Supplementary Material for
Partially Linear Functional Additive Models for Multivariate Functional Data

Raymond K. W. Wong[1], Yehua Li[2] and Zhengyuan Zhu[2]

[1]Department of Statistics, Texas A&M University, College Station, TX 77843

[2]Department of Statistics & Statistical Laboratory, Iowa State University, Ames, IA 50011

The supplementary material is organized as follows. We provide a proof for Proposition 1 in Section A, theory for PLFAM (including proofs of Theorems 1 and 2) in Section B, additional simulation and data analysis results in Sections C and D, and the bootstrap procedure for standard error estimation in Section E.

A   Theory for mFPCA

Proof of Proposition 1 We use C as a generic notation for positive constant. For two sequences $\{a_n\}$ and $\{b_n\}$, we use $a_n \lesssim b_n$ to denote that a_n is bounded by b_n omitting some negligible terms. Recall that $\Delta = n^{1/2}(\widehat{\mathcal{C}} - \mathcal{C})$, and under Assumption 2 we have $\mathbb{E}\|\Delta\|_{\text{op}}^2 < \infty$.

Asymptotic expansions for the empirical eigenfunctions and eigenvalues similar to (2.8) and (2.9) in Hall and Hosseini-Nasab (2006) also hold for multivariate FPCA. For any k such that $\delta_k > n^{-1/2}\|\Delta\|_{\text{op}}$,

$$\widehat{\lambda}_k - \lambda_k = n^{-1/2}\langle\Delta\boldsymbol{\psi}_k, \boldsymbol{\psi}_k\rangle_{\mathbb{X}} + \Lambda_{nk} \times \{1 + o_p(1)\},$$

$$\widehat{\boldsymbol{\psi}}_k(t) - \boldsymbol{\psi}_k(t) = \left\{n^{-1/2}\sum_{j\neq k}(\lambda_k - \lambda_j)^{-1}\boldsymbol{\psi}_j\langle\Delta\boldsymbol{\psi}_k, \boldsymbol{\psi}_k\rangle_{\mathbb{X}}\right\} \times \{1 + \mathcal{O}_p(n^{-1/2}\delta_k^{-1})\}, \text{ (S.1)}$$

where $\Lambda_{nk} = n^{-1}\sum_{j\neq k}(\lambda_k - \lambda_j)^{-1}(\langle\Delta\psi_j, \psi_k\rangle_{\mathbb{X}})^2 = n^{-1}\sum_{j\neq k}(\lambda_k - \lambda_j)^{-1}(n^{-1/2}\sum_{i=1}^n \xi_{ij}\xi_{ik})^2$.

It is easy to see that $\mathbb{E}|\Lambda_{nk}| \leq (n\delta_k)^{-1}\sum_{j\neq k}\lambda_j\lambda_k \leq C(n\delta_k)^{-1}\lambda_k$ for all k.

Since $\widehat{\xi}_{ik} = \langle\boldsymbol{x}_i, \boldsymbol{\psi}_k\rangle_{\mathbb{X}}$, by the expansion (S.1),

$$\widehat{\xi}_{ik} - \xi_{ik} = A_{ik} \times \{1 + o_p(1)\} \quad \text{for all } k \leq J_n,$$

where $A_{ik} = n^{-1/2}\sum_{j\neq k}(\lambda_k - \lambda_j)^{-1}\xi_{ij}\langle\Delta\psi_k, \psi_j\rangle_{\mathbb{X}} = \sum_{j\neq k}(\lambda_k - \lambda_j)^{-1}\xi_{ij}(\frac{1}{n}\sum_{i_1=1}^n \xi_{i_1k}\xi_{i_1j})$.

Next, we calculate the order of A_{ik}. Denote $[x]$ as the integer part of x. By Assumption 1, $\lambda_j - \lambda_{j+1} \geq C_\lambda^{-1}j^{-\alpha-1}$,

$$\lambda_j - \lambda_k \geq C_\lambda^{-1}\sum_{l=j}^{k-1}l^{-\alpha-1} \geq C_\lambda^{-1}\int_j^k x^{-\alpha-1}dx \geq \frac{1}{C_\lambda\alpha}(j^{-\alpha} - k^{-\alpha}) \qquad \text{for } j < k;$$

$$\lambda_k - \lambda_j \geq C_\lambda^{-1}\sum_{l=k}^{j-1}l^{-\alpha-1} \geq C_\lambda^{-1}\int_k^j x^{-\alpha-1}dx \geq \frac{1}{C_\lambda\alpha}(k^{-\alpha} - j^{-\alpha}) \qquad \text{for } j > k. \text{ (S.2)}$$

By Assumption 2 $\mathbb{E}(\frac{1}{n}\sum_{i_1=1}^n \xi_{i_1 k}\xi_{i_1 j})^2 \leq C\lambda_k\lambda_j/n$ for all k and j, and by (S.2)

$$
\begin{aligned}
\mathbb{E}(A_{ik}^2) &\lesssim \frac{C}{n}\sum_{j\neq k}(\lambda_k-\lambda_j)^{-2}\lambda_k\lambda_j^2 \\
&= \frac{C}{n}(\sum_{j<k}+\sum_{j>k})(\lambda_k-\lambda_j)^{-2}\lambda_k\lambda_j^2 \\
&\leq \frac{C\lambda_k}{n}\Bigg\{ \sum_{j=1}^{[(1-a)k]}\left(\frac{C_\lambda^2\alpha j^{-\alpha}}{j^{-\alpha}-k^{-\alpha}}\right)^2 + \Bigg(\sum_{j=[(1-a)k]+1}^{k-1}+\sum_{j=k+1}^{[(1+b)k]}\Bigg)\frac{C_\lambda^2 j^{-2\alpha}}{C_\lambda^{-2}k^{-2\alpha-2}} \\
&\qquad + \sum_{j=[(1+b)k]+1}^{\infty}\left(\frac{C_\lambda^2\alpha j^{-\alpha}}{k^{-\alpha}-j^{-\alpha}}\right)^2 \Bigg\} \qquad \text{(for some } a,b\in(0,1)\,) \\
&\lesssim \frac{C\lambda_k}{n}\Bigg\{ \sum_{j=1}^{[(1-a)k]}\left(\frac{1}{1-(j/k)^\alpha}\right)^2 + \sum_{j=[(1+b)k]+1}^{\infty}\left(\frac{1}{(j/k)^\alpha-1}\right)^2 + [(a+b)k]k^2 \Bigg\} \\
&\lesssim \frac{Ck\lambda_k}{n}\Bigg\{ \int_0^{1-a}(1-x^\alpha)^{-2}dx + \int_{(1+b)}^{\infty}(x^\alpha-1)^{-2}dx \Bigg\} + C(a+b)k^{3-\alpha}/n \\
&\lesssim \frac{Ck^{1-\alpha}}{n}\Bigg\{ \int_0^{(1-a)}(1-y)^{-2}dy + \int_{(1+b)}^{\infty}(y-1)^{-2}dy \Bigg\} + C(a+b)k^{3-\alpha}/n \\
&\lesssim \frac{Ck^{1-\alpha}}{n}\{a^{-1}-1+b^{-1}+(a+b)k^2\}.
\end{aligned}
$$

We select $a\sim k^{-1}$ and $b\sim k^{-1}$, we get $\mathbb{E}A_{ik}^2\leq Ck^{2-\alpha}/n$ for all k. This implies $\widehat{\xi}_{ik}-\xi_{ik}=\mathcal{O}_p(n^{-1/2}k^{1-\alpha/2})$ uniformly for $k\leq J_n$.

On the other hand, by (S.1) we can show

$$
\mathbb{E}\left|\widehat{\lambda}_k-\lambda_k-n^{-1/2}\langle\Delta\boldsymbol{\psi}_k,\boldsymbol{\psi}_k\rangle\right| \lesssim \mathbb{E}|\Lambda_{nk}|\leq Cn^{-1}\lambda_k\delta_k^{-1},
$$

$$
\mathbb{E}(n^{-1/2}\langle\Delta\boldsymbol{\psi}_k,\boldsymbol{\psi}_k\rangle)^2 = \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n(\xi_{ik}^2-\lambda_k)\right\}^2 \leq C\lambda_k^2 n^{-1}.
$$

This also means $\widehat{\lambda}_k-\lambda_k=\mathcal{O}_p(n^{-1/2}\lambda_k)$ uniformly for all $k\leq J_n$. Since $\Phi(\cdot)$ is differentiable

transformation function, using the delta method

$$
\begin{aligned}
\widehat{\zeta}_{ik} &\approx \Phi[(\xi_{ik} + A_{ik})\{\lambda_k^{-1/2} - (1/2)\lambda_k^{-3/2}(\widehat{\lambda}_k - \lambda_k)\}] \\
&\approx \zeta_{ik} + \Phi'(\xi_{ik}\lambda_{ik}^{-1/2})\{\lambda_k^{-1/2}A_{ik} - \frac{1}{2}\xi_{ik}\lambda_k^{-3/2}(\widehat{\lambda}_k - \lambda_k)\} \qquad \text{(S.3)} \\
&= \zeta_{ik} + \mathcal{O}_p(n^{-1/2}k).
\end{aligned}
$$

By the assumption that $|\Phi'(x)| < C$ for all x and the mean-value theorem, one can verify that $\mathbb{E}(\widehat{\zeta}_{ik} - \zeta_{ik})^2 \leq Cn^{-1}k^2$ uniformly for all $k \leq J_n$.

B Theory for PLFAM

Throughout the theoretical development, we utilize the following representation of a generic function $m \in \mathbb{M}$:

$$
m(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\boldsymbol{\nu} + h(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\boldsymbol{\nu} + \sum_{k=1}^{s} h_k(\boldsymbol{u}, \zeta_k),
$$

where $h_k \in \mathbb{H}_k = \{h_k \in \mathbb{I} \oplus \bar{\mathbb{F}}_k : \sum_{i=1}^{n} h_k(\boldsymbol{u}_i, \widehat{\zeta}_{ik})u_{ij} = 0, j = 1\ldots, p+1\}$ for $k = 1, \ldots, s$. Note that the set $\mathbb{H}_k$ depends on $\{\boldsymbol{u}_i\}$ and $\{\widehat{\boldsymbol{\zeta}}_i\}$ and thus is a random set with randomness inherited from them. Write $\boldsymbol{U} = [u_{ij}]_{i=1,\ldots,n,j=1,\ldots,p+1}$. Given $m(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\boldsymbol{\theta} + \sum_{k=1}^{s} f_k(\zeta_k)$, where $f_k \in \bar{\mathbb{F}}_k$, one can transform it into the aforementioned representation by setting $\boldsymbol{\nu} = \boldsymbol{\theta} - \sum_{k=1}^{s} \boldsymbol{\omega}_k$ and $h_k(\boldsymbol{u}, \zeta_k) = \boldsymbol{u}^{\mathsf{T}}\boldsymbol{\omega}_k + f_k(\zeta_k)$, where $\boldsymbol{\omega}_k$ fulfills

$$
\frac{1}{n}\boldsymbol{U}^{\mathsf{T}}\boldsymbol{U}\boldsymbol{\omega}_k = -\frac{1}{n}\boldsymbol{U}^{\mathsf{T}}(f_k(\widehat{\zeta}_{k1}), \ldots, f_k(\widehat{\zeta}_{kn}))^{\mathsf{T}}.
$$

Similarly, $m_0(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\boldsymbol{\nu}_0 + h_0(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\boldsymbol{\nu}_0 + \sum_{k=1}^{s} h_{0k}(\boldsymbol{u}, \zeta_k)$ and $\widehat{m}(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\widehat{\boldsymbol{\nu}} + \widehat{h}(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}\widehat{\boldsymbol{\nu}} + \sum_{k=1}^{s} \widehat{h}_k(\boldsymbol{u}, \zeta_k)$. Moreover, write $\mathbb{H} = \sum_{k=1}^{s} \mathbb{H}_k$.

