Supplemental Material for “Model-based joint curve registration and classification”

Lin Tang[a,1], Pengcheng Zeng[b,1], Jian Qing Shi [*c], and Won-Seok Kim[d]

[a]Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming, Yunnan, China

[b]Institute of Mathematical Sciences, ShanghaiTech University, Shanghai, China

[c]Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, China

[d]Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, South Korea

Assumptions for Theorems

Let $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ be the $L^2(\mathcal{T})$ inner product and norm, respectively. For notational simplicity, let $\tilde{\boldsymbol{x}}_{ai}(t) = \boldsymbol{x}_{ai}(g^{-1}(t))$ and define the covariance of the curve $\tilde{\boldsymbol{x}}_{ai}(t)$ as $K_{x_a}(s,t) = \mathrm{Cov}[\tilde{\boldsymbol{x}}_{ai}(s), \tilde{\boldsymbol{x}}_{ai}(t)]$, where $K_{x_a}(s,t)$ is continuous on the interval $[0,1]$. Then by Mercer's Theorem we have

$$K_{x_a}(s,t) = \sum_{l=1}^{\infty} \lambda_{al}\phi_{al}(s)\phi_{al}(t),$$

where $\lambda_{a1} > \lambda_{a2} > \ldots > 0$ are eigenvalues and $\phi_{a1}(t), \phi_{a2}(t) \ldots$ are eigenfunctions of the covariance operator corresponding to $K_{x_a}(s,t)$. Then by the Karhunen-Loève representation, the process $\tilde{\boldsymbol{x}}_{ai}$

[1]Both authors contributed equally to this work.

[*]Corresponding author, email: shijq@sustech.edu.cn

and the functional coefficient $\boldsymbol{\beta}_a(t)$ follows the following eigen decompositions

$$\tilde{\boldsymbol{x}}_{ai}(t) = \sum_{l=1}^{\infty} p_{ail}\phi_{al}(t) \quad \boldsymbol{\beta}_a(t) = \sum_{l=1}^{\infty} e_{al}\phi_{al}(t),$$

where $p_{ail} = \langle \tilde{\boldsymbol{x}}_{ai}, \phi_{al} \rangle$ are uncorrelated random variables with zero mean and variance λ_{al} and $e_{al} = \langle \boldsymbol{\beta}_a, \phi_{al} \rangle$.

In practice, $K_{x_a}(s,t)$ is unknown and we can take the empirical version by

$$\hat{K}_{x_a}(s,t) = \sum_{l=1}^{\infty} \hat{\lambda}_{al}\hat{\phi}_{al}(s)\hat{\phi}_{al}(t),$$

where $(\hat{\lambda}_{al}, \hat{\phi}_{al})$ is the estimator of $(\lambda_{al}, \phi_{al})$ with $\hat{\lambda}_{a1} \geq \hat{\lambda}_{a2} \ldots \geq 0$.

Therefore, the systematic component in equation (12) can be rewritten as

$$\eta_i = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b} + \int_0^1 \tilde{\boldsymbol{x}}_i(t)\boldsymbol{\beta}(t)dt = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b} + \sum_{a=1}^{2}\sum_{l=1}^{\infty} \tilde{p}_{ail}e_{al} = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b} + \sum_{a=1}^{2}\sum_{l=1}^{K_x} \tilde{p}_{ail}e_{al} + R_i,$$

where $\tilde{p}_{ail} = \langle \tilde{\boldsymbol{x}}_{ai}, \hat{\phi}_{al} \rangle$, $R_i = \sum\limits_{a=1}^{2}\sum\limits_{l=K_x+1}^{\infty} \tilde{p}_{ail}e_{al}$ and K_x is the tuning parameter which is set to be sufficiently large.

For notational simplicity, we define $\tilde{\boldsymbol{p}}_i = (\tilde{\boldsymbol{p}}_{1i}^{\mathrm{T}}, \tilde{\boldsymbol{p}}_{2i}^{\mathrm{T}})^{\mathrm{T}} = (\tilde{p}_{1i1}, \ldots, \tilde{p}_{1iK_x}, \tilde{p}_{2i1}, \ldots, \tilde{p}_{2iK_x})^{\mathrm{T}}$, $\boldsymbol{e} = (\boldsymbol{e}_1^{\mathrm{T}}, \boldsymbol{e}_2^{\mathrm{T}})^{\mathrm{T}} = (e_{11}, \ldots, e_{1K_x}, e_{21}, \ldots, e_{2K_x})^{\mathrm{T}}$, $\boldsymbol{Y} = (y_1, \ldots, y_N)^{\mathrm{T}}$, $\boldsymbol{v} = (\boldsymbol{v}_1^{\mathrm{T}}, \ldots, \boldsymbol{v}_N^{\mathrm{T}})^{\mathrm{T}}$, $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1(t), \ldots, \tilde{\boldsymbol{x}}_N(t))$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)^{\mathrm{T}}$.

