Partially Linear Additive Functional Regression

Xiaohui Liu[1], Wenqi Lu[2,3], Heng Lian[2], Yuzi Liu[1] and Zhongyi Zhu[3]

Jiangxi University of Finance and Economics[1], *City University of Hong Kong*[2],

Fudan University[3]

Supplementary Material

S1. Proof for Theorem 1

We define $\widehat{H}(\boldsymbol{\theta}) = K^{-1/2}\widehat{F}(\boldsymbol{\theta})$, $H_0 = K^{-1/2}F_0$. We have $\|F\|_{\mathcal{H}} = \|H\|$ when $F \in \mathcal{H}$ and $H = K^{-1/2}F$. Furthermore, the prediction risk can be written as $\|T^{1/2}(\widehat{H} - H_0)\|$.

Since the objective function (2.1) can be written as

$$\sum_{i=1}^{n} \left(Y_i - \langle H, U_i\rangle - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2 + n\lambda\|H\|^2,$$

we have, for given $\boldsymbol{\theta}$,

$$\widehat{H}(\boldsymbol{\theta}) = (T_n + \lambda I)^{-1}\frac{\sum_{i=1}^{n} U_i(Y_i - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta})}{n}, \tag{S1.1}$$

where $T_n = \sum_i U_i \otimes U_i/n$ is a simple moment estimator of $T = E[U \otimes U]$.

Plugging (S1.1) into (2.2), we get that $\widehat{\boldsymbol{\theta}}$ is the minimizer of

$$\begin{aligned}
&\sum_{i=1}^{n}\left(Y_i - \langle U_i, (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j(Y_j - \mathbf{Z}_j^{\mathrm{T}}\boldsymbol{\theta})}{n}\rangle - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2 \\
= &\sum_{i=1}^{n}\left(\epsilon_i + \langle U_i, H_0\rangle - \langle U_i, (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j(\epsilon_j + \langle U_j, H_0\rangle - \mathbf{Z}_j^{\mathrm{T}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0))}{n}\rangle \right. \\
&\left. -\mathbf{Z}_i^{\mathrm{T}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right)^2.
\end{aligned}$$

Thus we have

$$\begin{aligned}
\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = &\frac{1}{n}\left(\frac{\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^{\mathrm{T}}}{n}\right)^{-1} \\
&\left(\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\left(\epsilon_i + \langle U_i, H_0\rangle - \langle U_i, (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j(\epsilon_j + \langle U_j, H_0\rangle)}{n}\rangle\right)\right),
\end{aligned}$$

where $\tilde{\mathbf{Z}}_i = \mathbf{Z}_i - \langle U_i, (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j\mathbf{Z}_j}{n}\rangle$. Here $\langle U_i, (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j\mathbf{Z}_j}{n}\rangle$ denotes the p-vector with components $\langle U_i, (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j Z_{jk}}{n}\rangle$ with Z_{jk} being the k-th component of $\mathbf{Z}_j$ for $k = 1, 2, \ldots, p$.

As mentioned before, we define $\widehat{F}^*$ to be the estimator of F_0 assuming $\boldsymbol{\theta}_0$ is known. Thus we have $\widehat{H}^* := K^{-1/2}\widehat{F}^* = (T_n+\lambda I)^{-1}\frac{\sum_{i=1}^{n} U_i(\epsilon_i + \langle U_i, H_0\rangle)}{n}$. Note that $\widehat{F}^*$ is exactly the regularized estimator of F_0 in a functional linear model without the multivariate part. Also, we define $\widehat{\mathbf{g}} := K^{-1/2}\widehat{\boldsymbol{\gamma}} = (K^{-1/2}\widehat{\gamma}_1, \ldots, K^{-1/2}\widehat{\gamma}_p)^{\mathrm{T}} = (T_n+\lambda I)^{-1}\frac{\sum_{j=1}^{n} U_j\mathbf{Z}_j}{n}$, and actually $\widehat{\boldsymbol{\gamma}}$ is an estimator of $\boldsymbol{\gamma}_0$ defined in (2.7) based on the RKHS approach. Then

