Supplementary File for the Manuscript Titled “Semiparametric Mixture of Binomial Regression” by J. Cao and W. Yao

Proofs

Let $g(t)$ be the density function for $t$. The following technical conditions are imposed in this section. They are not the weakest possible conditions, but they are imposed to facilitate the proofs.

Technical Conditions:

A $\pi_1(t)$ and $p(t)$ has continuous second derivative at $t_0$ and $0 < \pi_1(t_0) < 1$ and $0 < p(t_0) < 1$. (For the constant proportion semiparametric mixture model (3), we use the same assumption for $p(t)$ and assume $0 < \pi_1 < 1$.)

B $g(t)$ has continuous second derivative at the point $t_0$ and $g(t_0) > 0$.

C $K(\cdot)$ is a symmetric (about 0) kernel density with compact support $[-1, 1]$.

D The bandwidth $h$ tends to zero such that $nh \to \infty$.

Let $\alpha_n = (nh)^{-1/2} + h^2, \theta_0 = \{\pi_1(t_0), p(t_0)\}$,

$$f(x, \theta) = \pi_1 I(x = 0) + \pi_2 \left( \frac{N}{x} \right) p^x \{1 - p\}^{N-x},$$

$l(x, \theta) = \log f(x, \theta)$, where $\theta = (\pi_1, p)$. Then the objective function (4) can be written as

$$\ell(\theta) = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \log f(x_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) l(x, \theta).$

Define

$$l_1(x, \theta) = \frac{\partial}{\partial \theta} l(x, \theta) \quad \text{and} \quad l_2(x, \theta) = \frac{\partial^2}{\partial \theta \partial \theta^T} l(x, \theta),$$
\[ G(t) = \mathbb{E}\{l_1(X, \theta_0) \mid t\} \text{ and } I(t) = -\mathbb{E}\{l_2(X, \theta_0) \mid t\}. \] The moments of \( K \) and \( K^2 \) are denoted respectively by

\[
\mu_j = \int t^j K(t) dt \quad \text{and} \quad \nu_j = \int t^j K^2(t) dt.
\]

By some simple calculations, we can get the following results.

**Lemma 1.** Assume that the regularity conditions A–C hold. We have the following results

1. The \( G(t) \) has continuous second derivative at \( t_0 \) and \( \mathbb{E}\{l_1(X, \theta_0)^2 \mid t\} \) is continuous at \( t_0 \).

2. The \( \partial^3 \ell(\theta_0)/(\partial \theta_i \partial \theta_j \partial \theta_k) \) is a bounded function for all \( \theta \) in a neighborhood of \( \theta_0 \) and all \( x \).

3. \( I(t) \) is continuous at \( t_0 \) and positive definite at \( t_0 \) and

\[ I(t_0) = \mathbb{E}\{l_1(X, \theta_0)l_1(X, \theta_0)^T \mid t_0\}. \]

**Proof of Theorem 2.1.**

Note that

\[
\ell(\theta) = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \log f(x_i, \theta).
\]

Hence,

\[
\ell(\theta^{(k+1)}) - \ell(\theta^{(k)}) = \sum_{i=1}^{n} \log \left( \frac{f(x_i, \theta^{(k+1)})}{f(x_i, \theta^{(k)})} \right) K_h(t_i - t_0)
= \sum_{i=1}^{n} \log \left( \frac{\pi_{1}^{(k)} B(x_i, N, 0) \pi_{1}^{(k+1)} B(x_i, N, 0)}{f(x_i, \theta^{(k)}) \pi_{1}^{(k)} B(x_i, N, 0)} + \frac{\pi_{2}^{(k)} B(x_i, N, p^{(k)}) \pi_{2}^{(k+1)} B(x_i, N, p^{(k+1)})}{f(x_i, \theta^{(k)}) \pi_{2}^{(k)} B(x_i, N, p^{(k)})} \right) K_h(x_i - x_0)
= \sum_{i=1}^{n} \log \left( \frac{\pi_{1}^{(k+1)} B(x_i, N, 0)}{\pi_{1}^{(k)} B(x_i, N, 0)} + \frac{\pi_{2}^{(k+1)} B(x_i, N, p^{(k+1)})}{\pi_{2}^{(k)} B(x_i, N, p^{(k)})} \right) K_h(x_i - x_0)
\]

