Statistica Sinica: Supplement

# DEPENDENT FUNCTIONAL LINEAR MODELS WITH APPLICATIONS TO MONITORING STRUCTURAL CHANGE

Alexander Aue
Siegfried Hörmann
Lajos Horváth
Marie Hušková

*University of California, Davis, Université Libre de Bruxelles, University of Utah, Charles University*

**Supplementary Material**

## S1 Additional figures and tables

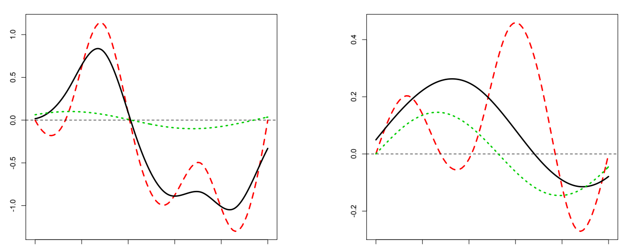

Figure 1: Predictor functions $X_k$ (dashed), error functions $\varepsilon_k$ (dotted) and response functions $Y_k$ (solid) from the functional linear model generated by (4.2)–(4.4). The bandwidth parameter of the regression operator $\Psi$ is $b = 0.2$ in the left panel and $b = 0.5$ in the right panel.

## S2 Approximation for partial sums of random vectors

In this section, we provide strong approximations for sums of weakly dependent random variables which are used in the paper to establish Lemma 5.4, but which may also be of independent interest.

**Theorem S2.1** *Let* $(\mathbf{Y}_i : i \in \mathbb{Z})$ *be an* $\mathbb{R}^d$*-valued random sequence, such that*

$$\mathbf{Y}_i = f(\varepsilon_i, \varepsilon_{i-1}, \ldots),$$

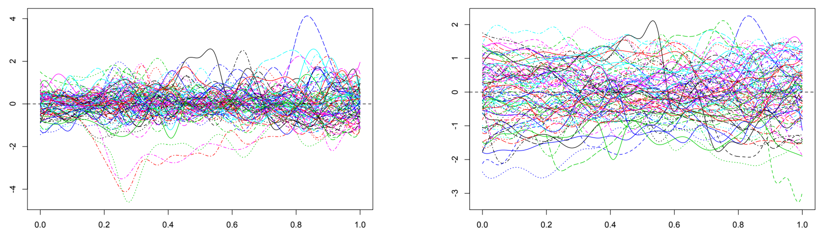


Figure 2: Predictor functions $X_k$ giving NOx concentration (left) and response functions $Y_k$ giving PM10 concentrations (right), both measured at Völkermarkterstraße.

*where* $(\varepsilon_i : i \in \mathbb{Z})$ *is an i.i.d. sequence with values in some arbitrary metric space. We assume that* $E[\mathbf{Y}_0] = \mathbf{0}$ *and* $E[|\mathbf{Y}_0|^\theta] < \infty$ *for some* $\theta > 2$*. (Here and in the sequel* $|\boldsymbol{u}|$ *denotes the Euclidean norm of the vector* $\boldsymbol{u}$*.) Let* $(\varepsilon_i^* : i \in \mathbb{Z})$ *be an independent copy of* $(\varepsilon_i : i \in \mathbb{Z})$ *and define*

$$\mathbf{Y}_{0m} = f(\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-m}, \varepsilon^*_{-m-1}, \varepsilon^*_{-m-2}, \ldots).$$

*If, for some* $A > 2$,

$$\Big(E\big[|\mathbf{Y}_0 - \mathbf{Y}_{0m}|^\theta\big]\Big)^{1/\theta} \le C_0 m^{-A}, \tag{S2.1}$$

*then*

$$\Gamma = \sum_{h \in \mathbb{Z}} E[\mathbf{Y}_0 \mathbf{Y}_h']$$

*converges (coordinatewise) absolutely and* $(\mathbf{Y}_i : i \in \mathbb{Z})$ *can be redefined on a new probability space together with a sequence of i.i.d. normal random vectors* $(\boldsymbol{N}_i : i \in \mathbb{Z})$ *with* $N_i \sim N(\mathbf{0}, \Gamma)$*, such that*

$$\left| \sum_{i=1}^k \mathbf{Y}_i - \sum_{i=1}^k \boldsymbol{N}_i \right| = O\Big(k^{1/2-\mu}\Big) \quad \text{a.s.} \tag{S2.2}$$

*The constant* $\mu > 0$ *depends on* $A$*,* $d$ *and* $\theta$*.*

The proof of Theorem S2.1 requires some auxiliary lemmas.

**Lemma S2.1** *Let* $(\mathbf{Y}_i : i \in \mathbb{Z})$ *satisfy the assumption of Theorem 2..1 Set* $\boldsymbol{S}_k = \sum_{i=1}^k \mathbf{Y}_i$*. Then* $E[\max_{1\le k\le n} |\boldsymbol{S}_k|^\theta] \le C_\theta n^{\theta/2}$*. The constant* $C_\theta$ *does not depend on* $n$*.*

*Proof.* Let $\mathbf{Y}_0(v)$, $1 \le v \le d$, be a coordinate of $\mathbf{Y}_0$. Then (S2.1) obviously implies that $\delta_\theta(m) = (E[|\mathbf{Y}_0(v) - \mathbf{Y}_{0m}(v)|^\theta])^{1/\theta} = O(m^{-A})$. Since $\sum_{m=1}^\infty \delta_\theta(m) < \infty$ we infer from Corollary 1 in Berkes et al. (2011) that

$$E\left[\max_{1\le k\le n} |\boldsymbol{S}_k(v)|^\theta\right] \le C_\theta n^{\theta/2}.$$

Utilizing the inequality

$$\max_{1\le k\le n} |\boldsymbol{S}_k|^\theta \le d\sum_{v=1}^{d} \max_{1\le k\le n} |\boldsymbol{S}_k(v)|^\theta,$$

the assertion follows. □

For the next lemma we introduce new variables:

$$\mathbf{Y}_k^{(m)} = f(\varepsilon_k, \varepsilon_{k-1}, \ldots, \varepsilon_{k-m}, \varepsilon_{k-m-1}^{(k)}, \varepsilon_{k-m-2}^{(k)}, \ldots),$$

where for each $k \in \mathbb{Z}$ the sequences $(\varepsilon_i^{(k)} : i \in \mathbb{Z})$ are mutually independent copies of $(\varepsilon_i : i \in \mathbb{Z})$. It follows that the $\mathbf{Y}_k^{(m)}$ have the same marginal distribution as the $\mathbf{Y}_k$ and that the $(\mathbf{Y}_k^{(m)} : k \in \mathbb{Z})$ are $m$-dependent.

