Supplemental Materials for "Cluster Analysis with Regression of Non-Gaussian Functional Data on Covariates"

A: Notation

Let

$$
\begin{aligned}
&Pf = \int f(x)dP(x),\ \ P_n f = \frac{1}{n}\sum_{i=1}^{n} f(x_i),\\
&\ell_i(\boldsymbol{\theta}_n; H) = \log\left\{\sum_{k=1}^{K}\pi_k f_{nk}(H(\mathbf{y}_i)|\mathbf{X}_i, \mathbf{t}_i)\right\},\ \ \tilde{\ell}_i(\boldsymbol{\theta}_n; H) = \ell_i(\boldsymbol{\theta}_n; H) - \lambda\sum_{k=1}^{K}\log\{\frac{\epsilon+\pi_k}{\epsilon}\},\\
&\mathcal{G}_n = \left\{\boldsymbol{\alpha}' B_n(t) : \boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{q_n})' \in R^{q_n}, \max_{1\le i\le q_n}|\alpha_i| \le L, t\in[0,1]\right\},\\
&\boldsymbol{\Theta}_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \mathbf{g}) \in R^{Kp}\otimes[0,1]^K\otimes R^{Km}\otimes\mathcal{G}_n^K, \|\boldsymbol{\beta}\| + \|\boldsymbol{\pi}\| + \|\boldsymbol{\gamma}\| \le M\},\\
&Q_n(\boldsymbol{\theta}_n; H) = \sum_{i=1}^{n}\ell_i(\boldsymbol{\theta}_n; H) - n\lambda\sum_{k=1}^{K}\log\left\{\frac{\epsilon+\pi_k}{\epsilon}\right\},\\
&Q_n^*(\boldsymbol{\theta}_n; H) = Q_n(\boldsymbol{\theta}_n; H) - \lambda_1(\sum_{k=1}^{K}\pi_k - 1),\\
&\psi_{ij}(H(y); \boldsymbol{\theta}_n) = I\left(Y_{ij}\le y\right) - \sum_{k=1}^{K}\pi_k\Phi\left\{\frac{H(y) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}}\right\},\\
&\Psi_n(H(y); \boldsymbol{\theta}_n) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\psi_{ij}(H(y); \boldsymbol{\theta}_n),\\
&\Psi(H(y); \boldsymbol{\theta}_n) = \sum_{k=1}^{K} E\Bigg[\pi_{k0}\Phi\left(\frac{H_0(y) - g_{k0}(t_{ij}) - X_{i(j)}\beta_{k0}}{\sqrt{\sigma_{kj0}}}\right)\\
&\qquad\qquad\qquad - \pi_k\Phi\left(\frac{H(y) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}}\right)\Bigg],
\end{aligned}
$$

$\Upsilon = \{\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \pi_k, k = 1, \cdots, s_0\}$ and $\widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)$ is the estimator of $H(y)$ given $\boldsymbol{\theta}_n$ and is defined by Equation (15) in the main paper.

Four Essential Lemmas

Before presenting the lemmas, we first define the covering number of the class $\mathcal{L}_n = \{\tilde{\ell}\big(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)\big) : \boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n\}$. In particular, for any $\epsilon > 0$ define the covering number $N(\epsilon, \mathcal{L}_n, L_1(P_n))$ as the smallest value of κ for which there exist $\{\boldsymbol{\theta}_{n,j} \in \boldsymbol{\Theta}_n, j = 1, \cdots, \kappa\}$ such that

$$\min_{j\in\{1,\cdots,\kappa\}} \frac{1}{n}\sum_{i=1}^{n} |\tilde{\ell}_i\big(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)\big) - \tilde{\ell}_i(\boldsymbol{\theta}_{n,j}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_{n,j}))| < \epsilon,$$

for all $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$. If no such κ exists, define $N(\epsilon, \mathcal{L}_n, L_1(P_n)) = \infty$. We remark that our theory relies on modern empirical process theory (Van der Varrt & Weller, 1996). For the situation that we consider here, to establish the asymptotic normality we employ the Riesz representation theorem. Next we state and also prove four lemmas.

Lemma 1 *The covering number of the class $\boldsymbol{\Theta}_n$ satisfies*

$$N(\epsilon, \boldsymbol{\Theta}_n, L_2) \prec \epsilon^{-(q_n+p+m+1)K},$$

where the sign " $\prec$ " indicates that the function on its left-hand side is bounded by a positive constant times the function on its right-hand side.

Proof: Applying Lemma 2.5 and Corollary 2.6 in Van de Geer (2000), we can complete the proof.

Lemma 2 *Under conditions (A1)-(A5), $\widehat{H}_n(y;\boldsymbol{\theta}_n)$ which is defined in Equation (15) in the main paper satisfies satisfies*

$$\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n, y\in[\underline{y},\bar{y}]} |\widehat{H}_n(y;\boldsymbol{\theta}_n) - H(y;\boldsymbol{\theta}_n)| \to 0,$$

where $H(y;\boldsymbol{\theta}_n)$ satisfies

$$\Psi(H(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n) = 0, \tag{S.1}$$

for given $y \in [\underline{y}, \bar{y}]$ and $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$.

Proof: It follows from the Law of Large Numbers and the monotonicity of $H_0(y)$ that for given $\zeta \geq 0$, $y \in [\underline{y}, \bar{y}]$, $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$,

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij} \leq y\right) - \sum_{k=1}^{K}\pi_k \Phi\left\{\frac{H_0(y) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}} - \zeta\right\}\right]$$
$$\to E\sum_{k=1}^{K}\left\{\pi_{k0}\Phi\left(\frac{H_0(y) - g_{k0}(t_{ij}) - X_{i(j)}\beta_{k0}}{\sqrt{\sigma_{kj0}}}\right)\right.$$
$$\left. -\pi_k \Phi\left(\frac{H_0(y) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}} - \zeta\right)\right\}, \qquad \text{(S.2)}$$

almost surely as $n \to \infty$, where $N = \sum_{i=1}^{n} n_i$.

