Supplementary Appendix to “Testing the Number of Components in Normal Mixture Regression Models”

Hiroyuki Kasahara∗
Vancouver School of Economics
University of British Columbia
hkasahar@mail.ubc.ca

Katsumi Shimotsu
Faculty of Economics
University of Tokyo
shimotsu@e.u-tokyo.ac.jp

May 23, 2014

This supplementary appendix contains the following details omitted from the main paper due to space constraints: (A) proofs of the propositions in the paper, (B) auxiliary results and their proofs, (C) details of computer experiments to obtain the empirical formula in (27), and (D) additional results from empirical examples.

A Proof of propositions

A.1 Proof of Proposition 1

The stated result follows from Proposition D with \( m_0 = 1 \) and \( m = 2 \). □

A.2 Proof of Proposition 2

We suppress the subscript \( \alpha \) from \( \psi_\alpha \). For a vector \( \mathbf{x} \) and a function \( f(\mathbf{x}) \), let \( \nabla_{\mathbf{x}^k} f(\mathbf{x}) \) denote its \( k \)th derivative with respect to \( \mathbf{x} \), which can be a multidimensional array. Observe that, for any finite \( k \) and for a neighborhood \( \mathcal{N} \) of \( \psi^* \), we obtain

\[
E\left\| \nabla_{\psi^k} g(Y_i | \mathbf{X}_i, \mathbf{Z}_i; \psi^*, \alpha) / g(Y_i | \mathbf{X}_i, \mathbf{Z}_i; \psi^*, \alpha) \right\|^2 < \infty,
\]

\[
E\left\| \sup_{\psi \in \Theta \cap \mathcal{N}} \nabla_{\psi^k} \ln g(Y_i | \mathbf{X}_i, \mathbf{Z}_i; \psi, \alpha) \right\|^2 < \infty,
\]  

(28)
because each element of $\nabla \psi^* \ln g(y|\mathbf{x}, \mathbf{z}; \psi, \alpha)$ is written as a sum of products of Hermite polynomials.

First, we prove part (a). Collect $\lambda_\sigma$ and $\lambda_\beta$ into one vector as $\lambda_{\sigma\beta} = (\lambda_0, \lambda_1, \ldots, \lambda_q)^T := (\lambda_\sigma, \lambda_\beta)^T$. Propositions B and C(a)(b)(e)(f)(g) and (28) imply the followings:

for $k = 1, 2, 3$ and $\ell = 0, 1, \ldots$, $\nabla \lambda_{k, 0} L_n(\psi^*, \alpha) = 0$;

for $k = 4, 5, 6, 7$, $\nabla \lambda_{k, 0} L_n(\psi^*, \alpha) = O_p(n^{1/2})$;

for $\ell = 0, 1, \ldots$, $\nabla \lambda_{\sigma, 0} L_n(\psi^*, \alpha) = 0$;

for $k = 2, 3$, $\nabla \lambda_{k, \beta} L_n(\psi^*, \alpha) = O_p(n^{1/2})$;

$\nabla \lambda_{k, \sigma} L_n(\psi^*, \alpha) = O_p(n^{1/2})$;

for $k = 1, \ldots, 4$, $\nabla \lambda_{k, \sigma} L_n(\psi^*, \alpha) = O_p(n^{1/2})$.

Expanding $L_n(\psi, \alpha)$ nine times around $\psi^*$ and using (28) and (29), we can write $L_n(\psi, \alpha) - L_n(\psi^*, \alpha)$ as the sum of relevant terms and remainder term as

$$L_n(\psi, \alpha) - L_n(\psi^*, \alpha) =$$
$$\nabla_\eta L_n^*(\eta - \eta^*) + \frac{1}{2!} (\eta - \eta^*)^T \nabla_\eta \eta^* L_n^*(\eta - \eta^*)$$