Similar to P_n, we write $P_{n,*}$ as the empirical distributions of $(\boldsymbol{Z}, \widehat{\boldsymbol{\zeta}})$. That is, $P_{n,*} =$

$\sum_{i=1}^{n} \delta_{\boldsymbol{z}_i, \widehat{\boldsymbol{\zeta}}_i}/n$. Moreover, we define the corresponding version of (squared) empirical norm and inner product as

$$\|m_1\|_{n,*}^2 = \int m_1^2 dP_{n,*} \quad \text{and} \quad (m_1, m_2)_{n,*} = \int m_1 m_2 dP_{n,*}, \qquad \text{for any } m_1, m_2 \in \mathbb{M}.$$

First, we prove the following proposition about the convergence with respect to the empirical norm $\|\cdot\|_{n,*}$ rather than the intended $\|\cdot\|_n$.

Proposition 2 *Suppose* $s = \mathcal{O}_p(n^{1/\{2(1+\alpha)\}})$ *and* $\mathbb{E}(\widehat{\zeta}_{ik} - \zeta_{ik})^2 \leq Cn^{-1}k^{2\beta}$ *uniformly for all* $k \leq s$. *Further, assume* $J(m_0) < \infty$ *and* $\boldsymbol{\Sigma}$ *is non-singular. If* $\tau_n^{-1} = \mathcal{O}_p(\min\{n^{2/5}s^{-6/5}, n^{1/2}s^{-(\frac{1}{2}+\beta)}\})$, *we have* $\|\widehat{m} - m_0\|_{n,*} = \mathcal{O}_p(\tau_n)$ *and* $J(\widehat{m}) = \mathcal{O}_p(1)$. *If* $J(m_0) = 0$ *and* $\tau_n \asymp n^{-1/4}s^3$, $\|\widehat{m} - m_0\|_{n,*} = \mathcal{O}_p(n^{-1/2})$ *and* $J(\widehat{m}) = \mathcal{O}_p(n^{-1/2}s^{-6})$.

The proof of Proposition 2 is given in Section B.1. By Taylor expansion arguments and convergence of $\widehat{\boldsymbol{\zeta}}$, the convergence results based on $\|\cdot\|_n$ (Theorem 1) is implied by those based on $\|\cdot\|_{n,*}$ (Proposition 2). See Section B.2 for the proof of Theorem 1. With convergence of $\widehat{m}$, we study the parametric part in details and obtain the optimal $\sqrt{n}$-consistency for $\widehat{\boldsymbol{\gamma}}$. The details is shown in Section B.3.

For ease of reading, we collect all other lemmas that are used throughout the subsequent proofs here. Their proofs are deferred to Section B.4.

Lemma 2 *For any* $f(\boldsymbol{\zeta}) = \sum_{k=1}^{s} f_k(\zeta_k) \in \sum_{k=1}^{s} \mathbb{F}_k$, *there exists* C_2 *(independent of s) such that*

$$\max_{1 \leq k \leq s} \sup_{\zeta_k \in [0,1]} \left|\frac{\partial f_k(\zeta_k)}{\partial \zeta_k}\right| \bigg/ \|f_k\| \leq C_2. \tag{S.4}$$

Lemma 3 (Entropy result) *Assume* $\boldsymbol{\Sigma}$ *is non-singular. Then there exists constants* C_1 *and* C_1' *such that the events*

$$\liminf_n \left\{\sup_{\delta > 0} \delta^{1/2} H_\infty(\delta, \{h_k \in \mathbb{H}_k : J(h_k) \leq 1\}) \leq C_1\right\},$$

S.5

$$\liminf_n \left\{ \sup_{\delta > 0} \delta^{1/2} H_\infty(\delta, \{h \in \mathbb{H} : J(h) \le 1\}) \le C_1 s^{3/2} \right\}$$

and

$$\liminf_n \left\{ \sup_{h \in \mathbb{H}: J(h) \le 1} |h|_\infty \le C_1' s \right\}$$

are of probability 1.

Lemma 4 *Assume $\boldsymbol{\Sigma}$ is non-singular. We have*

$$\sup_{h \in \mathbb{H}} \frac{|(\varepsilon, h - h_0)_{n,*}|}{\|h - h_0\|_{n,*}^{3/4} \{J(h) + J(h_0)\}^{1/4}} = \mathcal{O}_p(n^{-1/2} s^{3/2}),$$

where $m_0(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^\mathsf{T} \boldsymbol{\nu}_0 + h_0(\boldsymbol{u}, \boldsymbol{\zeta})$ with $\boldsymbol{\nu}_0 \in \mathbb{R}^{p+1}$ and $h_0 \in \mathbb{H}$.

B.1 Proof of Proposition 2

Proof of Proposition 2

Expanding the objective function, we have

$$
\begin{aligned}
\ell(m) &= \frac{1}{n} \sum_{i=1}^n \{y_i - m(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(m) \\
&= \frac{1}{n} \sum_{i=1}^n \{\boldsymbol{u}_i^\mathsf{T} \boldsymbol{\nu}_0 + h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i - \boldsymbol{u}_i^\mathsf{T} \boldsymbol{\nu} - h(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(h) \\
&= \frac{1}{n} \sum_{i=1}^n \{\boldsymbol{u}_i^\mathsf{T}(\boldsymbol{\nu}_0 - \boldsymbol{\nu})\}^2 + \frac{2}{n} \sum_{i=1}^n \{\boldsymbol{u}_i^\mathsf{T}(\boldsymbol{\nu}_0 - \boldsymbol{\nu})\}\{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i\} \\
&\quad + \frac{1}{n} \sum_{i=1}^n \{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i - h(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(h).
\end{aligned}
$$

Minimizing ℓ is equivalent to the following two minimizations:

$$\widehat{\boldsymbol{\nu}} = \arg\min_{\boldsymbol{\nu}\in\mathbb{R}^{p+1}} \left\{ \frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{u}_i^{\mathsf{T}}(\boldsymbol{\nu}_0 - \boldsymbol{\nu})\}^2 + \frac{2}{n}\sum_{i=1}^{n}\{\boldsymbol{u}_i^{\mathsf{T}}(\boldsymbol{\nu}_0 - \boldsymbol{\nu})\}\{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i\} \right\},$$

$$\widehat{h} = \arg\min_{h\in\mathbb{H}} \left\{ \frac{1}{n}\sum_{i=1}^{n}\{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i - h(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(h) \right\}.$$

The first one leads to

$$\frac{1}{n}\boldsymbol{U}^{\mathsf{T}}\boldsymbol{U}(\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0) = \frac{1}{n}\boldsymbol{U}^{\mathsf{T}}(\boldsymbol{h}_0 + \boldsymbol{\varepsilon}),$$

where $\boldsymbol{U} = [u_{ij}]_{i=1,\dots,n,j=1,\dots,p+1}$, $\boldsymbol{h}_0 = (h_0(\boldsymbol{u}_1, \boldsymbol{\zeta}_1), \dots, h_0(\boldsymbol{u}_n, \boldsymbol{\zeta}_n))^{\mathsf{T}}$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^{\mathsf{T}}$. By Taylor expansion of h_0 with respect to $\boldsymbol{\zeta}$ at $\widehat{\boldsymbol{\zeta}}_i$ and the fact that $D_{\boldsymbol{\zeta}}h_0 = D_{\boldsymbol{\zeta}}f_0$,

$$\frac{1}{n}\sum_{i=1}^{n}u_{ij}h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) = \frac{1}{n}\sum_{i=1}^{n}u_{ij}D_{\boldsymbol{\zeta}}f_0(\boldsymbol{\zeta}_i^*)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i) = J(f_0)\,\mathcal{O}_p(n^{-1/2}s^{\frac{1}{2}+\beta}) \tag{S.5}$$

where $\boldsymbol{\zeta}_i^*$ lies on the line segment joining $\boldsymbol{\zeta}_i$ and $\widehat{\boldsymbol{\zeta}}_i$; and the last equality follows from the assumption $\mathbb{E}(\widehat{\zeta}_{ik} - \zeta_{ik})^2 \le Cn^{-1}k^{2\beta}$, Lemma 2 and the following calculation

$$
\begin{aligned}
|D_{\boldsymbol{\zeta}}f_0(\boldsymbol{\zeta}_i^*)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)| &= \left| \sum_{k=1}^{s} \frac{\partial}{\partial \zeta_k} f_{0k}(\zeta_{ik}^*)(\widehat{\zeta}_{ik} - \zeta_{ik}) \right| \\
&\le \left\{ \sum_{k=1}^{s} \left| \frac{\partial}{\partial \zeta_k} f_{0k}(\zeta_{ik}^*) \right|^2 \right\}^{1/2} \left\{ \sum_{k=1}^{s} (\widehat{\zeta}_{ik} - \zeta_{ik})^2 \right\}^{1/2} \\
&\le \left(\sum_{k=1}^{s} \|f_{0k}\|^2 \right)^{1/2} \times \left\{ \mathcal{O}_p\left(\sum_{k=1}^{s} n^{-1}k^{2\beta} \right) \right\}^{1/2} \\
&= \|f_0\| \times \mathcal{O}_p(n^{-1/2}s^{\beta+\frac{1}{2}}).
\end{aligned}
$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^{n}u_{ij}\varepsilon_i = \mathcal{O}_p(n^{-1/2}).$$

Since $\boldsymbol{U}^{\mathsf{T}}\boldsymbol{U}/n \to \boldsymbol{\Sigma}$ almost surely (element-wisely) and $\boldsymbol{\Sigma}$ is non-singular, we have $\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E = \mathcal{O}_p(n^{-1/2}s^{\frac{1}{2}+\beta})$. Note that if $J(f_0) = 0$, we have $\boldsymbol{U}^{\mathsf{T}}\boldsymbol{h}_0 = \boldsymbol{0}$ and $\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E =$

$\mathcal{O}_p(n^{-1/2})$.