$$\begin{aligned}
& \sum_{i=1}^{n} \tilde{\mathbf{Z}}_i \left(\langle U_i, H_0 \rangle - \langle U_i, (T_n + \lambda I)^{-1} \frac{\sum_{j=1}^{n} U_j (\epsilon_j + \langle U_j, H_0 \rangle)}{n} \rangle \right) \\
= & \sum_{i=1}^{n} \tilde{\mathbf{Z}}_i \langle U_i, H_0 - \widehat{H}^* \rangle \\
= & \sum_{i=1}^{n} (\mathbf{Z}_i - \langle U_i, \widehat{\mathbf{g}} \rangle) \langle U_i, H_0 - \widehat{H}^* \rangle \\
= & \sum_{i=1}^{n} (\mathbf{Z}_i - \langle U_i, \mathbf{g}_0 \rangle) \langle U_i, H_0 - \widehat{H}^* \rangle + \sum_{i=1}^{n} \langle U_i, \mathbf{g}_0 - \widehat{\mathbf{g}} \rangle \langle U_i, H_0 - \widehat{H}^* \rangle.
\end{aligned}$$

Using that $(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0 \rangle) U_i = (\mathbf{Z}_i - \int \boldsymbol{\gamma}_0(t, X(t)) dt) U_i = \boldsymbol{\eta}_i U_i$ has mean zero, the first term above is $O_p(\sqrt{n} \| H_0 - \widehat{H}^* \|) = O_p(\sqrt{n} \| F_0 - \widehat{F}^* \|_{\mathcal{H}})$, which is $O_p(\sqrt{n})$ when $r = 0$ and $o_p(\sqrt{n})$ when $r > 0$ (see (S1.3) in Theorem 1 in the Appendix B). The second term above is, by Cauchy-Schwarz inequality and (S1.4) in the appendix,

$$O_p(n \| T_n^{1/2} (\mathbf{g}_0 - \widehat{\mathbf{g}}) \| \| T_n^{1/2} (H_0 - \widehat{H}^*) \|) = O_p \left(n^{1 - \frac{(1+r)\alpha}{(1+r)\alpha + 1}} \right) = o_p(\sqrt{n})$$

when $\alpha > 1$ and $r \in [0,1]$. Furthermore,

$$\begin{aligned}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\epsilon_i &= \sum_{i=1}^{n}(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)\epsilon_i + \sum_{i=1}^{n}\langle U_i, \mathbf{g}_0 - \widehat{\mathbf{g}}\rangle\epsilon_i \\ &= \sum_{i=1}^{n}(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)\epsilon_i + \langle\sum_{i=1}^{n}U_i\epsilon_i, \mathbf{g}_0 - \widehat{\mathbf{g}}\rangle \\ &= \sum_{i=1}^{n}(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)\epsilon_i + O_p(\sqrt{n})\|\mathbf{g}_0 - \widehat{\mathbf{g}}\|,\end{aligned}$$

$$\begin{aligned}&\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^{\mathrm{T}} - \sum_{i=1}^{n}(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)^{\mathrm{T}} \\ = &\sum_{i=1}^{n}(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)\langle U_i, \mathbf{g}_0 - \widehat{\mathbf{g}}\rangle^{\mathrm{T}} + \sum_{i=1}^{n}\langle U_i, \mathbf{g}_0 - \widehat{\mathbf{g}}\rangle(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)^{\mathrm{T}} \\ &+ \sum_{i=1}^{n}\langle U_i, \mathbf{g}_0 - \widehat{\mathbf{g}}\rangle\langle U_i, \mathbf{g}_0 - \widehat{\mathbf{g}}\rangle^{\mathrm{T}} \\ = &O_p(\sqrt{n})\|\mathbf{g}_0 - \widehat{\mathbf{g}}\| + O_p(n\|T_n^{1/2}(\mathbf{g}_0 - \widehat{\mathbf{g}})\|^2) \\ = &o_p(n).\end{aligned}$$

Thus $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(n^{-1/2})$ if $r = 0$, while if $r > 0$ the dominant term in $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ is $\frac{1}{n}\Sigma_1^{-1}\sum_{i=1}^{n}(\mathbf{Z}_i - \langle U_i, \mathbf{g}_0\rangle)\epsilon_i$ which established the asymptotic normality property.

