Supplement to: Functional Principal Component Analysis of Spatially Correlated Data

Chong Liu
State Street Global Advisors, Boston
and
Surajit Ray
School of Mathematics and Statistics, University of Glasgow
and
Giles Hooker
Department of Statistical Science and
Department of Biological Statistics and Computational Biology, Cornell University

1 More on Consistency and Proof of Theorem 1^*

This section starts with a brief review of results in Yao et al. (2005) and then discusses the development of asymptotic properties of SPACE model components. In subsequent discussions, we change some notations used in Yao et al. (2005) to accommodate existing ones in our work. Specifically, We denote the number of curves by N whereas Yao et.al. used n. For the number of observations on curve i, we use n_i whereas Yao et al. (2005) used N_i. Yao et al. (2005) assumed N_i follows the distribution of random variable N which is denoted by N^* in this work. We denote all variables which depend on the BLUP of fPC scores by check sign instead of tilde used in Yao et al. (2005). We keep the same notations as in Yao et al. (2005) for other variables unless stated otherwise.

Yao et al. (2005) stated five theorems as well as a series of lemmas and corollaries to demonstrate asymptotic properties of PACE estimators. Theorem 1 established the uniform convergence rates in probability of local linear estimates of mean $\mu(t)$ and covariance function $G(s, t)$. Corollary 1 stated the convergence rate in probability of measurement error variance σ^2 and was derived by applying the results in Theorem 1. Theorem 2 established the convergence in probability of the eigen-structure of the covariance kernel, which included the convergence of eigenvalues λ_k of multiplicity 1, L^2 and uniform convergence of eigenfunctions $\phi_k(t)$'s associated with λ_k of multiplicity 1. Based on the results in Theorem 1 and 2, derived in Theorem 3 were the convergence of estimated fPC score $\hat{\xi}_{ik}$ towards the conditional expectation of true fPC score (BLUP) $\check{\xi}_{ik}$, and the pointwise convergence of reconstructed curves $\widehat{X}_{K,i}(t)$ based on K-dimensional approximation towards its associated infinite dimensional target $\check{X}_i(t)$ constructed from $\{\check{\xi}_{ik}\}_{i=1,k=1}^{N,\infty}$. With Gaussian assumption, Theorem 4 established the pointwise asymptotic distribution of the standardized deviance between reconstructed curve $\widehat{X}_{K,i}(t)$ and true curve $X_i(t)$. In Theorem 5, simultaneous confidence region of $X_{K,i}(t)$ over $\mathcal{T}$ in terms of reconstructed curves $\widehat{X}_{K,i}(t)$ was derived. Finally, the simultaneous region of K-dimensional vector $\boldsymbol{\xi}_{K,i}$ in terms of estimated fPC score vector $\widehat{\boldsymbol{\xi}}_{K,i}$ was derived in Corollary 2.

The above results held under a series of conditions and assumptions. We keep those assumptions and introduce new ones related to spatial correlation. Specifically, both observation time T_{ik} and the observed value $Y_{ik} = Y_i(T_{ik})$ are random variables. T_{ik} is assumed to follow the same marginal distribution T with density $f(t)$, data pair (T_{ik}, Y_{ik}) is assumed to follow the same joint distribution (T, Y) with density $g(t, y)$, for any given curve i and time index k. For each spatial separation vector $\boldsymbol{\Delta}$ and any location pairs $(i, j) \in \mathcal{N}(\boldsymbol{\Delta})$, we assume $(T_{ik}, T_{jl}, Y_{ik}, Y_{jl})$ follows the same distribution with density function $g_{\boldsymbol{\Delta}}(t_1, t_2, y_1, y_2)$. We assume the same regularity conditions on $g_{\boldsymbol{\Delta}}(t_1, t_2, y_1, y_2)$ as those on $g(t_1, t_2, y_1, y_2)$ in Yao et al. (2005).In addition, observation times T_{ik} is assumed to be independent across curves and within individual curves.

Moreover, Yao et al. (2005) also described three groups of conditions. In group A, series (A1) pertained to the number of observations per curve. It was assumed to be a discrete random variable following the distribution N^*. N^* was assumed to have finite mean and takes value greater than 1 with positive probability, and was independent of all T_{ik}'s and Y_{ik}'s. Specifically, assumption (A1.1) and (A1.2) state the following.