Let $\boldsymbol{D} = (\boldsymbol{v}, \tilde{\boldsymbol{X}})$ and $\boldsymbol{\theta} = (\boldsymbol{b}^{\mathrm{T}}, \boldsymbol{e}^{\mathrm{T}})^{\mathrm{T}}$ be the unknown parameter vector of the models defined in equation (11) and equation (12). Then the estimation $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{b}}^{\mathrm{T}}, \hat{\boldsymbol{e}}^{\mathrm{T}})^{\mathrm{T}}$ is obtained by maximizing the following log-likelihood function

$$\ell(\boldsymbol{\theta}) \triangleq \ell(\boldsymbol{Y};\boldsymbol{\eta}) = \sum_{i=1}^{N} \ell_i(y_i;\eta_i) = \sum_{i=1}^{N} \left\{ y_i\boldsymbol{\zeta}(\boldsymbol{v}_i, \boldsymbol{x}_i) + \mathcal{B}[\boldsymbol{\zeta}(\boldsymbol{v}_i, \boldsymbol{x}_i)] + \mathcal{C}(y_i) \right\}$$

For simplicity, let C be a constant whose value might change according to different circumstances. Denote $\dot{\ell}_{i,d}(y_i;d)$ and $\ddot{\ell}_{i,d}(y_i;d)$ as the first- and second-order derivative of $\ell_i(y_i;d)$ with respect to d, respectively. Also, similar to [1], we define

$$\mathrm{I}(\boldsymbol{D}) = -\mathrm{E}\big[\ddot{\ell}_{\eta}(\boldsymbol{Y};\eta)|\boldsymbol{D}\big] = -\mathrm{E}\big[\sum_{i=1}^{N} \ddot{\ell}_{i,\eta_i}(y_i;\eta_i)|\boldsymbol{D}\big],$$

where $\eta_i = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b} + \int_0^1 \tilde{\boldsymbol{x}}_i(t)\boldsymbol{\beta}(t)dt$.

In model (11), the response variable y_i is related to both the scalar variables and the functional variables. So, the main complicated issue comes from the dependence between $\boldsymbol{v}_i$ and $\tilde{\boldsymbol{x}}_i(t)$. To solve this problem, similar to [2], we define

$$\boldsymbol{G}(\tilde{\boldsymbol{X}}) = \frac{\mathrm{E}(\boldsymbol{v}\ddot{\ell}_\eta(\boldsymbol{Y};\eta)|\tilde{\boldsymbol{X}})}{\mathrm{E}(\ddot{\ell}_\eta(\boldsymbol{Y};\eta)|\tilde{\boldsymbol{X}})}, \text{ and } \boldsymbol{v} = \tilde{\boldsymbol{v}} + \boldsymbol{G}(\tilde{\boldsymbol{X}}),$$

where $\tilde{\boldsymbol{v}} = (\tilde{v}_1, \dots, \tilde{v}_{p+1})^{p+1}$ is a zero mean $(p+1)$-dimensional random vector and $\boldsymbol{G}(\tilde{\boldsymbol{X}}) = (G_1(\tilde{\boldsymbol{X}}), \dots, G_{p+1}(\tilde{\boldsymbol{X}}))^{\mathrm{T}}$ is a $(p+1)$-dimensional functional vector with $G_j(\tilde{\boldsymbol{X}}) \in L^2(\mathcal{T})$ for $j = 1, \dots, p+1$.

Suppose that the following assumptions hold

(A1) For each $i \in \{1, \dots, N\}$, $\mathrm{E}\|\tilde{\boldsymbol{x}}_i\|^4 < \infty$.

(A2) For each $a \in [1,2]$ and l, $\mathrm{E}(p_{ail}^4) \le C\lambda_{al}^2$ with the eigenvalue λ_{al} satisfies $C^{-1}l^{-\alpha} \le \lambda_{al} \le Cl^{-\alpha}$ and $\lambda_{al} - \lambda_{a(l+1)} \ge C^{-1}l^{-\alpha-1}$ for $l \ge 1$ and some constant $\alpha > 1$. In addition, $|e_{al}^*| \le Cl^{-\gamma}$ for some constant $\gamma > \alpha/2 + 1$.

(A3) The tuning parameter K_x satisfies

$$K_x \asymp N^{\frac{1}{\alpha+2\gamma}},$$

where the notation $a_N \asymp b_N$ means that there exist constants $0 < L < M < \infty$ such that $L \le a_N/b_N \le M$ for all N.

(A4) For each i, the scalar covariates $\boldsymbol{v}_i$ satisfies $\mathrm{E}\|\boldsymbol{v}_i\|^4 < \infty$.

(A5) $\mathrm{E}(\tilde{\boldsymbol{v}}) = \boldsymbol{0}$,

$$\boldsymbol{\Omega}_1 = \mathrm{E}\big\{\sum_{i=1}^N \ddot{\ell}_{i,\eta_i^*}(y_i;\eta_i^*)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}\big\} \text{ and } \boldsymbol{\Omega}_2 = \mathrm{E}\big\{\sum_{i=1}^N \dot{\ell}_{i,\eta_i^*}^2(y_i;\eta_i^*)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}\big\},$$

where $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are assumed to be positive definite matrices and $\eta_i^* = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b}^* + \int_0^1 \tilde{\boldsymbol{x}}_i(t)\boldsymbol{\beta}^*(t)$.

(A6) $|I(\boldsymbol{D})| < C$ and $I(\boldsymbol{D})$ satisfies the first-order Lipschitz condition.

(A7) The true value $\boldsymbol{\theta}_x^*$ of $\boldsymbol{\theta}_x$ is unique and $\hat{\boldsymbol{\theta}}_x \xrightarrow{p} \boldsymbol{\theta}_x^*$ where $\hat{\boldsymbol{\theta}}_x$ is the MLE of $\boldsymbol{\theta}_x$.