**Lemma S2.2** *We work under the conditions of Theorem S2.1. Define*

$$\Gamma_n^{(m)} = \frac{1}{n}\mathrm{Var}\Big(\mathbf{Y}_1^{(m)} + \cdots + \mathbf{Y}_n^{(m)}\Big).$$

*Then*

$$\big|\Gamma - \Gamma_n^{(m)}\big|_\infty \le C_1\big(n^{-1} + m^{-A+1}\big),$$

*where $C_1$ depends only on $E[|\mathbf{Y}_0|^2]$, $C_0$ and $A$. Here $|M|_\infty = \max_{1\le i,j\le d} m_{ij}$ for some matrix $M$ with entries $m_{ij}$.*

*Proof.* We have, for all $1 \le v_1, v_2 \le d$,

$$\begin{aligned}
\big|E[\mathbf{Y}_0(v_1)\mathbf{Y}_i(v_2)]\big| &= \big|E\big[\mathbf{Y}_0(v_1)\big(\mathbf{Y}_i(v_2) - \mathbf{Y}_i^{(i)}(v_2)\big)\big]\big| \\
&\le \big(E\big[|\mathbf{Y}_0(v_1)|^2\big]\big)^{1/2}\Big(E\big[|\mathbf{Y}_i(v_2) - \mathbf{Y}_i^{(i)}(v_2)|^2\big]\Big)^{1/2} \\
&\le \big(E[|\mathbf{Y}_0|^2]\big)^{1/2}\Big(E\big[\big|\mathbf{Y}_i - \mathbf{Y}_i^{(i)}\big|^\theta\big]\Big)^{1/\theta} \\
&\le C_0\left(E[|\mathbf{Y}_0|^2]\right)^{1/2} i^{-A}.
\end{aligned}$$

This shows that the series for $\Gamma$ is absolutely convergent. Using the strict stationarity of $(\mathbf{Y}_i : i \in \mathbb{Z})$ we obtain

$$\Gamma_n(v_1, v_2) = \frac{1}{n}\mathrm{Cov}\Big(\boldsymbol{S}_n(v_1), \boldsymbol{S}_n(v_2)\Big) = \sum_{|i|<n}\Big(1 - \frac{|i|}{n}\Big)\mathrm{Cov}(\mathbf{Y}_0(v_1), \mathbf{Y}_i(v_2)).$$

Thus, since $A > 2$,

$$\begin{aligned}\left|\Gamma(v_1,v_2)-\Gamma_n(v_1,v_2)\right| &\le \left|\sum_{|i|\ge n}\mathrm{Cov}(\mathbf{Y}_0(v_1),\mathbf{Y}_i(v_2))\right| + \left|\sum_{|i|<n}\frac{|i|}{n}\mathrm{Cov}(\mathbf{Y}_0(v_1),\mathbf{Y}_i(v_2))\right| \\ &\le C_0\left(E[|\mathbf{Y}_0|^2]\right)^{1/2}\left(\sum_{|i|\ge n}|i|^{-A}+\frac{1}{n}\sum_{|i|<n}|i|^{-A+1}\right) \\ &\le C_1 n^{-1},\end{aligned}$$

where the constant $C_1$ depends on $C_0$, $A$ and $E[|\mathbf{Y}_0|^2]$ but not on $v_1$ and $v_2$. Note that the sequence $(\mathbf{Y}_k^{(m)} : k \in \mathbb{Z})$ is also strictly stationary and therefore some routine calculations imply that, for $n \le m$,

$$\begin{aligned}&\left|\Gamma_n(v_1,v_2)-\Gamma_n^{(m)}(v_1,v_2)\right| \\ &\quad\le \sum_{|i|<n}\left(1-\frac{|i|}{n}\right)\left|\mathrm{Cov}(\mathbf{Y}_0(v_1),\mathbf{Y}_i(v_2))-\mathrm{Cov}(\mathbf{Y}_0^{(m)}(v_1),\mathbf{Y}_i^{(m)}(v_2))\right| \\ &\quad\le \sum_{|i|<n}2\left(1-\frac{|i|}{n}\right)\left(E[|\mathbf{Y}_0|^2]\right)^{1/2}\left(E\left[\left|\mathbf{Y}_0-\mathbf{Y}_0^{(m)}\right|^2\right]\right)^{1/2} \\ &\quad\le 4C_0\left(E[|\mathbf{Y}_0|^2]\right)^{1/2} nm^{-A}.\end{aligned}$$

If $n > m$ we have, for $i > m$, $\mathrm{Cov}(\mathbf{Y}_0^{(m)}(v_1),\mathbf{Y}_i^{(m)}(v_2)) = 0$ and $|\mathrm{Cov}(\mathbf{Y}_0(v_1),\mathbf{Y}_i(v_2))| \le C_0(E[|\mathbf{Y}_0|^2])^{1/2} i^{-A}$. □

The following Lemma is a special case of Theorem 1 in Berkes and Philipp (1978). It is the crucial ingredient for the construction of the approximating random vectors $\boldsymbol{N}_i$ in (S2.2).