(A1.1)

The number of observations made for the ith subject or cluster is a random variable denoted by n_i with $n_i \overset{i.i.d.}{\sim} N^*$, where $N^* > 0$ is a positive discrete random variable, with

$EN^* < \infty$ and $P\{N^* > 1\} > 0$.

(A1.2)

The observation times and measurements are assumed to be independent of the number of measurements, i.e., for any subset $J_i \subseteq \{1, \cdots, n_i\}$ and for all $i = 1, \cdots, N$, $(\{T_{ij} : j \in J_i\}, \{Y_{ij} : j \in J_i\})$ is independent of n_i.

To accommodate spatial correlation, we need the following in addition in (A1.2). For any subset $J_k \subseteq \{1, \cdots, n_k\}$ and $J_l \subseteq \{1, \cdots, n_l\}$ and for all $k, l = 1, \cdots, N$, $(\{T_{kj} : j \in J_k\}, \{T_{lj} : j \in J_l\}, \{Y_{kj} : j \in J_k\}, \{Y_{lj} : j \in J_l\})$ is independent of n_k and n_l. Since the modification is marginal, we don't list it separately in a different assumption.

Series (A2) specified the convergence rates of smoothing bandwidths and their relative rates with respect to the number of curves N. This series of conditions ensures the convergence rate derived in Theorem 1. As $N \to \infty$, the original (A2) series requires the following.

(A2.1) $h_\mu \to 0$, $Nh_\mu^4 \to \infty$ and $Nh_\mu^6 < \infty$.

(A2.2) $h_G \to 0$, $Nh_G^6 \to \infty$ and $Nh_G^8 < \infty$.

(A2.3) $h_V \to 0$, $Nh_V^4 \to \infty$ and $Nh_V^6 < \infty$

Series (A3) and (A4) assumed favorable properties of smoothing kernel functions and finite fourth centered moments of Y, such that the auxiliary results in Lemma A.1 and A.2 held. Define the Fourier transforms of $\kappa_1(u), \kappa_2(u, v)$ by $\zeta_1(t) = \int e^{-iut}\kappa_1(u)du$ and $\zeta_2(t, s) = \int e^{-(iut+ivs)}\kappa_2(u, v)dudv$. Original (A3) and (A4) state the following.

(A3.1) $\zeta_1(t)$ is absolutely integrable, i.e., $\int |\zeta_1(t)dt| < \infty$

(A3.2) $\zeta_2(t, s)$ is absolutely integrable, i.e., $\int\int |\zeta_2(t, s)dtds| < \infty$

(A4) $E\left[(Y - \mu(T))^4\right] < \infty$

Condition (A5) assumed Gaussian distribution of fPC scores required by Theorem 4 and 5 and Corollary 2. Condition (A6) assumed that the data asymptotically follow a linear scheme. Condition (A7) assumed that the pointwise limit of $\boldsymbol{\phi}_{K,s}\boldsymbol{\Sigma}(\check{\boldsymbol{\xi}}_{K,i} - \boldsymbol{\xi}_{K,i})\boldsymbol{\phi}_{K,t} \triangleq \boldsymbol{\phi}_{K,s}\boldsymbol{\Omega}_K\boldsymbol{\phi}_{K,t}$ as $K \to \infty$ existed such that Theorem 4 and 5 held.

Group B specifies properties of density and smoothing kernel functions for Lemma A.1 and A.2 to hold. Specifically, original group B conditions state the following.

(B1.1) $(d^l/dt^l)f(t)$ exists and is continuous on $\mathcal{T}$ with $f(t) > 0$ on $\mathcal{T}$;

(B1.2) $(d^l/dt^l)g(t,y)$ exists and is uniformly continuous on $\mathcal{T} \times \mathcal{R}$;

(B1.3) $(d^l/dt_1^{l_1}dt_2^{l_2})g(t_1,t_2,y_1,y_2)$ exists and is uniformly continuous on $\mathcal{T}^2 \times \mathcal{R}^2$;

(B2.1a) kernel κ_1 is compactly supported, $\|\kappa_1\|^2 = \int \kappa_1^2(u)du < \infty$;

(B2.2a) κ_1 is a kernel function of order (ν, l);

(B2.1b) κ_2 is compactly supported, $\|\kappa_2\|^2 = \int\int \kappa_2^2(u,v)dudv < \infty$;

(B2.2b) κ_2 is a kernel function of order (ν, l).

