Supporting Information for "Cluster Non-Gaussian Functional Data" by

Qingzhi Zhong[1], Huazhen Lin[1]*, Yi Li[2]

1. Center of Statistical Research and School of Statistics,

Southwestern University of Finance and Economics, Chengdu, China 611130

2. Department of Biostatistics,

University of Michigan, Ann Arbor, MI48109, USA

**email*: linhz@swufe.edu.cn

1 Web Appendix A: Notations

$$
\begin{aligned}
&U_n = \{\boldsymbol{\alpha}'\mathbf{B}_n(t) : \boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{q_n})' \in \mathcal{R}^{q_n}, \max_{1\le i\le q_n} |\alpha_i| \le c_0, t \in [0,1]\},\\
&\Phi_n = \{\boldsymbol{\beta}'\mathbf{B}_n(t) : \boldsymbol{\beta} = (\beta_1, \cdots, \beta_{q_n})' \in \mathcal{R}^{q_n}, \max_{1\le i\le q_n} |\beta_i| \le c_0, t \in [0,1]\},\\
&\boldsymbol{\Omega}_n^* = \{\boldsymbol{\Omega}_n = (\boldsymbol{\Lambda}', \boldsymbol{\sigma}^{2\prime}, \boldsymbol{\pi}', \boldsymbol{\mu}', \boldsymbol{\phi}')' \in R_+^{\sum_{g=1}^C K_g} \otimes R_+^C \otimes [0,1]^C \otimes U_n^C \otimes \Phi_n^{\sum_{g=1}^C K_g}\},\\
&d(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = \left(\|\boldsymbol{\Lambda}_1 - \boldsymbol{\Lambda}_2\|^2 + \|\boldsymbol{\sigma}_1^2 - \boldsymbol{\sigma}_2^2\|^2 + \|\boldsymbol{\pi}_1 - \boldsymbol{\pi}_2\|^2 + \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2^2 + \|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2\|_2^2\right)^{1/2},\\
&\sigma_{ij,g}(\boldsymbol{\Omega}_n) = \boldsymbol{\phi}_g(t_{ij})'\boldsymbol{\Lambda}_g\boldsymbol{\phi}_g(t_{ij}) + \sigma_g^2,\\
&G(\omega; y, \boldsymbol{\Omega}_n) = E\left[\sum_{g=1}^C \pi_{g0}\Phi\left\{\frac{H_0(y) - \mu_{g0}(t_{ij})}{\sqrt{\sigma_{ij,g}(\boldsymbol{\Omega}_0)}}\right\} - \pi_g\Phi\left\{\frac{\omega - \mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\boldsymbol{\Omega}_n)}}\right\}\right],\\
&G_n(\omega; y, \boldsymbol{\Omega}_n) = \frac{1}{\sum_{i=1}^C n_i}\sum_{i=1}^n\sum_{j=1}^{n_i}\left[I\left(Y_{ij} \le y\right) - \sum_{g=1}^C \pi_g\Phi\left\{\frac{\omega - \mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\boldsymbol{\Omega}_n)}}\right\}\right],\\
&l_i(\boldsymbol{\Omega}_n; H) = \log\left\{\sum_{g=1}^C \pi_g f_{gi}(\boldsymbol{\Omega}_n; H)\right\} - \lambda\sum_{g=1}^C \log\left\{\frac{\epsilon + \pi_g}{\epsilon}\right\}\\
&Pl(\boldsymbol{\Omega}_n; H) = El_i(\boldsymbol{\Omega}_n; H), P_n l(\boldsymbol{\Omega}_n; H) = \frac{1}{n}\sum_{i=1}^n l_i(\boldsymbol{\Omega}_n; H).
\end{aligned}
$$

Writing $\sqrt{n}(P_n - P)l(\boldsymbol{\Omega}_n; H)$ for the empirical process indexed by $l_i(\boldsymbol{\Omega}_n; H)$. $\widehat{H}_n(y; \boldsymbol{\Omega}_n)$ is the estimator of $H(y)$ given $\boldsymbol{\Omega}_n$ and is the solution of (3.17) with respect to $H(y)$.

2 Web Appendix B: Lemmas

In this section, we sketch the proofs of Theorems 1-3 of the paper. To prove Theorems 1-3, we will employ the theory of empirical processes, and some techniques commonly used in semiparametric literature. Define the class of functions $\mathcal{L}_n = \{l_i(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) : \boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*\}$. For any $\varepsilon > 0$, the $L_1(P_n)$ covering number $N(\varepsilon, \mathcal{L}_n, L_1(P_n))$ of $\mathcal{L}_n$ is the smallest value κ for which there exist $\{\boldsymbol{\Omega}_{n,j} \in \boldsymbol{\Omega}_n^*, j = 1, \cdots, \kappa\}$, such that for any $\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*$,

$$\min_{j\in\{1,\cdots,\kappa\}} \frac{1}{n}\sum_{i=1}^{n} |l_i(\mathbf{\Omega}_n; \widehat{H}_n(\cdot;\mathbf{\Omega}_n)) - l_i(\mathbf{\Omega}_{n,j}; \widehat{H}_n(\cdot;\mathbf{\Omega}_{n,j}))| \le \varepsilon.$$

We will define $N(\varepsilon, \mathcal{L}_n, L_1(P_n)) = \infty$ if no such κ exists. To obtain the proofs, we first give some lemmas.

Lemma 1 *By Lemma 2.5 and Corollary 2.6 in Van de Geer (2000), we can get the covering number of* $\mathbf{\Omega}_n^*$ *satisfying*

$$N(\varepsilon, \mathbf{\Omega}_n^*, L_2) \le c_1 c_0^K M_n^d \cdot \varepsilon^{-(d+K+2C)},$$

where c_0, c_1 *are finite constants independent of* n, K*, and* $K = \sum_{g=1}^{C} K_g$*,* $d = (K+C)q_n$*.*

Lemma 2 *Assume that Conditions (C1)-(C5) hold, then*

$$\sup_{\mathbf{\Omega}_n\in\mathbf{\Omega}_n^*, y\in[y_1,y_2]} |\widehat{H}_n(y;\mathbf{\Omega}_n) - H(y;\mathbf{\Omega}_n)| \to 0,$$

where $H(y;\mathbf{\Omega}_n)$ *is the solution to*

$$G(H(y;\mathbf{\Omega}_n); y, \mathbf{\Omega}_n) = 0, \tag{2.1}$$

for any given $\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*, y \in [y_1, y_2]$*.*

Proof: We first prove the existence of $\widehat{H}_n(y;\mathbf{\Omega}_n)$, then uniform convergence of $\widehat{H}_n(y;\mathbf{\Omega}_n)$ to $H(y;\mathbf{\Omega}_n)$ for $y \in [y_1, y_2]$ and $\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*$. By the Lemma 1 and Theorem 19.4 of Van der Vaart (1998), $\mathbf{\Omega}_n^*$ is P-Glivenko-Cantelli class. Since $\pi_g\Phi\left\{\frac{H_0(y)-\mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_n)}} - \vartheta\right\}$ is a continuous function on $\mathbf{\Omega}_n^*$ and the indicator function class is VC, and both are bounded by 1. Then by Theorem 3.7 of Van de Geer (2000) and the monotonicity of $H_0(y)$, we have

$$\frac{1}{\sum_{i=1}^{C} n_i}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij} \le y\right) - \sum_{g=1}^{C}\pi_g\Phi\left\{\frac{H_0(y)-\mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_n)}} - \vartheta\right\}\right]$$
$$\to \sum_{g=1}^{C} E\left[\pi_{g0}\Phi\left\{\frac{H_0(y)-\mu_{g0}(t_{ij})}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_0)}}\right\} - \pi_g\Phi\left\{\frac{H_0(y)-\mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_n)}} - \vartheta\right\}\right], \tag{2.2}$$

almost surely as $n \to \infty$, for any $\vartheta \geq 0$, and uniformly for $\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*$ and $y \in [y_1, y_2]$.

By (2.2), for large n and sufficiently large ϑ,

$$\frac{1}{\sum_{i=1}^{C} n_i} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[I\left(Y_{ij} \leq y\right) - \sum_{g=1}^{C} \pi_g \Phi \left\{ \frac{H_0(y) - \mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_n)}} - \vartheta \right\} \right] > 0, \quad (2.3)$$

$$\frac{1}{\sum_{i=1}^{C} n_i} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[I\left(Y_{ij} \leq y\right) - \sum_{g=1}^{C} \pi_g \Phi \left\{ \frac{H_0(y) - \mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_n)}} + \vartheta \right\} \right] < 0, \quad (2.4)$$

This together with the monotonicity and continuity of Φ implies that there exists a unique $\widehat{H}_n(y; \mathbf{\Omega}_n)$ such that

$$\frac{1}{\sum_{i=1}^{C} n_i} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[I\left(Y_{ij} \leq y\right) - \sum_{g=1}^{C} \pi_g \Phi \left\{ \frac{\widehat{H}_n(y; \mathbf{\Omega}_n) - \mathbf{B}_n(t_{ij})'\boldsymbol{\alpha}_g}{\sqrt{\sigma_{ij,g}(\mathbf{\Omega}_n)}} \right\} \right] = 0. \quad (2.5)$$

Furthermore, by Lemma 1 and the uniform strong law of large numbers, we have

$$G_n(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n) \to G(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n),$$

uniformly for $y \in [y_1, y_2]$ and $\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*$.

Denote $\zeta_n = \sup_{\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*, y \in [y_1, y_2]} \|G_n(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n)\|$, by the definition of $G(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n)$ in Supplementary 1 and the definition of $H(y; \mathbf{\Omega}_n)$ in (2.1), we have $G(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n) = 0$, hence $\zeta_n \to 0$. Note that

$$G_n(\widehat{H}_n(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n) = G_n(\widehat{H}_n(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n) - G_n(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n) + G_n(H(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n),$$

we have

$$0 = \|G_n(\widehat{H}_n(y; \mathbf{\Omega}_n); y, \mathbf{\Omega}_n)\| \geq M \|\widehat{H}_n(y; \mathbf{\Omega}_n) - H(y; \mathbf{\Omega}_n)\| - \zeta_n,$$

and hence $\widehat{H}_n(y; \mathbf{\Omega}_n) \to H(y; \mathbf{\Omega}_n)$ uniformly in $y \in [y_1, y_2]$ and $\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*$.