**Lemma S2.3** *Let $(\mathbf{X}_\ell : \ell \ge 1)$ be a sequence of independent $\mathbb{R}^d$-valued random vectors with characteristic functions $f_\ell(\boldsymbol{u})$, $\boldsymbol{u} \in \mathbb{R}^d$, and let $(G_\ell : \ell \ge 1)$ be a sequence of probability distributions on $\mathbb{R}^d$ with characteristic functions $g_\ell(\boldsymbol{u})$, $\boldsymbol{u} \in \mathbb{R}^d$. Suppose that, for some nonnegative numbers $\lambda_\ell$, $\delta_\ell$ and $W_\ell \ge 10^8 d$,*

$$|f_\ell(\boldsymbol{u}) - g_\ell(\boldsymbol{u})| \le \lambda_\ell$$

*for all $\boldsymbol{u}$ with $|\boldsymbol{u}| \le W_\ell$ and*

$$G_\ell\Big(\boldsymbol{u} : |\boldsymbol{u}| > W_\ell/4\Big) \le \delta_\ell.$$

*Then without changing its distribution we can redefine the sequence $(\mathbf{X}_\ell : \ell \ge 1)$ on a richer probability space together with a sequence $(\mathbf{Y}_\ell : \ell \ge 1)$ of independent random variables such that $\mathbf{Y}_\ell \overset{\mathcal{D}}{=} G_\ell$ and*

$$P\Big(|\mathbf{X}_\ell - \mathbf{Y}_\ell| \ge \alpha_\ell\Big) \le \alpha_\ell \qquad \text{for } \ell \in \mathbb{N},$$

*where* $\alpha_1 = 1$ *and*

$$\alpha_\ell = 16dW_\ell^{-1}\log W_\ell + 4\lambda_\ell^{1/2}W_\ell^d + \delta_\ell \qquad \text{for } \ell \geq 2.$$

The next Lemma is due to von Bahr (1967).

**Lemma S2.4** *Let* $(\boldsymbol{Z}_i\colon i \in \mathbb{Z})$ *be a sequence of i.i.d. random vectors with* $E[\boldsymbol{Z}_1] = \mathbf{0}$ *and* $E[|\boldsymbol{Z}_1|^\theta] < \infty$, $\theta > 2$, *and* $\Sigma = E[\boldsymbol{Z}_1\boldsymbol{Z}_1']$. *Further let* $f_n(\boldsymbol{u})$ *be the characteristic function of* $n^{-1/2}(\boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_n)$. *Then*

$$\left| f_n(\boldsymbol{u}) - \exp\big(-\frac{1}{2}\,\boldsymbol{u}'\Sigma\,\boldsymbol{u}\big)\right| \leq C_3 n^{-(\theta-2)/2}|\boldsymbol{u}|^\theta \exp(-C_4|\boldsymbol{u}|^2)$$

*for all* $\boldsymbol{u} \in \mathbb{R}^d$ *with* $|\boldsymbol{u}| \leq C_5 n^{1/2}$. *The constants* $C_3, C_4$ *and* $C_5$ *only depend on* $d$ *and the moments of* $\boldsymbol{Z}_1$.

For the proof of Theorem S2.1 we use a blocking argument. We introduce some further notation. Let $(t_k\colon k \in \mathbb{N})$ be an integer-valued sequence with $t_1 = 1$ and $t_k - t_{k-1} = \lfloor k^a \rfloor$. The constant $a$ will be defined later. Note that

$$k^a \ll t_k \ll k^{a+1}, \tag{S2.3}$$

where here and in the sequel $a_k \ll b_k$ means $\limsup_k |a_k/b_k| < \infty$. We set $T_k = \{t_{k-1}+1, \ldots, t_k\}$ and divide $T_k$ into "short blocks" $J_{k_l}$ and "long blocks" $I_{k_l}$, so that

$$T_k = J_{k_1} \cup I_{k_1} \cup J_{k_2} \cup I_{k_2} \cup \ldots \cup J_{k_n} \cup I_{k_n} \cup R_k,$$

where $|J_{k_l}| = \lfloor k^\beta \rfloor$ and $|I_{k_l}| = \lfloor k^b \rfloor$ with $0 < \beta < b < a$. (As usual $|\mathcal{S}|$ denotes the cardinality of a set $\mathcal{S}$.) Clearly $n = n(k) \sim k^{a-b}$ and the interval $R_k$ contains the remaining integers which are not contained in some interval $J_{k_l} \cup I_{k_l}$. If $i \in T_k$ we set $\mathbf{Y}_i^* = \mathbf{Y}_i^{(m)}$ with $m = |J_{k_1}|$. This implies that $\boldsymbol{Z}_{k_j} = |I_{k_1}|^{-1/2}\sum_{i \in I_{k_j}} \mathbf{Y}_i^*$, $j = 1, \ldots, n$, are independent and identically distributed.

Now we define the constants $a, b, \beta$ and $\mu$. We let $\rho > 0$ be a fixed but arbitrary small number. Then

$$a - b = \frac{8(d+1)(\rho+1)}{\theta - 2}, \tag{S2.4}$$

$$\mu = \frac{1}{1+\rho}\frac{1}{1+a}\Big(\frac{1}{2} - \frac{1}{\theta}\Big), \tag{S2.5}$$

$$\beta = b - 1. \tag{S2.6}$$

Finally $\beta$ is chosen such that

$$\beta > \max\left\{4(d+2)(1+\rho), \frac{a+1}{A}\right\}. \tag{S2.7}$$

*Proof of Theorem S2.1.* The proof of Theorem S2.1 is divided into two steps. In the first step we reduce (S2.2) to an alternative approximation problem. In the second step we will construct the approximating normal sequence.

*Step 1:* We claim that it is sufficient to show that

$$\left|\sum_{\ell=1}^{k}\sum_{j=1}^{n}\sum_{i\in I_{\ell_j}}(\mathbf{Y}_i^*-\boldsymbol{N}_i)\right| = O\Big(t_k^{1/2-\mu}\Big) \quad \text{a.s.} \tag{S2.8}$$

Essentially this means that we only need to prove (S2.2) for the perturbed sequence $(\mathbf{Y}_i^*)$ along the subsequence $(t_k)$. Further we need to show that the "short block" sums are negligible.