In sequel, we focus on the second optimization. Since $\widehat{h}$ is the minimizer,

$$\frac{1}{n}\sum_{i=1}^{n}\{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i - \widehat{h}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(\widehat{h}) \leq \frac{1}{n}\sum_{i=1}^{n}\{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) + \varepsilon_i - h_0(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 + \tau_n^2 J(h_0),$$

which leads to

$$\|h_0 - \widehat{h}\|_{n,*}^2 + \tfrac{2}{n}\sum_{i=1}^{n}\{h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) - h_0(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}\{h_0(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i) - \widehat{h}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\} + \tau_n^2 J(\widehat{h})$$

$$\leq (\varepsilon, \widehat{h} - h_0)_{n,*} + \tau_n^2 J(h_0). \tag{S.6}$$

Now, we utilize the previous Taylor expansions: For $i = 1, \ldots, n$,

$$h_0(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) = h_0(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i) + D_{\boldsymbol{\zeta}} f_0(\boldsymbol{\zeta}_i^*)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i).$$

Thus (B.1) becomes

$$\|\widehat{h} - h_0\|_{n,*}^2 + \frac{2}{n}\sum_{i=1}^{n}\{\widehat{h}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i) - h_0(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}\{D_{\boldsymbol{\zeta}} f_0(\boldsymbol{\zeta}_i^*)(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)\} + \tau_n^2 J(\widehat{h}) \leq 2(\varepsilon, \widehat{h} - h_0)_{n,*} + \tau_n^2 J(h_0). \tag{S.7}$$

Now we derive asymptotic order of the following two terms:

$$\left| \frac{2}{n}\sum_{i=1}^{n}\{\widehat{h}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i) - h_0(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}\{D_{\boldsymbol{\zeta}} f_0(\boldsymbol{\zeta}_i^*)(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_i)\} \right|$$

$$\leq 2\|\widehat{h} - h_0\|_{n,*}\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{k=1}^{s}\frac{\partial f_{0k}(\zeta_{ik}^*)}{\partial \zeta_{ik}}(\widehat{\zeta}_{ik} - \zeta_{ik})\right\}^2\right)^{1/2}$$

$$\leq J(h_0)\|\widehat{h} - h_0\|_{n,*}\,\mathcal{O}_p(n^{-1/2}s^{\frac{1}{2}+\beta});$$

S.8

and by Lemma 4,

$$2(\varepsilon, \widehat{h} - h_0)_{n,*} = \mathcal{O}_p(n^{-1/2}s^{3/2})\|\widehat{h} - h_0\|_{n,*}^{3/4}\{J(\widehat{h}) + J(h_0)\}^{1/4}.$$

Collecting the above results, (S.7) leads to

$$\|\widehat{h} - h_0\|_{n,*}^2 + \tau_n^2 J(\widehat{h}) \le \mathcal{O}_p(n^{-1/2}s^{3/2})\|\widehat{h} - h_0\|_{n,*}^{3/4}\{J(\widehat{h}) + J(h_0)\}^{1/4} + \tau_n^2 J(h_0)$$
$$+ J(h_0)\|\widehat{h} - h_0\|_{n,*}\mathcal{O}_p(n^{-1/2}s^{\frac{1}{2}+\beta}).$$

Next, we investigate the following three scenarios where one particular term on the right hand side dominates the other two.

(A) The term $\mathcal{O}_p(n^{-1/2}s^{3/2})\|\widehat{h} - h_0\|_{n,*}^{3/4}\{J(\widehat{h}) + J(h_0)\}^{1/4}$ is the largest: Thus

$$\|\widehat{h} - h_0\|_{n,*}^2 + \tau_n^2 J(\widehat{h}) \le \mathcal{O}_p(n^{-1/2}s^{3/2})\|\widehat{h} - h_0\|_{n,*}^{3/4}\{J(\widehat{h}) + J(h_0)\}^{1/4}.$$

If $J(\widehat{h}) \ge J(h_0)$, one can deduce that $\|\widehat{h} - h_0\|_{n,*} = \mathcal{O}_p(n^{-2/3}s^2)\tau_n^{-2/3}$ and $J(\widehat{h}) = \mathcal{O}_p(n^{-4/3}s^4)\tau_n^{-10/3}$. As for $J(\widehat{h}) < J(h_0)$, we have $\|\widehat{h} - h_0\|_{n,*} = \mathcal{O}_p(n^{-2/5}s^{6/5})J(h_0)^{1/5}$ and $J(\widehat{h}) = \mathcal{O}_p(1)J(h_0)$.

(B) The term $\tau_n^2 J(h_0)$ is the largest: Thus

$$\|\widehat{h} - h_0\|_{n,*}^2 + \tau_n^2 J(\widehat{h}) \le \tau_n^2 J(h_0)\,\mathcal{O}_p(1),$$

which leads to $\|\widehat{h} - h_0\|_{n,*} = \mathcal{O}_p(\tau_n)J^{1/2}(h_0)$ and $J(\widehat{h}) = \mathcal{O}_p(1)J(h_0)$.

(C) The term $J(h_0)\|\widehat{h} - h_0\|_{n,*}\mathcal{O}_p(n^{-1/2}s^{\frac{1}{2}+\beta})$ is the largest: Thus

$$\|\widehat{h} - h_0\|_{n,*}^2 + \tau_n^2 J(\widehat{h}) \le J(h_0)\|\widehat{h} - h_0\|_{n,*}\mathcal{O}_p(n^{-1/2}s^{\frac{1}{2}+\beta}),$$

which leads to

$$
\begin{cases}
\|\widehat{h} - h_0\|_{n,*} \leq J(h_0)\, \mathcal{O}_p(n^{-1/2} s^{\frac{1}{2}+\beta}), \\[2mm]
\tau_n^2 J(\widehat{h}) \leq \|\widehat{h} - h_0\|_{n,*}\, \mathcal{O}_p(n^{-1/2} s^{\frac{1}{2}+\beta}).
\end{cases}
$$

Thus $\|\widehat{h} - h_0\|_{n,*} = \mathcal{O}_p(n^{-1/2} s^{\frac{1}{2}+\beta}) J(h_0)$ and $J(\widehat{h}) = \mathcal{O}_p(n^{-1} s^{(1+2\beta)}) \tau_n^{-2} J^2(h_0)$.

By carefully comparing the stochastic orders of terms arising from the above three cases, if $\tau_n^{-1} = \mathcal{O}_p(\min\{n^{2/5} s^{-6/5}, n^{1/2} s^{-(\frac{1}{2}+\beta)}\})$, we have $\|\widehat{h} - h_0\|_{n,*} = \mathcal{O}_p(\tau_n)$ and $J(\widehat{h}) = \mathcal{O}_p(1)$. If $J(h_0) = 0$ and $\tau_n \asymp n^{-1/4} s^3$, $\|\widehat{h} - h_0\|_{n,*} = \mathcal{O}_p(n^{-1/2})$ and $J(\widehat{h}) = \mathcal{O}_p(n^{-1/2} s^{-6})$.

B.2 Proof of Theorem 1

Proof of Theorem 1. Let $q = \widehat{m} - m_0$. By Taylor expansion,

$$
\begin{aligned}
\|q\|_n^2 &= \frac{1}{n} \sum_{i=1}^n \{q(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i) + D_{\boldsymbol{\zeta}} q(\boldsymbol{u}_i, \widetilde{\boldsymbol{\zeta}}_i)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)\}^2 \\
&= \|q\|_{n,*}^2 + \frac{1}{n} \sum_{i=1}^n \{D_{\boldsymbol{\zeta}} q(\boldsymbol{u}_i, \widetilde{\boldsymbol{\zeta}}_i)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)\}^2 + \frac{1}{n} \sum_{i=1}^n 2q(\widehat{\boldsymbol{\zeta}}_i)\{D_{\boldsymbol{\zeta}} q(\boldsymbol{u}_i, \widetilde{\boldsymbol{\zeta}}_i)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)\}
\end{aligned}
$$

where $\widetilde{\boldsymbol{\zeta}}_i$ lies in the line segment joining $\boldsymbol{\zeta}_i$ and $\widehat{\boldsymbol{\zeta}}_i$. By calculation similar to (S.5), we have

$$
\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \{D_{\boldsymbol{\zeta}} q(\boldsymbol{u}_i, \widetilde{\boldsymbol{\zeta}}_i)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)\}^2 &= J(q)\, \mathcal{O}_p(n^{-1} s^{(1+2\beta)}), \\
\frac{1}{n} \sum_{i=1}^n 2q(\widehat{\boldsymbol{\zeta}}_i)\{Dq(\boldsymbol{u}_i, \widetilde{\boldsymbol{\zeta}}_i)(\boldsymbol{\zeta}_i - \widehat{\boldsymbol{\zeta}}_i)\} &= \|q\|_{n,*} J(q)\, \mathcal{O}_p(n^{-1/2} s^{\frac{1}{2}+\beta}).
\end{aligned}
$$

By Proposition 2, if $\tau_n^{-1} = \mathcal{O}_p(\min\{n^{2/5} s^{-6/5}, n^{1/2} s^{-(\frac{1}{2}+\beta)}\})$, $J(\widehat{m}) = \mathcal{O}_p(1)$ and

$$
\|q\|_n^2 = \|q\|_{n,*}^2 + \mathcal{O}_p(n^{-1} s^{(1+2\beta)}) + \|q\|_{n,*} \mathcal{O}_p(n^{-1/2} s^{\frac{1}{2}+\beta}) = \mathcal{O}_p(\tau_n^2).
$$

If $J(m_0) = 0$ and $\tau_n \asymp n^{-1/4} s^3$, $J(\widehat{m}) = \mathcal{O}_p(n^{-1/2} s^{-6})$ from Proposition 2. Similarly as the proof of Proposition 2, write $\widehat{m}(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}} \widehat{\boldsymbol{\nu}} + \widehat{h}(\boldsymbol{u}, \boldsymbol{\zeta})$. In its proof, we show that $\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E = \mathcal{O}_p(n^{-1/2})$ and $\|h_0\|_n = 0$ (due to $\boldsymbol{U}^{\mathsf{T}} \boldsymbol{h}_0 = \boldsymbol{0}$). By Lemma 3, we have $|\widehat{h}|_\infty =$

$J(\widehat{h})\,\mathcal{O}_p(1) = \mathcal{O}_p(n^{-1/2}s^{-6})$ since $J(\widehat{h}) = J(\widehat{m})\,\mathcal{O}_p(n^{-1/2}s^{-6})$. Since $\boldsymbol{u} \in [0,1]^{p+1}$, $\|q\|_n \leq \|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E + \|\widehat{h}\|_n = \mathcal{O}_p(n^{-1/2})$.

B.3   Proof of Theorem 2

We first introduce a few Lemmas, the proof of which is relegated to Section B.4.

Lemma 5 *Under the conditions of Theorem 2, $\|\widehat{m} - m_0\|_2 = \mathcal{O}_p(n^{-1/4})$, where $\|\cdot\|_2$ represents the $L_2(P)$-norm..*

Lemma 6 *For any $k = 1, \ldots, s$ and $g_k \in \overline{\mathbb{F}}_k$, we have*

$$\sup_{g_k \in \overline{\mathbb{F}}_k} \frac{\left| \|g_k^{(1)}\|_n^2 - \|g_k^{(1)}\|_2^2 \right|}{\|g_k\|^2} = \mathcal{O}_p(1).$$

Lemma 7 *Under the conditions of Theorem 2, $\|\widehat{f}_k' - f_{0k}'\|_n^2 = \mathcal{O}_p(1)$ for all $k = 1, \ldots, s$.*

Proof of Theorem 2.   Write $\widehat{m}(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{z}^{\intercal}\widehat{\boldsymbol{\gamma}} + \widehat{g}(\boldsymbol{\zeta})$ and $m_0(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{z}^{\intercal}\boldsymbol{\gamma}_0 + g_0(\boldsymbol{\zeta})$ where $\widehat{g}, g_0 \in \sum_{k=1}^{s} \mathbb{F}_k$ and $\boldsymbol{u} = (1, \boldsymbol{z}^{\intercal})^{\intercal}$. We also write $\widehat{g}_k = \mathcal{P}_k\widehat{g} \in \overline{\mathbb{F}}_k$ and $g_{0k} = \mathcal{P}_k g_0 \in \overline{\mathbb{F}}_k$ for $k = 1, \ldots, s$. Note that $\widehat{g}$ and $\sum_{k=1}^{s} \widehat{g}_k$ may differ by a constant. Similarly for g_0 and $\sum_{k=1}^{s} g_{0k}$.