Then, we show that Equation ((S.2) holds uniformly on $y \in [\underline{y}, \bar{y}]$ and $\theta_n \in \boldsymbol{\Theta}_n$. By Lemma 1 and Theorem 19.4 of Van der Varrt(1998), $\boldsymbol{\Theta}_n$ are P-Glivenko-Cantelli class. Since $\pi_k\Phi\left\{\frac{H_0(y)-\boldsymbol{\alpha}_k' B_n(t_{ij})-X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}} - \zeta\right\}$ is a continuous function on $\boldsymbol{\Theta}_n$ and is bounded by 1. Moreover the indicator function class $\{I\left(Y_{ij} \leq y\right)\}$ also belongs to the VC class; thus, the uniform convergence of the result in Equation (S.2) follows from Van de Geer (2000).

Furthermore, it also follows from Equation (S.2) that for large ζ,

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij} \leq y\right) - \sum_{k=1}^{K}\pi_k \Phi\left\{\frac{H_0(y) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}} - \zeta\right\}\right] > 0, \quad \text{(S.3)}$$
$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij} \leq y\right) - \sum_{k=1}^{K}\pi_k \Phi\left\{\frac{H_0(y) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}} + \zeta\right\}\right] < 0. \quad \text{(S.4)}$$

Together with the monotonicity and continuity of Φ, these results imply that there exists a unique $\widehat{H}_n(y;\boldsymbol{\theta}_n)$ such that

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[I\left(Y_{ij} \leq y\right) - \sum_{k=1}^{K}\pi_k \Phi\left\{\frac{\widehat{H}_n(y;\boldsymbol{\theta}_n) - \boldsymbol{\alpha}_k' B_n(t_{ij}) - X_{i(j)}\boldsymbol{\beta}_k}{\sqrt{\sigma_{kj}}}\right\}\right] = 0, \quad \text{(S.5)}$$

for given y and $\boldsymbol{\theta}_n$. Similarly, there is a unique function $H(y;\boldsymbol{\theta}_n)$ that satisfies Equation (S.1) for given y and $\boldsymbol{\theta}_n$. Note that

$$\Psi_n\big(\widehat{H}_n(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big) = \{\Psi_n\big(\widehat{H}_n(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big) - \Psi_n\big(H(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big)\} + \Psi_n\big(H(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big),$$

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

$$\Psi_n\big(H(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big) \to \Psi(H(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n) = 0,$$

almost surely uniformly in $y \in [\underline{y}, \bar{y}]$ and $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$. Then it follows that

$$0 = \|\Psi_n\big(\widehat{H}_n(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big)\| \geq C\|\widehat{H}_n(y;\boldsymbol{\theta}_n) - H(y;\boldsymbol{\theta}_n)\| - \xi_n, \tag{S.6}$$

where $C > 0$ does not depend on y, and

$$\xi_n = \sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n, y \in [\underline{y}, \bar{y}]} \|\Psi_n\big(H(y;\boldsymbol{\theta}_n);\boldsymbol{\theta}_n\big)\| \to 0.$$

Hence, Equation (S.6) implies that $\widehat{H}_n(y;\boldsymbol{\theta}_n)$ converges to $H(y;\boldsymbol{\theta}_n)$ uniformly in $y \in [\underline{y}, \bar{y}]$ and $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$.

Lemma 3 *Assume that Conditions (A1)-(A5) hold. Then the covering number of the class $\mathcal{L}_n$ satisfies*

$$N(\epsilon, \mathcal{L}_n, L_1(P_n)) \prec \epsilon^{-(q_n+p+m+1)K}.$$

Proof: For any $\boldsymbol{\theta}^{(1)} = \cup_{k=1}^K\{\beta_k^{(1)}, \pi_k^{(1)}, \gamma_k^{(1)}, g_k^{(1)}\} \in \boldsymbol{\Theta}_n$, $\boldsymbol{\theta}^{(2)} = \cup_{k=1}^K\{\beta_k^{(2)}, \pi_k^{(2)}, \gamma_k^{(2)}, g_k^{(2)}\} \in \boldsymbol{\Theta}_n$, we have

$$\begin{aligned} & P_n\tilde{\ell}\big(\boldsymbol{\theta}^{(1)};\widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}^{(1)})\big) - P_n\tilde{\ell}\big(\boldsymbol{\theta}^{(2)};\widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}^{(2)})\big) \\ = \;& P_n\big\{\tilde{\ell}\big(\boldsymbol{\theta}^{(1)};H(\,\cdot\,;\boldsymbol{\theta}^{(1)})\big) - \tilde{\ell}\big(\boldsymbol{\theta}^{(2)};H(\,\cdot\,;\boldsymbol{\theta}^{(2)})\big)\big\} \\ & +P_n\big\{\tilde{\ell}\big(\boldsymbol{\theta}^{(1)};\widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}^{(1)})\big) - \tilde{\ell}\big(\boldsymbol{\theta}^{(1)};H(\,\cdot\,;\boldsymbol{\theta}^{(1)})\big)\big\} \\ & -P_n\big\{\tilde{\ell}\big(\boldsymbol{\theta}^{(2)};\widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}^{(2)})\big) - \tilde{\ell}\big(\boldsymbol{\theta}^{(2)};H(\,\cdot\,;\boldsymbol{\theta}^{(2)})\big)\big\}. \end{aligned} \tag{S.7}$$

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

$$\begin{aligned} P_n\big\{\tilde{\ell}\big(\boldsymbol{\theta}^{(1)};\widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}^{(1)})\big) - \tilde{\ell}\big(\boldsymbol{\theta}^{(1)};H(\,\cdot\,;\boldsymbol{\theta}^{(1)})\big)\big\} &= o_p(1), \\ P_n\big\{\tilde{\ell}\big(\boldsymbol{\theta}^{(2)};\widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}^{(2)})\big) - \tilde{\ell}\big(\boldsymbol{\theta}^{(2)};H(\,\cdot\,;\boldsymbol{\theta}^{(2)})\big)\big\} &= o_p(1). \end{aligned} \tag{S.8}$$

Using a Taylor series expansion, under condition A(1) we obtain

$$
\begin{aligned}
&|\tilde{\ell}(\boldsymbol{\theta}^{(1)}; H(\,\cdot\,; \boldsymbol{\theta}^{(1)})) - \tilde{\ell}(\boldsymbol{\theta}^{(2)}; H(\,\cdot\,; \boldsymbol{\theta}^{(2)}))| \\
&\qquad \le c_0 \sum_{k=1}^{K} (\|\beta_k^{(1)} - \beta_k^{(2)}\| + \|\pi_k^{(1)} - \pi_k^{(2)}\| + \|\gamma_k^{(1)} - \gamma_k^{(2)}\| + \|g_k^{(1)} - g_k^{(2)}\|_\infty). \quad \text{(S.9)}
\end{aligned}
$$