Lemma 3 *Under Conditions (C1)-(C5), the covering number of $\mathcal{L}_n$ satisfies*

$$N(\varepsilon, \mathcal{L}_n, L_1(P_n)) \leq c_1 c_0^K M_n^d \cdot \varepsilon^{-(d+K+2C)}.$$

Proof: For any $\mathbf{\Omega}^{(1)} = (\mathbf{\Lambda}^{(1)'}, \boldsymbol{\sigma}^{(1)2'}, \boldsymbol{\pi}^{(1)'}, \boldsymbol{\mu}^{(1)'}, \boldsymbol{\phi}^{(1)'})', \mathbf{\Omega}^{(2)} = (\mathbf{\Lambda}^{(2)'}, \boldsymbol{\sigma}^{(2)2'}, \boldsymbol{\pi}^{(2)'}, \boldsymbol{\mu}^{(2)'}, \boldsymbol{\phi}^{(2)'})' \in \mathbf{\Omega}_n^*$. By using the Taylor expansion and $\lambda\sqrt{n} \to 0$, we have

$$
\begin{aligned}
|l_i(\mathbf{\Omega}^{(1)}; H(\cdot; \mathbf{\Omega}^{(1)})) - l_i(\mathbf{\Omega}^{(2)}; H(\cdot; \mathbf{\Omega}^{(2)}))| \leq M(\|\mathbf{\Lambda}^{(1)} - \mathbf{\Lambda}^{(2)}\| + \|{\boldsymbol{\sigma}^{(1)}}^2 - {\boldsymbol{\sigma}^{(2)}}^2\| \\
+ \|\boldsymbol{\pi}^{(1)} - \boldsymbol{\pi}^{(2)}\| + \sum_{g=1}^{C} \|\mu_g^{(1)} - \mu_g^{(2)}\|_\infty + \sum_{g=1}^{C} \sum_{k=1}^{K_g} \|\phi_{gk}^{(1)} - \phi_{gk}^{(2)}\|_\infty).
\end{aligned}
\tag{2.6}
$$

Note that

$$
\begin{aligned}
\|\mu_g^{(1)} - \mu_g^{(2)}\|_\infty = & \sup_t |\sum_{j=1}^{q_n} \alpha_{gj}^{(1)} b_j(t) - \sum_{j=1}^{q_n} \alpha_{gj}^{(2)} b_j(t)| \\
\leq & M \max_{1 \leq j \leq q_n} |\alpha_{gj}^{(1)} - \alpha_{gj}^{(2)}| = M\|\boldsymbol{\alpha}_g^{(1)} - \boldsymbol{\alpha}_g^{(2)}\|_\infty.
\end{aligned}
$$

Similarly, we can have $\|\phi_{gk}^{(1)} - \phi_{gk}^{(2)}\|_\infty \leq M\|\boldsymbol{\beta}_{gk}^{(1)} - \boldsymbol{\beta}_{gk}^{(2)}\|_\infty$. Then by (2.6), we have that

$$
\begin{aligned}
P_n|l(\mathbf{\Omega}^{(1)}; H(\cdot; \mathbf{\Omega}^{(1)})) - l(\mathbf{\Omega}^{(2)}; H(\cdot; \mathbf{\Omega}^{(2)}))| \leq M(\|\mathbf{\Lambda}^{(1)} - \mathbf{\Lambda}^{(2)}\| + \|{\boldsymbol{\sigma}^{(1)}}^2 - {\boldsymbol{\sigma}^{(2)}}^2\| \\
+ \|\boldsymbol{\pi}^{(1)} - \boldsymbol{\pi}^{(2)}\| + \sum_{g=1}^{C} \|\boldsymbol{\alpha}_g^{(1)} - \boldsymbol{\alpha}_g^{(2)}\|_\infty + \sum_{g=1}^{C} \sum_{k=1}^{K_g} \|\boldsymbol{\beta}_{gk}^{(1)} - \boldsymbol{\beta}_{gk}^{(2)}\|_\infty).
\end{aligned}
\tag{2.7}
$$

By the uniform convergence of $\widehat{H}_n(y; \mathbf{\Omega}_n)$ to $H(y; \mathbf{\Omega}_n)$ in Lemma 2, we have

$$
P_n l(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) = P_n l(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n)) + o_p(1).
$$

This combining with (2.7) implies that, given $\mathbf{\Omega} \in \mathbf{\Omega}_n^*$, there exists $\mathbf{\Omega}^{(j)} = (\mathbf{\Lambda}^{(j)'}, \boldsymbol{\sigma}^{(j)2'}, \boldsymbol{\pi}^{(j)'}, \boldsymbol{\mu}^{(j)'}, \boldsymbol{\phi}^{(j)'})'$ $\in \mathbf{\Omega}_n^*$ such that

$$
\begin{aligned}
P_n|l((\mathbf{\Omega}; \widehat{H}_n(\cdot; \mathbf{\Omega})) - l(\mathbf{\Omega}^{(j)}; \widehat{H}_n(\cdot; \mathbf{\Omega}^{(j)}))| \leq M(\|\mathbf{\Lambda} - \mathbf{\Lambda}^{(j)}\| + \|\boldsymbol{\sigma}^2 - {\boldsymbol{\sigma}^{(j)}}^2\| + \|\boldsymbol{\pi} - \boldsymbol{\pi}^{(j)}\| \\
+ \sum_{g=1}^{C} \|\boldsymbol{\alpha}_g - \boldsymbol{\alpha}_g^{(j)}\|_\infty + \sum_{g=1}^{C} \sum_{k=1}^{K_g} \|\boldsymbol{\beta}_{gk} - \boldsymbol{\beta}_{gk}^{(j)}\|_\infty).
\end{aligned}
\tag{2.8}
$$

By (2.8) and Lemma 1 and following the calculation on page 94 of Van der Varrt & Weller(1996), we have

$$N(\varepsilon, \mathcal{L}_n, L_1(P_n)) \leq c_1 c_0^K M_n^d \cdot \varepsilon^{-(d+K+2C)}.$$

This completes the proof.

Lemma 4 *Suppose that Conditions (C1)-(C5) hold. Then*

$$\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))| \to 0 \textit{ almost surely.}$$

Proof: Note that

$$\begin{aligned} P_n l(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n)) &= P_n l(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) - P_n l(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n)) \\ &+ P_n l(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n)). \end{aligned}$$

According to uniform convergence of $\widehat{H}_n(y; \boldsymbol{\Omega}_n)$ in Lemma 2, we have

$$\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))| \leq \sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))| + o_p(1).$$

Hence, we only need to prove $\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))| \to 0$. By Condition (C5), let $(v+e)/2 < \phi < 1/2$ and $\alpha_n = n^{-1/2+\phi}(\log n)^{1/2}$. Here, $\{\alpha_n\}$ is a non-increasing positive numbers sequence, and for the given $\varepsilon > 0$ in Lemma 1, let $\varepsilon_n = \varepsilon \alpha_n$. Under Condition (C2), $Pl^2(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))$ is bounded. Then for any $\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*$ and sufficiently large n, we have

$$\frac{\mathrm{var}[P_n l(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))]}{(4\varepsilon_n)^2} \leq \frac{(1/n) Pl^2(\boldsymbol{\Omega}_n; H(\cdot; \boldsymbol{\Omega}_n))}{16\varepsilon^2 \alpha_n^2} \leq \frac{c_0}{16 n \varepsilon^2 \alpha_n^2} \leq \frac{1}{16 \varepsilon^2 n^{2\phi} \log n} \leq \frac{1}{2},$$

where c_0 is finite constant independent of n.

Furthermore, by the inequality (31) and Lemma 33 of Pollard (1984) and Lemma

3, we have

$$
\begin{aligned}
&P[\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n))| > 8\varepsilon_n] \\
&\le 8N(\varepsilon_n, \mathcal{L}_n, L_1(P_n)) \exp(-n\varepsilon_n^2/128) P[\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l^2(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n))| \le 64] \\
&\quad + P[\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l^2(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n))| > 64] \\
&\le M c_0^K M_n^d \cdot \varepsilon_n^{-(d+K+2C)} \exp(-n\varepsilon_n^2/128) \\
&\le M \exp[(d+K+2C)v \log n - (d+K+2C)\log\{\varepsilon n^{-1/2+\phi}(\log n)^{1/2}\} - n\varepsilon^2 n^{-1+2\phi} \log n/128] \\
&= M \exp[(d+K+2C)\{(v-1/2+\phi)\log n - \log\log n/2 - \log\varepsilon\} - \varepsilon^2 n^{2\phi}\log n/128] \\
&\le M \exp(-M n^{2\phi} \log n),
\end{aligned}
$$

where M is a constant. Hence, $\sum_{n=1}^{\infty} P[\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n))| > 8\varepsilon_n] < \infty$. By the Borel-Cantelli lemma (Chandra, 2012, pp. 15), we have $\sup_{\boldsymbol{\Omega}_n \in \boldsymbol{\Omega}_n^*} |P_n l(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n)) - Pl(\boldsymbol{\Omega}_n; H(\cdot;\boldsymbol{\Omega}_n))| \to 0$ almost sure. This completes the proof.

3 Web Appendix C: Proof of Theorems 1-3

Proof of Theorem 1

The proof of Theorem 1 is split into three steps. The first step proves the consistency of $\widehat{\boldsymbol{\Omega}}_n$ and $\widehat{H}_n(\cdot;\widehat{\boldsymbol{\Omega}}_n)$. The second step consists of the convergence rate of $\widehat{\boldsymbol{\Omega}}_n$. Finally, we obtain the selection consistency for the cluster number.

