NONPARAMETRIC CLUSTER ANALYSIS ON MULTIPLE OUTCOMES OF LONGITUDINAL DATA

Yang Lv[1], Xiaolu Zhu[2], Zhongyi Zhu[3] and Annie Qu[4]

[1] *School of Statistics, Capital University of Economics and Business, Beijing, China*

[2] *Amazon.com Inc., Seattle, Washington, U.S.*

[3] *Department of Statistics, School of Management, Fudan University, Shanghai, China*

[4] *Department of Statistics, University of Illinois at Urbana-Champaign*

Supplementary Material

In this supplement, we provide simulation results under more settings, and give the technical proofs for Lemma 1-2, Theorem 1 and Corollary 1.

S1 Other Simulations

In this section, we let the random error $\boldsymbol{\varepsilon_{im}} = (\varepsilon_{i1m}, \ldots, \varepsilon_{i10m})^T$ follow non-normal distributions such as the exponential distribution, t-distribution, or mixture distribution in the following Case 3–5:

Case 3 : $\boldsymbol{\varepsilon_{im}} = exp(\xi_{im}) - 1$, where $\xi_{im} \sim N(0, 0.25R)$, and the correlation matrix R is the same as in Case 2.

Case 4 : $\boldsymbol{\varepsilon_{im}} \sim t_3(0, 0.25R)$, where the correlation R is the same as in Case 2.

Case 5 : $\boldsymbol{\varepsilon_{i1}} \sim N(0, 0.25R)$ and $\boldsymbol{\varepsilon_{i2}} \sim t_3(0, 0.04R)$, where the correlation R is the same as in Case 2.

The simulation results based on 100 simulation runs are provided in the following Tables S1–S3, which show that the proposed method is robust against any parametric assumption and performs the best compared to other methods.

S2 Proof of Lemma 1

Note that we can individually obtain $\tilde{\boldsymbol{\beta}}_{\boldsymbol{i}}$ and $\tilde{b}_i$ as

$$\tilde{\boldsymbol{\beta}}_{\boldsymbol{i}} = (\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i} + \lambda_1\mathbf{D_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}\boldsymbol{y_i}, \tag{S2.1}$$

$$\tilde{b}_i = (\mathbf{1}_{\boldsymbol{n_i}}^T\mathbf{\Sigma_i}^{-1}\mathbf{1}_{\boldsymbol{n_i}} + \lambda_2)^{-1}\mathbf{1}_{\boldsymbol{n_i}}^T\mathbf{\Sigma_i}^{-1}(\boldsymbol{y_i} - \mathbf{B_i}\tilde{\boldsymbol{\beta}}_{\boldsymbol{i}}), \tag{S2.2}$$

where $\mathbf{W_i} = (\mathbf{\Sigma_i} + \frac{1}{\lambda_2}\mathbf{1}_{\boldsymbol{n_i}}\mathbf{1}_{\boldsymbol{n_i}}^T)^{-1}$. Then we have

$$\begin{aligned}\tilde{\mathbf{f}}_{\mathbf{i}} &= \mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i} + \lambda_1\mathbf{D_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}\boldsymbol{y_i} \qquad \text{(S2.3)}\\ &= \mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}\mathbf{f_i} - \mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i} + \lambda_1\mathbf{D_i})^{-1}\lambda_1\mathbf{D_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i})^{-1}\\ &\quad \mathbf{B_i}^T\mathbf{W_i}\mathbf{f_i} + \mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i} + \lambda_1\mathbf{D_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}(\mathbf{1}_{\mathbf{n_i}}b_i + \boldsymbol{\varepsilon_i}).\end{aligned}$$