(A8) For $i = 1, \dots, N$, the likelihood function $\ell_i(\boldsymbol{\theta}_x)$ is thrice continuously differentiable with respect to $\boldsymbol{\theta}_x$.

(A9) There exist positive definite matrices $\boldsymbol{A}(\boldsymbol{\theta}_x^*)$ and $\boldsymbol{B}(\boldsymbol{\theta}_x^*)$ such that

$$\lim_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N}\partial^2_{\theta_x}\ell_i(\boldsymbol{\theta}_x^*) = \boldsymbol{A}(\boldsymbol{\theta}_x^*),\ \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\partial_{\theta_x}\ell_i(\boldsymbol{\theta}_x^*)\partial_{\theta_x}\ell_i(\boldsymbol{\theta}_x^*)^{\mathrm{T}} = \boldsymbol{B}(\boldsymbol{\theta}_x^*).$$

Remark 1. *Assumptions (A1)-(A3) are quite usual in the settings of functional linear model(see [3] and [4]). Assumption (A1) requires that $\tilde{\boldsymbol{x}}_i$ has a finite fourth moment and is necessary to get the L^2 convergence of the estimated functional coefficients. In particular, the condition $E(p_{ail}^4) \leq C\lambda_{al}^2$ in Assumption (A2) holds if the random process $\tilde{\boldsymbol{x}}_i$ is a Gaussian Process. The requirements of the eigenvalues in Assumption (A2) prevent the spacings between adjacent eigenvalues from being too small and ask that the slope function $\beta(t)$ is smoother than the sample path of $\tilde{\boldsymbol{x}}_i$ [2]. They are of great importance in ensuring the rate of convergence in Theorem 2.3. The last part of Assumptions (A2) prevents the coefficients e_{al}^* from decreasing too slowly. To optimize the convergence rate of the functional coefficients, requirement of smoothing parameter in Assumption (A3) is needed. Assumptions (A4)-(A6) are used to deal with the linear part with scalar variables. Assumption (A4) is analogy to Assumption (A1). The condition of the Fisher information $I(\boldsymbol{D})$ in Assumption (A6) is analogous to [1]. Assumptions (A7)-(A8) are commonly used conditions in parametric models, and they are applied here to develop the asymptotical consistency of the MLE for the functional nonlinear mixed effects model for curve alignment.*

Technical lemmas and proofs.

Let $\tilde{\eta}_i = \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b}^* + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^*$, where $\boldsymbol{b}^*$ and $\boldsymbol{e}^*$ denote the true values of $\boldsymbol{b}$ and $\boldsymbol{e}$, respectively.

Lemma 1. *Let $R_i = \int_0^1 \tilde{\boldsymbol{x}}_i(t)\boldsymbol{\beta}^*(t)dt - \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^*$ for $i = 1, \ldots, N$. Then under Assumptions (A1)-(A6), we have*

$$\|R_i\|^2 = \|\eta_i^* - \tilde{\eta}_i\|^2 = O_p(N^{-(2\gamma+\alpha-1)/(\alpha+2\gamma)}).$$

Thus

$$\eta_i^* - \tilde{\eta}_i = O_p(N^{-(2\gamma+\alpha-1)/2(\alpha+2\gamma)}).$$

Proof. Note that

$$
\begin{aligned}
R_i &= \int_0^1 \tilde{\boldsymbol{x}}_i(t)\boldsymbol{\beta}^*(t)dt - \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^* \\
&= \sum_{a=1}^{2}\sum_{l=1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}\rangle\langle\boldsymbol{\beta}_a^*,\phi_{al}\rangle - \sum_{a=1}^{2}\sum_{l=1}^{K_x}\langle\tilde{\boldsymbol{x}}_{ai},\hat{\phi}_{al}\rangle\langle\boldsymbol{\beta}_a^*,\phi_{al}\rangle \\
&= \sum_{a=1}^{2}\sum_{l=1}^{K_x}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}\rangle\langle\boldsymbol{\beta}_a^*,\phi_{al}\rangle - \sum_{a=1}^{2}\sum_{l=1}^{K_x}\langle\tilde{\boldsymbol{x}}_{ai},\hat{\phi}_{aj}\rangle\langle\boldsymbol{\beta}_a^*,\phi_{al}\rangle \\
&+ \sum_{a=1}^{2}\sum_{l=K_x+1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}\rangle\langle\boldsymbol{\beta}_a^*,\phi_{al}\rangle \\
&= \sum_{a=1}^{2}\sum_{l=1}^{K_x}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}-\hat{\phi}_{al}\rangle e_{al}^* + \sum_{a=1}^{2}\sum_{l=K_x+1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}\rangle e_{al}^* \\
&= I_1 + I_2
\end{aligned}
$$

where $I_1 = \sum_{a=1}^{2}\sum_{l=1}^{K_x}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}-\hat{\phi}_{al}\rangle e_{al}^*$, and
$I_2 = \sum_{a=1}^{2}\sum_{l=K_x+1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}\rangle e_{al}^* = \sum_{a=1}^{2}\sum_{l=K_x+1}^{\infty} p_{ail}e_{al}^*$.

