

Supplementary Material for:
**‘Likelihood Ratio Tests for the Mean Structure of Correlated
 Functional Processes’**

Ana-Maria Staicu, Yingxing Li, Ciprian M. Crainiceanu, and David
 Ruppert

This Web Supplement contains two sections. Section A.1 discusses the proof of Propositions 2.1, 3.1, 3.2, and 3.3. Section A.2 presents hypothesis testing for the structure of the mean difference in two independent sets of curves irrespective of their sampling design, as well as in two dependent sets of curves densely sampled and the proof of Proposition 5.1, in the setting of sparsely sampled curves.

A.1 One sample of functional data. Proofs

Proof of Proposition 2.1: Recall that the pseudo log-likelihood $\log \widehat{L}_{\widehat{\Psi}}(\boldsymbol{\beta}, \lambda)$ is defined in equation (3). Solving the first order condition for $\boldsymbol{\beta}$, we get the maximum pseudo profile likelihood estimation $\widehat{\boldsymbol{\beta}}(\lambda) = (\widehat{X}^T \widehat{H}_{\lambda}^{-1} \widehat{X})^{-1} \widehat{X}^T \widehat{H}_{\lambda}^{-1} \widehat{\mathbf{Y}}$. Let $\log \widehat{L}_{\widehat{\Psi}}(\lambda)$ be the pseudo profile likelihood when $\boldsymbol{\beta}$ is maximized out, i.e.,

$$2 \log \widehat{L}_{\widehat{\Psi}}(\lambda) = 2 \log \widehat{L}_{\widehat{\Psi}}\{\widehat{\boldsymbol{\beta}}(\lambda), \lambda\} = -[\log |\widehat{H}_{\lambda}| + \{\widehat{\mathbf{Y}} - \widehat{X}\widehat{\boldsymbol{\beta}}(\lambda)\}^T \widehat{H}_{\lambda}^{-1} \{\widehat{\mathbf{Y}} - \widehat{X}\widehat{\boldsymbol{\beta}}(\lambda)\}].$$

Let $\log \widehat{L}^{0,N}$ be the maximum pseudo log-likelihood under the null hypothesis (2). Then we can decompose $pLRT_N$ into two parts, i.e.,

$$pLRT_N = 2 \sup_{\lambda \geq 0} \{\log \widehat{L}_{\widehat{\Psi}}(\lambda) - \log \widehat{L}_{\widehat{\Psi}}(0)\} + 2\{\log \widehat{L}_{\widehat{\Psi}}(0) - \log \widehat{L}^{0,N}\}, \quad (\text{A.1})$$

where the first part corresponds to testing for $\lambda = 0$ and the second part corresponds to testing for the fixed effects $\beta_q = 0$ for $q \in Q$. In what follows we take each part at a time.

First part of (A.1). We first rewrite this part in a more convenient way and then prove that it converges weakly to $LRT_{\infty}(\lambda')$, which is defined by (4). Recall that $\widehat{H}_{\lambda} = I_N + \lambda \widehat{Z}\widehat{Z}^T$.

Define $\widehat{\xi}_{k,N}$ as the k th eigenvalues of $N^{-\varrho}\widehat{Z}^T\widehat{Z}$ for $k = 1, \dots, K$. Since $\widehat{Z}\widehat{Z}^T$ has the same nonzero eigenvalues as $\widehat{Z}^T\widehat{Z}$, we have, $\log|\widehat{H}_\lambda| = \sum_{k=1}^K \log(1 + \lambda N^\varrho \widehat{\xi}_{k,N})$.

According to Patterson and Thompson (1971), there exists an $N \times (N - p - 1)$ matrix \widehat{W} such that $\widehat{W}\widehat{W}^T = I_N - \widehat{X}(\widehat{X}^T\widehat{X})^{-1}\widehat{X}^T$, $\widehat{W}^T\widehat{W} = I_{N-p-1}$. We have $-2\log\widehat{L}_{\widehat{Y}}(0) = \widehat{Y}^T\widehat{W}\widehat{W}^T\widehat{Y}$. By an application of the Woodbury matrix identity (Woodbury, 1950); see also Harville (1997, p. 424) we can write

$$\begin{aligned} 2\log\widehat{L}_{\widehat{Y}}(\lambda) - 2\log\widehat{L}_{\widehat{Y}}(0) &= \lambda\widehat{Y}^T\widehat{W}\widehat{W}^T\widehat{Z}(I_K + \lambda\widehat{Z}^T\widehat{W}\widehat{W}^T\widehat{Z})^{-1}\widehat{Z}^T\widehat{W}\widehat{W}^T\widehat{Y} \\ &\quad - \sum_{k=1}^K \log(1 + \lambda N^{-\varrho}\widehat{\xi}_{k,N}). \end{aligned} \quad (\text{A.2})$$

Define $\widehat{\zeta}_{k,N}$ as the k th eigenvalue of $N^{-\varrho}\widehat{Z}^T\widehat{W}\widehat{W}^T\widehat{Z}$ and let $\widehat{U}_{\widehat{Z}\widehat{W}}$ be the $K \times K$ matrix whose k th column is the eigenvector associated with $\widehat{\zeta}_{k,N}$. Note that $\widehat{Z}^T\widehat{W}\widehat{W}^T\widehat{Z} = \widehat{U}_{\widehat{Z}\widehat{W}}\text{diag}(N^\varrho\widehat{\zeta}_{1,N}, \dots, N^\varrho\widehat{\zeta}_{K,N})\widehat{U}_{\widehat{Z}\widehat{W}}^T$ and thus

$$2\log\widehat{L}_{\widehat{Y}}(\lambda) - 2\log\widehat{L}_{\widehat{Y}}(0) = \sum_{k=1}^K \frac{\lambda N^\varrho \widehat{w}_{k,N}^2}{1 + \lambda N^\varrho \widehat{\zeta}_{k,N}} - \sum_{k=1}^K \log(1 + \lambda N^\varrho \widehat{\xi}_{k,N}), \quad (\text{A.3})$$

where $\widehat{w}_{k,N}$ is the k th component of the column vector $\widehat{\mathbf{w}}_N = N^{-\varrho/2}\widehat{U}_{\widehat{Z}\widehat{W}}^T\widehat{Z}^T\widehat{W}\widehat{W}^T\widehat{Y}$.

For simplicity of exposition we define

$$\widehat{f}_N(\lambda') = \sum_{k=1}^K \frac{\lambda' \widehat{w}_{k,N}^2}{1 + \lambda' \widehat{\zeta}_{k,N}} - \sum_{k=1}^K \log(1 + \lambda' \widehat{\xi}_{k,N}) \quad (\text{A.4})$$

and write $2\sup_{\lambda \geq 0} \{\log\widehat{L}_{\widehat{Y}}(\lambda) - \log\widehat{L}_{\widehat{Y}}(0)\} = \sup_{\lambda' \geq 0} \widehat{f}_N(\lambda')$. We will show that $\sup_{\lambda' \geq 0} \widehat{f}_N(\lambda') \Rightarrow \sup_{\lambda' \geq 0} LRT_\infty(\lambda')$ in two steps:

(S1) $\widehat{f}_N(\lambda')$ converges weakly to $LRT_\infty(\lambda')$ on the space of $C[0, M]$, for $M < \infty$;

(S2) a continuous mapping theorem type result holds for $\sup_{\lambda' \geq 0} \widehat{f}_N(\lambda')$.

We show (S1) in two parts: 1) first prove that $\widehat{f}_N(\lambda') \xrightarrow{D} LRT_\infty(\lambda')$ and 2) then show that $\widehat{f}_N(\lambda')$ is a tight sequence (Billingsley 1968, p54). The definition of $LRT_\infty(\lambda')$ is similarly to that of $\widehat{f}_N(\lambda')$ except that $\widehat{\zeta}_{k,N}$'s, $\widehat{\xi}_{k,N}$'s and $\widehat{w}_{k,N}$'s are replaced by ζ_k 's, ξ_k 's and w_k 's. For the first part, it is sufficient to prove that $\widehat{w}_{k,N} \xrightarrow{D} w_k$, $\widehat{\zeta}_{k,N} \xrightarrow{P} \zeta_k$, and $\widehat{\xi}_{k,N} \xrightarrow{P} \xi_k$ for all k ; Lemma A.1.1, discusses these results next. The weak convergence of $\widehat{f}_N(\lambda')$ to $LRT_\infty(\lambda')$, or in general, the finite dimensional convergence $\{\widehat{f}_N(\lambda'_1), \dots, \widehat{f}_N(\lambda'_L)\}$ to $\{LRT_\infty(\lambda'_1), \dots, LRT_\infty(\lambda'_L)\}$ follows by an application of the continuous mapping theorem.