When $\lambda_1 = 0$ in (S2.3), it refers to a regression spline estimator. More precisely, the regression spline estimator $\tilde{\mathbf{f}}_{\mathbf{i}}^{\mathbf{reg}} = \mathbf{B_i}\tilde{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\boldsymbol{reg}}$ is the minimizer of

$$(\boldsymbol{y_i} - \mathbf{B_i}\tilde{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\boldsymbol{reg}})^T\mathbf{W_i}(\boldsymbol{y_i} - \mathbf{B_i}\tilde{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\boldsymbol{reg}}), \tag{S2.4}$$

thus $E(\tilde{\mathbf{f}}_{\mathbf{i}}^{\mathbf{reg}}) = \mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}\mathbf{f_i}$. Furthermore, we denote $\mathbf{f}_{\mathbf{i}}^{\mathbf{s}} = \mathbf{B_i}\boldsymbol{\beta_i}$ as the best L_∞ approximation to $\mathbf{f_i}$, where $\boldsymbol{\beta_i} \in \mathcal{R}^p$. Thus, we obtain

$$\begin{aligned}||\tilde{\mathbf{f}}_{\mathbf{i}} - \mathbf{f_i}||_{n_i}^2 &\leq ||E(\tilde{\mathbf{f}}_{\mathbf{i}}^{\mathbf{reg}}) - \mathbf{f_i}||_{n_i}^2 + ||\mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i} + \lambda_1\mathbf{D_i})^{-1}\lambda_1\mathbf{D_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i})^{-1}\\ &\quad \mathbf{B_i}^T\mathbf{W_i}\mathbf{f_i}||_{n_i}^2 + ||\mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i} + \lambda_1\mathbf{D_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}(\mathbf{1}_{\mathbf{n_i}}b_i + \boldsymbol{\varepsilon_i})||_{n_i}^2\\ &= I_1 + I_2 + I_3,\end{aligned}$$

where the definitions of $I_j, j = 1, 2, 3$, should be apparent from the context. Under Assumption A1-A3, when $k \to \infty$ and $k^4 = o(n_0)$, then from Theorem 1 in Zhu, Fung, and He (2008), it is straightforward to show that $I_1 = O_p(h^{2r})$.

Furthermore, let $\mathbf{F_i} = \frac{1}{n_i}\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i}$, $\mathbf{H_i} = \mathbf{F_i} + \frac{\lambda_1}{n_i}\mathbf{D_i}$. From Zhu, Fung, and He (2008) Lemma A1, we have $||\mathbf{F_i}^{-1}||_\infty = O(h^{-1})$. Similarly to Lemma 6.2 in Cardot (2000), it follows that $||\mathbf{D_i}||_\infty = O(h^{1-2d})$. Since $\frac{\lambda_1 h^{-2d}}{n_0} = o(1)$, then $||\mathbf{H_i}^{-1}||_\infty = ||\mathbf{F_i}^{-1} - (\mathbf{I_p} + \mathbf{F_i}^{-1}\frac{\lambda_1}{n_i}\mathbf{D_i})^{-1}\mathbf{F_i}^{-1}\frac{\lambda_1}{n_i}\mathbf{D_i}\mathbf{F_i}^{-1}||_\infty = O(h^{-1})$. We can write $I_2 = ||\mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i}+\lambda_1\mathbf{D_i})^{-1}\lambda_1\mathbf{D_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}(\mathbf{f_i}-\mathbf{f_i^s}+\mathbf{f_i^s})||^2_{n_i} = ||\frac{\lambda_1}{n_i}\mathbf{B_i}\mathbf{H_i}^{-1}\mathbf{D_i}\mathbf{F_i}^{-1}\frac{1}{n_i}\mathbf{B_i}^T\mathbf{W_i}(\mathbf{f_i}-\mathbf{f_i^s})+\frac{\lambda_1}{n_i}\mathbf{B_i}\mathbf{H_i}^{-1}\mathbf{D_i}\mathbf{F_i}^{-1}\mathbf{B_i}^T\mathbf{W_i}\mathbf{f_i^s}/n_i||^2_{n_i}$. According to the proof of Theorem 1 in Chen and Wang (2011), we can show that

$$-\frac{\lambda_1}{n_i}||\mathbf{B_i}\mathbf{H_i}^{-1}\mathbf{D_i}\mathbf{F_i}^{-1}\mathbf{B_i}^T\mathbf{W_i}\mathbf{f_i^s}/n_i||_\infty = O(\lambda_1 n_i^{-1}h^{-d}),$$
$$-\frac{\lambda_1}{n_i}||\mathbf{B_i}\mathbf{H_i}^{-1}\mathbf{D_i}\mathbf{F_i}^{-1}\frac{1}{n_i}\mathbf{B_i}^T\mathbf{W_i}(\mathbf{f_i}-\mathbf{f_i^s})||_\infty = o(\lambda_1 n_i^{-1}h^{r-2d}).$$