Group C assumed favorable continuity and boundedness properties of functions ψ_p and θ_p crucial for Lemma A.1 and A.2. We keep the same set of assumptions for the development of our results.

As the cornerstone of all other results, Theorem 1 was derived based on Lemma A.1 and A.2. Theorem 1* in Section 4 of the paper extends Theorem 1 and serves as the foundation of our work.To prove Theorem 1*, we first adapt Lemma A.1 and A.2 to correlated functional data. To that end, we require $\psi_p(t)$ and $\theta_p(t,s,y_1,y_2)$ to satisfy the following additional conditions respectively:

(C3.1a)

$$\frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} E\left|Cov\big(\psi_p(T_{i_1},Y_{i_1}),\psi_p(T_{i_2},Y_{i_2})\ \big|\ T_{i_1},T_{i_2}\big)\right| \to 0$$

and

(C3.1b)

$$\frac{1}{N(\mathbf{\Delta})^2}\sum_{\substack{(i_1,j_1)\in\mathcal{N}(\mathbf{\Delta})\\(i_2,j_2)\in\mathcal{N}(\mathbf{\Delta})\\(i_1,j_1)\ne(i_2,j_2)}} E\left|Cov\big(\theta_p(T_{i_1},T_{j_1},Y_{i_1},Y_{j_1}),\theta_p(T_{i_2},T_{j_2},Y_{i_2},Y_{j_2})\ \big|\ T_{i_1},T_{i_2},T_{j_1},T_{j_2}\big)\right| \to 0$$

Counterparts of Lemma A.1 and Lemma A.2 are stated respectively in Lemma A.1* and Lemma A.2*.

*Lemma A.1** Under (A1.1),(A1.2),(A3.1),(B1.1),(B1.2),(B2.1a),(B2.2a),(C1.1a)-C(1.3a),(C2.1a) in Yao et al. (2005) and (C3.1a), $\tau_{pn} = \sup_{t\in\mathcal{T}}|\Psi_{pN} - \mu_p| = O_p(1/(\sqrt{N}h_\mu^{\nu+1}))$

Proof. The key step of achieving the stated rate is to show that $Var(\varphi_{pN}(u)) = O(1/N)$. With the presence of spatial correlation, we have,

$$Var(\varphi_{pN}(u)) \quad = \quad \frac{1}{N^2}Var\left(\sum_{i=1}^{N}\frac{1}{EN^*}\sum_{k=1}^{n_i}e^{-iuT_{ik}}\psi_p(T_{ik},Y_{ik})\right)$$

$$
\begin{aligned}
&= \frac{1}{N} Var\left(\frac{1}{EN^*}\sum_{k=1}^{N^*} e^{-iuT_k}\psi_p(T_k,Y_k)\right) \\
&\quad + \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} Cov\Big(\frac{1}{EN^*}\sum_{k_1=1}^{n_{i_1}} e^{-iuT_{i_1k_1}}\psi_p(T_{i_1k_1},Y_{i_1k_1}), \\
&\qquad\qquad \frac{1}{EN^*}\sum_{k_2=1}^{n_{i_2}} e^{-iuT_{i_2k_2}}\psi_p(T_{i_2k_2},Y_{i_2k_2})\Big) \\
&\triangleq \frac{1}{N} Var\left(\frac{1}{EN^*}\sum_{k}^{N^*} e^{-iuT_k}\psi_p(T_k,Y_k)\right) + \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} Cov(Z_{i_1},Z_{i_2}) \\
&\le \frac{1}{N} E\left[\left(\frac{1}{EN^*}\sum_{k=1}^{N^*} e^{-iuT_k}\psi_p(T_k,Y_k)\right)^2\right] + \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} Cov(Z_{i_1},Z_{i_2}) \\
&\le \frac{1}{N(EN^*)^2} E\left[\left(\sum_{k=1}^{N^*} e^{-i2uT_k}\right)\left(\sum_{k=1}^{N^*}\psi_p^2(T_k,Y_k)\right)\right] + \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} Cov(Z_{i_1},Z_{i_2}) \\
&\le \frac{1}{N(EN^*)^2} E\left[N^*\left(\sum_{k=1}^{N^*}\psi_p^2(T_k,Y_k)\right)\right] + \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} Cov(Z_{i_1},Z_{i_2}) \\
&= \frac{EN^{*2}}{N(EN^*)^2} E(\psi_p^2(T,Y)) + \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} Cov(Z_{i_1},Z_{i_2})
\end{aligned}
$$