LEMMA A.1.1. Let $\widehat{\xi}_{k,N}$, $\widehat{\zeta}_{k,N}$, and \widehat{w}_N be defined as above. Assume conditions (C2) and (C3) of Proposition 2.1 are true. Then:

- (a) For each $k = 1, \dots, K$, as $N \rightarrow \infty$ we have $\widehat{\xi}_{k,N} \xrightarrow{P} \xi_k$, $\widehat{\zeta}_{k,N} \xrightarrow{P} \zeta_k$.
(b) If in addition condition (C1) is true, then $\widehat{w}_N \xrightarrow{D} w$, where $\mathbf{w} = (w_1, \dots, w_K)$ is defined by Proposition 2.1.

Proof: To show the results of Lemma A.1.1 we need to introduce additional notation. Let $\widetilde{\mathbf{Y}}$, \widetilde{X} , \widetilde{Z} and \widetilde{W} be defined similarly to $\widehat{\mathbf{Y}}$, \widehat{X} , \widehat{Z} and \widehat{W} but with the $\widehat{\Sigma}$ replaced by Σ , and similarly define $\widetilde{\xi}_{k,N}$, $\widetilde{\zeta}_{k,N}$, and $\widetilde{\mathbf{w}}_N$ corresponding to $\widehat{\xi}_{k,N}$, $\widehat{\zeta}_{k,N}$, and $\widehat{\mathbf{w}}_N$; for example $\widetilde{\xi}_{k,N}$ and $\widetilde{\zeta}_{k,N}$ are exactly the quantities introduced by the Proposition 2.1 with the same notation.

(a) We will show that $\widehat{\xi}_k - \widetilde{\xi}_k = o_p(1)$ and $\widehat{\zeta}_k - \widetilde{\zeta}_k = o_p(1)$; then the result (a) follows by applying Slutsky's theorem and assuming the condition (C3). By using Theorem 8.1-6 of Golub and Van Loan (1983) it is sufficient to prove that

$$\|N^{-\varrho} \widehat{Z}^T \widehat{Z} - N^{-\varrho} \widetilde{Z}^T \widetilde{Z}\| = o_p(1) \text{ and} \quad (\text{A.5})$$

$$\|N^{-\varrho} \widehat{Z}^T \widehat{W} \widehat{W}^T \widehat{Z} - N^{-\varrho} \widetilde{Z}^T \widetilde{W} \widetilde{W}^T \widetilde{Z}\| = o_p(1), \quad (\text{A.6})$$

where $\|A\| = \sqrt{\sum_i \sum_j a_{ij}^2}$ is the Frobenius norm of some matrix $A = (a_{ij})_{i,j}$. These results follow from applications of norm inequalities as well as continuous mapping theorem.

(b) Next, we prove the convergence in distribution of $\widehat{\mathbf{w}}_N$. The idea is first to show that $\widetilde{\mathbf{w}}_N \xrightarrow{D} \mathbf{w}$ under the null hypothesis and then to show that $\|\widehat{\mathbf{w}}_N - \widetilde{\mathbf{w}}_N\| = o_p(1)$.

Under the null hypothesis, we have $\widetilde{\mathbf{Y}} = \widetilde{X}\boldsymbol{\beta} + \widetilde{\mathbf{e}}$ and thus $\widetilde{\mathbf{w}}_N = N^{-\varrho/2} \widetilde{U}_{\widetilde{Z}\widetilde{W}}^T \widetilde{Z}^T \widetilde{W} \widetilde{W}^T \widetilde{\mathbf{e}}$, since $\widetilde{W}^T \widetilde{\mathbf{Y}} = \widetilde{W}^T \widetilde{\mathbf{e}}$. Because $\widetilde{\mathbf{e}} = \Sigma^{-1/2} \mathbf{e}$ it follows that $\widetilde{\mathbf{e}} \sim N(\mathbf{0}_{N \times 1}, I_N)$ and thus $\widetilde{\mathbf{w}}_N$ has mean-zero multivariate normal distribution, since it is a linear combination of independent normal variables. The result, $\widetilde{\mathbf{w}}_N \xrightarrow{D} \mathbf{w}$, is concluded by assuming condition (C3), using Cramér-Wold device and an application of the (Lévy's) continuity theorem.

We prove next that $\|\widehat{\mathbf{w}}_N - \widetilde{\mathbf{w}}_N\| = o_p(1)$. Recall that $\widehat{\mathbf{w}}_N - \widetilde{\mathbf{w}}_N = N^{-\varrho/2} \widehat{U}_{\widehat{Z}\widehat{W}}^T \widehat{Z}^T \widehat{W} \widehat{W}^T \widehat{\mathbf{e}} - N^{-\varrho/2} \widetilde{U}_{\widetilde{Z}\widetilde{W}}^T \widetilde{Z}^T \widetilde{W} \widetilde{W}^T \widetilde{\mathbf{e}}$ and thus we have:

$$\begin{aligned} \|\widehat{\mathbf{w}}_N - \widetilde{\mathbf{w}}_N\| &\leq \|\widehat{U}_{\widehat{Z}\widehat{W}}^T - \widetilde{U}_{\widetilde{Z}\widetilde{W}}^T\| \|N^{-\varrho/2} \widehat{Z}^T \widehat{W} \widehat{W}^T \widehat{\mathbf{e}}\| \\ &\quad + \|\widetilde{U}_{\widetilde{Z}\widetilde{W}}^T\| \|N^{-\varrho/2} (\widehat{Z}^T \widehat{W} \widehat{W}^T \widehat{\mathbf{e}} - \widetilde{Z}^T \widetilde{W} \widetilde{W}^T \widetilde{\mathbf{e}})\| \end{aligned} \quad (\text{A.7})$$

Using norm and matrix manipulation one can show that $\|\widetilde{U}_{\widetilde{Z}\widetilde{W}}\| = O_p(1)$, $\|\widehat{U}_{\widehat{Z}\widehat{W}} - \widetilde{U}_{\widetilde{Z}\widetilde{W}}\| = o_p(1)$, $\|N^{-\varrho/2} \widehat{Z}^T \widehat{W} \widehat{W}^T \widehat{\mathbf{e}}\| = O_p(1)$, and $\|N^{-\varrho/2} \widehat{Z}^T \widehat{W} \widehat{W}^T \widehat{\mathbf{e}} - N^{-\varrho/2} \widetilde{Z}^T \widetilde{W} \widetilde{W}^T \widetilde{\mathbf{e}}\| =$

$o_p(1)$. This concludes the proof of Lemma A.1.1. #

Next, we prove the second part of (S1). Using Theorem 8.3 of Billingsley (1968), it suffices to show that for every ε' and $\eta' > 0$, there exists $\delta_0 > 0$ and N_0 such that for $N \geq N_0$,

$$\frac{1}{\delta_0} P\left\{ \sup_{t \leq t' \leq t + \delta_0} |\widehat{f}_N(t') - \widehat{f}_N(t)| \geq \varepsilon' \right\} \leq \eta'. \quad (\text{A.8})$$

It is noteworthy to point out that for every $\delta > 0$, and every $0 \leq t \leq t' \leq t + \delta$ we have

$$|\widehat{f}_N(t) - \widehat{f}_N(t')| \leq \sum_{k=1}^K |t - t'| \widehat{w}_{k,N}^2 + \sum_{k=1}^K \log \left\{ 1 + \frac{(t' - t) \widehat{\xi}_{k,N}}{1 + t \widehat{\xi}_{k,N}} \right\} \leq \sum_{k=1}^K \delta \widehat{w}_{k,N}^2 + \sum_{k=1}^K \delta \widehat{\xi}_{k,N},$$

since $\widehat{\xi}_{k,N}$'s, $\widehat{\zeta}_{k,N}$'s are nonnegative, it holds true $\log(1 + x) < x$ for $x > 0$.

Let ε' and η' be arbitrary but fixed positive values. Then for every $\delta > 0$ we have

$$P \left\{ \sup_{t \leq t' \leq t + \delta} |f_N(t') - f_N(t)| \geq \varepsilon' \right\} \leq \sum_{k=1}^K P\{\widehat{w}_{k,N}^2 \geq \varepsilon'/(2K\delta)\} + \sum_{k=1}^K P\{\widehat{\xi}_{k,N} \geq \varepsilon'/(2K\delta)\} \quad (\text{A.9})$$

which follows from an application of the Bonferroni inequality, $P(\sum_{i=1}^{2K} A_i \geq a) \leq \sum_{i=1}^{2K} P\{A_i \geq a/(2K)\}$, along with the observation that if $\sum_{i=1}^{2K} A_i \geq a$ holds, for variables $\{A_i : i = 1, \dots, 2K\}$ then we must have that $A_i \geq a/(2K)$, for some i . It is sufficient to show that there exists $\delta_0 = \delta_0(\varepsilon', \eta') > 0$ and $N_0 = N_0(\varepsilon', \eta') \geq 1$ such that the the right hand expression of the above inequality, with δ replaced by δ_0 , is bounded by $\delta_0 \eta'$ for all $N \geq N_0$.