Then $I_2 = O_p(\frac{\lambda_1^2}{n_i^2}h^{-2d})$. Next, it can be shown that

$$\begin{aligned}
&E[\mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i}+\lambda_1\mathbf{D_i})^{-1}\mathbf{B_i}^T\mathbf{W_i}(\mathbf{1_{n_i}}b_i+\boldsymbol{\varepsilon_i})]^T[\mathbf{B_i}(\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i}+\lambda_1\mathbf{D_i})^{-1} \\
&\mathbf{B_i}^T\mathbf{W_i}(\mathbf{1_{n_i}}b_i+\boldsymbol{\varepsilon_i})] \\
&\le \lambda_{max}(\mathbf{W_i}(\sigma_b^2\mathbf{1}_{\boldsymbol{n_i}}\mathbf{1}_{\boldsymbol{n_i}}^T+\mathbf{\Sigma_i^0}))tr\{\mathbf{B_i}^T\mathbf{W_i}\mathbf{B_i}n_i^{-1}\mathbf{H_i}^{-1}\mathbf{B_i}^T\mathbf{B_i}n_i^{-1}\mathbf{H_i}^{-1}\} \\
&\le \lambda_{max}(\mathbf{W_i}(\sigma_b^2\mathbf{1}_{\boldsymbol{n_i}}\mathbf{1}_{\boldsymbol{n_i}}^T+\mathbf{\Sigma_i^0}))\lambda_{max}(\mathbf{F_i}\mathbf{H_i}^{-1})\lambda_{max}(\mathbf{B_i}^T\mathbf{B_i})tr\{n_i^{-1}\mathbf{H_i}^{-1}\} \\
&= O(n_i^{-1}h^{-1}).
\end{aligned}$$

The above inequalities are obtained from Zhou, Shen, and Wolfe (1998) Lemma 6.5. Since $\lambda_{max}(\mathbf{W_i}(\sigma_b^2\mathbf{1}_{\boldsymbol{n_i}}\mathbf{1}_{\boldsymbol{n_i}}^T+\mathbf{\Sigma_i^0})) < C$, and for any $x \in [0,1]$, $0 \le \pi_{im}(x) \le 1$, thus $tr\{\mathbf{B_i}^T\mathbf{B_i}\} \le 1$ and $\lambda_{max}(\mathbf{B_i}^T\mathbf{B_i}) \le 1$. Then $I_3 = O_p(n_i^{-1}h^{-1})$. Consequently, we can show that for all $i = 1, \cdots, n$,

$$\begin{aligned}
||\tilde{\mathbf{f}}_\mathbf{i} - \mathbf{f_i}||^2_{n_i} &\le O_p(k^{-2r}) + O_p(\frac{\lambda_1^2}{n_i^2}k^{2d}) + O_p(\frac{k}{n_i}) \\
&\le O_p(k^{-2r}) + O_p(\frac{\lambda_1^2}{n_0^2}k^{2d}) + O_p(\frac{k}{n_0}).
\end{aligned}$$

Then, it follows that

$$||\tilde{\mathbf{f}} - \mathbf{f}||^2_N = \frac{1}{N}\sum_{i=1}^n(\tilde{\mathbf{f}}_\mathbf{i} - \mathbf{f_i})^T(\tilde{\mathbf{f}}_\mathbf{i} - \mathbf{f_i}) \le O_p(k^{-2r}) + O_p(\frac{\lambda_1^2}{n_0^2}k^{2d}) + O_p(\frac{k}{n_0}).$$

