \\ = & \, r_h(\lambda)t_{h\lambda} + r_{\boldsymbol{\theta}}(\lambda)t_{\boldsymbol{\theta}\lambda} + \frac{1}{2}t_{\boldsymbol{\theta}\lambda} + \frac{1}{2}t_{h\lambda} \\ = & \left(r_h(\lambda) + \frac{1}{2}\right)t_{h\lambda} + \left(r_{\boldsymbol{\theta}}(\lambda) + \frac{1}{2}\right)t_{\boldsymbol{\theta}\lambda} \\ < & \, t_{h\alpha} + t_{\boldsymbol{\theta}\alpha}.\end{aligned}$$

Now for step 2, by Taylor expansion, we get that for $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda})$, $g_1, g_2 \in S_h(t_{h\lambda})$,

$$\begin{aligned}Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_2, h_0 + g_2) &= Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_1, h_0 + g_1) \\ &\quad + \int_0^1 D_{\boldsymbol{\theta}} Z_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_1 + t(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1), h_0 + g_1 + t(g_2 - g_1)\right](\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1) \\ &\qquad + D_h Z_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_1 + t(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1), h_0 + g_1 + t(g_2 - g_1)\right](g_2 - g_1)\, dt.\end{aligned}$$

Applying Taylor expansion again to the terms inside the integral and letting $\boldsymbol{\zeta}^* = \boldsymbol{\zeta}_1 + t(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1)$, $g^* = g_1 + t(g_2 - g_1)$, we have

$$
\begin{aligned}
&Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_2, h_0 + g_2) - Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_1, h_0 + g_1) \\
&\quad = U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1) + U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0)(g_2 - g_1) \\
&\qquad + \int_0^1 \int_0^1 \left[D^2_{\boldsymbol{\theta\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + t'\boldsymbol{\zeta}^*, h_0 + t'g^*)\boldsymbol{\zeta}^* + D^2_{h\boldsymbol{\theta}} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + t'\boldsymbol{\zeta}^*, h_0 + t'g^*)g^* \right] (\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1)\ dt'dt \\
&\qquad + \int_0^1 \int_0^1 \left[D^2_{\boldsymbol{\theta}h} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + t'\boldsymbol{\zeta}^*, h_0 + t'g^*)\boldsymbol{\zeta}^* + D^2_{hh} Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + t'\boldsymbol{\zeta}^*, h_0 + t'g^*)g^* \right] (g_2 - g_1)\ dt'dt.
\end{aligned}
$$

Note that for $0 \le t \le 1$, $\boldsymbol{\zeta}^* = \boldsymbol{\zeta}_1 + t(\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1) \in S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda})$, $g^* = g_1 + t(g_2 - g_1) \in S_h(t_{h\lambda})$ by convexity of $S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda})$ and $S_h(t_{h\lambda})$. Since

$$
\begin{aligned}
F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}_1, g_1) - F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}_2, g_2) = (\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2) - G^{-1}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0, h_0) \{ &[Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_1, h_0 + g_1) - Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_2, h_0 + g_2)] \\
&- U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_0, h_0) G_h^{-1}(\boldsymbol{\theta}_0, h_0) [Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_1, h_0 + g_1) - Z_h(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}_2, h_0 + g_2)] \},
\end{aligned}
$$

similar algebraic manipulations as in the proof of step 1 show that

$$
\begin{aligned}
\| F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}_1, g_1) - F_{\boldsymbol{\theta}}(\boldsymbol{\zeta}_2, g_2) \|_{\mathbb{R}^p} \le &\left(K^3_{\boldsymbol{\theta}} \| g^* \|_{\mathcal{H},1} + K^4_{\boldsymbol{\theta}} \| \boldsymbol{\zeta}^* \|_{\mathbb{R}^p} \right) \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p} \\
&+ \left(K^2_{\boldsymbol{\theta}} \| \boldsymbol{\zeta}^* \|_{\mathbb{R}^p} + K^1_{\boldsymbol{\theta}} \| g^* \|_{\mathcal{H},1} \right) \| g_2 - g_1 \|_{\mathcal{H},1}.
\end{aligned}
$$