For the estimator $\widehat{H}$, we have

$$
\begin{aligned}
& \widehat{H} - H_0 \\
= \;& (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n U_i (Y_i - \mathbf{Z}_i^{\mathrm{T}} \widehat{\boldsymbol{\theta}})}{n} - H_0 \\
= \;& (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n U_i (Y_i - \mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\theta}_0)}{n} - H_0 - (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n U_i \mathbf{Z}_i^{\mathrm{T}} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}{n} \\
= \;& (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n U_i (Y_i - \mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\theta}_0)}{n} - H_0 - \widehat{\mathbf{g}}^{\mathrm{T}} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).
\end{aligned}
$$

Note the first term above is just $\widehat{H}^* = K^{-1/2} \widehat{F}^*$. Furthermore,

$$
\begin{aligned}
& T^{1/2} \widehat{\mathbf{g}}^{\mathrm{T}} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
= \;& T^{1/2} (\widehat{\mathbf{g}} - \mathbf{g}_0)^{\mathrm{T}} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + T^{1/2} \mathbf{g}_0^{\mathrm{T}} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\
= \;& O_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|) = O_p(1/\sqrt{n}),
\end{aligned}
$$

which finished the proof for $\|T^{1/2}(\widehat{H} - H_0)\|$. □

S1.1 Results for purely functional additive regression

In this part, we consider the purely functional model $Y_i = \langle U_i, H_0 \rangle + \epsilon_i$ under the same conditions (A1)-(A3) (except that all assumptions involving $\mathbf{Z}$ now becomes void). With abuse of notation, the estimator is

$$
\widehat{H} = \operatorname*{argmin}_H \sum_i (Y_i - \langle U_i, H \rangle)^2 + n\lambda \|H\|^2
$$

(instead of writing it as $\widehat{H}^*$ as in the main text).

Theorem 1. *Assume (A1)-(A3). If $s_j \asymp j^{-\alpha}$ for some constant $\alpha > 1$ and $r \in [0,1]$ in assumptions (A3), by setting $\lambda \asymp n^{-\alpha/((1+r)\alpha+1)}$, we have*

$$E^*\langle \widehat{H} - H_0, U^* \rangle^2 = \|T^{1/2}(\widehat{H} - H_0)\|^2 = O_p\left(n^{-\frac{(1+r)\alpha}{(1+r)\alpha+1}}\right), \tag{S1.2}$$

and

$$\|\widehat{F} - F_0\|_{\mathcal{H}}^2 = \|\widehat{H} - H_0\|^2 = O_p\left(n^{-\frac{r\alpha}{(1+r)\alpha+1}}\right). \tag{S1.3}$$

Furthermore, we have

$$(1/n)\sum_{i=1}^{n}\langle \widehat{H} - H_0, U_i \rangle^2 = \|T_n^{1/2}(\widehat{H} - H_0)\|^2 = O_p(n^{-1/2}) \text{ and } o_p(n^{-1/2}), \tag{S1.4}$$

when $r = 0$ and $r > 0$, respectively

Remark 1. *The first rate above with $r = 0$ (assuming $F_0 \in \mathcal{H}$ without further smoothness assumptions) is the same as the bound in Cai and Yuan (2012). Note that when $r = 0$, we only have $\|\widehat{F} - F_0\|_{\mathcal{H}}^2 = \|\widehat{H} - H_0\|^2 = O_p(1)$ meaning the estimator is inconsistent under this $\|.\|_{\mathcal{H}}$ error measure. Both convergence rates become faster as r increases.*