Towards the proof of *Step 1*, let

$$A_k = \left\{\max_{t_{k-1}+1\le\ell\le t_k}\left|\sum_{i=t_{k-1}+1}^{\ell}\mathbf{Y}_i\right| > t_{k-1}^{1/2-\mu}\right\}$$

and

$$\bar{A}_k = \left\{\max_{t_{k-1}+1\le\ell\le t_k}\left|\sum_{i=t_{k-1}+1}^{\ell}\boldsymbol{N}_i\right| > t_{k-1}^{1/2-\mu}\right\}.$$

From Lemma S2.1, the Markov inequality and (S2.3) we infer

$$\begin{aligned}\sum_{k\ge1}P(A_k) &\le \sum_{k\ge1}C_\theta(t_k-t_{k-1})^{\theta/2}t_{k-1}^{-(1/2-\mu)\theta}\\ &\ll \sum_{k\ge1}k^{a\theta/2-(a+1)(1/2-\mu)\theta}\\ &= \sum_{k\ge1}k^{-\frac{1+(\theta/2)\rho}{1+\rho}} < \infty.\end{aligned}$$

The last line in the display above follows from (S2.5) and the requirement $\theta > 2$. Hence by the Borel-Cantelli Lemma $P(A_k \text{ i.o.}) = 0$. A similar argument gives $P(\bar{A}_k \text{ i.o.}) = 0$. We conclude that (S2.2) follows from

$$\left|\sum_{i=1}^{t_k}(\mathbf{Y}_i-\boldsymbol{N}_i)\right| = O\Big(t_k^{1/2-\mu}\Big) \quad \text{a.s.}$$

Next we show that the contribution of the "short block" sums is negligible. Let

$$B_k = \left\{\left|\sum_{\ell=1}^{k}\sum_{j=1}^{n}\sum_{i\in J_{\ell_j}}\mathbf{Y}_i\right| > t_{k-1}^{1/2-\mu}\right\}$$

and

$$\bar{B}_k = \left\{\left|\sum_{\ell=1}^{k}\sum_{j=1}^{n}\sum_{i\in J_{\ell_j}}\boldsymbol{N}_i\right| > t_{k-1}^{1/2-\mu}\right\}.$$

Using similar arguments as before and (S2.6) we obtain

$$\begin{aligned}\sum_{k\geq 1} P(B_k) &\leq \sum_{k\geq 1} E\Big[\Big|\sum_{\ell=1}^{k}\sum_{j=1}^{n}\sum_{i\in J_{\ell_j}} \mathbf{Y}_i\Big|^{\theta}\Big] t_{k-1}^{-(1/2-\mu)\theta} \\ &\ll \sum_{k\geq 1}\left(\sum_{\ell=1}^{k}\ell^{a-b+\beta}\right)^{\theta/2} k^{-(1+a)(1/2-\mu)\theta} \\ &\ll \sum_{k\geq 1} k^{-\frac{1+(\theta/2)\rho}{1+\rho}} < \infty.\end{aligned}$$

We infer that $P(B_k \text{ i.o.}) = P(\bar{B}_k \text{ i.o.}) = 0$.

To complete *Step 1* we note that

$$\sum_{i\geq 1}(\mathbf{Y}_i - \mathbf{Y}_i^*) < \infty \quad \text{a.s.} \tag{S2.9}$$

This implies that it is sufficient to work with the sequence $(\mathbf{Y}_i^*)$. Relation (S2.9) follows from

$$E\Big[\Big|\sum_{i\geq 1}(\mathbf{Y}_i - \mathbf{Y}_i^*)\Big|\Big] \leq \sum_{i\geq 1}\left(E\big[|\mathbf{Y}_i - \mathbf{Y}_i^*|^{\theta}\big]\right)^{1/\theta} \leq \sum_{k\geq 1}|T_k||J_{k_1}|^{-A} \ll \sum_{k\geq 1} k^{a-A\beta} < \infty.$$

The convergence of the series follows from (S2.6) and (S2.7).
*Step 2:* We define $\mathbf{X}_\ell = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\boldsymbol{Z}_{\ell_j}$. Further we let $f_{\mathbf{X}_\ell}(\boldsymbol{u})$, $f_{N(\mathbf{0},\Gamma)}(\boldsymbol{u})$ and $f_{N(\mathbf{0},\Gamma_\ell)}(\boldsymbol{u})$ be the characteristic functions of $\mathbf{X}_\ell$, $\boldsymbol{N}_1$ and a normal random vector with zero expectation and variance $\Gamma_\ell = \mathrm{Var}(\boldsymbol{Z}_{\ell_1})$. Notice again that $\ell^{(a-b)} \ll n \ll \ell^{(a-b)}$. Thus by (S2.4) and Lemmas S2.2 and S2.4 we get for $|\boldsymbol{u}| \leq C_5\ell^{2(1+\rho)}$

$$\begin{aligned}&\Big|f_{\mathbf{X}_\ell}(\boldsymbol{u}) - f_{N(\mathbf{0},\Gamma)}(\boldsymbol{u})\Big| \\ &\quad\leq \Big|f_{\mathbf{X}_\ell}(\boldsymbol{u}) - f_{N(\mathbf{0},\Gamma_\ell)}(\boldsymbol{u})\Big| + \Big|f_{N(\mathbf{0},\Gamma_\ell)}(\boldsymbol{u}) - f_{N(\mathbf{0},\Gamma)}(\boldsymbol{u})\Big| \\ &\quad\leq C_3 n^{-(\theta-2)/2}|\boldsymbol{u}|^{\theta}\exp(-C_4|\boldsymbol{u}|^2) + \Big|\boldsymbol{u}'(\Gamma_\ell - \Gamma)\boldsymbol{u}\Big| \\ &\quad\ll n^{-(\theta-2)/2} + |\boldsymbol{u}|^2\Big|\Gamma_\ell - \Gamma\Big|_\infty \\ &\quad\ll n^{-(\theta-2)/2} + \ell^{4(1+\rho)}\Big(\ell^{-b} + (\ell^{\beta})^{-A+1}\Big) \\ &\quad\ll \ell^{\gamma},\end{aligned}$$

where

$$\gamma = \max\{-(\theta-2)(a-b)/2,\, 4(1+\rho) - \beta\}.$$

The crucial step is now the application of Lemma S2.3. Let $W_\ell = \max\{C_5\ell^{2(1+\rho)}, 10^8 d\}$ and

$$\lambda_\ell = \sup_{|\boldsymbol{u}|\le W_\ell} \Big| f_{\mathbf{X}_\ell}(\boldsymbol{u}) - f_{N(\mathbf{0},\Gamma)}(\boldsymbol{u})\Big|.$$