Since $\|\hat{\phi}_{al}-\phi_{al}\|^2 = O_p(N^{-1}l^2)$, then it follows from Assumptions (A1)-(A3) that

$$
\begin{aligned}
\|I_1\|^2 &= \|\sum_{a=1}^{2}\sum_{l=1}^{K_x}\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}-\hat{\phi}_{al}\rangle e_{al}^*\|^2 \\
&\leq 2K_x\sum_{l=1}^{K_x}\|\langle\tilde{\boldsymbol{x}}_{ai},\phi_{al}-\hat{\phi}_{al}\rangle e_{al}^*\|^2 \\
&\leq 2K_x\sum_{l=1}^{K_x}\|\phi_{al}-\hat{\phi}_{al}\|^2|e_{al}^*|^2 \\
&\leq O_P(K_x)\sum_{l=1}^{K_x}N^{-1}l^2l^{-2\gamma} \\
&\leq O_P(N^{-1}K_x)\sum_{l=1}^{K_x}l^{-2\gamma+2} \\
&\leq O_P(N^{-1}K_x) = O_P(N^{-\frac{2\gamma+\alpha+1}{\alpha+2\gamma}})
\end{aligned}
$$

For I_2, note that p_{ail} are uncorrelated random variables with zero mean and variance λ_{al}, then we have

$$
\begin{aligned}
\mathrm{E}(\|I_2\|^2) &= \mathrm{E}(\sum_{a=1}^{2}\sum_{l=K_x+1}^{\infty} p_{ail}e_{al}^*)^2 \\
&= \sum_{a=1}^{2}\sum_{l=K_x+1}^{\infty} e_{al}^{*2}\lambda_{al} \\
&\leq 2C\sum_{l=K_x+1}^{\infty} l^{-(2\gamma+\alpha)} \\
&= O_p(K_x^{-(2\gamma+\alpha-1)}) = O_p(N^{-\frac{2\gamma+\alpha-1}{\alpha+2\gamma}})
\end{aligned}
$$

Then $R_i = I_1 + I_2 = O_p(N^{-\frac{2\gamma+\alpha-1}{2(\alpha+2\gamma)}}) + O_p(N^{-\frac{2\gamma+\alpha-1}{2(\alpha+2\gamma)}}) = O_p(N^{-\frac{2\gamma+\alpha-1}{2(\alpha+2\gamma)}})$ holds, which indicates that $\|R_i\|^2 = \|\eta_i^* - \tilde{\eta}_i\|^2 = O_p(N^{-(2\gamma+\alpha-1)/(\alpha+2\gamma)})$, thus Lemma 1 holds. □

Let $\tilde{\boldsymbol{v}}_i = \boldsymbol{v}_i - \boldsymbol{G}(\tilde{\boldsymbol{x}}_i)$ and

$$
\breve{\boldsymbol{b}} = \arg\max_{\boldsymbol{b}} \sum_{i=1}^{N} \ell_i(y_i; \tilde{\boldsymbol{v}}_i^{\mathrm{T}}\boldsymbol{b} + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^* + \boldsymbol{G}(\tilde{\boldsymbol{x}}_i)\boldsymbol{b}^*).
$$

Then the following Lemma says that the estimation $\breve{\boldsymbol{b}}$ is asymptotically distributed as normal distribution.

Lemma 2. *Under Assumptions (A1)-(A9), we have*

$$
\sqrt{N}(\breve{\boldsymbol{b}} - \boldsymbol{b}^*) \to N(\boldsymbol{0}, \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1}),
$$

where $\boldsymbol{\Omega}_1 = E\{\sum_{i=1}^{N} \ddot{\ell}_{i,\eta_i^*}(y_i;\eta_i^*)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}\}$, $\boldsymbol{\Omega}_2 = E\{\sum_{i=1}^{N} \dot{\ell}_{i,\eta_i^*}^2(y_i;\eta_i^*)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}\}$ *and* $\tilde{\boldsymbol{v}}_i = \boldsymbol{v}_i - \boldsymbol{G}(\tilde{\boldsymbol{x}}_i)$.

Proof. Let $\boldsymbol{\omega} = \sqrt{N}(\boldsymbol{b} - \boldsymbol{b}^*)$ and $\breve{\boldsymbol{\omega}} = \sqrt{N}(\breve{\boldsymbol{b}} - \boldsymbol{b}^*)$, which according to the definition of $\breve{\boldsymbol{b}}$, is obtained by maximizing the following function

$$
M(\boldsymbol{\omega}) = \sum_{i=1}^{N} \ell_i(y_i; \tilde{\boldsymbol{v}}_i^{\mathrm{T}}\boldsymbol{b}^* + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^* + \boldsymbol{G}(\tilde{\boldsymbol{x}}_i)\boldsymbol{b}^* + \tilde{\boldsymbol{v}}_i^{\mathrm{T}}\boldsymbol{\omega}/\sqrt{N}) - \sum_{i=1}^{N} \ell_i(y_i; \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b}^* + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^*).
$$

Taking a second-order Taylor's expansion of $M(\boldsymbol{\omega})$ yields

$$
M(\boldsymbol{\omega}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \dot{\ell}_{i,\tilde{\eta}_i}(y_i;\tilde{\eta}_i)\tilde{\boldsymbol{v}}_i^{\mathrm{T}}\boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega},
$$

where

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} \ddot{\ell}_{i,\tilde{\eta}_i}(y_i; \tilde{\eta}_i + \nu_i)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}$$

with ν_i lies between 0 and $\tilde{\boldsymbol{v}}_i\boldsymbol{\omega}/\sqrt{N}$. It follows from [5] that $\boldsymbol{\Sigma} = -\boldsymbol{\Omega}_1 + o_p(1)$.

