Semiparametric Mixtures of Nonparametric Regressions

Sijia Xiang [*], Weixin Yao [†]

Proofs

In this section, the conditions required by Theorems 1, 2, 3 and 4 are listed. They are not the weakest sufficient conditions, but could easily facilitate the proofs. The proofs of Theorems 1, 2, 3 and 4 are also presented in this section.

Technical Conditions:

(C1) $nh^4 \to 0$ and $nh^2 \log(1/h) \to \infty$ as $n \to \infty$ and $h \to 0$.

(C2) $nh \to \infty$ as $n \to \infty$ and $h \to 0$.

(C3) The sample $\{(X_i, Y_i), i = 1, ..., n\}$ are independently and identically distributed from $f(x, y)$ with finite sixth moments. The support for x, denoted by $\mathscr{X} \in \mathbb{R}$, is bounded and closed.

(C4) $f(x, y) > 0$ in its support and has continuous first derivative.

(C5) $|\partial^3 \ell(\boldsymbol{\theta}, x, y)/\partial\theta_i\partial\theta_j\partial\theta_k| \leq M_{ijk}(x, y)$, where $E(M_{ijk}(x, y))$ is bounded for all i, j, k and all X, Y.

(C6) The unknown functions $m_j(x)$, $j = 1, ..., k$, have continuous second derivative.

(C7) $\sigma_j^2 > 0$ and $\pi_j > 0$ for $j = 1, ..., k$ and $\sum_{j=1}^k \pi_j = 1$.

(C8) $E(X^{2r}) < \infty$ for some $\epsilon < 1 - r^{-1}$, $n^{2\epsilon-1}h \to \infty$.

(C9) $I_\theta(x)$ and $I_m(x)$ are positive definite.

(C10) The kernel function $K(\cdot)$ is symmetric, continuous with compact support.

(C11) The marginal density $f(x)$ of X is Lipschitz continuous and bounded away from 0. X has a bounded support $\mathscr{X}$.

(C12) $t^3K(t)$ and $t^3K'(t)$ are bounded and $\int t^4K(t)dt < \infty$.

(C13) $E|q_\theta|^4 < \infty$, $E|q_m|^4 < \infty$, where q_θ and q_m are defined in the proof of Theorem 2.5.

[*]Corresponding author, School of Mathematics and Statistics, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang 310018, P. R. China. E-mail address: sjxiang@zufe.edu.cn.

[†]Department of Statistics, University of California, Riverside, CA 92887, U.S.A.

The next lemma is from Fan and Huang (2005), and will be used throughout the rest of the proofs.

Lemma 1. Let $\{(X_i, Y_i), i = 1, ..., n\}$ be i.i.d random vectors from (X, Y), where X is a random vector and Y is a scalar random variable. Let f be the joint density of (X, Y), and further assume that $E|Y|^r < \infty$ and $\sup_x \int |y|^r f(x, y)dy < \infty$. Let $K(\cdot)$ be a bounded positive function with bounded support, satisfying a Lipschitz condition. Then,

$$\sup_{x\in\mathscr{X}} \left| \frac{1}{n} \sum_{i=1}^{n} [K_h(X_i - x)Y_i - E\{K_h(X_i - x)Y_i\}] \right| = O_p(\gamma_n \log^{1/2}(1/h)),$$

given $n^{2\epsilon-1}h \to \infty$, for some $\epsilon < 1 - 1/r$, where $\gamma_n = (nh)^{-1/2}$.

In order to prove the asymptotic properties of $\{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\sigma}}^2\}$, we first need to study the asymptotic property of $\{\tilde{\boldsymbol{\pi}}, \tilde{\boldsymbol{m}}, \tilde{\boldsymbol{\sigma}}^2\}$, which is the maximum local log-likelihood estimator of (5).

Define

$$\tilde{\pi}_j^* = \sqrt{nh}\{\tilde{\pi}_j - \pi_j\},\ \tilde{m}_j^* = \sqrt{nh}\{\tilde{m}_j - m_j\},\ \tilde{\sigma}_j^{2*} = \sqrt{nh}\{\tilde{\sigma}_j^2 - \sigma_j^2\}.$$

Let $\tilde{\boldsymbol{\pi}}^* = (\tilde{\pi}_1^*, ..., \tilde{\pi}_{k-1}^*)^T$, $\tilde{\boldsymbol{m}}^* = (\tilde{m}_1^*, ..., \tilde{m}_k^*)^T$, and $\tilde{\boldsymbol{\sigma}}^{2*} = (\tilde{\sigma}_1^{2*}, ..., \tilde{\sigma}_k^{2*})^T$. Furthermore, define $\tilde{\boldsymbol{\theta}}^* = ((\tilde{\boldsymbol{m}}^*)^T, (\tilde{\boldsymbol{\pi}}^*)^T, (\tilde{\boldsymbol{\sigma}}^{2*})^T)^T$, $\boldsymbol{\beta} = ((\tilde{\boldsymbol{\pi}})^T, (\tilde{\boldsymbol{\sigma}}^{2*})^T)^T$.