Step 1: consistency

Under Conditions (C1) and (C2), by Corollary 6.21 of Schumaker (1981), there exist $\mu_{n,g} = \boldsymbol{\alpha}'_{g0}\mathbf{B}_n(t)$ and $\phi_{n,gk} = \boldsymbol{\beta}'_{gk0}\mathbf{B}_n(t)$ such that

$$\sup_{t\in[0,1]} |\mu_{n,g}(t) - \mu_{g0}(t)| = O(q_n^{-r}), and$$
$$\sup_{t\in[0,1]} |\phi_{n,gk}(t) - \phi_{gk0}(t)| = O(q_n^{-r}),$$

where $\mu_{g0}(\cdot)$ and $\phi_{gk0}(\cdot)$ denote the true functions of $\mu_g(\cdot)$ and $\phi_{gk}(\cdot)$, respectively, for $g = 1, \cdots, C, k = 1, \cdots, K_g$. Let $\boldsymbol{\Omega}_{n0} = (\boldsymbol{\Lambda}'_0, \boldsymbol{\sigma}_0^{2'}, \boldsymbol{\pi}'_0, \boldsymbol{\mu}'_{n0}, \boldsymbol{\phi}'_{n0})'$, where $\boldsymbol{\mu}_{n0} = (\mu_{n,1}, \cdots, \mu_{n,C})'$, $\boldsymbol{\phi}_{n0} = (\phi_{n,gk}, g = 1, \cdots, C, k = 1, \cdots, K_g)'$. Then we have

$$d(\boldsymbol{\Omega}_0, \boldsymbol{\Omega}_{n0}) = O(n^{-rv+e/2}). \tag{3.1}$$

Let $\boldsymbol{\Theta_0} = (\boldsymbol{\pi}_0', \boldsymbol{\mu}_0', \boldsymbol{\Sigma}_0')'$ denote the true value of $\boldsymbol{\Theta}$, $\boldsymbol{\Theta_{n0}} = (\boldsymbol{\pi}_0', \boldsymbol{\mu}_{n0}', \boldsymbol{\Sigma}_{n0}')'$, and

$$\Sigma_{g0}(s,t) = \sum_{k=1}^{K_g} \phi_{gk0}(t)\lambda_{gk0}\phi_{gk0}(s) + \sigma_{g0}^2 I(s=t),$$
$$\Sigma_{n,g0}(s,t) = \sum_{k=1}^{K_g} \phi_{n,gk}(t)\lambda_{gk0}\phi_{n,gk}(s) + \sigma_{g0}^2 I(s=t).$$

From (3.1), we have that

$$\begin{aligned}
& \sup_{(t,s)\in[0,1]^2} |\Sigma_{g0}(s,t) - \Sigma_{n,g0}(s,t)| \\
\leq \; & \sum_{k=1}^{K_g} \lambda_{gk0} \sup_{(t,s)\in[0,1]^2} |\phi_{gk0}(t)\phi_{gk0}(s) - \phi_{n,gk}(t)\phi_{n,gk}(s)| \\
\leq \; & \sum_{k=1}^{K_g} \lambda_{gk0} \{ \sup_{(t,s)\in[0,1]^2} |\phi_{gk0}(s)||\phi_{gk0}(t) - \phi_{n,gk}(t)| + \sup_{(t,s)\in[0,1]^2} |\phi_{n,gk}(t)||\phi_{gk0}(s) - \phi_{n,gk}(s)| \} \\
= \; & O(K^{1/2} q_n^{-r}).
\end{aligned}$$

Then

$$d(\boldsymbol{\Theta_{n0}}, \boldsymbol{\Theta_0}) = O(n^{-rv+e/2}). \tag{3.2}$$

Let $M_i(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) = -l_i(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n))$, and $PM(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) = E\{M_i(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n))\}$, where "E" is expection over i. $P_n M(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n)) = \frac{1}{n}\sum_{i=1}^n M_i(\boldsymbol{\Omega}_n; \widehat{H}_n(\cdot; \boldsymbol{\Omega}_n))$,

$K_\varsigma = \{\mathbf{\Omega}_n : d(\mathbf{\Omega}_n, \mathbf{\Omega}_{n0}) \geq \varsigma, \mathbf{\Omega}_n \in \mathbf{\Omega}_n^*\}$ for $\varsigma > 0$. Then, we have

$$\begin{aligned}\inf_{K_\varsigma} PM(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) &= \inf_{K_\varsigma} \Big[PM(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) - P_n M(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) + P_n M(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n))\Big] \\ &\leq \zeta_{1n} + \inf_{K_\varsigma} P_n M(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)), \end{aligned} \tag{3.3}$$

where $\zeta_{1n} = \sup_{\mathbf{\Omega}_n \in \mathbf{\Omega}_n^*} |P_n M(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) - PM(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n))|$.

If $\widehat{\mathbf{\Omega}}_n \in K_\varsigma$, one can show that

$$\begin{aligned}&\inf_{K_\varsigma} P_n M(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) = P_n M(\widehat{\mathbf{\Omega}}_n; \widehat{H}_n(\cdot; \widehat{\mathbf{\Omega}}_n)) \\ &\quad \leq P_n M(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0})) = \zeta_{2n} + PM(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0})), \end{aligned} \tag{3.4}$$

with $\zeta_{2n} = P_n M(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0})) - PM(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0}))$. By (3.3) and (3.4), we have

$$\inf_{K_\varsigma} PM(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) \leq \zeta_{1n} + \zeta_{2n} + PM(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0})) = \zeta_n + PM(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0})),$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. It is obvious that $\zeta_n \geq \delta_\varsigma \widehat{=} \inf_{K_\varsigma} PM(\mathbf{\Omega}_n; \widehat{H}_n(\cdot; \mathbf{\Omega}_n)) - PM(\mathbf{\Omega}_{n0}; \widehat{H}_n(\cdot; \mathbf{\Omega}_{n0}))$ which is larger than zero under condition $\lambda\sqrt{n} \to 0$ when n is large enough. Hence

$$\{\widehat{\mathbf{\Omega}}_n \in K_\varsigma\} \subseteq \{\zeta_n \geq \delta_\varsigma\}. \tag{3.5}$$

By Lemma 2 and Law of Large Numbers, we have $\zeta_{1n} \to 0$ and $\zeta_{2n} \to 0$ then $\zeta_n \to 0$ almost surely. Hence, when n is large enough $\bigcup_{i=1}^\infty \bigcap_{n=i}^\infty \{\zeta_n \geq \delta_\varsigma\}$ is null set. By (3.5), we have $\bigcup_{i=1}^\infty \bigcap_{n=i}^\infty \{\widehat{\mathbf{\Omega}}_n \in K_\varsigma\}$ is null set. Coupling with the definition of K_ς, we have

$$d(\widehat{\mathbf{\Omega}}_n, \mathbf{\Omega}_{n0}) \to 0, \tag{3.6}$$

almost surely as $n \to \infty$. This together with (3.1), we have $d(\widehat{\mathbf{\Omega}}_n, \mathbf{\Omega}_0) \to 0$. Furthermore, since

$$\begin{aligned}\widehat{\Sigma}_g(s,t)-\Sigma_{n,g0}(s,t) &= \sum_{k=1}^{K_g}\Big\{\widehat{\lambda}_{gk}\widehat{\phi}_{gk}(t)\widehat{\phi}_{gk}(s)-\lambda_{gk0}\phi_{n,gk}(t)\phi_{n,gk}(s)\Big\}+(\widehat{\sigma}_g^2-\sigma_{g0}^2)I(s=t)\\ &= \sum_{k=1}^{K_g}\widehat{\lambda}_{gk}\Big[\widehat{\phi}_{gk}(t)\big\{\widehat{\phi}_{gk}(s)-\phi_{n,gk}(s)\big\}+\phi_{n,gk}(s)\big\{\widehat{\phi}_{gk}(t)-\phi_{n,gk}(t)\big\}\Big]\\ &\quad +\{\widehat{\lambda}_{gk}-\lambda_{gk0}\}\phi_{n,gk}(t)\phi_{n,gk}(s)+(\widehat{\sigma}_g^2-\sigma_{g0}^2)I(s=t), \qquad (3.7)\end{aligned}$$

therefore $\widehat{\Sigma}_g(s,t)\to\Sigma_{n,g0}(s,t)$ almost surely for $g=1,\cdots,C$ by (3.6). Then $d(\widehat{\mathbf{\Theta}}_n,\mathbf{\Theta}_{n0})\to 0$, together with (3.2), we prove $d(\widehat{\mathbf{\Theta}}_n,\mathbf{\Theta}_0)\to 0$. Since $\widehat{H}_n(y)\equiv\widehat{H}_n(y;\widehat{\mathbf{\Omega}}_n)$, coupling with Lemma 2, we have $\widehat{H}_n(y)\to H_0(y)$ uniformly in $y\in[y_1,y_2]$.

Step 2: convergence rate

To show the convergence rate, for any $\eta>0$, define class $\mathcal{L}_\eta=\{l_i(\mathbf{\Omega}_n;\widehat{H}_n(\cdot;\mathbf{\Omega}_n))-l_i(\mathbf{\Omega}_{n0};\widehat{H}_n(\cdot;\mathbf{\Omega}_{n0})):\mathbf{\Omega}_n\in\mathbf{\Omega}_n^*,\eta/2\le d(\mathbf{\Omega}_n,\mathbf{\Omega}_{n0})\le\eta\}$, and $P[l(\mathbf{\Omega}_n;\widehat{H}_n(\cdot;\mathbf{\Omega}_n))-l(\mathbf{\Omega}_{n0};\widehat{H}_n(\cdot;\mathbf{\Omega}_{n0}))]=E[l_i(\mathbf{\Omega}_n;\widehat{H}_n(\cdot;\mathbf{\Omega}_n))-l_i(\mathbf{\Omega}_{n0};\widehat{H}_n(\cdot;\mathbf{\Omega}_{n0}))$. Following the calculation of Shen and Wong (1994,p.597), for $0<\varepsilon<\eta$, we can establish that $\log N_{[]}(\varepsilon,\mathcal{L}_\eta,L_2(P))\le MKq_n\log(\eta/\varepsilon)$. Moreover, for large n and $\lambda\sqrt{n}\to 0$, we have $P[l(\mathbf{\Omega}_n;\widehat{H}_n(\cdot;\mathbf{\Omega}_n))-l(\mathbf{\Omega}_{n0};\widehat{H}_n(\cdot;\mathbf{\Omega}_{n0}))]\le M\eta^2$, for any $l(\mathbf{\Omega}_n;\widehat{H}_n(\cdot;\mathbf{\Omega}_n))-l(\mathbf{\Omega}_{n0};\widehat{H}_n(\cdot;\mathbf{\Omega}_{n0}))\in\mathcal{L}_\eta$. Therefore, by Lemma 3.4.2 of Van der Varrt and Wellner (1996) and the condition $\lambda\sqrt{n}\to 0$, we have

$$E_P\|\sqrt{n}(P_n-P)\|_{\mathcal{L}_\eta}\le MJ_{[]}(\eta,\mathcal{L}_\eta,L_2(P))\left[1+\frac{J_{[]}(\eta,\mathcal{L}_\eta,L_2(P))}{n^{\frac{1}{2}}\eta^2}\right], \qquad (3.8)$$