On the other hand,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} \dot{\ell}_{i,\tilde{\eta}_i}(y_i; \tilde{\eta}_i)\tilde{\boldsymbol{v}}_i^{\mathrm{T}} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \dot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i^{\mathrm{T}} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \ddot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i^{\mathrm{T}}(\tilde{\eta}_i - \eta_i^*)\\ &+ o_p(1)\\ &= I_3 + I_4 + o_p(1),\end{aligned}$$

where

$$I_3 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \dot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i^{\mathrm{T}} \text{ and } I_4 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \ddot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i^{\mathrm{T}}(\tilde{\eta}_i - \eta_i^*).$$

It is easy to find that $I_4 = o_p(1)$. By the Lindeberg-Feller central limit theory, we have $I_3 \to \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_2)$. Then,

$$M(\boldsymbol{\omega}) = I_3^{\mathrm{T}}\boldsymbol{\omega} - \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Omega}_1\boldsymbol{\omega} + o_p(1).$$

The results of [6] and [7] show that

$$\breve{\boldsymbol{\omega}} = \boldsymbol{\Omega}_1^{-1}I_3 + o_p(1),$$

then Lemma 2 holds from the Slutsky Theorem. □

Lemma 3. *Under Assumptions (A1)-(A9), we have*

$$\|\hat{\boldsymbol{\theta}} - \breve{\boldsymbol{\theta}}\|^2 = O_p(N^{-\frac{2\gamma-1}{\alpha+2\gamma}}),$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{b}}^{\mathrm{T}}, \hat{\boldsymbol{e}}^{\mathrm{T}})^{\mathrm{T}}$ *and* $\breve{\boldsymbol{\theta}} = (\breve{\boldsymbol{b}}^{\mathrm{T}}, \boldsymbol{e}^{*\mathrm{T}})^{\mathrm{T}}$.

Proof. Taking a first-order Taylor's expansion of $\dot{\ell}(\hat{\boldsymbol{\theta}}) = \frac{\partial \ell(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}$ at $\breve{\boldsymbol{\theta}}$ yields

$$0 = \dot{\ell}(\hat{\boldsymbol{\theta}}) = \dot{\ell}(\breve{\boldsymbol{\theta}}) + \ddot{\ell}(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \breve{\boldsymbol{\theta}}) + o_p(1),$$

where $\bar{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}$ and $\breve{\boldsymbol{\theta}}$, $\dot{\ell}(\breve{\boldsymbol{\theta}}) = \frac{\partial \ell(\theta)}{\partial \theta}|_{\theta=\breve{\theta}}$ and $\dot{\ell}(\bar{\boldsymbol{\theta}}) = \frac{\partial \ell(\theta)}{\partial \theta}|_{\theta=\bar{\theta}}$.

Then we have

$$\hat{\boldsymbol{\theta}} - \breve{\boldsymbol{\theta}} = -[\ddot{\ell}(\bar{\boldsymbol{\theta}})]^{-1}\dot{\ell}(\breve{\boldsymbol{\theta}}).$$

Denote

$$\dot{\ell}(\breve{\boldsymbol{\theta}}) = \left\{(\frac{\partial \ell(\breve{\boldsymbol{\theta}})}{\partial \boldsymbol{b}})^{\mathrm{T}}, (\frac{\partial \ell(\breve{\boldsymbol{\theta}})}{\partial \boldsymbol{e}})^{\mathrm{T}}\right\}^{\mathrm{T}} = \sum_{i=1}^{N} \dot{\ell}_{i,\breve{\eta}_i}(y_i;\breve{\eta}_i)(\boldsymbol{v}_i^{\mathrm{T}}, \tilde{\boldsymbol{p}}_i^{\mathrm{T}})^{\mathrm{T}},$$

where $\breve{\eta}_i = \boldsymbol{v}_i^{\mathrm{T}}\breve{\boldsymbol{b}} + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\boldsymbol{e}^*$.

Note that

$$\frac{\partial \ell(\breve{\boldsymbol{\theta}})}{\partial \boldsymbol{b}} = \sum_{i=1}^{N} \dot{\ell}_{i,\breve{\eta}_i}(y_i;\breve{\eta}_i)\boldsymbol{v}_i = \sum_{i=1}^{N} \dot{\ell}_{i,\eta_i^*}(y_i;\eta_i^*)\boldsymbol{v}_i + \sum_{i=1}^{N} \ddot{\ell}_{i,\bar{\eta}_i}(y_i;\bar{\eta}_i)(\boldsymbol{v}_i^{\mathrm{T}}(\breve{\boldsymbol{b}} - \boldsymbol{b}^*) + R_i)\boldsymbol{v}_i,$$

and

$$\frac{\partial \ell(\breve{\boldsymbol{\theta}})}{\partial \boldsymbol{e}} = \sum_{i=1}^{N} \dot{\ell}_{i,\breve{\eta}_i}(y_i;\breve{\eta}_i)\tilde{\boldsymbol{p}}_i = \sum_{i=1}^{N} \dot{\ell}_{i,\eta_i^*}(y_i;\eta_i^*)\tilde{\boldsymbol{p}}_i + \sum_{i=1}^{N} \ddot{\ell}_{i,\bar{\eta}_i}(y_i;\bar{\eta}_i)(\tilde{\boldsymbol{p}}_i^{\mathrm{T}}(\breve{\boldsymbol{e}} - \boldsymbol{e}^*) + R_i)\tilde{\boldsymbol{p}}_i,$$

where $\bar{\eta}_i$ lies between η_i^* and $\breve{\eta}_i$.

