Asymmetric estimation for varying-coefficient additive model with functional response in reproducing kernel Hilbert space

Yi Liu[1], Wei Tu[2], Yanchun Bao[3], Bei Jiang[4], Linglong Kong[4]

[1] *York University,* [2] *Queen's University,* [3] *University of Essex and* [4] *Univertsity of Alberta*

Supplementary Material

This supplemental material contains the technical proofs for Theorems 1–4.

S1 Optimal Rate of Convergence under random design

Proof of Theorem 1. Note that any lower bound for a specific case yields a lower bound for a general case. We set $\tau = \frac{1}{2}$ and assume $(\epsilon_i(\mathcal{T}_{i1}), \ldots, \epsilon_i(\mathcal{T}_{im}))$ follows a normal distribution with zero mean and $\operatorname{Cov}(\epsilon_i(\mathcal{T}_{ij}), \epsilon_i(\mathcal{T}_{ik})) = C_0(\mathcal{T}_{ij}, \mathcal{T}_{ik})$ where $C_0(s,s) = \sigma_0^2$.

The Varshamov-Gilbert bound shows that for any $m \geq 8$, there exists a set $\mathcal{B} = \left\{b^{(0)}, b^{(1)}, \ldots, b^{(N_0)}\right\} \subset \{0,1\}^m$ such that

1. $b^{(0)} = (0, \ldots, 0)'$

2. $H(b,b') > m/8$ for any $b \neq b' \in \mathcal{B}$, where $H(\cdot,\cdot) = \sum_{i=1}^m |b_i - b_i'|$ is the Hamming distance;

3. $N_0 \geq 2^{m/8}$.

Let $N = \lceil mn^{1/(2r+1)} \rceil$. Define $\boldsymbol{\beta} = \sqrt{N^{-1}} \sum_{k=N+1}^{2N} b_k \sqrt{\rho_k} \phi_k$.

Then $\|\boldsymbol{\beta}_b\|_{\mathcal{H}^p}^2 = N^{-1} \sum_{k=1}^N b_k \leq 1$ and $\|\boldsymbol{\beta}_b\|_2^2 = N^{-1} \sum_{k=N+1}^{2N} b_k \rho_k$.

Notice that,

$$\begin{cases} N^{-1} H(b,b') \rho_{2N} \lesssim N^{-1} \displaystyle\sum_{k=N+1}^{2N} b_k \rho_k \lesssim N^{-1} H(b,b') \rho_N, \\ \dfrac{1}{8} \leq \dfrac{H(b,b')}{N} \leq 1. \end{cases} \tag{S1.1}$$

Combine S1.1 with condition the eigenvalue condition $\rho_k \approx k^{-2r}$, we have

$$N^{-r} \lesssim \|\boldsymbol{\beta} - \boldsymbol{\beta}_{b'}\|_2^2 \lesssim N^{-2r}. \tag{S1.2}$$

By S1.2 and condition (A1) (A2),

$$\begin{aligned} \mathrm{KL}\left(\Pi_{Y|\boldsymbol{\beta}_b} \mid \Pi_{Y|\boldsymbol{\beta}_{b'}}\right) &= n\mathbb{E}\left[(x\boldsymbol{\beta}_b - x\boldsymbol{\beta}_{b'})' \Sigma^{-1}(T) (x\boldsymbol{\beta}_b - x\boldsymbol{\beta}_{b'})\right] \\ &\lesssim nm\sigma_0^{-2} \|\boldsymbol{\beta}_b - \boldsymbol{\beta}_{b'}\|_2^2 \\ &\lesssim nm\sigma_0^{-2} N^{-2r}. \end{aligned}$$

Fano's lemma now yields

$$\begin{aligned} \max_{1\leq j\leq M} \mathbb{E}_{\boldsymbol{\beta}^{(j)}} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(j)}\|_2^2 &\gtrsim N^{-r} \left(1 - \frac{\log\left(c_1 nm\sigma_0^{-2} N^{-2r}\right) + \log 2}{\log M}\right) \\ &\approx (nm)^{-r/(2r+1)}. \end{aligned}$$

This gives:

$$\lim_{a\to 0}\lim_{n\to\infty}\sup_{\boldsymbol{\beta}_0}\mathbb{P}\left(\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|_2^2 \geq a\left((nm)^{-r/(2r+1)}\right)\right)=1.$$

Let $(\epsilon_i(\mathcal{T}_{i1}),\ldots,\epsilon_i(\mathcal{T}_{im}))$ follows a normal distribution with unknown mean μ and $\mathrm{Cov}(\epsilon_i(\mathcal{T}_{ij}),\epsilon_i(\mathcal{T}_{ik}))=\sigma_0^2$. Then the estimation of $\boldsymbol{\beta}$ is equivalent to the estimation of μ from n i.i.d samples.