$$+ \frac{1}{2!} \left\{ 2 \nabla \lambda_{\mu} \lambda_{\sigma} L_n^* \lambda_{\mu} \lambda_{\sigma} + \lambda_{\sigma}^T \nabla \lambda_{\sigma} \lambda_{\sigma}^T L_n^* \lambda_{\sigma} \right\}$$

$$+ \frac{1}{3!} \left\{ 3 \sum_{i=0}^{q} \sum_{j=0}^{q} \nabla \lambda_{i, j} \eta^T L_n^* \lambda_{i} \lambda_{j} (\eta - \eta^*) + 6 \sum_{i=0}^{q} \nabla \lambda_{i} \eta^T L_n^* \lambda_{i} (\eta - \eta^*) \right\}$$

$$+ \frac{1}{4!} \left\{ \nabla \lambda_{i} \lambda_{\mu} L_n^* \lambda_{\mu}^4 + \sum_{i=0}^{q} \sum_{j=0}^{q} \sum_{k=0}^{q} \sum_{\ell=0}^{q} \nabla \lambda_{i} \lambda_{k} \lambda_{\ell} L_n^* \lambda_{i} \lambda_{k} \lambda_{\ell} \right\}$$

$$+ 4 \sum_{i=1}^{q} \sum_{j=1}^{q} \sum_{k=1}^{q} \nabla \lambda_{i} \lambda_{k}^2 L_n^* \lambda_{i} \lambda_{k} + 6 \lambda_{\sigma}^T \nabla \lambda_{\sigma} \lambda_{\sigma}^T \lambda_{\sigma}^2 L_n^* \lambda_{\sigma} \lambda_{\sigma}$$

$$+ \frac{5}{5!} \nabla \lambda_{\mu} \eta^T L_n^* \lambda_{\mu}^4 (\eta - \eta^*) + \frac{6}{6!} \nabla \lambda_{\mu} \lambda_{\sigma} L_n^* \lambda_{\mu}^5 \lambda_{\sigma}$$

$$+ \frac{1}{8!} \nabla \lambda_{\mu} L_n^* \lambda_{\mu}^8 + R_{1n}(\psi, \alpha),$$

2
where $\nabla L_\ast^n$ denotes the derivative of $L_n(\psi, \alpha)$ evaluated at $(\psi^\ast, \alpha)$, and

$$R_{1n}(\psi, \alpha) =$$

$$O_p(n^{1/2})(||\lambda_{\sigma\beta}||^3 + \lambda_\mu^2||\lambda_{\sigma\beta}|| + \lambda_\mu||\lambda_{\sigma\beta}||^2) + O_p(n)||\hat{\eta}||^3 + O_p(n)(\lambda_\mu + ||\lambda_{\sigma\beta}||)||\hat{\eta}||^2$$

$$+ O_p(n^{1/2})\lambda_\mu^3||\lambda_{\sigma\beta}|| + \sum_{(p,q,r)\in D(4)} O_p(n)\lambda_\mu^p||\lambda_{\sigma\beta}||^q||\hat{\eta}||^r$$

$$+ O_p(n^{1/2})\lambda_\mu^5 + O_p(n^{1/2})\lambda_\mu^4||\lambda_{\sigma\beta}|| + \sum_{(p,q,r)\in D(5)} O_p(n)\lambda_\mu^p||\lambda_{\sigma\beta}||^q||\hat{\eta}||^r$$

$$+ O_p(n^{1/2})\lambda_\mu^6 + \sum_{(p,q,r)\in D(6)} O_p(n)\lambda_\mu^p||\lambda_{\sigma\beta}||^q||\hat{\eta}||^r$$

$$+ \sum_{(p,q,r)\in D(7)} O_p(n)\lambda_\mu^p||\lambda_{\sigma\beta}||^q||\hat{\eta}||^r + \sum_{(p,q,r)\in D(9)} O_p(n)\lambda_\mu^p||\lambda_{\sigma\beta}||^q||\hat{\eta}||^r,$$