Similarly for F_h, we get

$$
\begin{aligned}
\| F_h(\boldsymbol{\zeta}_1, g_1) - F_h(\boldsymbol{\zeta}_2, g_2) \|_{\mathcal{H},1} \le &\left(K^3_h \| g^* \|_{\mathcal{H},1} + K^4_h \| \boldsymbol{\zeta}^* \|_{\mathbb{R}^p} \right) \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p} \\
&+ \left(K^2_h \| \boldsymbol{\zeta}^* \|_{\mathbb{R}^p} + K^1_h \| g^* \|_{\mathcal{H},1} \right) \| g_2 - g_1 \|_{\mathcal{H},1}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\left\| \vec{F}(\boldsymbol{\zeta}_1, g_1) - \vec{F}(\boldsymbol{\zeta}_2, g_2) \right\|_{\mathcal{Q},*1} \\
&\quad \le (K^1_h + K^1_{\boldsymbol{\theta}}) \| g^* \|_{\mathcal{H},1} \| g_2 - g_1 \|_{\mathcal{H},1} + (K^2_h + K^2_{\boldsymbol{\theta}}) \| \boldsymbol{\zeta}^* \|_{\mathbb{R}^p} \| g_2 - g_1 \|_{\mathcal{H},1} \\
&\qquad + (K^3_h + K^3_{\boldsymbol{\theta}}) \| g^* \|_{\mathcal{H},1} \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p} + (K^4_h + K^4_{\boldsymbol{\theta}}) \| \boldsymbol{\zeta}^* \|_{\mathbb{R}^p} \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p} \\
&\quad = 2 \left[(K^1_h + K^1_{\boldsymbol{\theta}}) d_h(\lambda) + (K^2_h + K^2_{\boldsymbol{\theta}}) d_{\boldsymbol{\theta}}(\lambda) \right] \| g_2 - g_1 \|_{\mathcal{H},1} \\
&\qquad + 2 \left[(K^3_h + K^3_{\boldsymbol{\theta}}) d_h(\lambda) + (K^4_h + K^4_{\boldsymbol{\theta}}) d_{\boldsymbol{\theta}}(\lambda) \right] \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p} \\
&\quad = 2 r_h(\lambda) \| g_2 - g_1 \|_{\mathcal{H},1} + 2 r_{\boldsymbol{\theta}}(\lambda) \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p} \\
&\quad \le C_1 \| g_2 - g_1 \|_{\mathcal{H},1} + C_2 \| \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1 \|_{\mathbb{R}^p},
\end{aligned}
$$

where $0 < C_1 < 1$, $0 < C_2 < 1$, so $\vec{F} = (F_{\boldsymbol{\theta}}, F_h)$ is a contraction map on $S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})$. By the contraction mapping theorem (Theorem 9.23 in Rudin (1976)), there exists a unique $(\boldsymbol{\zeta}_\lambda, g_\lambda) \in S_{\boldsymbol{\theta}}(t_{\boldsymbol{\theta}\lambda}) \times S_h(t_{h\lambda})$ such that $\vec{F}(\boldsymbol{\zeta}_\lambda, g_\lambda) = (\boldsymbol{\zeta}_\lambda, g_\lambda)$. Let $\boldsymbol{\theta}_\lambda = \boldsymbol{\theta}_0 + \boldsymbol{\zeta}_\lambda$, $h_\lambda = h_0 + g_\lambda$. Then $\boldsymbol{\theta}_\lambda \in S_{\boldsymbol{\theta},\boldsymbol{\theta}_0}(t_{\boldsymbol{\theta}\lambda})$, $h_\lambda \in S_{h,h_0}(t_{\boldsymbol{\theta}\lambda})$, and $(\boldsymbol{\theta}_\lambda, h_\lambda)$ are the unique solutions to $Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) = 0$, $Z_h(\boldsymbol{\theta}_\lambda, h_\lambda) = 0$.

For step 3, note that

$$\begin{aligned}(\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_\lambda, \bar{h}_\lambda - h_\lambda) &= (\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_0, \bar{h}_\lambda - h_0) - (\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}_0, h_\lambda - h_0) \\ &= \vec{F}(0,0) - \vec{F}(\boldsymbol{\zeta}_\lambda, g_\lambda).\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\bar{\boldsymbol{\theta}}_\lambda - \boldsymbol{\theta}_\lambda\right\|_{\mathbb{R}^p} + \left\|\bar{h}_\lambda - h_\lambda\right\|_{\mathcal{H},1} &= \left\|\vec{F}(\boldsymbol{\zeta}_\lambda, g_\lambda) - \vec{F}(0,0)\right\|_{\mathcal{Q},*1} \\ &\leq 2r_h(\lambda)\left\|g_\lambda\right\|_{\mathcal{H},1} + 2r_{\boldsymbol{\theta}}(\lambda)\left\|\boldsymbol{\zeta}_\lambda\right\|_{\mathbb{R}^p} \\ &\leq 4\left[r_h(\lambda)d_h(\lambda) + r_{\boldsymbol{\theta}}(\lambda)d_{\boldsymbol{\theta}}(\lambda)\right].\end{aligned}$$