S3 Proof of Lemma 2

When the true group memberships $\mathcal{G}_1, \cdots, \mathcal{G}_G$ are known, the oracle approximation for $\mathbf{f}$ conditional on the oracle estimate of random effects $\tilde{\boldsymbol{b}}^{\boldsymbol{or}}$ is denoted as $\tilde{\mathbf{f}}^{\mathbf{or}} = \mathbf{B}\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}$, where

$$\begin{aligned}\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}_{\boldsymbol{(g)}} &= \arg\min_{i\in\mathcal{G}_g} Q_{(g)}(\boldsymbol{\beta}_{\boldsymbol{(g)}}, \boldsymbol{b}) \\ &= \arg\min_{i\in\mathcal{G}_g} \frac{1}{2}\sum_{i=1}^{|\mathcal{G}_g|}\{(\boldsymbol{y_i} - \mathbf{B_i}\boldsymbol{\beta}_{\boldsymbol{(g)}} - \mathbf{1}_{\mathbf{n_i}} b_i)^T \boldsymbol{\Sigma}_{\mathbf{i}}^{\mathbf{-1}}(\boldsymbol{y_i} - \mathbf{B_i}\boldsymbol{\beta}_{\boldsymbol{(g)}} - \mathbf{1}_{\mathbf{n_i}} b_i) + \\ &\quad \lambda_1 \boldsymbol{\beta}^T_{\boldsymbol{(g)}}\mathbf{D_i}\boldsymbol{\beta}_{\boldsymbol{(g)}} + \lambda_2 b_i^2\}.\end{aligned}$$

Similarly to the proof of Lemma 1, we can show under Assumptions $A1 - A5$ and $\frac{\lambda_1 h^{-2d}}{N_0} = o(1)$ that

$$||\tilde{\mathbf{f}}^{\mathbf{or}} - \mathbf{f}||_N^2 \leq O_p(k^{-2r}) + O_p(\frac{\lambda_1^2}{N_0^2}k^{2d}) + O_p(\frac{k}{N_0}).$$

S4 Proof of Theorem 1

By the triangular inequality, $\|\hat{\mathbf{f}} - \mathbf{f}\|_N^2 = \|\hat{\mathbf{f}} - \tilde{\mathbf{f}}^{\mathbf{or}} + \tilde{\mathbf{f}}^{\mathbf{or}} - \mathbf{f}\|_N^2 \leq \|\hat{\mathbf{f}} - \tilde{\mathbf{f}}^{\mathbf{or}}\|_N^2 + \|\tilde{\mathbf{f}}^{\mathbf{or}} - \mathbf{f}\|_N^2$.

The proposed objective function is

$$\begin{aligned}H(\boldsymbol{\beta}, \boldsymbol{b}) &= \frac{1}{2}(\boldsymbol{Y} - \mathbf{B}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \mathbf{B}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b}) + \frac{1}{2}\lambda_1\boldsymbol{\beta}^T\mathbf{D_d}\boldsymbol{\beta} + \frac{1}{2}\lambda_2||\boldsymbol{b}||_2^2 + \\ &\quad \sum_{i,j\in\mathcal{L}}\rho(|\boldsymbol{\beta_i} - \boldsymbol{\beta_j}|, \lambda_3). \qquad \text{(S4.1)}\end{aligned}$$

We can obtain the estimate of random effects $\boldsymbol{b}$ by minimizing (S4.1), then

$$\hat{\boldsymbol{b}} = (\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \lambda_2\mathbf{I_n})^{-1}\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}\boldsymbol{\beta}).$$

Replacing $\boldsymbol{b}$ in (S4.1) by $\hat{\boldsymbol{b}}$, we can obtain the profiled objective function

$$\begin{aligned}H_n(\boldsymbol{\beta}) &= \frac{1}{2}(\boldsymbol{Y} - \mathbf{B}\boldsymbol{\beta})^T\mathbf{W}(\boldsymbol{Y} - \mathbf{B}\boldsymbol{\beta}) + \frac{1}{2}\lambda_1\boldsymbol{\beta}^T\mathbf{D_d}\boldsymbol{\beta} + \sum_{i,j\in\mathcal{L}}\rho(|\boldsymbol{\beta_i} - \boldsymbol{\beta_j}|, \lambda_3) \\ &= Q_n(\boldsymbol{\beta}) + S_{\lambda_3}(\boldsymbol{\beta}), \qquad \text{(S4.2)}\end{aligned}$$