By expanding $\|\widehat{m} - m_0\|_2^2 = \|\widetilde{\boldsymbol{w}}^{\intercal}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\|_2^2 + \|\boldsymbol{w}^{\intercal}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \widehat{g} - g_0\|_2^2$, we show that $\|\widetilde{\boldsymbol{w}}^{\intercal}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\|_2^2 = \mathcal{O}_p(n^{-1/4})$ using Lemma 5. By the condition that $\boldsymbol{M}$ is non-singular, we have

$$\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|_E = \mathcal{O}_p(n^{-1/4}) \qquad \text{and} \qquad \|\widehat{g} - g_0\|_2 = \mathcal{O}_p(n^{-1/4}). \tag{S.8}$$

Recall that $\widetilde{\boldsymbol{w}}(\boldsymbol{z}, \boldsymbol{\zeta}) = \boldsymbol{z} - \boldsymbol{w}(\boldsymbol{\zeta})$. We then define

$$\widehat{m}_{\boldsymbol{\rho}}(\boldsymbol{z}, \boldsymbol{\zeta}) = \widehat{m}(\boldsymbol{u}, \boldsymbol{\zeta}) + \boldsymbol{\rho}^{\intercal}\widetilde{\boldsymbol{w}}(\boldsymbol{z}, \boldsymbol{\zeta}) = \boldsymbol{z}^{\intercal}(\widehat{\boldsymbol{\gamma}} + \boldsymbol{\rho}) + \{\widehat{g}(\boldsymbol{\zeta}) - \boldsymbol{\rho}^{\intercal}\boldsymbol{w}(\boldsymbol{\zeta})\},$$

for $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_p)^\mathsf{T} \in \mathbb{R}^p$. Note that we assume that $w_j \in \sum_{k=1}^s \mathbb{F}_k$ and hence $\widetilde{w}_j \in \mathbb{I} + \sum_{k=1}^s \mathbb{F}_k$. Since $\widehat{m}_{\boldsymbol{\rho}} \in \mathbb{I} + \sum_{k=1}^s \mathbb{F}_k$, there exists a subgradient $\boldsymbol{c} = (c_1, \ldots, c_p)^\mathsf{T}$ of $J(\widehat{m}_{\boldsymbol{\rho}})$ with respect to $\boldsymbol{\rho}$ at $\boldsymbol{\rho} = \boldsymbol{0}$ such that

$$\frac{\partial}{\partial \boldsymbol{\rho}} \left[\frac{1}{n} \sum_{i=1}^n \{y_i - \widehat{m}_{\boldsymbol{\rho}}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2 \right] \Bigg|_{\boldsymbol{\rho}=\boldsymbol{0}} + \tau_n^2 \boldsymbol{c} = \boldsymbol{0}. \tag{S.9}$$

We first analyze the order of the subgradient $\boldsymbol{c}$. Note that $J(\widehat{m}_{\boldsymbol{\rho}}) = \sum_{k=1}^s \|\widehat{g}_k - \sum_{j=1}^p \rho_j w_{jk}\|$ where $w_{jk} = \mathcal{P}_k w_j$. Now we study two cases, $\|\widehat{g}_k\| > 0$ and $\|\widehat{g}_k\| = 0$, separately.

Suppose $\|\widehat{g}_k\| > 0$. Then $\|\widehat{g}_k - \sum_{j=1}^p \rho_j w_{jk}\|$ is differentiable at $\boldsymbol{\rho} = \boldsymbol{0}$ and its partial derivative with respect to ρ_l at $\boldsymbol{\rho} = \boldsymbol{0}$ is

$$-\frac{\int_0^1 \widehat{g}_k(t)dt \int_0^1 w_{lk}(t)dt + \int_0^1 \widehat{g}_k'(t)dt \int_0^1 w_{lk}'(t)dt + \int_0^1 \widehat{g}_k''(t)w_{lk}''(t)dt}{\|\widehat{g}_k\|},$$

for $l = 1, \ldots, p$. The numerator is less than or equal to $\|\widehat{g}_k\|\|w_{lk}\|$. Hence the absolute value of this partial derivative is smaller than or equal to $\|w_{lk}\| < \infty$ by the assumption that $J(w_l) < \infty$.

Suppose $\|\widehat{g}_k\| = 0$, which implies that $\widehat{g}_k = 0$. Then

$$\left\| \widehat{g}_k - \sum_{j=1}^p \rho_j w_{jk} \right\|^2 = \left(\int_0^1 \sum_{j=1}^p \rho_j w_{jk}(t)dt \right)^2 + \left(\int_0^1 \sum_{j=1}^p \rho_j w_{jk}'(t)dt \right)^2 + \int_0^1 \left(\sum_{j=1}^p \rho_j w_{jk}''(t) \right)^2 dt$$
$$= \boldsymbol{\rho}^\mathsf{T} \boldsymbol{N}_k \boldsymbol{\rho},$$

where $\boldsymbol{N}_k$ is a $p \times p$ matrix with (i, j)-entry being $\int w_{ik} \int w_{jk} + \int w_{ik}' \int w_{jk}' + \int w_{ik}'' w_{jk}''$. Note that $\boldsymbol{N}_k$ is positive semi-definite. Using subgradient chain rule and the subgradient

S.12

formulation of Euclidean norm, the subgradient of $\sqrt{\boldsymbol{\rho}^{\intercal}\boldsymbol{N}_k\boldsymbol{\rho}}$ with respect to $\boldsymbol{\rho}$ is

$$
\begin{cases}
\dfrac{\boldsymbol{N}_k\boldsymbol{\rho}}{\|\boldsymbol{N}_k^{1/2}\boldsymbol{\rho}\|_E}, & \text{if } \boldsymbol{N}_k^{1/2}\boldsymbol{\rho} \neq \boldsymbol{0}; \\[2ex]
\in \{\boldsymbol{N}_k^{1/2}\boldsymbol{a} : \|\boldsymbol{a}\|_E \leq 1\}, & \text{otherwise.}
\end{cases}
$$

Recall that we are interested in the case of $\boldsymbol{\rho} = \boldsymbol{0}$. For any $\boldsymbol{a} = (a_1, \ldots, a_p)^{\intercal}$ such that

$\|\boldsymbol{a}\|_E \leq 1$, $\|\boldsymbol{N}_k^{1/2}\boldsymbol{a}\|_\infty \leq \|\boldsymbol{N}_k^{1/2}\boldsymbol{a}\|_E = \|\sum_{j=1}^p a_j w_{jk}\| \leq \sum_{j=1}^p |a_j|\|w_{jk}\| \leq \sum_{j=1}^p \|w_{jk}\| < \infty$,

where $\|\cdot\|_\infty$ is the max norm of a vector. Combining results from both cases, $\|\widehat{g}_k\| > 0$ and

$\|\widehat{g}_k\| = 0$, we conclude that all entries of $\boldsymbol{c}$ are $\mathcal{O}(1)$.

Now, we go back to (S.9) and study the first term on the right hand side. For $l = 1, \ldots, p$,

$$
\frac{1}{2}\frac{\partial}{\partial \rho_l}\left[\frac{1}{n}\sum_{i=1}^n \{y_i - \widehat{m}_{\boldsymbol{\rho}}(\boldsymbol{z}_i, \widehat{\boldsymbol{\zeta}}_i)\}^2\right]\Bigg|_{\boldsymbol{\rho}=\boldsymbol{0}} = -\frac{1}{n}\sum_{i=1}^n \{y_i - \widehat{m}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}\widetilde{w}_l(\boldsymbol{z}_i, \widehat{\boldsymbol{\zeta}}_i)
$$

$$
= -\frac{1}{n}\sum_{i=1}^n \left[\{y_i - m(\boldsymbol{u}_i, \boldsymbol{\zeta}_i)\} + \{m(\boldsymbol{u}_i, \boldsymbol{\zeta}_i) - m(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\} + \{m(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i) - \widehat{m}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)\}\right]\widetilde{w}_l(\boldsymbol{z}_i, \widehat{\boldsymbol{\zeta}}_i),
$$

$$
= -(\varepsilon, \widetilde{w}_l)_n + ((\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^{\intercal}\boldsymbol{w}, \widetilde{w}_l)_n + ((\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^{\intercal}\widetilde{\boldsymbol{w}}, \widetilde{w}_l)_n + (\widehat{g} - g_0, \widetilde{w}_l)_{n,*}
$$

$$
+ \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^s \widetilde{w}_l(\boldsymbol{z}_i, \boldsymbol{\zeta}_i)f'_{0k}(\zeta_{ik})(\widehat{\zeta}_{ik} - \zeta_{ik}) + \mathcal{O}_p(n^{-1})
$$

$$
= -\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \mathrm{V} + \mathcal{O}_p(n^{-1}).
$$

By the asymptotic expansions (S.1) and (S.3),

$$
\begin{aligned}
\mathrm{V} &= \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^s \widetilde{w}_l(\boldsymbol{z}_i, \boldsymbol{\zeta}_i)f'_{0k}(\zeta_{ik})\Phi'(\zeta_{ik})\Bigg\{n^{-1/2}\sum_{j\neq k}\frac{\zeta_{ij}\lambda_j^{1/2}}{(\lambda_k - \lambda_j)\lambda_k^{1/2}}\langle\Delta\psi_k, \psi_j\rangle \\
&\qquad\qquad -\frac{1}{2}n^{-1/2}\zeta_{ik}\lambda_k^{-1}\langle\Delta\psi_k, \psi_k\rangle\Bigg\} \\
&= n^{-1/2}\sum_{k=1}^s \langle\Delta\psi_k, \boldsymbol{\varpi}_{k,l}\rangle \times \{1 + \mathcal{O}_p(n^{-1/2})\},
\end{aligned}
$$

where

$$\boldsymbol{\varpi}_{k,l} = \sum_{j\neq k} \mathbb{E}\{\widetilde{w}_l(\boldsymbol{z}_1, \boldsymbol{\zeta}_1) f'_{0k}(\zeta_{1k}) \Phi'(\zeta_{1k}) \zeta_{1j}\} \lambda_j^{1/2} \lambda_k^{-1/2} (\lambda_k - \lambda_j)^{-1} \boldsymbol{\psi}_j$$
$$- \frac{1}{2} \mathbb{E}\{\widetilde{w}_l(\boldsymbol{z}_1, \boldsymbol{\zeta}_1) f'_{0k}(\zeta_{1k} \Phi'(\zeta_{1k}) \zeta_{1k}\} \lambda_k^{-1} \boldsymbol{\psi}_k. \tag{S.10}$$

Since Δ converge weakly to a Gaussian random field, it is easy to see that $V = \mathcal{O}_p(n^{-1/2})$ and is asymptotically normal.

By (10), $\mathbb{E}\{w_j(\boldsymbol{\zeta})\widetilde{w}_l(\boldsymbol{Z}, \boldsymbol{\zeta})\} = 0$,

$$\text{II} = \sum_{j=1}^{p} (\widehat{\gamma}_j - \gamma_{0j})(w_j, \widetilde{w}_l)_n = \sum_{j=1}^{p} \mathcal{O}_p(n^{-1/2})(\widehat{\gamma}_j - \gamma_{0j}).$$

Similarly, by law of large numbers,

$$\text{III} = \sum_{j=1}^{p} (\widehat{\gamma}_j - \gamma_{0j})(\widetilde{w}_j, \widetilde{w}_l)_n = \sum_{j=1}^{p} (M_{lj} + o_p(1))(\widehat{\gamma}_j - \gamma_{0j}).$$

Similarly as before, we can show that the event $\liminf_n \{|\widehat{g} - g_0|_\infty / (1 + J(\widehat{g}) + J(g_0)) \leq C_1 + 1\}$ is of probability 1.