Based on the Jensen’s inequality, we have

\[
\ell(\theta^{(k+1)}) - \ell(\theta^{(k)}) \geq \sum_{i=1}^{n} r_{i1}^{(k+1)} \log \left( \frac{\pi_{1}^{(k+1)} B(x_i, N, 0)}{\pi_{1}^{(k)} B(x_i, N, 0)} \right) K_h(x_i - x_0)
+ r_{i2}^{(k+1)} \log \left( \frac{\pi_{2}^{(k+1)} B(x_i, N, p^{(k+1)})}{\pi_{2}^{(k)} B(x_i, N, p^{(k)})} \right) K_h(x_i - x_0)
\]
Based on the property of M-step of (5), we have
\[ \ell(\theta^{(k+1)}) - \ell(\theta^{(k)}) \geq 0. \]

**Proof of Theorem 3.1.** Denote \( \alpha_n = (nh)^{-1/2} + h^2 \). It is sufficient to show that for any given \( \eta > 0 \), there exists a large constant \( c \) such that
\[ P\{ \sup_{\|u\|=c} \ell(\theta_0 + \alpha_n u) < \ell(\theta_0) \} \geq 1 - \eta, \quad (13) \]
where \( \ell(\theta) \) is defined in (4).

By using Taylor expansion, it follows that
\[
\ell(\theta_0 + \alpha_n u) - \ell(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \{ l(x_i, \theta_0 + \alpha_n u) - l(x_i, \theta_0) \}
= \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \left\{ l_1(x_i, \theta_0)^T u \alpha_n + u^T l_2(x_i, \theta_0) u \alpha_n^2 + \alpha_n^3 q(x_i, \hat{\theta}) \right\}
= I_1 + I_2 + I_3,
\]
where \( \| \hat{\theta} - \theta_0 \| \leq c \alpha_n \) and
\[
q(x_i, \hat{\theta}) = \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} \left[ \frac{\partial^3 l(x_i, \hat{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k} \right] u_i u_j u_k,
\]
where \( u = (u_1, u_2) \).

By directly calculating the mean and variance and note that \( G(t_0) = 0 \), we obtain
\[
E(I_1) = \alpha_n E \left\{ K_h(t - t_0) G(t)^T u \right\} = O(c \alpha_n h^2);
\]
\[
\text{var}(I_1) = n^{-1} \alpha_n^2 \text{var}[K_h(t_i - t_0) l_i(\theta_0, x_i)^T u] = O(c^2 \alpha_n^2 (nh)^{-1}).
\]
Hence
\[
I_1 = O(c \alpha_n h^2) + \alpha_n c O_p((nh)^{-1/2}) = O_p(c \alpha_n^2).
\]

Similarly,
\[
I_3 = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \alpha_n^3 \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} \left[ \frac{\partial^3 l(x_i, \hat{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k} \right] u_i u_j u_k
= O_p(\alpha_n^3).
\]

And
\[
I_2 = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) u^T l_2(x_i, \theta_0) u \alpha_n^2 + \alpha_n^3 q(x_i, \hat{\theta}) = -\alpha_n^2 g(t_0) u^T T(t_0) u (1 + o_p(1)).
\]

Noticing that \( T(t_0) \) is a positive matrix, \( \|u\| = c \), we can choose \( c \) large enough such that \( I_2 \) dominates both \( I_1 \) and \( I_3 \) with probability at least \( 1 - \eta \). Thus (13) holds. Hence with probability approaching 1 (wpa1), there exists a local maximizer \( \hat{\theta} \) such that
\[ \| \hat{\theta} - \theta_0 \| \leq \alpha_n c, \]
where \( \alpha_n = (nh)^{-1/2} + h^2 \). Based on the definition of \( \theta \), we can also get, wpa1, \( \hat{\theta}(t_0) - \pi(t_0) = O_p((nh)^{-1/2} + h^2) \) and \( \hat{\theta}(t_0) - p(t_0) = O_p((nh)^{-1/2} + h^2) \).
Proof of Theorem 3.2.