Proof of Theorem 1. In the proofs we use C to denote a generic positive constant. Using $Y_i = \langle H_0, U_i \rangle + \epsilon_i$, we have

$$\begin{aligned}
& \widehat{H} - H_0 \\
= \; & (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n \langle H_0, U_i \rangle U_i}{n} + (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n \epsilon_i U_i}{n} - H_0 \\
= \; & \left((T_n + \lambda I)^{-1} T_n - I \right) H_0 + (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n \epsilon_i U_i}{n} \\
= \; & -\lambda (T_n + \lambda I)^{-1} H_0 + (T_n + \lambda I)^{-1} \frac{\sum_{i=1}^n \epsilon_i U_i}{n} \\
=: \; & A_1 + A_2.
\end{aligned}$$

Furthermore, we decompose

$$\begin{aligned}
A_1 &= -\lambda (T + \lambda I)^{-1} H_0 - \lambda (T_n + \lambda I)^{-1} (T - T_n)(T + \lambda I)^{-1} H_0 \\
&= -\lambda (T + \lambda I)^{-1} H_0 - \lambda (T + \lambda I)^{-1} (T - T_n)(T + \lambda I)^{-1} H_0 \\
&\quad + \lambda (T + \lambda I)^{-1} (T_n - T)(T_n + \lambda I)^{-1} (T - T_n)(T + \lambda I)^{-1} H_0 \\
&= A_{11} + A_{12} + A_{13},
\end{aligned}$$

where in both the first equality and the second, we used the identity $B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1} = A^{-1}(A - B)B^{-1}$ with $A = T + \lambda I$ and $B = T_n + \lambda I$.

For simplicity, we define $H_0 = \sum_{k=1}^\infty b_k e_k$, $S_1 := \sum_{j=1}^\infty b_j^2/(s_j + \lambda)$ and

$S_2 = \sum_{j=1}^{\infty} s_j/(s_j+\lambda)$. Obviously, from Lemma 1, it follows

$$\begin{aligned}\|A_{11}\|^2 &= \left\|-\lambda(T+\lambda I)^{-1}H_0\right\|^2 \\ &= \lambda^2 \left\| \left(\sum_{j=1}^{n} (s_j+\lambda)^{-1}(e_j \otimes e_j) \right) \sum_{k=1}^{n} b_k e_k \right\|^2 \leq \lambda S_1.\end{aligned}$$

For A_{12}, similar to Lemma 2, we have

$$\begin{aligned}\|A_{12}\|^2 &\leq \|A_{11}\|^2 \|(T+\lambda I)^{-1}(T-T_n)\|_{hs}^2 \\ &= O(\lambda S_1) \cdot O_p\left(\frac{1}{n} \sum_{k=1}^{\infty} \frac{s_k}{(s_k+\lambda)^2} \right) \\ &= O_p\left(\frac{S_1 S_2}{n} \right).\end{aligned}$$

For A_{13}, we have

$$\begin{aligned}\|A_{13}\|^2 &\leq \|(T+\lambda I)^{-1}(T_n-T)\|_{hs}^2 \times \|\lambda(T_n+\lambda I)^{-1}\|_{op}^2 \times \\ &\quad \|(T-T_n)(T+\lambda I)^{-1/2}\|_{hs}^2 \times \|(T+\lambda I)^{-1/2}H_0\|^2 \\ &= O_p\left(\frac{S_2}{n\lambda}\right) \cdot 1 \cdot O_p\left(\frac{S_2}{n}\right) \cdot S_1 \\ &= O_p\left(\frac{S_1 S_2^2}{n\lambda}\right).\end{aligned}$$

Now, write $A_2 = (T+\lambda I)^{-1}\frac{\sum_{i=1}^{n}\epsilon_i U_i}{n} + (T+\lambda I)^{-1}(T-T_n)(T_n+\lambda I)^{-1}\frac{\sum_{i=1}^{n}\epsilon_i U_i}{n} =:$