The first term in the last line is $O(1/N)$ provided that $E(\psi_p^2(T,Y))<\infty$ and $E(N^*)^2/(EN^*)^2$ is O(1) as $N\to\infty$. $\sum_{1\le i_1\ne i_2\le N} Cov(Z_{i_1},Z_{i_2})/N^2$ is the extra term due to spatial correlations. Assuming (C3.1a) we have

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} |Cov(Z_{i_1},Z_{i_2})| \\
&= \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} \Big|Cov\Big(\frac{1}{EN^*}\sum_{k_1=1}^{n_{i_1}} e^{-iuT_{i_1k_1}}\psi_p(T_{i_1k_1},Y_{i_1k_1}), \frac{1}{EN^*}\sum_{i_2=1}^{n_{i_2}} e^{-iuT_{i_2k_2}}\psi_p(T_{i_2k_2},Y_{i_2k_2})\Big)\Big| \\
&= \frac{EN^{*2}}{N^2(EN^*)^2}\sum_{1\le i_1\ne i_2\le N} \Big|Cov\left(e^{-iuT_{i_1}}\psi_p(T_{i_1},Y_{i_1}), e^{-iuT_{i_2}}\psi_p(T_{i_2},Y_{i_2})\right)\Big| \\
&\triangleq \frac{EN^{*2}}{N^2(EN^*)^2}\sum_{1\le i_1\ne i_2\le N} \Big|Cov\left(W_{i_1},W_{i_2}\right)\Big| \\
&= \frac{EN^{*2}}{N^2(EN^*)^2}\sum_{1\le i_1\ne i_2\le N} \Big|E\big(Cov(W_{i_1},W_{i_2})|T_{i_1},T_{i_2}\big) + Cov\big(E(W_{i_1}|T_{i_1},T_{i_2}), E(W_{i_2}|T_{i_1},T_{i_2})\big)\Big| \\
&= \frac{EN^{*2}}{N^2(EN^*)^2}\sum_{1\le i_1\ne i_2\le N} \Big|E\big(e^{-iu(T_{i_1}+T_{i_2})}Cov(\psi_p(T_{i_1},Y_{i_1}),\psi_p(T_{i_2},Y_{i_2})|T_{i_1},T_{i_2})\big)
\end{aligned}
$$

$$
\begin{aligned}
& + Cov\Big(e^{-iuT_{i_1}}E(\psi_p(T_{i_1},Y_{i_1})|T_{i_1}), e^{-iuT_{i_2}}E(\psi_p(T_{i_2},Y_{i_2})|T_{i_2})\Big)\Big| \\
= & \frac{EN^{*2}}{N^2(EN^*)^2}\sum_{1\le i_1\ne i_2\le N}\Big|E\Big(e^{-iu(T_{i_1}+T_{i_2})}Cov(\psi_p(T_{i_1},Y_{i_1}),\psi_p(T_{i_2},Y_{i_2})|T_{i_1},T_{i_2})\Big)\Big| \\
\le & \frac{EN^{*2}}{N^2(EN^*)^2}\sum_{1\le i_1\ne i_2\le N}E\Big|Cov(\psi_p(T_{i_1},Y_{i_1}),\psi_p(T_{i_2},Y_{i_2})|T_{i_1},T_{i_2})\Big| = O(1/N)
\end{aligned}
$$

Lemma A.2* is designed to cover cross-covariance smoothers as well and thus define

$$
\begin{aligned}
\Theta_{p\boldsymbol{\Delta}} &= \Theta_{p\boldsymbol{\Delta}}(t,s) \\
&= \frac{1}{N(\boldsymbol{\Delta})h_{G_{\boldsymbol{\Delta}}}^{|\nu|+2}}\sum_{(i,j)\in\mathcal{N}(\boldsymbol{\Delta})}\frac{1}{EN^*(EN^*-1)} \\
&\quad\times\sum_{1\le k\le n_i}\sum_{1\le l\le n_j}\theta_p(T_{ik},T_{jl},Y_{ik},Y_{jl})\kappa_2\left(\frac{t-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}},\frac{s-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right)
\end{aligned}
$$