Lemma 2. Suppose that conditions (C2)-(C10) are satisfied, then,

$$\sup_{x\in\mathscr{X}} \left| \tilde{\boldsymbol{\theta}}^* - f^{-1}(x) I_\theta^{-1}(x) S_n \right| = O_p(h^2 + \gamma_n \log^{1/2}(1/h)),$$

where S_n is defined in (3). □

Proof of Lemma 2.

Since $\{\tilde{\boldsymbol{\pi}}, \tilde{\boldsymbol{m}}, \tilde{\boldsymbol{\sigma}}^2\}$ maximizes $\ell_1(\boldsymbol{\pi}, \boldsymbol{m}, \boldsymbol{\sigma}^2)$ defined in (5), it is easy to see that $\tilde{\boldsymbol{\theta}}^*$ maximizes

$$\begin{aligned} \ell_n^*(\boldsymbol{\theta}^*) &= h \sum_{i=1}^{n} \{\ell(\boldsymbol{\theta}(x) + \gamma_n \boldsymbol{\theta}^*, Y_i) - \ell(\boldsymbol{\theta}(x), Y_i)\} K_h(X_i - x), \\ &= S_n \boldsymbol{\theta}^* + \frac{1}{2} \boldsymbol{\theta}^{*T} W_n \boldsymbol{\theta}^* + o_p(\|\boldsymbol{\theta}^*\|^2), \end{aligned} \tag{1}$$

where

$$S_n = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{\partial \ell(\boldsymbol{\theta}(x), Y_i)}{\partial \boldsymbol{\theta}} K_h(X_i - x),\ W_n = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ell(\boldsymbol{\theta}(x), Y_i)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} K_h(X_i - x), \tag{2}$$

and the second equality holds by Taylor expansion. It is easy to see that $W_n = -f(x)I_\theta(x) + o_p(1)$, and therefore,

$$\ell_n^*(\boldsymbol{\theta}^*) = S_n \boldsymbol{\theta}^* - \frac{1}{2} f(x) \boldsymbol{\theta}^{*T} I_\theta(x) \boldsymbol{\theta}^* + o_p(\|\boldsymbol{\theta}^*\|^2). \tag{3}$$

By Lemma 1 and assumption (C9), it can be shown that for all $x \in \mathscr{X}$, W_n converges to $-f(x)I_\theta(x)$ uniformly. From (3) and assumption (C7) and (C9), we know that $-\ell_n^*(\boldsymbol{\theta}^*)$ is convex function defined on a convex open set, when n is large enough. Therefore, by the convexity lemma (Pollard, 1991),

$$\sup_{x\in\mathscr{X}}\left|(S_n\boldsymbol{\theta}^*+\frac{1}{2}\boldsymbol{\theta}^{*T}W_n\boldsymbol{\theta}^*)-[S_n\boldsymbol{\theta}^*-\frac{1}{2}f(x)\boldsymbol{\theta}^{*T}I_\theta(x)\boldsymbol{\theta}^*]\right|\xrightarrow{P}0$$

holds uniformly for all $x\in\mathscr{X}$ and $\boldsymbol{\theta}^*$ in any compact set. We know that $-f^{-1}(x)I_\theta^{-1}(x)S_n$ is a unique maximizer of (3), and by definition, $\tilde{\boldsymbol{\theta}}^*$ is a maximizer of (1), then, by Lemma A.1 of Carroll et al. (1997), $\sup_{x\in\mathscr{X}}\left|\tilde{\boldsymbol{\theta}}^*-f^{-1}(x)I_\theta^{-1}(x)S_n\right|\xrightarrow{P}0$, which also implies that