$A_{21} + A_{22}$. We have

$$\begin{aligned} E\|A_{21}\|^2 &= E[tr(A_{21} \otimes A_{21})] \\ &= \frac{1}{n} tr\left((T+\lambda I)^{-1} E[\epsilon_1^2 (U_1 \otimes U_1)](T+\lambda I)^{-1}\right) \\ &= O\left(\frac{1}{n} tr(T(T+\lambda I)^{-2})\right) \\ &= O\left(\frac{S_2}{n\lambda}\right). \end{aligned}$$

Furthermore, we denote $\mathcal{C} = (T+\lambda I)^{-1}(T - T_n)(T_n + \lambda I)^{-1}$ to avoid lengthy expressions below and consider the conditional expectation as follows.

$$\begin{aligned} & E[\|A_{22}\|^2 | \{X_i\}_{i=1}^n] \\ = \;& E[tr(A_{22} \otimes A_{22}) | \{X_i\}_{i=1}^n] \\ = \;& O_p\left(\frac{1}{n} tr(\mathcal{C} T_n \mathcal{C})\right) \\ = \;& O_p\left(\frac{1}{n} \|T_n^{1/2} \mathcal{C}\|_{hs}^2\right) \\ = \;& O_p\left(\frac{1}{n} \|T_n^{1/2}(T_n+\lambda I)^{-1}(T - T_n)(T+\lambda I)^{-1}\|_{hs}^2\right) \\ \leq \;& O_p\left(\frac{1}{n} \|T_n^{1/2}(T_n+\lambda I)^{-1/2}\|_{op}^2 \|(T_n+\lambda I)^{-1/2}\|_{op}^2 \|(T - T_n)(T+\lambda I)^{-1}\|_{hs}^2\right) \\ = \;& O_p\left(\frac{1}{n\lambda}\right) \cdot O_p\left(\frac{S_2}{n\lambda}\right). \end{aligned}$$

The rate (S1.3) is obtained by combining the bounds for $A_{11}, A_{12}, A_{13}, A_{21}, A_{22}$ above. More specifically, using the bounds $S_1 = O(\lambda^{r-1})$ and $S_2 = O(\lambda^{-1/\alpha})$

in Lemma 1, choosing λ to balance the two terms in $\lambda S_1 + S_2/(n\lambda)$, we get $\lambda \asymp n^{-\alpha/((1+r)\alpha+1)}$. With this choice of λ, other terms in the bounds above are of smaller order and we get the rates are dominated by $\lambda S_1 + S_2/(n\lambda)$ which is as in (S1.3).

The proof of (S1.2) is similar to the proof of (S1.3) with some changes. First we note that the prediction risk can be written as $\|T^{1/2}(\widehat{H} - H_0)\|$. We have

$$
\begin{aligned}
& T^{1/2}\widehat{H} - T^{1/2}H_0 \\
= \; & A^*_{11} + A^*_{12} + A^*_{13} + A^*_{21} + A^*_{22}
\end{aligned}
$$

where

$$
\begin{aligned}
A^*_{11} &= -\lambda T^{1/2}(T + \lambda I)^{-1}H_0 \\
A^*_{12} &= -\lambda T^{1/2}(T + \lambda I)^{-1}(T - T_n)(T + \lambda I)^{-1}H_0 \\
A^*_{13} &= \lambda T^{1/2}(T + \lambda I)^{-1}(T_n - T)(T_n + \lambda I)^{-1}(T - T_n)(T + \lambda I)^{-1}H_0 \\
A^*_{21} &= T^{1/2}(T + \lambda I)^{-1}\frac{\sum_{i=1}^n \epsilon_i U_i}{n} \\
A^*_{22} &= T^{1/2}(T + \lambda I)^{-1}(T - T_n)(T_n + \lambda I)^{-1}\frac{\sum_{i=1}^n \epsilon_i U_i}{n}.
\end{aligned}
$$

