**Supplementary Material for:
Longitudinal Functional Data Analysis**

SY Park and AM Staicu

Department of Statistics, North Carolina State University

July 18, 2015

Section S1 summarizes the proposed estimation method in steps and provides a list available softwares for implementation. Section S2 provides detailed proofs of the theoretical results given in Section 4. Specifically Section S2.1 corresponds to Section 4.1 and includes proofs of the theoretical results for the case when $Y_{ij}(s) = X_{ij}(s)$. Whereas Section S2.2 corresponds to Section 4.2 and includes proofs for the case when $Y_{ij}(s) = X_{ij}(s) + \epsilon_{1,ij}(s)$. Section S3 includes additional simulation results for the case when data are corrupted only with white noise error, and results obtained using Chen & Müller (2012)'s method. Section S4 includes additional figures for the DTI data analysis.

R code implementing the simulation study is available at http://www4.stat.ncsu.edu/~staicu/software/MLFD_Sims_Rcode.zip. An illustration of the main ideas is available at http://www4.stat.ncsu.edu/~staicu/software/MLFD_Rcode.zip

S1. Implementation using available softwares

An important advantage of the proposed approach is that its implementation can be carried using available software.

Step 1. Estimate the smooth mean function $\widehat{\mu}(s, T)$ using the sandwich smoother (Xiao et al., 2013) (the `fbps` function in `R` (R Core Team, 2014) package `refund` (Ciprian Crainiceanu et al., 2014)) or using the penalized tensor product spline smoothing (the `gam` and `te` functions in `R` (R Core Team, 2014) package `mgcv` (Wood, 2011)).

Step 2. Estimate the smooth covariance function $\widehat{\Xi}(s, s')$ with the demeaned data (dense design) using the sandwich smoother (Xiao et al., 2015) and get the eigenfunctions $\widehat{\phi}_k(s)$ using the `fpca.face` function in the `refund` package (Ciprian Crainiceanu et al., 2014). The default option of this function also provides $\widetilde{\xi}_{W,ijk}$'s.

Step 3. For each k, carry out FPCA of $\{T_{ij}, \widetilde{\xi}_{W,ijk} : i, j\}$. There are several available options for implementation: `fpca.sc` function in the `refund` package (Ciprian Crainiceanu et al., 2014) and `fpca.mle` and `fpca.pred` functions in the `FPCA` package (Peng & Paul, 2009; James et al., 2000) in `R`. Alternatively one can use the `FPCA` function (Yao et al., 2005) in the `MATLAB` (MATLAB, 2014) package `PACE` (Yao et al., 2005) available at `http://www.stat.ucdavis.edu/PACE/`.

Step 4. Determine the predicted trajectories, $\widehat{Y}_i(s, T) = \widehat{\mu}(s, T) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \widehat{\zeta}_{ikl} \widehat{\psi}_{kl}(T) \widehat{\phi}_k(s)$. .

S2. Proofs of theoretical results given in Section 4

S2.1. Case when response curves are measured without error (Section 4.1)

Here we consider the case when $Y_{ij}(s) = X_{ij}(s)$; thus in this section, $Y(s, T)$ satisfies all of the assumptions made on $X(s, T)$. We first show the following corollaries:

Corollary 4.0.1
Under the assumptions (A1.) and (A2.), the marginal covariance function, $\Sigma(s, s') = \int c\{(s, T), (s', T)\}g(T)dT$, (i) is symmetric, (ii) is positive definite, and (iii) has eigenvalues satisfying that $\sum_{k=1}^{\infty} \lambda_k$ is finite.

Proof. We assume that $X_i(s, T)$ is a realization of a random process, X, that satisfies the assumption (A1.). By definition, the covariance operator of X is $C(y) = E[< X, y > X]$, $y \in L^2(\mathcal{S} \times \mathcal{T})$, and it follows that

$$C(y)(s, T) = \iint c((s, T), (s', T'))y(s', T')ds'dT',$$

where $c((s, T), (s', T')) = E[X(s, T)X(s', T')]$. Since the operator, C, is a properly defined covariance operator, it (i) is symmetric, i.e. $c((s, T), (s', T')) = c((s', T'), (s, T))$, (ii) is positive definite, i.e. $\iiiint c((s, T), (s', T'))y(s, T)y(s', T')dsds'dTdT'$, and (iii) has the sum of its eigenvalues being finite, i.e. $\iint c((s, T), (s, T))dsdT < \infty$ (Horváth & Kokoszka, 2012). In the following, we show that $\Sigma(s, s') = \int c\{(s, T), (s', T)\}g(T)dT$ has the same properties as $c((s, T), (s', T'))$ has, and thus is also a proper covariance function.

(i) $\Sigma(s, s')$ is symmetric.

$$\begin{aligned}\Sigma(s, s') &= \int c\{(s, T), (s', T)\}g(T)dT = \int E\{X_i(s, T)X_i(s', T)\}g(T)dT \\ &= \int E\{X_i(s', T)X_i(s, T)\}g(T)dT = \int c\{(s', T), (s, T)\}g(T)dT = \Sigma(s', s)\end{aligned} \tag{S1}$$

(ii) $\Sigma(s, s')$ is positive definite.

To show $\Sigma(s, s')$ is positive definite we need to prove the following holds:

$$\begin{aligned}\iint \Sigma(s, s')z(s)z(s')dsds' &= \iint \Big[\int c\{(s, T), (s', T)\}g(T)dT\Big]z(s)z(s')dsds' \\ &\overset{\text{Fubini's Th}}{=} \int g(T)\Big[\iint c\{(s, T), (s', T)\}z(s)z(s')dsds'\Big]dT \geq 0.\end{aligned}$$

And because a density function, $g(T)$, is non-negative by definition, it is sufficient to show that $\iint c\{(s, T), (s', T)\}z(s)z(s')dsds' \geq 0$ for any $z(\cdot) \in L^2([0, 1])$ and it is equivalent to show that $X(s, T^*)$ for arbitrary *fixed* $T^* \in \mathcal{T}$ is square integrable.

We prove this by contradiction. Let $Z(T) = \int X^2(s, T)ds$. Assume that $\mathrm{E}[Z(T)] = \infty$ for $T \in \mathcal{T}_0 \subset \mathcal{T}$ such that $\int_{\mathcal{T}_0} g(T)dT > 0$. It follows that $\int_{\mathcal{T}} Z(T)dT > \int_{\mathcal{T}_0} Z(T)dT$, because $\mathcal{T}_0$ is a subset of $\mathcal{T}$. As expectation is a linear operator, $\mathrm{E}\int_{\mathcal{T}} Z(T)dT > \mathrm{E}\int_{\mathcal{T}_0} Z(T)dT$ still holds. It implies that $\mathrm{E}\int_{\mathcal{T}} Z(T)dT$ is infinite as we assume that $\mathrm{E}[Z(T)] = \infty$ for $T \in \mathcal{T}_0$. However because $X(s,T)$ is square integrable and $E\{\int_{\mathcal{T}} Z(T)dT\} = E\{\iint X^2(s,T)dsdT\}$, $E\{\int_{\mathcal{T}} Z(T)dT\}$ is finite. Thus by contradiction, we show that $E[Z(T)]$ is finite and $X(s,T^*)$ is square integrable for *fixed* $T^* \in \mathcal{T}$. And it follows that $\Sigma(s,s')$ is positive definite.

(iii) Eigenvalues of $\Sigma(s,s')$ satisfy $\sum_{l=1}^{\infty} \lambda_l < \infty$, which is equivalent to $\int \Sigma(s,s)ds < \infty$.

Because $c((s,T),(s',T'))$ is a continuous function in $\mathcal{S}\times\mathcal{T}$ and the intervals, $\mathcal{S}$ and $\mathcal{T}$, are compact, $\iint c((s,T),(s,T))dsdT$ is finite. Hence under the assumptions (A1.) and (A2.),

$$\begin{aligned}\int \Sigma(s,s)ds &= \iint c((s,T),(s,T))g(T)dTds \\ &\overset{\text{Fubini's Th}}{=} \int g(T)\Big[\int c((s,T),(s,T))ds\Big]dT \\ &\le \int \sup_{T\in\mathcal{T}} g(T)\Big[\int c((s,T),(s,T))ds\Big]dT = \sup_{T\in\mathcal{T}} g(T)\iint c((s,T),(s,T))dsdT < \infty,\end{aligned}$$

and equivalently, the sum of eigenvalues of $\Sigma(s,s')$ is finite.

Thus we obtain that $\Sigma(s,s') = \int c((s,T),(s',T)g(T)dT$ is a *proper* covariance function. The reason why we refer $\Sigma(s,s')$ as the *marginal* covariance function of $X_i(s,T)$ is as follows. Assume that there is a true latent process, $U_i(\cdot)$, such that $X_i(\cdot,T_{ij})$ is $U_i(\cdot)$ given that $T = T_{ij}$. Then the marginal covariance function, $\Sigma(s,s')$, is the covariance function of the true latent process, $U_i(s)$; specifically,

$$\begin{aligned}&\mathrm{cov}\{U_i(s),U_i(s')\} \\ &\quad= \mathrm{E}_T[\mathrm{cov}\{U_i(s),U_i(s')|T=T_{ij}\}] + \mathrm{cov}_T[\mathrm{E}\{U_i(s)|T=T_{ij}\},\mathrm{E}\{U_i(s')|T=T_{ij}\}] \\ &\quad= \mathrm{E}_T[\mathrm{cov}\{X_i(s,T_{ij}),X_i(s',T_{ij})\}] + \mathrm{cov}_T[\mathrm{E}\{X_i(s,T_{ij})\},\mathrm{E}\{X_i(s',T_{ij})\}],\end{aligned}$$

where the first term is

$$\begin{aligned}\mathrm{E}_T[\mathrm{cov}\{Y_i(s,T_{ij}),Y_i(s',T_{ij})\}] &= \int c\{(s,T),(s',T)\}g(T)dT \\ &= \Sigma(s,s'),\end{aligned}$$

and the second term is equal to 0.

Let $\{T_d : d = 1,\ldots,D\}$ be a set of unique T_{ij}'s for all i and j in an increasing order, such that $T_1 < T_2 < \ldots, < T_D$ and assume $T_0 = 0$ and $T_D = 1$.

Corollary 4.0.2
$\widehat{\Sigma}(s,s')$ is an unbiased estimator of the marginal covariance function, $\Sigma(s,s')$.

Proof. Here we show that that the sample covariance, $\widehat{\Sigma}(s,s') = \sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')/(\sum_{i=1}^{n} m_i)$, is an unbiased estimator of $\Sigma(s,s')$, i.e. $\mathrm{E}\{\widehat{\Sigma}(s,s')\} = \Sigma(s,s')$. Recall that here we are assuming $Y_{ij}(s) = X_{ij}(s)$. The proof is as follows:

$$
\begin{aligned}
\mathrm{E}\{\widehat{\Sigma}(s,s')\} &= \mathrm{E}\Big\{\frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')\Big\}\\
&= \mathrm{E}\Big\{\frac{1}{\sum_{i=1}^n m_i}\sum_{d=1}^D\sum_{i=1}^n\sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')I(T_{ij}\in(T_{d-1},T_d])\Big\}\\
&= \frac{1}{\sum_{i=1}^n m_i}\sum_{d=1}^D\sum_{i=1}^n\sum_{j=1}^{m_i} \mathrm{E}\Big\{Y_{ij}(s)Y_{ij}(s')I(T_{ij}\in(T_{d-1},T_d])\Big\}\\
&= \frac{1}{\sum_{i=1}^n m_i}\sum_{d=1}^D\sum_{i=1}^n\sum_{j=1}^{m_i} \mathrm{E}\Big\{Y_i(s,T_d)Y_i(s',T_d)\Big\}\mathrm{E}\Big\{I(T_{ij}\in(T_{d-1},T_d])\Big\} \qquad (\text{because } T\perp\!\!\!\perp Y_i(\cdot,T=t))\\
&= \frac{1}{\sum_{i=1}^n m_i}\sum_{d=1}^D\sum_{i=1}^n\sum_{j=1}^{m_i} c((s,T_d),(s',T_d))\mathrm{P}(T\in(T_{d-1},T_d]) \qquad (\text{because } Y_i(s,T_d)=X_i(s,T_d))\\
&= \sum_{d=1}^D c((s,T_d),(s',T_d))\mathrm{P}(T\in(T_{d-1},T_d]) = \int c((s,t),(s',t))dF(t) = \Sigma(s,s'),
\end{aligned}
$$

where F is a true sampling distribution function of T.

Theorem 4.1.1 (Theorem 1 in Section 4.1)
Assume (A1.) - (A3.) hold. Then $|\widehat{\Sigma}(s,s')-\Sigma(s,s')| \xrightarrow{p} 0$ *as* n *diverges. If in addition (A4.) holds, then*

$$\|\widehat{\Sigma}(\cdot,\cdot)-\Sigma(\cdot,\cdot)\|_s \xrightarrow{p} 0 \text{ as } n\longrightarrow\infty, \tag{S2}$$

where $\|k(\cdot,\cdot)\|_s = \{\iint k^2(s,s')dsds'\}^{1/2}$ *is the Hilbert-Schmidt norm of* $k(\cdot,\cdot)$.

Proof. We prove this theorem in two steps: first we show (i) the pointwise consistency and then we show (ii) the Hilbert-Schmidt norm consistency.

(i) The pointwise consistency

Let $M_1=\sum_{i=1}^n m_i$, $M_2=\sum_{i=1}^n m_i^2$, and $P_d=P(T\in(T_{d-1},T_d])$. Let $A_i(T)$ be a realization of $A(T)=Y(s,T)Y(s',T)$ for fixed s and s', and let $V_d=(D/M_1)\sum_{i=1}^n\sum_{j=1}^{m_i}A_i(T_d)I(T_{ij}\in(T_{d-1},T_d])$. Then the estimator, $\widehat{\Sigma}(s,s')$, can be written as $\overline{V}=D^{-1}\sum_{d=1}^D V_d$;

$$
\begin{aligned}
\widehat{\Sigma}(s,s') &= \frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')\\
&= \frac{1}{D}\sum_{d=1}^D\frac{D}{\sum_{i=1}^n m_i}\sum_{i=1}^n\sum_{j=1}^{m_i} Y_i(s,T_d)Y_i(s',T_d)I(T_{ij}\in(T_{d-1},T_d])\\
&= \frac{1}{D}\sum_{d=1}^D\frac{D}{M_1}\sum_{i=1}^n\sum_{j=1}^{m_i} A_i(T_d)I(T_{ij}\in(T_{d-1},T_d]) = \frac{1}{D}\sum_{d=1}^D V_d.
\end{aligned}
$$

Notice that V_d's are correlated over d. To show the consistency of the average of the correlated random variables, we first obtain the covariance of V_d and $V_{d'}$, $\mathrm{cov}(V_d, V_{d'}) = \mathrm{E}[V_d V_{d'}] - \mathrm{E}[V_d]\mathrm{E}[V_{d'}]$ and then use Theorem 5.3 of Boos & Stefanski (2013, p.208). For completeness the theorem we used is given below:

Theorem 5.3 (Boos & Stefanski, 2013, p.208) If $X_1, \ldots, X_n$ are random variable with finite means $\mu_i = E(X_i)$, variance $E(X_i - \mu_i)^2 = \sigma_i^2$, and covariances $E(X_i - \mu_i)(X_j - \mu_j) = \sigma_{ij}$ such that

$$\mathrm{var}(\bar{\mathrm{X}}) = \frac{1}{\mathrm{n}^2}\left[\sum_{\mathrm{i}=1}^{\mathrm{n}} \text{œ}_{\mathrm{i}}^2 + 2\sum_{\mathrm{i}=1}^{\mathrm{n}-1}\sum_{\mathrm{j}=\mathrm{i}+1}^{\mathrm{n}} \text{œ}_{\mathrm{ij}}\right] \to 0 \text{ as } \mathrm{n} \to \infty,$$

then $\bar{X} - \bar{\mu} \xrightarrow{p} 0$ as $n \to \infty$, where $\bar{\mu} = n^{-1}\sum_{i=1}^{n} \mu_i$

In the following we obtain $\mathrm{cov}(V_d, V_{d'})$ by getting $\mathrm{E}[V_d]$ and $\mathrm{E}[V_d V_{d'}]$:

$$\begin{aligned}
\mathrm{E}[V_d] &= \mathrm{E}\left[\frac{D}{M_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_i(s, T_d)Y_i(s', T_d)I(T_{ij} \in (T_{d-1}, T_d])\right] \\
&= D \cdot \mathrm{E}[Y_i(s, T_d)Y_i(s', T_d)] \cdot \mathrm{E}[I(T_{ij} \in (T_{d-1}, T_d])] \qquad (\text{because } T \perp\!\!\!\perp Y(\cdot, T = t)) \\
&= D \cdot c((s, T_d), (s', T_d)) \cdot P_d \qquad (\text{because } Y_i(s, T) = X_i(s, T))
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[V_d V_{d'}] &= \mathrm{E}\left[\left\{\frac{D}{M_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} A_i(T_d)I(T_{ij} \in (T_{d-1}, T_d])\right\}\left\{\frac{D}{M_1}\sum_{i'=1}^{n}\sum_{j'=1}^{m_i'} A_{i'}(T_{d'})I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\right\}\right] \\
&= \left(\frac{D}{M_1}\right)^2 \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{i'=1}^{n}\sum_{j'=1}^{m_i'} A_i(T_d)I(T_{ij} \in (T_{d-1}, T_d])A_{i'}(T_{d'})I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\right] \\
&= \left(\frac{D}{M_1}\right)^2 \mathrm{E}\left[\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{m_i}\sum_{j'=1}^{m_i'} A_i(T_d)A_{i'}(T_{d'})I(T_{ij} \in (T_{d-1}, T_d])I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\right] \\
&= \left(\frac{D}{M_1}\right)^2 \sum_{i,i'}\sum_{j,j'} \mathrm{E}\left[A_i(T_d)A_{i'}(T_{d'})I(T_{ij} \in (T_{d-1}, T_d])I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\right] \\
&= \left(\frac{D}{M_1}\right)^2 \sum_{i,i'}\sum_{j,j'} \mathrm{E}\left[A_i(T_d)A_{i'}(T_{d'})\right]\mathrm{E}\left[I(T_{ij} \in (T_{d-1}, T_d])I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\right] \qquad \text{(S3)}
\end{aligned}$$

Here we consider Equation (S3) for two cases; when $d = d'$ and when $d \neq d'$.