For the first term of (A.9) let $F_{k,N}(t)$ and $F_k(t)$ be the cumulative distribution functions of $\widehat{w}_{k,N}$ and w_k respectively, for $k = 1, \dots, K$. Because $\widehat{\mathbf{w}}_{k,N} \Rightarrow \mathbf{w}_k$, it follows that $|F_{k,N}(t) - F_k(t)| \rightarrow 0$ for all t . Then it is not hard to show that for every $\delta < \widetilde{\delta}^0$ there is $N_{*\delta} > 0$ such that $\sum_{k=1}^K P\{\widehat{w}_{k,N}^2 \geq \varepsilon'/(2K\delta)\} < \delta \eta'/2$ for all $N > N_{*\delta}$.

Consider now the second sum of (A.9). For every summand k we have:

$$P\{\widehat{\xi}_{k,N} \geq \varepsilon'/(2K\delta)\} \leq P\{\widehat{\xi}_{k,N} > \varepsilon'/(2K\delta), |\widehat{\xi}_{k,N} - \xi_k| \leq \varepsilon'\} + P(|\widehat{\xi}_{k,N} - \xi_k| > \varepsilon')$$

which is less or equal than $P(|\widehat{\xi}_{k,N} - \xi_k| > \varepsilon')$, for any $\delta < \varepsilon' / \{2K(\xi_k + \varepsilon')\}$; since for this choice $|\widehat{\xi}_{k,N} - \xi_k| > \varepsilon'$ and the first term equals zero. Because $\widehat{\xi}_{k,N} \Rightarrow \xi_k$ and ξ_k is constant, we have that $\widehat{\xi}_{k,N} \rightarrow \xi_k$ in probability. It follows that for every $\delta < \varepsilon' / [2K\{\varepsilon' + \max_k(\xi_k)\}]$ we have that $\sum_{k=1}^K P\{\widehat{\xi}_{k,N} \geq \varepsilon'/(2K\delta)\} \leq \delta \eta'/2$ for all $N > N_{\delta}^{**}$.

Combining the two findings, one can find suitable δ_0 and N_0 so that expression (A.9) holds. This concludes the tightness proof of the sequence $\widehat{f}_N(\lambda')$.

To show (S2) we apply the continuous mapping theorem as in Crainiceanu (2003) and use weak convergence result of (S1) for all $M < \infty$; to save space we omit the details.

Second part of (A.1). Next we will show that there exists independent standard normal variables $\nu_1, \dots, \nu_{\#Q}$ such that

$$2 \log \widehat{L}_{\widehat{\mathbf{Y}}}(0) - 2 \log \widehat{L}^{0,N} \Rightarrow \sum_{i=1}^{\#Q} \nu_i^2, \quad (\text{A.10})$$

where $\#Q$ is the cardinality of the set Q in the null hypothesis (2). We discuss the case when $\#Q > 0$; if $\#Q = 0$ then equation (A.10) is trivial.

Before we simplify the left hand side of equation (A.10), we introduce the following definition. Partition $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^T | \boldsymbol{\beta}_{(2)}^T)^T$, where $\boldsymbol{\beta}_{(2)}$ contains all β_q for $q \in Q$. Similarly, partition $X = (X_{(1)} | X_{(2)})$ according to the partition of $\boldsymbol{\beta}$. We define $\widehat{X}_{(i)} = \widehat{\Sigma}^{-1/2} X_{(i)}$. For any matrix A with linearly independent columns, denote by $S_A = A(A^T A)^{-1} A^T$ the projection matrix onto the space spanned by the columns of A . In the special case when $\#Q = p + 1$, $X_{(2)} = X$ and $X_{(1)}$ does not exist, we use the convention that $S_{\widehat{X}_{(1)}} = \mathbf{0}_{N \times N}$.

Under the null hypothesis we have that $\widehat{\mathbf{Y}} = \widehat{X}_{(1)} \boldsymbol{\beta}_{(1)} + \widehat{\mathbf{e}}$. Then $2 \log \widehat{L}^{0,N} = -\widehat{\mathbf{Y}}^T (I_N - S_{\widehat{X}_{(1)}}) \widehat{\mathbf{Y}} = -\widehat{\mathbf{e}}^T (I_N - S_{\widehat{X}_{(1)}}) \widehat{\mathbf{e}}$, and $2 \log \widehat{L}_{\widehat{\mathbf{Y}}}(0) = -\widehat{\mathbf{e}}^T (I_N - S_{\widehat{X}}) \widehat{\mathbf{e}}$. It follows $2 \log \widehat{L}_{\widehat{\mathbf{Y}}}(0) - 2 \log \widehat{L}^{0,N} = \widehat{\mathbf{e}}^T (S_{\widehat{X}} - S_{\widehat{X}_{(1)}}) \widehat{\mathbf{e}}$, where $S_{\widehat{X}} - S_{\widehat{X}_{(1)}}$ is a projection matrix. There exists a $N \times \#Q$ matrix such that $\widehat{W}_0 \widehat{W}_0^T = S_{\widehat{X}} - S_{\widehat{X}_{(1)}}$ and $\widehat{W}_0^T \widehat{W}_0 = I_{\#Q}$. Denote $\widehat{\boldsymbol{\omega}} = \widehat{W}_0^T \widehat{\mathbf{e}}$. Applying the same arguments as for the Lemma A.1.1, part (b) we can conclude that $\widehat{\boldsymbol{\omega}}_i$'s are asymptotically standard normal independent variables assuming (C1), that the null hypothesis is true. It implies that equation (A.10) holds, and in turn that (4) holds.

The proof is now concluded, as independence between $\sup LRT_{\infty}(\lambda)$ and $\sum_{i=1}^{\#Q} \nu_i^2$ can be established using the same techniques as in the proof of Theorem 3 of Crainiceanu (2003). Hence Proposition 2.1 holds. #

Proof of Proposition 3.1: First the following regularity conditions are imposed:

- (A1) The true parameter of the correlation structure $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ lies in the interior of a compact set.
- (A2) For any $\boldsymbol{\theta} \in \Theta$, the functions $\partial \varphi(t, t'; \boldsymbol{\theta}) / \partial \theta_{\ell}$, and $\partial^2 \varphi(t, t'; \boldsymbol{\theta}) / \partial \theta_{\ell} \partial \theta'_{\ell}$ are bounded bivariate functions of t, t' .

(A3) For any $\boldsymbol{\theta} \in \Theta$, the eigenvalues of the correlation matrix $C_i(\boldsymbol{\theta})$ of a generic subject i are between $0 < \varrho_0 < \varrho_1 < \infty$.

Let $C(\boldsymbol{\theta}) = \text{diag}\{C_1(\boldsymbol{\theta}), \dots, C_n(\boldsymbol{\theta})\}$ be the true correlation matrix, and denote by $C(\widehat{\boldsymbol{\theta}})$ its estimate. Condition (C2) entails two parts: (a) $\widehat{\sigma}_e^{-2} \mathbf{a}^T C(\widehat{\boldsymbol{\theta}})^{-1} \mathbf{a} - \sigma_e^{-2} \mathbf{a}^T C(\boldsymbol{\theta})^{-1} \mathbf{a} = o_p(1)$, and (b) $\widehat{\sigma}_e^{-2} \mathbf{a}^T C(\widehat{\boldsymbol{\theta}})^{-1} \mathbf{e} - \sigma_e^{-2} \mathbf{a}^T C(\boldsymbol{\theta})^{-1} \mathbf{e} = o_p(1)$ where $\mathbf{a} \in R^N$, $\|\mathbf{a}\| = 1$, \mathbf{e} is the N -dimensional vector of \mathbf{e}_i and $N = \sum_{i=1}^n m_i$.