where $J_{[]}(\eta,\mathcal{L}_\eta,L_2(P))=\int_0^\eta\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{L}_\eta,L_2(P))}d\varepsilon\le M(Kq_n)^{1/2}\eta$. The right hand of (3.8) yields that the key function $\phi_n(\eta)=\sqrt{Kq_n}\eta+Kq_n/n^{1/2}$. Note that, $\phi_n(\eta)/\eta$ is decreasing in η, and $r_n^2\phi_n(1/r_n)=r_n\sqrt{Kq_n}+r_n^2Kq_n/n^{1/2}\le Mn^{1/2}$,where $r_n=(Kq_n)^{-1/2}n^{1/2}=n^{(1-v-e)/2}$. Hence $n^{(1-v-e)/2}d(\widehat{\mathbf{\Omega}}_n,\mathbf{\Omega}_{n0})=O_p(1)$ by Theorem 3.4.1 of Van der Varrt and Wellner (1996). This with (3.1), yields that $d(\widehat{\mathbf{\Omega}}_n,\mathbf{\Omega}_0)=O_p\{n^{-\min(rv-e/2,(1-v-e)/2)}\}$. Furthermore, by $d(\widehat{\mathbf{\Omega}}_n,\mathbf{\Omega}_{n0})=O_p\{n^{-(1-v-e)/2}\}$ and (3.7), we have $d(\widehat{\mathbf{\Theta}}_n, \mathbf{\Theta}_{n0}) = O_p\{n^{-(1-v-e)/2}\}$, combining with (3.2), yields that

$$d(\widehat{\mathbf{\Theta}}_n, \mathbf{\Theta}_0) = O_p\{n^{-(1-v-e)/2} + n^{-rv+e/2}\} = O_p\{n^{-\min(rv-e/2,(1-v-e)/2)}\}. \tag{3.9}$$

Step 3: selection consistency

Denote $\dot{Q}_{ng}(\mathbf{\Omega}_n; H(\cdot)) = \frac{\partial Q_n(\mathbf{\Omega}_n; H(\cdot))}{\partial \pi_g}$. Following the lines of the proof for Theorem 3, for any given number of clusters C, we have that the estimates of parametric part have an $n^{1/2}$ rate of convergence, therefore, $\pi_g = \pi_{g0} + O_p(n^{-1/2})$ for $g = 1, \cdots, C$ with the condition $\lambda\sqrt{n} \to 0$. According to the proof of Huang et al. (2017), we only need to consider the solution of $\dot{Q}_{ng}(\mathbf{\Omega}_n; \widehat{H}(\cdot; \mathbf{\Omega}_n)) = 0$ with $d(\mathbf{\Omega}_n, \mathbf{\Omega}_0) = O_p\{n^{-\min(rv-e/2,(1-v-e)/2)}\}$, $\pi_g < \frac{1}{\sqrt{n}\log(n)}$ and $g > C_0$. According to Lemma 2, it is sufficient to show that

$$\dot{Q}_{ng}(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n)) < 0 \qquad for \quad 0 < \pi_g < \frac{1}{\sqrt{n}\log(n)} \text{ and } g > C_0, \tag{3.10}$$

with probability tending to 1 as $n \to \infty$. With the constraint $\sum_{g=1}^{C} \pi_g = 1$, we can write

$$\dot{Q}_{ng}(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n)) = \sum_{i=1}^{n} \frac{f_{gi}(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n))}{\sum_{g=1}^{C} \pi_g f_{gi}(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n))} - \sum_{i=1}^{n} \frac{f_{1i}(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n))}{\sum_{g=1}^{C} \pi_g f_{gi}(\mathbf{\Omega}_n; H(\cdot; \mathbf{\Omega}_n))}$$
$$-n\lambda \frac{1}{\epsilon + \pi_g} + n\lambda \frac{1}{\epsilon + \pi_1} \widehat{=} R_1 - R_2 - R_3 + R_4. \tag{3.11}$$

By the law of large numbers, it is obvious that R_1 and R_2 are of order $O_p(n)$. Furthermore, we have $\pi_g = \pi_{g0} + O_p(n^{-1/2}) > \frac{1}{2} \cdot \min\{\pi_{10}, \cdots, \pi_{C_0,0}\}$ when $g = 1$. Hence, R_4 should be $O_p(n\lambda) = o_p(n)$ by the condition $\lambda = o(1)$.

Because $\epsilon = o\{\frac{1}{\sqrt{n}\log(n)}\}$ in Theorem 1 when $g > C_0$ and $0 < \pi_g < \frac{1}{\sqrt{n}\log(n)}$, we have $R_3 = O_p(n\lambda\sqrt{n}\log n)$, with probability tending to one. Since R_1 and R_2 are of order $O_p(n)$ and R_4 is of order $o_p(n)$, hence, R_3 dominates R_1, R_2 and R_4 with the condition $\lambda\sqrt{n}\log n \to \infty$. Therefore, we prove (3.10), or equivalently $\pi_g = 0$ for $g > C_0$ with probability tending to one when $n \to \infty$.

Proof of Theorem 2

The proof of Theorem 2 follows by Steps 1 and 2 in the proof of Theorem 1.

Proof of Theorem 3

Let $\mathbf{\Omega}^*_{\backslash 0}$ to be $\mathbf{\Omega}^*$ excluding the true value $\mathbf{\Omega}_0$, where $\mathbf{\Omega}^*$ denotes the parameter space. Denote $\delta_n = n^{-\min(rv-e/2,(1-v-e)/2)}$. Let V denote the linear span of $\mathbf{\Omega}^*_{\backslash 0}$ and define the Fisher inner product on the space V as $< \boldsymbol{v}, \widetilde{\boldsymbol{v}} >= P\left\{\dot{l}_i(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}]\dot{l}_i(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\widetilde{\boldsymbol{v}}]\right\}$ for $\boldsymbol{v}, \widetilde{\boldsymbol{v}} \in V$, the Fisher norm $\|\boldsymbol{v}\|^2 =< \boldsymbol{v}, \boldsymbol{v} >$, where $\dot{l}_i(\mathbf{\Omega}; H(\cdot;\mathbf{\Omega}))[\boldsymbol{v}] = \frac{dl_i(\mathbf{\Omega}+s\boldsymbol{v}; H(\cdot;\mathbf{\Omega}))}{ds}\Big|_{s=0}$ is the first order directional derivative of $l_i(\mathbf{\Omega}; H(\cdot;\mathbf{\Omega}))$ at the direction $\boldsymbol{v} \in V$, and $\mathbf{\Omega} \in \{\mathbf{\Omega} \in \mathbf{\Omega}^* : d(\mathbf{\Omega}, \mathbf{\Omega}_0) = O(\delta_n)\}$. Let $\overline{V}$ be the closed linear span of V under the Fisher norm, then $(\overline{V}, \|\cdot\|)$ is a Hilbert space. Let $\dot{l}_{i,\mathbf{\Omega}}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}] = \dot{l}_i(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}] + \dot{l}_{i,H}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}]$, where $\dot{l}_{i,H}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}] = \frac{dl_i(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0+s\boldsymbol{v}))}{ds}\Big|_{s=0}$. Denote $< \boldsymbol{v}, \widetilde{\boldsymbol{v}} >_{\mathbf{\Omega}}= P\left\{\dot{l}_i(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}]\dot{l}_{i,\mathbf{\Omega}}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\widetilde{\boldsymbol{v}}]\right\}$. For a $2C_0$-dimensional vector $\mathbf{\Gamma} = (\mathbf{\Gamma}_1', \mathbf{\Gamma}_2')'$ with $\|\mathbf{\Gamma}\| \leq 1$, let $\rho(\mathbf{\Omega}; H(\cdot;\mathbf{\Omega})) = \mathbf{\Gamma}'\boldsymbol{\theta} = \boldsymbol{\sigma}^{2\prime}\mathbf{\Gamma}_1 + \boldsymbol{\pi}'\mathbf{\Gamma}_2$ denote a smooth functional of $\mathbf{\Omega}$ and $H(\cdot;\mathbf{\Omega})$, and $\dot{\rho}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\boldsymbol{v}] = \frac{d\rho(\mathbf{\Omega}_0+s\boldsymbol{v}; H(\cdot;\mathbf{\Omega}_0+s\boldsymbol{v}))}{ds}\Big|_{s=0}$. Note that $\rho(\mathbf{\Omega}; H(\cdot;\mathbf{\Omega})) - \rho(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0)) = \dot{\rho}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\mathbf{\Omega} - \mathbf{\Omega}_0]$. According to the Riesz representation theorem, for any given $\boldsymbol{v} \in V$, there exists $\boldsymbol{v}^* \in \overline{V}$ such that $\dot{\rho}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[v] =< \boldsymbol{v}^*, \boldsymbol{v} >_{\mathbf{\Omega}}$. Thus, according to the Cramér-Wold device, to prove Theorem 3, it suffices to show that

$$\sqrt{n} < \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0 >_{\mathbf{\Omega}} \xrightarrow{d} N\{0, \mathbf{\Gamma}'\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{\Gamma}\}, \tag{3.12}$$

due to $\mathbf{\Gamma}'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \rho(\widehat{\mathbf{\Omega}}_n; H(\cdot;\widehat{\mathbf{\Omega}}_n)) - \rho(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0)) = \dot{\rho}(\mathbf{\Omega}_0; H(\cdot;\mathbf{\Omega}_0))[\widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] =< \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0 >_{\mathbf{\Omega}}$. In fact, (3.12) holds when $\sqrt{n} < \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0 >_{\mathbf{\Omega}} \xrightarrow{d} N(0, \|\boldsymbol{v}^*\|^2)$ and $\|\boldsymbol{v}^*\|^2 = \mathbf{\Gamma}'\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\mathbf{\Gamma}$.