Case 1: $d = d'$

This is equivalent to say that $T_d = T_{d'}$ as $\{T_d : d = 1, \ldots, D\}$ is a set of unique values. Let $T^* = T_d = T_{d'}$.

$$\mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] = P(T_{ij} \in (T_{d-1}, T_d] \text{ and } T_{i'j'} \in (T_{d-1}, T_d])$$
$$= \begin{cases} P(T_{ij} \in (T_{d-1}, T_d])P(T_{i'j'} \in (T_{d-1}, T_d]) = P_d^2 & \text{for } i \neq i', j, j' \\ P(T_{ij} \in (T_{d-1}, T_d]) = P_d & \text{for } i = i', j = j' \\ 0 & \text{for } i = i', j \neq j' \end{cases}$$

$$\mathrm{E}\Big[A_i(T_d)A_{i'}(T_{d'})\Big] = \mathrm{E}\Big[Y_i(s, T_d)Y_i(s', T_d)Y_{i'}(s, T_{d'})Y_{i'}(s', T_{d'})\Big]$$
$$= \mathrm{E}\Big[Y_i(s, T^*)Y_i(s', T^*)Y_{i'}(s, T^*)Y_{i'}(s', T^*)\Big]$$
$$= \begin{cases} \mathrm{E}\Big[Y_i(s, T^*)Y_i(s', T^*)\Big]\mathrm{E}\Big[Y_{i'}(s, T^*)Y_{i'}(s', T^*)\Big] \\ \quad = c((s, T^*), (s', T^*))^2 \qquad \text{for } i \neq i', j, j' \\ \quad (\text{because } Y_i(s, T_d) = X_i(s, T_d)) \\ \mathrm{E}\Big[\Big\{Y_i(s, T^*)Y_i(s', T^*)\Big\}^2\Big] \\ \quad = \mathrm{var}\Big\{Y(s, T^*)Y(s', T^*)\Big\} + \Big\{c((s, T^*), (s', T^*))\Big\}^2 \qquad \text{for } i = i', j, j' \\ \quad (\text{because } Y_i(s, T_d) = X_i(s, T_d)) \end{cases}$$

By summing over all i, i', j, and j', Equation (S3) for the case of $d = d'$ equals to

$$\mathrm{E}[V_d V_{d'}] = \Big(\frac{D}{M_1}\Big)^2 \Big\{\sum_{i=1}^{n} m_i(M_1 - m_i)\Big\}\Big\{P_d \cdot c((s, T^*), (s', T^*))\Big\}^2$$
$$+ \Big(\frac{D}{M_1}\Big)^2 M_1 \cdot P_d \cdot \Bigg[\mathrm{var}\Big\{Y(s, T^*)Y(s', T^*)\Big\} + \Big\{c((s, T^*), (s', T^*))\Big\}^2\Bigg]$$
$$= \Big(D^2 - \frac{M_2 D^2}{M_1^2}\Big)\Big\{P_d \cdot c((s, T^*), (s', T^*))\Big\}^2$$
$$+ \Big(\frac{D^2}{M_1}\Big) \cdot P_d \cdot \Bigg[\mathrm{var}\Big\{Y(s, T^*)Y(s', T^*)\Big\} + \Big\{c((s, T^*), (s', T^*))\Big\}^2\Bigg] \qquad \text{for } d = d'$$

Case 2: $d \neq d'$

$$\mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] = P(T_{ij} \in (T_{d-1}, T_d] \text{ and } T_{i'j'} \in (T_{d'-1}, T_{d'}])$$
$$= \begin{cases} P(T_{ij} \in (T_{d-1}, T_d])P(T_{i'j'} \in (T_{d'-1}, T_{d'}]) = P_d P_{d'} & \text{for } i \neq i', j, j' \\ P(T_{ij} \in (T_{d-1}, T_d] \text{ and } T_{ij'} \in (T_{d'-1}, T_{d'}]) = P_d P_{d'} & \text{for } i = i', j \neq j' \quad (\text{because} T_{ij} \perp\!\!\!\perp T_{ij'}) \\ 0 \qquad \text{for } i = i' \text{ and } j = j' \end{cases}$$

For the last case ($i = i'$ and $j = j'$), $P(T_{ij} \in (T_{d-1}, T_d]$ and $T_{i'j'} \in (T_{d'-1}, T_{d'}])$ is equal to 0 because it is not possible that one subject is observed at two different visit times, T_d and $T_{d'}$, with the same index for visit, $j = j'$.

$$\mathrm{E}\Big[A_i(T_d)A_{i'}(T_{d'})\Big] = \mathrm{E}\Big[Y_i(s, T_d)Y_i(s', T_d)Y_{i'}(s, T_{d'})Y_{i'}(s', T_{d'})\Big]$$
$$= \begin{cases} \mathrm{E}\Big[Y_i(s, T_d)Y_i(s', T_d)\Big]\mathrm{E}\Big[Y_{i'}(s, T_{d'})Y_{i'}(s', T_{d'})\Big] \\ \qquad = c((s, T_d), (s', T_d)) \cdot c((s, T_{d'}), (s', T_{d'})) & \text{for } i \neq i', j, j' \\ \qquad (\text{because } Y_i(s, T_d) = X_i(s, T_d)) \\ \mathrm{E}\Big[Y_i(s, T_d)Y_i(s', T_d)Y_i(s, T_{d'})Y_i(s', T_{d'})\Big] & \text{for } i = i', j \neq j' \\ 0 \qquad \text{for } i = i', j = j' \end{cases}$$

By summing over i, i', j, and j', Equation (S3) for the case of $d \neq d'$ equals to

$$\begin{aligned} \mathrm{E}[V_d V_{d'}] =& \left(\frac{D}{M_1}\right)^2 \left\{\sum_{i=1}^n m_i(M_1 - m_i)\right\}\Big\{P_d \cdot P_{d'} \cdot c((s, T_d), (s', T_d)) \cdot c((s, T_{d'}), (s', T_{d'})\Big\} \\ &+ \left(\frac{D}{M_1}\right)^2 \left\{\sum_{i=1}^n m_i(m_i - 1)\right\}\Big\{P_d \cdot P_{d'} \cdot \mathrm{E}\Big[Y_i(s, T_d)Y_i(s', T_d)Y_i(s, T_{d'})Y_i(s', T_{d'})\Big]\Big\} \\ =& \left(D^2 - \frac{M_2 D^2}{M_1^2}\right) \Big\{P_d \cdot P_{d'} \cdot c((s, T_d), (s', T_d)) \cdot c((s, T_{d'}), (s', T_{d'})\Big\} \\ &+ \left(\frac{M_2 D^2}{M_1^2} - \frac{D^2}{M_1}\right) \Big\{P_d \cdot P_{d'} \cdot \mathrm{E}\Big[Y_i(s, T_d)Y_i(s', T_d)Y_i(s, T_{d'})Y_i(s', T_{d'})\Big]\Big\} \qquad \text{for } d \neq d' \end{aligned}$$

In summary, we have

$$\mathrm{E}[V_d] = D \cdot c((s, T_d), (s', T_d)) \cdot P_d \qquad (\text{because } Y_i(s, T) = X_i(s, T))$$

and

$$\mathrm{E}[V_d V_{d'}] = \begin{cases} \Big\{D^2 - \dfrac{M_2 D^2}{M_1^2}\Big\}\Big\{P_d \cdot c((s,T_d),(s',T_d))\Big\}^2 \\ \qquad +\Big\{\dfrac{D^2}{M_1}\Big\}\Big\{P_d \cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\Big\} & \text{for } d = d' \\ \\ \Big\{D^2 - \dfrac{M_2 D^2}{M_1^2}\Big\}\Big\{P_d P_{d'} \cdot c((s,T_d),(s',T_d)) \cdot c((s,T_{d'}),(s',T_{d'})\Big\} \\ \qquad +\Big\{\dfrac{M_2 D^2}{M_1^2} - \dfrac{D^2}{M_1}\Big\}\Big\{P_d P_{d'} \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\Big\} & \text{for } d \neq d' \end{cases}$$

Finally we obtain the covariance of V_d and $V_{d'}$; that is,

$$\begin{aligned} \mathrm{cov}(V_d, V_{d'}) &= \mathrm{E}[V_d V_{d'}] - \mathrm{E}[V_d]\mathrm{E}[V_{d'}] \\ &= \begin{cases} \Big\{\dfrac{D^2}{M_1}\Big\}\Big\{P_d \cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\Big\} \\ \qquad\qquad - \Big\{\dfrac{M_2 D^2}{M_1^2}\Big\}\Big\{P_d \cdot c((s,T_d),(s',T_d))\Big\}^2 & \text{for } d = d' \\ \\ \Big\{\dfrac{M_2 D^2}{M_1^2} - \dfrac{D^2}{M_1}\Big\}\Big\{P_d P_{d'} \cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\Big\} \\ \qquad\qquad - \Big\{\dfrac{M_2 D^2}{M_1^2}\Big\}\Big\{P_d P_{d'} \cdot c((s,T_d),(s',T_d)) \cdot c((s,T_{d'}),(s',T_{d'}))\Big\} & \text{for } d \neq d' \end{cases} \end{aligned}$$

For simplicity in notation, denote the variance of V_d with σ_d^2, and the covariance of V_d and $V_{d'}$ with $\sigma_{d,d'}$. Under the assumptions (A1.) - (A3.), $\mathrm{E}[V_d]$, σ_d^2 and $\sigma_{d,d'}$ are finite. Following Theorem 5.3 of Boos & Stefanski (2013, p.208), we show the pointwise consistency of $\widehat{\Sigma}(s,s') = \overline{V}$ by proving that the following holds;

$$\frac{1}{D^2}\Big[\sum_{d=1}^{D}\sigma_d^2 + 2\sum_{d=1}^{D-1}\sum_{d'=d+1}^{D}\sigma_{dd'}\Big] = \frac{1}{D^2}\Big[\sum_{d=1}^{D}\sigma_d^2 + \sum_{d=1}^{D}\sum_{d'\neq d}\sigma_{dd'}\Big] \xrightarrow{\text{p}} 0 \qquad \text{as } D \to \infty.$$

By plugging in $\mathrm{cov}(V_d, V_{d'})$ that we obtained earlier, we get the first term, $\frac{1}{D^2}\sum_{d=1}^{D}\sigma_d^2$, equal to

$$\begin{aligned} &\frac{1}{D^2}\sum_{d=1}^{D}\Bigg[\Big\{\frac{D^2}{M_1}\Big\}\Big\{P_d \cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\Big\} - \Big\{\frac{D^2 M_2}{M_1^2}\Big\}\Big\{P_d \cdot c((s,T_d),(s',T_d))\Big\}^2\Bigg] \\ &= \Bigg[\frac{1}{M_1}\sum_{d=1}^{D}\Big\{P_d \cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\Big\}\Bigg] - \Bigg[\frac{M_2}{M_1^2}\sum_{d=1}^{D}\Big\{P_d \cdot c((s,T_d),(s',T_d))\Big\}^2\Bigg] \\ &= \Bigg[\frac{1}{M_1}\mathrm{E}_T[E[A(T)^2|T]]\Bigg] - \Bigg[\frac{M_2}{M_1^2}\sum_{d=1}^{D}\Big\{P_d \cdot c((s,T_d),(s',T_d))\Big\}^2\Bigg] \end{aligned} \tag{S4}$$

and the second term, $\frac{1}{D^2}\sum_{d=1}^{D}\sum_{d'\neq d}\sigma_{dd'}$, equal to

$$\frac{1}{D^2}\sum_{d=1}^{D}\sum_{d'\neq d}\left[\left\{\frac{M_2D^2}{M_1^2}-\frac{D^2}{M_1}\right\}\left\{P_dP_{d'}\cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\right\}\right.$$
$$\left.-\left\{\frac{M_2D^2}{M_1^2}\right\}\left\{P_dP_{d'}\cdot c((s,T_d),(s',T_d))\cdot c((s,T_{d'}),(s',T_{d'}))\right\}\right]$$

$$=\frac{1}{D^2}\sum_{d=1}^{D}\sum_{d'\neq d}\left[\left\{\frac{M_2D^2}{M_1^2}-\frac{D^2}{M_1}\right\}\left\{P_dP_{d'}\cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\right\}\right] \tag{S5}$$

$$-\frac{1}{D^2}\sum_{d=1}^{D}\sum_{d'\neq d}\left[\left\{\frac{M_2D^2}{M_1^2}\right\}\left\{P_dP_{d'}\cdot c((s,T_d),(s',T_d))\cdot c((s,T_{d'}),(s',T_{d'}))\right\}\right], \tag{S6}$$

where

$$\text{Term (S5)}=\left\{\frac{M_2}{M_1^2}-\frac{1}{M_2}\right\}\sum_{d=1}^{D}\sum_{d'\neq d}P_dP_{d'}\cdot \mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]$$
$$=\left\{\frac{M_2}{M_1^2}-\frac{1}{M_1}\right\}\sum_{d=1}^{D}P_d\left[\sum_{d'=1}^{D}P_{d'}\mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\right.$$
$$\left.-P_d\mathrm{E}\Big[Y_i(s,T_d)Y_i(s',T_d)Y_i(s,T_{d'})Y_i(s',T_{d'})\Big]\right]$$
$$=\left[\left\{\frac{M_2}{M_1^2}-\frac{1}{M_1}\right\}\sum_{d=1}^{D}P_d\cdot \mathrm{E}_{T'}[\mathrm{E}[A(T)A(T')|T=T_d,T'=T_{d'}]]\right]$$
$$-\left[\left\{\frac{M_2}{M_1^2}-\frac{1}{M_1}\right\}\sum_{d=1}^{D}P_d^2\mathrm{E}[A(T)^2|T=T_d]\right]$$
$$=\left[\left\{\frac{M_2}{M_1^2}-\frac{1}{M_1}\right\}\mathrm{E}_T[\mathrm{E}_{T'}[\mathrm{E}[A(T)A(T')|T,T']]]\right]$$
$$-\left[\left\{\frac{M_2}{M_1^2}-\frac{1}{M_1}\right\}\sum_{d=1}^{D}P_d^2\cdot\mathrm{E}[A(T)^2|T=T_d]\right],$$

and

$$\text{Term (S6)}=-\frac{M_2}{M_1^2}\sum_{d=1}^{D}\sum_{d'\neq d}\left[\left\{P_dP_{d'}\cdot c((s,T_d),(s',T_d))\cdot c((s,T_{d'}),(s',T_{d'}))\right\}\right]$$
$$=-\frac{M_2}{M_1^2}\sum_{d=1}^{D}P_d\cdot c((s,T_d),(s',T_d))\sum_{d'\neq d}P_{d'}\cdot c((s,T_{d'}),(s',T_{d'}))$$

$$= -\frac{M_2}{M_1^2}\sum_{d=1}^{D} P_d \cdot c((s,T_d),(s',T_d)) \times \left[\sum_{d'=1}^{D} P_{d'}\cdot c((s,T_{d'}),(s',T_{d'})) - P_d\cdot c((s,T_d),(s',T_d))\right]$$

$$= -\frac{M_2}{M_1^2}\{\Sigma(s,s')\}^2 + \frac{M_2}{M_1^2}\sum_{d=1}^{D}\left\{P_d\cdot c((s,T_d),(s',T_d))\right\}^2$$

By combining everything together, we have

$$\frac{1}{D^2}\left[\sum_{d=1}^{D}\sigma_d^2 + \sum_{d=1}^{D}\sum_{d'\neq d}\sigma_{dd'}\right] = \left[\frac{1}{M_1}\mathrm{E}_{\mathcal{T}}[E[A(T)^2|T]]\right] + \left[\left\{\frac{M_2}{M_1^2} - \frac{1}{M_1}\right\}\mathrm{E}_{\mathcal{T}}[\mathrm{E}_{\mathcal{T}'}[\mathrm{E}[A(T)A(T')|T,T']]]\right] - \left[\left\{\frac{M_2}{M_1^2} - \frac{1}{M_1}\right\}\sum_{d=1}^{D} P_d^2\cdot \mathrm{E}[A(T)^2|T=T_d]\right] - \frac{M_2}{M_1^2}\{\Sigma(s,s')\}^2. \qquad \text{(S7)}$$

As discussed before, to prove the pointwise consistency of $\widehat{\Sigma}(s,s')$ it is sufficient to show that Equation (S7) converges to 0 as D diverges; we show this by showing each of the terms in Equation (S7) converges to zero as D diverges.

