An RKHS Approach to Robust Functional Linear Regression

Samsung Software R&D Center and Hankuk University of Foreign Studies

Supplementary Material

We prove the main results in this supplementary note.

S1 Proof of Proposition 1

Differentiating (5) with respect to $\boldsymbol{c} = (c_1, \ldots, c_n)^T$ yields

$$-\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{y_i - \alpha - \sum_{l=1}^{L} d_l \int_{\mathcal{T}} x_i(t)\theta_l(t)dt - \sum_{j=1}^{n} c_j \langle \xi_i, \xi_j\rangle_{\mathcal{H}}}{\hat{\sigma}}\right)\frac{\langle \xi_i, \xi_k\rangle_{\mathcal{H}}}{\hat{\sigma}} + 2\lambda\sum_{i=1}^{n} c_i \langle \xi_i, \xi_k\rangle_{\mathcal{H}} = 0 \tag{S1.1}$$

for $k = 1, \ldots, n$. For $\beta \in S_n$, (S1.1) is written as

$$\left\langle -\frac{1}{n}\sum_{i=1}^{n}\xi_i\psi\left(\frac{y_i - \alpha - \langle \eta_i, \beta\rangle_{\mathcal{H}}}{\hat{\sigma}}\right)\frac{1}{\hat{\sigma}} + 2\lambda P_1\beta, \xi_k\right\rangle_{\mathcal{H}} = 0, \quad k = 1, \ldots, n,$$

which implies that $-(n\hat{\sigma})^{-1}\sum_{i=1}^{n}\xi_i\psi\left(\frac{y_i - \alpha - \langle \eta_i, \beta\rangle_{\mathcal{H}}}{\hat{\sigma}}\right) + 2\lambda P_1\beta$ is an element in $\mathcal{H}_1$ perpendicular to $\xi_1, \ldots, \xi_n$. However, $-(n\hat{\sigma})^{-1}\sum_{i=1}^{n}\xi_i\psi\left(\frac{y_i - \alpha - \langle \eta_i, \beta\rangle_{\mathcal{H}}}{\hat{\sigma}}\right) + 2\lambda P_1\beta$ belongs to span$\{\xi_1, \ldots, \xi_n\}$, so

$$-\frac{1}{n}\sum_{i=1}^{n}\xi_i\psi\left(\frac{y_i - \alpha - \langle \eta_i, \beta\rangle_{\mathcal{H}}}{\hat{\sigma}}\right)\frac{1}{\hat{\sigma}} + 2\lambda P_1\beta = 0 \tag{S1.2}$$

holds for a minimizer $\hat{\beta}_{n\lambda}$. Also, differentiating (5) with respect to $\boldsymbol{d} = (d_1, \ldots, d_L)^T$ yields

$$\sum_{i=1}^{n}\psi\left(\frac{y_i - \alpha - \langle \eta_i, \beta\rangle_{\mathcal{H}}}{\hat{\sigma}}\right)\frac{\int_{\mathcal{T}} x_i(t)\theta_l(t)dt}{\hat{\sigma}} = 0, \quad l = 1, \ldots, L. \tag{S1.3}$$

Combining (S1.2) and (S1.3) with the fact that $\eta_i(t) = \int_{\mathcal{T}} x_i(u)K(u,t)du = \sum_{l=1}^{L}\{\int_{\mathcal{T}} x_i(u)\theta_l(u)du\}\theta_l(t) + \xi_i(t)$, we have

$$-\frac{1}{n}\sum_{i=1}^{n}\eta_i\psi\left(\frac{y_i - \alpha - \langle \eta_i, \beta\rangle_{\mathcal{H}}}{\hat{\sigma}}\right)\frac{1}{\hat{\sigma}} + 2\lambda P_1\beta = 0$$

for $\beta \in S_n$. Therefore, a minimizer $\hat{\beta}_{n\lambda}$ satisfies (6). □

S2 Proof of Theorem 1

Let us define the norm $|||f|||_{n\lambda} = \|\mathcal{G}_{n\lambda}^{-1} f\|_{\Gamma}$ for $f \in S_n$ and constant $B_n = \sup_{1\le j\le n} E|||\eta_j|||_{n\lambda}^2$. Note that $\mathcal{G}_{n\lambda}^{-1} f$ is a function of $\eta_1, \ldots, \eta_n$ and so $\mathcal{G}_{n\lambda}^{-1}\eta_1, \ldots, \mathcal{G}_{n\lambda}^{-1}\eta_n$ are dependent, but the $\mathcal{G}_{n\lambda}^{-1}\eta_j$ are identically distributed. This means that the random variables $|||\eta_j|||_{n\lambda}$ are not independent but identically distributed. Thus, $B_n = \sup_{1\le j\le n} E|||\eta_j|||^2_{n\lambda} = E|||\eta_1|||^2_{n\lambda}$. Also, note that, by Lemma 1 and (14),

$$\lim_{n\to\infty} n^{-1}B_n = \lim_{n\to\infty} B_n C_n = 0. \tag{S2.1}$$

By Mean-value theorem, we have

$$\psi\left(\frac{y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}}}{\hat\sigma}\right)\frac{1}{\hat\sigma} = \psi\left(\frac{y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}}}{\sigma}\right)\frac{1}{\sigma}$$
$$- (\hat\sigma - \sigma)\left\{\psi'\left(\frac{y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}}}{\sigma_n}\right)\left(\varepsilon_i + \frac{\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}}{\sigma_n}\right) + \psi\left(\frac{y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}}}{\sigma_n}\right)\right\}\frac{1}{\sigma_n^2}$$

with $\sigma_n = \inf_{1\le i\le n}\sigma_{in}$ for σ_{in} lying between $\hat\sigma$ and σ. Taking a second-order Taylor expansion of $\psi((y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}})/\sigma)$ around β_0, observe that

$$|||\Psi_{n\lambda}\beta - \sigma^2\Phi_{n\lambda}(\beta,\hat\sigma)/E\psi'|||_{n\lambda} \le T_1 + T_2 + T_3,$$

where

$$T_1 = |||n^{-1}\sum_{i=1}^n \eta_i\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}(\psi'(\varepsilon_i) - E\psi')/E\psi'|||_{n\lambda},$$

and

$$T_2 = |||(2n\sigma)^{-1}\sum_{i=1}^n \eta_i\psi''(\varepsilon_i + a_i)\langle\eta_i,\beta_0-\beta\rangle^2_{\mathcal{H}}/E\psi'|||_{n\lambda}$$

for a random variable a_i that is between 0 and $\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}/\sigma$. Also,

$$T_3 = |||\frac{(\hat\sigma-\sigma)}{\sigma_n^2}\frac{\sigma^2}{E\psi'}\frac{1}{n}\sum_{i=1}^n \eta_i\left\{\psi'\left(\frac{y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}}}{\sigma_n}\right)\left(\varepsilon_i + \frac{\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}}{\sigma_n}\right) + \psi\left(\frac{y_i - \langle\eta_i,\beta\rangle_{\mathcal{H}}}{\sigma_n}\right)\right\}|||_{n\lambda}.$$