It is easy to see that

$$\text{IV} = (\widehat{g} - g_0, \widetilde{w}_l)_n + \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{s} \{\widehat{f}'_k(\zeta_{ik} - f'_{k0}(\zeta_{ik})\} \widetilde{w}_l(\zeta_{ik})(\widehat{\zeta}_{ik} - \zeta_{ik}) + o_p(n^{-1/2})$$
$$= (\widehat{g} - g_0, \widetilde{w}_l)_n + o_p(n^{-1/2}) \qquad (\text{ by Lemma 7}).$$

Next, we study the behavior of $\sqrt{n}(g - g_0, \widetilde{w}_l)_n$ as a function of $\|(g - g_0)\widetilde{w}_l\|_2$. We are going to apply Theorem 2.4 of Mammen and van de Geer (1997). To prepare this, we first derive

some entropy results. Let

$$\mathcal{K} = \left\{ (g - g_0)\widetilde{w}_l : J(g - g_0) \le 1, g \in \sum_{k=1}^{s} \mathbb{F}_k \right\}.$$

Since $\widetilde{w}_l \in \mathbb{M}$, write $K_6 = |\widetilde{w}_l|_\infty < \infty$. Therefore

$$H_\infty(\delta, \mathcal{K}) \le H_\infty\left(\frac{\delta}{K_6}, \widetilde{\mathcal{K}}\right) \quad \text{with } \widetilde{\mathcal{K}} = \left\{ g - g_0 : J(g - g_0) \le 1, g \in \sum_{k=1}^{s} \mathbb{F}_k \right\}.$$

For any $m \in \mathcal{M}$, we can write it in two ways:

$$m(\boldsymbol{u}, \boldsymbol{\zeta}) - m_0(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{z}^{\mathsf{T}}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + g(\boldsymbol{\zeta}) - g_0(\boldsymbol{\zeta}) = \boldsymbol{u}^{\mathsf{T}}(\boldsymbol{\nu} - \boldsymbol{\nu}_0) + h(\boldsymbol{u}, \boldsymbol{\zeta}) - h_0(\boldsymbol{u}, \boldsymbol{\zeta}). \quad \text{(S.11)}$$

Note that $J(m - m_0) = J(g - g_0) = J(h - h_0)$. If $J(m - m_0) \le 1$, we can represent $g - g_0$ and $h - h_0$ uniquely as follows:

$$g(\boldsymbol{\zeta}) - g_0(\boldsymbol{\zeta}) = \mu + \sum_{k=1}^{s} \widetilde{r}_k(\zeta_k), \quad \text{(S.12)}$$

$$h(\boldsymbol{u}, \boldsymbol{\zeta}) - h_0(\boldsymbol{u}, \boldsymbol{\zeta}) = \sum_{k=1}^{s} \widetilde{h}_k(\boldsymbol{u}, \zeta_k) \quad \text{with} \quad \widetilde{h}_k(\boldsymbol{u}, \zeta_k) = \boldsymbol{u}^{\mathsf{T}}\widetilde{\boldsymbol{\omega}}_k + \widetilde{r}_k(\zeta_k) \in \mathbb{H}_k,$$

where $\widetilde{r}_k \in \bar{\mathbb{F}}_k$ such that $\sum_{i=1}^{n} \widetilde{r}_k(\zeta_{ik}) = 0$ and $J(\widetilde{r}_k) \le 1$. Plugging them into (S.11), we show that μ is the first element of $\boldsymbol{\nu} - \boldsymbol{\nu}_0 + \sum_{k=1}^{s} \widetilde{\boldsymbol{\omega}}_k$. Write $\widehat{\mu}$ as μ in (S.12) for $\widehat{g} - g_0$. Recall that the event $\liminf\{\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E \le K_1\}$ is of probability 1. Moreover, from the proof of Lemma 3, we have the event $\liminf\{\max_{k=1,\dots,s} \|\boldsymbol{\omega}_k\|_E \le L\}$ is of probability 1. Thus $\liminf\{|\widehat{\mu}| \le K_7\}$ for some constant K_7. Thus we focus on the set

$$\bar{\mathcal{K}} = \left\{ g(\boldsymbol{\zeta}) - g_0(\boldsymbol{\zeta}) = \mu + \sum_{k=1}^{s} \widetilde{r}_k(\zeta_k) : |\mu| \le K_7, J(g - g_0) \le 1, g \in \sum_{k=1}^{s} \mathbb{F}_k \right\},$$

where, with probability 1, $\widehat{g} - g_0$ will eventually falls into. We use similar trick in (S.16)

S.15

to derive the entropy result for $\widetilde{\mathcal{K}}$ by the decomposition (S.12). It suffices to obtain bound for $H_\infty(\cdot, \{g(\boldsymbol{\zeta}) = \mu : |\mu| \le K_7\})$ and $H_\infty(\cdot, \{\sum_{k=1}^s \widetilde{r}_k : \boldsymbol{u}^\mathsf{T}\widetilde{\boldsymbol{\omega}}_k + \widetilde{r}_k(\zeta_k) \in \mathbb{H}_k, \sum_{i=1}^n \widetilde{r}_k(\zeta_{ik}) = 0, J(\widetilde{r}_k) \le 1\})$. The bound for the first entropy is from Lemma 2.5 of van de Geer (2000), while that for the second entropy is derived similarly in the proof of Lemma 3. For simplicity, we skip those details. In the end, we get the event $\liminf_n\{\sup_{\delta>0} \delta^{1/2} H_\infty(\delta, \bar{\mathcal{K}}) \le K_8\}$ is of probability 1. Combining with the above results, we obtain an entropy bound for the set

$$\widehat{\mathcal{K}} = \left\{ \frac{(g - g_0)\widetilde{w}_l}{1 + J(g) + J(g_0)} : g - g_0 = \mu + \sum_{k=1}^s \widetilde{r}_k, |\mu| \le K_7, g \in \sum_{k=1}^s \mathbb{F}_k \right\}.$$

That is, the event $\liminf_n\{\sup_{\delta>0} \delta^{1/2} H_\infty(\delta, \widehat{\mathcal{K}}) \le K_9\}$ is of probability 1.

Note that $\mathbb{E}(g - g_0, \widetilde{w}_l)_n = 0$ since $\mathbb{E}(g(\boldsymbol{\zeta})\widetilde{w}_l(\boldsymbol{z}, \boldsymbol{\zeta})) = 0$ for any $g \in \sum_{k=1}^s \mathbb{F}_k$. Applying Theorem 2.4 of Mammen and van de Geer (1997) to $\widehat{\mathcal{K}}$, we have

$$\mathrm{IV} = o_p(n^{-1/2}),$$

since $J(\widehat{g}) = \mathcal{O}_p(1)$ (Theorem 1) and $\|\widehat{g} - g_0\|_2 = o_p(1)$.

Also, it is simple to show that $\sum_{i=1}^n (y_i - \widehat{m}(\boldsymbol{u}_i, \widehat{\boldsymbol{\zeta}}_i)) \mathcal{O}_p(n^{-1/2})/n = \mathcal{O}_p(n^{-1/2})$ since $\|\widehat{m} - m_0\|_n = \mathcal{O}_p(\tau_n) = \mathcal{O}_p(n^{-1/4})$. Collecting all the above results, we have, for $l = 1, \dots, p$,

$$-(\varepsilon, \widetilde{w}_l)_n + \sum_{j=1}^p (M_{lj} + o_p(1))(\widehat{\gamma}_j - \gamma_{0j}) + n^{-1/2} \sum_{k=1}^s \langle \Delta\boldsymbol{\psi}_k, \boldsymbol{\varpi}_{k,l} \rangle + 2\tau_n^2 c_l + o_p(n^{-1/2}) = 0,$$

with $c_l = \mathcal{O}(1)$. Since $\boldsymbol{M}$ is non-singular, we have

$$n^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \boldsymbol{M}^{-1}(\boldsymbol{q}_1 + \boldsymbol{q}_2) + o_p(1), \tag{S.13}$$

where $\boldsymbol{q}_j = (q_{j1}, \ldots, q_{jp})^\intercal$, $j = 1, 2$, with $q_{1l} = n^{1/2}(\varepsilon, \widetilde{w}_l)_n$, $q_{2l} = -\sum_{k=1}^{s} \langle \Delta\boldsymbol{\psi}_k, \boldsymbol{\varpi}_{k,l} \rangle$. Put

$$\boldsymbol{V}_1 = \mathrm{cov}(\boldsymbol{q}_1) \quad \text{and} \quad \boldsymbol{V}_2 = \mathrm{cov}(\boldsymbol{q}_2), \tag{S.14}$$

by the central limit theorem $\boldsymbol{q}_1 \to Normal(\boldsymbol{0}, \boldsymbol{V}_1)$ in distribution, and since Δ converge weakly to a Gaussian random field (Dauxois et al., 1982), $\boldsymbol{q}_2 \to Normal(\boldsymbol{0}, \boldsymbol{V}_2)$ in distribution. It is easy to see that $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are asymptotically independent because ε and Δ are independent. The results of the theorem follows from (S.13).

B.4   Proofs of Lemmas

Proof of Lemma 2. For $f_k \in \mathbb{F}_k$ which is a RKHS with the reproducing kernel $R_k(\cdot, \cdot)$

$$\left| \frac{\partial f_k(\zeta_k)}{\partial \zeta_k} \right| = \left| \left\langle f_k(\cdot), \frac{\partial R_k(\zeta_k, \cdot)}{\partial \zeta_k} \right\rangle \right| \leq \|f_k\| \left\| \frac{\partial R_k(\zeta_k, \cdot)}{\partial \zeta_k} \right\|.$$

The reproducing kernel of 2nd order Sobolev Hilbert spaces are $R_k(s, t) = h_1(s)h_1(t) + h_2(s)h_2(t) - h_4(|s - t|)$ where $h_1(t) = t - 1/2$, $h_2(t) = \{h_1^2(t) - 1/12\}/2$ and $h_4(t) = \{h_1^4(t) - h_1^2(t)/2 + 7/240\}/24$. Note that

$$\frac{\partial^2 R_k(s, t)}{\partial s \partial t} = \frac{13}{12} + \left(s - \frac{1}{2} \right) \left(t - \frac{1}{2} \right) - \frac{1}{2}|s - t| + \frac{1}{2}(s - t)^2. \tag{S.15}$$

Now, for any $k \leq s$,

$$\sup_{\zeta \in [0,1]} \left\| \frac{\partial R_k(\zeta, \cdot)}{\partial \zeta} \right\|^2 = \sup_{\zeta \in [0,1]} \left\langle \frac{\partial R_k(\zeta, \cdot)}{\partial \zeta}, \frac{\partial R_k(\zeta, \cdot)}{\partial \zeta} \right\rangle = \sup_{\zeta \in [0,1]} \left. \frac{\partial^2 R_k(s, t)}{\partial s \partial t} \right|_{s=t=\zeta} \leq \frac{4}{3}.$$

Proof of Lemma 3. We will study the entropy result for $\widetilde{\mathbb{H}}_k := \{h_k \in \mathbb{H}_k : J(h_k) \leq 1\}$ first. For $h_k \in \widetilde{\mathbb{H}}_k$, we can represent it uniquely as $h_k(\boldsymbol{u}, \zeta) = \boldsymbol{u}^\intercal \boldsymbol{\omega}_k + r_k(\zeta)$, where $\sum_{i=1}^{n} r_k(\zeta_{ik}) = 0$ and $r_k \in \bar{\mathbb{F}}_k$ with $J(r_k) \leq 1$. Note that if $\mathcal{S}_1$ and $\mathcal{S}_2$ are two sets of functions, we can bound

the uniform entropy of $\mathcal{S}_1 + \mathcal{S}_2$:

$$H_\infty(\delta, \mathcal{S}_1 + \mathcal{S}_2) \leq H_\infty(\delta/2, \mathcal{S}_1) + H_\infty(\delta/2, \mathcal{S}_2). \tag{S.16}$$

Take

$$\mathcal{S}_{k,1} = \left\{ r_k : h_k(\boldsymbol{u}, \zeta) = \boldsymbol{u}^\mathsf{T}\boldsymbol{\omega}_k + r_k(\zeta), \sum_{i=1}^n r_k(\zeta_{ik}) = 0, h_k \in \widetilde{\mathbb{H}}_k \right\}$$

and

$$\mathcal{S}_{k,2} = \left\{ g(\boldsymbol{u}) = \boldsymbol{u}^\mathsf{T}\boldsymbol{\omega} : h_k(\boldsymbol{u}, \zeta) = \boldsymbol{u}^\mathsf{T}\boldsymbol{\omega}_k + r_k(\zeta), \sum_{i=1}^n r_k(\zeta_{ik}) = 0, h_k \in \widetilde{\mathbb{H}}_k \right\}.$$

Note that $\widetilde{\mathbb{H}}_k \subseteq \mathcal{S}_{k,1} + \mathcal{S}_{k,2}$ and thus $H_\infty(\delta, \widetilde{\mathbb{H}}_k) \leq H_\infty(\delta/2, \mathcal{S}_{k,1}) + H_\infty(\delta/2, \mathcal{S}_{k,2})$. By the

proof of Lemma A.1 in Lin and Zhang (2006), $|r_k|_\infty \leq 1$ and there exists a constant A such

that $H_\infty(\delta, \mathcal{S}_{k,1}) \leq A\delta^{-1/2}$ for all $\delta > 0$.