The main modification is simply to use $\|(T + \lambda I)^{-1/2}T^{1/2}\|_{op} \leq 1$ and the fact that for two Hilbert-Schmidt operators A and B, $\|AB\|_{hs} \leq \|A\|_{hs}\|B\|_{op}$. More specifically, we have

$$\|A_{11}^*\|^2 = \lambda^2 S_1,$$

$$\begin{aligned}
\|A_{12}^*\|^2 &\leq \|\lambda(T+\lambda I)^{-1/2}\|_{op}^2 \|(T+\lambda I)^{-1/2}H_0\|^2 \|(T-T_n)(T+\lambda I)^{-1/2}\|_{hs}^2 \\
&= O_p\left(\frac{\lambda}{n} S_1 S_2\right),
\end{aligned}$$

by Lemma 2, and

$$\begin{aligned}
\|A_{13}^*\|^2 &\leq \|(T+\lambda I)^{-1/2}(T-T_n)\|_{hs}^2 \times \|\lambda(T_n+\lambda I)^{-1}\|_{op}^2 \times \\
&\quad \|(T-T_n)(T+\lambda I)^{-1/2}\|_{hs}^2 \times \|(T+\lambda I)^{-1/2}H_0\|^2 \\
&= O_p\left(\frac{S_1 S_2^2}{n^2}\right).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
E\|A_{21}^*\|^2 &= E[tr(A_{21}^* \otimes A_{21}^*)] \\
&= \frac{1}{n} tr(T(T+\lambda I)^{-1} E[\epsilon_i^2 (U_i \otimes U_i)](T+\lambda I)^{-1}) \\
&= O_p\left(\frac{1}{n} tr(T^2(T+\lambda I)^{-2})\right) \\
&= O_p\left(\frac{1}{n} \sum_{j=1}^{n} \frac{s_j^2}{(s_j+\lambda)^2}\right) \\
&= O_p\left(\frac{S_2}{n}\right),
\end{aligned}$$

where in the last step we used $s_j/(s_j+\lambda)\le 1$, and also

$$\begin{aligned}
& E[\|A_{22}^*\|^2|\{X_i\}_{i=1}^n] \\
= \; & E[tr(A_{22}^*\otimes A_{22}^*)|\{X_i\}_{i=1}^n] \\
= \; & O_p\left(\frac{1}{n}\|T_n^{1/2}(T_n+\lambda I)^{-1}(T-T_n)(T+\lambda I)^{-1}T^{1/2}\|_{hs}^2\right) \\
= \; & O_p\left(\frac{1}{n}\|T_n^{1/2}(T_n+\lambda I)^{-1}\|_{op}^2\|(T-T_n)(T+\lambda I)^{-1/2}\|_{hs}^2\right) \\
= \; & O_p\left(\frac{1}{n\lambda}\right)\cdot O_p\left(\frac{S_2}{n}\right).
\end{aligned}$$

To see (S1.2), using the bounds for $A_{11}^*,\dots,A_{22}^*$ above, the same value of λ balances the two terms in $\lambda^2 S_1 + S_2/n$ and other terms in the rate are of smaller order and thus the convergence rate is $O_p(n^{-(1+r)\alpha/((1+r)\alpha+1)})$.

Finally, to establish (S1.4), we use

$$\begin{aligned}
& \|T_n^{1/2}(\widehat{H}-H_0)\|^2 \\
= \; & \|T^{1/2}(\widehat{H}-H_0)\|^2 - \langle \widehat{H}-H_0,(T-T_n)(\widehat{H}-H_0)\rangle \\
\le \; & \|T^{1/2}(\widehat{H}-H_0)\|^2 + O_p(n^{-1/2}\|\widehat{H}-H_0\|^2).
\end{aligned}$$

By (S1.2) and (S1.3), the above is $O_p(n^{-1/2})$ if $r=0$ and $o_p(n^{-1/2})$ if $r>0$. □