*Lemma A.2** Under (A1.1),(A1.2),(A3.2),(B1.1),(B1.2),(B2.1b),(B2.2b),(C1.1b)-C(1.3b),(C2.1b) in Yao et al. (2005) and (C3.1b), $\vartheta_{p\boldsymbol{\Delta}} = \sup_{t,s\in\mathcal{T}}|\Theta_{p\boldsymbol{\Delta}}-\varrho_p| = O_p\left(1/\left(\sqrt{N(\boldsymbol{\Delta})}h_{G_{\boldsymbol{\Delta}}}^{|\nu|+2}\right)\right)$

Proof. This is analogous to the proof of Lemma A.1*.

Equipped with Lemma A.1* and A.2*, we prove Theorem 1* of the paper next.

Proof of Theorem 1*

This proof adopts the same techniques employed in Yao et al. (2005) for the main reasoning process and we focus on the part where spatial correlation makes a difference. We first look at the convergence of local linear estimator $\hat{\mu}(t)$ of mean function to its true value $\mu(t)$. $\hat{\mu}(t)$ is written as

$$
\begin{aligned}
\hat{\mu}(t) &= \hat{\beta}_0(t) \\
&= \frac{\sum_i \frac{1}{EN^*}\sum_k w_{ik}Y_{ik}}{\sum_i\frac{1}{EN^*}\sum_k w_{ik}} - \frac{\sum_i\frac{1}{EN^*}\sum_k w_{ik}(T_{ik}-t)}{\sum_i\frac{1}{EN^*}\sum_k w_{ik}}\hat{\beta}_1(t), \qquad (1.1)
\end{aligned}
$$

where

$$
\hat{\beta}_1(t) = \frac{\sum_i\frac{1}{EN^*}\sum_k w_{ik}(T_{ik}-t)Y_{ik} - \frac{\sum_i\frac{1}{EN^*}\sum_k w_{ik}(T_{ik}-t)\sum_i\frac{1}{EN^*}\sum_k w_{ik}Y_{ik}}{\sum_i\frac{1}{EN^*}\sum_k w_{ik}}}{\sum_i\frac{1}{EN^*}\sum_k w_{ik}(T_{ik}-t)^2 - \frac{\left(\sum_i\frac{1}{EN^*}\sum_k w_{ik}(T_{ik}-t)\right)^2}{\sum_i\frac{1}{EN^*}\sum_k w_{ik}}},
$$

and

$$w_{ik} = \frac{1}{Nh_\mu}\kappa_1\left(\frac{t-T_{ik}}{h_\mu}\right)$$

Let the first term in (1.1) be $\hat{\mu}_{NW}(t)$ and $\hat{f}(t) = \sum_i \sum_k w_{ik}/EN^*$ be an estimator of T's density function $f(t)$. Key steps in Yao et al. (2005) are showing

$$\sup_{t\in\mathcal{T}} |\hat{\mu}_{NW}(t) - \mu(t)| = O_p(1/\sqrt{N}h_\mu)$$
$$\sup_{t\in\mathcal{T}} |\hat{f}(t) - f(t)| = O_p(1/\sqrt{N}h_\mu)$$
$$\sup_{t\in\mathcal{T}} |\hat{\beta}_1(t) - \mu'(t)| = O_p(1/\sqrt{N}h_\mu^2)$$

The convergence of $\hat{f}(t)$ can be shown by applying Lemma A.1* on $\psi_2(u,y) \equiv 1$ with κ_1 of order (0,2); The convergence of $\hat{\mu}_{NW}(t)$ follows by applying Lemma A.1* on $\psi_1(u,y) = y$ with κ_1 of order (0,2) and the convergence of $\hat{f}(t)$; The convergence of $\hat{\beta}_1(t)$ can be established by applying Lemma A.1* on $\psi_1(u,y) = y$, $\psi_2(u,y) \equiv 1$ and $\psi_3(u,y) = u - t$ with $\tilde{\kappa}_1 = -t\kappa_1(t)/\int u^2\kappa_1(u)du$ of order (1,3). We show below that under (D1), $\psi_1(u,y)$, $\psi_2(u,y)$ and $\psi_3(u,y)$ all satisfy (C3.1a).