First notice that M_1 diverges with D because $D \leq M_1$ and $M_1/M_2 = O(1)$. And recall that here we are considering the case when $Y(s,T) = X(s,T)$, and thus $Y(s,T)$ satisfies all of the assumptions made on $X(s,T)$. Because integration of a continuous function in a compact interval is finite, under the assumptions (A2.) and (A3.) it is easy to show that $\mathrm{E}_{\mathcal{T}}[\mathrm{E}[A(T)^2|T]]$, $\mathrm{E}_{\mathcal{T}}[\mathrm{E}_{\mathcal{T}'}[\mathrm{E}[A(T)A(T')|T,T']]]$, and $\Sigma(s,s')$ are finite. For example $\mathrm{E}_{\mathcal{T}}[\mathrm{E}[A(T)^2|T]] = \int_{\mathcal{T}} g(T)\mathrm{E}[A(T)^2|T]dT \leq \sup_{\mathcal{T}} g(T)\int_{\mathcal{T}}\mathrm{E}[A(T)^2|T]dT$ is finite because (i) by the assumption (A2.) $\sup_{\mathcal{T}} g(T)$ is finite and (ii) $\int_{\mathcal{T}}\mathrm{E}[A(T)^2|T]dT$ is finite because by the assumption (A3.) $\mathrm{E}[A(T)^2|T]$ is continuous and finite in the compact interval $\mathcal{T} = [0,1]$. And with the same reasoning we can also show that $\sum_{d=1}^{D} P_d^2\cdot\mathrm{E}[A(T)^2|T=T_d]$ is finite because $\sum_{d=1}^{D} P_d^2\cdot\mathrm{E}[A(T)^2|T=T_d] < \sup_d P_d \mathrm{E}_{\mathcal{T}}[E[A(T)^2|T]]$. Thus we show that each of the terms in Equation (S7) converges to zero as D diverges.

It follows that

$$\frac{1}{D^2}\left[\sum_{d=1}^{D}\sigma_d^2 + \sum_{d=1}^{D}\sum_{d'\neq d}\sigma_{dd'}\right] \xrightarrow{\text{p}} 0 \text{ as } D\to\infty,$$

and by *Theorem 5.3* (*Boos & Stefanski, 2013*), $\widehat{\Sigma}(s,s') = \overline{V} = D^{-1}\sum_{d=1}^{D} V_d$ converges in probability to $D^{-1}\sum_{d=1}^{D}\mathrm{E}[V_d] = \sum_{d=1}^{D} P_d\cdot c((s,T_d),(s',T_d)) = \Sigma(s,s')$ as n diverges; recall that we assume D diverges with the sample size n.

(ii) The Hilbert-Schmidt norm consistency

Because the Hilbert-Schmidt norm consistency is implied by the convergence of $\mathrm{E}\left[\left\|\widehat{\Sigma}(s,s') - \Sigma(s,s')\right\|_s^2\right]$ using Markov inequality, it is sufficient to show that $\mathrm{E}\left[\left\|\widehat{\Sigma}(s,s') - \mathrm{E}[\widehat{\Sigma}(s,s')]\right\|_s^2\right] = \mathrm{E}\left[\left\|\widehat{\Sigma}(s,s') - \Sigma(s,s')\right\|_s^2\right]$ converges to 0 as n diverges and we prove it by showing that the upper bound given in Equation (S10) converges to 0 as n diverges.

The sample covariance operator associated with $\widehat{\Sigma}(s,s')$ is

$$H(x) = D^{-1}\sum_{d=1}^{D}\frac{D}{M_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} < Y_{ij}, x > Y_{ij} I(T_{ij} \in (T_{d-1}, T_d])$$
$$= D^{-1}\sum_{d=1}^{D}\frac{D}{M_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} < Y_i(\cdot, T_d), x > Y_i(\cdot, T_d) I(T_{ij} \in (T_{d-1}, T_d]), \qquad x \in L^2$$

and using the covariance operator, $H(\cdot)$, $\mathrm{E}\Big[\|\widehat{\Sigma}(s,s') - \mathrm{E}[\widehat{\Sigma}(s,s')]\|_s^2\Big]$ can be written as

$$\mathrm{E}\Big[\|\widehat{\Sigma}(s,s') - \mathrm{E}[\widehat{\Sigma}(s,s')]\|_s^2\Big] = \mathrm{E}\Big[\|H - \mathrm{E}[H]\|_s^2\Big] = \mathrm{E}\Big[\sum_{v=1}^{\infty}\|H(e_v) - \mathrm{E}[H(e_v)]\|_s^2\Big]$$
$$= \mathrm{E}\Bigg[\sum_{v=1}^{\infty}\Big\|D^{-1}\sum_{d=1}^{D}\frac{D}{M_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big\{ < Y_{ij}, e_v > Y_{ij} I(T_{ij} \in (T_{d-1}, T_d])$$
$$- \mathrm{E}[< Y_{ij}, e_v > Y_{ij} I(T_{ij} \in (T_{d-1}, T_d])]\Big\}\Big\|_s^2\Bigg],$$

where $\{e_v : v \geq 1\}$ is any orthonormal basis (Horváth & Kokoszka, 2012, p.22).

Let $R_{ijdv} =< Y_{ij}, e_v > Y_{ij} I(T_{ij} \in (T_{d-1}, T_d]) =< Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d) I(T_{ij} \in (T_{d-1}, T_d])$. Then it follows that

$$\mathrm{E}\Big[\|\widehat{\Sigma}(s,s') - \mathrm{E}[\widehat{\Sigma}(s,s')]\|_s^2\Big] = \mathrm{E}\Big[\|H - \mathrm{E}[H]\|_s^2\Big]$$
$$= \mathrm{E}\Big[\sum_{v=1}^{\infty}\Big\|D^{-1}\sum_{d=1}^{D}\frac{D}{M_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big\{R_{ijdv} - \mathrm{E}[R_{ijdv}]\Big\}\Big\|_s^2\Big]$$
$$= \mathrm{E}\Big[\sum_{v=1}^{\infty} D^{-2}\sum_{d=1}^{D}\sum_{d'=1}^{D}\frac{D^2}{M_1^2} < \sum_{i}^{n}\sum_{j}^{m_i}\{R_{ijdv} - \mathrm{E}[R_{ijdv}]\}, \sum_{i'}^{n}\sum_{j'}^{m_{i'}}\{R_{i'j'd'v} - \mathrm{E}[R_{i'j'd'v}]\} > \Big]$$
$$= D^{-2}\sum_{d=1}^{D}\sum_{d'=1}^{D}\frac{D^2}{M_1^2}\Bigg[\sum_{i}^{n}\sum_{i'}^{n}\sum_{j}^{m_i}\sum_{j'}^{m_{i'}}\sum_{v=1}^{\infty}\mathrm{E}\Big\{ < R_{ijdv} - \mathrm{E}[R_{ijdv}], R_{i'j'd'v} - \mathrm{E}[R_{i'j'd'v}] > \Big\}\Bigg]$$
$$= M_1^{-2}\sum_{d=1}^{D}\sum_{d'=1}^{D} Q_{dd'}, \tag{S8}$$

where we define

$$Q_{dd'} = \sum_{i}^{n}\sum_{i'}^{n}\sum_{j}^{m_i}\sum_{j'}^{m_{i'}}\sum_{v=1}^{\infty}\mathrm{E}\Big\{ < R_{ijdv} - \mathrm{E}[R_{ijdv}], R_{i'j'd'v} - \mathrm{E}[R_{i'j'd'v}] > \Big\}$$
$$= \sum_{i}^{n}\sum_{i'}^{n}\sum_{j}^{m_i}\sum_{j'}^{m_{i'}}\sum_{v=1}^{\infty}\Big\{\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \Big\}. \tag{S9}$$

Furthermore because $Q_{dd'} \leq |Q_{dd'}|$ and $|a+b| \leq |a| + |b|$ for any $a, b \in \mathbb{R}$,

$$0 \leq \mathrm{E}\Big[\|H - \mathrm{E}[H]\|_s^2\Big] = M_1^{-2}\sum_{d=1}^{D}\sum_{d'=1}^{D} Q_{dd'} \leq M_1^{-2}\sum_{d=1}^{D}\sum_{d'=1}^{D}|Q_{dd'}|$$

$$= M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} \left| \sum_{i}^{n} \sum_{i'}^{n} \sum_{j}^{m_i} \sum_{j'}^{m_{i'}} \sum_{v=1}^{\infty} \left\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \right\} \right|$$

$$\leq M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} \sum_{i}^{n} \sum_{i'}^{n} \sum_{j}^{m_i} \sum_{j'}^{m_{i'}} \sum_{v=1}^{\infty} \left| \left\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \right\} \right|. \qquad \text{(S10)}$$

As discussed, we show the convergence of $\mathrm{E}\left[\left\|\widehat{\Sigma}(s,s') - \mathrm{E}[\widehat{\Sigma}(s,s')]\right\|_s^2\right] = M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} Q_{dd'}$ by showing that the upper bound given in Equation (S10) converges to 0. In the following we first simplify the summand in Equation (S10) and then study the sum.

Let $G_{id}(x) = < Y_i(\cdot, T_d), x > Y_i(\cdot, T_d)$ for $x \in L_2$. Notice that the summand in Equation (S9), $\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] >]$, can be rewritten as

$$\begin{aligned}
&\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] >] \\
&= \mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d]) \cdot I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] \times \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{i'd'}(e_v)\big\rangle\Big] \\
&\quad - \mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])\Big] \cdot \mathrm{E}\Big[I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] \times \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle,
\end{aligned} \qquad \text{(S11)}$$

because of the following:

$$\begin{aligned}
&\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] \\
&= \mathrm{E}\Big[\big\langle < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d) I(T_{ij} \in (T_{d-1}, T_d]), < Y_{i'}(\cdot, T_{d'}), e_v > Y_{i'}(\cdot, T_d') I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\big\rangle\Big] \\
&= \mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d]) \cdot I(T_{i'j'} \in (T_{d'-1}, T_{d'}]) \\
&\qquad\qquad \times \big\langle < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d), < Y_{i'}(\cdot, T_{d'}), e_v > Y_{i'}(\cdot, T_d')\big\rangle\Big] \\
&= \mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d]) \cdot I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] \\
&\qquad\qquad \times \mathrm{E}\Big[\big\langle < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d), < Y_{i'}(\cdot, T_{d'}), e_v > Y_{i'}(\cdot, T_d')\big\rangle\Big]
\end{aligned}$$

and

$$\begin{aligned}
&< \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \\
&= \Big\langle \mathrm{E}\Big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d) I(T_{ij} = T_d)\Big],\ \mathrm{E}\Big[< Y(\cdot, T_{d'}), e_v > Y(\cdot, T_{d'}) I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big]\Big\rangle \\
&= \Big\langle \mathrm{E}\Big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d)\Big] \cdot \mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])\Big], \\
&\qquad\qquad \mathrm{E}\Big[< Y(\cdot, T_{d'}), e_v > Y(\cdot, T_{d'})\Big] \cdot \mathrm{E}\Big[I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big]\Big\rangle \\
&= \mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])\Big] \cdot \mathrm{E}\Big[I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] \\
&\qquad\qquad \times \Big\langle \mathrm{E}\Big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d)\Big],\ \mathrm{E}\Big[< Y(\cdot, T_{d'}), e_v > Y(\cdot, T_{d'})\Big]\Big]\Big\rangle
\end{aligned}$$

In the following we consider Equation (S10) for two cases, when $d = d'$ and when $d \neq d'$.

Case 1: $d = d'$

This case is equivalent to the case of $T_d = T_{d'}$ as T_d's are unique values. Let $T^* = T_d = T_{d'}$. First observe the following:

$$\mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d]) \cdot I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] = \begin{cases} P_d^2 & \text{for } i \neq i', j, j' \\ 0 & \text{for } i = i', j \neq j' \\ P_d & \text{for } i = i', j = j' \end{cases}$$

$$\mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])\Big] \cdot \mathrm{E}\Big[I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] = P_d^2 \qquad \text{for all } i, i', j, j'$$

Case 1(a): $d = d', i \neq i', j, j'$ (Using independence)

$$\begin{aligned}
&\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \qquad \text{from Equation (S11)} \\
&= P_d^2 \cdot \Big\{ \mathrm{E}\Big[\Big\langle G_{id}(e_v), G_{i'd'}(e_v)\Big\rangle\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle \Big\} \\
&= P_d^2 \cdot \Big\{ \mathrm{E}\Big[\Big\langle G_{id}(e_v), G_{i'd}(e_v)\Big\rangle\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle \Big\} \\
&= P_d^2 \cdot \Big\{ \mathrm{E}\Big[\int G_{id}(e_v)(s) \cdot G_{i'd}(e_v)(s)\ ds\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle \Big\} \\
&= P_d^2 \cdot \Big\{ \mathrm{E}\Big[\int < Y_i(\cdot, T_d), e_v > Y_i(s, T_d) < Y_{i'}(\cdot, T_d), e_v > Y_{i'}(s, T_d) ds\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle \Big\} \\
&= P_d^2 \cdot \Big\{ \int \mathrm{E}\Big[< Y_i(\cdot, T_d), e_v > Y_i(s, T_d) < Y_{i'}(\cdot, T_d), e_v > Y_{i'}(s, T_d)\Big] ds - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle \Big\}
\end{aligned}$$

(by commutative property of expectation with bounded operators (Horváth & Kokoszka, 2012, p.23))

$$\begin{aligned}
&= P_d^2 \cdot \Big\{ \int \mathrm{E}\Big[< Y_i(\cdot, T_d), e_v > Y_i(s, T_d)\Big] \mathrm{E}\Big[< Y_{i'}(\cdot, T_d), e_v > Y_{i'}(s, T_d)\Big] ds \\
&\qquad\qquad - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle \Big\}
\end{aligned}$$

(by independence)

$$= P_d^2 \cdot \Big\{ \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd}(e_v)\Big]\Big\rangle \Big\} = 0$$

Case 1(b): $d = d', i = i', j \neq j'$

$$\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] >$$

$$
\begin{aligned}
&= 0 \cdot \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{i'd'}(e_v)\big\rangle\Big] - P_d^2 \cdot \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle \\
&= -P_d^2 \cdot \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{id}(e_v)\Big]\Big\rangle \\
&= -P_d^2 \cdot \Big\|\mathrm{E}\big[G_{id}(e_v)\big]\Big\|^2 \\
&= -P_d^2 \cdot \Big\|\mathrm{E}\big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d))\big]\Big\|^2
\end{aligned}
$$

Because P_d^2 and $\Big\|\mathrm{E}\big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d))\big]\Big\|^2$ are nonnegative, it is implied that

$$
\begin{aligned}
\Big|\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \Big| &= P_d^2 \cdot \Big\|\mathrm{E}\big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d))\big]\Big\|^2 \\
&\le P_d^2 \cdot \mathrm{E}\Big[\| < Y(\cdot, T_d), e_v > Y(\cdot, T_d))\|^2\Big] \\
&\quad \text{(by Jensen's inequality; } \|\cdot\|^2 \text{ is a convex fucntion)} \\
&= P_d^2 \cdot \mathrm{E}\Big[| < Y(\cdot, T_d), e_v > |^2 \cdot \|Y(\cdot, T_d))\|^2\Big]
\end{aligned}
$$