where $\mathbf{W} = (\mathbf{\Sigma} + \frac{1}{\lambda_2}\mathbf{Z}\mathbf{Z}^T)^{-1}$, $\mathbf{W_i} = (\mathbf{\Sigma_i} + \frac{1}{\lambda_2}\mathbf{1}_{\boldsymbol{n_i}}\mathbf{1}_{\boldsymbol{n_i}}^T)^{-1}$, $Q_n(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^n(\boldsymbol{y_i} - \mathbf{B_i}\boldsymbol{\beta_i})^T\mathbf{W_i}(\boldsymbol{y_i} - \mathbf{B_i}\boldsymbol{\beta_i}) + \frac{1}{2}\sum_{i=1}^n \lambda_1\boldsymbol{\beta_i}^T\mathbf{D_i}\boldsymbol{\beta_i}$ and $S_{\lambda_3}(\boldsymbol{\beta}) = \sum_{i,j\in\mathcal{L}}\rho(|\boldsymbol{\beta_i} - \boldsymbol{\beta_j}|, \lambda_3)$. Denote

$$\begin{aligned} \mathbf{D_n^s} &= \partial^2 Q_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta} = \mathbf{B}^T\mathbf{W}\mathbf{B} + \lambda_1\mathbf{D_d}, && \text{(S4.3)} \\ \mathbf{M_n^s} &= Cov(\partial Q_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}) \\ &= \sum_{i=1}^n \mathbf{B_i}^T\mathbf{W_i}(\sigma_b^2\mathbf{1}_{\boldsymbol{n_i}}\mathbf{1}_{\boldsymbol{n_i}}^T + \mathbf{\Sigma_i^0})\mathbf{W_i}^T\mathbf{B_i}. && \text{(S4.4)} \end{aligned}$$

Let $\tau_n^s = \lambda_{min}(\mathbf{D_n^s}(\mathbf{M_n^s})^{-1}\mathbf{D_n^s})$, and $\mathbf{B}\boldsymbol{\beta^*} = \mathbf{B}\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}} + (\tau_n^s)^{-\frac{1}{2}}\boldsymbol{u}$, where $||\boldsymbol{u}||_N = d_u$.

Note that

$$\begin{aligned} (\tau_n^s)^{-1} &= \lambda_{max}((\mathbf{D_n^s})^{-1}(\mathbf{M_n^s})(\mathbf{D_n^s})^{-1}) \\ &\leq C_1\lambda_{max}\{E[\mathbf{B}(\mathbf{B}^T\mathbf{W}\mathbf{B} + \lambda_1\mathbf{D_d})^{-1}\mathbf{B}^T\mathbf{W}\mathbf{Y} - E(\mathbf{B}\tilde{\boldsymbol{\beta}})]^T[\mathbf{B}(\mathbf{B}^T\mathbf{W}\mathbf{B} \\ &\quad +\lambda_1\mathbf{D_d})^{-1}\mathbf{B}^T\mathbf{W}\mathbf{Y} - E(\mathbf{B}\tilde{\boldsymbol{\beta}})]\} \\ &= C_1E(\tilde{\mathbf{f}} - \mathbf{f^s})^T(\tilde{\mathbf{f}} - \mathbf{f^s}) \\ &\leq C_1N||\tilde{\mathbf{f}} - \mathbf{f}||_N^2 + C_1N||\mathbf{f} - \mathbf{f^s}||_N^2, && \text{(S4.5)} \end{aligned}$$

where the first inequality in (S4.5) can be derived from Lemma A.4 in Zhu and Qu (2018). Thus, from Lemma 1, $(\tau_n^s)^{-1} \leq A_1(\frac{k}{n_0} + \frac{\lambda_1^2}{n_0^2}k^{2d} + k^{-2r})$ with A_1 sufficiently large.

If N is sufficiently large, we have

$$\begin{aligned} \|\mathbf{B}(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}_{\boldsymbol{(g)}} - \tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}_{\boldsymbol{(g')}})\|_N &= \|\tilde{\mathbf{f}}^{\mathbf{or}}_{\mathbf{(g)}} - \tilde{\mathbf{f}}^{\mathbf{or}}_{\mathbf{(g')}}\|_N \\ &= \|\mathbf{f_{(g)}} - \mathbf{f_{(g')}} - (\mathbf{f_{(g)}} - \tilde{\mathbf{f}}^{\mathbf{or}}_{\mathbf{(g)}}) + (\mathbf{f_{(g')}} - \tilde{\mathbf{f}}^{\mathbf{or}}_{\mathbf{(g')}})\|_N \\ &\geq \|\mathbf{f_{(g)}} - \mathbf{f_{(g')}}\|_N - \|\mathbf{f_{(g)}} - \tilde{\mathbf{f}}^{\mathbf{or}}_{\mathbf{(g)}}\|_N - \|\mathbf{f_{(g')}} - \tilde{\mathbf{f}}^{\mathbf{or}}_{\mathbf{(g')}}\|_N \\ &\geq d_f. \end{aligned}$$