A routine calculation shows that for some sufficiently large $\varrho > 0$

$$\delta_\ell := P\Big(|N(0,\Gamma)| > W_\ell/4\Big) \ll \exp(-\ell^\varrho).$$

Thus we have

$$\alpha_\ell = 16dW_\ell^{-1}\log W_\ell + 4\lambda_\ell^{1/2}W_\ell^d + \delta_\ell.$$

Using the definitions of the constants $a$, $b$ and $\beta$ it can be easily shown that $\alpha_\ell \ll \frac{\log \ell}{\ell^{2(1+\rho)}}$.

Since the $\mathbf{X}_\ell$ are by definition independent we conclude with Lemma S2.3 that there exists a sequence $(\boldsymbol{M}_\ell\colon \ell\in\mathbb{N})$ of i.i.d. normal random variables, with $\boldsymbol{M}_1 \sim N(\mathbf{0},\Gamma)$, such that, for some large enough constant $C_7$,

$$P\Big(|\mathbf{X}_\ell - \boldsymbol{M}_\ell| \ge C_7\ell^{-(2+\rho)}\Big) \le C_7\ell^{-(2+\rho)}.$$

The latter inequality implies that

$$P\bigg(\Big|\sum_{\ell=1}^k\sum_{j=1}^n\sum_{i\in I_{\ell_j}}\mathbf{Y}_i^* - \sum_{\ell=1}^k(n|I_{\ell_1}|)^{1/2}\boldsymbol{M}_\ell\Big| \ge C_7\sum_{\ell=1}^k(n|I_{\ell_1}|)^{1/2}\ell^{-(2+\rho)}\bigg) \le C_7k^{-(1+\rho)}. \tag{S2.10}$$

Further enlarging the probability space we can write

$$(n|I_{\ell_1}|)^{1/2}\boldsymbol{M}_\ell = \sum_{j=1}^n\sum_{i\in I_{\ell_j}}\boldsymbol{N}_i. \tag{S2.11}$$

Some algebra shows that

$$\sum_{\ell=1}^k(n|I_{\ell_1}|)^{1/2}\ell^{-(2+\rho)} \ll t_{k-1}^{1/2-\mu}. \tag{S2.12}$$

The Borel-Cantelli Lemma and (S2.10)–(S2.12) imply (S2.8). □

## S3 Some technical lemmas

To establish Lemma 6.2, several additional results concerned with a decomposition of $\hat{\Xi}'_{n,N}\hat{\boldsymbol{E}}_{n,N}$ are needed. These are provided in this section. Using the definitions of $\hat{\Xi}_{n,N}$ and $\hat{\boldsymbol{E}}_{n,N}$ we have that

$$\hat{\Xi}'_{n,N}\hat{\boldsymbol{E}}_{n,N} = \mathrm{vec}(\hat{G}), \tag{S3.1}$$

where the matrix $\hat{G} = (\hat{G}_{i,j} : i = 1, \ldots, p, j = 1, \ldots, q)$ is given by the entries $\hat{G}_{i,j} = \sum_{k=n+1}^{N} \hat{\xi}_{k,i}\hat{\eta}_{k,j}$. For the individual terms in the latter sum we obtain from (2.4) the expression

$$\hat{\xi}_{k,i}\hat{\eta}_{k,j} = \hat{\xi}_{k,i}\big(\langle \varepsilon_k, \hat{w}_j\rangle + \langle \rho_{k,1}, \hat{w}_j\rangle + \langle \rho_{k,2}, \hat{w}_j\rangle + \langle \phi_{k,1}, \hat{w}_j\rangle + \langle \phi_{k,2}, \hat{w}_j\rangle\big) \tag{S3.2}$$

writing $\phi_{k,1} = \phi_{k,1}^{(m)}$ and $\phi_{k,2} = \phi_{k,2}^{(m)}$ to simplify notation. In the following series of lemmas, we establish the large-sample behavior of the partial sums of the terms on the right-hand side of (S3.2). Let $\hat{F}_{n,N} = \frac{1}{N-n}\sum_{k=n+1}^{N} X_k \otimes \varepsilon_k$. This is a random operator estimating the zero cross-covariances between $\varepsilon_k$ and $X_k$. Recall that $x \otimes y(v) = \langle x, v\rangle y$.

**Lemma S3.1** *If Assumptions 2.1–2.5 hold, then*

$$\left| \sum_{k=1}^{m} \hat{\xi}_{k,i}\langle \varepsilon_k, \hat{w}_j\rangle - T_{0,m}^{(1)}(i,j) \right| = O_P(1) \tag{S3.3}$$

*and*

$$\sup_{\ell > 1} \frac{1}{\sqrt{\ell(\log \ell)^{\beta}}} \left| \sum_{k=m+1}^{m+\ell} \hat{\xi}_{k,i}\langle \varepsilon_k, \hat{w}_j\rangle - T_{m,m+k}^{(1)}(i,j) \right| = O_P\left(\frac{1}{\sqrt{m}}\right) \tag{S3.4}$$

*for all $\beta > 3$, where*

$$T_{n,N}^{(1)}(i,j) = \hat{c}_i\hat{d}_j \sum_{k=n+1}^{N} \xi_{k,i}\eta_{k,j}^{*}.$$