To prove part (a) it is sufficient to show that: (a1) $\mathbf{a}^T C^{-1}(\boldsymbol{\theta}) \mathbf{a} = O(1)$ and (a2) $\mathbf{a}^T \{C^{-1}(\boldsymbol{\theta}) - C^{-1}(\widehat{\boldsymbol{\theta}})\} \mathbf{a} = o_p(1)$. The result then follows by using $\widehat{\sigma}_e^2 - \sigma_e^2 = o_p(1)$, the observation $\widehat{\sigma}_e^{-2} = O_p(1)$ and an application of the triangle inequality. Showing (a1) is easy by using the regularity condition (A3). Consider now (a2). Let $\varphi_{\boldsymbol{\theta}}(t, t'; \boldsymbol{\theta}) = \partial \varphi(t, t'; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, for some time points t, t' and let $\partial C_i(\boldsymbol{\theta}) / \partial \theta_l$ be the matrix with components $\varphi_{\theta_l}(t_{ij}, t_{ij'}; \boldsymbol{\theta})$ for $j, j' = 1, \dots, m_i$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$. Fix $l = 1, \dots, d$, and denote by $D_{i,l}(\boldsymbol{\theta}) = C_i^{-1}(\boldsymbol{\theta}) \{\partial C_i(\boldsymbol{\theta}) / \partial \theta_l\} C_i^{-1}(\boldsymbol{\theta})$, which is $\partial C_i^{-1}(\boldsymbol{\theta}) / \partial \theta_l$. Using the regularity conditions (A1)-(A3) is easy to see that $\|C_i^{-1}(\widehat{\boldsymbol{\theta}}) - C_i^{-1}(\boldsymbol{\theta})\| = O_p(n^{-1/2})$, where $\|A\|$ is the Frobenius norm of matrix A , defined above. Moreover one can show that $\|C_i^{-1}(\widehat{\boldsymbol{\theta}}) - C_i^{-1}(\boldsymbol{\theta})\| \leq \sum_{l=1}^d |\widehat{\theta}_l - \theta_l| M_l$, for some M_l such that $\sup_{\boldsymbol{\theta}} \|D_{i,l}(\boldsymbol{\theta})\| < M_l$ for all $i = 1, \dots, n$ and $l = 1, \dots, d$. It follows that $\|C^{-1}(\boldsymbol{\theta}) - C^{-1}(\widehat{\boldsymbol{\theta}})\|_2 = \max_i \|C_i^{-1}(\boldsymbol{\theta}) - C_i^{-1}(\widehat{\boldsymbol{\theta}})\|_2 = o_p(1)$.

To prove part (b) it is sufficient to show that: (b1) $\mathbf{a}^T C^{-1}(\boldsymbol{\theta}) \mathbf{e} = O_p(1)$ and (b2) $\mathbf{a}^T \{C^{-1}(\boldsymbol{\theta}) - C^{-1}(\widehat{\boldsymbol{\theta}})\} \mathbf{e} = o_p(1)$. To show (b1) it suffices to show that $\text{var}\{\mathbf{a}^T C^{-1}(\boldsymbol{\theta}) \mathbf{e}\} = O(1)$. This is easy to check since $\|C_i^{-1}(\boldsymbol{\theta})\|_2 < \infty$ for all i and $\|\mathbf{a}\| = 1$. Consider now part (b2). Let $f_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{a}_i^T C_i^{-1}(\boldsymbol{\theta}) \mathbf{e}_i$. Using the first order Taylor expansion of the function $f_{\mathbf{a},\mathbf{e}}(\widehat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$ we obtain;

$$f_{\mathbf{a},\mathbf{e}}(\widehat{\boldsymbol{\theta}}) = f_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) + (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T f'_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) + o_p(1), \quad (\text{A.11})$$

where $f'_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) = \partial f_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. We show that the vector $n^{-1/2} f'_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) = o_p(1)$, by proving that each of its components, $n^{-1/2} \partial f_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) / \partial \theta_l = o_p(1)$, is $o_p(1)$. Using condition (A3) we have

$$\text{var}\left\{\sum_{i=1}^n \mathbf{a}_i^T D_{i,l}(\boldsymbol{\theta}) \mathbf{e}_i\right\} = \sum_{i=1}^n \sigma_e^2 \mathbf{a}_i^T D_{i,l}(\boldsymbol{\theta}) C_i(\boldsymbol{\theta}) D_{i,l}^T(\boldsymbol{\theta}) \mathbf{a}_i$$

which is finite, since $m_i < \infty$ and $\|C_i(\boldsymbol{\theta})\| < \infty$. The result (b2) follows easily since $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T f'_{\mathbf{a},\mathbf{e}}(\boldsymbol{\theta}) = o_p(1)$, by employing the assumption (L3) of the proposition. $\#$

Proof of Proposition 3.2: For illustration simplicity we assume that m satisfies assumption (F2) and thus $\tilde{m} = m$, and $\tilde{t}_l = t_l$. Under the dense design, \mathbf{X}_i and \mathbf{Z}_i do not depend on

i , and $\boldsymbol{\Sigma} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}_0$, where the (j, j') th element of $\boldsymbol{\Sigma}_0$ is $\Gamma(t_j, t_{j'}) + \sigma_\epsilon^2 \mathbf{1}(t_j = t_{j'})$. It suffices to show that (C2') holds for any m -dimensional vector, \mathbf{a} , $\|\mathbf{a}\| = 1$, and $\mathbf{e}_0 = n^{-1/2} \sum_{i=1}^n \mathbf{e}_i$, i.e.

$$\|\widehat{\boldsymbol{\Sigma}}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}\|_2 = o_p(1) \quad (\text{A.12})$$

$$\mathbf{a}^T \widehat{\boldsymbol{\Sigma}}_0^{-1} \mathbf{e}_0 - \mathbf{a}^T \boldsymbol{\Sigma}_0^{-1} \mathbf{e}_0 = o_p(1). \quad (\text{A.13})$$

Consider equation (A.12). Let Γ^0 be the $m \times m$ covariance matrix obtained from the covariance function $\Gamma(t, t')$ evaluated over the set of observed points $\{t_1, \dots, t_m\}^2$. Furthermore denote by $\mathbf{G} = (g_{jk})_{1 \leq j \leq m, 1 \leq k \leq M}$ the $m \times M$ matrix of eigenvectors and by $\boldsymbol{\Lambda} = \text{diag}\{\sigma_1^2, \dots, \sigma_M^2\}$ the $M \times M$ diagonal matrix of associated eigenvalues corresponding to Γ^0 . It follows that $g_{jk} \approx \theta_k(t_j)/\sqrt{m}$ for $k = 1, \dots, M$ and $j = 1, \dots, m$ since $\sum_{j=1}^m \theta_k^2(t_j)/m$ is the Riemann approximation of the unit-valued integral $\int_{\mathcal{T}} \theta_k(t)^2 dt = 1$. In the new notation we write $\boldsymbol{\Sigma}_0 = \sigma_\epsilon^2 \mathbf{I}_m + m \mathbf{G} \boldsymbol{\Lambda} \mathbf{G}^T$ and by using the Woodbury matrix identity it follows that $\boldsymbol{\Sigma}_0^{-1} = \sigma_\epsilon^{-2} \mathbf{I}_m - \sigma_\epsilon^{-2} \mathbf{G} \text{diag}\{\sigma_k^2/(\sigma_k^2 + \sigma_\epsilon^2/m)\}_k \mathbf{G}^T$.

In a similar way, define $\widehat{\mathbf{G}}$ and $\widehat{\boldsymbol{\Lambda}}$ the estimated quantities corresponding to \mathbf{G} and $\boldsymbol{\Lambda}$, respectively, for $k = 1, \dots, M$ and $\widehat{\boldsymbol{\Sigma}}_0 = \widehat{\sigma}_\epsilon^2 \mathbf{I}_m + m \widehat{\mathbf{G}} \widehat{\boldsymbol{\Lambda}} \widehat{\mathbf{G}}^T$. Following the assumption $\|\widehat{\theta}_k - \theta_k\| = O_p(n^{-\alpha})$ we have $\|\widehat{\mathbf{g}}_k - \mathbf{g}_k\| = \|\widehat{\theta}_k - \theta_k\| = O_p(n^{-\alpha})$, where $\widehat{\mathbf{g}}_k$ and \mathbf{g}_k are the k th columns of $\widehat{\mathbf{G}}$ and \mathbf{G} respectively.