This completes the proof of Theorem S.2. ■

Next, we consider the existence of $(\hat{\boldsymbol{\theta}}, \hat{h}) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$. Define

$$\begin{aligned}d_{n\boldsymbol{\theta}}(\lambda) &= \left\|\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda\right\|_{\mathbb{R}^p}, \\ d_{nh}(\lambda) &= \left\|\bar{h}_{n\lambda} - h_\lambda\right\|_{\mathcal{H},1}, \\ r_{n\boldsymbol{\theta}}(\lambda) &= E^2_{n\boldsymbol{\theta}} + E^2_{nh} + (K^3_{n\boldsymbol{\theta}} + K^3_{nh})d_{nh}(\lambda) + (K^4_{n\boldsymbol{\theta}} + K^4_{nh})d_{n\boldsymbol{\theta}}(\lambda), \\ r_{nh}(\lambda) &= E^1_{n\boldsymbol{\theta}} + E^1_{nh} + (K^1_{n\boldsymbol{\theta}} + K^1_{nh})d_{nh}(\lambda) + (K^2_{n\boldsymbol{\theta}} + K^2_{nh})d_{n\boldsymbol{\theta}}(\lambda).\end{aligned}$$

We get the following existence theorem for $(\hat{\boldsymbol{\theta}}, \hat{h}) \in \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$.

Theorem S.3. *Suppose λ_n is a sequence such that for all n sufficiently large, $\boldsymbol{\theta}_{\lambda_n} \in \mathcal{N}_{\boldsymbol{\theta}_0}$, $h_{\lambda_n} \in \mathcal{N}_{h_0}$, and*

$$\begin{aligned}d_{n\boldsymbol{\theta}}(\lambda_n) &\xrightarrow{P} 0, & d_{nh}(\lambda_n) &\xrightarrow{P} 0, \\ r_{n\boldsymbol{\theta}}(\lambda_n) &\xrightarrow{P} 0, & r_{nh}(\lambda_n) &\xrightarrow{P} 0.\end{aligned}$$

Then, with probability tending to unity as $n \to \infty$, there is a unique root $(\hat{\boldsymbol{\theta}}, \hat{h})$ of $Z_{n\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \hat{h}) = 0$, $Z_{nh}(\hat{\boldsymbol{\theta}}, \hat{h}) = 0$ in $S_{\boldsymbol{\theta}, \boldsymbol{\theta}_{\lambda_n}}(2d_{n\boldsymbol{\theta}}(\lambda_n)) \times S_{h,h_{\lambda_n}}(2d_{nh}(\lambda_n)) \subset \mathcal{N}_{\boldsymbol{\theta}_0} \times \mathcal{N}_{h_0}$. In addition, as $n \to \infty$ and $\lambda_n \to 0$,

$$\left\|\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{n\lambda_n}\right\|_{\mathbb{R}^p} + \left\|\hat{h} - \bar{h}_{n\lambda_n}\right\|_{\mathcal{H},1} \leq 4r_{n\boldsymbol{\theta}}(\lambda_n)d_{n\boldsymbol{\theta}}(\lambda_n) + 4r_{nh}(\lambda_n)d_{nh}(\lambda_n).$$

Proof. For convenience, we drop the subscript on λ_n and let $t_{n\boldsymbol{\theta}\lambda} = 2d_{n\boldsymbol{\theta}}(\lambda)$, $t_{nh\lambda} = 2d_{nh}(\lambda)$. Let

$$\begin{aligned}
F_{n\boldsymbol{\theta}}(\boldsymbol{\zeta}, g) &= \boldsymbol{\zeta} - G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\left[Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\right. \\
&\qquad \left. - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\right], \\
F_{nh}(\boldsymbol{\zeta}, g) &= g - G_{hh}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\left[Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\right. \\
&\qquad \left. - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda)U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\right].
\end{aligned}$$