Now, it remains to obtain results about $H_\infty(\delta, S_{k,2})$. The constraints of $\mathbb{H}_k$ can be written

as

$$\frac{1}{n}\boldsymbol{U}^\mathsf{T}\boldsymbol{U}\boldsymbol{\omega}_k = -\frac{1}{n}\boldsymbol{U}^\mathsf{T}(r_k(\widehat{\zeta}_{k1}), \ldots, r_k(\widehat{\zeta}_{kn}))^\mathsf{T}$$

where $\boldsymbol{U} = [u_{ij}]_{i=1,\ldots,n,j=1,\ldots,p+1}$. Note that $\boldsymbol{U}^\mathsf{T}\boldsymbol{U}/n \to \boldsymbol{\Sigma}$ almost surely (element-wisely)

and $\boldsymbol{\Sigma}$ is non-singular. Write the smallest eigenvalue of $\boldsymbol{\Sigma}$ as σ_1. Let $\mathcal{E}_n$ be the event that

$\max_{k=1,\ldots,s} \|\boldsymbol{\omega}_k\|_E \leq L = 2\sqrt{p+1}/\sigma_1$. Combining with $|r_k|_\infty \leq 1$ and $|u_{ij}| \leq 1$, we have

$$\left\|\frac{1}{n}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^\mathsf{T}(r_k(\zeta_{k1}), \ldots, r_k(\zeta_{kn}))^\mathsf{T}\right\|_E \leq \frac{1}{\sigma_1}\left\|\frac{1}{n}\boldsymbol{U}^\mathsf{T}(r_k(\zeta_{k1}), \ldots, r_k(\zeta_{kn}))^\mathsf{T}\right\|_E \leq \frac{\sqrt{p+2}}{\sigma_1}$$

for all k. Therefore $P(\liminf_{n\to\infty} \mathcal{E}_n) = 1$. We note that this result hinges on the convergence

of $\boldsymbol{U}^\mathsf{T}\boldsymbol{U}/n$, which does not depend on s, and thus still holds even s grows with n. Next, for

any $\boldsymbol{u} \in [0,1]^{p+1}$ and $\boldsymbol{\omega}, \boldsymbol{\omega}^* \in \mathbb{R}^{p+1}$, $|\boldsymbol{u}^\mathsf{T}\boldsymbol{\omega} - \boldsymbol{u}^\mathsf{T}\boldsymbol{\omega}^*| \leq \sqrt{p+1}\|\boldsymbol{\omega} - \boldsymbol{\omega}^*\|_E$. Therefore, on $\mathcal{E}_n$,

$H_\infty(\delta, \mathcal{S}_{k,2}) \leq H(\delta/\sqrt{p+1}, \{\boldsymbol{\omega} : \|\boldsymbol{\omega}\|_E \leq L\}, \|\cdot\|_E)$. From Lemma 2.5 of van de Geer (2000),

there exists a constant B such that $H(\delta/\sqrt{p+1}, \{\boldsymbol{\omega} : \|\boldsymbol{\omega}\|_E \leq L\}, \|\cdot\|_E) \leq (p+1)\log(1+$

$4L\sqrt{p+1}/\delta) \leq B\delta^{-1/2}$. Thus $H_\infty(\delta, \{h_k \in \mathbb{H}_k : J(h_k) \leq 1\}) \leq (A+B)\sqrt{2}\delta^{-1/2} = C_1\delta^{-1/2}$

where $C_1 = sqrt2(A+B)$. As a result, on $\mathcal{E}_n$, $H_\infty(\delta, \{\delta, \{h \in \mathbb{H} : J(h) \leq 1\}\}) \leq C_1 s^{3/2}\delta^{-1/2}$

since $J(h) \leq 1$ implies $J(h_k) \leq 1$ for all k. Moreover, on $\mathcal{E}_n$, $\sup_{\{h \in \mathbb{H}: J(h)\leq 1\}} |h|_\infty < sC_1'$ due

to $|h_k| \leq C_1' := \sqrt{p+1}L + 1$ for all k.

Proof of Lemma 4. Suppose

$$H_\infty(\delta, \{h \in \mathbb{H} : J(h) \leq 1\}) \leq C_1 s^{3/2}\delta^{-1/2}, \tag{S.17}$$

for all $\delta > 0$, $n \geq 1$ and some constant $C_1 > 0$ not depending on n and s. Then,

$$H\left(\delta, \left\{\frac{h - h_0}{J(h) + J(h_0)} : h \in \mathbb{H}\right\}, \|\cdot\|_{n,*}\right)$$

has the same entropy bound (S.17). The rest follows from the proof of Lemma 8.4 in van de

Geer (2000) and Lemma 3 that (S.17) holds eventually with probability 1.

Proof of Lemma 5. By Theorem 1, we have $\|\widehat{m} - m_0\|_n = \mathcal{O}_p(n^{-1/4})$. We will show that

$\|\widehat{m} - m_0\|_n$ and $\|\widehat{m} - m_0\|_2$ have the same order.

Recall that, in the proof of Proposition 2, we write $\widehat{m}(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^\mathsf{T}\widehat{\boldsymbol{\nu}} + \widehat{h}(\boldsymbol{u}, \boldsymbol{\zeta})$ and $m(\boldsymbol{u}, \boldsymbol{\zeta}) =$

$\boldsymbol{u}^\mathsf{T}\boldsymbol{\nu} + h(\boldsymbol{u}, \boldsymbol{\zeta})$. In its proof, using strong laws of large number, we show that $\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E$

converges to zero almost surely and hence the event $\liminf_n\{\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E \leq K_1\}$ is of probability

1 for some constant K_1. Consider the set

$$\mathcal{J} = \{m - m_0 : \|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|_E \leq K_1, J(h - h_0) \leq 1, m(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^\mathsf{T}\boldsymbol{\nu} + h(\boldsymbol{u}, \boldsymbol{\zeta}) \in \mathbb{M}\}.$$

We can use the similar trick in (S.16) to derive the entropy result for $\mathcal{J}$ by decomposing a

function in $\mathcal{J}$: $m - m_0 = \boldsymbol{u}^\mathsf{T}(\boldsymbol{\nu} - \boldsymbol{\nu}_0) + h - h_0$. Next, it suffices to derive uniform entropies

$H_\infty(\cdot, \{\boldsymbol{u}^\mathsf{T}(\boldsymbol{\nu} - \boldsymbol{\nu}_0) : \|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|_E \leq K_1, \boldsymbol{\nu} \in \mathbb{R}^p\})$ and $H_\infty(\cdot, \{h - h_0 : J(h - h_0) \leq 1, h \in \mathbb{H}\})$.

S.19

The first one can be handled by Lemma 2.5 of van de Geer (2000) similary as in the proof of Lemma 3 while the second one can be handled by Lemma 3. For simplicity, we skip those details. In the end, we have $\liminf_n \{\sup_{\delta>0} \delta^{1/2} H_\infty(\delta, \mathcal{J}) \leq K_2\}$ is of probability 1 for some constant K_2. And this implies the entropy results for the set

$$\widetilde{\mathcal{J}} = \left\{ \frac{m-m_0}{1+J(m)+J(m_0)} : \|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|_E \leq K_1, m(\boldsymbol{u}, \boldsymbol{\zeta}) = \boldsymbol{u}^\mathsf{T}\boldsymbol{\nu} + h(\boldsymbol{u}, \boldsymbol{\zeta}) \in \mathbb{M} \right\}.$$

Namely, $\liminf_n\{\sup_{\delta>0} \delta^{1/2} H_\infty(\delta, \widetilde{\mathcal{J}}) \leq K_3\}$ is of probability 1 for some constant K_3.

Using Lemma 3, we can show that the event $\liminf_n\{|\widehat{h} - h_0|_\infty / (1 + J(\widehat{h}) + J(h_0)) \leq K_4\}$ is of probability 1 for some constant K_4. (Note that s is assumed to be fixed and thus is assimilated into the constant.) Combining with $\mathbb{P}(\liminf_n\{\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}_0\|_E \leq K_1\}) = 1$, we can simply focus on the set

$$\bar{\mathcal{J}} = \left\{ \frac{m-m_0}{1+J(m)+J(m_0)} : \|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|_E \leq K_1, \frac{|\widehat{h} - h_0|_\infty}{1 + J(\widehat{h}) + J(h_0)} \leq K_4, m \in \mathbb{M} \right\}, \quad \text{(S.18)}$$

where, with probability 1, $(\widehat{m} - m_0)/(1 + J(\widehat{m}) + J(m_0))$ will eventually fall into. Clearly, we also have that $\liminf_n\{\sup_{\delta>0} \delta^{1/2} H_\infty(\delta, \bar{\mathcal{J}}) \leq K_3\}$ is of probability 1. It is also easy to show that $\bar{\mathcal{J}}$ is uniformly bounded.

From Theorem 1, we have $\|\widehat{m} - m\|_n = \mathcal{O}_p(n^{-1/4})$. Hence, by applying Lemma 5.16 of van de Geer (2000) on $\bar{\mathcal{J}}$, with $\delta_n = K_5 n^{-2/5}$ for some constant K_5, we can show that $\|\widehat{m} - m_0\|_n$ and $\|\widehat{m} - m_0\|_2$ have the same order and thus $\|\widehat{m} - m_0\|_2 = \mathcal{O}_p(1)$.

Proof of Lemma 6.

Consider $\widehat{\mathbb{F}}'_k = \{f^{(1)}/\|f\| : f \in \bar{\mathbb{F}}_k\}$. By Lemma 2, we have the uniform boundedness of $\widehat{\mathbb{F}}'_k$: $\sup_{f \in \widehat{\mathbb{F}}'_k} \sup_{t \in [0,1]} |f(t)| \leq C_2$. Using Lemma 2.4 of van de Geer (2000), it is easy to show that there exists a constant C_3 such that $\sup_{\delta>0} \delta H_\infty(\delta, \widehat{\mathbb{F}}'_k) \leq C_3$. Owing to the uniform boundedness of $\widehat{\mathbb{F}}'_k$, $\sup_{\delta>0} \delta H_\infty(\delta, \{f^2 : f \in \widehat{\mathbb{F}}'_k\}) \leq 2C_2 C_3$. The desired result then follows

from Lemma 3.6 of van de Geer (2000).