Let $\boldsymbol{\alpha}_k^{(j)} = (\alpha_{k1}^{(j)}, \cdots, \alpha_{kq_n}^{(j)}), j = 1, 2$ be the spline coefficients of $g_k^{(j)}, j = 1, 2$, respectively. We have

$$
\|g_k^{(1)} - g_k^{(2)}\|_\infty \le \max_{1 \le i \le q_n} |\alpha_{ki}^{(1)} - \alpha_{ki}^{(2)}| := \|\boldsymbol{\alpha}_k^{(1)} - \boldsymbol{\alpha}_k^{(2)}\|_\infty. \qquad \text{(S.10)}
$$

By Combining (S.9) and (S.10) we obtain

$$
\begin{aligned}
&|\tilde{\ell}(\boldsymbol{\theta}^{(1)}; H(\,\cdot\,; \boldsymbol{\theta}^{(1)})) - \tilde{\ell}(\boldsymbol{\theta}^{(2)}; H(\,\cdot\,; \boldsymbol{\theta}^{(2)}))| \\
&\qquad \le C \sum_{k=1}^{K} (\|\boldsymbol{\beta}_k^{(1)} - \boldsymbol{\beta}_k^{(2)}\| + \|\pi_k^{(1)} - \pi_k^{(2)}\| + \|\boldsymbol{\gamma}_k^{(1)} - \boldsymbol{\gamma}_k^{(2)}\| + \|\boldsymbol{\alpha}_k^{(1)} - \boldsymbol{\alpha}_k^{(2)}\|_\infty). \text{(S.11)}
\end{aligned}
$$

Next, using Equations (S.7), (S.8), (S.11) together with Lemma 1, and mimicking the calculation on page 94 of Van der Varrt & Weller(1996), we have $N(\epsilon, \mathcal{L}_n, L_1(P_n)) \prec \epsilon^{-(q_n+p+m+1)K}$. This completes the proof.

Lemma 4 *Assume that Conditions (A1)-(A5) hold. Then we have*

$$
\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n \tilde{\ell}\big(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)\big) - P\tilde{\ell}\big(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)\big)| \to 0 \textit{ almost surely.}
$$

Proof: By lemma 2, we have $\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n\{\tilde{\ell}\big(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)\big) - \tilde{\ell}\big(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)\big)\}| \to 0$. And it is obvious that $P_n \tilde{\ell}(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)) - P\tilde{\ell}(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)) = P_n \ell(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)) - P\ell(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n))$. Then we need to prove

$$
\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n \ell(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)) - P\ell(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n))| \to 0
$$

. Note that $|\ell(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n))|$ is bounded under Conditions (A1)-(A4). So, without loss of generality, we assume $\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |\ell(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n))| \le 1$. Then $P\ell^2(\boldsymbol{\theta}_n; H(\,\cdot\,; \boldsymbol{\theta}_n)) \le$

$P(\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n}|\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n))|)^2 \le 1$. Let $\alpha_n = n^{-1/2+\phi_1}(\log n)^{1/2}$ with $\nu/2 < \phi_1 < 1/2$. Obviously $\{\alpha_n\}$ is a non-increasing sequence of positive numbers. Also for a given $\epsilon > 0$, let $\epsilon_n = \epsilon\alpha_n$. Then for sufficiently large n and any $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$, we have

$$\begin{aligned}\frac{\mathrm{var}(P_n\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)))}{(4\epsilon_n)^2} &\le \frac{(1/n)P\ell^2(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n))}{16\epsilon^2\alpha_n^2} \\ &\le \frac{1}{16n\epsilon^2\alpha_n^2} \le \frac{1}{16\epsilon^2 n^{2\phi_1}\log n} \le \frac{1}{2}.\end{aligned}$$

Let P_n^o denote the signed measure that places mass $\pm\frac{1}{n}$ at each of the observations $\{O_1,\cdots,O_n\}$, with the plus and minus signs themselves independent of the O_i. Then from page 31 of Pollard (1984) and $\mathrm{var}(P_n\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)))/(4\epsilon_n)^2 \le 1/2$, the following symmetrization inequality holds:

$$\begin{aligned}&P(\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n}|P_n\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - P\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n))| > 8\epsilon_n) \\ &\qquad \le 4P(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_n}|P_n^o\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n))| > 2\epsilon_n). \qquad \text{(S.12)}\end{aligned}$$

Let $\mathcal{O} = \{O_1,\cdots,O_n\}$ represent the observed data. Given $\mathcal{O}$, select $(\boldsymbol{\theta}_n^{(1)},\cdots,\boldsymbol{\theta}_n^{(\kappa)})$, where $\kappa = N(\epsilon_n/2,\mathcal{L}_n,L_1(P_n))$, such that

$$\min_{j\in\{1,\cdots,\kappa\}} P_n|\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - \ell(\boldsymbol{\theta}_n^{(j)};H(y;\boldsymbol{\theta}_n^{(j)}))| \le \frac{\epsilon_n}{2},$$

for all $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$. For each $\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n$, let

$$\boldsymbol{\theta}_n^* = \arg\min_{\boldsymbol{\theta}_n^{(j)}} P_n|\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - \ell(\boldsymbol{\theta}_n^{(j)};H(y;\boldsymbol{\theta}_n^{(j)}))|.$$

Note that

$$\begin{aligned}&|P_n^o\big(\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - \ell(\boldsymbol{\theta}_n^*;H(y;\boldsymbol{\theta}_n^*))\big)| \\ =\ & \Big|\frac{1}{n}\sum_{i=1}^n \pm(\ell_i(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - \ell_i(\boldsymbol{\theta}_n^*;H(y;\boldsymbol{\theta}_n^*)))\Big| \\ \le\ & \frac{1}{n}\sum_{i=1}^n \big|(\ell_i(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - \ell_i(\boldsymbol{\theta}_n^*;H(y;\boldsymbol{\theta}_n^*)))\big| \\ =\ & P_n\big|\ell(\boldsymbol{\theta}_n;H(\cdot;\boldsymbol{\theta}_n)) - \ell(\boldsymbol{\theta}_n^*;H(y;\boldsymbol{\theta}_n^*))\big|. \qquad \text{(S.13)}\end{aligned}$$