*Proof.* We prove only (S3.4) as (S3.3) can be established in a similar fashion. In a first step, observe that $\hat{\xi}_{k,i}\langle \varepsilon_k, \hat{w}_j\rangle = \hat{c}_i\hat{d}_j\xi_{k,i}\eta_{k,j}^{*} + (\hat{\xi}_{k,i} - \hat{c}_i\xi_{k,i})\langle \varepsilon_k, \hat{w}_j\rangle + \hat{c}_i\xi_{k,i}(\langle \varepsilon_k, \hat{w}_j\rangle - \hat{d}_j\eta_{k,j}^{*})$. Summing over the first term on the right-hand side gives rise to $\hat{c}_i\hat{d}_j T_{m,m+\ell}^{(1)}(i,j)$. For the sum of the second term, we note that

$$\sum_{k=n+1}^{N} (\hat{\xi}_{k,i} - \hat{c}_i\xi_{k,i})\langle \varepsilon_k, \hat{w}_j\rangle = \left\langle \left(\sum_{k=n+1}^{N} X_k \otimes \varepsilon_k\right)(\hat{v}_i - \hat{c}_i v_i), \hat{w}_j \right\rangle.$$

Hence it follows that

$$\begin{aligned}
&\sup_{\ell > 1} \frac{1}{\sqrt{\ell(\log \ell)^{\beta}}} \left| \sum_{k=m+1}^{m+\ell} (\hat{\xi}_{k,i} - \hat{c}_i\xi_{k,i})\langle \varepsilon_k, \hat{w}_j\rangle \right| \\
&\quad \le \sup_{\ell > 1} \frac{1}{\sqrt{\ell(\log \ell)^{\beta}}} \|\ell \hat{F}_{m,m+\ell}\|_{\mathcal{S}} \|\hat{v}_i - \hat{c}_i v_i\| \|\hat{w}_j\| \\
&\quad = O_P\left(\frac{1}{\sqrt{m}}\right),
\end{aligned}$$

using the Cauchy-Schwarz inequality, Lemma 5.1 and the same arguments used in Lemma 5.2 to bound $\|\hat{F}_{m,m+\ell}\|_{\mathcal{S}}$. Similarly one shows that

$$\sup_{\ell>1} \frac{1}{\sqrt{\ell(\log \ell)^\beta}} \left| \sum_{k=m+1}^{m+\ell} \hat{c}_i \xi_{k,i}(\hat{\eta}_{k,j} - \hat{d}_j \eta_{k,j}) \right| = O_P\left(\frac{1}{\sqrt{m}}\right).$$

This completes the proof of (S3.4). □

**Lemma S3.2** *If Assumptions 2.1–2.5 hold, then*

$$\left| \sum_{k=1}^{m} \hat{\xi}_{k,i} \langle \rho_{k,1}, \hat{w}_j \rangle - \left( T^{(2)}_{0,m}(i,j) + mR^{(3)}_m(i,j) \right) \right| = O_P(1) \tag{S3.5}$$

*and*

$$\sum_{k=m+1}^{m+\ell} \hat{\xi}_{k,i} \langle \rho_{k,1}, \hat{w}_j \rangle = T^{(2)}_{m,m+\ell}(i,j) + \ell R^{(3)}_m(i,j) + U_{m,1}(\ell) + U_{m,2}(\ell), \tag{S3.6}$$

*where* $T^{(2)}_{n,N}(i,j) = \hat{c}_i \hat{d}_j \sum_{k=n+1}^{N} \xi_{k,i}\theta_{k,j}$, $R^{(1)}_m(i,j) = \hat{d}_j \ell \, \langle C(\hat{v}_i - \hat{c}_i v_i), u_j \rangle$,

$$\sup_{\ell>1} \frac{|U_{m,1}(\ell)|}{\sqrt{\ell(\log \ell)^\beta}} = O_P\left(\frac{1}{\sqrt{m}}\right) \qquad \textit{and} \qquad \sup_{\ell>1} \frac{|U_{m,2}(\ell)|}{\ell} = O_P\left(\frac{1}{m}\right).$$

*Proof.* Let $\varrho^{(1)}_{k,j} = \langle \rho_{k,1}, w_j \rangle$ and $\hat{\varrho}^{(1)}_{k,j} = \langle \rho_{k,1}, \hat{w}_j \rangle$. Then,

$$\hat{\xi}_{k,i}\hat{\varrho}^{(1)}_{k,j} = \hat{c}_i\hat{d}_j\xi_{k,i}\varrho^{(1)}_{k,j} + \hat{c}_i\xi_{k,i}(\hat{\varrho}^{(1)}_{k,j} - \hat{d}_j\varrho^{(1)}_{k,j}) + \hat{d}_j(\hat{\xi}_{k,i} - \hat{c}_i\xi_{k,i})\varrho^{(1)}_{k,j} + (\hat{\xi}_{k,i} - \hat{c}_i\xi_{k,i})(\hat{\varrho}^{(1)}_{k,j} - \hat{d}_j\varrho^{(1)}_{k,j}).$$

In the following, we will estimate the sums of the terms on the right-hand side. Notice first that the orthogonality of the $(w_j : j \in \mathbb{N})$ and (3.4) imply that

$$\varrho^{(1)}_{k,j} = \langle \rho_{k,1}, w_j \rangle = \sum_{i'=p+1}^{\infty} \sum_{j'=1}^{q} \psi_{i',j'} \xi_{k,i'} \langle w_{j'}, w_j \rangle = \sum_{i'=p+1}^{\infty} \psi_{i',j} \xi_{k,i'} = \langle X_k, u_j \rangle = \theta_{k,j}.$$

Therefore $\hat{c}_i \hat{d}_j \sum_{k=m+1}^{m+\ell} \xi_{k,i} \varrho^{(1)}_{k,j} = T^{(2)}_{m,m+\ell}(i,j)$. To study the remaining terms, we introduce

$$A^{(1)}_m(\ell) = \hat{c}_i \sum_{k=m+1}^{m+\ell} \xi_{k,i}(\hat{\varrho}^{(1)}_{k,j} - \hat{d}_j \varrho^{(1)}_{k,j}),$$

$$A^{(2)}_m(\ell) = \sum_{k=m+1}^{m+\ell} \hat{d}_j(\hat{\xi}_{k,i} - \hat{c}_i \xi_{k,i})\varrho^{(1)}_{k,j},$$

$$A^{(3)}_m(\ell) = \sum_{k=m+1}^{m+\ell} (\hat{\xi}_{k,i} - \hat{c}_i \xi_{k,i})(\hat{\varrho}^{(1)}_{k,j} - \hat{d}_j \varrho^{(1)}_{k,j}).$$