Similar to [8], we have

$$\mathrm{E}\big(\| \sum_{i=1}^{N} \dot{\ell}_{i,\eta_i^*}(y_i;\eta_i^*)\boldsymbol{v}_i\|\big) = O(N^{1/2}). \tag{S1}$$

The Assumption (A2) and Lemma 1 indicate that

$$\begin{aligned} \|\sum_{i=1}^{N} \ddot{\ell}_{i,\bar{\eta}_i}(y_i;\bar{\eta}_i)(\boldsymbol{v}_i^{\mathrm{T}}(\breve{\boldsymbol{b}} - \boldsymbol{b}^*) + R_i)\boldsymbol{v}_i\| &= O_p(N^{1/2}) + O_p(N \cdot N^{-(2\gamma+\alpha-1)/2(\alpha+2\gamma)}) \\ &= O_p(\sqrt{NK_x}). \end{aligned} \tag{S2}$$

Equation (S1) and (S2) show that $\partial \ell(\breve{\boldsymbol{\theta}})/\partial \boldsymbol{b} = O_p(\sqrt{NK_x})$. Similarly, we have $\partial \ell(\breve{\boldsymbol{\theta}})/\partial \boldsymbol{e} = O_p(\sqrt{NK_x})$. Therefore, $\dot{\ell}(\breve{\boldsymbol{\theta}}) = O_p(\sqrt{NK_x})$.

Similar to Lemma A.3 of [9], we have $\big\|\big(\frac{1}{N}\ddot{\ell}(\bar{\boldsymbol{\theta}})\big)^{-1}\big\| = O_p(\lambda_{K_x}^{-1/2}) = O_p(K_x^{\alpha/2})$, which yields

$$\begin{aligned} \|\hat{\boldsymbol{\theta}} - \breve{\boldsymbol{\theta}}\| &\leq \big\|\big(\tfrac{1}{N}\ddot{\ell}(\bar{\boldsymbol{\theta}})\big)^{-1}\big\|\big\|\tfrac{1}{N}\dot{\ell}(\breve{\boldsymbol{\theta}})\big\| \\ &= O_p(K_x^{\alpha/2})O_p(N^{-(2\gamma+\alpha-1)/2(\alpha+2\gamma)}) \\ &= O_p(N^{-(2\gamma-1)/2(\alpha+2\gamma)}) \end{aligned}$$

Thus, the result $\|\hat{\boldsymbol{\theta}} - \breve{\boldsymbol{\theta}}\|^2 = O_p(N^{-(2\gamma-1)/(\alpha+2\gamma)})$ holds. □

Proof of Theorem 2.1. The identifiability proof goes as follows. By assumptions in model (1), the random effects and the random error can be integrated into a new Gaussian process error denoted as $\epsilon_i^*(t)$ with $E\{\epsilon_i^*(t)\} = 0$. Then we have that $\epsilon_i^*(t) = x_i(t) - \tau(g_i(t))$, thus there is no ambiguity about the error term. Suppose that $E\{x_i(t)\} = \tau_1(g_{1i}(t)) = \tau_2(g_{2i}(t))$ for all $i = 1, \ldots, N$, then we have $\tau_1(t) = \tau_2(g_{2i}(g_{1i}^{-1}(t)))$. Since the left-hand side of this equation doesn't depend on i, we have that $g_{2i}(g_{1i}^{-1}(t)) = l(t)$ for all i some function $l(\cdot)$. Then $g_{1i}^{-1}(t) = g_{2i}^{-1}(l(t))$ for all i. The assumption in Theorem 1 shows that $E(g_{1i}) = E(g_{2i})$, then we have $l(t) = t$. Therefore, $g_{1i}(t) = g_{2i}(t)$ for all i and the warping functions are identifiable. $\square$

Proof of Theorem 2.2. Let $\hat{\eta}_i = \boldsymbol{v}_i^{\mathrm{T}}\hat{\boldsymbol{b}} + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\hat{\boldsymbol{e}}$ and for any $\boldsymbol{z} \in \mathbb{R}^{p+1}$, define $\hat{\eta}_i(\boldsymbol{z}) = \boldsymbol{v}_i^{\mathrm{T}}\hat{\boldsymbol{b}} + \tilde{\boldsymbol{p}}_i^{\mathrm{T}}\hat{\boldsymbol{e}} + \tilde{\boldsymbol{v}}_i^{\mathrm{T}}\boldsymbol{z}$, where $\tilde{\boldsymbol{v}}_i = \boldsymbol{v}_i - \boldsymbol{G}(\tilde{\boldsymbol{x}}_i)$. Obviously, when $\boldsymbol{z} = \boldsymbol{0}$,

$$\hat{\eta}(\boldsymbol{z}) = \arg\max_{\eta(\boldsymbol{z})} \ell(\boldsymbol{Y}; \eta(\boldsymbol{z})).$$