Then by the definition of $\boldsymbol{\theta}_n^*$ and (S.13) we have

$$\begin{aligned}
& P(\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n} |P_n^o \ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| > 2\epsilon_n|\mathcal{O}) \\
\leq\ & P(\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n} \big[|P_n^o \ell(\boldsymbol{\theta}_n^*; H(y;\boldsymbol{\theta}_n^*))| + P_n^o|\ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n)) - \ell(\boldsymbol{\theta}_n^*; H(y;\boldsymbol{\theta}_n^*))|\big] > 2\epsilon_n|\mathcal{O}) \\
\leq\ & P(\max_j |P_n^o \ell(\boldsymbol{\theta}_n^{(j)}; H(y;\boldsymbol{\theta}_n^{(j)}))| > \frac{3\epsilon_n}{2}|\mathcal{O}) \\
\leq\ & N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n)) \max_j P(|P_n^o \ell(\boldsymbol{\theta}_n^{(j)}; H(y;\boldsymbol{\theta}_n^{(j)}))| > \frac{3\epsilon_n}{2}|\mathcal{O}). \qquad \text{(S.14)}
\end{aligned}$$

According to the definition of the covering number $N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n))$, for each $\boldsymbol{\theta}_n^{(j)}$, there exists $\check{\boldsymbol{\theta}}_n^{(j)}$ such that $P_n|\ell(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)})) - \ell(\boldsymbol{\theta}_n^{(j)}; H(y;\boldsymbol{\theta}_n^{(j)}))| \leq \frac{\epsilon_n}{2}$. Therefore

$$\begin{aligned}
& P(|P_n^o \ell(\boldsymbol{\theta}_n^{(j)}; H(y;\boldsymbol{\theta}_n^{(j)}))| > \frac{3\epsilon_n}{2}|\mathcal{O}) \\
\leq\ & P([|P_n^o \ell(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)}))| \\
& \qquad + P_n|\ell(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)})) - \ell(\boldsymbol{\theta}_n^{(j)}; H(y;\boldsymbol{\theta}_n^{(j)}))|] > \frac{3\epsilon_n}{2}|\mathcal{O}) \\
\leq\ & P(|P_n^o \ell(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)}))| > \epsilon_n|\mathcal{O}). \qquad \text{(S.15)}
\end{aligned}$$

By Hoeffding's inequality, we have

$$\begin{aligned}
P(|P_n^o \ell(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)}))| > \epsilon_n|\mathcal{O}) &= P(|\sum_{i=1}^n \pm \ell_i(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)}))| > n\epsilon_n|\mathcal{O}) \\
&\leq 2\exp\Big[-2(n\epsilon_n)^2 / \sum_{i=1}^n (2\ell_i(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)})))^2\Big] \\
&\leq 2\exp(-n\epsilon_n^2/2). \qquad \text{(S.16)}
\end{aligned}$$

The last inequality in Equation (S.16) holds because $|\ell_i(\check{\boldsymbol{\theta}}_n^{(j)}; H(y;\check{\boldsymbol{\theta}}_n^{(j)}))| \leq 1$. Combining the the results found in Equations (S.14)-(S.16) and Lemma 3, we obtain

$$\begin{aligned}
P(\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n} |P_n^o \ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| > 2\epsilon_n|\mathcal{O}) &\leq 2N(\epsilon_n/2, \mathcal{L}_n, L_1(P_n)) \exp(-n\epsilon_n^2/2) \\
&\leq C(\epsilon_n/2)^{-(q_n+p+m+1)K} \exp(-n\epsilon_n^2/2).
\end{aligned}$$

Note that the right-hand side does not depend on the observations $\mathcal{O}$; then, by taking expectations over $\mathcal{O}$, we have

$$P(\sup_{\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_n} |P_n^o \ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| > 2\epsilon_n) \leq C(\epsilon_n/2)^{-(q_n+p+m+1)K} \exp(-n\epsilon_n^2/2). \qquad \text{(S.17)}$$

Combining Equations (S.12) and (S.17) we obtain

$$
\begin{aligned}
& P(\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n \ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n)) - P\ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| > 8\epsilon_n) \\
& \quad \leq \; 4P(\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n^o \ell(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| > 2\epsilon_n) \\
& \quad \leq \; (\epsilon_n/2)^{-(q_n+p+m+1)K} \exp(-n\epsilon_n^2/2) \\
& \quad \leq \; C\exp(-C\epsilon^2 n^{2\phi_1}\log n).
\end{aligned}
$$

Hence $\sum_{n=1}^{\infty} P(\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n\tilde{\ell}(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n)) - P\tilde{\ell}(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| > 8\epsilon_n) < \infty$. By the Borel-Cantelli lemma, we will have $\sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n\tilde{\ell}(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n)) - P\tilde{\ell}(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))| \to 0$ almost surely, which completes the proof of Lemma 4.

Proofs of Theorems 1–3

Proof of Theorem 1. Our proof of Theorem 1 relies on three steps. The first step establishes the consistency of $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{H}_n(y;\widehat{\boldsymbol{\theta}}_n)$. Next we determine the rate of convergence of $\widehat{\boldsymbol{\theta}}_n$. Finally, by utilizing the two previous steps we obtain the selection consistency for the cluster number.

Step 1 (Consistency). Under condition (A2) and by Corollary 6.21 of Schumaker (1981), there exist $g_{nk0} = \boldsymbol{\alpha}_{k0}' B_n(t)$ such that

$$\sup_{t\in[0,1]} |g_{nk0}(t) - g_{k0}(t)| = O(q_n^{-r}) = O(n^{-rv}),$$

where $g_{k0}(\cdot)$ denotes the true function of $g_k(\cdot)$, $k = 1, \cdots, K$. Let $\boldsymbol{\theta}_{n0} = (\boldsymbol{\beta}_0, \boldsymbol{\pi}_0, \boldsymbol{\gamma}_0, \mathbf{g}_{n0})$, where $\mathbf{g}_{n0} = (g_{1n0}, \cdots, g_{Kn0})$. Then,

$$d(\boldsymbol{\theta}_{n0}, \boldsymbol{\theta}_0) = O(n^{-rv}). \tag{S.18}$$

Let $M(\boldsymbol{\theta}_n; \widehat{H}_n(y;\boldsymbol{\theta}_n)) = -\tilde{\ell}(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n))$, $K_\epsilon = \{\boldsymbol{\theta}_n : d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) \geq \epsilon, \boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n\}$ for