$$\tilde{\boldsymbol{\theta}}^*=f^{-1}(x)I_\theta^{-1}(x)S_n+o_p(1). \tag{4}$$

Since $\tilde{\boldsymbol{\theta}}^*$ maximizes (1),

$$\begin{aligned}0=&h\gamma_n\sum_{i=1}^n\frac{\partial\ell(\boldsymbol{\theta}(x)+\gamma_n\tilde{\boldsymbol{\theta}}^*,Y_i)}{\partial\boldsymbol{\theta}}K_h(X_i-x)\\=&h\gamma_n\sum_{i=1}^n\frac{\partial\ell(\boldsymbol{\theta}(x),Y_i)}{\partial\boldsymbol{\theta}}K_h(X_i-x)+h\gamma_n^2\tilde{\boldsymbol{\theta}}^*\sum_{i=1}^n\frac{\partial^2\ell(\boldsymbol{\theta}(x),Y_i)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}K_h(X_i-x)+O_p(\gamma_n\|\tilde{\boldsymbol{\theta}}^*\|^2),\end{aligned}$$

that is, $W_n\tilde{\boldsymbol{\theta}}^*+O_p(\gamma_n\|\tilde{\boldsymbol{\theta}}^*\|^2)=-S_n$. Therefore,

$$\{W_n-E(W_n)\}\tilde{\boldsymbol{\theta}}^*+O_p(\gamma_n\|\tilde{\boldsymbol{\theta}}^*\|^2)=-S_n-E(W_n)\tilde{\boldsymbol{\theta}}^*=-S_n+f(x)I_\theta(x)\tilde{\boldsymbol{\theta}}^*. \tag{5}$$

From (4) and (9), it is easy to show that $\sup_{x\in\mathscr{X}}|\tilde{\boldsymbol{\theta}}^*|=O_p(1)$. By Lemma 1, $\sup_{x\in\mathscr{X}}|W_n-E(W_n)|=O_p\{h^2+\gamma_n\log^{1/2}(1/h)\}$, thus $\{W_n-E(W_n)\}\tilde{\boldsymbol{\theta}}^*+O_p(\gamma_n\|\tilde{\boldsymbol{\theta}}^*\|^2)=O_p\{h^2+\gamma_n\log^{1/2}(1/h)\}$. Combined with (5), we have

$$\sup_{x\in\mathscr{X}}\left|-S_n+f(x)I_\theta(x)\tilde{\boldsymbol{\theta}}^*\right|=O_p\{h^2+\gamma_n\log^{1/2}(1/h)\}.$$

Since $f(x)$ and $I_\theta(x)$ are bounded and continuous functions in a closed set of $\mathscr{X}$ and $I_\theta(x)$ is positive definite,

$$\sup_{x\in\mathscr{X}}\left|\tilde{\boldsymbol{\theta}}^*-f^{-1}(x)I_\theta^{-1}S_n\right|=O_p\{h^2+\gamma_n\log^{1/2}(1/h)\}.$$

□

Proof of Theorem 1.
Define $\hat{\boldsymbol{\beta}}^*=\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, where $\hat{\boldsymbol{\beta}}$ maximizes $\ell_2(\boldsymbol{\beta})$ in (6). Let

$$\ell(\tilde{\boldsymbol{m}}(X_i),\boldsymbol{\beta},Y_i)=\log\{\sum_{j=1}^k\pi_j\phi(Y_i|\tilde{m}_j(X_i),\sigma_j^2\},$$

$$\ell(\tilde{\boldsymbol{m}}(X_i),\boldsymbol{\beta}+\boldsymbol{\beta}^*/\sqrt{n},Y_i)=\log\{\sum_{j=1}^k(\pi_j+\pi_j^*/\sqrt{n})\phi(Y_i|\tilde{m}_j(X_i),\sigma_j^2+\sigma_j^{2*}/\sqrt{n}\}.$$

Since $\hat{\boldsymbol{\beta}}$ maximizes ℓ_2, it is easy to see that $\hat{\boldsymbol{\beta}}^*$ maximizes

$$\ell_n(\boldsymbol{\beta}^*) = \sum_{i=1}^{n}\{\ell(\tilde{\boldsymbol{m}}(X_i), \boldsymbol{\beta} + \boldsymbol{\beta}^*/\sqrt{n}, Y_i) - \ell(\tilde{\boldsymbol{m}}(X_i), \boldsymbol{\beta}, Y_i\} = A_n\boldsymbol{\beta}^* + \frac{1}{2}\boldsymbol{\beta}^{*T}B_n\boldsymbol{\beta}^* + o_p(\|\boldsymbol{\beta}^*\|^2),$$