Proof of Lemma 7. Put $q_k = \widehat{f}_k - f_{0k}$, then $\widehat{f} - f_0 = \sum_{j=1}^s q_j$. Since $q_j \in \bar{\mathbb{F}}_j$, $\int_0^1 q_j(t)dt = 0$, and therefore $\|\widehat{f} - f_0\|_{L^2[0,1]^s}^2 = \sum_{j=1}^s \|q_j\|_{L^2[0,1]}^2$. By (S.8), $\|\widehat{g} - g_0\|_2 = \mathcal{O}_p(1)$. By the assumption that $\boldsymbol{\zeta}$ has non-degenerate, bounded joint density on $[0,1]^s$, $\|\cdot\|_2$ and $\|\cdot\|_{L^2[0,1]^s}$ are equivalent norms, and therefore $\|q_j\|_{L^2[0,1]} = \mathcal{O}_p(1)$ for $j = 1, \ldots, s$. By Gagliardo-Nirenberg interpolation inequality (Nirenberg (1959) and Brezis (2010, pp. 313-314)), there exists a constant C_4 such that

$$\|q_k^{(1)}\|_{L^2[0,1]} \le C_4 \|q_k\|^{1/2} \|q_k\|_{L^2[0,1]}^{1/2}.$$

By Theorem 1, $J(\widehat{m}) = \mathcal{O}_p(1)$ and therefore $\|q_k\| = \mathcal{O}_p(1)$. Therefore $\|q_k^{(1)}\|_{L^2[0,1]} = \mathcal{O}_p(1)$. Again, because $\|\cdot\|_{L^2[0,1]}$ and $\|\cdot\|_2$ are equivalent norms, $\|q_k^{(1)}\|_2 = \mathcal{O}_p(1)$. Finally, by Lemma 6 and $\|q_k\| = \mathcal{O}_p(1)$, we have $\|q_k^{(1)}\|_n^2 = \|q_k^{(1)}\|_{2,k}^2 + \|q_k^{(1)}\|_n^2 - \|q_k^{(1)}\|_{2,k}^2 = \|q_k^{(1)}\|_{2,k}^2 + \|q_k\|^2 \mathcal{O}_p(1) = \mathcal{O}_p(1)$.

C   Additional results for Section 5

Following the suggestion of a referee, we also provide results when s is set to recover 90% of the total variation in $\{\boldsymbol{x}_i\}$, instead of 99.9%. The results are presented in Tables S.1-S.4, which should be compared with Tables 1-4 in the main text. When such a smaller percentage is used, the 4th component, which is related to Y, is near the cut-off point and often not included in the model. As a result, f_4 is often falsely excluded from the model (see Table S.2), and there is a much lower chance for COSSO to select the correct model. We also see much bigger prediction errors in Table S.4 than those in Table 4. Our conclusion is it is best to include as many components as possible and let the model selection mechanism of COSSO determine the size of the model.

D Additional Results for Section 6

In Figure S.1, we show 50 randomly selected trajectories for daily maximum and daily minimum temperature. Since the two functional predictors in our real data are strongly correlated, we also compare the prediction performance for models using only one functional predictor. Recall that $X_1(t)$ and $X_2(t)$ are the daily maximum and daily minimum temperature trajectories respectively. We denote by $\bar{X}(t) = \{X_1(t) + X_2(t)\}/2$ the mean trajectory. In addition to the models presented in Section 6, we also compare the yield prediction performance of the following 12 models, which use only one of $X_1(t)$, $X_2(t)$ and $\bar{X}(t)$ as the functional predictor. In the prediction experiment described in Section 6.1, the prediction errors of these 12 models are presented in Table S.5. As we can see, the models using only one functional predictor or the average yield higher prediction errors than PLFAM(joint) which jointly model both functional predictors.

1. PLFAM(max): PLFAM based on univariate FPCA scores from X_1;

2. FAM(max): FAM based on univariate FPCA scores from X_1;

3. FLM-Cov(max): FLM based on univariate FPCA scores from X_1, with covariate effects;

4. FLM(max): FLM based on univariate FPCA scores from X_1 (without $\boldsymbol{Z}$);

5. PLFAM(min): PLFAM based on univariate FPCA scores from X_2;

6. FAM(min): FAM based on univariate FPCA scores from X_2;

7. FLM-Cov(min): FLM based on univariate FPCA scores from X_2, with covariate effects;

8. FLM(min): FLM based on univariate FPCA scores from X_2 (without $\boldsymbol{Z}$);

9. PLFAM(mean): PLFAM based on univariate FPCA scores from $\bar{X}$;

10. FAM(mean): FAM based on univariate FPCA scores from $\bar{X}$;

11. FLM-Cov(mean): FLM based on univariate FPCA scores from $\bar{X}$, with covariate effects;

12. FLM(mean): FLM based on univariate FPCA scores from $\bar{X}$ (without $\boldsymbol{Z}$).

We also made the assumption that crop yields in different counties and years are conditional independent given the local meteorology information. To check for possible spatial dependency, we calculate the spatial variograms for each year based on the residuals from the fitted yield prediction model; to check for possible temporal dependency, we also calculate the autocorrelation function (ACF) for each county. Because of limited space, we show the spatial variograms for the first 4 years in Figure S.2 and ACF for the first 4 counties in Figure S.3. These plots are based on the residuals of the corn yield prediction model. Plots for other years and counties and those based on the soybean prediction model are similar. All variograms and ACF's are contained in the confidence band based on the assumption of no dependency, which supports the conditional independence assumption that we make.

E   Standard Error Estimation by Bootstrap

To quantify the uncertainties in the estimated model, we estimate the standard errors of both $\widehat{\boldsymbol{\theta}}$ and $\widehat{f}(\boldsymbol{\zeta})$ using bootstrap. In addition to the uncertainties in the regression step, our bootstrap procedure also takes into account the variation in mFPCA. The bootstrap samples are obtained by resampling residuals from both the observations on the functional covariates and the response variables. The procedure is as follows.

1. (Resampling the functional covariates) Recall that the discrete noisy observations on $\boldsymbol{x}_i$ are

$$w_{ijk} = x_{ij}(t_{ijk}) + e_{ijk}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, d, \quad k = 1, \ldots, N_{ij},$$

and the recovered functions from the discrete observations are $\tilde{x}_{ij}(t)$. Let $\widehat{e}_{ijk} = w_{ijk} - \tilde{x}_{ij}(t_{ijk})$ and resample with replacement e^*_{ijk} from $\{\widehat{e}_{ijk} : k = 1, \ldots, N_{ij}\}$ to obtain a bootstrap sample $w^*_{ijk} = \tilde{x}_{ij}(t_{ijk}) + e^*_{ijk}$. Repeat for all i, j, k, to obtain the bootstrap sample $\mathcal{W}^* = \{w^*_{ijk} : i = 1, \ldots, n, \ j = 1, \ldots, d, \ k = 1, \ldots, N_{ijk}\}$ for the functional data.

2. (Resampling the response) Denote $\widehat{y}_i$ as the fitted value of y_i from the original data and define the residuals $\widehat{\varepsilon}_i = \pi_i^{1/2}(y_i - \widehat{y}_i)$. Sample with replacement ε^*_i uniformly from $\{\widehat{\varepsilon}_i : i = 1, \ldots, n\}$ to obtain a bootstrap sample $y^*_i = \widehat{y}_i + \pi_i^{-1/2}\varepsilon^*_i$ of y_i. Denote the bootstrap sample as $\mathcal{Y}^* = \{y^*_i : i = 1, \ldots, n\}$.

3. Apply the mFPCA procedure on $\mathcal{W}^*$ to obtained mFPC scores $\boldsymbol{\zeta}^*$, and then fit the propose PLFAM to $\mathcal{Y}^*$ using $\boldsymbol{\zeta}^*$ and the original $\boldsymbol{Z}$. Denote the estimates from the bootstrap sample as $\widehat{\boldsymbol{\theta}}^*$ and $\widehat{f}^*(\boldsymbol{\zeta})$.

4. Repeat Steps 1- 3 a large number of times and use the sample standard deviations of $\widehat{\boldsymbol{\theta}}^*$ and $\widehat{f}^*(\boldsymbol{\zeta})$ as estimates of the standard errors for $\widehat{\boldsymbol{\theta}}$ and $\widehat{f}(\boldsymbol{\zeta})$.

Table S.1: Percentages of fitted model sizes.

Setting	Model	\% for the following model sizes							
		1	2	3	4	5	6	7	8
{(i), (I)}	FAM	1	40	58.5	0.5	0	0	0	0
	PLFAM	1	40	58.5	0.5	0	0	0	0
{(ii), (I)}	FAM	2.5	95.5	2	0	0	0	0	0
	PLFAM	2.5	95.5	2	0	0	0	0	0
{(i), (II)}	FAM	5.5	50	42	2.5	0	0	0	0
	PLFAM	0	37.5	62.5	0	0	0	0	0
{(ii), (II)}	FAM	15	84	1	0	0	0	0	0
	PLFAM	2	97	1	0	0	0	0	0

Table S.2: Percentages of selected components and, correct and super selection.

Setting	Model	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_3$	$\widehat{f}_4$	$\widehat{f}_5$	$\widehat{f}_6$	$\widehat{f}_7$	$\widehat{f}_8$	% correct set	% super set
{(i), (I)}	FAM	100	99	2.5	57	0	0	0	0	56.5	57
	PLFAM	100	99	2.5	57	0	0	0	0	56.5	57
{(ii), (I)}	FAM	100	97.5	2	0	0	0	0	0	0	0
	PLFAM	100	97.5	2	0	0	0	0	0	0	0
{(i), (II)}	FAM	100	81	3	57.5	0	0	0	0	41.5	44
	PLFAM	100	100	2.5	60	0	0	0	0	60	60
{(ii), (II)}	FAM	100	85	1	0	0	0	0	0	0	0
	PLFAM	100	98	1	0	0	0	0	0	0	0

Table S.3: Averaged integrated squared errors.

Setting	Model	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_3$	$\widehat{f}_4$	$\widehat{f}_5$	$\widehat{f}_6$	$\widehat{f}_7$	$\widehat{f}_8$	$\widehat{f}$
{(i), (I)}	FAM	0.0257	0.0903	0.0020	0.4682	0.0000	0.0000	0.0000	0.0000	0.5861
	PLFAM	0.0258	0.0907	0.0018	0.4682	0.0000	0.0000	0.0000	0.0000	0.5865
{(ii), (I)}	FAM	0.0321	0.1364	0.0026	0.9508	0.0000	0.0000	0.0000	0.0000	1.1219
	PLFAM	0.0324	0.1352	0.0027	0.9508	0.0000	0.0000	0.0000	0.0000	1.1210
{(i), (II)}	FAM	0.0439	0.2211	0.0056	0.4902	0.0000	0.0000	0.0000	0.0000	0.7609
	PLFAM	0.0252	0.0855	0.0015	0.4348	0.0000	0.0000	0.0000	0.0000	0.5470
{(ii), (II)}	FAM	0.0423	0.2158	0.0014	0.9508	0.0000	0.0000	0.0000	0.0000	1.2102
	PLFAM	0.0278	0.1341	0.0009	0.9508	0.0000	0.0000	0.0000	0.0000	1.1136

Table S.4: Prediction errors and mean squared errors for FAM and PLFAM, using separate univariate FPCA scores (columns labelled "separate") or mFPCA scores (columns labelled "joint"). For prediction errors, means are presented with corresponding standard deviations in parentheses.