The proof proceeds in three steps, similar to the proof of Theorem S.2, with additional terms introduced in approximating $D_{\boldsymbol{\theta}}Z_{n\boldsymbol{\theta}}$ and $D_h Z_{nh}$ by $D_{\boldsymbol{\theta}}Z_{\boldsymbol{\theta}}$ and $D_h Z_h$, respectively. Take n large enough so that $S_{\boldsymbol{\theta}, \boldsymbol{\theta}_\lambda}(t_{n\boldsymbol{\theta}\lambda}) \subset \mathcal{N}_{\boldsymbol{\theta}_0}$, $S_{h,h_\lambda}(t_{nh\lambda}) \subset \mathcal{N}_{h_0}$ and $r_{n\boldsymbol{\theta}}(\lambda) < \frac{1}{2}$, $r_{nh}(\lambda) < \frac{1}{2}$.

First, we show that $\vec{F}_n(\boldsymbol{\zeta}, g) = (F_{n\boldsymbol{\theta}}(\boldsymbol{\zeta}, g), F_{nh}(\boldsymbol{\zeta}, g))$ maps $S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})$ to itself, i.e., $\vec{F}_n\left(S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})\right) \subset S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})$. By definition, for $(\boldsymbol{\zeta}, g) \in S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})$, we have

$$\left\|\vec{F}_n(\boldsymbol{\zeta}, g)\right\|_{\mathcal{Q},*1} = \|F_{n\boldsymbol{\theta}}(\boldsymbol{\zeta}, g)\|_{\mathbb{R}^p} + \|F_{nh}(\boldsymbol{\zeta}, g)\|_{\mathcal{H},1}.$$

For $F_{n\boldsymbol{\theta}}$, by the triangle inequality, we get

$$\begin{aligned}
\|F_{n\boldsymbol{\theta}}(\boldsymbol{\zeta}, g)\|_{\mathbb{R}^p} \leq &\left\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\left[Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\right.\right. \\
&\left. - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\right] \\
&\left. - (\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda)\right\|_{\mathbb{R}^p} + \left\|\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda\right\|_{\mathbb{R}^p}
\end{aligned}$$

Using the definition of $\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda$, $G_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}, h)$, the Taylor expansions of $Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)$

and $Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)$, and the remainder bound (S.14), we get

$$\begin{aligned}
&\big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta\theta}}^{-1}((\boldsymbol{\theta}_\lambda, h_\lambda)\,[Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g) \\
&\qquad - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\big] - (\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda)\big\|_{\mathbb{R}^p} \\
&= \big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,\{[Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g) - Z_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)] \\
&\qquad - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,[Z_{nh}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g) - Z_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda)]\}\big\|_{\mathbb{R}^p} \\
&= \big\|\boldsymbol{\zeta} - G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,\big[U_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) - (U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)\big]\,\boldsymbol{\zeta} \\
&\qquad - G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,[e_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)]\,\boldsymbol{\zeta} g \\
&\qquad + G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,[e_h(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda)]\,\boldsymbol{\zeta} g\big\|_{\mathbb{R}^p} \\
&= \big\|G_{\boldsymbol{\theta\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,\{[e_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{n\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda)]\,\boldsymbol{\zeta} g \\
&\qquad - U_{h\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda, h_\lambda) G_h^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,[e_h(\boldsymbol{\theta}_\lambda, h_\lambda) + R_{nh}(\boldsymbol{\theta}_\lambda, h_\lambda)]\,\boldsymbol{\zeta} g\}\big\|_{\mathbb{R}^p} \\
&\le E_{n\boldsymbol{\theta}}^1 \|g\|_{\mathcal{H},1} + E_{n\boldsymbol{\theta}}^2 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p} + \frac{1}{2}\left(K_{n\boldsymbol{\theta}}^1 \|g\|_{\mathcal{H},1} + K_{n\boldsymbol{\theta}}^2 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right)\|g\|_{\mathcal{H},1} \\
&\qquad + \frac{1}{2}\left(K_{n\boldsymbol{\theta}}^3 \|g\|_{\mathcal{H},1} + K_{n\boldsymbol{\theta}}^4 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right)\|\boldsymbol{\zeta}\|_{\mathbb{R}^p}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\|F_{nh}(\boldsymbol{\zeta}, g)\|_{\mathcal{H},1} \le \|g - G_{hh}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,&[Z_h(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g) \\
&- U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\big] \\
&- (\bar{h}_{n\lambda} - h_\lambda)\|_{\mathcal{H},1} + \big\|\bar{h}_{n\lambda} - h_\lambda\big\|_{\mathcal{H},1},
\end{aligned}$$