We have that for any $\beta \in S_n$,

$$\begin{aligned}
ET_1^2 &= E|||n^{-1}\sum_{i=1}^n \eta_i\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}(\psi'(\varepsilon_i) - E\psi')/E\psi'|||^2_{n\lambda}\\
&= E\|n^{-1}\sum_{i=1}^n \mathcal{G}_{n\lambda}^{-1}\eta_i\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}(\psi'(\varepsilon_i) - E\psi')/E\psi'\|^2_{\Gamma}\\
&= n^{-2}\sum_{i=1}^n E\left[\|\mathcal{G}_{n\lambda}^{-1}\eta_i\|^2_{\Gamma}\langle\eta_i,\beta_0-\beta\rangle^2_{\mathcal{H}}(\psi'(\varepsilon_i) - E\psi')^2/(E\psi')^2\right]\\
&\quad + n^{-2}\sum\sum\nolimits_{i\ne j} E\left[\langle\mathcal{G}_{n\lambda}^{-1}\eta_i, \mathcal{G}_{n\lambda}^{-1}\eta_j\rangle_{\Gamma}\langle\eta_i,\beta_0-\beta\rangle_{\mathcal{H}}\langle\eta_j,\beta_0-\beta\rangle_{\mathcal{H}}\right]\\
&\qquad\qquad \times E[(\psi'(\varepsilon_i) - E\psi')]E[(\psi'(\varepsilon_j) - E\psi')]/(E\psi')^2\\
&= n^{-2}\sum_{i=1}^n E\left[|||\eta_i|||^2_{n\lambda}\langle\eta_i,\beta_0-\beta\rangle^2_{\mathcal{H}}\right]\operatorname{Var}(\psi')/(E\psi')^2\\
&\le n^{-1}CE|||\eta_1|||^2_{n\lambda}\left\{n^{-1}\sum_{i=1}^n E\langle\eta_i,\beta_0-\beta\rangle^2_{\mathcal{H}}\right\}\operatorname{Var}(\psi')/(E\psi')^2\\
&= n^{-1}CB_n\|\beta_0-\beta\|^2_{\Gamma}\operatorname{Var}(\psi')/(E\psi')^2
\end{aligned}$$

because the x_i and ε_i are independent, the ε_i are independent and identically distributed, the x_i are independent and identically distributed, and $E\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2 = E\left(\int_{\mathcal{T}} x_i(t)(\beta_0(t) - \beta(t))dt\right)^2 = \|\beta_0 - \beta\|_\Gamma^2$. Remark that the expectation for T_1 is taken with respect to the sample $x_1, \ldots, x_n$ and $\varepsilon_1, \ldots, \varepsilon_n$. Note that the inequality above is obtained by Cauchy-Schwarz inequality, (A7) and Lemma 2, where we have

$$\begin{aligned} E\left[|||\eta_i|||_{n\lambda}^2 \langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2\right] &\le \left\{E|||\eta_i|||_{n\lambda}^4 E\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^4\right\}^{1/2} \\ &\le CE|||\eta_i|||_{n\lambda}^2 E\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2. \end{aligned}$$

We have that for any $\beta \in S_n$,

$$\begin{aligned} T_2 &= |||(2n\sigma)^{-1} \sum_{i=1}^n \eta_i \psi''(\varepsilon_i + a_i)\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2 / E\psi'|||_{n\lambda} \\ &\le \frac{1}{2n\sigma} \frac{\sup|\psi''|}{|E\psi'|} ||| \sum_{i=1}^n \eta_i \langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2 |||_{n\lambda} \\ &\le \frac{1}{2n\sigma} \frac{\sup|\psi''|}{|E\psi'|} \sum_{i=1}^n |||\eta_i|||_{n\lambda} \langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2, \end{aligned}$$

and so

$$ET_2 \le \frac{1}{2\sigma} \frac{\sup|\psi''|}{|E\psi'|} C^{1/2} B_n^{1/2} \|\beta_0 - \beta\|_\Gamma^2$$

because, by Cauchy-Schwarz inequality and (A7),

$$E\left[|||\eta_i|||_{n\lambda} \langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^2\right] \le \left\{E|||\eta_i|||_{n\lambda}^2 E\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}^4\right\}^{1/2} \le C^{1/2}\left\{E|||\eta_i|||_{n\lambda}^2\right\}^{1/2} \|\beta_0 - \beta\|_\Gamma^2.$$

Also, for $\beta \in S_n$,

$$\begin{aligned} T_3 \le \frac{|\hat\sigma - \sigma|}{\sigma_n^2} \frac{\sigma^2}{|E\psi'|} \Bigg\{ &\sup|\psi'| \left(n^{-1} \sum_{i=1}^n |||\eta_i|||_{n\lambda} |\varepsilon_i| + \sigma_n^{-1} n^{-1} \sum_{i=1}^n |||\eta_i|||_{n\lambda} |\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}| \right) \\ &+ \sup|\psi| \left(n^{-1} \sum_{i=1}^n |||\eta_i|||_{n\lambda} \right) \Bigg\}. \end{aligned}$$

From the fact that $(\sum_{i=1}^n a_i)^2 \le n \sum_{i=1}^n a_i^2$, observe that $E\left(n^{-1}\sum_{i=1}^n |||\eta_i|||_{n\lambda}\right)^2 \le E|||\eta_1|||_{n\lambda}^2$,

$$E\left(n^{-1} \sum_{i=1}^n |||\eta_i|||_{n\lambda} |\varepsilon_i|\right)^2 \le n^{-1} \sum_{i=1}^n E|||\eta_i|||_{n\lambda}^2 E\varepsilon_i^2 = E|||\eta_1|||_{n\lambda}^2$$

and

$$\begin{aligned} E\left(n^{-1} \sum_{i=1}^n |||\eta_i|||_{n\lambda} |\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}|\right)^2 &\le n^{-1} \sum_{i=1}^n E\left[|||\eta_i|||_{n\lambda}^2 |\langle \eta_i, \beta_0 - \beta\rangle_{\mathcal{H}}|^2\right] \\ &\le CE|||\eta_1|||_{n\lambda}^2 \|\beta_0 - \beta\|_\Gamma^2. \end{aligned}$$

Thus, by Cauchy-Schwarz inequality and (A3), we have

$$ET_3 \le C_1(n^{-1}B_n)^{1/2} + C_2(n^{-1}B_n)^{1/2}\|\beta_0 - \beta\|_\Gamma$$

for some positive constants C_1 and C_2.