We will take two steps to prove (3.12). First, we prove $\sqrt{n} < \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0 >_{\mathbf{\Omega}} \xrightarrow{d} N(0, \|\boldsymbol{v}^*\|^2)$. By Corollary 6.21 of Schumaker (1981), there exists $\Pi_n \boldsymbol{v}^* \in \mathbf{\Omega}^*_n$ such that $\|\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*\| = o(1)$ and $\delta_n \|\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*\| = o(n^{-1/2})$. For any $\mathbf{\Omega} \in \{\mathbf{\Omega} \in$

$\mathbf{\Omega}^* \; : \; d(\mathbf{\Omega}, \mathbf{\Omega}_0) = O(\delta_n)\}$, define $\ddot{l}_i(\mathbf{\Omega}; H(\cdot; \mathbf{\Omega}))[\boldsymbol{v}, \widetilde{\boldsymbol{v}}] = \left.\frac{d\dot{l}_i(\mathbf{\Omega}+\widetilde{s}\widetilde{\boldsymbol{v}}; H(\cdot;\mathbf{\Omega}+\widetilde{s}\widetilde{\boldsymbol{v}}))[\boldsymbol{v}]}{d\widetilde{s}}\right|_{\widetilde{s}=0}$, and $r_i[\Pi_n \boldsymbol{v}^*, \mathbf{\Omega} - \mathbf{\Omega}_0] = \dot{l}_i(\mathbf{\Omega}; H(\cdot; \mathbf{\Omega}))[\Pi_n \boldsymbol{v}^*] - \dot{l}_i(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\Pi_n \boldsymbol{v}^*]$.

Since $P_n \dot{l}(\widehat{\mathbf{\Omega}}_n; \widehat{H}(\cdot; \widehat{\mathbf{\Omega}}_n))[\Pi_n \boldsymbol{v}^*] = 0$ by the definition of $\widehat{\mathbf{\Omega}}_n$. Then by Lemma 2, we have

$$
\begin{aligned}
0 &= P_n \dot{l}(\widehat{\mathbf{\Omega}}_n; H(\cdot; \widehat{\mathbf{\Omega}}_n))[\Pi_n \boldsymbol{v}^*] = P_n \dot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\Pi_n \boldsymbol{v}^*] + P_n r[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] \\
&= P_n \dot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\boldsymbol{v}^*] + P_n \dot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*] + (P_n - P) r[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] \\
&\quad + P r[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] \\
&= P_n \dot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\boldsymbol{v}^*] + I_1 + I_2 + I_3. \qquad (3.13)
\end{aligned}
$$

We will investigate the asymptotic behavior of I_1, I_2 and I_3. For I_1, by Chebyshev inequality, $\lambda\sqrt{n} \to 0$, $P\dot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*] = 0$ and $\|\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*\| = o(1)$, we have that

$$
I_1 = o_p(n^{-1/2}). \qquad (3.14)
$$

For I_2, by the definition of $r[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0]$, we can get

$$
I_2 = (P_n - P)\left\{\dot{l}(\widehat{\mathbf{\Omega}}_n; H(\cdot; \widehat{\mathbf{\Omega}}_n)) - \dot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))\right\}[\Pi_n \boldsymbol{v}^*].
$$

By Theorem 2.8.3 of Van der Varrt and Wellner (1996), $\{\dot{l}(\mathbf{\Omega}; H(\cdot; \mathbf{\Omega}))[\Pi_n \boldsymbol{v}^*] : \|\mathbf{\Omega} - \mathbf{\Omega}_0\| = O_p(\delta_n)\}$ is a Donsker class. Hence by Theorem 2.11.23 of Van der Varrt and Wellner (1996), we have

$$
I_2 = o_p(n^{-1/2}). \qquad (3.15)
$$

For I_3, by the definition of $r[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0]$ and the Mean Value Theorem (Flett, 1958), we have

$$
\begin{aligned}
I_3 &= P\left\{\ddot{l}(\widetilde{\mathbf{\Omega}}; H(\cdot; \widetilde{\mathbf{\Omega}})) - \ddot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))\right\}[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] + P\ddot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\Pi_n \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] \\
&= o_p(n^{-1/2}) + P\ddot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] + P\ddot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0] \\
&= o_p(n^{-1/2}) + P\ddot{l}(\mathbf{\Omega}_0; H(\cdot; \mathbf{\Omega}_0))[\boldsymbol{v}^*, \widehat{\mathbf{\Omega}}_n - \mathbf{\Omega}_0], \qquad (3.16)
\end{aligned}
$$

where $\widetilde{\boldsymbol{\Omega}}$ is between $\widehat{\boldsymbol{\Omega}}_n$ and $\boldsymbol{\Omega}_0$, the second equation follows by using a Taylor expansion, Condition (C2) and $\|\Pi_n \boldsymbol{v}^*\|^2 \rightarrow \|\boldsymbol{v}^*\|^2$, the last equation holds since $\delta_n \|\Pi_n \boldsymbol{v}^* - \boldsymbol{v}^*\| = o(n^{-1/2})$. Hence, by (3.14), (3.13), (3.15), (3.16), together with $\lambda\sqrt{n} \rightarrow 0$, $P\dot{l}(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))[\boldsymbol{v}^*] = 0$ and the definition of $<\cdot,\cdot>_{\boldsymbol{\Omega}}$, we have that

$$
\begin{aligned}
0 &= P_n \dot{l}(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))[\boldsymbol{v}^*] - <\boldsymbol{v}^*, \widehat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_0>_{\boldsymbol{\Omega}} + o_p(n^{-1/2}) \\
&= (P_n - P)\dot{l}(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))[\boldsymbol{v}^*] - <\boldsymbol{v}^*, \widehat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_0>_{\boldsymbol{\Omega}} + o_p(n^{-1/2}).
\end{aligned}
$$

Hence

$$
\sqrt{n} <\boldsymbol{v}^*, \widehat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_0>_{\boldsymbol{\Omega}} = \sqrt{n}(P_n - P)\dot{l}(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))[\boldsymbol{v}^*] + o_p(1) \rightarrow N(0, \|\boldsymbol{v}^*\|^2),
$$

with $\|\boldsymbol{v}^*\|^2 = \|\dot{l}(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))[\boldsymbol{v}^*]\|^2$.

In the second step, we calculate $\|\boldsymbol{v}^*\|^2$. Rewrite $\boldsymbol{\theta} = (\boldsymbol{\sigma}^{2\prime}, \boldsymbol{\pi}')' = (\theta_1, \cdots, \theta_{2C_0})'$, let $K_0 = \sum_{g=1}^{C_0} K_g$ and $l_{i,b_j} = \frac{\partial l_i(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))}{\partial b_j}$. For each $\theta_q, q = 1, 2, \cdots, 2C_0$, denote $\psi_q^* = \{b_{1q}^*, b_{2q}^*, \cdots, b_{(C_0+K_0)q+K_0}^*, b_{H,q}^*\}$ be the minimizer

$$
E\left\{l_{i,\boldsymbol{\theta}} \cdot e_q - l_{i,b_1}[b_{1q}] - l_{i,b_2}[b_{2q}] - \cdots - l_{i,b_{C_0+2K_0}}[b_{(C_0+K_0)q+K_0}] - l_{i,b_H}[b_{H,q}]\right\}^2,
$$

with respect to $\psi_q = \{b_{1q}, b_{2q}, \cdots, b_{(C_0+K_0)q+K_0}, b_{H,q}\}$, where e_q is a $2C_0$-dimensional vector of zeros except the qth element equal to 1, $l_{i,\boldsymbol{\theta}} = (l'_{i,\boldsymbol{\sigma}^2}, l'_{i,\boldsymbol{\pi}})'$, $l_{i,\boldsymbol{\sigma}^2} = (l_{i,\sigma_1^2}, \cdots, l_{i,\sigma_{C_0}^2})'$, $l_{i,\boldsymbol{\pi}} = (l_{i,\pi_1}, \cdots, l_{i,\pi_{C_0}})'$, $l_{i,\sigma_g^2} = \frac{\partial l_i(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))}{\partial \sigma_g^2}$, $l_{i,\pi_g} = \frac{\partial l_i(\boldsymbol{\Omega}_0; H(\cdot;\boldsymbol{\Omega}_0))}{\partial \pi_g}$, $g = 1, \cdots, C_0$, and $(l_{i,b_j}[b_{jq}], j = 1, \cdots, C_0)$ is the directional derivative of $\boldsymbol{\mu}$, $(l_{i,b_j}[b_{jq}], j = C_0 + 1, \cdots, C_0 + K_0)$ is the directional derivative of $\boldsymbol{\phi}$, $(l_{i,b_j}[b_j], j = C_0 + K_0 + 1, \cdots, C_0 + 2K_0)$ is the directional derivative of $\boldsymbol{\Lambda}$, and $l_{i,b_H}[b_{H,q}]$ is the directional derivative of H. By the similar calculation of Chen et al. (2006), we can obtain $\|\boldsymbol{v}^*\|^2 = \boldsymbol{\Gamma}'[E\{S(\boldsymbol{\theta}_0)S(\boldsymbol{\theta}_0)'\}]^{-1}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'\mathbf{I}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}$, where $S(\boldsymbol{\theta}_0)$ is a $2C_0$-dimensional vector, with the qth element as $l_{i,\boldsymbol{\theta}} \cdot e_q - l_{i,b_1}[b_{1q}^*] - l_{i,b_2}[b_{2q}^*] - \cdots - l_{i,b_{C_0+2K_0}}[b_{(C_0+K_0)q+K_0}^*] - l_{i,b_H}[b_{H,q}^*]$. Then, we complete the proof of Theorem 3.

4 Web Appendix D: Tables 6-8 and Figures 3-9

Table 6: Results for Case 2 of Simulation 1.