and

$$\begin{aligned}
&\|g - G_{hh}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda)\,[Z_h(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g) \\
&\qquad - U_{\boldsymbol{\theta}h}(\boldsymbol{\theta}_\lambda, h_\lambda) U_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}_\lambda, h_\lambda) Z_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}, h_\lambda + g)\big] - (\bar{h}_{n\lambda} - h_\lambda)\|_{\mathcal{H},1} \\
&\le E_{nh}^1 \|g\|_{\mathcal{H},1} + E_{nh}^2 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p} + \frac{1}{2}(K_{nh}^1 \|g\|_{\mathcal{H},1} + K_{nh}^2 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p})\|g\|_{\mathcal{H},1} \\
&\qquad + \frac{1}{2}(K_{nh}^3 \|g\|_{\mathcal{H},1} + K_{nh}^4 \|\boldsymbol{\zeta}\|_{\mathbb{R}^p})\|\boldsymbol{\zeta}\|_{\mathbb{R}^p}.
\end{aligned}$$

Thus, for $(\boldsymbol{\zeta}, g) \in S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})$,

$$
\begin{aligned}
&\left\|\vec{F}(\boldsymbol{\zeta}, g)\right\|_{\mathcal{Q},*1} \\
&\leq \left[(E^1_{n\boldsymbol{\theta}} + E^1_{nh}) + \frac{1}{2}(K^1_{n\boldsymbol{\theta}} + K^1_{nh}) \|g\|_{\mathcal{H},1} + \frac{1}{2}(K^2_{n\boldsymbol{\theta}} + K^2_{nh}) \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right] \|g\|_{\mathcal{H},1} \\
&\quad + \left[(E^2_{n\boldsymbol{\theta}} + E^2_{nh}) + \frac{1}{2}(K^3_{n\boldsymbol{\theta}} + K^3_{nh}) \|g\|_{\mathcal{H},1} + \frac{1}{2}(K^4_{n\boldsymbol{\theta}} + K^4_{nh}) \|\boldsymbol{\zeta}\|_{\mathbb{R}^p}\right] \|\boldsymbol{\zeta}\|_{\mathbb{R}^p} \\
&\quad + \left\|\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda\right\|_{\mathbb{R}^p} + \left\|\bar{h}_{n\lambda} - h_\lambda\right\|_{\mathcal{H},1} \\
&\leq r_{nh}(\lambda) t_{nh\lambda} + r_{n\boldsymbol{\theta}}(\lambda) t_{n\boldsymbol{\theta}\lambda} + \frac{1}{2} t_{n\boldsymbol{\theta}\lambda} + \frac{1}{2} t_{nh\lambda} \\
&= \left[r_{nh}(\lambda) + \frac{1}{2}\right] t_{nh\lambda} + \left[r_{n\boldsymbol{\theta}}(\lambda) + \frac{1}{2}\right] t_{n\boldsymbol{\theta}\lambda} \\
&< t_{nh\lambda} + t_{n\boldsymbol{\theta}\lambda}.
\end{aligned}
$$

Therefore, we have shown that $\vec{F}_n(S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})) \subset S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})$.

Next, we show that $\vec{F}_n$ is a contraction map. By similar calculations as in the proof for Theorem S.2, after applying Taylor expansion twice, for $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda})$, $g_1, g_2 \in S_h(t_{nh\lambda})$, we get

$$
\begin{aligned}
\|F_{n\boldsymbol{\theta}}(\boldsymbol{\zeta}_1, g_1) - F_{n\boldsymbol{\theta}}(\boldsymbol{\zeta}_2, g_2)\|_{\mathbb{R}^p} \leq &\left(E^2_{n\boldsymbol{\theta}} + K^3_{n\boldsymbol{\theta}} t_{nh\lambda} + K^4_{n\boldsymbol{\theta}} t_{n\boldsymbol{\theta}\lambda}\right) \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|_{\mathbb{R}^p} \\
&+ \left(E^1_{n\boldsymbol{\theta}} + K^2_{n\boldsymbol{\theta}} t_{n\boldsymbol{\theta}\lambda} + K^1_{n\boldsymbol{\theta}} t_{nh\lambda}\right) \|g_1 - g_2\|_{\mathcal{H},1},
\end{aligned}
$$