Setting	Model	Prediction error separate	Prediction error joint	$\widehat{\theta}_1$ separate	$\widehat{\theta}_2$ separate	$\widehat{\theta}_3$ separate	$\widehat{\theta}_1$ joint	$\widehat{\theta}_2$ joint	$\widehat{\theta}_3$ joint
{(i), (I)}	FAM	1.68 (0.11)	1.68 (0.40)	-	-	-	-	-	-
	PLFAM	1.69 (0.11)	1.70 (0.41)	0.0763	0.0975	0.1097	0.0756	0.1047	0.1095
{(ii), (I)}	FAM	1.68 (0.10)	2.13 (0.12)	-	-	-	-	-	-
	PLFAM	1.69 (0.10)	2.15 (0.13)	0.0667	0.1108	0.0858	0.0767	0.1388	0.1100
{(i), (II)}	FAM	3.94 (0.24)	3.94 (0.36)	-	-	-	-	-	-
	PLFAM	1.71 (0.11)	1.69 (0.39)	0.0688	0.1091	0.0973	0.0686	0.1181	0.0818
{(ii), (II)}	FAM	3.91 (0.25)	4.29 (0.27)	-	-	-	-	-	-
	PLFAM	1.71 (0.11)	2.13 (0.13)	0.0675	0.0897	0.1156	0.079	0.1332	0.1284

Table S.5: Average of 5-year overall prediction errors.

		corn	soybean
(a) functional additive models	PLFAM(joint)	298.43	35.64
	PLFAM(separate)	306.50	38.85
	PLFAM(max)	324.27	38.22
	PLFAM(min)	338.51	44.09
	PLFAM(mean)	330.17	40.93
	FAM(joint)	830.17	48.54
	FAM(separate)	839.00	51.06
	FAM(max)	898.12	51.92
	FAM(min)	997.27	65.48
	FAM(mean)	916.80	57.79
(b) functional linear models	FLM-Cov(joint)	303.81	35.29
	FLM-Cov(separate)	308.57	35.69
	FLM-Cov(max)	317.83	37.52
	FLM-Cov(min)	338.88	42.43
	FLM-Cov(mean)	310.02	37.27
	FLM(joint)	704.19	47.31
	FLM(separate)	767.42	50.42
	FLM(max)	779.56	51.49
	FLM(min)	842.12	61.42
	FLM(mean)	790.96	52.38
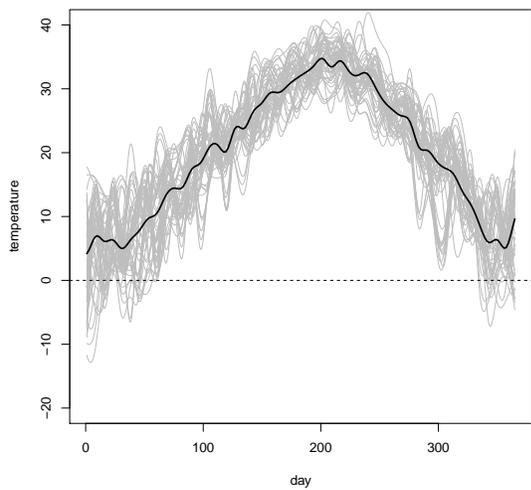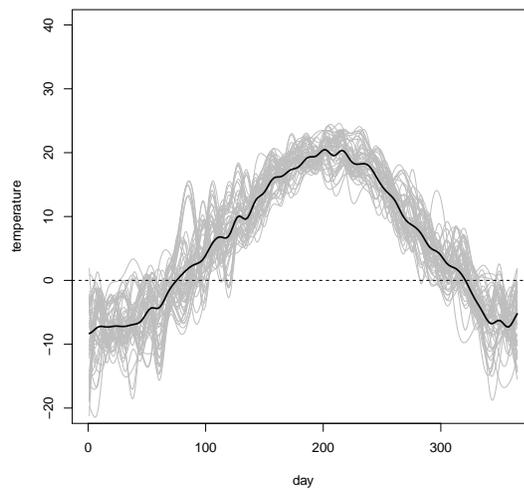
(a) $X_1(t)$ (b) $X_2(t)$

Figure S.1: 50 randomly selected trajectories for daily maximum and daily minimum temperature. The solid dark curve in each panel is the mean function.
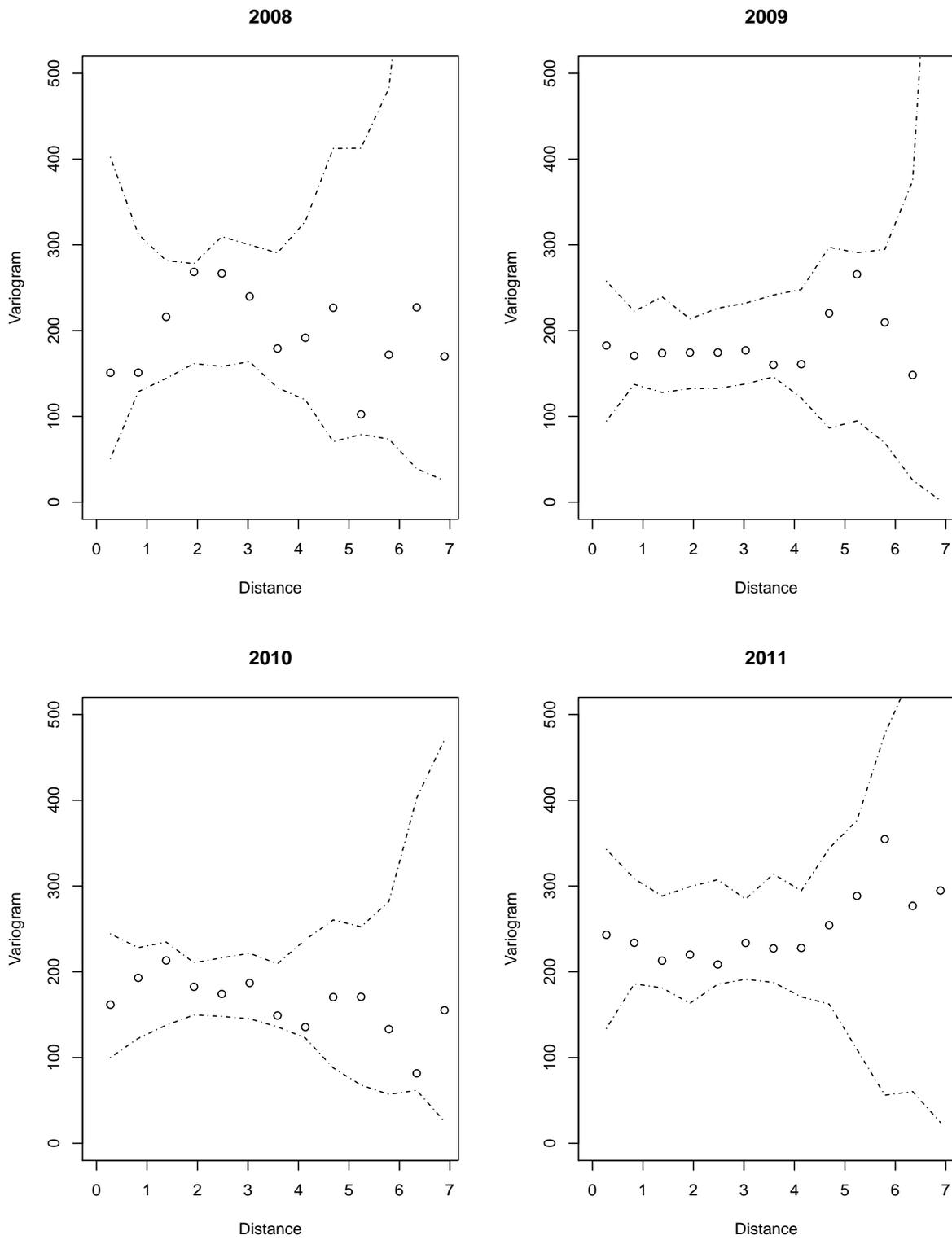
Figure S.2: Spatial variograms for each year from 2008 to 2011, based on the residuals from the corn yield prediction model. The unit in the horizontal axis is degree (in longitude or latitude). The dotted curves are confidence bands based on the assumption of no spatial dependency.
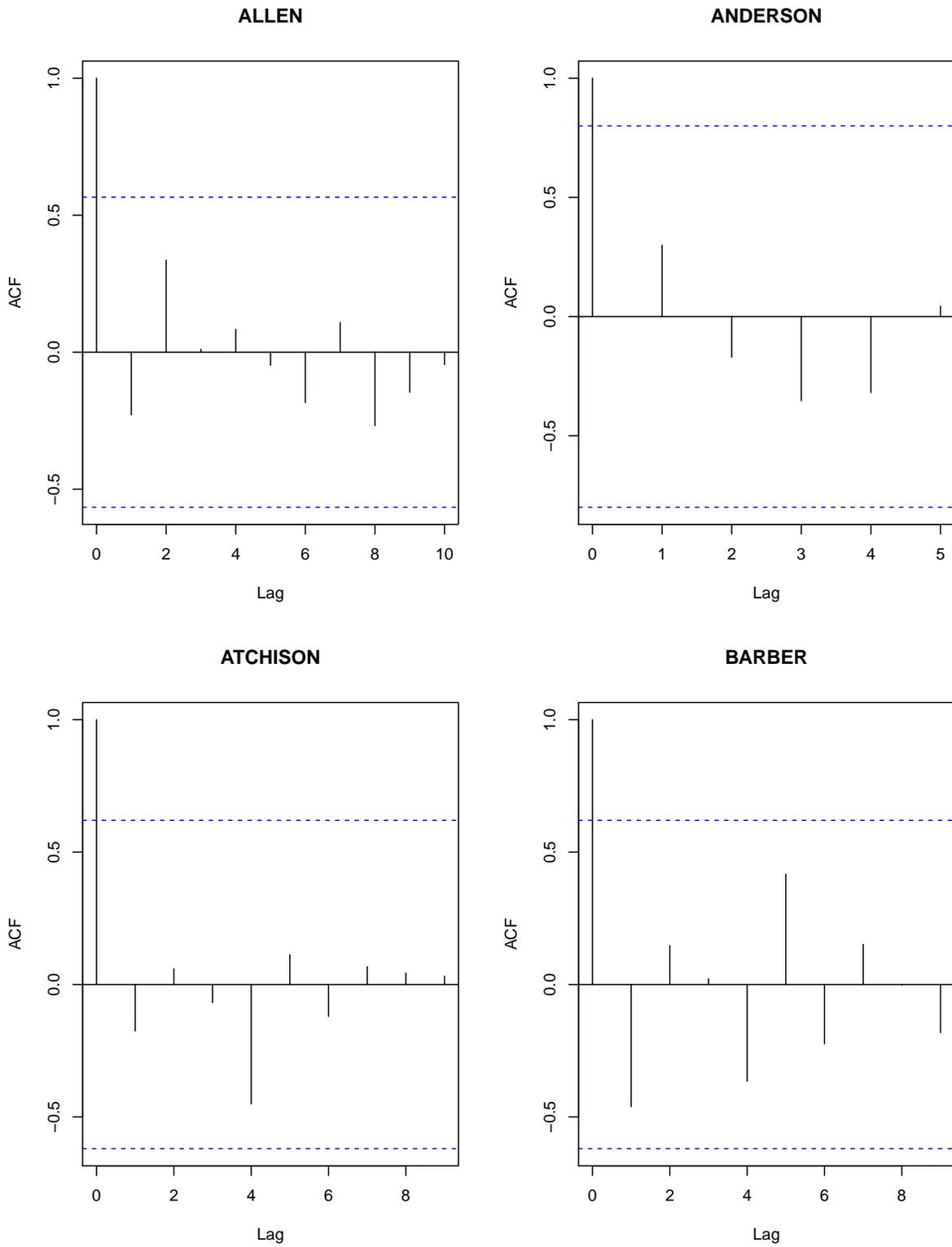
Figure S.3: The ACF plot for the first four counties, based on the residuals from the corn yield prediction model.

S.29