Now observe that for $A \geq 8/\delta$,

$$P[\|\tilde{\beta}_{n\lambda} - \beta_0\|_\Gamma < (1/2)(AC_n)^{1/2}] > 1 - \delta/2 \tag{S2.2}$$

by Markov's inequality. From Lemma 1 and (13), there is a constant C_3 such that $|B_n/(nC_n)| \leq C_3$ for sufficiently large n. Let $C_4 = C_1 C_3^{1/2}$. Define $F_n = \{\beta \in S_n : \|\beta - \beta_0\|_\Gamma^2 \leq (A^{1/2}C_4 + 1)^2 AC_n\}$. Then, for $\beta \in F_n$, we have

$$ET_1^2 \leq n^{-1}C(A^{1/2}C_4 + 1)^2 AB_nC_n \mathrm{Var}(\psi')/(E\psi')^2$$

and

$$ET_2 \leq \frac{1}{2\sigma}\frac{\sup|\psi''|}{|E\psi'|}C^{1/2}(A^{1/2}C_4 + 1)^2 AB_n^{1/2}C_n.$$

Also, for $\beta \in F_n$,

$$ET_3 \leq \left\{C_1A^{-1/2}(n^{-1}B_nC_n^{-1})^{1/2} + (A^{1/2}C_4 + 1)C_2(n^{-1}B_n)^{1/2}\right\}(AC_n)^{1/2}.$$

By (S2.1), for sufficiently large n, $ET_3 \leq C_4C_n^{1/2}$.

Letting $D_1 = \{8\delta^{-1}C(A^{1/2}C_4 + 1)^2\mathrm{Var}(\psi')/(E\psi')^2\}^{1/2}$ and $D_2 = 2\delta^{-1}\sigma^{-1}C^{1/2}(A^{1/2}C_4 + 1)^2(\sup|\psi'|/|E\psi'|)$, by Markov inequality, we have

$$P[T_1 \leq D_1(n^{-1}AB_nC_n)^{1/2}] > 1 - \delta/8 \tag{S2.3}$$

and

$$P[T_2 \leq D_2AB_n^{1/2}C_n] > 1 - \delta/4. \tag{S2.4}$$

Also,

$$P[T_3 \leq C_4AC_n^{1/2}] > 1 - \delta/8. \tag{S2.5}$$

Recall that $\tilde{\beta}_{n\lambda}$ is the solution of $\Psi_{n\lambda}\beta = 0$. From $\Psi_{n\lambda}\tilde{\beta}_{n\lambda} = 0$, we have $n^{-1}\sum_{i=1}^n \tilde{y}_i\eta_i = \mathcal{G}_{n\lambda}\tilde{\beta}_{n\lambda}$. So, for any $\beta \in S_n$, $\Psi_{n\lambda}\beta = -n^{-1}\sum_{i=1}^n \tilde{y}_i\eta_i + \mathcal{G}_{n\lambda}\beta = -\mathcal{G}_{n\lambda}\tilde{\beta}_{n\lambda} + \mathcal{G}_{n\lambda}\beta = \mathcal{G}_{n\lambda}(\beta - \tilde{\beta}_{n\lambda})$. Combining (S2.2), (S2.3), (S2.4) and (S2.5), we have an event of probability greater than $1 - \delta$ on which for all $\beta \in F_n$,

$$\begin{aligned}|||\sigma^2\Phi_{n\lambda}(\beta, \hat{\sigma})/E\psi' - \mathcal{G}_{n\lambda}(\beta - \beta_0)|||_{n\lambda} &\leq |||\sigma^2\Phi_{n\lambda}(\beta, \hat{\sigma})/E\psi' - \Psi_{n\lambda}\beta|||_{n\lambda} + |||\Psi_{n\lambda}\beta - \mathcal{G}_{n\lambda}(\beta - \beta_0)|||_{n\lambda} \\ &= |||\sigma^2\Phi_{n\lambda}(\beta, \hat{\sigma})/E\psi' - \Psi_{n\lambda}\beta|||_{n\lambda} + \|\tilde{\beta}_{n\lambda} - \beta_0\|_\Gamma \\ &\leq \{D_1(n^{-1}B_n)^{1/2} + D_2A^{1/2}B_n^{1/2}C_n^{1/2} + 1/2 + C_4A^{1/2}\}(AC_n)^{1/2}.\end{aligned}$$

By (S2.1), the quantity in braces will be less than or equal to $C_4A^{1/2} + 1$ for sufficiently large n. For such n, if $x \in F_n^*$ with $F_n^* = \{\beta - \beta_0 : \beta \in F_n\}$ and

$$U(x) = x - \sigma^2\mathcal{G}_{n\lambda}^{-1}\Phi_{n\lambda}(x + \beta_0, \hat{\sigma})/E\psi',$$

then $\|U(x)\|_\Gamma^2 \leq (A^{1/2}C_4 + 1)^2AC_n$, which means that the continuous function U maps the compact, convex set F_n^* into itself. By Brouwer's theorem, U has a fixed point x_0 in F_n^* such that $U(x_0) = x_0$, i.e., $\Phi_{n\lambda}(x_0 + \beta_0, \hat{\sigma}) = 0$. Taking $\hat{\beta}_{n\lambda} = x_0 + \beta_0$, $\Phi_{n\lambda}(\hat{\beta}_{n\lambda}, \hat{\sigma}) = 0$. Also, for such $\hat{\beta}_{n\lambda}$, $|||\Psi_{n\lambda}\hat{\beta}_{n\lambda} - \sigma^2\Phi_{n\lambda}(\hat{\beta}_{n\lambda}, \hat{\sigma})/E\psi'|||_{n\lambda} = |||\Psi_{n\lambda}\hat{\beta}_{n\lambda}|||_{n\lambda} = |||\mathcal{G}_{n\lambda}(\hat{\beta}_{n\lambda} - \tilde{\beta}_{n\lambda})|||_{n\lambda} = \|\hat{\beta}_{n\lambda} - \tilde{\beta}_{n\lambda}\|_\Gamma$. Thus, together with (S2.3), (S2.4) and (S2.5), we have

$$\begin{aligned}\|\hat{\beta}_{n\lambda} - \tilde{\beta}_{n\lambda}\|_\Gamma &= |||\Psi_{n\lambda}\hat{\beta}_{n\lambda} - \sigma^2\Phi_{n\lambda}(\hat{\beta}_{n\lambda}, \hat{\sigma})/E\psi'|||_{n\lambda} \\ &\leq \{D_1(n^{-1}B_n)^{1/2} + D_2A^{1/2}B_n^{1/2}C_n^{1/2} + C_4A^{1/2}\}(AC_n)^{1/2},\end{aligned}$$

where the inequality holds on an event of probability greater than $1 - \delta$. Applying (S2.1) completes the proof. $\square$

Lemma 1. *Under the assumptions (A4)-(A8), we have*

$$E|||\eta_j|||_{n\lambda}^2 = O(n^{1/(2r+2s+1)})$$

for $1 \leq j \leq n$*, where the norm* $|||\cdot|||_{n\lambda}$ *is defined as* $|||f|||_{n\lambda} = \|\mathcal{G}_{n\lambda}^{-1} f\|_\Gamma$ *for* $f \in S_n$*.*