It is easy to prove that there exists a constant c, such that $\|\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}_{\boldsymbol{(g)}} - \tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}_{\boldsymbol{(g')}}\|_n \geq cd_f$. From the definition of $\rho_\tau(t, \lambda_3) = 0$, when $t = 0$ and $\rho_\tau(t, \lambda_3) \geq 0$, when $t \neq 0$, and since the minimum distance d_f satisfies $cd_f \geq \tau\lambda_3$, we have $S_{\lambda_3}(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}) = 0$. By Taylor's expansion, we can obtain that

$$\begin{aligned} L_n(u) &= H_n(\boldsymbol{\beta}^*) - H_n(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}) \\ &= Q_n(\boldsymbol{\beta}^*) - Q_n(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}) + S_{\lambda_3}(\boldsymbol{\beta}^*) \\ &= (\tau_n^s)^{-\frac{1}{2}}\dot{Q}_n^T(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}})\boldsymbol{u} + \frac{1}{2}(\tau_n^s)^{-1}\boldsymbol{u}^T\ddot{Q}_n(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}})\boldsymbol{u} + S_{\lambda_3}(\boldsymbol{\beta}^*) \\ &= (\tau_n^s)^{-\frac{1}{2}}\dot{Q}_n^T(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}})\boldsymbol{u} + \frac{1}{2}(\tau_n^s)^{-1}\boldsymbol{u}^T\mathbf{D_n^s}\boldsymbol{u} + S_{\lambda_3}(\boldsymbol{\beta}^*), \end{aligned} \tag{S4.6}$$

where $\dot{Q}_n(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}) = \frac{\partial}{\partial\boldsymbol{\beta}}Q(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}}$, $\ddot{Q}_n(\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}) = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}}Q(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}^{\boldsymbol{or}}}$. Note that

$$\begin{aligned} (\mathbf{M_n^s})^{\frac{1}{2}} &= (\mathbf{D_n^s})^{\frac{1}{2}}(\mathbf{D_n^s})^{-\frac{1}{2}}(\mathbf{M_n^s})^{\frac{1}{2}}(\mathbf{D_n^s})^{-\frac{1}{2}}(\mathbf{D_n^s})^{\frac{1}{2}} \\ &\leq (\mathbf{D_n}^s)^{\frac{1}{2}}\lambda_{max}((\mathbf{D_n^s})^{-\frac{1}{2}}(\mathbf{M_n^s})^{\frac{1}{2}}(\mathbf{D_n^s})^{-\frac{1}{2}})(\mathbf{D_n^s})^{\frac{1}{2}} \\ &= \lambda_{min}((\mathbf{D_n^s})^{\frac{1}{2}}(\mathbf{M_n^s})^{-\frac{1}{2}}(\mathbf{D_n^s})^{\frac{1}{2}})^{-1}\mathbf{D_n^s} \\ &= (\tau_n^s)^{-\frac{1}{2}}\mathbf{D_n^s}, \end{aligned} \tag{S4.7}$$

and thus $(\tau_n^s)^{-\frac{1}{2}}(\mathbf{M_n^s})^{\frac{1}{2}} \leq (\tau_n^s)^{-1}\mathbf{D_n^s}$. Consequently, if d_u is sufficiently large, then the second term in $L_n(u)$ dominates the first term, which implies that, with probability tending to 1, $L_n(u) > 0$ at $||\boldsymbol{u}||_N = d_u$. Hence we have

$$P\{\inf_{||\boldsymbol{u}||_N=d_u} L_n(u) > 0\} \to 1,$$

which entails that with probability tending to 1, there exists a local minimum of $H_n(\boldsymbol{\beta})$ which lies in the ball $\boldsymbol{\mathcal{B}} = \{\boldsymbol{\beta} : \|\mathbf{B}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}^{\boldsymbol{or}})\|_N^2 = (\tau_n^s)^{-1}d_u^2 = A_2(\frac{k}{n_0} + \frac{\lambda_1^2}{n_0^2}k^{2d} + k^{-2r})\}$ with A_2 sufficiently large. And $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} H_n(\boldsymbol{\beta})$, then $\|\mathbf{B}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}^{\boldsymbol{or}})\|_n^2 = O_p(\frac{k}{n_0} + \frac{\lambda_1^2}{n_0^2}k^{2d} + k^{-2r})$. Thus by combining the result from Lemma 2, we can complete the proof.