It follows that Equation (S10) for the case of $d = d', i \neq i', j, j'$ is equal to

$$
\begin{aligned}
&{M_1}^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'\neq j} \sum_{v=1}^{\infty} \Big| \Big\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \Big\} \Big| \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'\neq j} \sum_{v=1}^{\infty} P_d^2 \cdot \mathrm{E}\Big[| < Y(\cdot, T_d), e_v > |^2 \cdot \|Y(\cdot, T_d))\|^2\Big] \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'\neq j} P_d^2 \cdot \mathrm{E}\Big[\|Y(\cdot, T_d)\|^2 \cdot \sum_{v=1}^{\infty} | < Y(\cdot, T_d), e_v > |^2\Big] \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'\neq j} P_d^2 \cdot \mathrm{E}\Big[\|Y(\cdot, T_d)\|^2 \cdot \|Y(\cdot, T_d)\|^2\Big], \text{ by Parseval's identity} \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'\neq j} P_d^2 \cdot \mathrm{E}\Big[\|Y(\cdot, T_d)\|^4\Big] \\
&= \frac{M_2 - M_1}{M_1^2} \sum_{d=1}^{D} P_d^2 \cdot \mathrm{E}\Big[\|Y(\cdot, T_d)\|^4\Big] \\
&\le \frac{(M_2 - M_1) \cdot \sup_d P_d}{M_1^2} \cdot\ \mathrm{E}_T\Bigg[\mathrm{E}\Big[\|Y(\cdot, T_d)\|^4\Big]\Bigg]
\end{aligned}
$$

<u>Case 1(c): $d = d', i = i', j = j'$</u>

$$
\begin{aligned}
&\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \\
&\quad = P_d \cdot \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{i'd'}(e_v)\big\rangle\Big] - P_d^2 \cdot \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle \\
&\quad = P_d \cdot \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{id}(e_v)\big\rangle\Big] - P_d^2 \cdot \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{id}(e_v)\Big]\Big\rangle
\end{aligned}
$$

$$
\begin{aligned}
&= P_d \cdot \mathrm{E}\Big[\big\|G_{id}(e_v)\big\|^2\Big] - P_d^2 \cdot \Big\|\mathrm{E}\big[G_{id}(e_v)\big]\Big\|^2 \\
&\leq P_d \cdot \mathrm{E}\Big[\big\| < Y(\cdot, T_d), e_v > Y(\cdot, T_d)\big\|^2\Big] \\
&\text{(because } P_d^2 \text{ and } \Big\|\mathrm{E}\big[G_{id}(e_v)\big]\Big\|^2 \text{ are nonnegative)} \\
&= P_d \cdot \mathrm{E}\Big[| < Y(\cdot, T_d), e_v > |^2 \cdot \big\|Y(\cdot, T_d)\big\|^2\Big],
\end{aligned}
$$

Furthermore in this case,

$$
\begin{aligned}
\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}], \mathrm{E}[R_{i'j'd'v}] > &= \mathrm{E}[< R_{ijdv}, R_{ijdv} >] - < \mathrm{E}[R_{ijdv}], \ \mathrm{E}[R_{ijdv}] > \\
&= \mathrm{E}\big[\|R_{ijdv}\|^2\big] - \big\|\mathrm{E}[R_{ijdv}]\big\|^2 = \mathrm{E}\Big[\big\|R_{ijdv} - \mathrm{E}[R_{ijdv}]\big\|^2\Big] \geq 0
\end{aligned}
$$

Thus it follows that Equation (S10) for the case of $d = d', i = i', j = j'$ is equal to

$$
\begin{aligned}
&M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} \Big| \Big\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}], \ \mathrm{E}[R_{i'j'd'v}] > \Big\} \Big| \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} \Big\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}], \ \mathrm{E}[R_{i'j'd'v}] > \Big\} \\
&\text{(because it is shown that the summand in this case is always nonnegative)} \\
&\leq M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} P_d \cdot \mathrm{E}\Big[| < Y(\cdot, T_d), e_v > |^2 \cdot \big\|Y(\cdot, T_d)\big\|^2\Big] \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'=j} P_d \cdot \mathrm{E}\Big[\big\|Y(\cdot, T_d)\big\|^2 \cdot \sum_{v=1}^{\infty} | < Y(\cdot, T_d), e_v > |^2\Big] \\
&= M_1^{-2} \sum_{d,d'=d} \sum_{i,i'=i} \sum_{j,j'=j} P_d \cdot \mathrm{E}\Big[\big\|Y(\cdot, T_d)\big\|^4\Big] \text{ (by Parseval's identity)} \\
&= M_1^{-1} \sum_{d=1}^{D} P_d \cdot \mathrm{E}\Big[\big\|Y(\cdot, T_d)\big\|^4\Big] = \frac{1}{M_1} \cdot \ \mathrm{E}_T\Big[\mathrm{E}\Big[\big\|Y(\cdot, T_d)\big\|^4\Big]\Big]
\end{aligned}
$$

Case 2: $d \neq d'$

First observe the following:

$$\mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d]) \cdot I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] = \begin{cases} P_d P_{d'} & \text{for } i \neq i', j, j' \\ P_d P_{d'} & \text{for } i = i', j \neq j' \\ 0 & \text{for } i = i', j = j' \end{cases}$$

$$\mathrm{E}\Big[I(T_{ij} \in (T_{d-1}, T_d])\Big] \cdot \mathrm{E}\Big[I(T_{i'j'} \in (T_{d'-1}, T_{d'}])\Big] = P_d P'_d \qquad \text{for all } i, i', j, j'$$

Case 2(a): $d \neq d', i \neq i', j, j'$ (Using independence)

$$\begin{aligned} &\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \\ &= P_d P_{d'} \cdot \Big\{ \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{i'd'}(e_v)\big\rangle\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle \Big\} \\ &\text{(by the same reasoning that we use for Case 1(a))} \\ &= P_d P_{d'} \cdot \Big\{ \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{i'd'}(e_v)\Big]\Big\rangle \Big\} \\ &= 0 \end{aligned}$$

Case 2(b): $d \neq d', i = i', j \neq j'$

$$\begin{aligned} &\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \\ &= P_d P_{d'} \cdot \Big\{ \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{id'}(e_v)\big\rangle\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{id'}(e_v)\Big]\Big\rangle \Big\} \end{aligned} \qquad \text{(S12)}$$

Here we consider lower and upper bounds for each term in Equation (S12). By Cauchy-Schwartz inequality,

$$L^{(1)}_{dd'v} \leq P_d P_{d'} \cdot \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{id'}(e_v)\big\rangle\Big] \leq U^{(1)}_{dd'v}, \qquad \text{(S13)}$$

where $U^{(1)}_{dd'v} = P_d P_{d'} \cdot \mathrm{E}\Big[\| < Y(\cdot, T_d), e_v > Y(\cdot, T_d)\| \ \| < Y(\cdot, T_{d'}), e_v > Y(\cdot, T_{d'})\|\Big]$, and $L^{(1)}_{dd'v} = -U^{(1)}_{dd'v}$. Similarly, by Cauchy-Schwartz inequality, the second term in Equation (S12) is bounded below and above:

$$L^{(2)}_{dd'v} \leq P_d P_{d'} \cdot \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{id'}(e_v)\Big]\Big\rangle \leq U^{(2)}_{dd'v}, \qquad \text{(S14)}$$

where $U^{(2)}_{dd'v} = P_d P_{d'} \cdot \Big\|\mathrm{E}\Big[< Y(\cdot, T_d), e_v > Y(\cdot, T_d)\Big]\Big\| \cdot \Big\|\mathrm{E}\Big[< Y(\cdot, T_{d'}), e_v > Y(\cdot, T_{d'})\Big]\Big\|$, and $L^{(2)}_{dd'v} = -U^{(2)}_{dd'v}$.

It follows that, for the case of $d \neq d', i = i', j \neq j'$,

$$L_{dd'v} \leq P_d P_{d'} \cdot \left\{ \mathrm{E}\Big[\big\langle G_{id}(e_v), G_{id'}(e_v) \big\rangle\Big] - \Big\langle \mathrm{E}\Big[G_{id}(e_v)\Big],\ \mathrm{E}\Big[G_{id'}(e_v)\Big] \Big\rangle \right\} \leq U_{dd'v},$$

where

$$U_{dd'v} = U^{(1)}_{dd'v} - L^{(2)}_{dd'v} = U^{(1)}_{dd'v} + U^{(2)}_{dd'v}, \text{ and}$$
$$L_{dd'v} = L^{(1)}_{dd'v} - U^{(2)}_{dd'v} = -U^{(1)}_{dd'v} - U^{(2)}_{dd'v} = -U_{dd'v}.$$

As the lower and upper bounds has the same magnitude, i.e. $U_{dd'v} = |L_{dd'v}|$, it follows that $\big|\mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \big| \leq U_{dd'v}$.

Furthermore using the fact that $(a^2 + b^2)/2 \geq ab$ for any $a, b \in \mathbb{R}$,

$$\begin{aligned}
\sum_{v=1}^{\infty} U^{(1)}_{dd'v} &= \sum_{v=1}^{\infty} P_d P_{d'} \cdot \mathrm{E}\Big[\big\|G_{id}(e_v)\big\| \cdot \big\|G_{id'}(e_v)\big\|\Big] \\
&= \sum_{v=1}^{\infty} P_d P_{d'} \cdot \mathrm{E}\Big[\big\| < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d)\big\| \cdot \big\| < Y_i(\cdot, T_{d'}), e_v > Y_i(\cdot, T_{d'})\big\|\Big] \\
&\leq \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \cdot\ \mathrm{E}\Big[\big\| < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d)\big\|^2 + \big\| < Y_{i'}(\cdot, T_{d'}), e_v > Y_{i'}(\cdot, T_{d'})\big\|^2\Big] \\
&= \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \cdot \left\{\ \mathrm{E}\Big[\big\| < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d)\big\|^2\Big] +\ \mathrm{E}\Big[\big\| < Y_i(\cdot, T_{d'}), e_v > Y_i(\cdot, T_{d'})\big\|^2\Big] \right\} \\
&= \frac{1}{2} P_d P_{d'} \cdot \left\{\ \mathrm{E}\Big[\big\|Y(\cdot, T_d)\big\|^4\Big] +\ \mathrm{E}\Big[\big\|Y(\cdot, T_{d'})\big\|^4\Big] \right\} \qquad \text{(S15)} \\
&\qquad \text{(by Parseval's identity)}
\end{aligned}$$

$$\begin{aligned}
\sum_{v=1}^{\infty} U^{(2)}_{dd'v} &= \sum_{v=1}^{\infty} P_d P_{d'} \cdot \Big\|\mathrm{E}\big[G_{id}(e_v)\big]\Big\| \cdot \Big\|\mathrm{E}\big[G_{id'}(e_v)\big]\Big\| \\
&\leq \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \cdot \left\{ \Big\|\ \mathrm{E}\big[< Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d)\big]\Big\|^2 + \Big\|\ \mathrm{E}\big[< Y_{i'}(\cdot, T_{d'}), e_v > Y_{i'}(\cdot, T_{d'})\big]\Big\|^2 \right\} \\
&\leq \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \cdot \left\{\ \mathrm{E}\Big[\big\| < Y_i(\cdot, T_d), e_v > Y_i(\cdot, T_d)\big\|^2\Big] +\ \mathrm{E}\Big[\big\| < Y_i(\cdot, T_{d'}), e_v > Y_i(\cdot, T_{d'})\big\|^2\Big] \right\} \\
&\qquad \text{(by Jensen's inequality; } \|\cdot\|^2 \text{ is a convex function))} \\
&= \frac{1}{2} P_d P_{d'} \cdot \left\{\ \mathrm{E}\Big[\big\|Y(\cdot, T_d)\big\|^4\Big] +\ \mathrm{E}\Big[\big\|Y(\cdot, T_{d'})\big\|^4\Big] \right\} \qquad \text{(S16)} \\
&\qquad \text{(by Parseval's identity)}
\end{aligned}$$

Then Equation (S10) for the case of $d \neq d', i = i', j \neq j'$ is equal to

$$
\begin{aligned}
& M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j' \neq j} \sum_{v=1}^{\infty} \left| \left\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \right\} \right| \\
& \leq M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j' \neq j} \sum_{v=1}^{\infty} U_{dd'v} = M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j' \neq j} \sum_{v=1}^{\infty} \left\{ U^{(1)}_{dd'v} + U^{(2)}_{dd'v} \right\} \\
& \leq M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j' \neq j} P_d P_{d'} \cdot \left\{ \mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] + \mathrm{E}\left[\|Y(\cdot, T_{d'})\|^4 \right] \right\} \\
& = \frac{M_2 - M_1}{M_1^2} \left\{ \sum_{d,d' \neq d} P_d P_{d'} \cdot \mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] + \sum_{d,d' \neq d} P_d P_{d'}\, \mathrm{E}\left[\|Y(\cdot, T_{d'})\|^4 \right] \right\} \\
& = \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d \cdot \mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \cdot \sum_{d' \neq d} P_{d'} \\
& \qquad + \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d \cdot \left\{ \sum_{d'=1}^{D} P_{d'}\, \mathrm{E}\left[\|Y(\cdot, T_{d'})\|^4 \right] - P_d \cdot \mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \right\} \\
& = \frac{M_2 - M_1}{M_1^2} \cdot \mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \right] - \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d^2 \cdot \mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \\
& \qquad + \frac{M_2 - M_1}{M_1^2} \cdot \mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \right] - \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d^2 \cdot \mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \\
& \leq \frac{2(M_2 - M_1)}{M_1^2} \cdot \mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot, T_d)\|^4 \right] \right]
\end{aligned}
$$

Case 2(c): $d \neq d', i = i', j = j'$

$$
\begin{aligned}
& \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \\
& \qquad = 0 \cdot \mathrm{E}\left[\langle G_{id}(e_v), G_{id'}(e_v) \rangle \right] - P_d P_{d'} \cdot \left\langle \mathrm{E}\left[G_{id}(e_v) \right],\ \mathrm{E}\left[G_{id'}(e_v) \right] \right\rangle \\
& \qquad = -P_d P_{d'} \cdot \left\langle \mathrm{E}\left[G_{id}(e_v) \right],\ \mathrm{E}\left[G_{id'}(e_v) \right] \right\rangle
\end{aligned}
$$

and it follows that

$$
\left| \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \right| = P_d P_{d'} \cdot \left| \left\langle \mathrm{E}\left[G_{id}(e_v) \right],\ \mathrm{E}\left[G_{id'}(e_v) \right] \right\rangle \right| \leq U^{(2)}_{dd'v}
$$

(by Cauchy-Schwartz inequality)

Then Equation (S10) for the case of $d \neq d', i = i', j = j'$ is equal to

$$
M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} \left| \left\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \right\} \right|
$$

$$
\begin{aligned}
&\leq M_1^{-2} \sum_{d,d'\neq d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} U^{(2)}_{dd'v} \\
&\leq M_1^{-2} \sum_{d,d'\neq d} \sum_{i,i'=i} \sum_{j,j'=j} \frac{1}{2} P_d P_{d'} \cdot \left\{ \mathrm{E}\left[\|Y(\cdot,T_d)\|^4\right] + \mathrm{E}\left[\|Y(\cdot,T_{d'})\|^4\right] \right\} \\
&= \frac{1}{2M_1} \left\{ \sum_{d,d'\neq d} P_d P_{d'} \cdot \mathrm{E}\left[\|Y(\cdot,T_d)\|^4\right] + \sum_{d,d'\neq d} P_d P_{d'} \, \mathrm{E}\left[\|Y(\cdot,T_{d'})\|^4\right] \right\} \\
&\leq \frac{1}{M_1} \cdot \mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot,T_d)\|^4\right]\right]
\end{aligned}
$$

Finally by combining all cases, we have

$$
\begin{aligned}
0 \leq \mathrm{E}\left[\|H - \mathrm{E}[H]\|_s^2\right] &= M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} Q_{dd'} \leq M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} |Q_{dd'}| \\
&\leq M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} \sum_{i}^{n} \sum_{i'}^{n} \sum_{j}^{m_i} \sum_{j'}^{m_{i'}} \sum_{v=1}^{\infty} \left| \left\{ \mathrm{E}[< R_{ijdv}, R_{i'j'd'v} >] - < \mathrm{E}[R_{ijdv}],\ \mathrm{E}[R_{i'j'd'v}] > \right\} \right| \\
&\leq \left\{ \frac{(M_2 - M_1)\cdot \sup_d P_d}{M_1^2} + \frac{1}{M_1} + \frac{2(M_2 - M_1)}{M_1^2} + \frac{1}{M_1} \right\} \cdot \mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot,T_d)\|^4\right]\right]. \qquad \text{(S17)}
\end{aligned}
$$

In the case of $Y(s,T) = X(s,T)$ it is implied by the assumptions (A1.), (A2.), and (A4.) that $\mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot,T_d)\|^4\right]\right]$ is finite. Thus Equation (S17) converges to 0 as n diverges. It follows the convergence of $\mathrm{E}\left[\|\widehat{\Sigma}(s,s') - \Sigma(s,s')\|_s^2\right]$ as well as the Hilbert-Schmidt norm consistency.