$$
\begin{aligned}
\|F_{nh}(\boldsymbol{\zeta}_1, g_1) - F_{nh}(\boldsymbol{\zeta}_2, g_2)\|_{\mathcal{H},1} \leq &\left(E^2_{nh} + K^3_{nh} t_{nh\lambda} + K^4_{nh} t_{n\boldsymbol{\theta}\lambda}\right) \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|_{\mathbb{R}^p} \\
&+ \left(E^1_{nh} + K^2_{nh} t_{n\boldsymbol{\theta}\lambda} + K^1_{nh} t_{nh\lambda}\right) \|g_1 - g_2\|_{\mathcal{H},1}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\left\|\vec{F}_n(\boldsymbol{\zeta}_1, g_1) - \vec{F}_n(\boldsymbol{\zeta}_2, g_2)\right\|_{\mathcal{Q},*1} \\
&\quad \leq \left[E^2_{n\boldsymbol{\theta}} + E^2_{nh} + (K^3_{n\boldsymbol{\theta}} + K^3_{nh}) t_{nh\lambda} + (K^4_{n\boldsymbol{\theta}} + K^4_{nh}) t_{n\boldsymbol{\theta}\lambda}\right] \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|_{\mathbb{R}^p} \\
&\qquad + \left[E^1_{n\boldsymbol{\theta}} + E^1_{nh} + (K^1_{n\boldsymbol{\theta}} + K^1_{nh}) t_{nh\lambda} + (K^2_{n\boldsymbol{\theta}} + K^2_{nh}) t_{n\boldsymbol{\theta}\lambda}\right] \|g_1 - g_2\|_{\mathcal{H},1} \\
&\quad \leq 2 r_{n\boldsymbol{\theta}}(\lambda) \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|_{\mathbb{R}^p} + 2 r_{nh}(\lambda) \|g_1 - g_2\|_{\mathcal{H},1}.
\end{aligned}
$$

Since $r_{n\lambda}(\lambda) < \frac{1}{2}$, $r_{nh}(\lambda) < \frac{1}{2}$, we have shown that $\vec{F}(\boldsymbol{\zeta}, g)$ is a contraction map on $S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}) \times S_h(t_{nh\lambda})$. By the contraction mapping theorem, there exists a unique $(\boldsymbol{\zeta}_{n\lambda}, g_{n\lambda}) \in S_{\boldsymbol{\theta}}(t_{n\boldsymbol{\theta}\lambda}, \alpha) \times$

$S_h(t_{nh\lambda}, \alpha)$ such that $\vec{F}_n(\boldsymbol{\zeta}_{n\lambda}, g_{n\lambda}) = (\boldsymbol{\zeta}_{n\lambda}, g_{n\lambda})$. Let $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_\lambda + \boldsymbol{\zeta}_{n\lambda} \in S_{\boldsymbol{\theta},\boldsymbol{\theta}_\lambda}(2d_{n\boldsymbol{\theta}}(\lambda))$ and $\hat{h} = h_\lambda + g_{n\lambda} \in S_{h,h_\lambda}(2d_{nh}(\lambda))$. Then $(\hat{\boldsymbol{\theta}}, \hat{h})$ is the unique root of $Z_{n\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \hat{h}) = 0$ and $Z_{nh}(\hat{\boldsymbol{\theta}}, \hat{h}) = 0$.

To get the upper bound, we observe that

$$\begin{aligned}(\bar{\boldsymbol{\theta}}_{n\lambda} - \hat{\boldsymbol{\theta}}, \bar{h}_{n\lambda} - \hat{h}) &= (\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_\lambda, \bar{h}_{n\lambda} - h_\lambda) - (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\lambda, \hat{h} - h_\lambda) \\ &= \vec{F}_n(0,0) - \vec{F}_n(\boldsymbol{\zeta}_{n\lambda}, g_{n\lambda}).\end{aligned}$$