S5 Proof of Corollary 1

From Theorem 1, there exists a local minimizer $\hat{\mathbf{f}} = \mathbf{B}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} \in \mathcal{B}$. For any pair i, j such that $G(i) = G(j)$, we have $\|\hat{\mathbf{f}}_\mathbf{i} - \hat{\mathbf{f}}_\mathbf{j}\|_N^2 = \|\hat{\mathbf{f}}_\mathbf{i} - \mathbf{f}_\mathbf{i} + \mathbf{f}_\mathbf{i} - \mathbf{f}_\mathbf{j} + \mathbf{f}_\mathbf{j} - \hat{\mathbf{f}}_\mathbf{j}\|_N^2 \leq 2 \max_i \|\hat{\mathbf{f}}_\mathbf{i} - \mathbf{f}_\mathbf{i}\|_N^2 + \|\mathbf{f}_\mathbf{i} - \mathbf{f}_\mathbf{j}\|_N^2 \leq O_p(\frac{k}{n_0} + \frac{\lambda_1^2}{n_0^2} k^{2d} + k^{-2r}) \to O_p(k^{-2r})$, as $n_0 \to \infty$. Thus $\hat{\mathbf{f}}_\mathbf{i}$ and $\hat{\mathbf{f}}_\mathbf{j}$ will be in the same group with probability tending to 1. On the other side, for any pair i, j such that $G(i) \neq G(j)$, we have $\|\hat{\mathbf{f}}_\mathbf{i} - \hat{\mathbf{f}}_\mathbf{j}\|_N^2 \geq \min \|\mathbf{f}_\mathbf{i} - \mathbf{f}_\mathbf{j}\|_N^2 - 2 \max_i \|\hat{\mathbf{f}}_\mathbf{i} - \mathbf{f}_\mathbf{i}\|_N^2 \geq d_f^2 - O_p(k^{-2r})$ as $n_0 \to \infty$, which indicates that $\hat{\mathbf{f}}_\mathbf{i}$ and $\hat{\mathbf{f}}_\mathbf{j}$ will be in different groups with probability tending to 1. Hence, $P(\hat{\mathcal{G}} = \mathcal{G}) \to 1$.

References

Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics 12*, pp. 503-538.

Chen, H. H. and Wang Y. J. (2011). A penalized spline approach to functional mixed effects model analysis. *Biometrics 67*, pp. 861-870.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics 26*, pp. 1760-1782.

Zhu, X. L. and Qu, A. (2018). Cluster analysis of longitudinal profiles with subgroups. *Electronic Journal of Statistics 12*, pp. 171-193.

Zhu, Z. Y., Fung, W. K. and He, X. M. (2008). On the asymptotics of marginal regression splines with longitudinal data. *Biometrika 95*, pp. 907-917.

Table S1: Case3: Comparison results from the proposed nonparametric pairwise-grouping with three different working correlation structures (NPGr-IN, NPGr-AR(1), NPGr-Ex), Gaussian Mixtures (bGM), K-means (bKmeans), SSClust, MixedEffects, and Kernel for balanced data.