with $\hat{\eta} := \eta - \eta^\ast$ and the sets $D(4) - D(9)$ defined by

$$D(4) := \{(p,q,r) : p + q + r = 4, r \neq 0, (p,q,r) \neq (3,0,1)\},$$

$$D(5) := \{(p,q,r) : p + q + r = 5, (p,q,r) \neq (5,0,0), (4,1,0), (4,0,1)\},$$

$$D(6) := \{(p,q,r) : p + q + r = 6, (p,q,r) \neq (6,0,0), (5,1,0)\},$$

$$D(7) := \{(p,q,r) : p + q + r = 7, (p,q,r) \neq (7,0,0)\},$$

$$D(8) := \{(p,q,r) : p + q + r = 8, (p,q,r) \neq (8,0,0)\},$$

$$D(9) := \{(p,q,r) : p + q + r = 9\}.$$ We prove part (a) by showing that the terms in (30)–(35) are written as

$$t_n(\psi, \alpha)^\top S_n - (1/2)t_n(\psi, \alpha)^\top I_n t_n(\psi, \alpha) + [O(||\psi - \psi^\ast||) + o(1)]O_p((1 + ||t_n(\psi, \alpha)||)^2),$$

Henceforth, we suppress $(\psi, \alpha)$ from $t_n(\psi, \alpha)$. The first term in (30), the terms in (31) and the first term in (33) are written as $t_n^\top S_n$, because $\nabla_{\lambda_\mu\lambda_\beta} L_\ast^n \lambda_\mu \lambda_\sigma = \alpha(1 - \alpha) \sum_{i=1}^n H_i^\ast x_i^\top \lambda_\mu \lambda_\beta$, $\nabla_{\lambda_\mu\lambda_\beta} L_\ast^n \mu \lambda_\beta = \alpha(1 - \alpha) \sum_{i=1}^n H_i^\ast x_i^\top \lambda_\mu \lambda_\beta$, $\nabla_{\lambda_\mu\lambda_\beta} L_\ast^n \lambda_\mu \lambda_\beta = \alpha(1 - \alpha) \sum_{i=1}^n H_i^\ast x_i^\top \lambda_\mu \lambda_\beta$, and

$$(1/2)!\lambda_\beta^\top \nabla_{\lambda_\mu\lambda_\beta} L_\ast^n \lambda_\beta = \alpha(1 - \alpha) \sum_{j=1}^q \sum_{k=1}^q \sum_{l=1}^n H_i^\ast x_j^\top x_k^\top x_k^\top = v_\beta(\lambda_\beta)^\top s_{\lambda_{\beta_1}}$$

in view of Propositions A and C(c)(f)(g). The other terms in (30)–(35) except for $R_{1n}(\psi, \alpha)$ are written as $-(1/2)t_n^\top I_n t_n + O_p(||\psi - \psi^\ast||||t_n||^2) + O_p(n^{-1/2}||t_n||^2)$ from a tedious but straightforward calculation in conjunction with Propositions A, B and C.
We complete the proof of part (a) by showing that $R_{1n}(\psi, \alpha)$ satisfies the order in (42). Note that

$$12\lambda^3 = \lambda [b(\alpha)\lambda^4 + 12\lambda^2] - (\lambda^3 b(\alpha))\lambda = O(n^{-1/2}||\lambda||||t_{\lambda n||}). \tag{43}$$

Therefore, the term with $\lambda^3$ in first term in (36) is $O_p(||\lambda||||t_n||)$. The other terms in (36) are either $O_p(||\lambda||||t_n||)$ or $O_p(||\lambda|| + ||\eta||)||t_n||^2$.