We show (A.12) by using the Woodbury matrix identity for $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ and $\boldsymbol{\Sigma}_0^{-1}$ and triangle inequality:

$$\begin{aligned} \|\widehat{\boldsymbol{\Sigma}}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}\|_2 &\leq \left\| \widehat{\sigma}_\epsilon^{-2} \widehat{\mathbf{G}} \text{diag}\left\{\frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m}\right\}_k \widehat{\mathbf{G}}^T - \sigma_\epsilon^{-2} \mathbf{G} \text{diag}\left\{\frac{\sigma_k^2}{\sigma_k^2 + \sigma_\epsilon^2/m}\right\}_k \mathbf{G}^T \right\|_2 \\ &\quad + |\widehat{\sigma}_\epsilon^{-2} - \sigma_\epsilon^{-2}|, \end{aligned}$$

since (a) $\widehat{\sigma}_\epsilon^{-2} - \sigma_\epsilon^{-2} = o_p(1)$, (b) $\|\widehat{\mathbf{G}} \text{diag}\{\widehat{\sigma}_k^2/(\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m)\}_k \widehat{\mathbf{G}}^T\|_2 = \max_k \{\widehat{\sigma}_k^2/(\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m)\}_k = \widehat{\sigma}_1^2/(\widehat{\sigma}_1^2 + \widehat{\sigma}_\epsilon^2/m) = O_p(1)$, (c) $\|\widehat{\mathbf{G}} - \mathbf{G}\| = O_p(n^{-\alpha})$, (d) $\|\text{diag}\{\widehat{\sigma}_k^2/(\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m) - \sigma_k^2/(\sigma_k^2 + \sigma_\epsilon^2/m)\}_k\| = O_p(n^{-\alpha} m^{-1})$, (e) $\sum_{k=1}^M \sigma_k^2 < \infty$ as well as the unitary invariance of 2-norms and the norm relationship $\|\cdot\|_2 \leq \|\cdot\|$.

We turn now to (A.13). Again we use Woodbury matrix identity and reduce (A.13) to

$$\mathbf{a}^T \left[\widehat{\mathbf{G}} \text{diag}\left\{\frac{\widehat{\sigma}_k^2}{\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m}\right\}_k \widehat{\mathbf{G}}^T - \mathbf{G} \text{diag}\left\{\frac{\sigma_k^2}{\sigma_k^2 + \sigma_\epsilon^2/m}\right\}_k \mathbf{G}^T \right] \mathbf{e} = o_p(1),$$

under the assumption that $\widehat{\sigma}_\epsilon^{-2} - \sigma_\epsilon^{-2} = o_p(n^{-\alpha})$, since $\mathbf{a}^T \mathbf{e} = O_p(m^{1/2})$. Moreover, if the

number of eigenvalues M is finite, and $(1 - \sigma_k^{-2}\sigma_\epsilon^2/m) = 1 + O(m^{-1})$ it suffices to show that

$$\{\widehat{\sigma}_k^2/(\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m) - \sigma_k^2/(\sigma_k^2 + \sigma_\epsilon^2/m)\}a^T \widehat{\mathbf{g}}_k \widehat{\mathbf{g}}_k^T \mathbf{e} = o_p(1) \quad (\text{A.14})$$

$$a^T (\widehat{\mathbf{g}}_k - \mathbf{g}_k) \mathbf{g}_k^T \mathbf{e} = O_p(n^{-\alpha} m^{1/2}) \quad (\text{A.15})$$

$$a^T \mathbf{g}_k (\widehat{\mathbf{g}}_k - \mathbf{g}_k)^T \mathbf{e} = O_p(n^{-\alpha} m^{1/2}). \quad (\text{A.16})$$

Relation (A.14) follows from application of Chebyshev's inequality and the following facts: $\widehat{\sigma}_k^2/(\widehat{\sigma}_k^2 + \widehat{\sigma}_\epsilon^2/m) - \sigma_k^2/(\sigma_k^2 + \sigma_\epsilon^2/m) = O_p(n^{-\alpha} m^{-1})$, $\|\widehat{\mathbf{g}}_k\| = 1$, $\|a\| = 1$ and $\|\mathbf{e}\| = O_p(m^{1/2})$.

Relation (A.15) is obvious since $\mathbf{g}_k^T \mathbf{e} = O_p(m^{1/2})$ and $|a^T (\widehat{\mathbf{g}}_k - \mathbf{g}_k)| = O_p(n^{-\alpha})$. To show (A.16) it is sufficient to show $(\widehat{\mathbf{g}}_k - \mathbf{g}_k)^T \mathbf{e} = O_p(n^{-\alpha} m^{-1/2})$ which follows from an application of Chebyshev's inequality.

The result now follows using the assumption that $O_p(n^{-\alpha} m^{1/2}) = o_p(1)$. This concludes our proof. #

Proof of Proposition 3.3: Under the sparse design, $\boldsymbol{\Sigma}$ is a block diagonal matrix, with the i th block equal to the $m_i \times m_i$ -dimensional matrix $\boldsymbol{\Sigma}_i = \sigma_\epsilon^2(\mathbf{I}_{m_i} + \mathbf{K}_i)$ where $\mathbf{K}_i = \mathbf{G}_i \boldsymbol{\Lambda} \mathbf{G}_i^T$, \mathbf{G}_i is $m_i \times M$ dimensional matrix with the (l, k) th element equal to $\theta_k(t_{il})$, and $\boldsymbol{\Lambda}$ is $M \times M$ block diagonal matrix whose (k, k) th component is $\sigma_k^2/\sigma_\epsilon^2$. Note that \mathbf{G}_i and $\boldsymbol{\Lambda}$ are defined differently from the corresponding ones defined in the proof of Proposition 3.2. Similarly, we can define $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\sigma}_\epsilon^2(\mathbf{I}_{m_i} + \widehat{\mathbf{K}}_i)$, where $\widehat{\mathbf{K}}_i = \widehat{\mathbf{G}}_i \widehat{\boldsymbol{\Lambda}} \widehat{\mathbf{G}}_i^T$. Partition \mathbf{a} and \mathbf{e} in condition (C2) into n vectors, \mathbf{a}_i and \mathbf{e}_i of length m_i . To prove condition (C2), it suffices to show

$$\|\text{diag}\{\widehat{\boldsymbol{\Sigma}}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}\}_i\|_2 = o_p(1), \quad (\text{A.17})$$

$$\sum_{i=1}^n \mathbf{a}_i^T \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{e}_i - \sum_{i=1}^n \mathbf{a}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{e}_i = o_p(1). \quad (\text{A.18})$$

Consider first equation (A.17). We make use of the well known inequalities: $\|\text{diag}\{A_i\}_i\|_2 \leq \max_i \|A_i\|_2$, and $\|A\|_2^2 \leq \|A\|_1 \|A\|_\infty$ for matrices $A = (a_{jj'})_{j,j'}$, A_i 's, where $\|A\|_1 = \max_{j'} \sum_j |a_{jj'}|$ and $\|A\|_\infty = \max_j \sum_{j'} |a_{jj'}|$ are the 1 and ∞ matrix norm induced. The result follows by noting that

$$\|\widehat{\boldsymbol{\Sigma}}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}\|_2 \leq |\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| + \sigma_\epsilon^2 \|(\mathbf{I}_{m_i} + \widehat{\mathbf{K}}_i)^{-1}\|_2 \|\widehat{\mathbf{K}}_i - \mathbf{K}_i\|_2 \|(\mathbf{I}_{m_i} + \mathbf{K}_i)^{-1}\|_2$$

and that $\max_i \|\widehat{\mathbf{K}}_i - \mathbf{K}_i\|_\infty = O_p(n^{-\alpha})$, $\max_i \|\widehat{\mathbf{K}}_i - \mathbf{K}_i\|_1 = O_p(n^{-\alpha})$. In addition $\|(\mathbf{I}_{m_i} + \widehat{\mathbf{K}}_i)^{-1}\| < 1$ and $\|(\mathbf{I}_{m_i} + \mathbf{K}_i)^{-1}\| < 1$ because both \mathbf{K}_i and $\widehat{\mathbf{K}}_i$ are positive definite matrices, and thus have positive eigenvalues.