where $A_n = \sqrt{\frac{1}{n}}\sum_{i=1}^n \frac{\partial\ell(\tilde{\boldsymbol{m}}(X_i),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{\beta}}$ and $B_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2\ell(\tilde{\boldsymbol{m}}(X_i),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}$. It can be easily seen that $B_n = -B + o_p(1)$ with $B = E\{I_\beta(X)\}$, therefore, by quadratic approximation lemma,

$$\hat{\boldsymbol{\beta}}^* = B^{-1}A_n + o_p(1). \tag{6}$$

Define $R_{1n} = \sqrt{\frac{1}{n}}\sum_{i=1}^n \frac{\partial^2\ell(\boldsymbol{m}(X_i),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{m}^T}(\tilde{\boldsymbol{m}}(X_i)-\boldsymbol{m}(X_i))$, then $A_n = \sqrt{\frac{1}{n}}\sum_{i=1}^n \frac{\partial\ell(\boldsymbol{m}(X_i),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{\beta}} + R_{1n} + O_p(\sqrt{\frac{1}{n}}\|\tilde{\boldsymbol{m}} - \boldsymbol{m}\|_\infty^2)$. Let $\varphi(X_t, Y_t)$ be a $k \times 1$ vector whose elements are the first k entries of $I_\theta^{-1}(X_t)\frac{\partial\ell(\boldsymbol{\theta}(X_t),Y_t)}{\partial\boldsymbol{\theta}}$. From assumption (C1), we know that $O_p\{n^{1/2}[\gamma_n h^2 + \gamma_n^2\log^{1/2}(1/h)]\} = o_p(1)$. By Lemma 3, $\tilde{\boldsymbol{\theta}}(X_i)-\boldsymbol{\theta}(X_i) = \frac{1}{n}f^{-1}(X_i)I_\theta^{-1}(X_i)\sum_{t=1}^n \frac{\partial\ell(\boldsymbol{\theta}(X_i),Y_t)}{\partial\boldsymbol{\theta}}K_h(X_t - X_i) + O_p\{\gamma_n h^2 + \gamma_n^2\log^{1/2}(1/h)\}$. Since $\boldsymbol{m}(X_i) - \boldsymbol{m}(X_t) = O(X_i - X_t)$,

$$\begin{aligned}R_{1n} &= n^{-3/2}\sum_{t=1}^{n}\sum_{i=1}^{n}\frac{\partial^2\ell(\boldsymbol{m}(X_i), \boldsymbol{\beta}, Y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{m}^T}f^{-1}(X_i)\varphi(X_t, Y_t)K_h(X_i - X_t) + O_p(n^{1/2}h^2)\\ &= R_{2n} + O_p(n^{1/2}h^2).\end{aligned}$$

It can be shown that $E[\frac{1}{n}\sum_{i=1}^n \frac{\partial^2\ell(\boldsymbol{m}(X_i),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{\beta}\partial\boldsymbol{m}^T}f^{-1}(X_i)K_h(X_i - X_t)] = I_{\beta m}(X_t)$. Let $\varpi(X_t, Y_t) = I_{\beta m}(X_t)\varphi(X_t, Y_t)$, and $R_{n3} = -n^{-1/2}\sum_{j=1}^n \varpi(X_t, Y_t)$, then $R_{n2} - R_{n3} \xrightarrow{P} 0$, and therefore

$$A_n = \sqrt{\frac{1}{n}}\sum_{i=1}^{n}\{\frac{\partial\ell(\boldsymbol{m}(X_i), \boldsymbol{\beta}, Y_i)}{\partial\boldsymbol{\beta}} - \varpi(X_i, Y_i)\} + o_p(1),$$

given $nh^4 \to 0$. Let $\Sigma = Var\{\frac{\partial\ell(\boldsymbol{\theta}(X),Y)}{\partial\boldsymbol{\beta}} - \varpi(X, Y)\}$, then $\mathrm{Var}(A_n) = \Sigma$. It can be easily seen that $E(A_n) = 0$, therefore by (6),

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(0, B^{-1}\Sigma B^{-1}).$$

□

Proof of Theorem 2.
Define $\hat{\boldsymbol{m}}^* = \sqrt{nh}(\hat{\boldsymbol{m}}(x) - \boldsymbol{m}(x))$, where $\hat{\boldsymbol{m}}(x)$ maximizes (7). It can be shown that

$$\hat{\boldsymbol{m}}^*(x) = f(x)^{-1}I_m(x)^{-1}\hat{S}_n + o_p(1), \tag{7}$$

where

$$\hat{S}_n = \sqrt{\frac{h}{n}}\sum_{i=1}^{n}\frac{\partial\ell(\boldsymbol{m}(x), \hat{\boldsymbol{\beta}}, Y_i)}{\partial\boldsymbol{m}}K_h(X_i - x). \tag{8}$$