	Proposed(C=7)		CT(C=3)		CT(C=7)		WoT(C=3)	
	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE
π_1	0.005(0.043)	0.043	0.002(0.031)	0.031	0.003(0.041)	0.041	0.058(0.054)	0.080
π_2	0.009(0.052)	0.053	0.000(0.032)	0.032	0.008(0.046)	0.047	0.053(0.085)	0.100
π_3	0.003(0.052)	0.052	0.002(0.030)	0.030	0.003(0.038)	0.038	0.111(0.082)	0.138
σ_1^2	0.002(0.016)	0.016	0.000(0.011)	0.011	0.001(0.011)	0.011	0.094(0.002)	0.094
σ_2^2	0.003(0.018)	0.018	0.001(0.014)	0.014	0.001(0.014)	0.014	0.141(0.002)	0.141
σ_3^2	0.002(0.020)	0.021	0.001(0.015)	0.015	0.001(0.019)	0.019	0.188(0.001)	0.188
λ_{11}	0.022(0.152)	0.153	0.017(0.132)	0.133	0.022(0.140)	0.141	0.960(0.008)	0.960
λ_{12}	0.009(0.037)	0.038	0.003(0.036)	0.036	0.003(0.036)	0.036	0.236(0.005)	0.236
λ_{21}	0.031(0.152)	0.155	0.006(0.130)	0.130	0.021(0.131)	0.133	1.044(0.014)	1.044
λ_{22}	0.001(0.033)	0.033	0.001(0.010)	0.010	0.001(0.016)	0.016	0.003(0.009)	0.010
λ_{31}	0.040(0.151)	0.156	0.031(0.105)	0.110	0.033(0.135)	0.139	0.937(0.009)	0.937
λ_{32}	0.030(0.086)	0.091	0.015(0.058)	0.060	0.015(0.061)	0.063	0.132(0.005)	0.132
$\mu_1(\cdot)$	0.007(0.131)	0.132	0.006(0.091)	0.091	0.007(0.093)	0.093	0.362(0.074)	0.369
$\mu_2(\cdot)$	0.013(0.135)	0.135	0.007(0.105)	0.105	0.012(0.109)	0.110	0.458(0.099)	0.468
$\mu_3(\cdot)$	0.012(0.133)	0.134	0.010(0.089)	0.090	0.011(0.115)	0.115	1.252(0.073)	1.255
$\phi_{11}(\cdot)$	0.007(0.069)	0.069	0.005(0.063)	0.063	0.007(0.065)	0.065	0.187(0.475)	0.511
$\phi_{12}(\cdot)$	0.008(0.081)	0.082	0.007(0.075)	0.075	0.008(0.077)	0.077	0.184(0.495)	0.528
$\phi_{21}(\cdot)$	0.016(0.061)	0.063	0.012(0.059)	0.060	0.016(0.061)	0.063	0.127(0.316)	0.341
$\phi_{22}(\cdot)$	0.008(0.090)	0.090	0.007(0.087)	0.087	0.008(0.089)	0.090	0.456(0.819)	0.937
$\phi_{31}(\cdot)$	0.024(0.093)	0.096	0.018(0.055)	0.058	0.018(0.060)	0.063	0.068(0.197)	0.208
$\phi_{32}(\cdot)$	0.007(0.046)	0.046	0.004(0.011)	0.011	0.006(0.041)	0.041	0.105(0.113)	0.154
$\Sigma_1(\cdot,\cdot)$	0.027(0.174)	0.176	0.022(0.154)	0.156	0.024(0.161)	0.163	0.996(0.092)	1.000
$\Sigma_2(\cdot,\cdot)$	0.043(0.172)	0.178	0.037(0.156)	0.160	0.041(0.159)	0.164	1.048(0.145)	1.058
$\Sigma_3(\cdot,\cdot)$	0.048(0.177)	0.184	0.031(0.135)	0.139	0.046(0.151)	0.158	1.582(0.067)	1.583
$\sharp$cluster	0.050(0.219)	0.224	-	-	0.003(0.208)	0.209	-	-

"−" not available

Table 7: Results for Simulation 2.

	Proposed(C=7)		CT(C=3)		CT(C=7)		WoT(C=3)	
	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE
π_1	0.024(0.091)	0.094	0.017(0.080)	0.081	0.022(0.084)	0.087	0.024(0.400)	0.401
π_2	0.032(0.090)	0.096	0.013(0.064)	0.066	0.013(0.068)	0.069	0.077(0.391)	0.399
π_3	0.027(0.116)	0.119	0.017(0.109)	0.111	0.026(0.114)	0.117	0.053(0.410)	0.413
σ_1^2	0.004(0.017)	0.018	0.003(0.013)	0.013	0.004(0.018)	0.018	0.020(0.003)	0.020
σ_2^2	0.016(0.090)	0.092	0.014(0.085)	0.086	0.014(0.085)	0.086	0.150(0.003)	0.150
σ_3^2	0.002(0.085)	0.085	0.001(0.078)	0.078	0.001(0.082)	0.082	0.237(0.003)	0.237
λ_{11}	0.017(0.087)	0.089	0.009(0.074)	0.074	0.015(0.083)	0.085	0.198(0.011)	0.198
λ_{31}	0.006(0.041)	0.041	0.003(0.029)	0.029	0.005(0.029)	0.030	0.137(0.013)	0.138
λ_{32}	0.008(0.038)	0.038	0.004(0.021)	0.021	0.007(0.032)	0.033	0.050(0.012)	0.052
$\mu_1(\cdot)$	0.003(0.033)	0.033	0.003(0.014)	0.014	0.003(0.014)	0.014	1.024(0.002)	1.024
$\mu_2(\cdot)$	0.004(0.023)	0.024	0.002(0.018)	0.018	0.002(0.019)	0.019	0.920(0.002)	0.920
$\mu_3(\cdot)$	0.002(0.042)	0.042	0.001(0.038)	0.038	0.001(0.042)	0.042	1.027(0.002)	1.027
$\phi_{11}(\cdot)$	0.006(0.038)	0.039	0.006(0.030)	0.031	0.006(0.031)	0.032	0.705(0.657)	0.964
$\phi_{31}(\cdot)$	0.017(0.069)	0.071	0.013(0.061)	0.062	0.016(0.062)	0.064	0.181(0.550)	0.579
$\phi_{32}(\cdot)$	0.005(0.037)	0.037	0.004(0.032)	0.032	0.004(0.032)	0.032	0.171(0.031)	0.174
$\Sigma_1(\cdot,\cdot)$	0.019(0.125)	0.127	0.012(0.095)	0.096	0.017(0.118)	0.119	0.929(0.819)	1.239
$\Sigma_2(\cdot,\cdot)$	0.016(0.090)	0.092	0.014(0.085)	0.086	0.014(0.085)	0.086	0.150(0.003)	0.150
$\Sigma_3(\cdot,\cdot)$	0.032(0.130)	0.134	0.029(0.088)	0.093	0.032(0.128)	0.132	0.982(0.707)	1.210
♯cluster	0.021(0.185)	0.186	-	-	0.007(0.144)	0.144	-	-

"−" not available

Table 8: The proposed estimators under gamma distributions for Simulation 3

τ	normal		100		10		5		1	
Skewness	0		0.2		0.63		0.89		2	
Excess kurtosis	0		0.06		0.6		1.2		6	
	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE	bias(sd)	RMSE
π_1	0.003(0.034)	0.034	0.004(0.034)	0.035	0.006(0.034)	0.035	0.007(0.034)	0.035	0.009(0.034)	0.035
π_2	0.009(0.039)	0.039	0.009(0.039)	0.039	0.009(0.040)	0.041	0.014(0.039)	0.041	0.030(0.053)	0.061
π_3	0.004(0.038)	0.038	0.006(0.038)	0.038	0.011(0.039)	0.041	0.017(0.040)	0.043	0.039(0.054)	0.067
σ_1^2	0.001(0.009)	0.009	0.001(0.009)	0.009	0.002(0.010)	0.010	0.001(0.010)	0.010	0.006(0.012)	0.013
σ_2^2	0.001(0.014)	0.014	0.004(0.019)	0.019	0.006(0.023)	0.024	0.008(0.024)	0.025	0.018(0.034)	0.038
σ_3^2	0.006(0.030)	0.031	0.007(0.032)	0.032	0.009(0.033)	0.034	0.009(0.035)	0.036	0.021(0.039)	0.044
λ_{11}	0.030(0.145)	0.148	0.040(0.164)	0.169	0.050(0.189)	0.195	0.063(0.204)	0.213	0.066(0.234)	0.243
λ_{12}	0.011(0.036)	0.038	0.012(0.041)	0.043	0.014(0.054)	0.055	0.016(0.056)	0.058	0.016(0.058)	0.061
λ_{21}	0.037(0.140)	0.145	0.043(0.177)	0.182	0.051(0.176)	0.183	0.073(0.187)	0.200	0.069(0.225)	0.235
λ_{22}	0.001(0.011)	0.011	0.006(0.035)	0.036	0.012(0.037)	0.038	0.015(0.039)	0.042	0.039(0.047)	0.062
λ_{31}	0.038(0.149)	0.154	0.041(0.170)	0.175	0.050(0.173)	0.180	0.067(0.178)	0.190	0.069(0.177)	0.190
λ_{32}	0.017(0.056)	0.058	0.021(0.057)	0.061	0.024(0.058)	0.063	0.035(0.061)	0.071	0.036(0.073)	0.081
$\mu_1(\cdot)$	0.020(0.130)	0.132	0.026(0.159)	0.161	0.029(0.165)	0.168	0.031(0.186)	0.189	0.035(0.222)	0.225
$\mu_2(\cdot)$	0.033(0.180)	0.183	0.040(0.230)	0.233	0.042(0.256)	0.259	0.047(0.272)	0.276	0.053(0.306)	0.310
$\mu_3(\cdot)$	0.049(0.184)	0.191	0.057(0.257)	0.263	0.059(0.251)	0.258	0.061(0.279)	0.285	0.071(0.293)	0.302
$\phi_{11}(\cdot)$	0.020(0.070)	0.073	0.022(0.088)	0.090	0.021(0.114)	0.116	0.020(0.138)	0.140	0.021(0.158)	0.159
$\phi_{12}(\cdot)$	0.013(0.083)	0.084	0.015(0.100)	0.101	0.018(0.135)	0.136	0.023(0.144)	0.146	0.023(0.165)	0.167
$\phi_{21}(\cdot)$	0.037(0.109)	0.115	0.040(0.107)	0.114	0.042(0.120)	0.127	0.045(0.160)	0.166	0.044(0.164)	0.170
$\phi_{22}(\cdot)$	0.014(0.115)	0.116	0.021(0.141)	0.142	0.029(0.152)	0.155	0.034(0.192)	0.195	0.082(0.308)	0.319
$\phi_{31}(\cdot)$	0.035(0.106)	0.112	0.042(0.118)	0.125	0.081(0.155)	0.175	0.085(0.173)	0.193	0.095(0.177)	0.201
$\phi_{32}(\cdot)$	0.023(0.107)	0.110	0.030(0.108)	0.112	0.034(0.117)	0.122	0.081(0.118)	0.144	0.091(0.134)	0.162
$\Sigma_1(\cdot,\cdot)$	0.061(0.267)	0.274	0.070(0.268)	0.277	0.082(0.276)	0.288	0.089(0.285)	0.298	0.095(0.461)	0.471
$\Sigma_2(\cdot,\cdot)$	0.072(0.271)	0.280	0.081(0.304)	0.314	0.092(0.297)	0.311	0.098(0.313)	0.328	0.101(0.398)	0.410
$\Sigma_3(\cdot,\cdot)$	0.073(0.293)	0.302	0.101(0.311)	0.327	0.114(0.353)	0.371	0.133(0.366)	0.389	0.142(0.372)	0.398