Next we prove the validity of (A.18). Using Woodbury matrix identity, we rewrite each term of the left hand side of (A.18) as

$$\begin{aligned}\sum_{i=1}^n \mathbf{a}_i^T \widehat{\Sigma}_i^{-1} \mathbf{e}_i &= \widehat{\sigma}_\epsilon^2 \sum_{i=1}^n \mathbf{a}_i^T \mathbf{e}_i - \widehat{\sigma}_\epsilon^2 \sum_{i=1}^n \mathbf{a}_i^T \widehat{\mathbf{G}}_i (\widehat{\Lambda}^{-1} + \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i)^{-1} \widehat{\mathbf{G}}_i^T \mathbf{e}_i; \\ \sum_{i=1}^n \mathbf{a}_i^T \Sigma_i^{-1} \mathbf{e}_i &= \sigma_\epsilon^2 \sum_{i=1}^n \mathbf{a}_i^T \mathbf{e}_i - \sigma_\epsilon^2 \sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i,\end{aligned}$$

where \mathbf{e}_i are independent m_i -dimensional vectors with distribution $N(0, \Sigma_i)$. By an application of the continuity theorem for $S_n = \sum_{i=1}^n \mathbf{a}_i^T \mathbf{e}_i = \sum_{i=1}^n |c_i| \epsilon_i$, where $c_i^2 = \mathbf{a}_i^T \Sigma_i \mathbf{a}_i$ and ϵ_i are independent standard normal variables we find $S_n = O_p(1)$, since $\|\Sigma_i\| < \sigma_\epsilon^2 (\max_i m_i) \{\text{tr}(\Lambda) + 1\}$ and $\sum_{i=1}^n \|\mathbf{a}_i\|^2 = 1$. Thus $(\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2) \sum_{i=1}^n \mathbf{a}_i^T \mathbf{e}_i = o_p(1)$. It implies that to prove (A.18) it is sufficient to show that

$$\sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i = O_p(1) \quad (\text{A.19})$$

$$\sum_{i=1}^n \mathbf{a}_i^T \{ \widehat{\mathbf{G}}_i (\widehat{\Lambda}^{-1} + \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i)^{-1} \widehat{\mathbf{G}}_i^T - \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \} \mathbf{e}_i = o_p(1). \quad (\text{A.20})$$

Consider equation (A.19). Because \mathbf{e}_i are multivariate normal with variance Σ_i , set c_i such that $c_i^2 = \mathbf{a}_i^T \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \Sigma_i \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{a}_i$ and rewrite (A.19) as $S_n = \sum_{i=1}^n |c_i| \epsilon_i$, for independent standard normal variables ϵ_i . It suffices to prove that $\sum_{i=1}^n c_i^2 < \infty$, the result then follows from an application of the continuity theorem. For this we show that there exists $L < \infty$ such that $\|\mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \Sigma_i \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T\| \leq L$ for all i . This is not hard to show, because $\|\mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \Sigma_i \mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T\| \leq \|\mathbf{G}_i (\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T\|^2 \|\Sigma_i\|$, and moreover $\|\mathbf{G}_i\|^2 = \text{tr}(\mathbf{G}_i^T \mathbf{G}_i) < M$, $\|(\Lambda^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1}\| \leq M^{1/2} \Lambda_{11} + M (\max_i m_i) \Lambda_{11}^2$ and $\|\Sigma_i\| \leq \sigma_\epsilon^2 (\max_i m_i) \{\text{tr}(\Lambda) + 1\}$. Then $L = \sigma_\epsilon^2 M^3 \Lambda_{11}^2 \{1 + M^{1/2} \Lambda_{11} (\max_i m_i)\}^2 (\max_i m_i) \{\text{tr}(\Lambda) + 1\}$ is finite and satisfies our inequality.

We show next (A.20). Because the eigenvalues σ_k^2 and the eigenfunctions $\theta_k(t)$ are consistently estimated, this reduces to showing that the dominant terms on the left hand side

of equation (A.20) are $o_p(1)$. Equivalently, we need to show that

$$\sum_{i=1}^n \mathbf{a}_i^T (\widehat{\mathbf{G}}_i - \mathbf{G}_i) (\boldsymbol{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i = o_p(1) \quad (\text{A.21})$$

$$\sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i (\boldsymbol{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} (\widehat{\mathbf{G}}_i - \mathbf{G}_i)^T \mathbf{e}_i = o_p(1) \quad (\text{A.22})$$

$$\sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i \{ (\widehat{\boldsymbol{\Lambda}}^{-1} + \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i)^{-1} - (\boldsymbol{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \} \mathbf{G}_i^T \mathbf{e}_i = o_p(1). \quad (\text{A.23})$$

Consider now equation (A.21). The key idea is to use assumption (F1'), that for each subject i , the observation time points t_{ij} are generated uniformly from (t_1, \dots, t_m) . Note that if $t_{il} = t_{i'l'} = t_j$, then $\widehat{g}_{ik,l} = \widehat{g}_{ik,l} = \widehat{\theta}_k(t_j)$, where $\widehat{g}_{ik,l}$ and $g_{ik,l}$ are the l th element of \widehat{g}_{ik} and g_{ik} . The left hand-side of expression (A.21) can be written as

$$\begin{aligned} & \sum_{i=1}^n \mathbf{a}_i^T (\widehat{\mathbf{G}}_i - \mathbf{G}_i) (\boldsymbol{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i \\ &= \sum_{i=1}^n \sum_{l=1}^{m_i} \sum_{k=1}^M a_{il} (\widehat{g}_{ik,l} - g_{ik,l}) e'_{ik} \\ &= \sum_{k=1}^M \sum_{j=1}^m \sum_{i=1}^n \sum_{l=1}^{m_i} a_{il} e'_{ik} (\widehat{g}_{ik,l} - g_{ik,l}) 1(t_{il} = t_j) \\ &= \sum_{k=1}^M \sum_{j=1}^m \{ \widehat{\theta}_k(t_j) - \theta_k(t_j) \} \sum_{i=1}^n \sum_{l=1}^{m_i} a_{il} e'_{ik} 1(t_{il} = t_j), \end{aligned} \quad (\text{A.24})$$

where e'_{ik} is the k th element of $\mathbf{e}'_i = (\boldsymbol{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i$ and $1(t_{il} = t_j)$ is equal to 1 if $t_{il} = t_j$ and 0 otherwise. Set $B_{n,j} = \sum_{i=1}^n \sum_{l=1}^{m_i} a_{il} e'_{ik} 1(t_{il} = t_j)$. Because M is finite, it suffices to show that, for each k

$$E \left[\sum_{j=1}^m \{ \widehat{\theta}_k(t_j) - \theta_k(t_j) \} B_{n,j} \right]^2 = o(1); \quad (\text{A.25})$$

the result that expression (A.24) is $o_p(1)$ follows then from an application of Bonferroni and Chebychev's inequalities. Simple algebra calculations points out that

$$E \left[\sum_{j=1}^m \{ \widehat{\theta}_k(t_j) - \theta_k(t_j) \} B_{n,j} \right]^2 \leq n^{-2\alpha} E \left[\sum_{j=1}^m |B_{n,j}| \right]^2 \leq n^{-2\alpha} m \sum_{j=1}^m E(B_{n,j}^2),$$

using (F2'), that $\sup_{t \in \mathcal{T}} |\widehat{\theta}_k(t) - \theta_k(t)| = O_p(n^{-\alpha})$; thus to show (A.25) it suffices to show that $n^{-2\alpha} m \sum_{j=1}^m E(B_{n,j}^2) = o(1)$ as $n \rightarrow \infty$. This follows from $O(n^{-2\alpha} m) = o(1)$ and

$$E(B_{n,j}^2) = E \left\{ \sum_{i=1}^n \sum_{l=1}^{m_i} a_{il}^2 (e'_{ik})^2 1(t_{il} = t_j) \right\} = m^{-1} E \left\{ \sum_{i=1}^n \sum_{l=1}^{m_i} a_{il}^2 (e'_{ik})^2 \right\} = O(m^{-1}), \quad (\text{A.26})$$

using the independence between e'_{ik} 's and t_{ik} 's, and the fact that $E\{1(t_{il} = t_j)\} = m^{-1}$. For the last equality of (A.26) we used the following observations: 1) $\|\mathbf{a}\| = O(1)$, 2) $E(e'_{ik}) < \text{tr}\{\text{cov}(\mathbf{e}'_i)\}$, and 3) $\|\text{cov}(\mathbf{e}'_i)\| < \infty$, where $\text{cov}(\mathbf{e}'_i) = (\mathbf{\Lambda} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{\Sigma}_i \mathbf{G}_i (\mathbf{\Lambda} + \mathbf{G}_i^T \mathbf{G}_i)^{-1}$.

Next we show (A.22) holds. Following a similar rationale, we rewrite equation (A.22) as

$$\begin{aligned} \sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} (\widehat{\mathbf{G}}_i - \mathbf{G}_i)^T \mathbf{e}_i &= \sum_{i=1}^n \sum_{l=1}^{m_i} \sum_{k=1}^M a'_{ik} (\widehat{g}_{ik,l} - g_{ik,l})^T e_{il} \\ &= \sum_{k=1}^M \sum_{j=1}^m \{\widehat{\theta}_k(t_j) - \theta_k(t_j)\} \sum_{i=1}^n \sum_{l=1}^{m_i} a'_{ik} e_{il} 1(t_{il} = t_j), \end{aligned}$$

where a'_{ik} is the k th element of $\mathbf{a}'_i = \mathbf{a}_i^T \mathbf{G}_i (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1}$. Set $C_{n,j} = \sum_{i=1}^n \sum_{l=1}^{m_i} a'_{ik} e_{il} 1(t_{il} = t_j)$, and denote by \mathbf{a}' the vector obtained by stacking \mathbf{a}'_i over $i = 1, \dots, n$. we have that $\|\mathbf{a}'\| = O(1)$. Using similar arguments as above, we obtain $EC_{n,j}^2 = O(m^{-1})$ for all j and furthermore conclude that $E[\sum_{j=1}^m \{\widehat{\theta}_k(t_j) - \theta_k(t_j)\} C_{n,j}]^2 = o(1)$ and thus equation (A.22) holds.