This gives:

$$\lim_{a\to 0}\lim_{n\to\infty}\sup_{\boldsymbol{\beta}_0}\mathbb{P}\left(\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|_2^2 \geq an^{-1}\right)=1,$$

as desired. □

Proof of Theorem 2. Define:

$$\begin{cases}
\ell_\infty(g)=\mathbb{E}\left(\dfrac{1}{nm}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m}\rho\left(Y_{ij}-x_i^T g\left(\mathcal{T}_{ij}\right)\right)\right)\\
\ell_{mn}(g)=\dfrac{1}{nm}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m}\rho\left(Y_{ij}-x_i^T\left(\mathcal{T}_{ij}\right)\right)\\
\ell_{mn,\lambda}(g)=\ell_{mn}(g)+\lambda\|g\|_{\mathcal{H}^p}^2,\\
\ell_{\infty,\lambda}(g)=\ell_\infty(g)+\lambda\|g\|_{\mathcal{H}^p}^2.\\
\boldsymbol{\beta}_\lambda^\infty=\operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathcal{H}(K)}\left\{\ell_\infty(\boldsymbol{\beta})+\lambda\|\boldsymbol{\beta}\|_{\mathcal{H}^p}^2\right\}\\
\boldsymbol{\beta}_\lambda^{mn}=\operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathcal{H}(K)}\left\{\ell_{mn}(\boldsymbol{\beta})+\lambda\|\boldsymbol{\beta}\|_{\mathcal{H}^p}^2\right\}
\end{cases}$$

Notice that under (C4) when $\lambda=0$, $g=\boldsymbol{\beta}_0$ minimizes all $\rho\left(Y_{ij}^k-x_i^T g^k\left(\mathcal{T}_{ij}\right)\right)$, therefore $\boldsymbol{\beta}_\lambda^\infty|_{\lambda=0}=\boldsymbol{\beta}_0$.

First we bound the regulated estimator and the unregulated estimator with infinite samples: $\|\boldsymbol{\beta}_\lambda^\infty - \boldsymbol{\beta}_0\|_2^2$. Consider the following eigen decomposition:

$$\begin{cases} \boldsymbol{\beta}_0 = \sum_{v\geq 1} a_v \phi_v, \\ \boldsymbol{\beta}_\lambda^\infty = \sum_{v\geq 1} b_v \phi_v. \end{cases}$$

Define

$$\begin{aligned} \ell_\infty^2(g) &= \mathbb{E}\left(\frac{1}{nm}\sum_{i=1}^n\sum_{j=1}^m \left(Y_{ij} - x_i^T g\left(\mathcal{T}_{ij}\right)\right)^2\right) \\ &= \mathbb{E}\left(\left[Y_{11} - x^T g_0\left(\mathcal{T}_{11}\right)\right]^2\right) + \mathbb{E}\left(\|x^T g - x^T g_0\|_2^2\right), \end{aligned}$$

Let $\tau_0 = min(\tau, 1-\tau)$ and $c_0 = \tau_0 \mathbb{E}\left(\|x^T x\|_2^2\right)$.

First we prove $|\frac{a_k}{1+\lambda c_0^{-1}\rho_k^{-1}}| \leq |b_k| \leq |a_k|$ and $a_k b_k \geq 0$ by contradiction:

Define $f_\epsilon = \boldsymbol{\beta}_\lambda^\infty + \epsilon\phi_k$, by the fact that $c_0 \leq \frac{\ell_\infty(f) - \ell_\infty(g)}{\ell_\infty^2(f) - \ell_\infty^2(g)} \leq 1$.