$$\begin{aligned}
& \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} E\big|Cov(Y_{i_1}, Y_{i_2}|T_{i_1}, T_{i_2})\big| \\
= \; & \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} E\big|\boldsymbol{\phi}'(T_{i_1})Cov(\boldsymbol{\xi}_{i_1}, \boldsymbol{\xi}_{i_2})\boldsymbol{\phi}(T_{i_2})|T_{i_1}, T_{i_2}\big| \\
= \; & \frac{1}{N^2}\sum_{1\le i_1\ne i_2\le N} E\big|\boldsymbol{\phi}'(T_{i_1})diag(\rho_1^*(\boldsymbol{\Delta}_{i_1i_2}), \rho_2^*(\boldsymbol{\Delta}_{i_1i_2}), \cdots, \rho_K^*(\boldsymbol{\Delta}_{i_1i_2}))\boldsymbol{\phi}(T_{i_2})|T_{i_1}, T_{i_2}\big| \\
\le \; & \frac{1}{N^2}\max_{t\in\mathcal{T}}(|\boldsymbol{\phi}(t)|^2)\sum_{1\le i_1\ne i_2\le N}\sum_{k=1}^K |\rho_k^*(\boldsymbol{\Delta}_{i_1i_2})| \to 0
\end{aligned}$$

The last line holds under (D1). Furthermore, noting that

$$\sum_{1\le i_1\ne i_2\le N} E\big|Cov(1, 1|T_{i_1}, T_{i_2})\big| = 0$$
$$\sum_{1\le i_1\ne i_2\le N} E\big|Cov(T_{i_1} - t, T_{i_2} - t|T_{i_1}, T_{i_2})\big| = 0$$

The result (4.1) in Theorem 1* in the paper follows.

We proceed to show (4.2) of the paper. Define

$$\kappa_2\left(\frac{s-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}}, \frac{t-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right) = W_{ikjl}$$

$$
\begin{aligned}
&\kappa_2\left(\frac{s-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}}, \frac{t-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right)(s-T_{ik}) = B1_{ikjl} \\
&\kappa_2\left(\frac{s-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}}, \frac{t-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right)(t-T_{jl}) = B2_{ikjl} \\
&\kappa_2\left(\frac{s-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}}, \frac{t-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right)(s-T_{ik})^2 = C1_{ikjl} \\
&\kappa_2\left(\frac{s-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}}, \frac{t-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right)(t-T_{jl})^2 = C2_{ikjl} \\
&\kappa_2\left(\frac{s-T_{ik}}{h_{G_{\boldsymbol{\Delta}}}}, \frac{t-T_{jl}}{h_{G_{\boldsymbol{\Delta}}}}\right)(s-T_{ik})(t-T_{jl}) = E_{ikjl} \\
&D_{ij}(t_{ik}, t_{jl}) = (Y_i(t_{ik}) - \hat{\mu}(t_{ik}))(Y_j(t_{jl}) - \hat{\mu}(t_{jl})) = D_{ikjl} \\
&D_{ij}(t_{ik}, t_{jl}) = (Y_i(t_{ik}) - \mu(t_{ik}))(Y_j(t_{jl}) - \mu(t_{jl})) = D^0_{ikjl}
\end{aligned}
$$

Denote $\sum_{(i,j)\in\mathcal{N}(\boldsymbol{\Delta})}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}$ by $\sum_{ikjl}$. The local linear estimator $\widehat{G}_{\boldsymbol{\Delta}}(s,t)$ of the cross-covariance $G_{\boldsymbol{\Delta}}(s,t)$ based on D^0_{ikjl} can be written as

$$
\begin{aligned}
\widehat{G}_{\boldsymbol{\Delta}}(s,t) &= \hat{\beta}_0(s,t) \\
&= \frac{\sum_{ikjl} W_{ikjl} D^0_{ikjl}}{\sum_{ikjl} W_{ikjl}} - \hat{\beta}_1(s,t)\frac{\sum_{ikjl} B1_{ikjl}}{\sum_{ikjl} W_{ikjl}} - \hat{\beta}_2(s,t)\frac{\sum_{ikjl} B2_{ikjl}}{\sum_{ikjl} W_{ikjl}}, \qquad (1.2)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\beta}_1(s,t) &= \frac{\sum W \sum B1D^0 - \sum B1 \sum WD^0 - \hat{\beta}_2(s,t)(\sum E \sum W - \sum B1 \sum B2)}{\sum W \sum C1 - (\sum B1)^2} \\
\hat{\beta}_2(s,t) &= \frac{\sum W \sum B2D^0 - \sum B2 \sum WD^0 - \hat{\beta}_1(s,t)(\sum E \sum W - \sum B1 \sum B2)}{\sum W \sum C1 - (\sum B2)^2}.
\end{aligned}
$$