	Methods	$\widehat{K}$	Rand	aRand	Jaccard	AMSE
AR(1)	NPGr-IN	3.02	0.9998	0.9995	0.9993	0.0420
	NPGr-AR(1)	3.00	1.0000	1.0000	1.0000	0.0375
	NPGr-Ex	3.00	1.0000	1.0000	1.0000	0.0380
	bGM	3.17	0.9941	0.9846	0.9819	0.0463
	bKmeans	3.00	0.9302	0.8606	0.8739	2.5818
	SSClust	7.98	0.8301	0.5504	0.4799	0.3549
	MixedEffects	4.68	0.9406	0.8555	0.8182	0.1199
	Kernel	4.24	0.9481	0.8743	0.8412	0.4984
Ex	NPGr-IN	3.10	0.9990	0.9978	0.9970	0.0476
	NPGr-AR(1)	3.00	1.0000	1.0000	1.0000	0.0369
	NPGr-Ex	3.00	1.0000	1.0000	1.0000	0.0372
	bGM	3.06	0.9986	0.9968	0.9957	0.0420
	bKmeans	3.00	0.9360	0.8721	0.8842	2.3726
	SSClust	7.69	0.8297	0.5483	0.4786	0.2954
	MixedEffects	4.22	0.9524	0.8849	0.8544	0.0958
	Kernel	4.37	0.9433	0.8621	0.8264	0.5371

Table S2: Case4: Comparison results from the proposed nonparametric pairwise-grouping with three different working correlation structures (NPGr-IN, NPGr-AR(1), NPGr-Ex), Gaussian Mixtures (bGM), K-means (bKmeans), SSClust, MixedEffects, and Kernel for balanced data.

	Methods	$\widehat{K}$	Rand	aRand	Jaccard	AMSE
AR(1)	NPGr-IN	3.70	0.9925	0.9826	0.9770	0.1601
	NPGr-AR(1)	3.02	0.9998	0.9995	0.9993	0.0635
	NPGr-Ex	3.03	0.9996	0.9992	0.9989	0.0679
	bGM	3.76	0.9788	0.9456	0.9351	0.1542
	bKmeans	3.00	0.9217	0.8438	0.8576	2.9920
	SSClust	6.42	0.8745	0.6782	0.6157	0.2549
	MixedEffects	5.18	0.9309	0.8323	0.7889	0.7447
	Kernel	5.29	0.9365	0.8454	0.8055	0.9188
Ex	NPGr-IN	3.71	0.9925	0.9826	0.9769	0.1529
	NPGr-AR(1)	3.06	0.9993	0.9984	0.9979	0.0800
	NPGr-Ex	3.03	0.9996	0.9992	0.9989	0.0651
	bGM	3.57	0.9876	0.9685	0.9619	0.1383
	bKmeans	3.00	0.9260	0.8522	0.8654	2.8116
	SSClust	6.74	0.8600	0.6377	0.5713	0.3023
	MixedEffects	5.18	0.9258	0.8186	0.7730	0.6701
	Kernel	5.08	0.9353	0.8423	0.8019	0.9801

Table S3: Case5: Comparison results from the proposed nonparametric pairwise-grouping with three different working correlation structures (NPGr-IN, NPGr-AR(1), NPGr-Ex), Gaussian Mixtures (bGM), K-means (bKmeans), SSClust, MixedEffects, and Kernel for balanced data.

	Methods	$\widehat{K}$	Rand	aRand	Jaccard	AMSE
AR(1)	NPGr-IN	3.00	1.0000	1.0000	1.0000	0.0329
	NPGr-AR(1)	3.00	1.0000	1.0000	1.0000	0.0327
	NPGr-Ex	3.00	1.0000	1.0000	1.0000	0.0328
	bGM	3.17	0.9965	0.9916	0.9892	0.0545
	bKmeans	3.00	0.9321	0.8639	0.8760	2.5680
	SSClust	8.40	0.8170	0.5093	0.4397	0.2825
	MixedEffects	4.68	0.9439	0.8632	0.8283	0.1128
	Kernel	3.94	0.9593	0.9015	0.8753	0.4525
Ex	NPGr-IN	3.00	1.0000	1.0000	1.0000	0.0334
	NPGr-AR(1)	3.00	1.0000	1.0000	1.0000	0.0333
	NPGr-Ex	3.00	1.0000	1.0000	1.0000	0.0333
	bGM	3.13	0.9977	0.9946	0.9930	0.0510
	bKmeans	3.02	0.9284	0.8561	0.8682	2.6681
	SSClust	8.26	0.8082	04837	0.4127	0.2573
	MixedEffects	3.92	0.9769	0.9456	0.9294	0.0644
	Kernel	4.01	0.9582	0.8992	0.8720	0.4672