Corollary 4.1.1 (Corollary 1 in Section 4.1)
Under the assumptions (A1.)-(A5.), for each k we have $|\widehat{\lambda}_k - \lambda_k| \xrightarrow{p} 0$, *and* $\|\widehat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s \xrightarrow{p} 0$ *as n diverges.*

Proof. We prove this corollary using the Hilbert-Schmidt norm consistency result obtained in Theorem 1 of Section 4.1 and Theorem 4.4 and Lemma 4.3 of Bosq (2000, p.104). For completeness the theorem we used is given below:

Theorem 4.4 and Lemma 4.3 (Bosq, 2000, p.104) Let λ_j *and* v_j *be eigenvalues and eigenfunctions of the operator C. Let* λ_{jn} *and* v_{jn} *be the corresponding empirical eigen-elements, such that the operator* $C_n(v_{jn}) = \lambda_{jn} v_{jn}$. *Let* $v'_{jn} = sgn < v_{jn}, v_j > v_j,\ j \geq 1$.

$$
sup_{j\geq 1}|\lambda_{jn} - \lambda_j| \leq \|C_n - C\|_L \leq \|C_n - C\|_s \qquad eq.\ (4.43)
$$

$$
\|v_{jn} - v'_{jn}\| \leq a_j \|C_n - C\|_L \qquad eq.\ (4.44)
$$

where $a_j = 2\sqrt{2}\ max[(\lambda_{j-1} - \lambda_j)^{-1}, (\lambda_j - \lambda_{j+1})^{-1}]$ *if* $j \geq 2$, *and* $a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$.

By Theorem 4.4 of Bosq (2000, p.104),

$$
0 \leq \sup_{k\geq 1}|\widehat{\lambda}_k - \lambda_k| \leq \|\widehat{\Sigma}(s,s') - \Sigma(s,s')\|_L \leq \|\widehat{\Sigma}(s,s') - \Sigma(s,s')\|_s, \text{ and} \qquad \text{(S18)}
$$

$$0 \le \|\widehat{\phi}_k(s) - \phi_k(s)\|_s \le 2\sqrt{2}a_k^{-1}\|\widehat{\Sigma}(s,s') - \Sigma(s,s')\|_s, \tag{S19}$$

where $a_k = (\lambda_1 - \lambda_2)$ for $k = 1$ and $\max[(\lambda_{k-1} - \lambda_k), (\lambda_k - \lambda_{k+1})]$ otherwise. Thus under the assumption (A5.) the following consistency results are implied by the Hilbert-Schmidt norm consistency:

$$|\widehat{\lambda}_k - \lambda_k| \xrightarrow{p} 0 \text{ and } \|\widehat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s \xrightarrow{p} 0, \text{ as } n \to \infty \tag{S20}$$

Theorem 4.1.2 (Theorem 2 in Section 4.1)
Under the assumptions (A1.) - (A6.), for each k $\sup_j|\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0$ *and* $\|\widehat{G}_k(\cdot,\cdot) - G_k(\cdot,\cdot)\|_s \xrightarrow{p} 0$ *as n diverges. In fact a stronger result also holds, namely* $\sup_T|\widehat{G}_k(T,T') - G_k(T,T')| \xrightarrow{p} 0$ *as n diverges.*

Proof. We prove this theorem in two steps: we first show that (i) $\sup_j|\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0$ and then we use the result to show that (ii) $\sup_T|\widehat{G}_k(T,T') - G_k(T,T')| \xrightarrow{p} 0$, which implies the Hilbert-Schmidt norm consistency of $\widehat{G}_k(\cdot,\cdot)$.

(i) The consistency of $\widetilde{\xi}_{W,ijk}$

For each k

$$\sup_j|\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0, \tag{S21}$$

as n diverges, because

$$\begin{aligned}
\sup_j|\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| &= \sup_j\left|\int Y_i(s,T_{ij})\widehat{\phi}_k(s)ds - \int Y_i(s,T_{ij})\phi_k(s)ds\right| \\
&= \sup_j\left|\int Y_i(s,T_{ij})\{\widehat{\phi}_k(s) - \phi_k(s)\}ds\right| \\
&\le \sup_j \int \left|Y_i(s,T_{ij})\right|\left|\{\widehat{\phi}_k(s) - \phi_k(s)\}\right|ds \\
&\le \sup_j \sup_{s\in[0,1]}\left|Y_i(s,T_{ij})\right| \int \left|\{\widehat{\phi}_k(s) - \phi_k(s)\right|\}ds \\
&\le \sup_{j,s}\left|Y_i(s,T_{ij})\right| \times \left\{\int \{\widehat{\phi}_k(s) - \phi_k(s)\}^2 ds\right\}^{1/2} \qquad \text{(by Cauchy-Schwartz ineq.)} \\
&= \sup_{j,s}\left|Y_i(s,T_{ij})\right| \times \|\widehat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s,
\end{aligned} \tag{S22}$$

where $\sup_{j,s}\left|Y_i(s,T_{ij})\right|$ is absolutely bounded almost surely under the assumption (A6.) and $\|\widehat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s$ converges to 0 as obtained in Corollary 1.

(ii) The consistency of $\widehat{G}_k(T,T')$

Recall that $G_k(T,T') = \text{cov}(\xi_{il}(T), \xi_{ik}(T'))$ is the true covariance. It is already shown in Yao et al. (2005) the uniform consistency of its local linear estimator, denoted by $\widehat{G}_{W,k}(T,T')$, when $\widehat{G}_{W,k}(T,T')$ is obtained with $\xi_{W,ijk}$'s; specifically, $\widehat{G}_{W,k}(T,T')$ is obtained by the local linear smoothing of $\{\widetilde{G}_{W,ik}(T_{ij}, T_{ij'}) = \xi_{W,ijk}\xi_{W,ij'k} : j \ne j'\}$. And Yao et al. (2005) showed that $\widehat{G}_{W,k}(T,T')$ is uniformly consistent to the true covariance function, $G_k(T,T')$. Here we show a similar result when the local linear estimator is obtained with $\widetilde{\xi}_{W,ijk}$ instead of $\xi_{W,ijk}$;

The sample covariance of $\widetilde{\xi}_{W,ijk}$'s is as follows:

$$\widetilde{G}_{ik}(T_{ij}, T_{ij'}) = \widetilde{\xi}_{W,ijk}\widetilde{\xi}_{W,ij'k}$$

$$= \{(\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}) + \xi_{W,ijk})\} \times \{(\widetilde{\xi}_{W,ij'k} - \xi_{W,ij'k}) + \xi_{W,ij'k}\}$$
$$= \widetilde{G}_{W,ik}(T_{ij}, T_{ij'}) + (\widetilde{\xi}_{W,ij'k} - \xi_{W,ij'k})(\widetilde{\xi}_{W,ijk} - \xi_{W,ijk})$$
$$+ \xi_{W,ij'k}(\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}) + \xi_{W,ijk}(\widetilde{\xi}_{W,ij'k} - \xi_{W,ij'k})$$

By the uniform consistency of $\widetilde{\xi}_{W,ijk}$ given in Equation (S21), the local linear estimator, $\widehat{G}_k(T, T')$, obtained by smoothing $\widetilde{G}_{ik}(T_{ij}, T_{ij'})$, is asymptotically equivalent to the local linear estimator, $\widehat{G}_{W,k}(T, T')$, obtained by smoothing $\widetilde{G}_{W,ik}(T_{ij}, T_{ij'})$. Thus the uniform consistency of $\widehat{G}_k(T, T')$ is implied by the uniform consistency of $\widehat{G}_{W,k}(T, T')$ shown in Yao et al. (2005). Finally the uniform consistency implies the Hilbert-Schmidt norm consistency.

Corollary 4.1.2 (Corollary 2 in Section 4.1)
Assume (A1.) - (A8.) hold for each k and l. Then the eigenvalues $\widehat{\eta}_{kl}$ and eigenfunctions $\widehat{\psi}_{kl}(\cdot)$ of $\widehat{G}_k(\cdot,\cdot)$ satisfy $|\widehat{\eta}_{kl} - \eta_{kl}| \xrightarrow{p} 0$, and $\|\widehat{\psi}_{kl}(\cdot) - \psi_{kl}(\cdot)\|_s \xrightarrow{p} 0$ as n diverges. Uniform convergence of $\widehat{\psi}_{kl}(\cdot)$ also holds: $\sup_T|\widehat{\psi}_{kl}(T) - \psi_{kl}(T)| \xrightarrow{p} 0$. Furthermore, as n diverges, we have $|\widehat{\sigma}^2_{e,k} - \sigma^2_{e,k}| \xrightarrow{p} 0$ and $|\widehat{\zeta}_{ikl} - \widetilde{\zeta}_{ikl}| \xrightarrow{p} 0$, where $\widetilde{\zeta}_{ikl} = E[\zeta_{ikl}|\boldsymbol{\xi}_{W,ik}]$ and $\boldsymbol{\xi}_{W,ik}$ is the m_i-dimensional column vector of $\xi_{W,ijk}$'s.

Proof. With the same reasoning that we use to show Corollary 2, we can show the consistency of the estimators of eigen-elements from the second FPCA. By Theorem 4.4 of Bosq (2000, p.104),

$$0 \le \sup_{l\ge1}|\widehat{\eta}_{kl} - \eta_{kl}| \le \|\widehat{G}_k(T, T') - G_k(T, T')\|_L \le \|\widehat{G}_k(T, T') - G_k(T, T')\|_s, \text{ and} \tag{S23}$$

$$0 \le \|\widehat{\psi}_{kl}(T) - \psi_{kl}(T)\|_s \le 2\sqrt{2}b_{kl}^{-1}\|\widehat{G}_k(T, T') - G(T, T')\|_s, \tag{S24}$$

where $b_{kl} = (\eta_{k1} - \eta_{k2})$ for $l = 1$ and $\max[(\eta_{k,l-1} - \eta_{k,l}), (\eta_{k,l} - \lambda_{k,l+1})]$ otherwise. Under the assumption (A7.), the consistency of $\|\widehat{G}_k(T, T') - G_k(T, T')\|_s$ implies that the following holds:

$$|\widehat{\eta}_{kl} - \eta_{kl}| \xrightarrow{p} 0, \text{ and } \|\widehat{\psi}_{kl}(\cdot) - \psi_{kl}(\cdot)\|_s \xrightarrow{p} 0 \tag{S25}$$

as n diverges.

Furthermore using Corollary 1, Theorems 2 and 3 of Yao et al. (2005), the consistency of the estimators of $\sigma^2_{e,k}$ and $\widetilde{\zeta}_{ikl}$ as well as the uniform consistency of $\psi_{kl}(\cdot)$ can be shown with the uniform consistency result of $\widehat{G}_k(T, T')$ obtained in Theorem 2 of this paper. The following proofs are similar or identical to the corresponding ones given in Yao et al. (2005).

First we show the uniform consistency of $\psi_{kl}(\cdot)$; the proof is the same as one given in Yao et al. (2005, p. 589). For fixed l,

$$|\widehat{\eta}_{kl}\widehat{\psi}_{kl}(T) - \eta_{kl}\psi_{kl}(T)| = \left|\int \widehat{G}_k(T, T')\widehat{\psi}_{kl}(T')dT' - \int G_k(T, T')\psi_{kl}(T')dT'\right|$$
$$\le \int |\widehat{G}_k(T, T') - G_k(T, T')||\widehat{\psi}_{kl}(T')|dT' + \int |G_k(T, T')||\widehat{\psi}_{kl}(T') - \psi_{kl}(T')|dT'$$
$$\le \sqrt{\int (\widehat{G}_k(T, T') - G_k(T, T'))^2 dT'} + \sqrt{\int (G_k(T, T'))^2 dT'}\|\widehat{\psi}_{kl} - \psi_{kl}\|_s,$$

and the consistency results of $\widehat{G}_k(T, T')$ and $\widehat{\psi}_{kl}(T)$ that we obtained previously imply that $\sup_T|\widehat{\eta}_{kl}\widehat{\psi}_{kl}(T) - \eta_{kl}\psi_{kl}(T)|$ converges to 0 and so does $\sup_T|\widehat{\eta}_{kl}\widehat{\psi}_{kl}(T)/\eta_{kl} - \psi_{kl}(T)|$ as n diverges. By Slutsky's theorem and the consistency of $\widehat{\eta}_{kl}$, it is easy to show that $\sup_T |\widehat{\eta}_{kl}\widehat{\psi}_{kl}(T)/\eta_{kl} - \psi_{kl}(T)|$ is asymptotically equivalent to $\sup_T |\widehat{\psi}_{kl}(T) - \psi_{kl}(T)|$ and thus the uniform consistency of $\widehat{\psi}_{kl}(T)$ holds.

Secondly the consistency of the estimator, $\widehat{\sigma}^2_{e,k}$, follows from the uniform consistency of $\widehat{G}(T,T')$ and that of $\widehat{V}_k(T)$; it is shown in Yao et al. (2005) that the local linear smoother, $\widehat{V}_k(T)$, is a uniform consistent estimator of $G_k(T,T)+\sigma^2_{e,k}$. This result is analogous to Corollary 1 of Yao et al. (2005).

Lastly by Slutsky's theorem and the consistency results of the estimators, $\widehat{\eta}_{kl}$, $\widehat{\boldsymbol{\psi}}_{ikl}$, $\widehat{\Sigma}_{,ik}$ and $\widetilde{\boldsymbol{\xi}}_{W,ik}$, we have

$$|\widehat{\zeta}_{ikl} - \widetilde{\zeta}_{ikl}| = \left|\widehat{\mathrm{E}[\zeta_{ikl}|\widetilde{\boldsymbol{\xi}}_{W,ik}]} - \mathrm{E}[\zeta_{ikl}|\boldsymbol{\xi}_{W,ik}]\right| \xrightarrow{p} 0 \tag{S26}$$

as n diverges.

Theorem 4.1.3 (Theorem 3 in Section 4.1)
Assume (A1.) - (A8.), for each $(s,T)\in\mathcal{S}\times\mathcal{T}$. Then $\widehat{Y}_i(s,T) \xrightarrow{p} \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)$ as n, K and L_k's $\to\infty$.

Proof. Let $\widetilde{Y}_i(s,T)=\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)$ be the full (not-truncated) true trajectory. And let $\widetilde{Y}_i^K(s,T)=\sum_{k=1}^{K}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)$ and $\widetilde{Y}_i^{KL}(s,T)=\sum_{k=1}^{K}\sum_{l=1}^{L_k}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)$ be the truncated versions of $\widetilde{Y}_i(s,T)$. Because $|\widehat{Y}_i(s,T)-\widetilde{Y}_i(s,T)|\le|\widehat{Y}_i(s,T)-\widetilde{Y}_i^{KL}(s,T)|+|\widetilde{Y}_i^{KL}(s,T)-\widetilde{Y}_i(s,T)|$, we prove the consistency of $\widehat{Y}_i(s,T)$ by showing that each of $|\widehat{Y}_i(s,T)-\widetilde{Y}_i^{KL}(s,T)|$ and $|\widetilde{Y}_i^{KL}(s,T)-\widetilde{Y}_i(s,T)|$ converges to 0 in probability. Using Slutsky's theorem and the consistency results obtained for the model components estimators, it is easy to show that $|\widehat{Y}_i(s,T)-\widetilde{Y}_i^{KL}(s,T)|$ converges to 0 in probability as n diverges.

In the following we show that the second term, $|\widetilde{Y}_i(s,T)-\widetilde{Y}_i^{KL}(s,T)|$, converges to 0 in probability by proving that $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]$ converges to 0 as K and L_k's diverge. Because

$$\begin{aligned}\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2 &= \|\{\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{K}(\cdot,\cdot)\}+\{\widetilde{Y}_i^{K}(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\}\|^2\\ &\le 2\big\{\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{K}(\cdot,\cdot)\|^2+\|\widetilde{Y}_i^{K}(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big\},\end{aligned}$$

we prove the convergence of $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]$ to 0 by proving that each of $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{K}(\cdot,\cdot)\|^2\big]$ and $\mathrm{E}\big[\|\widetilde{Y}_i^{K}(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]$ converges to 0 as K and L_k's diverge.