All the terms in (37)–(41) with $r \geq 2$ are $O_p(||\psi - \psi^*||n||\eta||^2) = O_p(||\psi - \psi^*||||t_n||^2)$. Hence, we only need to show that the terms in (37)–(41) with $r \leq 1$ are $O_p(||\psi - \psi^*||||t_n||^2)$. The first term in (37) is $O_p(\lambda^2||t_{\lambda n||})$. Of the other terms in $D(4)$ in (37) with $r = 1$, the ones with $\lambda^2 ||\lambda||^3||\eta||$ and $\lambda^2 ||\lambda||^2||\eta||$ are $O_p(||\lambda||n||\lambda||||\eta||) = O_p(||\lambda||||t_n||^2)$, and, similarly, the ones with $||\lambda||^3||\eta||$ are $O_p(||\lambda||||t_n||^2)$ because $\lambda^3 = O(n^{-1/2}||\lambda||||t_{\lambda n||})$, as shown in (43).

Note that $\lambda^5 = (\lambda /b(\alpha))[b(\alpha)\lambda^4 + 12\lambda^2] - (\lambda^3 b(\alpha))\lambda = O(n^{-1/2}||\lambda||||t_{\lambda n||})$. Therefore, the first term in (38) is $O_p(||\lambda||||t_n||)$ and so are the first terms in (39) and (40). The second term in (38) is dominated by the first term in (37). Of the terms in $D(5)$ in (38), the ones with $r = 0$ are those with $\lambda^2 ||\lambda||^2$, $\lambda^3 ||\lambda||^3$, $\lambda^2 ||\lambda||^4$, and $||\lambda||^5$. The term with $\lambda^5$ is $O_p(||\lambda||||t_n||^2)$ because $12\lambda^5 = \lambda^3 [b(\alpha)\lambda^4 + 12\lambda^2] - (\lambda^3 b(\alpha))\lambda = O(n^{-1/2}||\lambda||||t_{\lambda n||}^2)$ while the other terms in $D(5)$ with $r = 0$ are $O_p(||\lambda||||t_n||^2)$ because, for example, the terms with $\lambda^2 ||\lambda||^2$ and $\lambda^2 ||\lambda||^3$ are $O_p(||\lambda||n||\lambda||||t_n||^2) = O_p(||\lambda||||t_n||^2)$. The terms in $D(5)$ with $r = 1$ are $O_p(||\psi - \psi^*||||t_n||^2)$ from a simple calculation. Of the terms in $D(6)$ in (39) with $r = 0, 1$, the one with $\lambda^5 ||\eta||$ is $O_p(||\psi - \psi^*||||t_n||^2)$ because $\lambda^5 = O(n^{-1/2}||\lambda||||t_{\lambda n||})$, and the other terms in $D(6)$ with $r = 0, 1$ are bounded by those in $D(5)$. Of the terms in $D(7)$ in (40), the ones with $\lambda^6 ||\lambda||$ is $O_p(||\lambda||||t_n||^2)$ because $\lambda^6 = O(n^{-1/2}||\lambda||||t_{\lambda n||})$, and the other terms in $D(7)$ are bounded by those in $D(6)$. The terms in $D(8)$ are bounded by those in $D(7)$. Of the terms in $D(9)$, the one with $\lambda^9$ is $O_p(||\lambda||||t_n||^2)$ because $\lambda^9 = (\lambda/ b(\alpha))[b(\alpha)\lambda^4 + 12\lambda^2] - 12 (\lambda^3 /b(\alpha))\lambda = \lambda^5 O(n^{-1/2}||t_{\lambda n||}) + \lambda^3 O(n^{-1}||t_{\lambda n||^2}) = O(n^{-1/2}||\lambda||||t_{\lambda n||}^2)$, and the other terms in $D(9)$ are bounded by those in $D(8)$. This proves part (a).

Part (b) follows from the central limit theorem. Part (c) follows from the law of large numbers, where the nonsingularity of $I$ holds under Assumption 2(b) because the off-diagonal elements of $I = E[s_i s_j^T]$ that involve the interaction terms of $H_i^c$ and $H_j^c$ are zero for $j \neq k$ by the property of Hermite polynomial.
A.3 Proof of Proposition 3

We suppress the subscript $\alpha$ from $\psi_\alpha, \hat{\psi}_\alpha$, and $\psi^*_\alpha$. We suppress $(\psi, \alpha)$ from $t_n(\psi, \alpha)$, and let $\tilde{t}_n := t_n(\hat{\psi}; \alpha)$.