Then the following equation follows from a Taylor's expansion

$$\begin{aligned} 0 = \tfrac{\partial \ell(\hat{\eta}(\boldsymbol{z}))}{\partial \boldsymbol{z}}|_{\boldsymbol{z}=\boldsymbol{0}} &= \sum_{i=1}^N \dot{\ell}_{i,\hat{\eta}_i}(y_i; \hat{\eta}_i)\tilde{\boldsymbol{v}}_i \\ &= \sum_{i=1}^N \dot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i + \sum_{i=1}^N \ddot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i(\hat{\eta}_i - \eta_i^*) + o_p(1), \end{aligned}$$

Applying Lemma 2 and Lemma 3, the second term on the right hand side of the above equation can be rewritten as

$$\frac{1}{N}\sum_{i=1}^N \ddot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i(\hat{\eta}_i - \eta_i^*) = \frac{1}{N}\sum_{i=1}^N \ddot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}(\hat{\boldsymbol{b}} - \boldsymbol{b}^*) + o_p(N^{1/2}).$$

Then we have

$$(\hat{\boldsymbol{b}} - \boldsymbol{b}^*) = -\big[\frac{1}{N}\sum_{i=1}^N \ddot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i\tilde{\boldsymbol{v}}_i^{\mathrm{T}}\big]^{-1}\big[\frac{1}{N}\sum_{i=1}^N \dot{\ell}_{i,\eta_i^*}(y_i; \eta_i^*)\tilde{\boldsymbol{v}}_i\big] + o_p(N^{-1/2}).$$

Therefore, by Central Limit Theory and Slutsky's Theorem, we have

$$\sqrt{N}(\hat{\boldsymbol{b}} - \boldsymbol{b}^*) \to \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1}),$$

where $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are defined in Lemma 2. $\square$

Proof of Theorem 2.3. Similar to [2], for any $a \in [1, 2]$, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\beta}}_a(t) - \boldsymbol{\beta}_a^*\| &= \Big\| \sum_{l=1}^{K_x} \hat{e}_{al}\hat{\phi}_{al} - \sum_{l=1}^{\infty} e_{al}^*\phi_{al} \Big\|^2 \\
&\leq 2\Big\| \sum_{l=1}^{K_x} \hat{e}_{al}\hat{\phi}_{al} - \sum_{l=1}^{K_x} e_{al}^*\phi_{al} \Big\|^2 + 2\Big\| \sum_{l=K_x+1}^{\infty} e_{al}^*\phi_{al} \Big\|^2 \\
&\leq 4\Big\| \sum_{l=1}^{K_x} (\hat{e}_{al} - e_{al}^*)\hat{\phi}_{al} \Big\|^2 + 4\Big\| \sum_{l=1}^{K_x} e_{al}^*(\hat{\phi}_{al} - \phi_{al}) \Big\|^2 + 2\sum_{l=K_x+1}^{\infty} e_{al}^{*2} \\
&\leq 4\|\hat{\boldsymbol{e}}_a - \boldsymbol{e}_a^*\|^2 + 8K_x \sum_{l=1}^{K_x} e_{al}^{*2}\|\hat{\phi}_{al} - \phi_{al}\|^2 + 2\sum_{l=K_x+1}^{\infty} e_{al}^{*2}.
\end{aligned}$$

Note that

$$\sum_{l=K_x+1}^{\infty} e_{al}^{*2} \leq C \sum_{l=K_x+1}^{\infty} l^{-2\gamma} = O(K_x^{-(2\gamma-1)}) = O(N^{-(2\gamma-1)/(\alpha+2\gamma)}), \tag{S3}$$

and by $\|\hat{\phi}_{al} - \phi_{al}\|^2 = O_p(N^{-1}l^2)$, we have

$$8K_x \sum_{l=1}^{K_x} e_{al}^{*2}\|\hat{\phi}_{aj} - \phi_{aj}\|^2 \leq O_p(N^{-1}K_x) = o_p(N^{-(2\gamma-1)/(\alpha+2\gamma)}). \tag{S4}$$

In addition, from Lemma 3 we have $\|\hat{\boldsymbol{e}} - \boldsymbol{e}^*\|^2 = O_p(N^{-(2\gamma-1)/(\alpha+2\gamma)})$, then Theorem 2.3 holds by combining this with equations (S3) and (S4). $\square$

Proof of Corollary 2.4. Let $\hat{\eta}_i = \boldsymbol{v}_i^{\mathrm{T}}\hat{\boldsymbol{b}} + \int_0^1 \tilde{\boldsymbol{x}}_i(t)\hat{\boldsymbol{\beta}}(t)dt$, where $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\beta}}(t)$ are obtained from our proposed estimation procedure. Then, we have