$\epsilon > 0$ and

$$\zeta_{1n} = \sup_{\boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n} |P_n M(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) - PM(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n))|,$$
$$\zeta_{2n} = P_n M(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0})) - PM(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0})).$$

Then one can show that

$$\begin{aligned}
&\inf_{K_\epsilon} PM(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) \\
&\quad = \inf_{K_\epsilon} \{PM(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) - P_n M(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) + P_n M(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n))\} \\
&\quad \le \zeta_{1n} + \inf_{K_\epsilon} P_n M(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)). \qquad \text{(S.19)}
\end{aligned}$$

If $\widehat{\boldsymbol{\theta}}_n \in K_\epsilon$, then we have

$$\begin{aligned}
\inf_{K_\epsilon} P_n M(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) &= P_n M(\widehat{\boldsymbol{\theta}}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) \\
&\le P_n M(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0})) = \zeta_{2n} + PM(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0})) \qquad \text{(S.20)}
\end{aligned}$$

Let $\delta_\epsilon = \inf_{K_\epsilon} PM(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) - PM(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0}))$. One can easily verify that $\delta_\epsilon > 0$ under the conditions specified in Theorem 1 when n is large enough. By equations (S.19) and (S.20), we have

$$\begin{aligned}
\inf_{K_\epsilon} PM(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_n)) &\le \zeta_{1n} + \zeta_{2n} + PM(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0})) \\
&= \zeta_n + PM(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,; \boldsymbol{\theta}_{n0})),
\end{aligned}$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$, and hence $\zeta_n \ge \delta_\epsilon$ by the definition of δ_ϵ. Since $\{\widehat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subseteq \{\zeta_n \ge \delta_\epsilon\}$, then $\bigcup_{i=1}^\infty \bigcap_{n=i}^\infty \{\widehat{\boldsymbol{\theta}}_n \in K_\epsilon\} \subseteq \bigcup_{i=1}^\infty \bigcap_{n=i}^\infty \{\zeta_n \ge \delta_\epsilon\}$. By Lemma 2 and the strong law of large numbers, we have both $\zeta_{1n} \to 0$ and $\zeta_{2n} \to 0$ almost surely. Therefore, $\bigcup_{i=1}^\infty \bigcap_{n=i}^\infty \{\zeta_n \ge \delta_\epsilon\}$ is the null set when n is large enough, which proves that $d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) \to 0$ almost surely as $n \to \infty$. Combining this result with Equation (S.18), we have $d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \to 0$. Together with Lemma 2, we have $\widehat{H}_n(y) \equiv \widehat{H}_n(y; \widehat{\boldsymbol{\theta}}_n) \to H(y; \boldsymbol{\theta}_0) \equiv H_0(y)$ uniformly in $y \in [\underline{y}, \bar{y}]$.

Step 2 (Rate of convergence). We establish the convergence rate of $\widehat{\boldsymbol{\theta}}_n$ by using Theorem 3.4.1 of Van der Varrt & Weller (1996). For any $\eta > 0$, define the class of functions

$$\mathcal{F}_\eta = \{\tilde{\ell}(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)) - \tilde{\ell}(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_{n0})) : \boldsymbol{\theta}_n \in \boldsymbol{\Theta}_n, \eta/2 \le d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n0}) \le \eta\}.$$

For $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}_0$, the compactness of the parameter spaces implies that $P\{\tilde{\ell}(\boldsymbol{\theta}_0; H_0) - \tilde{\ell}(\boldsymbol{\theta}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}))\} \asymp d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. Hence

$$P(\tilde{\ell}(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_{n0})) - \tilde{\ell}(\boldsymbol{\theta}_0; H_0)) \asymp d^2(\boldsymbol{\theta}_{n0}, \boldsymbol{\theta}_0) \le Cn^{-2rv}. \tag{S.21}$$

Hence, for large n, by equation (S.21) we have

$$P(\tilde{\ell}(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)) - \tilde{\ell}(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_{n0}))) \;\; \le \;\; C\eta^2 + Cn^{-2rv} = O_p(\eta^2),$$

for any $\tilde{\ell}(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)) - \tilde{\ell}(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_{n0})) \in \mathcal{F}_\eta$. Following the calculations found on p.597 of Shen & Wong (1994), we can establish that for $0 < \varepsilon < \eta$, $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, L_2(P)) \le Cq_n \log(\eta/\varepsilon)$. Under Conditions (A1)-(A5), it is easy to see that $\mathcal{F}_\eta$ is uniformly bounded. Therefore, by Lemma 3.4.2 of Van der Varrt & Weller (1996), we obtain

$$E_P\|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \le CJ_{[]}(\eta, \mathcal{F}_\eta, L_2(P))\bigg\{1 + \frac{J_{[]}(\eta, \mathcal{F}_\eta, L_2(P))}{\eta^2\sqrt{n}}\bigg\},$$

where $J_{[]}(\eta, \mathcal{F}_\eta, L_2(P)) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, L_2(P))\}^{\frac{1}{2}} d\varepsilon \le C\sqrt{q_n}\eta$. Let $\phi_n(\eta) = \sqrt{q_n}\eta + q_n/\sqrt{n}$. It is easy to see that $\phi_n(\eta)/\eta$ is decreasing in η, and $r_n^2\phi_n(1/r_n) = r_n\sqrt{q_n} + r_n^2 q_n/n^{1/2} \le Cn^{1/2}$,where $r_n = N^{-1/2}n^{1/2} = n^{(1-v)/2}$. Noting that $P_n(\tilde{\ell}(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)) - \tilde{\ell}(\boldsymbol{\theta}_{n0}; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_{n0}))) \ge 0$ and $d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) \le d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) + d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{n0}) \to 0$ in probability, by applying Theorem 3.4.1 of Van der Varrt & Weller (1996), we have $n^{(1-v)/2} d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{n0}) = O_P(1)$. This result, together with $d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{n0}) = O^{-rv}$, shows that $d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_P(n^{-(1-v)/2} + n^{-rv}) = O(n^{-\min(\frac{1-v}{2}, rv)})$.