Notice that

$$\hat{S}_n = \sqrt{\frac{h}{n}}\sum_{i=1}^{n}\frac{\partial\ell(\boldsymbol{m}(x),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{m}}K_h(X_i-x)+\sqrt{\frac{h}{n}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\sum_{i=1}^{n}\frac{\partial^2\ell(\boldsymbol{m}(x),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{m}\partial\boldsymbol{\beta}^T}K_h(X_i-x)+o_p(1)$$
$$\equiv S_n + D_n + o_p(1).$$

where S_n is defined in (2). Since $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\ell(\boldsymbol{m}(x),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{m}\partial\boldsymbol{\beta}^T}K_h(X_i-x) = -f(x)I_{\beta m}^T(x)+o_p(1)$, then $D_n = \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\sqrt{h}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\ell(\boldsymbol{m}(x),\boldsymbol{\beta},Y_i)}{\partial\boldsymbol{m}\partial\boldsymbol{\beta}^T}K_h(X_i-x) = -\sqrt{h}f(x)I_{\beta m}^T(x)+o_p(1)$. Thus, from (7), $\hat{\boldsymbol{m}}^*(x) = f(x)^{-1}I_m(x)^{-1}S_n+o_p(1)$. Let $\Lambda(u|x) = E[\frac{\partial\ell(\boldsymbol{m}(x),\boldsymbol{\beta},Y)}{\partial\boldsymbol{m}}|X=u]$, it can be shown that

$$E(S_n) = \sqrt{nh}[\frac{1}{2}f(x)\Lambda''(x|x)+f'(x)\Lambda'(x|x)]\kappa_2h^2, Var(S_n) = f(x)I_m(x)\nu_0. \tag{9}$$

To complete the proof, let $\Delta(x) = I_m^{-1}(x)[\frac{1}{2}\Lambda''(x|x)+f^{-1}(x)f'(x)\Lambda'(x|x)]\kappa_2h^2$, and $\Delta_m(x)$ be a $k\times 1$ vector whose elements are the first k entries of $\Delta(x)$, then

$$\sqrt{nh}(\hat{\boldsymbol{m}}(x)-\boldsymbol{m}(x)-\Delta_m(x)+o_p(h^2))\xrightarrow{D} N(0, f^{-1}(x)I_m^{-1}(x)\nu_0).$$

□

Proof of Theorem 3.
(i) Assume the latent variables $\{Z_i, i=1,...,n\}$ be a random sample from population Z, then $P(Z_i=j|Y,\boldsymbol{\theta}) = \pi_j\phi(Y|m_j,\sigma_j^2)/\sum_{j=1}^{k}\pi_j\phi(Y|m_j,\sigma_j^2)$, and therefore,

$$\log\{\sum_{j=1}^{k}\pi_j\phi(Y_i|m_j,\sigma_j^2)\} = \log\{\pi_j\phi(Y_i|m_j,\sigma_j^2)\}-\log\{P(Z_i=j|Y,\boldsymbol{\theta})\}. \tag{10}$$

Given $\boldsymbol{\theta}^{(l)}(X_i) = (\boldsymbol{m}^{(l)}(X_i),\boldsymbol{\pi}^{(l)}(X_i),\boldsymbol{\sigma}^{2(l)}(X_i))$, for any $i=1,...,n$, $P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l)}(X_i)) = p_{ij}^{(l+1)}$ and $\sum_{j=1}^{k}p_{ij}^{(l+1)}=1$. Therefore, by (10)

$$\begin{aligned}\ell_1(\boldsymbol{\theta}) = &\sum_{i=1}^{n}\{\sum_{j=1}^{k}\log\{\pi_j\phi(Y_i|m_j,\sigma_j^2)\}p_{ij}^{(l+1)}\}K_h(X_i-x)\\ &-\sum_{i=1}^{n}\{\sum_{j=1}^{k}\log\{P(Z_i=j|Y,\boldsymbol{\theta})\}p_{ij}^{(l+1)}\}K_h(X_i-x).\end{aligned} \tag{11}$$