Proof. Recall that $\tilde{\beta}_{n\lambda} = \mathcal{G}_{n\lambda}^{-1}\left(n^{-1}\sum_{i=1}^n \tilde{y}_i\eta_i\right)$ from the fact that $\tilde{\beta}_{n\lambda}$ is the solution to $\Psi_{n\lambda}\beta = 0$. Then, $\mathcal{G}_{n\lambda}^{-1}\eta_j$ is obtained by taking $\tilde{y}_i = n\delta_{ij}$ so that $\tilde{\beta}_{n\lambda j} := \mathcal{G}_{n\lambda}^{-1}\eta_j$ is the minimizer over β of

$$\frac{1}{n}\sum_{i=1}^n (n\delta_{ij} - \langle \eta_i, \beta\rangle_{\mathcal{H}})^2 + 2\frac{\lambda\sigma^2}{E\psi'}\|P_1\beta\|_{\mathcal{H}}^2,$$

where δ_{ij} is the Kronecker's delta. This enables us to use the techniques in Yuan and Cai (2010) for getting the desired rate for $|||\eta_j|||_{n\lambda}^2$. Note that, if $\|P_1\beta\|_{\mathcal{H}}^2 = \int_{\mathcal{T}}[\beta^{(m)}(t)]^2dt$, then $\tilde{\beta}_{n\lambda j}$ is the least squares smoothing spline estimator for functional linear regression with impulse response.

Now let us bring some results and definitions from Yuan and Cai (2010). Let $\omega_k = \nu_k^{-1/2}R^{1/2}\zeta_k$, where $\nu_k = (1+\gamma_k^{-1})^{-1}$ and ζ_k are the eigenvalues and the corresponding eigenfunctions of the operator $R^{1/2}\Gamma R^{1/2}$. Then, it was shown in Yuan and Cai (2010) that for any $f \in \mathcal{H}$, $f = \sum_{k=1}^\infty f_k\omega_k$ with $f_k = \nu_k\langle f, \omega_k\rangle_R$, $\|f\|_\Gamma^2 = \sum_{k=1}^\infty f_k^2$, and $\|f\|_R^2 = \sum_{k=1}^\infty (1+\gamma_k^{-1})f_k^2$. For $0 \leq a \leq 1$, define the norm $\|\cdot\|_a$ by

$$\|f\|_a^2 = \sum_{k=1}^\infty (1+\gamma_k^{-a})f_k^2. \tag{S2.6}$$

Note that $\|f\|_0^2 = 2\|f\|_\Gamma^2$ and $\|f\|_1^2 = \|f\|_R^2$.

For $f \in \mathcal{H}$, define the operator G_λ by

$$G_\lambda f(\cdot) = \int_{\mathcal{T}}\int_{\mathcal{T}} f(s)\Gamma(s,t)K(\cdot,t)dsdt + 2\frac{\lambda\sigma^2}{E\psi'}P_1 f(\cdot).$$

From Lemma 3, we observe that the operator G_λ^{-1} given by

$$G_\lambda^{-1} f(\cdot) = \sum_{k=1}^\infty \left(1 + 2\frac{\lambda\sigma^2}{E\psi'}\gamma_k^{-1}\right)^{-1}\langle f, \omega_k\rangle_{\mathcal{H}}\omega_k(\cdot)$$

is the inverse operator of G_λ. Let $\tilde{\beta}_{n\lambda j}^* = G_\lambda^{-1}\eta_j$. Then, we have

$$E|||\eta_j|||_{n\lambda}^2 = E\|\tilde{\beta}_{n\lambda j}\|_\Gamma^2 \leq 2E\|\tilde{\beta}_{n\lambda j} - \tilde{\beta}_{n\lambda j}^*\|_\Gamma^2 + 2E\|\tilde{\beta}_{n\lambda j}^*\|_\Gamma^2.$$

We investigate the upper bounds for both terms in the right-hand side of the inequality above. Let $\lambda_0 = 2\lambda\sigma^2/E\psi'$. Since $\tilde{\beta}_{n\lambda j}^* = G_\lambda^{-1}\eta_j = \sum_{k=1}^\infty (1+\lambda_0\gamma_k^{-1})^{-1}\langle \eta_j, \omega_k\rangle_{\mathcal{H}}\omega_k$, we have

$$\|\tilde{\beta}_{n\lambda j}^*\|_a^2 = \sum_{k=1}^\infty (1+\gamma_k^{-a})\nu_k^2\langle \tilde{\beta}_{n\lambda j}^*, \omega_k\rangle_R^2 = \sum_{k=1}^\infty (1+\gamma_k^{-a})(1+\lambda_0\gamma_k^{-1})^{-2}\left(\int_{\mathcal{T}} x_j(t)\omega_k(t)dt\right)^2$$

using the fact that $\langle \eta_j, \omega_k\rangle_{\mathcal{H}} = \int_{\mathcal{T}} x_j(t)\omega_k(t)dt$ and $\langle \omega_k, \omega_l\rangle_R = \nu_k^{-1}\delta_{kl}$. Thus,

$$\begin{aligned}
E\|\tilde{\beta}^*_{n\lambda j}\|_a^2 &= \sum_{k=1}^{\infty}(1+\gamma_k^{-a})(1+\lambda_0\gamma_k^{-1})^{-2} \underbrace{E\left(\int_{\mathcal{T}} x_j(t)\omega_k(t)dt\right)^2}_{=\|\omega_k\|_\Gamma^2=1} \\
&= \sum_{k=1}^{\infty}(1+\gamma_k^{-a})(1+\lambda_0\gamma_k^{-1})^{-2} \\
&\le C_5\sum_{k=1}^{\infty}(1+k^{a(2r+2s)})(1+\lambda_0 k^{2r+2s})^{-2} \\
&\le C_6\lambda_0^{-(a+1/(2r+2s))}\int_{\lambda_0^{a+1/(2r+2s)}}^{\infty}\left(1+y^{(2r+2s)/(a(2r+2s)+1)}\right)^{-2}dy \\
&= O(\lambda^{-(a+1/(2r+2s))})
\end{aligned}$$

for some positive constants C_5 and C_6, so $E\|\tilde{\beta}^*_{n\lambda j}\|_0^2 = O(\lambda^{-1/(2r+2s)})$ by taking $a=0$, equivalently, we have $E\|\tilde{\beta}^*_{n\lambda j}\|_\Gamma^2 = O(\lambda^{-1/(2r+2s)})$. Next observe that $G_\lambda\tilde{\beta}^*_{n\lambda j} = \eta_j = \mathcal{G}_{n\lambda}\tilde{\beta}_{n\lambda j}$, so