Note that the estimate \( \hat{\theta} \) satisfies the equation

\[
0 = \ell'(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \left\{ l_1(x_i, \theta_0) + l_2(x_i, \theta_0)(\hat{\theta} - \theta_0) + O_p(||\hat{\theta} - \theta_0||^2) \right\} . \tag{14}
\]

The order of the third term could be derived from the (2) of Lemma 1. Let

\[
W_n = \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) l_1(x_i, \theta_0),
\]

\[
\Delta_n = -\frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) l_2(x_i, \theta_0).
\]

Note that

\[
E(W_n) = E\{K_h(t - t_0)G(t)\} = \frac{1}{2}(Gg)'(t_0)\mu_2 h^2(1 + o(1)),
\]

\[
\text{cov}(W_n) = n^{-1}\text{cov}\{K_h(t_i - t_0)l_1(x_i, \theta_0)\}
\]

\[
= n^{-1}\{EK_h^2(t_i - t_0)l_1(x_i, \theta_0)l_1(x_i, \theta_0)^T - E(W_n)^2\}
\]

\[
= (nh)^{-1}g(t_0)I(t_0)v_0(1 + o(1)), \tag{15}
\]

where \((Gg)''(t)\) is the second derivative of \(G(t)g(t)\), and

\[
E(\Delta_n) = E\{K_h(t - t_0)I(t)\} = I(t_0)g(t_0) + o(1),
\]

\[
\text{var}(\Delta_{n}(i,j)) \leq n^{-1}E \left[ K_h^2(t_i - t_0) \left( \frac{\partial^2 l_1(x_i, \theta_0)}{\partial \theta_i \partial \theta_j} \right)^2 \right]
\]

\[
= O\{(nh)^{-1}\} = o(1).
\]

Therefore, we have

\[
\Delta_n = I(t_0)g(t_0) + o_p(1).
\]

Note that \(\|\hat{\theta} - \theta_0\|^2 = o_p(W_n)\). Then from (14), we have

\[
\sqrt{n}h(\hat{\theta} - \theta_0) = g(t_0)^{-1}I(t_0)^{-1}\sqrt{nhW_n}(1 + o_p(1)). \tag{16}
\]

In order to prove the asymptotic normality of (16), we only need to establish the asymptotic normality of \(\sqrt{nhW_n}\). Next we show, for any unit vector \(d \in \mathbb{R}^2\), we prove

\[
\{d^T \text{cov}(W_n^*)d\}^{-\frac{1}{2}} \{d^T W_n^* - d^T E(W_n^*)\} \xrightarrow{L} N(0, 1),
\]