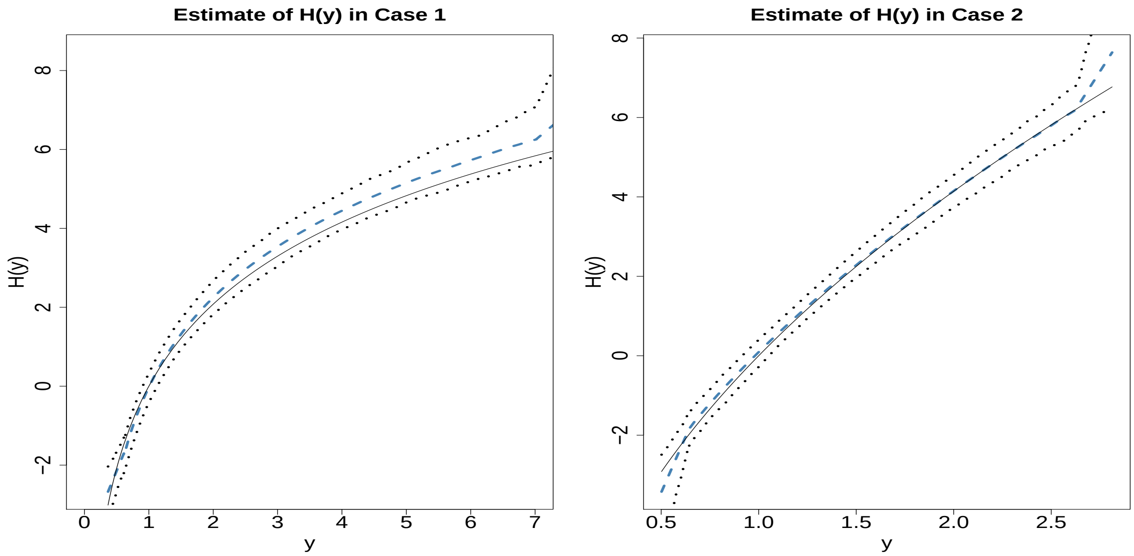

Figure 3: Estimated transformation functions for Cases 1 and 2 of Simulation 1 (solid–true function; dotted–95% confidence limits; dashed–average of the estimated function). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

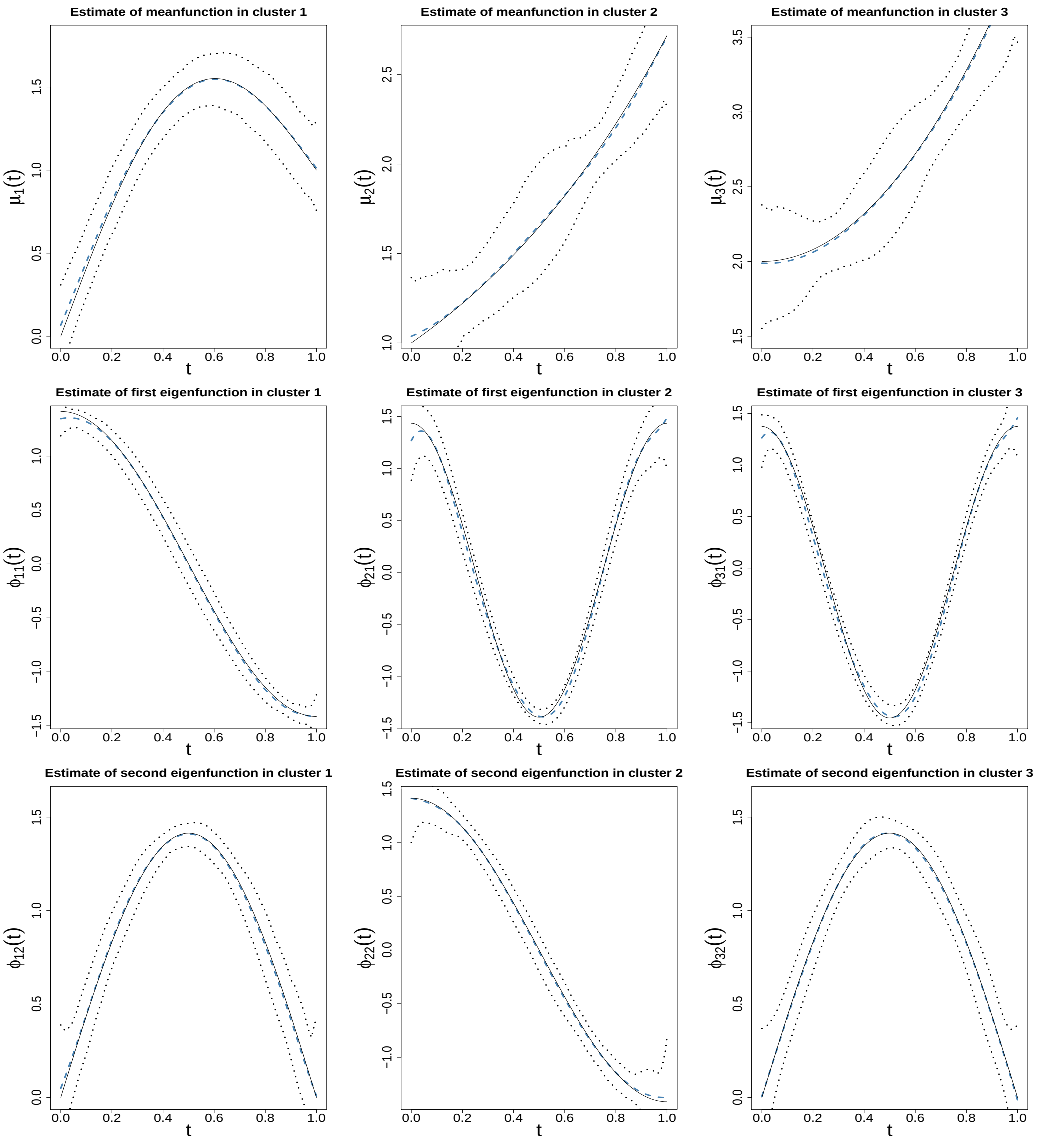

Figure 4: Estimated mean functions (top) and eigenfunctions (median and bottom) for Case 1 of Simulation 1 (solid–true function; dotted–95% confidence limits; dashed–average of the estimated function). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

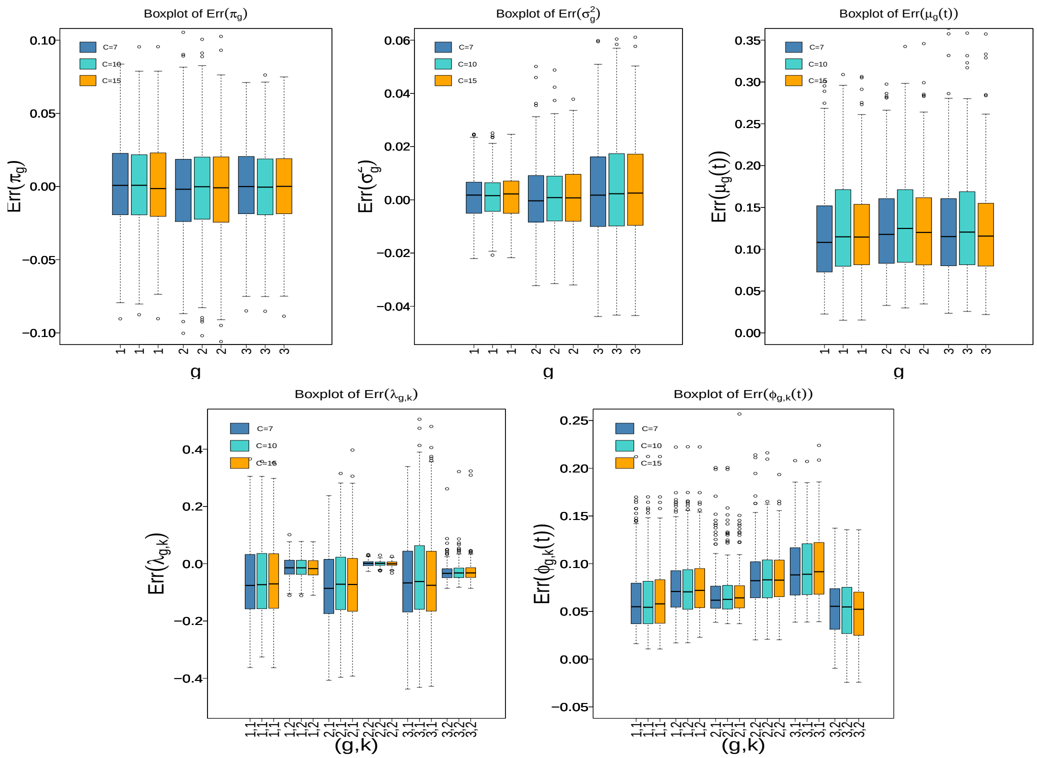

Figure 5: Comparison of the estimation error with different initial numbers of clusters, $C = 7, 10, 15,$ from left to right for Case 1 of Simulation 1. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