$$\begin{aligned} \ell_{\infty,\lambda}(\boldsymbol{\beta}_\lambda^\infty) - \ell_{\infty,\lambda}(f_\epsilon) &= \ell_\infty(f_\epsilon) - \ell_\infty(\boldsymbol{\beta}_\lambda^\infty) + \lambda(\|\boldsymbol{\beta}_\lambda^\infty\|_{\mathcal{H}^p}^2 - \|f_\epsilon\|_{\mathcal{H}^p}^2) \\ &\geq \tau_0\left(\ell_\infty^2(\boldsymbol{\beta}_\lambda^\infty) - \ell_\infty^2(f_\epsilon)\right) + \lambda(\|\boldsymbol{\beta}_\lambda^\infty\|_{\mathcal{H}^p}^2 - \|f_\epsilon\|_{\mathcal{H}^p}^2) \\ &\geq c_0((a_k - b_k)^2 - (a_k - b_k - \epsilon)^2) - \lambda\rho_k^{-1}(b_k^2 - (b_k + \epsilon)^2) \end{aligned}$$

If $a_k b_k < 0$, take $\epsilon = -b_k$ yields $\ell_{\infty,\lambda}(\boldsymbol{\beta}_\lambda^\infty) - \ell_{\infty,\lambda}(f_\epsilon) > 0$.

If $|b_k| > |a_k|$, take $\epsilon = a_k - b_k$ yields $\ell_{\infty,\lambda}(\boldsymbol{\beta}_\lambda^\infty) - \ell_{\infty,\lambda}(f_\epsilon) > 0$.

Otherwise let $\epsilon b_k < 0$ and sufficiently small in scale:

$$\begin{aligned}\ell_{\infty,\lambda}(\boldsymbol{\beta}_\lambda^\infty) - \ell_{\infty,\lambda}(f_\epsilon) &\geq c_0((a_k - b_k)^2 - (a_k - b_k - \epsilon)^2) - \lambda\rho_k^{-1}(b_k^2 - (b_k+\epsilon)^2) \\ &\approx \epsilon \frac{\partial[c_0(a_k - x)^2 + \lambda\rho_k x^2]}{\partial x}\Big|_{x=b_k} \\ &= \epsilon c_0 \left(b_k - \frac{a_k}{1+\lambda c_0^{-1}\rho_k^{-1}} \right) \\ &> 0.\end{aligned}$$

Therefore,

$$\|\boldsymbol{\beta}_\lambda^\infty - \boldsymbol{\beta}_0\|_2^2 = \sum_{k\geq 1}(a_k - b_k)^2 \leq \sum_{k\geq 1} a_k^2 (\frac{c_0^{-1}\rho_k^{-1}}{1+\lambda c_0^{-1}\rho_k^{-1}})^2 \lesssim \sum_{k\geq 1} a_k^2 \rho_k^{-1} \sup_{v\geq 1} \frac{\rho_k^{-1}}{\left(1+\lambda\rho_k^{-1}\right)^2} \leq \lambda \|\boldsymbol{\beta}_0\|_K^2.$$

Let

$$\tilde{\boldsymbol{\beta}_\lambda^\infty} = \boldsymbol{\beta}_\lambda^\infty - \frac{1}{2} G_\lambda^{-1} D\ell_{mn,\lambda}(\boldsymbol{\beta}_\lambda^\infty)$$

where D stands for the Fréchet derivative and $G_\lambda = (1/2)D^2\ell_{\infty,\lambda}(\boldsymbol{\beta}_\lambda^\infty)$.