Lemma 1. *For $r\in[0,1]$ and $s_j \asymp j^{-\alpha}$ for some $\alpha>1$, we have $S_1 := \sum_{j=1}^\infty b_j^2/(s_j+\lambda) = O(\lambda^{r-1})$ and $S_2 = \sum_{j=1}^\infty s_j/(s_j+\lambda) = O(\lambda^{-1/\alpha})$.*

Proof of Lemma 1. When $r \in [0,1]$, we have

$$\begin{aligned}
S_1^2 &= \sum_j \frac{b_j^2}{(s_j+\lambda)^2} \\
&= \sum_j \frac{b_j^2}{s_j^r} \frac{s_j^r}{(s_j+\lambda)^2} \\
&\leq C \max_j \frac{s_j^r}{(s_j+\lambda)^2} \\
&\leq C \max_j \frac{1}{(s_j+\lambda)^{2-r}} \\
&\leq C\lambda^{r-2}.
\end{aligned}$$

Let $J = \lambda^{-1/\alpha}$, by splitting the sum into $\sum_{j=1}^{J}$ and $\sum_{j=J+1}^{\infty}$, we have

$$\begin{aligned}
S_2 &= \sum_j \frac{s_j}{s_j+\lambda} \\
&\leq C\sum_{j=1}^{J} 1 + C\lambda^{-1} \sum_{j=J+1}^{\infty} j^{-\alpha} \\
&= O(J) + \lambda^{-1} O(J^{-\alpha+1}) \\
&= O(\lambda^{-1/\alpha}).
\end{aligned}$$

We complete the proof. □

Lemma 2. *Under the same assumptions as in Theorem 1, we have*

$$E\|(T+\lambda I)^{-1/2}(T-T_n)\|_{op}^2 = O\left(\frac{S_2}{n}\right). \tag{S1.5}$$

Proof of Lemma 2. Write $W_i = (T+\lambda I)^{-1/2}U_i \otimes U_i$ and $V_i = W_i - E(W_i)$. A simple derivation leads to that

$$\begin{aligned} E\|(T+\lambda I)^{-1/2}(T-T_n)\|_{hs}^2 &= E\left[\langle \frac{1}{n}\sum_{i=1}^{n} V_i, \frac{1}{n}\sum_{i=1}^{n} V_i\rangle_{hs}\right] \\ &= \frac{1}{n}E[\langle V_1, V_1\rangle_{hs}] \\ &= \frac{1}{n}E[\langle W_1, W_1\rangle_{hs}] - \frac{1}{n}[\langle E(W_1), E(W_1)\rangle_{hs}] \\ &\leq \frac{1}{n}E[\langle W_1, W_1\rangle_{hs}]. \end{aligned}$$

Without confusion, write the Karhunen-Loève expansion of U_1 as $U_1 = \sum_{j=1}^{\infty} \xi_j e_j$, where ξ_j denote the random coefficients satisfying that $E(\xi_j) = 0$, $E(\xi_j^2) = s_j$. Then, we have

$$\begin{aligned} \langle W_1, W_1\rangle_{hs} &= \sum_{j,k} \langle (T+\lambda I)^{-1/2}(U_1 \otimes U_1)e_j, e_k\rangle^2 \\ &= \sum_{j,k} \xi_j^2 \langle (T+\lambda I)^{-1/2}U_1, e_k\rangle^2 \\ &= \sum_{j,k} \xi_j^2 \langle U_1, (T+\lambda I)^{-1/2}e_k\rangle^2 \\ &= \sum_{j,k} \xi_j^2 \langle U_1, e_k/\sqrt{s_k+\lambda})\rangle^2 \\ &= \sum_{j,k} \frac{\xi_j^2 \xi_k^2}{s_k+\lambda}. \end{aligned}$$

This implies, by assumption (A2),

$$E[\langle W_1, W_1 \rangle_{hs}] = \sum_{j,k} \frac{s_j s_k}{s_k + \lambda} \leq C \sum_k \frac{s_k}{s_k + \lambda} = CS_2.$$

This lemma then follows immediately. □

Bibliography

Cai, T. and Yuan, M. (2012) Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, **107**, 1201–1216.