Let $\boldsymbol{\psi}_{ikl}=\{\widehat{\psi}_{kl}(T_{i1}),\ldots,\widehat{\psi}_{kl}(T_{im_i})\}^T$ be the m_i-dimensional column vector of the evaluations of $\psi_{kl}(\cdot)$ at $\{T_{ij}: j=1,\ldots,m_i\}$, $\Sigma_{\xi_{W,ik}}$ be a $m_i\times m_i$ - matrix with (j,j')th element equal to $G_k(T_{ij},T_{ij'})+\sigma^2_{e,k}$, for $j=j'$ and $G_k(T_{ij},T_{ij'})$ otherwise, and $\boldsymbol{\xi}_{W,ik}$ be the m_i - dimensional column vector of $\xi_{W,ijk}$'s; note that these are the true parameters corresponding to the estimators, $\widehat{\boldsymbol{\psi}}_{ikl}$, $\widehat{\Sigma}_{\xi_{W,ik}}$ and $\widetilde{\boldsymbol{\xi}}_{W,ik}$, defined in Section 3.3. Now consider the first term, $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{K}(\cdot,\cdot)\|^2\big]$:

$$\begin{aligned}\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^{K}(\cdot,\cdot)\|^2\big] &= \mathrm{E}\Big[\iint\Big\{\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)-\sum_{k=1}^{K}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)\Big\}^2 ds dT\Big]\\ &= \mathrm{E}\Big[\iint\Big\{\sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)\Big\}^2 ds dT\Big]\\ &= \mathrm{E}\Big[\sum_{k=K+1}^{\infty}\int\Big\{\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\Big\}^2 dT\Big] \qquad \text{(using the orthonormal property of } \phi_k(\cdot))\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\Big[\sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\{\widetilde{\zeta}_{ikl}\}^2\Big] \qquad \text{(using the orthonormal property of } \psi_{lk}(\cdot)) \\
&= \mathrm{E}\Big[\sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\big\{\mathrm{E}[\zeta_{ikl}|\boldsymbol{\xi}_{W,ik}]\big\}^2\Big] \\
&= \sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\mathrm{var}\big\{\mathrm{E}[\zeta_{ikl}|\boldsymbol{\xi}_{W,ik}]\big\} \qquad \text{(using the Monotone Convergence Theorem)} \\
&= \sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\eta_{kl}^2\boldsymbol{\psi}_{ikl}^T\Sigma_{\xi_{W,ik}}^{-1}\boldsymbol{\psi}_{ikl} \qquad \text{(under Gaussian assumption)} \\
&= \sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\eta_{kl}^2\cdot tr\Big[\boldsymbol{\psi}_{ikl}^T\Sigma_{\xi_{W,ik}}^{-1}\boldsymbol{\psi}_{ikl}\Big] \qquad \text{(because trace } tr(a)=a, a\in\mathbb{R}) \\
&= \sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\eta_{kl}^2\cdot tr\Big[\boldsymbol{\psi}_{ikl}^T\Big\{\sum_{l'=1}^{\infty}\eta_{kl'}\boldsymbol{\psi}_{ikl'}\boldsymbol{\psi}_{ikl'}^T+\sigma_{e,k}^2\mathbf{I}_{m_i}\Big\}^{-1}\boldsymbol{\psi}_{ikl}\Big] \\
&\qquad \text{(using the KL expansion of } G_k(T,T')) \\
&= \sum_{k=K+1}^{\infty} tr\Big[\Big\{\sum_{l'=1}^{\infty}\eta_{kl'}\boldsymbol{\psi}_{ikl'}\boldsymbol{\psi}_{ikl'}^T+\sigma_{e,k}^2\mathbf{I}_{m_i}\Big\}^{-1}\sum_{l=1}^{\infty}\eta_{kl}^2\boldsymbol{\psi}_{ikl}\boldsymbol{\psi}_{ikl}^T\Big] \\
&\qquad \text{(by the linear and cyclic properties of trace)} \\
&\le \sum_{k=K+1}^{\infty}\eta_{k1}\cdot tr\Big[\Big\{\sum_{l'=1}^{\infty}\eta_{kl'}\boldsymbol{\psi}_{ikl'}\boldsymbol{\psi}_{ikl'}^T+\sigma_{e,k}^2\mathbf{I}_{m_i}\Big\}^{-1}\sum_{l=1}^{\infty}\eta_{kl}\boldsymbol{\psi}_{ikl}\boldsymbol{\psi}_{ikl}^T\Big] \\
&\qquad \text{(because } \{\eta_{kl}\}_l \text{ is a strictly decreasing sequence for each } k) \\
&= \sum_{k=K+1}^{\infty}\eta_{k1}\cdot tr\Big[\mathbf{I}_{m_i}-\Big\{\sum_{l'=1}^{\infty}\eta_{kl'}\boldsymbol{\psi}_{ikl'}\boldsymbol{\psi}_{ikl'}^T+\sigma_{e,k}^2\mathbf{I}_{m_i}\Big\}^{-1}\sigma_{e,k}^2\mathbf{I}_{m_i}\Big] \\
&\le \sum_{k=K+1}^{\infty}\eta_{k1}tr(\mathbf{I}_{m_i}) \\
&\qquad \text{(because trace of a positive definite matrix is positive)} \\
&= \sum_{k=K+1}^{\infty}\eta_{k1}m_i \;\le\; m_i\sum_{k=K+1}^{\infty}\sum_{l=1}^{\infty}\eta_{kl}\longrightarrow 0,
\end{aligned}
$$

as K diverges because m_i is finite and the square integrable property of $X_i(s,T)$ in the assumption (A1.) ensures that $\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\eta_{kl}$ is finite. Thus we show that $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot)-\widetilde{Y}_i^K(\cdot,\cdot)\|^2\big]$ converges to 0 as K diverges.

We use the similar reasoning to show that $\mathrm{E}\big[\|\widetilde{Y}_i^K(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]$ converges to 0 as K and L_k's diverges. Specifically from the previous part, we use the fact that for each k $\sum_{l=1}^{\infty}\eta_{kl}^2\cdot tr\Big\{\big(\sum_{l'=1}^{\infty}\eta_{kl'}\boldsymbol{\psi}_{ikl'}\boldsymbol{\psi}_{ikl'}^T+\sigma_{e,k}^2\mathbf{I}_{m_i}\big)^{-1}\boldsymbol{\psi}_{ikl}\boldsymbol{\psi}_{ikl}^T\Big\}\le\eta_{k1}m_i$ is finite. Most steps are identical thus omitted.

$$
\mathrm{E}\big[\|\widetilde{Y}_i^K(\cdot,\cdot)-\widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]=\mathrm{E}\Big[\iint\Big\{\sum_{k=1}^{K}\sum_{l=1}^{\infty}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)-\sum_{k=1}^{K}\sum_{l=1}^{L_k}\widetilde{\zeta}_{ikl}\psi_{kl}(T)\phi_k(s)\Big\}^2 dsdT\Big]
$$

$$= \mathrm{E}\Big[\iint \Big\{ \sum_{k=1}^{K} \sum_{l=L_k+1}^{\infty} \widetilde{\zeta}_{ikl} \psi_{kl}(T) \phi_k(s) \Big\}^2 ds dT \Big]$$

$$\leq \sum_{k=1}^{K} \sum_{l=L_k+1}^{\infty} \eta_{kl}^2 \cdot tr\Big\{ \big(\sum_{l'=1}^{\infty} \eta_{kl'} \boldsymbol{\psi}_{ikl'} \boldsymbol{\psi}_{ikl'}^T + \sigma_{e,k}^2 \mathbf{I}_{m_i} \big)^{-1} \boldsymbol{\psi}_{ikl} \boldsymbol{\psi}_{ikl}^T \Big\} \longrightarrow 0,$$

as K and L_k's diverge because for each k, $\sum_{l=L_k+1}^{\infty} \eta_{kl}^2 \cdot tr\Big\{ \big(\sum_{l'=1}^{\infty} \eta_{kl'} \boldsymbol{\psi}_{ikl'} \boldsymbol{\psi}_{ikl'}^T + \sigma_{e,k}^2 \mathbf{I}_{m_i} \big)^{-1} \boldsymbol{\psi}_{ikl} \boldsymbol{\psi}_{ikl}^T \Big\}$ converges to 0 as L_k diverges.

The convergence of $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot) - \widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]$ to 0 follows from the fact that $\mathrm{E}\big[\|\widetilde{Y}_i(\cdot,\cdot) - \widetilde{Y}_i^{K}(\cdot,\cdot)\|^2\big]$ and $\mathrm{E}\big[\|\widetilde{Y}_i^{K}(\cdot,\cdot) - \widetilde{Y}_i^{KL}(\cdot,\cdot)\|^2\big]$ converge to 0. Furthermore by Markov inequality it is implied that $|\widetilde{Y}_i(s,T) - \widetilde{Y}_i^{KL}(s,T)|$ converges to 0 in probability as K and L_k's diverge.

S2.2. Case when response curves are measured with smooth error (Section 4.2)

Now we consider the case when data is corrupted with smooth error process, $\epsilon_{1,ij}(s)$; in other words, $Y_{ij}(s) = X_i(s, T_{ij}) + \epsilon_{1,ij}(s)$. The main difference from the case of having $Y_{ij}(s) = X_i(s, T_{ij})$ is that the sample covariance of $Y_{ij}(s)$ is no longer an estimator of the marginal covariance function, $\Sigma(s,s')$, but is an estimator of $\Xi(s,s') = \Sigma(s,s') + \Gamma(s,s')$, where $\Gamma(s,s')$ is a smooth covariance function of the error process, $\epsilon_{1,ij}(s)$.

As the additional error process, $\epsilon_{1,ij}(s)$, are correlated over s, but independent over i and j, the theoretical results associated with the longitudinal dynamics (the second FPCA step) and their proofs remain the same. In the following we provide proofs of the theoretical results given in Section 4.2 that correspond to the marginal FPCA step.

We start our proofs with showing that $\widehat{\Xi}(s,s')$ is an unbiased estimator of $\Xi(s,s')$.

Corollary 4.0.3
The sample covariance, $\widehat{\Xi}(s,s')$, is an unbiased estimator of the covariance function, $\Xi(s,s') = \Sigma(s,s') + \Gamma(s,s')$.

Proof. Proof is identical to that of Corollary 4.0.2 except that $\mathrm{E}\Big\{Y_i(s,T_d)Y_i(s',T_d)\Big\} = \mathrm{E}\Big[\{X_i(s,T_d) + \epsilon_{1,ij}(s)\}\{X_i(s',T_d) + \epsilon_{1,ij}(s')\}\Big] = c((s,T_d),(s',T_d)) + \Gamma(s,s')$, instead of $c((s,T_d),(s',T_d))$ only. Specifically,

$$\mathrm{E}\{\widehat{\Xi}(s,s')\} = \mathrm{E}\Big\{ \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(s) Y_{ij}(s') \Big\}$$

$$= \mathrm{E}\Big\{ \frac{1}{\sum_{i=1}^n m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(s) Y_{ij}(s') I(T_{ij} \in (T_{d-1}, T_d]) \Big\}$$

$$= \frac{1}{\sum_{i=1}^n m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathrm{E}\Big\{Y_i(s,T_d)Y_i(s',T_d)\Big\} \mathrm{E}\Big\{ I(T_{ij} \in (T_{d-1}, T_d]) \Big\} \qquad \text{(because } T \perp\!\!\!\perp Y_i(\cdot, T=t))$$

$$= \frac{1}{\sum_{i=1}^n m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathrm{E}\Big[\{X_i(s,T_d) + \epsilon_{1,ij}(s)\}\{X_i(s',T_d) + \epsilon_{1,ij}(s')\}\Big] \mathrm{E}\Big\{ I(T_{ij} \in (T_{d-1}, T_d]) \Big\}$$

$$= \frac{1}{\sum_{i=1}^n m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{c((s,T_d),(s',T_d)) + \Gamma(s,s')\} \mathrm{P}(T \in (T_{d-1}, T_d]) \qquad \text{(because } X_i(\cdot,\cdot) \perp\!\!\!\perp \epsilon_{1,ij}(\cdot))$$

$$= \sum_{d=1}^{D} \{c((s, T_d), (s', T_d)) + \Gamma(s, s')\} \mathrm{P}(T \in (T_{d-1}, T_d]) = \int \{c((s, T_d), (s', T_d)) + \Gamma(s, s')\} dF(t)$$
$$= \int c((s, T_d), (s', T_d)) dF(t) + \Gamma(s, s') \qquad \text{(because } \int dF(t) = 1)$$
$$= \Sigma(s, s') + \Gamma(s, s') = \Xi(s, s'),$$

where F is a true sampling distribution function of T.

Corollary 4.2.3 (Corollary 3 in Section 4.2)
Under the assumptions (A1.) - (A3.), and (A9.), for each (s, s'), $|\widehat{\Xi}(s, s') - \Xi(s, s')| \xrightarrow{p} 0$ *as n diverges. And under the assumptions (A1.), (A2.), (A4.), (A9.)-(A11.),* $\|\widehat{\Xi}(\cdot, \cdot) - \Xi(\cdot, \cdot)\|_s \xrightarrow{p} 0$ *and* $\sup_j |\widetilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0$ *as* $n \longrightarrow \infty$.

Proof. This corollary includes three consistency results that are corresponding to the ones given in Theorems 1 and 2. As proofs are very similar to the corresponding ones detailed earlier many steps are omitted. In the following we give a short proofs for (i) the pointwise consistency of $\widehat{\Xi}(s, s')$, (ii) the Hilbert-Schmidt norm consistency of $\widehat{\Xi}(s, s')$, and (iii) the uniform consistency of $\widetilde{\xi}_{W,ijk}$.

(i) The pointwise consistency of $\widehat{\Xi}(s, s')$

To show the pointwise consistency of $\widehat{\Xi}(s, s')$, we need to show that the following converges to 0 as n diverges:

$$\frac{1}{D^2}\Big[\sum_{d=1}^{D} \sigma_d^2 + \sum_{d=1}^{D} \sum_{d' \neq d} \sigma_{dd'}\Big] = \left[\frac{1}{M_1} \mathrm{E}_T[E[A(T)^2|T]]\right] + \left[\left\{\frac{M_2}{M_1^2} - \frac{1}{M_1}\right\} \mathrm{E}_T[\mathrm{E}_{T'}[\mathrm{E}[A(T)A(T')|T, T']]]\right]$$
$$- \left[\left\{\frac{M_2}{M_1^2} - \frac{1}{M_1}\right\} \sum_{d=1}^{D} P_d^2 \cdot \mathrm{E}[A(T)^2|T = T_d]\right] - \frac{M_2}{M_1^2}\{\Sigma(s, s') + \Gamma(s, s')\}^2, \qquad \text{(S27)}$$

which is analogous to Equation (S7).

Because $X_i(s, T_{ij})$'s and $\epsilon_{ij}(s)$'s are mutually independent and $\epsilon_{ij}(s)$'s are independent over i and j,

$$\mathrm{E}[A(T)A(T')|T, T'] = \mathrm{E}[Y_{ij}(s)Y_{ij}(s')Y_{ij'}(s)Y_{ij'}(s')]$$
$$= \mathrm{E}[X_i(s, T_{ij})X_i(s', T_{ij})X_i(s, T_{ij'})X_i(s', T_{ij'})] + \{\Gamma(s, s')\}^2$$
$$+ \Gamma(s, s')c((s, T_{ij}), (s, T_{ij'})) + \Gamma(s, s')c((s, T_{ij'}), (s', T_{ij'})).$$

Thus under the assumptions (A3.) and (A9.), $\mathrm{E}[A(T)A(T')|T, T']$ is finite. Then using the same argument that we used to show the convergence of Equation (S7), we can show that Equation (S27) converges to 0 as n diverges. It implies that for each (s, s'), $|\widehat{\Xi}(s, s') - \Xi(s, s')| \xrightarrow{p} 0$ as n diverges.