Since the $(v_i\colon i \in \mathbb{N})$ are the orthonormal eigenfunctions of the covariance operator $C$, its spectral decomposition yields that $\langle C(v_i), u_i\rangle$ for $i \le p$. Utilizing this, the definition of $\varrho_{k,j}^{(1)}$ and $\hat{\varrho}_{k,j}^{(1)}$, and a subsequent rearranging of terms leads to

$$\begin{aligned} A_m^{(1)}(\ell) &= \hat{c}_i \sum_{k=m+1}^{m+\ell} \sum_{j'=1}^{q} \langle X_k, v_i\rangle \langle X_k, u_{j'}\rangle \langle w_{j'}, \hat{w}_j - \hat{d}_j w_j\rangle \\ &= \hat{c}_i \sum_{j'=1}^{q} \left\langle \ell(\hat{C}_{m,m+\ell} - C)(v_i), u_{j'} \right\rangle \langle w_{j'}, \hat{w}_j - \hat{d}_j w_j\rangle \end{aligned}$$

where $\hat{C}_{m,m+\ell} = \frac{1}{\ell}\sum_{k=m+1}^{m+\ell} X_k \otimes X_k$ and $u_{j'}$ is defined in (3.4). Applications of the Cauchy-Schwarz inequality and Lemmas 5.1 and 5.2 yield

$$\sup_{\ell>1} \frac{|A_m^{(1)}(\ell)|}{\sqrt{\ell(\log \ell)^\beta}} \le \mathcal{I} \|\hat{w}_j - \hat{d}_j w_j\| \sup_{\ell>1} \sqrt{\frac{\ell}{(\log \ell)^\beta}} \|\hat{C}_{m,m+\ell} - C\|_{\mathcal{S}} = O_P\left(\frac{1}{\sqrt{m}}\right)$$

where $\mathcal{I} = \sum_{j'=1}^{q} \|u_{j'}\|$ is finite.

Using that $\varrho_{k,j}^{(1)} = \langle X_k, u_j\rangle$, write for the next term

$$\begin{aligned} A_m^{(2)}(\ell) &= \hat{d}_j \sum_{k=m+1}^{m+\ell} \langle X_k, \hat{v}_i - \hat{c}_i v_i\rangle \langle X_k, u_j\rangle \\ &= \hat{d}_j \ell \left\langle (\hat{C}_{m,m+\ell} - C)(\hat{v}_i - \hat{c}_i v_i), u_j \right\rangle + \hat{d}_j \ell \left\langle C(\hat{v}_i - \hat{c}_i v_i), u_j\right\rangle \\ &= A_{m,1}^{(2)}(\ell) + \ell R_m^{(1)}(i,j), \end{aligned}$$

where arguments similar to the ones applied to $A_m^{(1)}(\ell)$ also give that

$$\sup_{\ell>1} \frac{|A_{m,1}^{(2)}(\ell)|}{\sqrt{\ell(\log \ell)^\beta}} = O\left(\frac{1}{\sqrt{m}}\right).$$

Finally, along the same lines we get that

$$A_m^{(3)}(\ell) \le \mathcal{I} \|\hat{v}_i - \hat{c}_i v_i\| \|\hat{w}_j - \hat{d}_j w_j\| \|\ell \hat{C}_{m,m+\ell}\|_{\mathcal{S}}$$

with $\mathcal{I}$ from above, thus via Lemmas 5.1 and 5.2 implying that

$$\sup_{\ell>1} \frac{|A_m^{(3)}(\ell)|}{\ell} = \mathcal{I} \|\hat{v}_i - \hat{c}_i v_i\| \|\hat{w}_j - \hat{d}_j w_j\| \sup_{\ell>1} \|\hat{C}_{m,m+\ell}\|_{\mathcal{S}} = O_P\left(\frac{1}{m}\right).$$

Recognizing that we can set $U_{m,1}(\ell) = A_m^{(1)}(\ell) + A_{m,1}^{(2)}(\ell)$ and $U_{m,2}(\ell) = A_m^{(3)}(\ell)$, the proof of (S3.6) is complete. The proof of (S3.5) requires only minor modifications and is therefore omitted. □

**Lemma S3.3** *If Assumptions 2.1–2.5 hold, then*

$$\left|\sum_{k=1}^{m}\hat{\xi}_{k,i}\langle\rho_{k,2},\hat{w}_j\rangle - mR_m^{(2)}(i,j)\right| = O_P(1) \tag{S3.7}$$

*and*

$$\sum_{k=m+1}^{m+\ell}\hat{\xi}_{k,i}\langle\rho_{k,2},\hat{w}_j\rangle = \ell R_m^{(2)}(i,j) + U_{m,3}(\ell) + U_{m,4}(\ell), \tag{S3.8}$$

*where* $R_m^{(2)}(i,j) = \hat{c}_i\lambda_i\sum_{j'=q+1}^{\infty}\psi_{i,j'}\left\langle w_{j'},\hat{w}_j - \hat{d}_j w_j\right\rangle$,

$$\sup_{\ell>1}\frac{|U_{m,3}(\ell)|}{\sqrt{\ell(\log\ell)^{\beta}}} = O_P\left(\frac{1}{\sqrt{m}}\right) \qquad \text{and} \qquad \sup_{\ell>1}\frac{|U_{m,4}(\ell)|}{\ell} = O_P\left(\frac{1}{m}\right).$$