By Lax-Milgram theorem,

$$\begin{aligned}\mathbb{E}\|\tilde{\boldsymbol{\beta}_\lambda^\infty} - \boldsymbol{\beta}_\lambda^\infty\|_2^2 &= \mathbb{E}\left\| \frac{1}{2} G_\lambda^{-1} D\ell_{nm,\lambda}(\boldsymbol{\beta}_\lambda^\infty) \right\|_2^2 \\ &= \frac{1}{4}\mathbb{E}\left[\sum_{k\geq 1} \left(1+\lambda\rho_k^{-1}\right)^{-2} \left(D\ell_{nm,\lambda}(\boldsymbol{\beta}_\lambda^\infty)\phi_k\right)^2 \right]\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[D\ell_{mn,\lambda}(\boldsymbol{\beta}_\lambda^\infty)\phi_k\right]^2 &= \mathbb{E}\left[D\ell_{mn}(\boldsymbol{\beta}_\lambda^\infty)\phi_k - D\ell_\infty(\boldsymbol{\beta}_\lambda^\infty)\phi_k\right]^2 \\
&= \frac{4}{n^2m^2}\sum_{i=1}^n \mathrm{Var}\left[\sum_{j=1}^m \rho'\left(\left[Y_{ij} - x_i^T\boldsymbol{\beta}_\lambda^\infty\left(\mathcal{T}_{ij}\right)\right]\phi_k\left(\mathcal{T}_{ij}\right)\right)\right] \\
&\lesssim \frac{1}{n^2m^2}\left(\mathrm{Var}\left[\sum_{j=1}^{m_i}\left(\left[\boldsymbol{\beta}_0\left(\mathcal{T}_{ij}\right) - \tilde{\boldsymbol{\beta}_\lambda^\infty}\left(\mathcal{T}_{ij}\right)\right]\phi_k\left(\mathcal{T}_{ij}\right)\right)\right] + \mathbb{E}\left[\mathrm{Var}\left(\sum_{j=1}^m Y_{ij}\phi_k\left(\mathcal{T}_{ij}\right) \mid \mathcal{T}\right)\right]\right. \\
&\lesssim \frac{1}{n^2m^2}\left(m\|\boldsymbol{\beta}_0 - \boldsymbol{\beta}_\lambda^\infty\|_2^2\|\phi_k\|_2^2 + m\left(m-1\right)\int_{\mathcal{T}\times\mathcal{T}}\phi_k(s)C_0(s,t)\phi_k(t)dsdt + m\sigma_0^2\|\phi_k\|_2^2\right. \\
&+ m\int_{\mathcal{T}}\phi_k^2(s)C(s,s)ds\Big) \\
&\lesssim \frac{1}{mn^2} + \frac{c_k}{n},
\end{aligned}$$

where $\int_{\mathcal{T}\times\mathcal{T}}\phi_k(s)C_0(s,t)\phi_k(t)dsdt$. Therefore,

$$\mathbb{E}\|\tilde{\boldsymbol{\beta}_\lambda^\infty} - \boldsymbol{\beta}_\lambda^\infty\|_2^2 \lesssim \frac{1}{mn^2}\lambda^{-1/(2r)} + \frac{1}{n}$$

Next, consider $\|\boldsymbol{\beta}_\lambda^{mn} - \tilde{\boldsymbol{\beta}_\lambda^\infty}\|_2^2$. Direct computation shows,

$$\boldsymbol{\beta}_\lambda^{mn} - \tilde{\boldsymbol{\beta}_\lambda^\infty} = \frac{1}{2}G_\lambda^{-1}\left[D^2\ell_\infty(\boldsymbol{\beta}_\lambda^\infty)(\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}_\lambda^\infty) - D^2\ell_{mn}(\boldsymbol{\beta}_\lambda^\infty)(\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}_\lambda^\infty)\right]$$

By Lax-Milgram theorem,

$\|\boldsymbol{\beta}_\lambda^{mn} - \tilde{\boldsymbol{\beta}_\lambda^\infty}\|_2^2 \le \sum_{k\ge1}\left(1+\lambda\rho_k^{-1}\right)^{-2}\left[\frac{1}{nmp}\sum_{i=1}^n\sum_{j=1}^m\sum_{k=1}^p\left(\boldsymbol{\beta}_\lambda^{mn}\left(\mathcal{T}_{ij}\right) - \tilde{\boldsymbol{\beta}_\lambda^\infty}\left(\mathcal{T}_{ij}\right)\right)\phi_k\left(\mathcal{T}_{ij}\right) - \int_{\mathcal{T}}(\boldsymbol{\beta}_\lambda^{mn}(s) - \tilde{\boldsymbol{\beta}_\lambda^\infty}(s))\phi_k(s)ds\right]^2 \le \frac{1}{mn}\sum_{k\ge1}\left(1+\lambda\rho_k^{-1}\right)^{-2} \lesssim \frac{1}{mnr\lambda^{-1/2r}}$

Therefore

$$\|\boldsymbol{\beta}_\lambda^{mn} - \tilde{\boldsymbol{\beta}_\lambda^\infty}\|_2^2 \lesssim \frac{1}{mn\lambda^{-1/2r}}.$$

$$\|\boldsymbol{\beta}_\lambda^{mn}-\boldsymbol{\beta}_0\|_2^2 \lesssim \|\boldsymbol{\beta}_\lambda^{mn}-\tilde{\boldsymbol{\beta}_\lambda^{\infty}}\|_2^2+\|\tilde{\boldsymbol{\beta}_\lambda^{\infty}}-\boldsymbol{\beta}_\lambda^{\infty}\|_2^2+\|\boldsymbol{\beta}_\lambda^{\infty}-\boldsymbol{\beta}_0\|_2^2 \approx \frac{1}{mn\lambda^{-1/2r}}+\frac{1}{n}(\frac{1}{mn}\lambda^{-1/(2r)}+1)+\lambda.$$

Take $\lambda \approx (nm)^{-2r/(2r+1)}$, we have $\|\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}_0\|_2^2 \lesssim (nm)^{-2r/(2r+1)} + \frac{1}{n}$ as desired.