$$\begin{aligned}
\tilde{\beta}_{n\lambda j} - \tilde{\beta}^*_{n\lambda j} &= G_\lambda^{-1}G_\lambda(\tilde{\beta}_{n\lambda j} - \tilde{\beta}^*_{n\lambda j}) \\
&= G_\lambda^{-1}(G_\lambda\tilde{\beta}_{n\lambda j} - \mathcal{G}_{n\lambda}\tilde{\beta}_{n\lambda j}) \\
&= \sum_{k=1}^{\infty}(1+\lambda_0\gamma_k^{-1})^{-1}\left[\langle G_\lambda\tilde{\beta}_{n\lambda j}, \omega_k\rangle_{\mathcal{H}} - \langle \mathcal{G}_{n\lambda}\tilde{\beta}_{n\lambda j}, \omega_k\rangle_{\mathcal{H}}\right]\omega_k \\
&= \sum_{k=1}^{\infty}(1+\lambda_0\gamma_k^{-1})^{-1}\left[\int_{\mathcal{T}}\int_{\mathcal{T}}\tilde{\beta}_{n\lambda j}(s)\Gamma(s,t)\omega_k(t)dsdt\right. \\
&\qquad\left. - \int_{\mathcal{T}}\int_{\mathcal{T}}\tilde{\beta}_{n\lambda j}(s)\left(\frac{1}{n}\sum_{i=1}^{n}x_i(s)x_i(t)\right)\omega_k(t)dsdt\right]\omega_k.
\end{aligned}$$

Now write $\tilde{\beta}_{n\lambda j} = \sum_{k=1}^{\infty}\tilde{b}_{jk}\omega_k$. Then,

$$\begin{aligned}
\|\tilde{\beta}_{n\lambda j} - \tilde{\beta}^*_{n\lambda j}\|_a^2 &= \sum_{k=1}^{\infty}(1+\gamma_k^{-a})(1+\lambda_0\gamma_k^{-1})^{-2}\left[\sum_{l=1}^{\infty}\tilde{b}_{jl}\int_{\mathcal{T}}\int_{\mathcal{T}}\omega_l(s)\left(\frac{1}{n}\sum_{i=1}^{n}x_i(s)x_i(t) - \Gamma(s,t)\right)\omega_k(t)dsdt\right]^2 \\
&\le \sum_{k=1}^{\infty}(1+\gamma_k^{-a})(1+\lambda_0\gamma_k^{-1})^{-2}\underbrace{\left(\sum_{l=1}^{\infty}(1+\gamma_l^{-c})\tilde{b}_{jl}^2\right)}_{=\|\tilde{\beta}_{n\lambda j}\|_c^2} \\
&\quad\times\left(\sum_{l=1}^{\infty}(1+\gamma_l^{-c})^{-1}\left[\int_{\mathcal{T}}\int_{\mathcal{T}}\omega_l(s)\left(\frac{1}{n}\sum_{i=1}^{n}x_i(s)x_i(t) - \Gamma(s,t)\right)\omega_k(t)dsdt\right]^2\right).
\end{aligned}$$

Note that by the Cauchy-Schwarz inequality and (A7),

$$
\begin{aligned}
&E\left(\sum_{l=1}^{\infty}(1+\gamma_l^{-c})^{-1}\left[\int_{\mathcal{T}}\int_{\mathcal{T}}\omega_l(s)\left(\frac{1}{n}\sum_{i=1}^{n}x_i(s)x_i(t)-\Gamma(s,t)\right)\omega_k(t)dsdt\right]^2\right)\\
&=\frac{1}{n}\sum_{l=1}^{\infty}(1+\gamma_l^{-c})^{-1}\mathrm{Var}\left(\int_{\mathcal{T}}X(s)\omega_l(s)ds\int_{\mathcal{T}}X(t)\omega_k(t)dt\right)\\
&\leq\frac{1}{n}\sum_{l=1}^{\infty}(1+\gamma_l^{-c})^{-1}E\left[\left(\int_{\mathcal{T}}X(s)\omega_l(s)ds\int_{\mathcal{T}}X(t)\omega_k(t)dt\right)^2\right]\\
&\leq\frac{C}{n}\sum_{l=1}^{\infty}(1+\gamma_l^{-c})^{-1}\leq\frac{C_5}{n}\sum_{k=1}^{\infty}(1+k^{c(2r+2s)})^{-1}=O(n^{-1})
\end{aligned}
$$

for $c>1/(2r+2s)$. Thus,

$$\|\tilde{\beta}_{n\lambda j}-\tilde{\beta}^*_{n\lambda j}\|_a^2=O_p(n^{-1}\lambda^{-(a+1/(2r+2s))}\|\tilde{\beta}_{n\lambda j}\|_c^2).$$

Taking $a=c$ yields

$$\|\tilde{\beta}_{n\lambda j}-\tilde{\beta}^*_{n\lambda j}\|_c^2=O_p(n^{-1}\lambda^{-(c+1/(2r+2s))}\|\tilde{\beta}_{n\lambda j}\|_c^2).$$

If $n^{-1}\lambda^{-(c+1/(2r+2s))}\to 0$ as $n\to\infty$, then

$$\|\tilde{\beta}^*_{n\lambda j}\|_c\geq\|\tilde{\beta}_{n\lambda j}\|_c-\|\tilde{\beta}_{n\lambda j}-\tilde{\beta}^*_{n\lambda j}\|_c=(1-o_p(1))\|\tilde{\beta}_{n\lambda j}\|_c,$$

so $\|\tilde{\beta}_{n\lambda j}\|_c^2=O_p(\|\tilde{\beta}^*_{n\lambda j}\|_c^2)$. Since $\|\tilde{\beta}^*_{n\lambda j}\|_c^2=O_p(\lambda^{-(c+1/(2r+2s))})$ and $\|\cdot\|_\Gamma^2=\frac{1}{2}\|\cdot\|_0^2$,

$$\|\tilde{\beta}_{n\lambda j}-\tilde{\beta}^*_{n\lambda j}\|_\Gamma^2=O_p(n^{-1}\lambda^{-1/(2r+2s)}\|\tilde{\beta}_{n\lambda j}\|_c^2)=O_p(\lambda^{-1/(2r+2s)}n^{-1}\lambda^{-(c+1/(2r+2s))})=o_p(\lambda^{-1/(2r+2s)}).$$

Therefore, $E|||\eta_j|||_{n\lambda}^2=E\|\tilde{\beta}_{n\lambda j}\|_\Gamma^2=O(\lambda^{-1/(2r+2s)})$ for all j, so the proof is complete by (A6). □

Lemma 2. *Under the assumption (A7), we have*

$$E|||\eta_j|||_{n\lambda}^4\leq C\left\{E|||\eta_j|||_{n\lambda}^2\right\}^2$$

for $1\leq j\leq n$.