(ii) The Hilbert-Schmidt norm consistency of $\widehat{\Xi}(s, s')$

To show the Hilbert-Schmidt norm consistency, we need to show that the following converges to 0 as n diverges:

$$\left\{\frac{(M_2 - M_1) \cdot \sup_d P_d}{M_1^2} + \frac{1}{M_1} + \frac{2(M_2 - M_1)}{M_1^2} + \frac{1}{M_1}\right\} \cdot \ \mathrm{E}_T\left[\mathrm{E}\left[\|Y(\cdot, T_d)\|^4\right]\right]. \qquad \text{(S28)}$$

This is the same equation as Equation (S17). Because $X_i(s, T_{ij})$'s and $\epsilon_{ij}(s)$'s are mutually independent and $\epsilon_{ij}(s)$'s are independent over i and j,

$$\mathrm{E}\left[\|Y_{ij}(\cdot)\|^4 \mid T = T_{ij}\right] = \mathrm{E}\left[\|X_i(\cdot, T)\|^4 \mid T = T_{ij}\right] + \mathrm{E}\left[\|\epsilon_{1,ij}(\cdot)\|^4\right]$$

$$+\,\mathrm{E}\big[\|X_i(\cdot,T)\|^2 \mid T=T_{ij}\big]\cdot \mathrm{E}\big[\|\epsilon_{1,ij}(\cdot)\|^2\big] + 4\iint c((s,T_{ij}),(s',T_{ij})\Gamma(s,s')dsds',$$

where the first three terms are finite under the assumptions (A4.) and (A10.) and the last term is finite because by Cauchy-Schwarz inequality

$$\iint c((s,T_{ij}),(s',T_{ij})\Gamma(s,s')dsds' < \sqrt{\iint c((s,T_{ij}),(s',T_{ij}))^2dsds'}\sqrt{\iint \Gamma(s,s')^2dsds'}.$$

Thus $\mathrm{E}\big[\big\|Y(\cdot,T)\big\|^4|T\big]$ is also finite and under the assumptions (A1.), (A2.), (A4.) and (A10.) Equation (S28) converges to 0 as n diverges. Then it follows that $\|\widehat{\Xi}(\cdot,\cdot)-\Xi(\cdot,\cdot)\|_s$ converges to 0 in probability as n diverges.

(iii) The uniform consistency of $\widetilde{\xi}_{W,ijk}$

Lastly the uniform consistency of $\widetilde{\xi}_{W,ijk}$ holds because from Equation (S22) we have $\sup_j|\widetilde{\xi}_{W,ijk}-\xi_{W,ijk}|\leq \sup_{j,s}|Y_i(s,T_{ij})|\times\|\widehat{\phi}_k(\cdot)-\phi_k(\cdot)\|_s$, which is less than or equal to $\{\sup_{j,s}|X_i(s,T_{ij})|+\sup_s|\epsilon_{1,ij}(s)|\}\times\|\widehat{\phi}_k(\cdot)-\phi_k(\cdot)\|_s$, where $\sup_{j,s}|X_i(s,T_{ij})|$ and $\sup_s|\epsilon_{1,ij}(s)|$ are asymptotically bounded under the assumptions (A6.) and (A11.).

S2.3. Extensions

The theoretical results presented in Section 4 are based on the assumption that data $Y_{ij}(s)$'s are observed (i) fully, (ii) without white noise, $\epsilon_{2,ij}(s)\equiv 0$ for all s, and (iii) have mean zero. These assumptions are made for convenience, and they can be relaxed as we now explain.

(i) The assumption that $Y_{ij}(s)$'s are observed fully, as a continuous function, is quite common in theoretical study involving functional data; see for example Cardot et al. (2003, 2004); Chen & Müller (2012) among many others. One possibility to bypass this assumption is to use the corresponding smooth trajectories instead. (ii) Suppose that the profiles $Y_{ij}(\cdot)$'s are observed on dense grids of points and the measurements are additionally corrupted with white noise. Zhang & Chen (2007) showed that by smoothing each profile using local linear smoother, the true de-noised curves are recovered with asymptotically negligible error, in the case of independent curves. Another possibility to handle white noise is to use ideas similar to Yao et al. (2005). In Section 5 we illustrate numerically the effect of white noise on the performance accuracy. (iii) Finally, the theoretical properties of the model components estimators remain valid, when the mean function is non-zero and a consistent mean estimator is available; Chen & Müller (2012) had considered this problem and showed that under suitable assumptions such consistent mean estimator can be obtained using bivariate smoothing.

S3. Additional details for the simulation experiment

S3.1. Description of the study

Errors are generated from $\epsilon_{ij}(s) = e_{ij1}\phi_1(s) + e_{ij2}\phi_2(s) + \epsilon_{2,ij}(s)$, where e_{ij1}, e_{ij2} and $\epsilon_{2,ij}(s)$ are mutually independent with zero-mean and variances equal to $\sigma^2_{e,1}, \sigma^2_{e,2}$ and σ^2, respectively. The white noise variance, σ^2, is set based on the signal to noise ratio (SNR),

$$\mathrm{SNR} = \frac{\iint \mathrm{var}\{\mathrm{Y_i(s,T)}\}\mathrm{dsdT}}{(œ^2_{\mathrm{e},1} + œ^2_{\mathrm{e},2} + œ^2)} - 1. \quad \text{(S29)}$$

We consider the following experimental factors:

Case 1. covariance structure of the time-varying components:
(a) non-parametric covariance (NP): $\xi_{ik}(T) = \zeta_{ik1}\psi_{k1}(T) + \zeta_{ik2}\psi_{k2}(T)$, where
(i) $\psi_{11}(T) = \sqrt{2}\cos(2ßT)$, $\psi_{12}(T) = \sqrt{2}\sin(2ßT)$, $\zeta_{i11} \overset{\text{iid}}{\sim} \mathrm{N}(0,3)$, $\zeta_{i12} \overset{\text{iid}}{\sim} \mathrm{N}(0,1.5)$;
(ii) $\psi_{21}(T) = \sqrt{2}\cos(4ßT)$, $\psi_{22}(T) = \sqrt{2}\sin(4ßT)$, $\zeta_{i21} \overset{\text{iid}}{\sim} \mathrm{N}(0,2)$, $\zeta_{i22} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$.
(b) random effects model (REM): $\xi_{ik}(T) = b_{ik0} + b_{ik1}T$ with

$$\begin{pmatrix} b_{i10} \\ b_{i11} \end{pmatrix} \overset{\text{iid}}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.5 & 2 \\ 2 & 3 \end{pmatrix}\right], \begin{pmatrix} b_{i20} \\ b_{i21} \end{pmatrix} \overset{\text{iid}}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1.5 \end{pmatrix}\right],$$

(c) exponential autocorrelation model (Exp): $\xi_{ik}(T)$ is a Gaussian process with mean zero, variance λ_k and auto-correlation function $\mathrm{corr}\{ ,_{\mathrm{ik}}(\mathrm{T}), ,_{\mathrm{ik}}(\mathrm{T}')\} = æ_{\mathrm{k}}^{|\mathrm{T}-\mathrm{T}'|}$ denoted by $GP(\lambda_k, \rho_k)$. We set $\xi_{i1}(T) \overset{\text{iid}}{\sim} GP(4.5, 0.9)$ and $\xi_{i2}(T) \overset{\text{iid}}{\sim} GP(3, 0.5)$.
Note that regardless of the generating models for $\xi_{ik}(T)$, we have that $\int \mathrm{var}\{ ,_{\mathrm{ik}}(\mathrm{T})\}\mathrm{dT}$ is equal to 4.5 and 3 for $k = 1, 2$ respectively.

Case 2. number of repeated measurements per subject:
(a) $m_i \overset{\text{iid}}{\sim} \mathrm{Uniform}(\{8, 9, \ldots, 12\})$ (about 75% missing)
(b) $m_i \overset{\text{iid}}{\sim} \mathrm{Uniform}(\{15, 16, \ldots, 20\})$ (about 55% missing)

Case 3. variance of e_{ijk}:
(a) $(\sigma^2_{e,1}, \sigma^2_{e,2}) = (0, 0)$ (white noise only, i.e. $\epsilon_{ij}(s) \overset{\text{iid}}{\sim} N(0, \sigma^2)$)
(b) $(\sigma^2_{e,1}, \sigma^2_{e,2}) = (0.7, 0.3)$.

The simulation results for the Case 3.(a), i.e. no smooth error, are included in the Supplementary Material.

Case 4. signal to noise ratio: (a) SNR = 1 ($œ^2 = 6.5$), and (b) SNR = 5 ($œ^2 = 0.5$)

Case 5. number of subjects: (a) $n = 100$, (b) $n = 300$, and (c) $n = 500$

For each generated sample of size n we form a training set and a test set. To determine the test set we randomly select 10 subjects from the sample. The test set is formed by collecting these subjects' last functional observation; hence the test set contains 10 curves. The remaining functional observations for the 10 subjects and the data corresponding to the remaining subjects in the sample form the training set. Our model is fitted using the training set and the methods outlined in Section 3. To be more specific, the bivariate mean function, $\mu(s, T)$, is modeled using 50 cubic spline basis functions obtained from the tensor product of $d_s = 10$ basis functions in direction s and $d_T = 5$ in T. The smoothing parameters are selected via REML. The finite truncations K and L_k's are all estimated using the pre-specified level PVE = 0.95.

Estimation accuracy for the model components is evaluated using integrated mean squared errors (IMSE): specifically, for the bivariate mean function $\mathrm{IMSE}(\widehat{\ }) = \sum_{\mathrm{i_{sim}}=1}^{\mathrm{N_{sim}}} \iint \{\widehat{\mu}^{i_{sim}}(s,T) - \mu(s,T)\}^2 dsdT/N_{sim}$, and for the univariate eigenfunctions $\mathrm{IMSE}(\widehat{Œ}_\mathrm{k}) = \sum_{\mathrm{i_{sim}}=1}^{\mathrm{N_{sim}}} \int \{\widehat{\phi}_k^{i_{sim}}(s) - \phi_k(s)\}^2 ds/N_{sim}$, $k = 1, 2$. The prediction performance is assessed through the accuracy in predicting the time-varying model components, $\xi_{ik}(T)$, and in predicting the response curve, $Y_i(s,T)$. For the former assessment we use the in-sample integrated prediction errors (IPE) defined as $\mathrm{IPE}(,_\mathrm{k}) = \sum_{\mathrm{i_{sim}}=1}^{\mathrm{N_{sim}}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \int \{\widehat{,}_{\mathrm{ik}}^{\mathrm{i_{sim}}}(\mathrm{T}) - ,_{\mathrm{ik}}^{\mathrm{i_{sim}}}(\mathrm{T})\}^2 \mathrm{dT}/\ (nN_{sim})$, $k = 1, 2$. For the later assessment we use the in-sample IPE (IN-IPE) defined as $\mathrm{IN-IPE(Y)} = \sum_{\mathrm{i_{sim}}=1}^{\mathrm{N_{sim}}} \sum_{\mathrm{i}}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{m_i}} \int \{\widehat{\mathrm{Y}}_{\mathrm{ij}}^{\mathrm{i_{sim}}}(\mathrm{s}) - \mathrm{Y}_{\mathrm{ij}}^{*,\mathrm{i_{sim}}}(\mathrm{s})\}^2 \mathrm{ds}/(\mathrm{N_{sim}} \sum_{i=1}^{n} m_i)$, where $Y_{ij}^*(s)$ is the true signal, i.e. without measurement error $\epsilon_{ij}(s)$. Also we use the out-of-sample IPE (OUT-IPE) defined as $\mathrm{OUT-IPE(Y)} = \sum_{\mathrm{i_{sim}}=1}^{\mathrm{N_{sim}}} \sum_{\mathrm{i \in test\ set}} \int \{\widehat{\mathrm{Y}}_{\mathrm{im_i}}^{\mathrm{i_{sim}}}(\mathrm{s}) - \mathrm{Y}_{\mathrm{im_i}}^{*,\mathrm{i_{sim}}}(\mathrm{s})\}^2 \mathrm{ds}/(10\mathrm{N_{sim}})$ and $Y_{ij}^*(s)$ is the true signal in the test set. The results are based on $N_{sim} = 1000$ simulations.

In terms of estimation performance and prediction of $\xi_{ik}(T)$ there is no alternative approach. On the other hand, in terms of model prediction error and prediction of a subject's future curve there are two possible alternatives. One is the CM model of Chen & Müller (2012). However due to the high computational expense required by their method, we have to restrict our comparison to few scenarios only: $m_i \sim \{8, \ldots, 12\}$ number of repeated curves per subject, Case 3(b), and SNR $= 1$. The approach of Chen & Müller (2012) requires specification of several kernel bandwidths; due to the increased computation burden we use the pre-specified bandwidth $h = 0.1$ in smoothing both the mean and covariance functions. Even with these adjustments there is an order of magnitude difference in the computational cost (when $n = 100$ the method of Chen & Müller (2012) takes approximately 984 seconds, while our approach takes about 7 seconds). As well, we also used the pre-specified level PVE $= 0.95$ to be consistent with our approach. A second alternative approach for prediction of a subject's future visit trajectory is a rather naïve approach: let the future prediction equal the average of all previously observed profiles for that subject. For example, the naïve predictor of a profile of some subject in the test set is equal to the average of all profiles available in the training set for the corresponding subject.

S3.2. Additional results

Simulation results for the case when the error process has trivial covariance structure are presented in Table S1. The results show that the prediction accuracy is greatly improved when the error process is just white noise. The other factors we investigated in the study seem to have the similar effects on the estimation and prediction accuracies as they did for the case of having smooth error process (Table 1 in Section 5). Based on the results presented in both Tables 1 and S1, it seems that the exponential covariance structure (Exp) is most challenging; this most likely is due to the (very) large correlation coefficients used $\rho_1 = 0.9$ and $\rho_2 = 0.5$, which result in high temporal dependence even for the observations that are furthest apart $T = 0$ and $T = 1$, as $T \in [0, 1]$. For completeness the average computational times of the proposed method are studied for all the cases and they are presented in Tables S2 and S3.

Table S1. Simulation results for estimation and prediction accuracy based on $N_{sim} = 1000$ simulations (white noise only)

		μ	ϕ_1	ϕ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
$m_i \overset{iid}{\sim} \{8,\ldots,12\}$ and SNR $=1$										
NP (a)	$n=100$	0.088	0.005	0.014	0.169	0.150	0.173	7.779	0.697	11.114
	$n=300$	0.029	0.001	0.010	0.067	0.066	0.109	7.788	0.288	11.395
	$n=500$	0.017	0.001	0.010	0.048	0.047	0.092	7.778	0.196	11.173
REM (b)	$n=100$	0.110	0.035	0.043	0.299	0.303	0.193	1.197	0.705	2.140
	$n=300$	0.037	0.009	0.018	0.137	0.135	0.162	1.198	0.467	2.161
	$n=500$	0.022	0.005	0.014	0.108	0.102	0.156	1.198	0.408	2.131
Exp (c)	$n=100$	0.091	0.032	0.041	0.330	0.526	0.386	1.529	1.255	2.494
	$n=300$	0.030	0.009	0.018	0.209	0.358	0.353	1.529	1.000	2.472
	$n=500$	0.017	0.005	0.014	0.189	0.323	0.344	1.529	0.898	2.481
$m_i \overset{iid}{\sim} \{15,\ldots,20\}$ and SNR $=1$										
		μ	ϕ_1	ϕ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
NP (a)	$n=100$	0.073	0.002	0.011	0.092	0.062	0.119	7.803	0.318	10.739
	$n=300$	0.025	<0.001	0.009	0.037	0.029	0.073	7.798	0.145	10.937
	$n=500$	0.015	<0.001	0.009	0.026	0.023	0.063	7.791	0.112	10.620
REM (b)	$n=100$	0.093	0.032	0.041	0.251	0.269	0.174	0.895	0.493	1.818
	$n=300$	0.032	0.009	0.018	0.120	0.124	0.156	0.896	0.372	1.876
	$n=500$	0.019	0.005	0.014	0.095	0.094	0.151	0.897	0.347	1.811
Exp (c)	$n=100$	0.078	0.030	0.039	0.276	0.402	0.371	1.241	0.910	2.128
	$n=300$	0.026	0.009	0.018	0.196	0.274	0.357	1.243	0.797	2.145
	$n=500$	0.015	0.005	0.014	0.185	0.246	0.355	1.245	0.774	2.124
$m_i \overset{iid}{\sim} \{8,\ldots,12\}$ and SNR $=5$										
		μ	ϕ_1	ϕ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
NP (a)	$n=100$	0.087	0.009	0.011	0.160	0.135	0.123	7.173	0.568	10.430
	$n=300$	0.029	0.001	0.003	0.052	0.046	0.061	7.185	0.210	10.715
	$n=500$	0.017	0.001	0.003	0.034	0.030	0.046	7.176	0.134	10.490
REM (b)	$n=100$	0.109	0.048	0.049	0.335	0.356	0.151	0.591	0.603	1.454
	$n=300$	0.037	0.013	0.014	0.139	0.143	0.123	0.596	0.389	1.477
	$n=500$	0.022	0.007	0.009	0.107	0.103	0.119	0.596	0.342	1.447
Exp (c)	$n=100$	0.090	0.046	0.048	0.357	0.565	0.340	0.922	1.219	1.806
	$n=300$	0.030	0.012	0.014	0.207	0.358	0.309	0.927	0.995	1.787
	$n=500$	0.017	0.007	0.009	0.186	0.316	0.301	0.928	0.867	1.795
$m_i \overset{iid}{\sim} \{15,\ldots,20\}$ and SNR $=5$										
		μ	ϕ_1	ϕ_2	ξ_1	ξ_2	IN-IPE	IN-IPE$_{naive}$	OUT-IPE	OUT-IPE$_{naive}$
NP (a)	$n=100$	0.073	0.004	0.006	0.091	0.060	0.087	7.458	0.278	10.374
	$n=300$	0.025	0.001	0.003	0.030	0.022	0.039	7.455	0.105	10.574
	$n=500$	0.015	<0.001	0.002	0.019	0.015	0.029	7.452	0.072	10.197
REM (b)	$n=100$	0.093	0.048	0.049	0.303	0.338	0.143	0.550	0.440	1.451
	$n=300$	0.032	0.012	0.014	0.127	0.136	0.125	0.553	0.323	1.511
	$n=500$	0.019	0.007	0.009	0.097	0.099	0.120	0.554	0.298	1.442
Exp (c)	$n=100$	0.078	0.043	0.045	0.305	0.442	0.337	0.897	0.866	1.759
	$n=300$	0.026	0.012	0.014	0.200	0.277	0.325	0.900	0.763	1.777
	$n=500$	0.015	0.007	0.009	0.186	0.243	0.323	0.902	0.740	1.756