$$\begin{aligned}
\hat{\eta}_i - \eta_i^* &= \boldsymbol{v}_i^{\mathrm{T}}\hat{\boldsymbol{b}} + \int_0^1 \tilde{\boldsymbol{x}}_i(t)\hat{\boldsymbol{\beta}}(t)dt - \big[\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{b}^* + \int_0^1 \tilde{\boldsymbol{x}}_i(t)\boldsymbol{\beta}^*(t)dt\big] \\
&= \boldsymbol{v}_i^{\mathrm{T}}(\hat{\boldsymbol{b}} - \boldsymbol{b}^*) + \sum_{a=1}^{2}\sum_{l=1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai}, \phi_{al}\rangle\langle\hat{\boldsymbol{\beta}}_a, \hat{\phi}_{al}\rangle - \sum_{a=1}^{2}\sum_{l=1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai}, \phi_{al}\rangle\langle\boldsymbol{\beta}_a^*, \phi_{al}\rangle \\
&= \boldsymbol{v}_i^{\mathrm{T}}(\hat{\boldsymbol{b}} - \boldsymbol{b}^*) + \sum_{a=1}^{2}\sum_{l=1}^{\infty}\langle\tilde{\boldsymbol{x}}_{ai}, \phi_{al}\rangle(\hat{e}_{al} - e_{al}^*) \\
&= I_5 + I_6,
\end{aligned}$$

where $\hat{e}_{al} = \langle \hat{\boldsymbol{\beta}}_a, \hat{\phi}_{al} \rangle$, $e^*_{al} = \langle \boldsymbol{\beta}^*_a, \phi_{al} \rangle$, $I_5 = \boldsymbol{v}_i^{\mathrm{T}}(\hat{\boldsymbol{b}} - \boldsymbol{b}^*)$ and
$I_6 = \sum_{a=1}^{2}\sum_{l=1}^{\infty} \langle \tilde{\boldsymbol{x}}_{ai}, \phi_{al} \rangle (\hat{e}_{al} - e^*_{al}) = \sum_{a=1}^{2}\sum_{l=1}^{\infty} p_{ail}(\hat{e}_{al} - e^*_{al})$.

Given Theorem 2.2, the fact that $\|\hat{\boldsymbol{b}} - \boldsymbol{b}^*\| = O_p(N^{-1/2})$ holds. Then under Assumption (A4) we have $\mathrm{E}\|I_5\|^2 = \|\hat{\boldsymbol{b}} - \boldsymbol{b}^*\|^2 \mathrm{E}\boldsymbol{v}_i^2 = O_p(N^{-1})$.

Since $\langle \tilde{\boldsymbol{x}}_{ai}, \phi_{al} \rangle$ are uncorrelated random variables with zero mean and variance λ_{al}, and by Lemma 3 we have $\|\hat{\boldsymbol{e}} - \boldsymbol{e}^*\|^2 = O_p(N^{-(2\gamma-1)/(\alpha+2\gamma)})$. Then under Assumption (A2), we have

$$\mathrm{E}(\|I_6\|^2) = \sum_{a=1}^{2}\sum_{l=1}^{\infty} \lambda_{al}\|\hat{e}_{al} - e^*_{al}\|^2 \leq 2O_p(N^{\frac{2\gamma-1}{\alpha+2\gamma}})C\sum_{l=1}^{\infty} l^{-\alpha} = O_p(N^{\frac{2\gamma-1}{\alpha+2\gamma}}).$$

Thus

$$\hat{\eta}_i - \eta^* = I_5 + I_6 = O_p(N^{-1/2}) + O_p(N^{-(2\gamma-1)/(\alpha+2\gamma)}) = O_p(N^{-1/2}).$$

Note that the functional logistic regression model is a special case of model (11) and (12) with a logistic link function, i.e. $\eta_i = \mathrm{logit}(\pi_{\mathrm{i}})$. Since the inverse link function $h^{-1}(\eta_i)$ is continuous and differentiable in η_i, then $h^{-1}(\hat{\eta}_i) - h^{-1}(\eta^*_i) = O_p(N^{-1/2})$ holds, which indicate that $\hat{\pi}_i - \pi^*_i = O_p(N^{-1/2})$. □

Proof of Theorem 2.5. $\hat{\boldsymbol{\theta}}_x$ is the MLE of the second level model obtained through conditional models described in Section 2.2, then under Assumptions (A7)-(A9), the theorem follows from [10] immediately, to save space, we omit the proof here. □

References

[1] L. Ting and Z. Zhu, *Inference for generalized partial functional linear regression*, Statist. Sinica. 30(2020), pp. 1379–1397.

[2] H. Shin, *Partial functional linear regression*, J. Statist. Plann. Inference 139(2009), pp. 3405–3418.

[3] T. T. Cai, P. Hall and others, *Prediction in functional linear regression*, Ann. Statist. 34(2006), pp. 2159–2179.

[4] P. Hall, J. L. Horowitz and others, *Methodology and convergence rates for functional linear regression*, Ann. Statist. 35(2007), pp. 70–91.

[5] A. W. Van der Vaart, *Asymptotic Statistics*, Cambridge university press, Cambridge, 2000.

[6] C. J. Geyer, *On the convergence of Monte Carlo maximum likelihood calculations*, J. R. Stat. Soc. Ser. B. Stat. Methodol. 56 (1994), pp. 261–274.

[7] D. Pollard, *Asymptotics for least absolute deviation regression estimators*, Econometric Theory 7 (1991), pp. 186–199.

[8] R. Cao, J. Du, J. Zhou and T. Xie, *FPCA–based estimation for generalized functional partially linear models*, Statist. Papers 61 (2020), pp. 2715–2735.

[9] L. Wang, X. Liu, H. Liang and R.J. Carroll, *Estimation and variable selection for generalized additive partial linear models*, Ann. Statist. 39 (2011), pp. 1827–1851.

[10] H. White, *Estimation, inference and specification analysis*, Cambridge university press, Cambridge, 1996.