Proof. Recall that $\eta_j(t)=\int_{\mathcal{T}}x_j(u)K(u,t)du$. Observe that

$$|||\eta_j|||_{n\lambda}^2=\|\mathcal{G}_{n\lambda}^{-1}\eta_j\|_\Gamma^2=\int_{\mathcal{T}}\int_{\mathcal{T}}\mathcal{G}_{n\lambda}^{-1}\eta_j(s)\Gamma(s,t)\mathcal{G}_{n\lambda}^{-1}\eta_j(t)dsdt=\sum_{k=1}^{\infty}\pi_k\left(\int_{\mathcal{T}}\mathcal{G}_{n\lambda}^{-1}\eta_j(t)\phi_k(t)dt\right)^2,$$

where the (π_k,ϕ_k) are the pairs of the eigenvalue and eigenfunction of the covariance operator Γ. Letting $g_{n\lambda k}(u)=$

$\int_{\mathcal{T}} \mathcal{G}_{n\lambda}^{-1} K(u,t)\phi_k(t)dt$, $|||\eta_j|||_{n\lambda}^2 = \sum_{k=1}^{\infty} \pi_k \left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^2$ and so we have

$$
\begin{aligned}
E|||\eta_j|||_{n\lambda}^4 &= E\left[\sum_{k=1}^{\infty} \pi_k \left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^2\right]^2 \\
&= \sum_k \pi_k^2 E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^4 \\
&\quad + \sum\sum\nolimits_{k\neq l} \pi_k\pi_l E\left[\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^2 \left(\int_{\mathcal{T}} x_j(u) g_{n\lambda l}(u)du\right)^2\right] \\
&\leq \sum_k \pi_k^2 E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^4 \\
&\quad + \sum\sum\nolimits_{k\neq l} \pi_k\pi_l \left\{E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^4 E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda l}(u)du\right)^4\right\}^{1/2} \\
&\leq C\sum_k \pi_k^2 \left\{E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^2\right\}^2 \\
&\quad + C\sum\sum\nolimits_{k\neq l} \pi_k\pi_l E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^2 E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda l}(u)du\right)^2 \\
&= C\left\{\sum_k \pi_k E\left(\int_{\mathcal{T}} x_j(u) g_{n\lambda k}(u)du\right)^2\right\}^2
\end{aligned}
$$

by Cauchy-Schwarz inequality and the assumption (A7). The proof is complete. □

Lemma 3. *For $f \in \mathcal{H}$, define*

$$G_\lambda f(\cdot) = \int_{\mathcal{T}}\int_{\mathcal{T}} f(s)\Gamma(s,t)K(\cdot,t)dsdt + \lambda P_1 f(\cdot)$$

and

$$G_\lambda^{-1} f = \sum_{k=1}^{\infty} (1+\lambda\gamma_k^{-1})^{-1} \langle f, \omega_k\rangle_{\mathcal{H}} \omega_k.$$

Then, $G_\lambda^{-1} G_\lambda f = f = G_\lambda G_\lambda^{-1} f$.

Proof. We will first show that $G_\lambda^{-1} G_\lambda f = f$ for any $f \in \mathcal{H}$. For this, observe that

$$
\begin{aligned}
G_\lambda^{-1} G_\lambda f &= \sum_{k=1}^{\infty} (1+\lambda\gamma_k^{-1})^{-1} \langle G_\lambda f, \omega_k\rangle_{\mathcal{H}} \omega_k \\
&= \sum_{k=1}^{\infty} (1+\lambda\gamma_k^{-1})^{-1} \left[\int_{\mathcal{T}}\int_{\mathcal{T}} f(s)\Gamma(s,t)\omega_k(t)dsdt + \lambda\langle P_1 f, P_1\omega_k\rangle_{\mathcal{H}}\right]\omega_k
\end{aligned}
$$

because $\langle P_1 f, \omega_k\rangle_{\mathcal{H}} = \langle P_1 f, P_1\omega_k\rangle_{\mathcal{H}}$. Note that $\omega_k \in \mathcal{H}$ since we observe $\|\omega_k\|_{\mathcal{H}}^2 \leq \|\omega_k\|_R^2 < \infty$ from the definition of the norm $\|\cdot\|_R^2$. For any $f \in \mathcal{H}$, $f = \sum_{k=1}^{\infty} f_k\omega_k$ with $f_k = \nu_k\langle f, \omega_k\rangle_R$ and so

$$G_\lambda^{-1} G_\lambda f = \sum_{k=1}^{\infty} (1+\lambda\gamma_k^{-1})^{-1} \left[\sum_{j=1}^{\infty} f_j \left\{\int_{\mathcal{T}}\int_{\mathcal{T}} \omega_j(s)\Gamma(s,t)\omega_k(t)dsdt + \lambda\langle P_1\omega_j, P_1\omega_k\rangle_{\mathcal{H}}\right\}\right]\omega_k.$$

Now, from the definition of $\|\cdot\|_R^2$, we can observe that

$$\langle f,g\rangle_R = \int_{\mathcal{T}}\int_{\mathcal{T}} f(s)\Gamma(s,t)g(t)dsdt + \langle P_1 f, P_1 g\rangle_{\mathcal{H}},$$

so that we have

$$\begin{aligned}\int_{\mathcal{T}}\int_{\mathcal{T}} \omega_j(s)\Gamma(s,t)\omega_k(t)dsdt + \lambda\langle P_1\omega_j, P_1\omega_k\rangle_{\mathcal{H}} &= (1-\lambda)\int_{\mathcal{T}}\int_{\mathcal{T}} \omega_j(s)\Gamma(s,t)\omega_k(t)dsdt + \lambda\langle \omega_j,\omega_k\rangle_R \\ &= \{(1-\lambda)+\lambda\nu_k^{-1}\}\delta_{jk} \\ &= (1+\lambda\gamma_k^{-1})\delta_{jk}\end{aligned}$$

because $\langle \omega_j,\omega_k\rangle_R = \nu_k^{-1}\delta_{jk}$, $\int_{\mathcal{T}}\int_{\mathcal{T}} \omega_j(s)\Gamma(s,t)\omega_k(t)dsdt = \delta_{jk}$ and $\nu_k = (1+\gamma_k^{-1})^{-1}$. Thus,

$$G_\lambda^{-1}G_\lambda f = \sum_{k=1}^{\infty}(1+\lambda\gamma_k^{-1})^{-1}\left[\sum_{j=1}^{\infty} f_j(1+\lambda\gamma_k^{-1})\delta_{jk}\right]\omega_k = \sum_{k=1}^{\infty} f_k\omega_k = f.$$

To see that $G_\lambda G_\lambda^{-1} f = f$, observe that

$$\begin{aligned}G_\lambda G_\lambda^{-1}f(\cdot) &= \int_{\mathcal{T}}\int_{\mathcal{T}} G_\lambda^{-1}f(s)\Gamma(s,t)K(\cdot,t)dsdt + \lambda P_1 G_\lambda^{-1}f(\cdot) \\ &= \sum_{k=1}^{\infty}(1+\lambda\gamma_k^{-1})^{-1}\langle f,\omega_k\rangle_{\mathcal{H}}\left[\int_{\mathcal{T}}\int_{\mathcal{T}} \omega_k(s)\Gamma(s,t)K(\cdot,t)dsdt + \lambda P_1\omega_k(\cdot)\right].\end{aligned}$$