Based on the M-step of (8), (9) and (10), we have

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}\{\sum_{j=1}^{k}\log\{\pi_j^{(l+1)}(x)\phi(Y_i|m_j^{(l+1)}(x),\sigma_j^{2(l+1)}(x))\}p_{ij}^{(l+1)}\}K_h(X_i-x)\\ \geq &n^{-1}\sum_{i=1}^{n}\{\sum_{j=1}^{k}\log\{\pi_j^{(l)}(x)\phi(Y_i|m_j^{(l)}(x),\sigma_j^{2(l)}(x))\}p_{ij}^{(l+1)}\}K_h(X_i-x).\end{aligned}$$

To complete the proof, based on (11), we only need to show

$$\limsup_{n\to\infty} n^{-1}\sum_{i=1}^{n}\{\sum_{j=1}^{k}\log\{\frac{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l+1)}(x))}{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l)}(x))}\}p_{ij}^{(l+1)}\}K_h(X_i-x)\leq 0$$

in probability. Define

$$L = n^{-1}\sum_{i=1}^{n}\{\sum_{j=1}^{k}\log\{\frac{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l+1)}(x))}{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l)}(x))}\}p_{ij}^{(l+1)}\}K_h(X_i-x),$$

$$U = n^{-1}\sum_{i=1}^{n}\log\{\sum_{j=1}^{k}\{\frac{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l+1)}(x))}{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l)}(x))}\}p_{ij}^{(l+1)}\}K_h(X_i-x),$$

then, by Jensen's inequality, $L\leq U$. We complete the proof by showing that $U\xrightarrow{P}0$. Without loss of generality, assume that $P(Z_i=j|Y,\boldsymbol{\theta}^{(l)}(x))\geq\delta>0$ for some small value δ. Since $E(U)=E\{\log[\sum_{j=1}^{k}\frac{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l+1)}(x))}{P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l)}(x))}P(Z_i=j|Y_i,\boldsymbol{\theta}^{(l)}(X_i))]K_h(X_i-x)\}$, by similar argument as in the proof of Theorem 2 and Theorem 3, it can be shown that $E(U)\to 0$, and $\mathrm{Var}(U)=O_p((nh)^{-1})$. Therefore, by Chebyshv's inequality, $U=o_p(1)$, and thus completes the proof.

(ii) Notice that $P(Z_i=j|Y,\boldsymbol{m},\hat{\boldsymbol{\beta}})=\hat{\pi}_j\phi(Y|m_j,\hat{\sigma}_j^2)/\sum_{j=1}^{k}\hat{\pi}_j\phi(Y|m_j,\hat{\sigma}_j^2)$, $P(Z_i=j|Y_i,\boldsymbol{m}^{(l)}(X_i),\hat{\boldsymbol{\beta}})=p_{ij}^{(l+1)}$ and $\sum_{j=1}^{k}p_{ij}^{(l+1)}=1$, where $p_{ij}^{(l+1)}$ is defined in (11). The rest of the proof is in line with part (i), and thus is omitted here.

(iii) Notice that by fixing $\tilde{\boldsymbol{m}}(\cdot)=\boldsymbol{m}^{(l)}(\cdot)$, $\ell^*(\boldsymbol{\pi},\boldsymbol{m}^{(l)}(\cdot),\boldsymbol{\sigma}^2)=\ell_2(\boldsymbol{\pi},\boldsymbol{\sigma}^2)$. Therefore, by the ascent property of the ordinary EM algorithm,

$$\ell^*(\boldsymbol{\pi}^{(l+1)},\boldsymbol{m}^{(l)}(\cdot),\boldsymbol{\sigma}^{2(l+1)})=\ell_2(\boldsymbol{\pi}^{(l+1)},\boldsymbol{\sigma}^{2(l+1)})\geq\ell_2(\boldsymbol{\pi}^{(l)},\boldsymbol{\sigma}^{2(l)})=\ell^*(\boldsymbol{\pi}^{(l)},\boldsymbol{m}^{(l)}(\cdot),\boldsymbol{\sigma}^{2(l)}).$$

Thus, to complete the proof, we only need to show

$$\liminf_{n\to\infty} n^{-1}[\ell^*(\boldsymbol{\pi}^{(l+1)},\boldsymbol{m}^{(l+1)}(\cdot),\boldsymbol{\sigma}^{2(l+1)})-\ell^*(\boldsymbol{\pi}^{(l+1)},\boldsymbol{m}^{(l)}(\cdot),\boldsymbol{\sigma}^{2(l+1)})]\geq 0.$$