The proof of part (a) closely follows the proof of Theorem 1 of Andrews (1999) (A99, hereafter). Let $T_n := T_n^{1/2} \tilde{t}_n$. Then, in view of (14), we have

$$o_p(1) \leq L_n(\hat{\psi}, \alpha) - L_n(\psi^*, \alpha) = T_n' T_n^{-1/2} S_n - \frac{1}{2} ||T_n||^2 + R_n(\hat{\psi}, \alpha) = O_p(||T_n||) - \frac{1}{2} ||T_n||^2 + (1 + ||I_n^{-1/2} T_n||^2) o_p(1) = ||T_n|| O_p(1) - \frac{1}{2} ||T_n||^2 + o_p(||T_n||) + o_p(||T_n||^2) + o_p(1),$$

where the third equality holds because $T_n^{-1/2} S_n = O_p(1)$ and $R_n(\hat{\psi}, \alpha) = o_p((1 + ||I_n^{-1/2} T_n||)^2)$ from Propositions 1 and 2. Rearranging this equation yields $||T_n||^2 \leq 2 ||T_n|| O_p(1) + o_p(1)$. Denote the $O_p(1)$ term by $\zeta_n$. Then, $||T_n||^2 - \zeta_n^2 \leq \zeta_n^2 + o_p(1) = O_p(1)$; taking its square roots gives $||T_n|| \leq O_p(1)$. In conjunction with $I_n \to_p I$, we obtain $\tilde{t}_n = O_p(1)$, and part (a) follows.

For parts (b)(c), Define

$$W_n := I^{-1} S_n = \begin{bmatrix} W_{\eta m} \\ W_{\lambda \eta} \end{bmatrix}, \quad W_{\eta, \lambda} := W_{\eta m} - E[W_{\eta m} W_{\lambda \eta}^T] \text{Var}[W_{\lambda m}]^{-1} W_{\lambda m}, \quad t_{\eta, \lambda} := t_{\eta m} - E[W_{\eta m} W_{\lambda \eta}^T] \text{Var}[W_{\lambda m}]^{-1} t_{\lambda m}.$$

For any $\psi$ such that $t_n = O_p(1)$, we can write $2[L_n(\psi, \alpha) - L_n(\psi^*, \alpha)]$ as

$$2[L_n(\psi, \alpha) - L_n(\psi^*, \alpha)] = W_n^T I W_n - (t_n - W_n)^T I (t_n - W_n) + o_p(1) = A_n(t_{\eta, \lambda}) + B_n(t_{\lambda, \eta}) + o_p(1), \quad \text{(44)}$$

where

$$A_n(t_{\eta, \lambda}) = W_{\eta, \lambda}^T I_{\eta m} W_{\eta, \lambda} - (t_{\eta, \lambda} - W_{\eta, \lambda})^T I_{\eta m} (t_{\eta, \lambda} - W_{\eta, \lambda}), \quad B_n(t_{\lambda, \eta}) = W_{\lambda m}^T I_{\lambda m} W_{\lambda m} - (t_{\lambda m} - W_{\lambda m})^T I_{\lambda m} (t_{\lambda m} - W_{\lambda m}).$$

Note that $W_{\eta, \lambda} = I^{-1} S_{\eta m}, \nabla_{\eta} l(y|x; z; \psi^*, \alpha)$ equals the score of the one-component model as shown in (9), and the set of admissible values of $\tilde{t}_{\eta, \lambda}$ approaches to $\mathbb{R}^{q+2}$. Therefore,
When no conditioning variable $X$ is present, $B_n(\hat{t}_{\lambda n}) \to_d \chi^2(2)$ because the set of admissible values of $(\hat{t}_{\mu\sigma n}, \hat{t}_{\mu\tau n})^\top$ approaches to $\mathbb{R}^2$, and part (b) follows.