where \(W_n^* = \sqrt{nhW_n}\). Let

\[
\xi_i = \sqrt{h/n}K_h(t_i - t_0)d^T l_1(\theta_0, x_i).
\]
Then $d^TW^*_n = \sum_{i=1}^n \xi_i$. We check the Lyapunov’s condition. Based on (15), we can get $\text{cov}(W^*_n) = g(t_0)\mathcal{I}(t_0)\nu_0(1+o(1))$ and $\text{var}(d^TW^*_n) = d^T\text{cov}(W^*_n)d = g(t_0)\nu_0d^T\mathcal{I}(t_0)d(1+ o(1))$. So we only need to prove $nE|\xi_1|^3 \to 0$. Noticing that $l_1(\theta_0, x)$ is bounded for any $x$, and $K(\cdot)$ has compact support,
$$nE|\xi_1|^3 \leq O(nn^{-3/2}h^{3/2})E|K^3_n(t_i - t_0)| = O(n^{-1/2}h^{3/2})O(h^{-2}) = O((nh)^{-1/2}) \to 0.$$ 
So the asymptotic normality for $W^*_n$ holds such that
$$\sqrt{nh} \left\{ W_n - \frac{1}{2}(Gg)^\prime(t_0)\mu_2h^2 + o(h^2) \right\} \xrightarrow{D} N \{ 0, g(t_0)\mathcal{I}(t_0)\nu_0 \}.$$ 
Based on (16) and the Slutsky theorem, we can get the asymptotic result of $\hat{\theta}$
$$\sqrt{nh} \left\{ \hat{\theta} - \theta_0 - b(t_0)h^2 + o(h^2) \right\} \xrightarrow{D} N \{ 0, g^{-1}(t_0)\mathcal{I}^{-1}(t_0)\nu_0 \},$$
where
$$b(t_0) = \mathcal{I}^{-1}(t_0) \left\{ \frac{G'(t_0)g'(t_0)}{g(t_0)} + \frac{1}{2}G''(t_0) \right\} \mu_2.$$

**Proof of Theorem 3.3.**

Let
$$f(x_i, \pi_1, \hat{p}(t_i)) = \log \left[ \pi_1I(x_i = 0) + \pi_2 \left( \frac{N}{x_i} \right) \hat{p}(t_i)^{x_i} (1 - \hat{p}(t_i))^{N-x_i} \right].$$
Based on a Taylor expansion of (4), similar to the proof of Theorem 3.2, we have that
$$\sqrt{n}(\hat{\pi}_1 - \pi_1) = B_n^{-1}A_n + o_p(1).$$
where
$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial f(x_i, \pi_1, \hat{p}(t_i))}{\partial \pi_1}$$
$$B_n = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 f(x_i, \pi_1, \hat{p}(t_i))}{\partial \pi_1^2}$$
It can be shown that
$$B_n = -E \left\{ \frac{\partial^2 f(x_i, \pi_1, p(t_i))}{\partial \pi_1^2} \right\} + o_p(1)
= \mathcal{I}_{\pi_1} + o_p(1).$$
It can be shown that
$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial f(x_i, \pi_1, \hat{p}(t_i))}{\partial \pi_1} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial^2 f(x_i, \pi_1, p(t_i))}{\partial \pi_1 \partial p} \{ \hat{p}(t_i) - p(t_i) \} + O_p(d_{1n})
= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial f(x_i, \pi_1, p(t_i))}{\partial \pi_1} + S_{n1} + O_p(d_{1n}).$$
where \( d_{1n} = n^{-1/2}\|\tilde{\pi}_1 - \pi_1\|_\infty^2 = o_p(1) \). Based on the proof of Theorem 3.2, we have

\[
\hat{\theta}(t_i) - \theta(t_i) = \frac{1}{n} g(t_i) \mathcal{I}(t_i)^{-1} \sum_{j=1}^{n} K_h(t_j - t_i) l_1(x_j, \theta(t_i)) + O_p(d_{n2}),
\]

Based on Carroll et al. (1997) and Li and Liang (2008), we have that \( n^{1/2} d_{n2} = o_p(1) \) uniformly in \( t_i \), if \( n h^2 / \log(1/h) \to \infty \). Let \( \psi(t_j, x_j) \) be the second entry of \( \mathcal{I}(t_j)^{-1} l_1(x_j, \theta(t_j)) \). Since \( p(t_i) - p(t_j) = O(t_i - t_j) \) and \( K(\cdot) \) is symmetric about 0, we have

\[
S_{n1} = \frac{1}{n^{3/2}} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f(x_i, \pi_1, p(t_i))}{\partial \pi_1 \partial p} g(t_i)^{-1} \psi(t_j, x_j) K_h(t_j - t_i) + O_p(n^{1/2} h^2)
\]

\[
= S_{n2} + O_p(n^{1/2} h^2).
\]