□

S2 Optimal Rate of Convergence under fixed design

Proof of Theorem 3. Consider a fixed design with given sampling nodes $\mathbb{S} = \mathcal{T}_1, \cdots, \mathcal{T}_m$. There must exist an index i for which $\mathcal{T}_{i+1} - \mathcal{T}_i > \frac{1}{m}$. Let Ψ be a bump function defined as:

$$\Psi(x) = \begin{cases} \exp\left(-\frac{1}{1-x^2}\right), & x \in (-1,1) \\ 0, & \text{otherwise} \end{cases}$$

It can be readily observed that

$$\int_{\mathcal{T}} \left[\Psi^{(r)}(t)\right]^2 dt < \infty$$

for all $r \in \mathbb{N}$. Define $\boldsymbol{\beta}_1 = 0$ and $\boldsymbol{\beta}_2 = m^{-2r}\Psi\left(m(\cdot - \frac{\mathcal{T}_{i+1}+\mathcal{T}_i}{2})\right)$. Direct calculation yields:

$$\int_{\mathcal{T}} \left[\boldsymbol{\beta}_2^{(r)}(t)\right]^2 dt = \int_{\mathcal{T}} \left[\Psi^{(r)}(t)\right]^2 dt < \infty,$$

$$\|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\|_2^2 = \|\boldsymbol{\beta}_2\|_2^2 = m^{-2r}\|\Psi\|_2^2.$$

It is important to note that $\boldsymbol{\beta}_1(s_i) = \boldsymbol{\beta}_2(s_i)$ for all $i \in \{1, ..., m\}$. Thus, under fixed design, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are indistinguishable by any regressor. Therefore,

$$\lim_{n\to\infty} \inf_{\hat{\boldsymbol{\beta}}} \sup_{\boldsymbol{\beta}\in\mathcal{H}} P\left(\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|_2^2 > cm^{-2r}\right) > 0.$$

This concludes the proof, as desired.

$\square$

Proof of Theorem 4. Define:

$$\begin{cases} v_n(\boldsymbol{\beta}) = (mn)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau \left(y_i\left(\mathcal{T}_j\right) - \boldsymbol{x}_i^T \boldsymbol{\beta}\right), \\ \boldsymbol{\beta}_\lambda^{mn} = \arg\min_{\boldsymbol{\beta}\in\mathcal{H}^p} v_n(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_{\mathcal{H}^p}^2, \\ \|\boldsymbol{\beta}\|_s^2 = (mp)^{-1} \sum_{j=1}^{m} \left(\boldsymbol{\beta}\left(\mathcal{T}_j\right)\right)^T \left(\boldsymbol{\beta}\left(\mathcal{T}_j\right)\right). \end{cases}$$

First we consider a pointwise estimator $\hat{b} \in (\mathbb{R}^m)^p$, which minimizes

$$(mn)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \rho_\tau \left(y_i\left(\mathcal{T}_j\right) - \boldsymbol{x}_i^T b_j\right).$$

By Newey and Powell (1987) Theorem 3,

$$(mp)^{-1} \sum_{j=1}^{m} \sum_{k=1}^{p} \left(b_j^k - \boldsymbol{\beta}_0^k(\mathcal{T}_j)\right)^2 = O_p(\frac{1}{n}).$$

Let $\tilde{b} \in \mathcal{L}_2^p$ be the linear interpolation of $\hat{b}$, which is defined as:

$$\tilde{b}^k(t) = \begin{cases} b_1^k, & 0 \leq t \leq \mathcal{T}_1, \\ b_j^k \frac{\mathcal{T}_{j+1}-t}{\mathcal{T}_{j+1}-\mathcal{T}_j} + b_{j+1}^k \frac{t-\mathcal{T}_j}{\mathcal{T}_{j+1}-\mathcal{T}_j}, & \mathcal{T}_j \leq t \leq \mathcal{T}_{j+1}, \\ b_m^k, & \mathcal{T}_m \leq t \leq 1. \end{cases}$$