*Proof.* Let $\varrho_{k,j}^{(2)} = \langle\rho_{k,2},w_j\rangle$ and $\hat{\varrho}_{k,j}^{(2)} = \langle\rho_{k,2},\hat{w}_j\rangle$. Then, $\varrho_{k,j}^{(2)} = 0$ which follows from the orthogonality of the $(w_j\colon j\in\mathbb{N})$ as $\rho_{k,2}$ contains only $w_{j'}$ with $j'>q$, while $j\leq q$. Hence,

$$\hat{\xi}_{k,i}\hat{\varrho}_{k,j}^{(2)} = \hat{c}_i\xi_{k,i}(\hat{\varrho}_{k,j}^{(2)} - \hat{d}_j\varrho_{k,j}^{(2)}) + (\hat{\xi}_{k,i} - \hat{c}_i\xi_{k,i})(\hat{\varrho}_{k,j}^{(2)} - \hat{d}_j\varrho_{k,j}^{(2)}).$$

We proceed by estimating the sums of the terms on the right-hand side. Observe that

$$\begin{aligned}
A_m^{(4)}(\ell) &= \hat{c}_i\sum_{k=m+1}^{m+\ell}\xi_{k,i}(\hat{\varrho}_{k,j}^{(2)} - \hat{d}_j\varrho_{k,j}^{(2)})\\
&= \hat{c}_i\sum_{i'=1}^{\infty}\left\langle\ell(\hat{C}_{m,m+\ell}-C)(v_i),v_{i'}\right\rangle\sum_{j'=q+1}^{\infty}\psi_{i',j'}\left\langle w_{j'},\hat{w}_j-\hat{d}_j w_j\right\rangle\\
&\qquad + \ell\hat{c}_i\lambda_i\sum_{j'=q+1}^{\infty}\psi_{i,j'}\left\langle w_{j'},\hat{w}_j-\hat{d}_j w_j\right\rangle\\
&= A_{m,1}^{(4)}(\ell) + \ell R_m^{(2)}(i,j).
\end{aligned}$$

For the first term, the Cauchy-Schwarz inequality yields in combination with Lemmas 5.1 and 5.2 the estimate

$$\sup_{\ell>1}\frac{|A_{m,1}^{(4)}(\ell)|}{\sqrt{\ell(\log\ell)^{\beta}}} \leq \|\Psi\|_{\mathcal{S}}\sup_{\ell>1}\frac{1}{\sqrt{\ell(\log\ell)^{\beta}}}\ell\|\hat{C}_{m,m+\ell}-C\|_{\mathcal{S}}\|\hat{w}_j-\hat{d}_j w_j\|,$$

by recognizing that $\|\Psi\|_{\mathcal{S}}^2 = \sum_{i'=1}^{\infty}\sum_{j'=1}^{\infty}\psi_{i',j'}^2$.

As for $A_m^{(3)}(\ell)$ in the proof of Lemma S3.2 one can prove with the Cauchy-Schwarz inequality and Lemmas 5.1 and 5.2 that

$$\sup_{\ell>1}\frac{|A_m^{(5)}(\ell)|}{\ell} = \sup_{\ell>1}\frac{1}{\ell}\left|\sum_{k=m+1}^{m+\ell}(\hat{\xi}_{k,i}-\hat{c}_i\xi_{k,i})(\hat{\varrho}_{k,j}^{(2)} - \hat{d}_j\varrho_{k,j}^{(2)})\right| = O_P\left(\frac{1}{m}\right).$$

Setting $U_{m,3}(\ell) = A^{(4)}_{m,1}(\ell)$ and $U_{m,4}(\ell) = A^{(5)}_{m}(\ell)$, the proof of (S3.8) is complete. (S3.7) can be established similarly. □

The last two lemmas we state without proofs, as the arguments needed are now repetitive and very similar to the previous two lemmas.

**Lemma S3.4** *If Assumptions 2.1–2.5 hold, then*

$$\left|\sum_{k=1}^{m} \hat{\xi}_{k,i}\langle \phi_{k,1}, \hat{w}_j\rangle - mR^{(3)}_m(i,j)\right| = O_P(1) \tag{S3.9}$$

*and*

$$\sum_{k=m+1}^{m+\ell} \hat{\xi}_{k,i}\langle \phi_{k,1}, \hat{w}_j\rangle = \ell R^{(3)}_m(i,j) + U_{m,5}(\ell) + U_{m,6}(\ell) \tag{S3.10}$$

*where* $R^{(3)}_m(i,j) = \hat{c}_i\hat{d}_j\lambda_i \sum_{i'=1}^{p} \psi_{i',j}\langle \hat{c}_{i'}\hat{v}_{i'} - v_{i'}, v_i\rangle$,

$$\sup_{\ell>1} \frac{|U_{m,5}(\ell)|}{\sqrt{\ell(\log \ell)^\beta}} = O_P\left(\frac{1}{\sqrt{m}}\right) \qquad \textit{and} \qquad \sup_{\ell>1} \frac{|U_{m,6}(\ell)|}{\ell} = O_P\left(\frac{1}{m}\right).$$

**Lemma S3.5** *If Assumptions 2.1–2.5 hold, then*

$$\left|\sum_{k=1}^{m} \hat{\xi}_{k,i}\langle \phi_{k,2}, \hat{w}_j\rangle - mR^{(4)}_m(i,j)\right| = O_P(1) \tag{S3.11}$$

*and*

$$\sum_{k=m+1}^{m+\ell} \hat{\xi}_{k,i}\langle \phi_{k,2}, \hat{w}_j\rangle = \ell R^{(4)}_m(i,j) + U_{m,7}(\ell) + U_{m,8}(\ell), \tag{S3.12}$$

*where* $R^{(4)}_m(i,j) = \hat{c}_i\hat{d}_j\lambda_i \sum_{j'=1}^{q} \psi_{i,j'}\langle \hat{d}_{j'}\hat{w}_{j'} - \hat{d}_{j'}w_{j'}, w_j\rangle$,

$$\sup_{\ell>1} \frac{|U_{m,7}(\ell)|}{\sqrt{\ell(\log \ell)^\beta}} = O_P\left(\frac{1}{\sqrt{m}}\right) \qquad \textit{and} \qquad \sup_{\ell>1} \frac{|U_{m,8}(\ell)|}{\ell} = O_P\left(\frac{1}{m}\right).$$