Let

$$
\begin{aligned}
&\sum W \sum C1 - (\sum B1)^2 = a \\
&\sum W \sum C2 - (\sum B2)^2 = b \\
&\sum W \sum B1D^0 - \sum B1 \sum WD^0 = c \\
&\sum W \sum B2D^0 - \sum B2 \sum WD^0 = d \\
&\sum W \sum E - \sum B1 \sum B2 = e
\end{aligned}
$$

We have

$$
\hat{\beta}_1(s,t) = \frac{bc - de}{ab - e^2} = \left(\tilde{\beta}_1 - \frac{de}{ab}\frac{ab}{ab - e^2}\right)
$$

$$\hat{\beta}_2(s,t) = \frac{ad-ce}{ab-e^2} = \left(\tilde{\beta}_2 - \frac{ce}{ab}\frac{ab}{ab-e^2}\right).$$

Denote the first term in (1.2) by $\widehat{G}_{\boldsymbol{\Delta}NW}(s,t)$ and $\sum W$ by $\hat{f}(s,t)$. Analogous to the proof of (4.1) in the paper, we would like to show the following:

$$\begin{aligned}
\sup_{s,t\in\mathcal{T}} |\widehat{G}_{\boldsymbol{\Delta}NW}(s,t) - G_{\boldsymbol{\Delta}}(s,t)| &= O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^2_{G_{\boldsymbol{\Delta}}}) \\
\sup_{s,t\in\mathcal{T}} |\hat{f}(s,t) - f(s,t)| &= O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^2_{G_{\boldsymbol{\Delta}}}) \\
\sup_{s,t\in\mathcal{T}} |\hat{\beta}_1(s,t) - \frac{\partial G_{\boldsymbol{\Delta}}(s,t)}{\partial s}| &= O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^3_{G_{\boldsymbol{\Delta}}}) \\
\sup_{s,t\in\mathcal{T}} |\hat{\beta}_2(s,t) - \frac{\partial G_{\boldsymbol{\Delta}}(s,t)}{\partial t}| &= O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^3_{G_{\boldsymbol{\Delta}}})
\end{aligned}$$

and

$$D^0 \text{ is equivalent to } D.$$

Define

$$\begin{aligned}
\theta_1(u,v,y_1,y_2) &= (y_1-\mu(u))(y_2-\mu(v)) \\
\theta_2(u,v,y_1,y_2) &\equiv 1 \\
\theta_3(u,v,y_1,y_2) &= s-u \\
\theta_4(u,v,y_1,y_2) &= t-v
\end{aligned}$$

The convergence of $\hat{f}(s,t)$ follows by applying Lemma A.2* on θ_2 with κ_2 of order ((0,0),2). The convergence of $\widehat{G}_{\boldsymbol{\Delta}NW}(s,t)$ can be established based on the convergence of $\hat{f}(s,t)$ and by applying Lemma A.2* on θ_1 with κ_2. The convergence of $\hat{\beta}_1(s,t)$ and $\hat{\beta}_2(s,t)$ can be established by showing $|e| = O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^4_{G_{\boldsymbol{\Delta}}})$. Applying Lemma A.2* on θ_2 with kernel $\dot{\kappa}_2(s,t) = \dfrac{st\kappa_2(s,t)}{\int u^2v^2\kappa_2(u,v)dudv}$ of order ((1,1),4), we obtain $|E - f'(s)f'(t)| = O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^4_{G_{\boldsymbol{\Delta}}})$. Applying Lemma A.2* on θ_2 with kernel $\tilde{\kappa}_2(s,t) = \dfrac{s\kappa_2(s,t)}{\int u^2\kappa_2(u,v)dudv}$ of order ((1,0),3), we obtain $|B1 - f'(s)f(t)| = O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^4_{G_{\boldsymbol{\Delta}}})$. Applying Lemma A.2* on θ_2 with kernel $\check{\kappa}_2(s,t) = \dfrac{t\kappa_2(s,t)}{\int v^2\kappa_2(u,v)dudv}$ of order ((0,1),3), we obtain $|B2 - f(s)f'(t)| = O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^4_{G_{\boldsymbol{\Delta}}})$. By Slutsky's theorem, we have $|e| = \left|\sum W \sum E - \sum B1 \sum B2\right| = O_p(1/\sqrt{N(\boldsymbol{\Delta})}h^4_{G_{\boldsymbol{\Delta}}})$. The equivalence between D^0 and D can also be derived in a similar way by applying Lemma A.2* on θ_1, θ_3 and θ_4. We show below that under (D2), $\theta_1, \theta_2, \theta_3$ and θ_4 all satisfy (C3.1b).