It can be shown, by calculating the second moment, that

\[
S_{n2} - S_{n3} = o_p(1),
\]

where \( S_{n3} = -n^{-1/2} \sum_{j=1}^{n} \xi(t_j, x_j) \), with

\[
\xi(t_j, x_j) = -E \left\{ \frac{\partial^2 f(x, \pi_1, p(t_j))}{\partial p^2} \bigg| t = t_j \right\} \psi(t_j, x_j) = \mathcal{I}_{\pi, p}(t_j) \psi(t_j, x_j).
\]

By condition \( nh^4 \to 0 \), we know

\[
A_n = n^{-1/2} \sum_{i=1}^{n} \left\{ \frac{\partial f(x_i, \pi_1, p(t_i))}{\partial \pi_1} - \xi(t_i, x_i) \right\} + o_p(1).
\]

We can show that \( E(A_n) = 0 \). Define

\[
\Sigma = \text{var}(A_n) = \text{var} \left\{ \frac{\partial f(x, \pi_1, p(t))}{\partial \pi_1} - \xi(t, x) \right\}.
\]

Based on the central limit theorem, we can have

\[
\sqrt{n}(\tilde{\pi}_1 - \pi_1) \to N(0, \mathcal{I}_{\pi_1}^{-2} \Sigma).
\]

**Proof of Theorem 3.4.**

Based on a Taylor expansion of (7), similar to the proof of Theorem 3.2, we have

\[
\sqrt{n h}\{\tilde{p}(t_0) - p(t_0)\} = g(t_0)^{-1} \mathcal{I}_p(t_0)^{-1} \tilde{W}_n(1 + o_p(1)),
\]

where

\[
\mathcal{I}_p(t) = -E \left\{ \frac{\partial^2 f(x, \pi, p(t))}{\partial p^2} \bigg| t \right\}.
\]
and
\[ \hat{W}_n = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{\partial f(x_i, \pi_1, p(t_0))}{\partial p} K_h(t_i - t_0). \]

It can be calculated that
\[ \hat{W}_n = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{\partial f(x_i, \tilde{\pi}_1, p(t_0))}{\partial p} K_h(t_i - t_0) + C_n + o_p(1), \]
where
\[ C_n = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{\partial^2 f(x_i, \pi_1, p(t_0))}{\partial p \partial \pi_1} (\tilde{\pi}_1 - \pi_1) K_h(t_i - t_0). \]

Since \( \sqrt{n}(\tilde{\pi}_1 - \pi_1) = O_p(1) \), it can be shown that
\[ C_n = o_p(1). \]

Hence
\[ \sqrt{n} \{ \tilde{p}(t_0) - p(t_0) \} = g(t_0)^{-1} I_p(t_0)^{-1} W_n (1 + o_p(1)), \]
where
\[ W_n = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \frac{\partial f(x_i, \pi_1, p(t_0))}{\partial p} K_h(t_i - t_0). \]

Let
\[ \Gamma(t) = E \left\{ \frac{\partial f(x, \pi_1, p(t_0))}{\partial p} \mid t \right\}. \]

Note that \( \Gamma(t_0) = 0 \). We can show that
\[ \text{var}(W_n) = I_p(t_0) g(t_0) \nu_0 (1 + o_p(1)) \]
and
\[ E(W_n) = \frac{\sqrt{nh}}{2} \left\{ \Gamma''(t_0) g(t_0) + 2\Gamma'(t_0) g'(t_0) \right\} h^2 \mu_2 (1 + o_p(1)). \]

Similar to the proof of Theorem 3.2, we can prove the asymptotic normality of \( W_n \). Hence, we have
\[ \sqrt{n} h \{ \tilde{p}(t_0) - p(t_0) - \tilde{b}(t_0) h^2 \} \overset{D}{\rightarrow} N(0, g(t_0)^{-1} I_p(t_0)^{-1} \nu_0), \]
where
\[ \tilde{b}(t_0) = \frac{1}{2 g(t_0) I_p(t_0)} \left\{ \Gamma''(t_0) g(t_0) + 2 \Gamma'(t_0) g'(t_0) \right\} \mu_2. \]