Step 3 (Selection consistency). In the final step, according to the proof of Huang et al.(2017), we need only consider the maximizer of $Q_n(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n))$ with

$d(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) \le C_1/n^{\min(\frac{1-v}{2}, rv)}$ and $\widehat{\pi}_k < 1/(\sqrt{n}\log n)$ for $k > s_0$. We use a Lagrange multiplier λ_1 to account for the constraint $\sum_{k=1}^K \widehat{\pi}_k = 1$. Note $Q_n^*(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)) = Q_n(\boldsymbol{\theta}_n; \widehat{H}_n(\,\cdot\,;\boldsymbol{\theta}_n)) - \lambda_1(\sum_{k=1}^K \pi_k - 1)$. By Lemma 2, it is then sufficient to show that

$$\left.\frac{\partial Q_n^*(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))}{\partial \pi_k}\right|_{\pi=\widehat{\boldsymbol{\pi}}} < 0 \quad \text{for} \quad \widehat{\pi}_k < \frac{1}{\sqrt{n}\log n}, \tag{S.22}$$

with probability tending to one. To show equation (S.22),

$$\left.\frac{\partial Q_n^*(\boldsymbol{\theta}_n; H(\,\cdot\,;\boldsymbol{\theta}_n))}{\partial \pi_k}\right|_{\pi=\widehat{\boldsymbol{\pi}}} = \sum_{i=1}^n \frac{f_{nk}}{\sum_{k=1}^K \widehat{\pi}_k f_{nk}} - n\lambda \frac{1}{\varepsilon + \widehat{\pi}_k} - \lambda_1 = 0. \tag{S.23}$$

Obviously, the first term in the equation above is of order $O_p(n)$ by the law of large numbers. Given $k < s_0$, it is easy to establish that $\widehat{\pi}_k = \pi_{k0} + O_p(1/n^{\min(\frac{1-v}{2}, rv)}) > \frac{1}{2} \cdot \min\{\pi_{10}, \pi_{20}, \cdots, \pi_{s_0 0}\}$. Then the second term should be $O_p(n\lambda) = o_p(n)$, and moreover $\lambda_1 = O_p(n)$. Next, consider equation $(S.23)$ when $k > s_0$ and $\widehat{\pi}_k < \frac{1}{\sqrt{n}\log n}$. It is obvious that the first and third terms in Equation (S.23) are each of order $O_p(n)$. For the second term, because $\widehat{\pi}_k = O_p(\frac{1}{\sqrt{n}\log n})$, $\lambda\sqrt{n}\log n \to \infty$ and $\epsilon = o(\frac{1}{\sqrt{n}\log n})$, we have

$$n\lambda \frac{1}{\varepsilon + \widehat{\pi}_k}/n = \lambda \frac{1}{\varepsilon + \widehat{\pi}_k} = O_p(\lambda\sqrt{n}\log n) \longrightarrow \infty,$$

with probability tending to one. Hence the second term in Equation (S.23) dominates the first and third terms. Therefore we have proved Equation (S.22), or equivalently $\widehat{\pi}_k = 0$ for $k > s_0$ with probability tending to one when $n \to \infty$.

Proof of Theorem 2. The proof of Theorem 2 is completed by following by step 1 and step 2 in the proof of Theorem 1.

Proof of Theorem 3. Let $\Theta - \boldsymbol{\theta}_0$ to be Θ excluding $\boldsymbol{\theta}_0$. Let Ω denote the linear span of $\Theta - \boldsymbol{\theta}_0$ and define the Fisher inner product on the space Ω as $< v, \check{v} >= P\{\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[v]\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[\check{v}]\}$ for $v, \check{v} \in \Omega$ and the Fisher norm as $\|v\| =< v, v >$, where $\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[v] = \left.\frac{d\tilde{\ell}(\boldsymbol{\theta}_0 + sv; H(\,\cdot\,;\boldsymbol{\theta}_0))}{ds}\right|_{s=0}$ is the first order directional derivative of $\ell(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))$ at the direction $v \in \Omega$ (evaluated at $\boldsymbol{\theta}_0$).

Also let $\bar{\Omega}$ be the closed linear span of Ω under the Fisher norm. Then $(\bar{\Omega}, \|\cdot\|)$ is a Hilbert space. For a vector of $s_0(p+m+1)$-dimension $b = (b_1', b_2', b_3')'$ with $\|b\| \leq 1$ and for any $v \in \Omega$, define a smooth functional of $\boldsymbol{\theta}$ as $h(\boldsymbol{\theta}) = b'\Upsilon = b_1'\boldsymbol{\beta} + b_2'\boldsymbol{\gamma} + b_3'\boldsymbol{\pi}$ and $\dot{h}(\boldsymbol{\theta}_0)[v] = \frac{dh(\boldsymbol{\theta}_0+sv)}{ds}\Big|_{s=0}$, where $\Upsilon = \{\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k, \pi_k, k = 1, \cdots, s_0\}$, whenever the limit on the right-hand side is well-defined. According to the Riesz representation theorem, there exists $v^* \in \bar{\Omega}$ such that $\dot{h}(\boldsymbol{\theta}_0)[v] =< v, v^* >$ for all $v \in \bar{\Omega}$ and $\|v^*\| = \|\dot{h}(\boldsymbol{\theta}_0)\|$. Note that $h(\boldsymbol{\theta}) - h(\boldsymbol{\theta}_0) = \dot{h}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$. Thus according to Cramér-Wold device, in order to prove Theorem 3, it suffices to show that

$$\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > \xrightarrow{d} N(0, b'I^{-1}(\Upsilon_0)b), \tag{S.24}$$

since $b'\{(\widehat{\boldsymbol{\theta}}_n', \widehat{\pi}_n', \widehat{\gamma}_n')' - (\beta_0', \pi_0', \gamma_0')'\} = h(\widehat{\boldsymbol{\theta}}_n) - h(\boldsymbol{\theta}_0) = \dot{h}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) =< \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* >$. In fact, Equation (S.24) holds when $\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > \xrightarrow{d} N(0, \|v^*\|^2)$ and $\|v^*\|^2 = b'I^{-1}(\Upsilon_0)b$.