$$\begin{aligned}
& \frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} E\left|Cov\Big(\big(Y_{i_1}-\mu(T_{i_1})\big)\big(Y_{j_1}-\mu(T_{j_1})\big), \big(Y_{i_2}-\mu(T_{i_2})\big)\big(Y_{j_2}-\mu(T_{j_2})\big) \mid T_{i_1}, T_{i_2}, T_{j_1}, T\right. \\
= & \frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} E\left|Cov\Big(\boldsymbol{\phi}(T_{i_1})'\boldsymbol{\xi}_{i_1}\boldsymbol{\xi}_{j_1}'\boldsymbol{\phi}(T_{j_1}), \boldsymbol{\phi}(T_{i_2})'\boldsymbol{\xi}_{i_2}\boldsymbol{\xi}_{j_2}'\boldsymbol{\phi}(T_{j_2}) \mid T_{i_1}, T_{i_2}, T_{j_1}, T_{j_2}\Big)\right| \\
\leq & \frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} \sum_{k=1}^{K} E\left|Cov\Big(\phi_k(T_{i_1})\xi_{i_1k}\xi_{j_1k}\phi_k(T_{j_1}), \phi_k(T_{i_2})\xi_{i_2k}\xi_{j_2k}\phi_k(T_{j_2}) \mid T_{i_1}, T_{i_2}, T_{j_1}, T_{j_2}\Big)\right| \\
\leq & \frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} \sum_{k=1}^{K} \max_{t\in\mathcal{T}}(|\boldsymbol{\phi}(t)|^4)\Big|Cov(\xi_{i_1k}\xi_{j_1k}, \xi_{i_2k}\xi_{j_2k})\Big| \\
= & \frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} \sum_{k=1}^{K} \max_{t\in\mathcal{T}}(|\boldsymbol{\phi}(t)|^4)\Big|E(\xi_{i_1k}\xi_{j_1k}\xi_{i_2k}\xi_{j_2k}) - \big(\rho_k^*(\boldsymbol{\Delta})\big)^2\Big| \\
= & \frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} \sum_{k=1}^{K} \max_{t\in\mathcal{T}}(|\boldsymbol{\phi}(t)|^4)\Big|\rho_k^*(\boldsymbol{\Delta}_1)^2 + \rho_k^*(\boldsymbol{\Delta}_2)\rho_k^*(\boldsymbol{\Delta}_3)\Big| \\
\to & \ 0
\end{aligned}$$

where $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ and $\boldsymbol{\Delta}_3$ represent spatial separation vectors of location pairs (i_1, i_2), (i_1, j_2) and (j_1, i_2).

$$\frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} E\left|Cov\Big(\big(s-T_{i_1}\big), \big(s-T_{i_2}\big) \mid T_{i_1}, T_{i_2}, T_{j_1}, T_{j_2}\Big)\right| = 0$$

$$\frac{1}{N(\boldsymbol{\Delta})^2} \sum_{\substack{(i_1,j_1)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_2,j_2)\in\mathcal{N}(\boldsymbol{\Delta}) \\ (i_1,j_1)\neq(i_2,j_2)}} E\left|Cov\Big(\big(t-T_{j_1}\big), \big(s-T_{j_2}\big) \mid T_{i_1}, T_{i_2}, T_{j_1}, T_{j_2}\Big)\right| = 0$$

Therefore, the result in (4.2) of the paper follows.

2 R-Code

The R-codes for implementing all the analyses and simulations described in the paper are available at http://www.stats.gla.ac.uk/~sray/software/.

Bibliography

Yao, F., Müller, H. & Wang, J. (2005), 'Functional data analysis for sparse longitudinal data', Journal of the American Statistical Association **100**(470), 577–590.