If we fix $\hat{\boldsymbol{\pi}}=\boldsymbol{\pi}^{(l+1)}$ and $\hat{\boldsymbol{\sigma}}^2=\boldsymbol{\sigma}^{2(l+1)}$, then by part (ii), $\liminf_{n\to\infty}n^{-1}[\ell_3(\boldsymbol{m}^{(l+1)}(x))-\ell_3(\boldsymbol{m}^{(l)}(x))]\geq 0$ in probability for any $x\in\{X_t,t=1,...,n\}$. Therefore,

$$\liminf_{n\to\infty} n^{-2}\sum_{t=1}^{n}f(X_t)^{-1}[\ell_3(\boldsymbol{m}^{(l+1)}(X_t))-\ell_3(\boldsymbol{m}^{(l)}(X_t))]$$

$$\geq\liminf_{n\to\infty} n^{-1}\sum_{t=1}^{n}\liminf_{n\to\infty}n^{-1}f(X_t)^{-1}[\ell_3(\boldsymbol{m}^{(l+1)}(X_t))-\ell_3(\boldsymbol{m}^{(l)}(X_t))]\geq 0.$$

Since $K(\cdot)$ is symmetric about 0, $n^{-2}\sum_{t=1}^{n}f(X_t)^{-1}\ell_3(\boldsymbol{m}^{(l)}(X_t))=n^{-1}\sum_{i=1}^{n}\Gamma_i^{(l)}$, where

$$\Gamma_i^{(l)}=n^{-1}\sum_{t=1}^{n}f(X_t)^{-1}\log[\sum_{j=1}^{k}\hat{\pi}_j\phi(Y_i|m_j^{(l)}(X_t),\hat{\sigma}_j^2)]K_h(X_t-X_i).$$

It can be shown that $E(\Gamma_i^{(l)}|X_i,Y_i)=\log[\sum_{j=1}^k\hat{\pi}_j\phi(Y_i|m_j^{(l)}(X_i),\hat{\sigma}_j^2)](1+o_p(1))$, and $\text{Var}(\Gamma_i^{(l)}|X_i,Y_i)=O_p((nh)^{-1})$. The fact that $\sum_{i=1}^n E(\Gamma_i^{(l)}|X_i,Y_i)=\ell^*(\boldsymbol{\pi}^{(l+1)},\boldsymbol{m}^{(l)}(\cdot),\boldsymbol{\sigma}^{2(l+1)})$, and $\sum_{i=1}^n E(\Gamma_i^{(l+1)}|X_i,Y_i)=\ell^*(\boldsymbol{\pi}^{(l+1)},\boldsymbol{m}^{(l+1)}(\cdot),\boldsymbol{\sigma}^{2(l+1)})$ completes the proof. □

Proof of Theorem 4.
Since $\hat{\boldsymbol{\beta}}$ has faster convergence rate than $\hat{\boldsymbol{m}}(\cdot)$, $\hat{\boldsymbol{m}}(\cdot)$ has the same asymptotic properties as if $\boldsymbol{\beta}$ were known. Therefore, in the following proof, we study the property of $\hat{\boldsymbol{m}}(\cdot)$ assuming $\boldsymbol{\beta}$ to be known.

Define $\frac{\partial\ell(\theta(X_i),Y_i)}{\partial\theta}=q_{\theta i}$, $\frac{\partial^2\ell(\theta(X_i),Y_i)}{\partial\theta\partial\theta^T}=q_{\theta\theta i}$ and similarly, define q_{mi}, q_{mmi} and so on. Let $\tilde{\boldsymbol{\theta}}$ be the estimator under H_1 (Huang et al., 2013), and $\hat{\boldsymbol{m}}$ be the estimator under H_0 (model (1). From previous proof, we have

$$\tilde{\boldsymbol{\theta}}(X_i)-\boldsymbol{\theta}(X_i)=\frac{1}{n}f^{-1}(X_i)I_\theta^{-1}(X_i)\sum_{t=1}^n q_{\theta t}K_h(X_t-X_i)(1+o_p(1)), \tag{12}$$

$$\hat{\boldsymbol{m}}(X_i)-\boldsymbol{m}(X_i)=\frac{1}{n}f^{-1}(X_i)I_m^{-1}(X_i)\sum_{t=1}^n q_{mt}K_h(X_t-X_i)(1+o_p(1)). \tag{13}$$