In the following, we Equation (S.24) holds, we prove $\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > \rightarrow_d N(0, \|v^*\|^2)$. According to the result of Corollary 6.21 in Schumaker (1981), there exists $\Pi_n v^* \in \boldsymbol{\Theta}_n - \boldsymbol{\theta}_0$ such that $\|\Pi_n v^* - v^*\| = O(n^{-rv})$. In addition, under the assumptions $r \geq 2$ and $1/2 > v > 1/4r$, we have $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$ where $\delta_n = n^{-\min\{(1-v)/2, rv\}}$. For any $\boldsymbol{\theta} \in \{\boldsymbol{\theta} \in \Theta : d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = O(\delta_n)\}$, define the first- and second-order directional derivative in the directions $v, \check{v}$ to be

$$\dot{\tilde{\ell}}(\boldsymbol{\theta}; H(\cdot; \boldsymbol{\theta}))[v] = \frac{d\tilde{\ell}(\boldsymbol{\theta} + sv, O)}{ds}|_{s=0},$$
$$\ddot{\tilde{\ell}}(\boldsymbol{\theta}; H(\cdot; \boldsymbol{\theta}))[v, \check{v}] = \frac{d^2\tilde{\ell}(\boldsymbol{\theta} + sv + \check{s}\check{v}, O)}{d\check{s}ds}\Big|_{s=0,\check{s}=0} = \frac{d\dot{\tilde{\ell}}(\boldsymbol{\theta} + \check{s}\check{v}, O)[\check{v}]}{d\check{s}}\Big|_{\check{s}=0}.$$

Define $r(\boldsymbol{\theta} - \boldsymbol{\theta}_0; H(\cdot; \boldsymbol{\theta}_0)) = \tilde{\ell}(\boldsymbol{\theta}; H(\cdot; \boldsymbol{\theta})) - \tilde{\ell}(\boldsymbol{\theta}_0; H(\cdot; \boldsymbol{\theta}_0)) - \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot; \boldsymbol{\theta}_0))(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ and let $\varepsilon_n = o(n^{-1/2})$. Then by the definition of $\widehat{\boldsymbol{\theta}}_n$ and $P\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot; \boldsymbol{\theta}_0))[\prod_n v^*] = 0$, we have

$$
\begin{aligned}
0 &\le P_n\{\tilde{\ell}(\widehat{\boldsymbol{\theta}}_n; H(\,\cdot\,;\widehat{\boldsymbol{\theta}}_n)) - \tilde{\ell}(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^*; H(\,\cdot\,;\boldsymbol{\theta}))\} \\
&= P_n\big\{r(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0)) + \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&- r(\widehat{\boldsymbol{\theta}}_n + \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0)) - \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))(\widehat{\boldsymbol{\theta}}_n + \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0)\big\} \\
&= \mp \varepsilon_n P_n \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[\Pi_n v^*] \\
&\qquad + (P_n - P)\big\{r(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0)) - r(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))\big\} \\
&\qquad + P\big\{r(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0)) - r(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))\big\} \\
&= \mp \varepsilon_n P_n \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[v^*] \mp \varepsilon_n P_n \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[\Pi_n v^* - v^*] \\
&\qquad + (P_n - P)\{r(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0)) - r(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))\} \\
&\qquad + P\{r(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0)) - r(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))\} \\
&= \mp \varepsilon_n P_n \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[v^*] \mp I_1 + I_2 + I_3. \qquad \text{(S.25)}
\end{aligned}
$$

We will investigate the asymptotic behavior of I_1, I_2, I_3. For I_1, it follows from Conditions (A1)-(A5), the Chebyshev inequality and $\|\Pi_n v^* - v^*\| = o(1)$ that

$$
I_1 = \varepsilon_n \times o_p(n^{-1/2}). \qquad \text{(S.26)}
$$

For I_2, due to the mean value theorem, we obtain the result that

$$
\begin{aligned}
I_2 &= (P_n - P)\{\tilde{\ell}(\widehat{\boldsymbol{\theta}}_n; H(\,\cdot\,;\boldsymbol{\theta}_n)) \\
&\qquad - \tilde{\ell}(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^*; H(\,\cdot\,;\boldsymbol{\theta}_n)) \pm \varepsilon_n \tilde{\ell}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[\varepsilon_n \Pi_n v^*]\} \\
&= \mp \varepsilon_n (P_n - P)\big[\{\dot{\tilde{\ell}}(\tilde{\boldsymbol{\theta}}; H(\,\cdot\,;\boldsymbol{\theta})) - \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))\}[\Pi_n v^*]\big], \qquad \text{(S.27)}
\end{aligned}
$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^*$. By Theorem 2.8.3 of Van der Varrt & Weller (1996), we know that $\{\dot{\ell}(\boldsymbol{\theta}_0; H(\,\cdot\,;\boldsymbol{\theta}_0))[\Pi_n v^*] : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = O_p(\delta_n)\}$ belongs to the Donsker class. Hence by Theorem 2.11.23 of Van der Varrt & Weller (1996), we obtain $I_2 = \varepsilon_n \times o_p(n^{-1/2})$.

Since

$$
\begin{aligned}
& P(r[\boldsymbol{\theta}-\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0)]) \\
= \; & P\{\tilde{\ell}(\boldsymbol{\theta}; H(\cdot;\boldsymbol{\theta})) - \tilde{\ell}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0)) - \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0))[\boldsymbol{\theta}-\boldsymbol{\theta}_0]\} \\
= \; & \frac{1}{2}P\{\ddot{\tilde{\ell}}(\tilde{\boldsymbol{\theta}}; H(\cdot;\boldsymbol{\theta}))[\boldsymbol{\theta}-\boldsymbol{\theta}_0, \boldsymbol{\theta}-\boldsymbol{\theta}_0] - \ddot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0))[\boldsymbol{\theta}-\boldsymbol{\theta}_0, \boldsymbol{\theta}-\boldsymbol{\theta}_0]\} \\
& + \frac{1}{2}P\ddot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0))[\boldsymbol{\theta}-\boldsymbol{\theta}_0, \boldsymbol{\theta}-\boldsymbol{\theta}_0] \\
= \; & \frac{1}{2}P\ddot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0))[\boldsymbol{\theta}-\boldsymbol{\theta}_0, \boldsymbol{\theta}-\boldsymbol{\theta}_0] + \varepsilon_n \times o_p(n^{-1/2}),
\end{aligned}
$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ and the last equation follows from a Taylor expansion, conditions A(1)-A(5) and $r \geq 2$, $1/2 > \upsilon > 1/4r$. Therefore