Therefore,

$$\begin{aligned}\left\|\bar{\boldsymbol{\theta}}_{n\lambda} - \boldsymbol{\theta}_{n\lambda}\right\|_{\mathbb{R}^p} + \left\|\bar{h}_{n\lambda} - h_{n\lambda}\right\|_{\mathcal{H},1} &= \left\|\vec{F}_n(\boldsymbol{\zeta}_{n\lambda}, g_{n\lambda}) - \vec{F}_n(0,0)\right\|_{\mathcal{Q},1} \\ &\le 2r_{n\boldsymbol{\theta}}(\lambda) \left\|\boldsymbol{\zeta}_{n\lambda}\right\|_{\mathbb{R}^p} + 2r_{nh}(\lambda) \left\|g_{n\lambda}\right\|_{\mathcal{H},1} \\ &\le 4\left[r_{n\boldsymbol{\theta}}(\lambda) d_{n\boldsymbol{\theta}}(\lambda) + r_{nh}(\lambda) d_{nh}(\lambda)\right].\end{aligned}$$

■

S.6 Additional simulations

S.6.1 Additive case: Near Gumbel distribution

In this section, we consider simulations from the density function given by (4.2) in the main text with $\alpha_0(x;\mu,\sigma) = -[(x-\mu)/\sigma] - \exp\{(x-\mu)/\sigma\}$, which can be viewed as the logistic transformation of a truncated Gumbel distribution with $\mu = 0.5$ and $\sigma = 0.2$. We will consider the additive model (2.4) with $\alpha(x;\mu,\sigma) = -[(x-\mu)/\sigma] - \exp\{(x-\mu)/\sigma\}$ and $h \in W_2^2[0,1] \ominus \{1\}$. Note that α is nonlinear in $\boldsymbol{\theta} = (\mu,\sigma)$. Therefore, the method developed for linear additive models as in Yang (2009) does not apply. We compare our proposed method with the kernel, cubic spline and HG's method, where the starting parametric density f_0 in the HG's approach is a truncated Gumbel distribution. We use the profile likelihood approach as described in Section 2.2 to compute estimates of $\boldsymbol{\theta} = (\mu,\sigma)$ and h.

We consider three choices of a for (4.2), $a = 0.25, 1, 4$, and three sample sizes $n = 100, 200, 500$. For each simulation setting, we generate 100 simulated data sets. Table 1 shows that our semiparametric method has smaller KL distances when the true density is close to the truncated Gumbel (i.e. $a = 0.25$ and 1). When the true density is far away from the truncated Gumbel ($a = 4$), the cubic spline has smaller KL distances. Compared to the other semiparametric approach (HG) given by Hjort and Glad (1995), our semiparametric approach (SEMI) performs better for each simulation setting.

Table 1: Additive Case with $\alpha_0(x; \mu, \sigma) = -[(x - \mu)/\sigma] - \exp\{(x - \mu)/\sigma\}$, $\mu = 0.5$ and $\sigma = 0.2$ (Kullback-Leibler distances of different methods. Numbers inside parentheses of KLs are bias and variances.)

a	Method	KL $n = 100$	KL $n = 200$	KL $n = 500$
0.25	Kernel	3.50(0.41, 3.09)	2.60(0.36, 2.23)	1.72(0.21, 1.51)
	HG	3.20(1.12, 2.08)	1.89(0.68, 1.21)	1.09(0.49, 0.61)
	Cubic	2.12(0.59, 1.53)	1.32(0.42, 0.90)	0.63(0.19, 0.44)
	SEMI	1.58(0.16, 1.42)	0.65(0.01, 0.63)	0.45(0.04, 0.41)
1	Kernel	4.32(0.47, 3.85)	3.22(0.49, 2.73)	2.36(0.31, 2.04)
	HG	3.07(1.00, 2.07)	2.04(1.02, 1.02)	1.18(0.61, 0.58)
	Cubic	2.17(0.69, 1.48)	1.21(0.52, 0.69)	0.64(0.21, 0.43)
	SEMI	1.28(0.11, 1.17)	0.64(0.06, 0.58)	0.44(0.05, 0.40)
4	Kernel	11.47(1.86, 9.61)	9.67(1.88, 7.80)	7.68(1.53, 6.15)
	HG	7.83(5.85, 1.98)	5.27(4.08, 1.19)	3.51(3.04, 0.46)
	Cubic	2.00(0.65, 1.35)	1.08(0.37, 0.71)	0.52(0.19, 0.33)
	SEMI	2.33(0.58, 1.75)	1.31(0.32, 0.98)	0.65(0.13, 0.53)