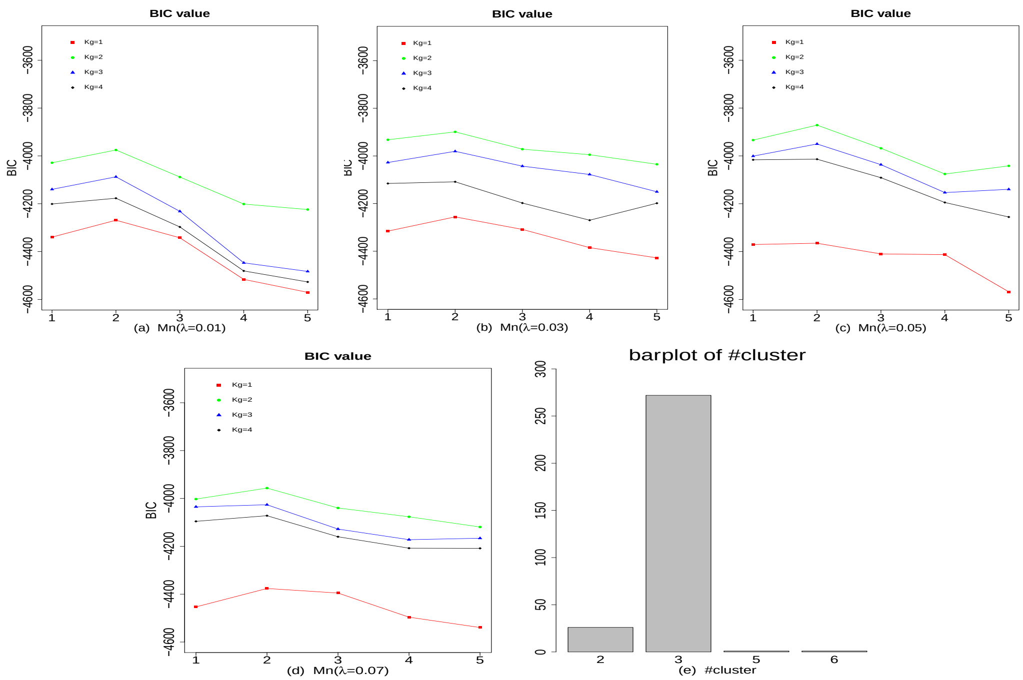

Figure 6: (a)- (d) The BIC values under various K_g, M_n and λ; (e) The barplot of estimated number of clusters based on the BIC criterion for Case 1 of Simulation 1. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

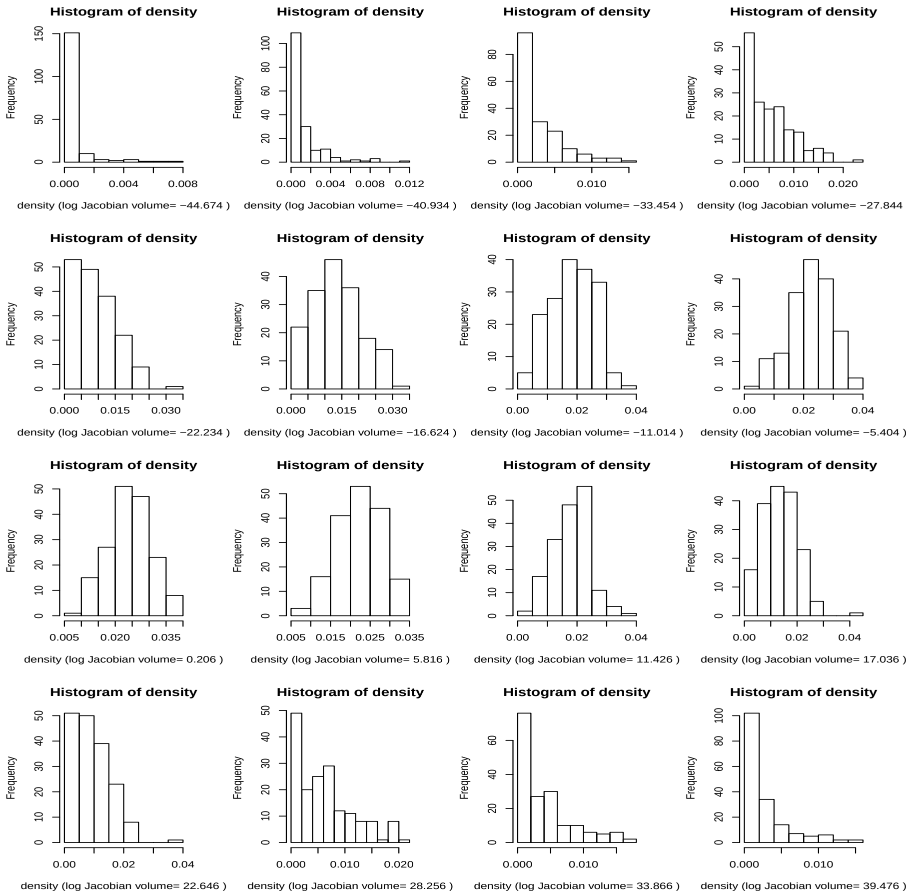

Figure 7: Histogram of density for cluster 1.

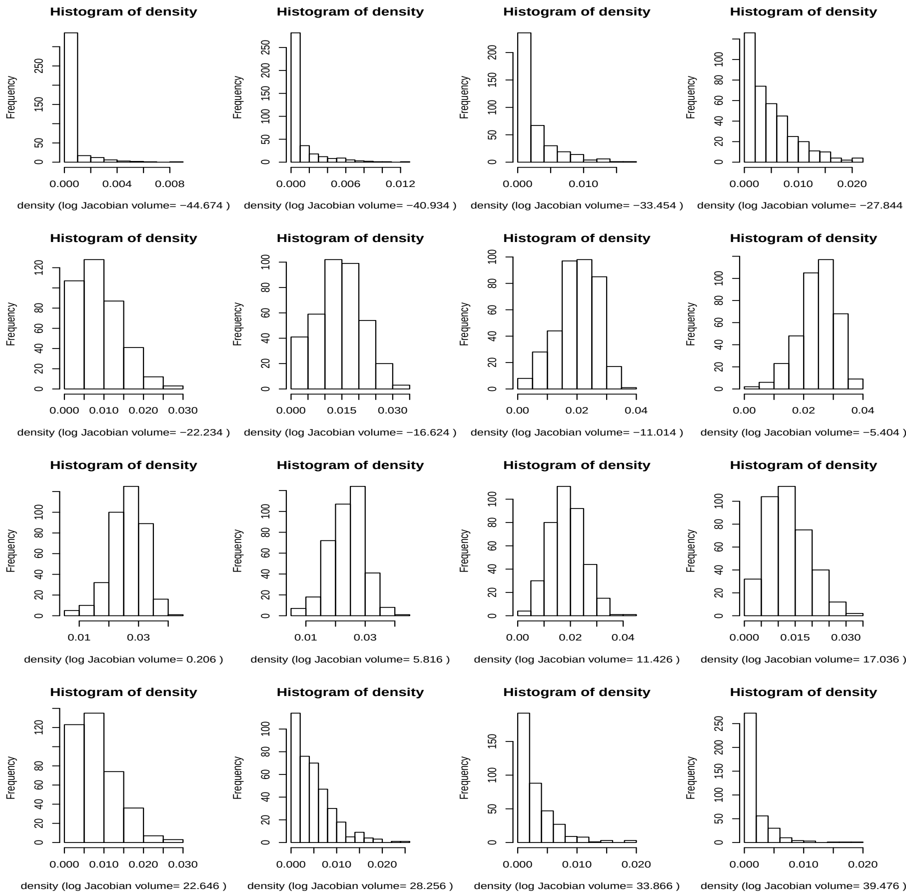

Figure 8: Histogram of density for cluster 2.

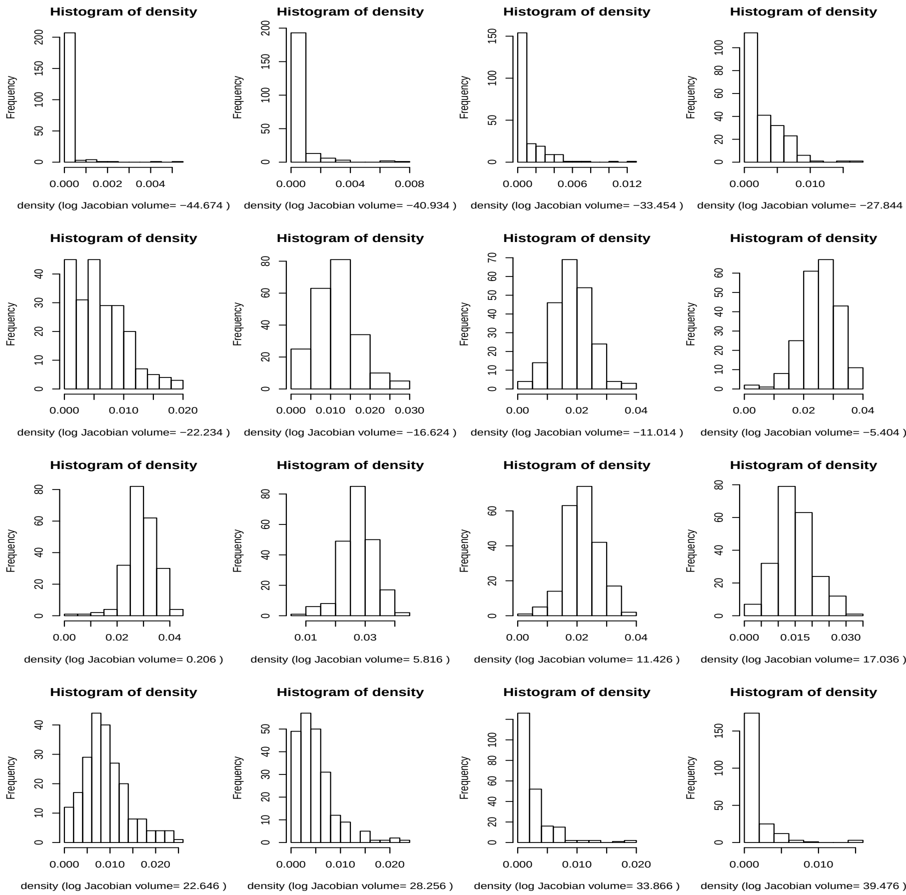

Figure 9: Histogram of density for cluster 3.

References

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, **101**, 1228-1240.

Chandra, T. K. (2012). The Borel-Cantelli Lemma. Springer Science & Business Media.

Flett, T. M. (1958). A mean value theorem. *The Mathematical Gazette*, **42**, 38-39.

Huang, T., Peng, H., and Zhang, K.(2017). Model selection for gaussian mixture models. *Statistica Sinica*, **27**, 147-169.

Pollard, D. (1984). Convergence of Stochastic Processes. Springer, New York.

Schumaker, L. L. Spline functions: basic theory. 1981. John Wiley & Sons, New York.

Shen, X., and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics*, **22**, 580-615.

Van de Geer, S. (2000). Empirical Processes in M-estimation. Cambridge University Press.

Van der Vaart, A. W. (1998). Asymptotic statistics (Vol. 3). Cambridge university press.

Van der Varrt, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes. Springer, New York.