Finally, we show (A.23) holds. Direct calculations show that

$$\begin{aligned} &\sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i \{(\widehat{\mathbf{\Lambda}}^{-1} + \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i)^{-1} - (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1}\} \mathbf{G}_i^T \mathbf{e}_i \\ &= \sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i (\widehat{\mathbf{\Lambda}}^{-1} + \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i)^{-1} (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i - \widehat{\mathbf{\Lambda}}^{-1} - \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i) (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i. \end{aligned}$$

Using again the consistency of the eigenvalues and eigenfunctions, it suffices to show that

$$\sum_{i=1}^n \mathbf{a}_i^T \mathbf{G}_i (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} (\widehat{\mathbf{\Lambda}}^{-1} + \widehat{\mathbf{G}}_i^T \widehat{\mathbf{G}}_i - \mathbf{\Lambda}^{-1} - \mathbf{G}_i^T \mathbf{G}_i) (\mathbf{\Lambda}^{-1} + \mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T \mathbf{e}_i = o_p(1). \quad (\text{A.27})$$

Use the notation of \mathbf{a}'_i and \mathbf{e}'_i above. Simple algebra points out that (A.27) follows from the following claims: 1) $\sum_{i=1}^n (\mathbf{a}'_i)^T (\widehat{\mathbf{G}}_i^T - \mathbf{G}_i^T) \mathbf{G}_i \mathbf{e}'_i = o_p(1)$, 2) $\sum_{i=1}^n (\mathbf{a}'_i)^T \mathbf{G}_i^T (\widehat{\mathbf{G}}_i - \mathbf{G}_i) \mathbf{e}'_i = o_p(1)$, and 3) $\sum_{i=1}^n (\mathbf{a}'_i)^T (\widehat{\mathbf{\Lambda}}^{-1} - \mathbf{\Lambda}^{-1}) \mathbf{e}'_i = o_p(1)$. We can use roughly the same ideas as earlier to justify 1) and 2). Claim 3) follows from simpler arguments, as we now show. We notice that 3) can be re-written as $\sum_{k=1}^M (\widehat{\sigma}_k^2 / \sigma_\epsilon^2 - \sigma_k^2 / \sigma_\epsilon^2) \sum_{i=1}^n \sum_{l=1}^{m_i} a'_{il} e'_{il} 1(\Lambda_{ll} = \sigma_k^2 / \sigma_\epsilon^2)$, which is $o_p(1)$, since for every k we have $(\widehat{\Lambda}_{kk} - \Lambda_{kk}) = O_p(n^{-\alpha})$ and $\sum_{i=1}^n a'_{ik} e'_{ik} = O_p(1)$.

It follows that equation (A.20) holds and furthermore that condition (C2) is satisfied. #

A.2 Two samples of functional data

AA.1 Independent samples of functional data

Let $Y_{idj} = Y_{id}(t_{idj})$ be the response at time point t_{idj} corresponding to the i th subject within the d th sample, for $d = 1, 2$, $i = 1, \dots, n_d$, and $j = 1, \dots, m_{id}$. As in Section 3.2 it is assumed that $t_{idj} \in \mathcal{T}$ for some bounded and closed interval \mathcal{T} . For simplicity we consider $n_1 = n_2$, but our results can be extended easily to the case when $n_1/n_2 \rightarrow a$ for $0 < a < \infty$. It is assumed that, for each $d = 1, 2$, the response $Y_{id}(t_{idj})$ can be modeled similarly to (6) as:

$$Y_{id}(t_{idj}) = \mu(t_{idj}) + \mu_d(t_{idj}) + \sum_{k \geq 1} \xi_{d,ik} \theta_{d,k}(t) + \epsilon_{idj}, \quad (\text{A.28})$$

where $\mu(\cdot)$ is the overall mean function, $\mu_d(\cdot)$ is the group specific mean deviation, and $\{\theta_{d,k}(t) : k \geq 1\}$ is the group specific orthogonal basis. For identifiability we assume that $\mu_1 + \mu_2 \equiv 0$. Moreover, the $\xi_{d,ik}$ are uncorrelated for all i, k and d , with mean zero and variance $E[\xi_{d,ik}^2] = \sigma_{d,k}^2$, and ϵ_{idj} are assumed independent and identically distributed with mean zero and variance $E[\epsilon_{dj}^2] = \sigma_{d,\epsilon}^2$. Denote by $\Gamma_d(\cdot, \cdot)$ the group d specific covariance function and consider its expansion in terms of orthogonal eigenfunctions, $\Gamma_d(t, t') = \sum_{k \geq 1} \sigma_{d,k}^2 \theta_{d,k}(t) \theta_{d,k}(t')$, where $\theta_{d,k}$ are eigenfunctions and $\sigma_{d,1}^2 > \sigma_{d,2}^2 > \dots$ are ordered eigenvalues for $d = 1, 2$. We assume that for each group the covariance function admits a finite number of non-zero eigenvalues. Our theoretical arguments are based on the additional assumption that $\{\xi_{d,ik}\}_k$ and ϵ_{dj}^2 are jointly Gaussian distributed.

The main objective is to test that the group mean functions are equal, or equivalently that $\mu_1 \equiv 0$. Irrespective of the sampling design (dense or sparse), we assume that the set of *pooled* time points, $\{t_{idj} : i, j\}$ is dense in \mathcal{T} for each d . Our methodology requires that the same sampling scheme is maintained for the two samples of curves, e.g., the curves are not densely observed in one sample and sparsely observed in the other sample. (One could extend the theory to the case of one sample being densely observed and the other sparse, but data of this type would be rare so we did not attempt such an extension.) We use quasi-residuals, $\tilde{Y}_{idj} = Y_{id}(t_{idj}) - \bar{\mu}(t_{idj})$, where $\bar{\mu} = (\tilde{\mu}_1 + \tilde{\mu}_2)/2$ is the average of the estimated mean functions, $\tilde{\mu}_d$ for $d = 1, 2$, which are obtained using the pooled data in each group. Because of the identifiability constraint, the estimated $\bar{\mu}$ can be viewed as a smooth estimate of the overall mean function μ . We assume that the overall mean function is estimated well enough (Kulasekera, 1995), so that \tilde{Y}_{idj} can be modeled similarly to (A.28), but without μ . Thus, we assume that $\mu \equiv 0$ and that the null hypothesis is $\mu_1 \equiv 0$. We

model $\mu_1(t)$ by p th truncated power polynomials: $\mu_1(t) = x_t\boldsymbol{\beta} + z_t\mathbf{b}$, where \mathbf{b} is $N(\mathbf{0}, \sigma_b^2 I_K)$. Let \mathbf{X}_{id} denote the $m_{id} \times (p+1)$ dimensional matrix with the j th row equal to $x_{t_{idj}}$, and let $\tilde{\mathbf{X}}_i = [\mathbf{X}_{i1}^T \mid -\mathbf{X}_{i2}^T]^T$, and analogously define the $m_{id} \times K$ matrices \mathbf{Z}_{id} 's for $d = 1, 2$ and construct $\tilde{\mathbf{Z}}_i = [\mathbf{Z}_{i1}^T \mid -\mathbf{Z}_{i2}^T]^T$ respectively. Here the vertical bar separates submatrices.

Denote by $\tilde{\mathbf{Y}}_i$ the m_i -dimensional vector obtained by stacking first \tilde{Y}_{i1j} 's over $j = 1, \dots, m_{i1}$, and then \tilde{Y}_{i2j} 's over $j = 1, \dots, m_{i2}$, where $m_i = m_{i1} + m_{i2}$. It follows that, the $m_i \times m_i$ -dimensional covariance matrix of $\tilde{\mathbf{Y}}_i$, denoted by $\boldsymbol{\Sigma}_i = \text{diag}\{\boldsymbol{\Sigma}_{i,1}, \boldsymbol{\Sigma}_{i,2}\}$ is a block diagonal $m_i \times m_i$ dimensional matrix, where $\boldsymbol{\Sigma}_{i,d}$ is $m_{id} \times m_{id}$ -dimensional matrix with the (j, j') element equal to $\Gamma_d(t_{idj}, t_{idj'}) + \sigma_{d,\epsilon}^2 1(j = j')$ for $d = 1, 2$. We can rewrite $\tilde{\mathbf{Y}}_i$ using a LMM framework as $\tilde{\mathbf{Y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{Z}}_i\mathbf{b} + \mathbf{e}_i$, where \mathbf{e}_i is m_i -dimensional vector, independent over i , with mean zero, and covariance matrix given by $\boldsymbol{\Sigma}_i$ described above.