S.6.2 Two-sample density estimation with Gumbel distribution

When $f(x; \mu, \sigma)$ in Section 4.2.2 is given by the density $f_G(x; \mu, \sigma)$ of the Gumbel distribution, i.e.

$$f_G(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\frac{x - \mu}{\sigma} + \exp\left[-\frac{x - \mu}{\sigma}\right]\right\},$$

for $x \in \mathbb{R}$. Simulation results are given in Table 2. As discussed in Section 4.2.2 in the main text, our semiparametric method provides better density estimation than the method in Potgieter and Lombard (2012) (CHAR) and the separate thin-spline estimates (TP). All methods improved as the sample size increases.

Table 2: Two-sample density estimation with simulations from logistic distribution (Kullback-Leibler distances and mean squared errors of different methods. Numbers inside parentheses of KLs are bias and variances. Numbers inside parentheses of MSEs are squared biases and variances.)

n_1	n_2	Method	Overall KL	KL_s	MSE($\hat{\mu}$)	MSE($\hat{\sigma}$)
100	100	SEMI	5.36(0.92, 4.44)	2.56(0.49, 2.07)	3.58(0.00, 3.58)	2.34(0.00, 2.34)
		CHAR	13.39(10.06, 3.33)	6.86(3.53, 3.33)	2.24(0.00, 2.24)	1.95(0.00, 1.95)
		TP	5.81(1.95, 3.86)			
	200	SEMI	4.27(0.79, 3.48)	2.56(0.41, 2.15)	2.31(0.01, 2.30)	1.41(0.00, 1.41)
		CHAR	11.07(8.82, 2.25)	6.26(4.02, 2.24)	1.58(0.00, 1.58)	1.34(0.00, 1.34)
		TP	4.95(1.61, 3.34)			
200	100	SEMI	3.84(0.87, 2.97)	1.60(0.41, 1.19)	2.63(0.11, 2.52)	1.29(0.07, 1.22)
		CHAR	10.67(8.83, 1.84)	4.69(2.86, 1.83)	1.52(0.04, 1.48)	0.97(0.03, 0.94)
		TP	4.58(1.57, 3.01)			
	200	SEMI	3.21(0.85, 2.36)	1.46(0.32, 1.14)	1.78(0.00, 1.78)	0.86(0.00, 0.86)
		CHAR	8.34(6.82, 1.52)	4.07(2.55, 1.52)	1.65(0.00, 1.65)	0.74(0.01, 0.73)
		TP	3.60(1.16, 2.44)			

References

Ambrosetti, A. and G. Prodi (1995). *A Primer of Nonlinear Analysis*, Volume 34. Cambridge University Press, Cambridge.

Aubin, J.-P. (1979). *Applied Functional Analysis*. John Wiley & Sons, New York-Chichester-Brisbane.

Cheng, G. and Z. Shang (2015). Joint asymptotics for semi-nonparametric regression models with partially linear structure. *Annals of Statistics 43*(3), 1351–1390.

Cox, D. D. (1988). Approximation of method of regularization estimators. *Annals of Statistics 16*(2), 694–712.

Cox, D. D. and F. O'Sullivan (1990). Asymptotic analysis of penalized likelihood and related estimators. *Annals of Statistics 18*, 1676–1695.

Gu, C. and C. Qiu (1993). Smoothing spline density estimation: theory. *Annals of Statistics 21*(1), 217–234.

Hjort, N. L. and I. K. Glad (1995). Nonparametric density estimation with a parametric start. *Annals of Statistics 23*(3), 882–904.

Ke, C. and Y. Wang (2004). Existence and uniqueness of penalized least square estimation for smoothing spline nonlinear nonparametric regression models. Unpublished Manuscript.

O'Sullivan, F. (1990). Convergence characteristics of methods of regularization estimators for nonlinear operator equations. *SIAM Journal on Numerical Analysis 27*(6), 1635–1649.

Potgieter, C. J. and F. Lombard (2012). Nonparametric estimation of location and scale parameters. *Computational Statistics & Data Analysis 56*(12), 4327–4337.

Rudin, W. (1976). *Principles of Mathematical Analysis* (Third ed.). McGraw-Hill Book Co., New York-Auckland-Düsseldorf.

Yang, Y. (2009). Penalized semiparametric density estimation. *Statistics and Computing 19*(4), 355–366.