Now we have

$$\begin{aligned}\langle G_\lambda G_\lambda^{-1}f,\omega_l\rangle_{\mathcal{H}} &= \sum_{k=1}^{\infty}(1+\lambda\gamma_k^{-1})^{-1}\langle f,\omega_k\rangle_{\mathcal{H}}\left[\int_{\mathcal{T}}\int_{\mathcal{T}} \omega_k(s)\Gamma(s,t)\omega_l(t)dsdt + \lambda\langle P_1\omega_k, P_1\omega_l\rangle_{\mathcal{H}}\right] \\ &= \sum_{k=1}^{\infty}(1+\lambda\gamma_k^{-1})^{-1}\langle f,\omega_k\rangle_{\mathcal{H}}(1+\lambda\gamma_k^{-1})\delta_{kl} \\ &= \langle f,\omega_l\rangle_{\mathcal{H}},\end{aligned}$$

which implies that $G_\lambda G_\lambda^{-1} f = f$ for $f\in\mathcal{H}$ since both $G_\lambda G_\lambda^{-1} f - f$ and ω_l are in $\mathcal{H}$. □

S3 Proof of Theorem 3

Now define the norm $|||f|||_{n\lambda} = \|\mathcal{G}_{n\lambda}^{-1}f\|$ for $f\in S_n$ and constants $B_n = \sup_{1\le j\le n} E|||\eta_j|||_{n\lambda}^2$ and $C_n = E\|\tilde{\beta}_{n\lambda}-\beta_0\|^2$. Since $\|\beta_0-\beta\|_\Gamma^2 = E\left(\int_{\mathcal{T}} X(t)(\beta_0(t)-\beta(t))dt\right)^2 \le E\|X\|^2\|\beta_0-\beta\|^2$, the proof of Theorem 1 goes exactly the same with respect to $\|\cdot\|$ and so Theorem 1 remains under the $\mathcal{L}_2$-norm $\|\cdot\|$. Thus, it is sufficient to show the condition on B_n and C_n in (S2.1) and Lemma 2 under the norm $\|\cdot\|$. From the fact that the norm in (S2.6) with $a = s/(r+s)$ is equivalent to $\|\cdot\|$, we can show that

$$B_n = \sup_{1\le j\le n} E|||\eta_j|||_{n\lambda}^2 = O(n^{(2s+1)/(2r+2s+1)})$$

in analogous to Lemma 1. Also, from Yuan and Cai (2010), we have that under the assumption (A4)-(A8),

$$C_n = E\|\tilde{\beta}_{n\lambda}-\beta_0\|^2 = O(n^{-2r/(2r+2s+1)}).$$

Consequently, the condition (S2.1) is met when $2r > 2s+1$.

Next we show that Lemma 2 holds with $|||f|||_{n\lambda} = \|\mathcal{G}_{n\lambda}^{-1} f\|$. For this, observe that $\mathcal{G}_{n\lambda}^{-1}\eta_j(t) = \int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}K(u,t)du = \sum_{k=1}^{\infty} \varsigma_k \int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du\vartheta_k(t)$, where the $(\varsigma_k, \vartheta_k)$ are the pairs of the eigenvalue and eigenfunction of the reproducing kernel K. Then,

$$\begin{aligned}
|||\eta_j|||_{n\lambda}^2 = \|\mathcal{G}_{n\lambda}^{-1}\eta_j\|^2 &= \int_{\mathcal{T}} \left(\sum_{k=1}^{\infty} \varsigma_k \int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du\vartheta_k(t) \right)^2 dt \\
&= \sum_k \varsigma_k^2 \left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \right)^2 \int_{\mathcal{T}} \vartheta_k^2(t)dt \\
&\quad + \sum\sum\nolimits_{k\neq l} \varsigma_k\varsigma_l \int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_l(u)du \int_{\mathcal{T}} \vartheta_k(t)\vartheta_l(t)dt \\
&= \sum_k \varsigma_k^2 \left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \right)^2.
\end{aligned}$$

In a similar way to the proof of Lemma 2, we have

$$\begin{aligned}
E|||\eta_j|||_{n\lambda}^4 &= E\left(\sum_k \varsigma_k^2 \left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \right)^2 \right)^2 \\
&= \sum_k \varsigma_k^4 E\left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \right)^4 \\
&\quad + \sum\sum\nolimits_{k\neq l} \varsigma_k^2\varsigma_l^2 E\left[\left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \right)^2 \left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_l(u)du \right)^2 \right] \\
&\leq C\left\{ \sum_k \varsigma_k^2 E\left(\int_{\mathcal{T}} x_j(u)\mathcal{G}_{n\lambda}^{-1}\vartheta_k(u)du \right)^2 \right\}^2 = C\left\{ E|||\eta_j|||_{n\lambda}^2 \right\}^2
\end{aligned}$$

by Cauchy-Schwarz inequality and (A7). Since $\|\hat{\beta}_{n\lambda} - \tilde{\beta}_{n\lambda}\|^2 = O_p(C_n)$ and $\|\hat{\beta}_{n\lambda} - \beta_0\|^2 \leq 2\|\tilde{\beta}_{n\lambda} - \beta_0\|^2 + 2\|\hat{\beta}_{n\lambda} - \tilde{\beta}_{n\lambda}\|^2$, the convergence rate for the estimation error of our M-type smoothing spline estimator $\hat{\beta}_{n\lambda}$ is the same for the least squares smoothing spline estimator $\tilde{\beta}_{n\lambda}$. $\square$

S4 Proof of Theorem 4

Recall that our M-estimate of scale is given in the form of $\sum_{i=1}^n \Psi(r_i^0, \sigma) = 0$ with $\Psi(r,\sigma) = \rho_0(|r|/\sigma) - \delta$ for some $\delta \in (0,1)$. Although $r_i^0 = y_i - \int_{\mathcal{T}} x_i(t)\hat{\beta}(t)dt$ with an estimator $\hat{\beta}$ obtained using all the observations in the sample, we take $r_i^0 = y_i - \int_{\mathcal{T}} x_i(t)\hat{\beta}_{-i}(t)dt$ with a leave-one-out estimator $\hat{\beta}_{-i}$, in sense that $\hat{\beta}_{-i} \approx \hat{\beta}$ for sufficiently large n.