Table S2. Computational time (seconds) corresponding to Table 1

$m_i \overset{iid}{\sim} \{8, ..., 12\}$andSNR $= 1$			
	$n = 100$	$n = 300$	$n = 500$
NP(a)	7.369	15.892	21.418
REM (b)	9.282	11.347	22.559
Exp (c)	7.514	16.229	17.109
$m_i \overset{iid}{\sim} \{15, ..., 20\}$andSNR $= 1$			
NP (a)	9.257	24.141	32.292
REM (b)	11.230	17.061	32.519
Exp (c)	11.096	23.341	33.279
$m_i \overset{iid}{\sim} \{8, ..., 12\}$andSNR $= 5$			
NP (a)	7.405	16.162	21.507
REM (b)	9.609	13.283	21.166
Exp (c)	8.412	15.446	17.653
$m_i \overset{iid}{\sim} \{15, ..., 20\}$andSNR $= 5$			
NP (a)	9.303	24.321	31.964
REM (b)	10.477	17.446	35.025
Exp (c)	10.683	18.844	32.851

Table S3. Computational time (seconds) corresponding to Table S1

$m_i \overset{iid}{\sim} \{8, ..., 12\}$andSNR $= 1$			
	$n = 100$	$n = 300$	$n = 500$
NP (a)	13.760	12.655	42.346
REM (b)	16.776	24.044	39.088
Exp (c)	10.541	20.195	30.071
$m_i \overset{iid}{\sim} \{15, ..., 20\}$andSNR $= 1$			
NP (a)	9.043	58.091	60.064
REM (b)	28.550	35.029	53.055
Exp (c)	13.641	30.569	43.809
$m_i \overset{iid}{\sim} \{8, ..., 12\}$andSNR $= 5$			
NP (a)	8.222	27.563	50.990
REM (b)	11.318	17.683	24.464
Exp (c)	10.888	18.627	24.939
$m_i \overset{iid}{\sim} \{15, ..., 20\}$andSNR $= 5$			
NP (a)	10.038	51.146	71.719
REM (b)	13.810	24.568	36.225
Exp (c)	13.415	25.850	37.066

S4. Additional figures for the DTI data analysis

This section includes figures that are discussed in the DTI data analysis.

Figure S1. Fitted varying coefficient model, $\widehat{\mu}(s,T) = \widehat{\mu}_0(s) + \widehat{\beta}_T(s)T$ (left), and estimated slope function, $\widehat{\beta}_T(s)$ (right)

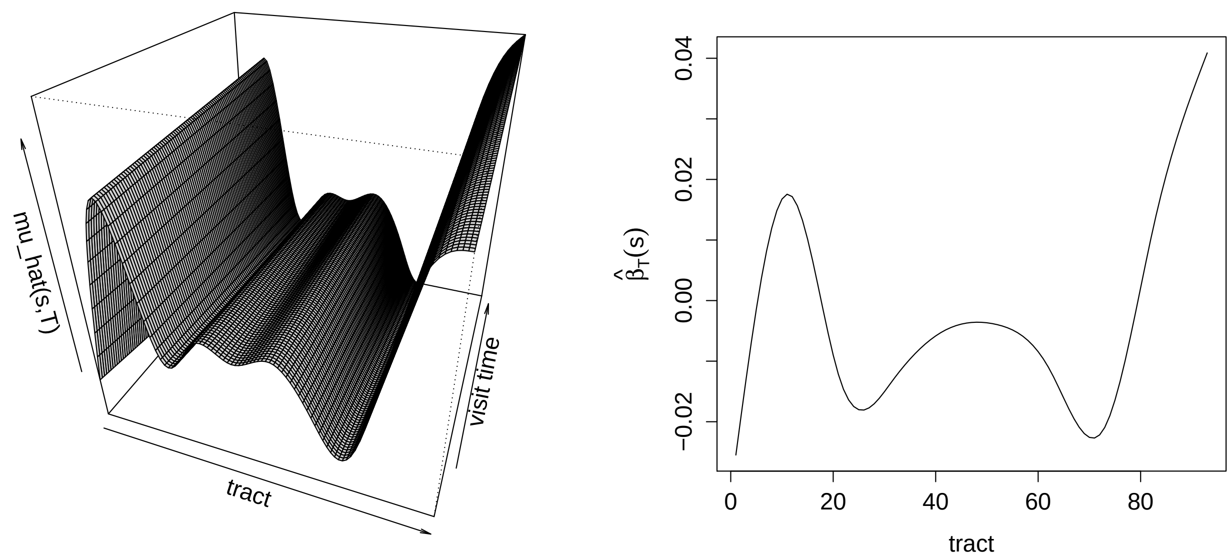


Figure S2. Scree plot of the marginal FPCA

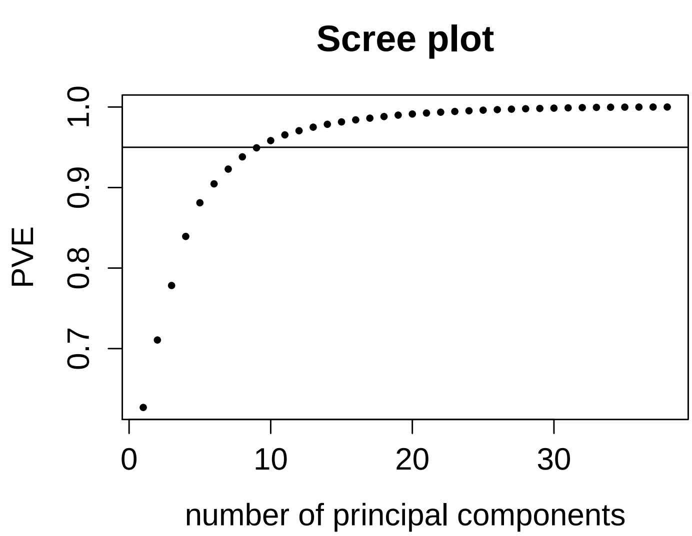


Figure S3. Estimated eigenfunctions of the marginal FPCA for FA with PVE = 95%

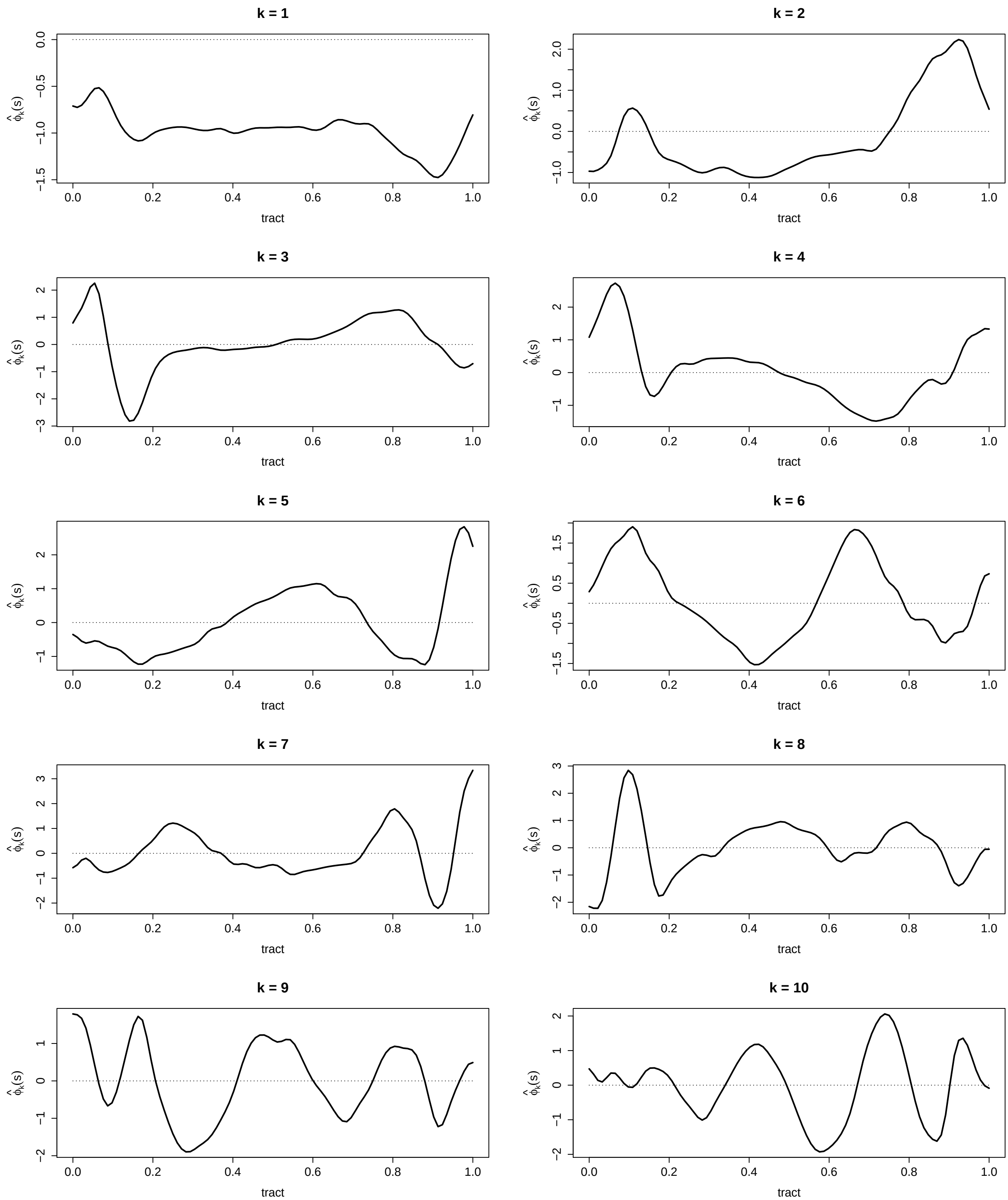
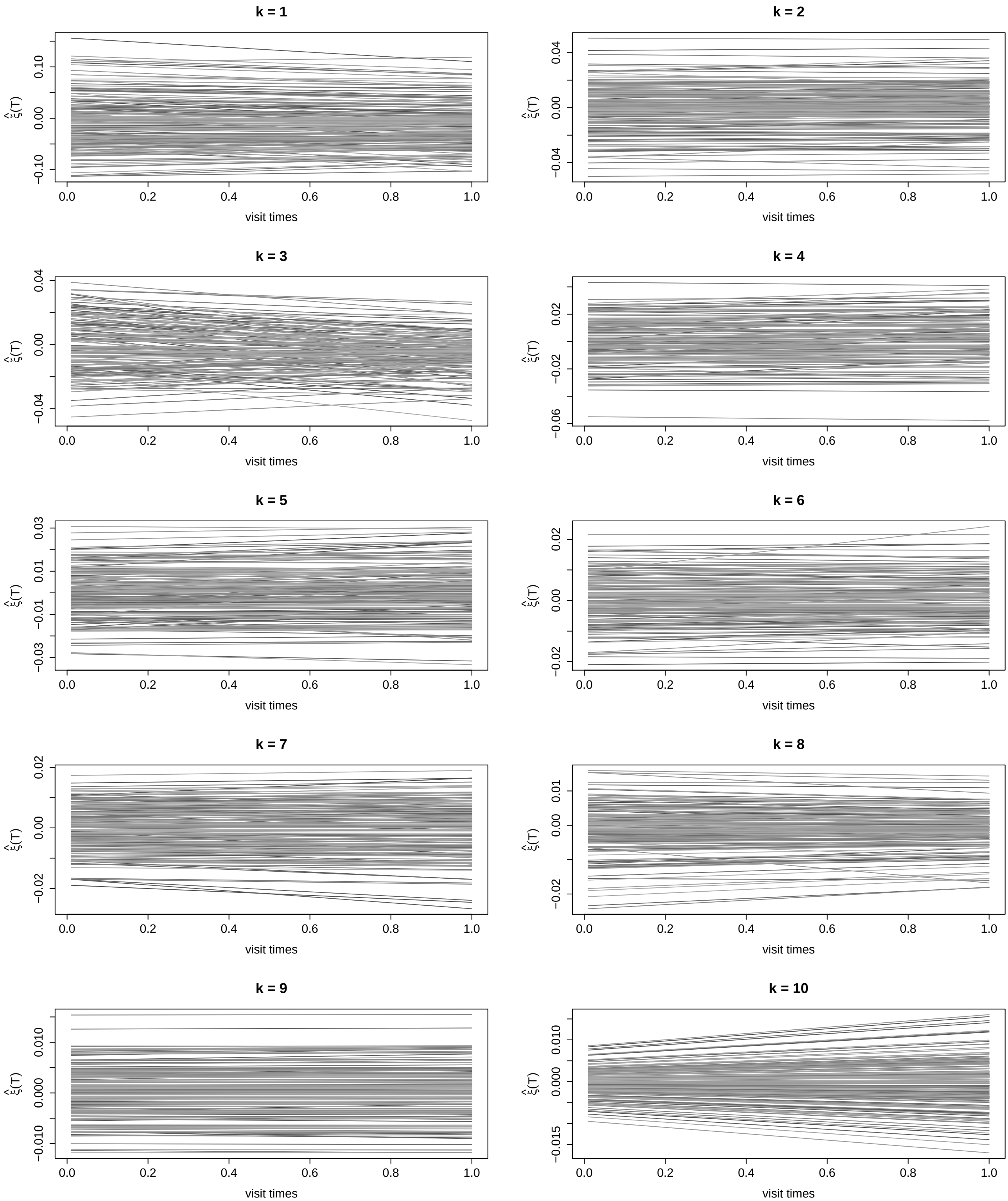


Figure S4. Estimated basis coefficient functions, $\hat{\xi}_{ik}(T)$ using a random coefficients linear model



References

Boos, DD & Stefanski, L (2013), *Essential Statistical Inference*, Springer.

Bosq, D (2000), *Linear processes in function spaces: theory and applications*, vol. 149, Springer.

Cardot, H, Ferraty, F, Mas, A & Sarda, P (2003), 'Testing hypotheses in the functional linear model,' *Scandinavian Journal of Statistics*, **30**(1), pp. 241–255.

Cardot, H, Goia, A & Sarda, P (2004), 'Testing for no effect in functional linear regression models, some computational approaches,' *Communications in Statistics-Simulation and Computation*, **33**(1), pp. 179–199.

Chen, K & Müller, HG (2012), 'Modeling repeated functional observations,' *Journal of the American Statistical Association*, **107**(500), pp. 1599–1609, doi:10.1080/01621459.2012.734196.

Ciprian Crainiceanu, Philip Reiss, Jeff Goldsmith, Lei Huang, Lan Huo & Fabian Scheipl (2014), *refund: Regression with Functional Data*, r package version 0.1-11.

Horváth, L & Kokoszka, P (2012), *Inference for functional data with applications*, vol. 200, Springer.

James, GM, Hastie, TJ & Sugar, CA (2000), 'Principal component models for sparse functional data,' *Biometrika*, **87**(3), pp. 587–602.

MATLAB (2014), *version 8.4.0 (R2014b)*, The MathWorks Inc., Natick, Massachusetts.

Peng, J & Paul, D (2009), 'A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data,' *Journal of Computational and Graphical Statistics*, **18**(4).

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Wood, SN (2011), 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models,' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(1), pp. 3–36.

Xiao, L, Li, Y & Ruppert, D (2013), 'Fast bivariate p-splines: the sandwich smoother,' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**(3), pp. 577–599.

Xiao, L, Ruppert, D, Zipunnikov, V & Crainiceanu, C (2015), 'Fast Covariance Estimation for High-dimensional Functional Data,' *Statistics and Computing*, to appear.

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

Zhang, JT & Chen, JW (2007), 'Statistical inferences for functional data,' *The Annals of Statistics*, **35**(3), pp. 1052–1079.