$$
\begin{aligned}
& I_3 = -\frac{1}{2}\{\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|^2 - \|\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^* - \boldsymbol{\theta}_0\|^2\} + \varepsilon_n \times o_p(n^{-1/2}) \\
& = \pm\varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \Pi_n v^* > + \frac{1}{2}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
& = \pm\varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \Pi_n v^* - v^* + v^* > + \frac{1}{2}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
& = \pm\varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > \pm \varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \Pi_n v^* - v^* > + \frac{1}{2}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
& = \pm\varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > + \frac{1}{2}\|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
& = \pm\varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > + \varepsilon_n \times o_p(n^{-1/2}),
\end{aligned} \tag{S.28}
$$

where the last equality holds due to the fact that $\delta_n\|\Pi_n v^* - v^*\| = o(n^{-1/2})$, the Cauchy-Schwartz inequality, and $\|\Pi_n v^*\|^2 \to \|v^*\|$. By (S.25), (S.26), (S.27), (S.28), By the results specified in Equations (S.25)–(S.28), combined with $P\ddot{\ell}(\boldsymbol{\theta}_0; H)[v^*] = 0$, we can establish that

$$
\begin{aligned}
0 \; \leq \; & P_n\{\tilde{\ell}(\widehat{\boldsymbol{\theta}}_n; H(\cdot;\boldsymbol{\theta}_n)) - \tilde{\ell}(\widehat{\boldsymbol{\theta}}_n \pm \varepsilon_n \Pi_n v^*; H(\cdot;\boldsymbol{\theta}))\} \\
= \; & \mp\varepsilon_n P_n \dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0))[v^*] \pm \varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > + \varepsilon_n \times o_p(n^{-1/2}) \\
= \; & \mp\varepsilon_n (P_n - P)\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H(\cdot;\boldsymbol{\theta}_0))[v^*] \pm \varepsilon_n < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > + \varepsilon_n \times o_p(n^{-1/2}).
\end{aligned}
$$

Therefore, we obtain $\mp\sqrt{n}(P_n - P)\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H)[v^*] \pm \sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* > + o_p(1) \geq 0$. By combining this result with the central limit theorem we have $\sqrt{n} < \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, v^* >= \sqrt{n}(P_n - P)\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H)[v^*] + o_p(1) \rightarrow N(0, \|v^*\|^2)$ and $\|v^*\|^2 = \|\dot{\tilde{\ell}}(\boldsymbol{\theta}_0; H)[v^*]\|^2$.

Now we calculate $\|v^*\|$. Rewrite $\Upsilon = (\beta', \gamma', \pi')' = (\Upsilon_1, \cdots, \Upsilon_{(p+m+1)s_0})$. For each component $\Upsilon_q, q = 1, 2, \cdots, (p+m+1)s_0$. Let $\psi_q^* = (b_{1q}^*, b_{2q}^*, \cdots, b_{s_0 q}^*)$ be the minimizer of $E\{\tilde{\ell}_\Upsilon \cdot e_q - \tilde{\ell}_{b_1}[b_{1q}] - \tilde{\ell}_{b_2}[b_{2q}] - \cdots - \tilde{\ell}_{b_{s_0}}[b_{s_0 q}]\}^2$ with respect to $\psi_q = (b_{1q}, b_{2q}, \cdots, b_{s_0 q})$, where $\tilde{\ell}_\Upsilon = (\tilde{\ell}'_\beta, \tilde{\ell}'_\gamma, \tilde{\ell}'_\pi)'$, $\tilde{\ell}_\beta = (\tilde{\ell}'_{\beta_1}, \cdots, \tilde{\ell}'_{\beta_{s_0}})'$, $\tilde{\ell}_\pi = (\tilde{\ell}_{\pi_1}, \cdots, \tilde{\ell}_{\pi_{s_0}})'$, $\tilde{\ell}_\gamma = (\tilde{\ell}'_{\gamma_1}, \cdots, \tilde{\ell}'_{\gamma_{s_0}})'$, $\tilde{\ell}_{\boldsymbol{\beta}_k} = \pi_k f_k(H(\mathbf{y})|\mathbf{x}) / \sum_{k=1}^{s_0} \pi_k f_k(H(\mathbf{y})|\mathbf{x})\mathbf{x}\Delta(\boldsymbol{\gamma}_k)^{-1}\{H(\mathbf{y}) - \mu_k\}$, $\tilde{\ell}_{\pi_k} = f_k(H(\mathbf{y})|\mathbf{x}) / \sum_{k=1}^{s_0} \pi_k f_k(H(\mathbf{y})|\mathbf{x}) - \frac{n\lambda}{\epsilon + \pi_k}$, $\tilde{\ell}_{\boldsymbol{\gamma}_k} = \pi_k / \sum_{k=1}^{s_0} \pi_k f_k(H(\mathbf{y})|\mathbf{x}) \frac{\partial f_k(H(\mathbf{y})|\mathbf{x})}{\partial \boldsymbol{\gamma}_k}$, $\tilde{\ell}_{b_k} = \pi_k f_k(H(\mathbf{y})|\mathbf{x}) / \sum_{k=1}^{s_0} \pi_k f_k(H(\mathbf{y})|\mathbf{x})\Delta(\boldsymbol{\gamma}_k)^{-1}\{H(\mathbf{y}) - \mu_k\}$, and e_q is a $(p+m+1)s_0$-dimensional vector of zeros except the q-th element equal to 1.

Define a vector S_Υ of dimension $(p+m+1)s_0$, with the q-th element as $\ell_\Upsilon \cdot e_q - \ell_{b_1}[b_{1q}^*] - \ell_{b_2}[b_{2q}^*] - \cdots - \ell_{b_{s_0}}[b_{s_0 q}^*]$, and $I(\Upsilon_0) = E(S_\Upsilon S'_\Upsilon)$. Furthermore, by following calculations similar to those found in Chen et al.(2006), we obtain

$$\|v^*\| = \|\dot{h}(\boldsymbol{\theta}_0)\| = \sup_{v \in \bar{V}: \|v\| > 0} \frac{|\dot{h}(\boldsymbol{\theta}_0)|^2}{v^2} = b'[E(S_\Upsilon S'_\Upsilon)]^{-1} b = b' I^{-1}(\Upsilon_0) b.$$

This completes the proof of Theorem 3.