Let $\breve{\sigma}$ be the solution of $\sum_{i=1}^n \Psi(\epsilon_i, \sigma) = 0$, where $\epsilon_i = y_i - \int \beta_0 x_i$. Then, we can show that

$$n^{-1}\sum_{i=1}^n \Psi(r_i^0, \sigma) = n^{-1}\sum_{i=1}^n \Psi(\epsilon_i, \sigma) + o_p(n^{-1/2}) \tag{S4.1}$$

implies $\hat{\sigma} - \breve{\sigma} = o_p(n^{-1/2})$. This is because

$$\begin{aligned}\left|n^{-1/2}\sum_{i=1}^{n}\Psi(r_i^0,\hat{\sigma}) - n^{-1/2}\sum_{i=1}^{n}\Psi(\epsilon_i,\hat{\sigma})\right| &= \left|n^{-1/2}\sum_{i=1}^{n}\Psi(\epsilon_i,\hat{\sigma})\right| \\ &= \left|n^{-1/2}\sum_{i=1}^{n}\Psi(\epsilon_i,\breve{\sigma}) + n^{-1/2}\sum_{i=1}^{n}\dot{\Psi}(\epsilon_i,\sigma_0)(\hat{\sigma}-\breve{\sigma})\right| \\ &= \left|n^{-1}\sum_{i=1}^{n}\dot{\Psi}(\epsilon_i,\sigma_0)\sqrt{n}(\hat{\sigma}-\breve{\sigma})\right|,\end{aligned}$$

where σ_0 is some value between $\hat{\sigma}$ and $\breve{\sigma}$ and $\dot{\Psi} = \partial\Psi/\partial\sigma$. Note that the second equation above yields by the first-order Talyor expansion of Ψ at σ. Since $n^{-1}\sum_{i=1}^{n}\dot{\Psi}(\epsilon_i,\sigma_0) \to E[\dot{\Psi}(\epsilon_1,\sigma_0)] < \infty$ in probability, (S4.1) implies that $\sqrt{n}(\hat{\sigma}-\breve{\sigma}) = o_p(1)$. Also, it can be shown that $\sqrt{n}(\breve{\sigma}-\sigma) = O_p(1)$ by asymptotic normality of M-estimates for scale in the form of $\sum_{i=1}^{n}\Psi(\epsilon_i,\sigma) = 0$. Thus, it suffices to show (S4.1) to verify (A3).

We now show (S4.1). A second-order Talyor expansion of ρ_0 at β yields

$$n^{-1}\sum_{i=1}^{n}\rho_0\left(\frac{r_i^0}{\sigma}\right) = n^{-1}\sum_{i=1}^{n}\rho_0\left(\frac{\epsilon_i}{\sigma}\right) + \sigma^{-1}V_1 + \sigma^{-2}V_2,$$

where $V_1 = n^{-1}\sum_{i=1}^{n}\rho_0'(\epsilon_i/\sigma)\int_{\mathcal{T}}(\hat{\beta}_{-i}-\beta_0)(t)x_i(t)dt$ and $V_2 = n^{-1}\sum_{i=1}^{n}\rho_0''\left(\frac{\epsilon_i+u_i}{\sigma}\right)\left(\int_{\mathcal{T}}(\hat{\beta}_{-i}-\beta_0)(t)x_i(t)dt\right)^2$ for some u_i in between 0 and $\int_{\mathcal{T}}(\beta_0-\hat{\beta}_{-i})(t)x_i(t)dt$. For V_2, observe that

$$E|V_2| \le M_1 n^{-1}\sum_{i=1}^{n}E\left[\left(\int_{\mathcal{T}}(\hat{\beta}_{-i}-\beta_0)(t)x_i(t)dt\right)^2\right] = M_1 E\|\hat{\beta}_{-i}-\beta_0\|_\Gamma^2$$

with $M_1 = \sup_t|\rho_0''(t)| < \infty$. Now let $D_i = \rho_0'(\epsilon_i/\sigma)\int_{\mathcal{T}}(\hat{\beta}_{-i}-\beta_0)(t)x_i(t)dt$. For V_1, we have

$$nEV_1^2 = n^{-1}\sum_{i=1}^{n}E[D_i^2] + n^{-1}\sum\sum\nolimits_{i\neq j}E[D_iD_j].$$

Observe that

$$n^{-1}\sum_{i=1}^{n}E[D_i^2] = n^{-1}\sum_{i=1}^{n}E\left[\left(\rho_0'(\epsilon_i/\sigma)\int(\hat{\beta}_{-i}-\beta_0)(t)x_i(t)dt\right)^2\right] = M_2 E\|\hat{\beta}_{-i}-\beta_0\|_\Gamma^2.$$

since $E[\rho_0'(\epsilon_i/\sigma)] = 0$ and $M_2 = \mathrm{Var}(\rho_0'(\epsilon_i/\sigma)) < \infty$. Let $D_{ij} = \rho_0'(\epsilon_i/\sigma)\int_{\mathcal{T}}(\hat{\beta}_{ij}-\beta_0)(t)x_i(t)dt$. Then, observe that

$$E[D_iD_{ji}] = E[D_{ji}E[D_i|(x_k,y_k), 1\le k\le n, k\neq i]] = 0$$

from the fact that $E[D_i|(x_k,y_k), 1\le k\le n, k\neq i] = \int(\hat{\beta}_{-i}-\beta_0)(t)E[\epsilon_i x_i(t)]dt = 0$. Similarly, $E[D_{ij}D_j] = E[D_{ij}D_{ji}] = 0$. Thus, we have $E[D_iD_j] = E[(D_i - D_{ij})(D_j - D_{ji})]$. Note that $\hat{\beta}_{ii} = \hat{\beta}_{-i}$, so $D_i = D_{ii}$. Since $|E[(D_i-D_{ij})(D_j-D_{ji})]| \le E[(D_i-D_{ij})^2]$ by Cauchy-Schwarz inequality and

$$E[(D_i-D_{ij})^2] = E\left[\left(\rho_0'(\epsilon_i/\sigma)\int_{\mathcal{T}}(\hat{\beta}_{-i}-\hat{\beta}_{ij})(t)x_i(t)dt\right)^2\right] = M_2 E\|\hat{\beta}_{-i}-\hat{\beta}_{ij}\|_\Gamma^2,$$

we have

$$
\begin{aligned}
nEV_1^2 &\le n^{-1}\sum_{i=1}^{n} E[D_i^2] + n^{-1}\sum\sum\nolimits_{i\neq j} E[(D_i - D_{ij})^2] \\
&= M_2 E\|\hat{\beta}_{-i} - \beta_0\|_\Gamma^2 + M_2 n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} E\|\hat{\beta}_{-i} - \hat{\beta}_{ij}\|_\Gamma^2.
\end{aligned}
$$

Under the assumptions (15) and (16), we have

$$
n^{-1}\sum_{i=1}^{n}\rho_0\left(\frac{r_i^0}{\sigma}\right) = n^{-1}\sum_{i=1}^{n}\rho_0\left(\frac{\epsilon_i}{\sigma}\right) + o_p(n^{-1/2}),
$$

which proves (S4.1). □