Consider the solution of the following optimization problem:

$$\text{minimize } \|f\|_{\mathcal{H}}$$

$$\text{subject to } f\left(x\right) = g\left(x\right) \text{ for all } x \in \mathbb{S}.$$

It is well known Green and Silverman (1993) that the solution of this optimization problem can be characterized as $T^r(g)$, where T^r is a bounded linear map from L_2 to $\mathcal{H}$ satisfying $\|T^r(f)-f\|_2^2 \lesssim \left(\max_{0\leq j\leq m} |\mathcal{T}_{j+1} - \mathcal{T}_j|^{2r}\right) \|f\|_{\mathcal{H}}^2$. We denote $T^r(\boldsymbol{\beta}) = T\left((\boldsymbol{\beta}^1, \cdots, \boldsymbol{\beta}^p)\right) = \left((T^r(\boldsymbol{\beta}^1), \cdots, T^r(\boldsymbol{\beta}^p))\right)$, it's easy to verify that the boundness still holds.

The boundness leads to $\|T^r(\boldsymbol{\beta}_0) - \boldsymbol{\beta}_0\|_2^2 \lesssim m^{-2r}$.

Define $\boldsymbol{\beta}^{mn} = T^r(\tilde{b})$ and $\boldsymbol{\beta}^{\infty} = T^r(\boldsymbol{\beta}_0)$.

$$\|\boldsymbol{\beta}^{mn}\|_{\mathcal{H}^p}^2 = \|T^r(\tilde{b} + \boldsymbol{\beta}_0 - \boldsymbol{\beta}_0)\|_{\mathcal{H}^p}^2 \leq \|\boldsymbol{\beta}_0\|_{\mathcal{H}^p}^2 + \|T^r(\tilde{b} - \boldsymbol{\beta}_0)\|_{\mathcal{H}^p}^2 < \infty.$$

Also, by DeVore and Lorentz (1993),

$$\|\boldsymbol{\beta}^{mn} - \boldsymbol{\beta}^{\infty}\|_2^2 \lesssim \|\tilde{b} - \boldsymbol{\beta}_0\|_2^2 \approx (mp)^{-1} \sum_{j=1}^{m} \sum_{k=1}^{p} \left(b_j^k - \boldsymbol{\beta}_0^k(\mathcal{T}_j)\right)^2 = O_p(\frac{1}{n}).$$

By the definition of Penalized expectile RKHS estimator and $\boldsymbol{\beta}_\lambda^{\infty}$,

$$\begin{aligned}
\|\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}^\infty\|_2^2 &= \|T^r(\boldsymbol{\beta}_\lambda^{mn}) - T^r(\boldsymbol{\beta}^\infty)\|_2^2 \\
&\lesssim \|\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}^\infty\|_s^2 \\
&\lesssim (mn)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_i\left(\mathcal{T}_j\right) - \boldsymbol{x}_i^T\boldsymbol{\beta}_\lambda^{mn}\right)^2 \\
&\approx (mn)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\rho_\tau\left(y_i\left(\mathcal{T}_j\right) - \boldsymbol{x}_i^T\boldsymbol{\beta}_\lambda^{mn}\right) \\
&\leq v_n(\boldsymbol{\beta}_\lambda^{mn}) + \lambda\|\boldsymbol{\beta}_\lambda^{mn}\|_{\mathcal{H}^p}^2 \\
&\leq v_n(\boldsymbol{\beta}^\infty) + \lambda\|\boldsymbol{\beta}^\infty\|_{\mathcal{H}^p}^2 \\
&= O_p(\lambda + n^{-1})
\end{aligned}$$

Therefore,

$$\|\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}_0\|_2^2 \leq \|\boldsymbol{\beta}_\lambda^{mn} - \boldsymbol{\beta}^\infty\|_2^2 + \|\boldsymbol{\beta}^\infty - \boldsymbol{\beta}_0\|_2^2 = O_p(m^{-2r} + n^{-1} + \lambda).$$

As long as $\lambda \approx m^{-2r} + n^{-1}$, the lower bound is reached. □

Bibliography

DeVore, R. A. and G. G. Lorentz (1993). Constructive approximation, Volume 303. Springer Science & Business Media.

Green, P. J. and B. W. Silverman (1993). Nonparametric regression and generalized linear models: a roughness penalty approach. Crc Press.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing. Econometrica: Journal of the Econometric Society, 819–847.