Thus the hypothesis $\mu_1 \equiv 0$ is equivalent to $H_0 : \boldsymbol{\beta} = \mathbf{0}$ and $\sigma_b^2 = 0$ in (2); the pseudo LRT can be applied as discussed in Section 3.2, where the estimator $\hat{\boldsymbol{\Sigma}}$ replaces $\boldsymbol{\Sigma}_{i,1}$ and $\boldsymbol{\Sigma}_{i,2}$ with their estimators, $\hat{\boldsymbol{\Sigma}}_{i,1}$ and $\hat{\boldsymbol{\Sigma}}_{i,2}$. The estimators $\hat{\boldsymbol{\Sigma}}_{i,1}$ and $\hat{\boldsymbol{\Sigma}}_{i,2}$ can be obtained as discussed in Section 3.2. Therefore, presuming that the data are densely sampled, condition (C2) of the Proposition 2.1 is met, under the assumption that (F1)–(F2) hold for each of the two samples. Likewise, in the sparse sampling design, (C2) is met under the assumption that (F1')–(F3') hold for each of the two samples. It follows that under these assumptions and the additional assumptions (C1) and (C3) of Proposition 2.1, the asymptotic null distribution of the pseudo LRT for testing the equality of the group mean function is given by (4).

AA.2 Dependent samples of functional data, densely sampled

Assume now two dependent sets of curves, and furthermore consider that in each set, the curves are densely sampled on a common grid of points $t_{idj} = t_j$ and $m_{id} = m$ for all i, d, j . Denote by $\{t_1, \dots, t_m\}$ the common grid of points at which every curve is measured, and denote by $\tilde{Y}_i(t_j) = Y_{i2}(t_j) - Y_{i1}(t_j)$ the i th pairwise difference. Using $\tilde{Y}_i(t_j)$ reduces the data to a one-sample problem and allows us to apply the theory in Section 3. Note that $\tilde{Y}_i(t_j)$ has a similar KL expansion as (6), $\tilde{Y}_i(t_j) = -2\mu_1(t_j) + \sum_{l \geq 1} \tilde{\zeta}_{il}\theta_{2,l}(t_j) + \tilde{\epsilon}_i(t_j)$, where $\tilde{\zeta}_{il}$'s can be viewed as principal component scores, are uncorrelated and have mean zero and variance equal to $2\sigma_{2,l}^2$. Also, $\tilde{\epsilon}_i(t_j)$'s are independent and identically distributed as $N(0, 2\sigma_\epsilon^2)$. To assess the hypothesis that $\mu_1 = 0$, one can apply the pseudo LRT, as discussed in Section 3.2. In particular the conditions required by Proposition 3.2 involve only the estimation of the covariance function Γ_2 and of the noise variance σ_ϵ^2 .

AA.3 Dependent samples of functional data, sparsely sampled

Proof of Proposition 5.1: Under the sparse design, Σ is a block diagonal matrix, whose i th block is the $(m_{i1} + m_{i2}) \times (m_{i1} + m_{i2})$ dimensional matrix

$$\Sigma_i = \begin{pmatrix} \Sigma_{i,11} & \Sigma_{i,12} \\ \Sigma_{i,12}^T & \Sigma_{i,22} \end{pmatrix}, \quad (\text{A.29})$$

where $\Sigma_{i,dd} = \sigma_\epsilon^2(\mathbf{I}_{m_{id}} + \mathbf{G}_{1,id}\mathbf{\Lambda}_1\mathbf{G}_{1,id}^T + \mathbf{G}_{2,id}\mathbf{\Lambda}_2\mathbf{G}_{2,id}^T)$ and $\Sigma_{i,12} = \sigma_\epsilon^2\mathbf{G}_{1,i1}\mathbf{\Lambda}_1\mathbf{G}_{1,i2}^T$. Here $\mathbf{G}_{1,id}$ and $\mathbf{G}_{2,id}$ are $m_{id} \times M_1$ and $m_{id} \times M_2$ dimensional matrices respectively, with the (j, k) th element and the (j, l) th element equal to the eigenfunctions $\theta_{1,k}(t_{idj})$, and $\theta_{2,l}(t_{idj})$ respectively, and $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are the $M_1 \times M_1$ and $M_2 \times M_2$ block diagonal matrices whose (k, k) th and (l, l) th components are $\sigma_{1,k}^2/\sigma_\epsilon^2$ and $\sigma_{2,l}^2/\sigma_\epsilon^2$ respectively. Similarly, define $\widehat{\Sigma}_i$ by replacing σ_ϵ^2 , $\mathbf{G}_{\iota,id}$, $\mathbf{\Lambda}_\iota$ with their respective estimators, for $\iota = 1, 2$. Partition \mathbf{a} and \mathbf{e} into n vectors, with $\mathbf{a}_i = (\mathbf{a}_{i1}^T, \mathbf{a}_{i2}^T)^T$ and $\mathbf{e}_i = (\mathbf{e}_{i1}^T, \mathbf{e}_{i2}^T)^T$ of length $m_{i1} + m_{i2}$. We want to prove condition (C2),

$$\sum_{i=1}^n \mathbf{a}_i^T \widehat{\Sigma}_i^{-1} \mathbf{a}_i - \sum_{i=1}^n \mathbf{a}_i^T \Sigma_i^{-1} \mathbf{a}_i = o_p(1), \quad (\text{A.30})$$

$$\sum_{i=1}^n \mathbf{a}_i^T \widehat{\Sigma}_i^{-1} \mathbf{e}_i - \sum_{i=1}^n \mathbf{a}_i^T \Sigma_i^{-1} \mathbf{e}_i = o_p(1). \quad (\text{A.31})$$

The equality (A.30) follows from assumption (M3'), which implies that $\max_i \|\widehat{\mathbf{G}}_{\iota,id} - \mathbf{G}_{\iota,id}\| = o_p(1)$, $\max_i \|\widehat{\mathbf{\Lambda}}_{\iota,ii} - \mathbf{\Lambda}_{\iota,ii}\| = o_p(1)$, and $\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 = o_p(1)$ and from the continuous mapping theorem, which implies that $\max_i \|\widehat{\Sigma}_i^{-1} - \Sigma_i^{-1}\| = o_p(1)$. The proof is concluded since $\sum_{i=1}^n \|\mathbf{a}_i\|^2 = O(1)$.

We turn to equality (A.31). We re-write Σ_i^{-1} using inverse of the partition matrix as follows:

$$\Sigma_i^{-1} = \begin{pmatrix} \mathbf{V}_{i1}^{-1} & -\Sigma_{i,11}^{-1}\Sigma_{i,12}\mathbf{V}_{i2}^{-1} \\ -\Sigma_{i,22}^{-1}\Sigma_{i,12}^T\mathbf{V}_{i1}^{-1} & \mathbf{V}_{i2}^{-1} \end{pmatrix} = \begin{pmatrix} [\Sigma_i^{-1}]_{11} & [\Sigma_i^{-1}]_{12} \\ [\Sigma_i^{-1}]_{21} & [\Sigma_i^{-1}]_{22} \end{pmatrix},$$

where $\mathbf{V}_{i1} = \Sigma_{i,11} - \Sigma_{i,12}\Sigma_{i,22}^{-1}\Sigma_{i,12}^T$, and $\mathbf{V}_{i2} = \Sigma_{i,22} - \Sigma_{i,12}^T\Sigma_{i,11}^{-1}\Sigma_{i,12}$. Similarly, we can define $\widehat{\Sigma}_i^{-1}$ by replacing the quantities with their estimates. The left hand side of equation (A.31) can be decomposed into $v_{11} + v_{12} + v_{21} + v_{22}$, where $v_{sl} = \sum_{i=1}^n \mathbf{a}_{is}^T([\widehat{\Sigma}_i^{-1}]_{sl} - [\Sigma_i^{-1}]_{sl})\mathbf{e}_{il}$. It is sufficient to show that $v_{sl} = o_p(1)$ for $1 \leq s, l \leq 2$. These results can be derived using similar techniques as in the proof of the Proposition 3.3, but they involve more tedious algebra. In the interest of space the details are omitted here; however for completeness they are publicly available on one of the authors's website. It follows that our proof is concluded. $\#$