By (12) and (13), we can obtain that

$$\sum_{i=1}^n\ell(\tilde{\boldsymbol{\theta}}(X_i),Y_i)-\sum_{i=1}^n\ell(\boldsymbol{\theta}(X_i),Y_i)=\{\frac{1}{n}\sum_{i,l}q_{\theta i}^Tf^{-1}(X_l)I_\theta^{-1}(X_l)q_{\theta l}K_h(X_i-X_l)$$
$$+\frac{1}{2n^2}\sum_{i,j,l}q_{\theta i}^Tf^{-2}(X_l)I_\theta^{-1}(X_l)q_{\theta\theta l}I_\theta^{-1}(X_l)q_{\theta j}K_h(X_i-X_l)K_h(X_j-X_l)\}(1+o_p(1)),$$
$$\sum_{i=1}^n\ell(\hat{\boldsymbol{m}}(X_i),Y_i)-\sum_{i=1}^n\ell(\boldsymbol{m}(X_i),Y_i)=\{\frac{1}{n}\sum_{i,l}q_{mi}^Tf^{-1}(X_l)I_m^{-1}(X_l)q_{ml}K_h(X_i-X_l)$$
$$+\frac{1}{2n^2}\sum_{i,j,l}q_{mi}^Tf^{-2}(X_l)I_m^{-1}(X_l)q_{mml}I_m^{-1}(X_l)q_{mj}K_h(X_i-X_l)K_h(X_j-X_l)\}(1+o_p(1)),$$

and so,

$$T=\frac{1}{n}\sum_{i,l}[q_{\theta i}^TI_\theta^{-1}(X_l)q_{\theta l}-q_{mi}^TI_m^{-1}(X_l)q_{ml}]f^{-1}(X_l)K_h(X_i-X_l)+\frac{1}{2n^2}\sum_{i,j,l}[q_{\theta i}^TI_\theta^{-1}(X_l)q_{\theta\theta l}$$
$$\times I_\theta^{-1}(X_l)q_{\theta j}-q_{mi}^TI_m^{-1}(X_l)q_{mml}I_m^{-1}(X_l)q_{mj}]f^{-2}(X_l)K_h(X_i-X_l)K_h(X_j-X_l)$$
$$\equiv\Lambda_n+\frac{1}{2}\Gamma_n.$$

By similar argument as Fan et al. (2001), it can be shown that under conditions (C9)-(C12), as $h\to 0$, $nh^{3/2}\to\infty$,

$$\Lambda_n=\frac{2k-1}{h}K(0)Ef(X)^{-1}+\frac{1}{n}\sum_{l\neq i}[q_{\theta i}^TI_\theta^{-1}(X_l)q_{\theta l}-q_{mi}^TI_m^{-1}(X_l)q_{ml}]f^{-1}(X_l)K_h(X_i-X_l)+o_p(h^{-1/2}),$$

$$\Gamma_n=-\frac{(2k-1)}{h}Ef(X)^{-1}\int K^2(t)dt-\frac{2}{n}\sum_{i<j}[q_{\theta i}^TI_\theta^{-1}(X_i)q_{\theta j}-q_{mi}^TI_m^{-1}(X_i)q_{mj}]f^{-1}(X_i)$$
$$\times K_h*K_h(X_i-X_j)+o_p(h^{-1/2}).$$

Therefore, $T = \mu_n + W_n/2\sqrt{h} + o_p(h^{-1/2})$, where $\mu_n = \frac{(2k-1)|\mathscr{X}|}{h}[K(0) - 0.5\int K^2(t)dt]$,

$$W_n = \frac{\sqrt{h}}{n}\sum_{i\neq j}\{q_{\theta i}^T I_\theta^{-1}(X_j)[2K_h(X_i - X_j) - K_h * K_h(X_i - X_j)]f^{-1}(X_j)q_{\theta j} - q_{mi}^T I_m^{-1}(X_j)[2K_h(X_i - X_j) - K_h * K_h(X_i - X_j)]f^{-1}(X_j)q_{mj}\}.$$

It can be shown that $\text{Var}(W_n) \to \zeta$, where $\zeta = 2(2k-1)Ef^{-1}(X)\int[2K(t) - K * K(t)]^2 dt$. Apply Proposition 3.2 in de Jong (1987), we obtain that

$$W_n \xrightarrow{D} N(0, \zeta),$$

and completes the proof. □

References

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M.P. (1997). Generalized partially Linear Single-index Models. *Journal of American Statistical Association*, 92, 477-489.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probablity Theory and Related Fields*, 75, 261-277.

Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annuals of Statistics*, 29, 153-193.

Fan, J. and Huang, T. (2005). Profile Likelihood Inference on Semiparametric Varying-Coefficient Partially Linear Models. *Bernoulli*, 11, 1031-1057.

Huang, M., Li, R., and Wang, S. (2013). Nonparametric mixture of regression models. *Journal of American Statistical Association*, 108, 929-941.