For part (c), consider the sets $\Theta_1^\lambda := \{\lambda \in \Theta_\lambda : |\lambda_\mu| \geq n^{-1/8}(\ln n)^{-1}\}$ and $\Theta_2^\lambda := \{\lambda \in \Theta_\lambda : |\lambda_\mu| < n^{-1/8}(\ln n)^{-1}\}$. For $j = 1, 2$, define $\hat{\psi}^j := \arg\max_{\psi \in \Theta(\epsilon_\alpha), \lambda \in \Theta_\lambda^j} L_n(\psi, \alpha)$ and $\hat{t}_n^j := t_n(\hat{\psi}^j, \alpha)$, which is $O_p(1)$ from part (a). From the same argument as in (45), we have

\[
2[L_n(\hat{\psi}, \alpha) - L_{0,n}(\hat{\gamma}_0, \hat{\theta}_0, \hat{\sigma}_0^2)] = \max\{B_n(\hat{t}_{\lambda n}^1), B_n(\hat{t}_{\lambda n}^2)\} + o_p(1). \tag{46}
\]

Observe that, because $\hat{t}_{\mu\sigma n}^1 = O_p(1)$ and $\hat{t}_{\beta\tau n}^1 = O_p(1)$, it follows from $|\hat{\lambda}_\sigma^1| \geq n^{-1/8}(\ln n)^{-1}$ that $\hat{\lambda}_\sigma^1 = O_p(n^{-3/8}\ln n)$ and $\hat{\lambda}_{\beta\lambda}^1 = O_p(n^{-3/8}\ln n)$. Consequently, $\hat{t}_{\lambda n}^1$ satisfies

\[
\hat{t}_{\beta\sigma n}^1 = o_p(1), \quad \hat{t}_{\beta\tau n}^1 = o_p(1), \quad \hat{t}_{\mu\tau n}^1 = n^{1/2}\alpha(1 - \alpha)b(\alpha)(\hat{\lambda}_\sigma^1)^4 + o_p(1). \tag{47}
\]

Define $\hat{t}_{\lambda n}^2 := \arg\max_{t_\lambda \in \hat{\Theta}_\lambda} B_n(t_\lambda)$, where $\hat{\Theta}_\lambda^1 := t_{\lambda n}(\Theta_\psi(\epsilon_\alpha), \alpha) \cap \{t_{\beta\sigma n} = 0, t_{\beta\tau n} = 0, t_{\mu\tau n} = n^{1/2}\alpha(1 - \alpha)b(\alpha)\hat{\lambda}_\sigma^1\}$. Then, we have $B_n(\hat{t}_{\lambda n}^1) \geq B_n(\hat{t}_{\lambda n}^2) + o_p(1)$ from the definition of $\hat{t}_{\lambda n}^1$, definition of $B_n(t_{\lambda n}^1)$, and (47). Note that $\hat{t}_{\lambda n}^2$ satisfies $\hat{t}_{\mu\tau n}^2 = n^{1/2}\alpha(1 - \alpha)12(\hat{\lambda}_\sigma^2)^2 + o_p(1)$ because $|\lambda_\mu| < n^{-1/8}(\ln n)^{-1}$ if $\lambda \in \Theta_\lambda^2$. Define $\hat{t}_{\lambda n}^2 := \arg\max_{t_\lambda \in \hat{\Theta}_\lambda^2} B_n(t_\lambda)$, where $\hat{\Theta}_\lambda^2 := t_{\lambda n}(\Theta_\psi(\epsilon_\alpha), \alpha) \cap \{t_{\mu\tau n} = n^{1/2}\alpha(1 - \alpha)12\hat{\lambda}_\sigma^2\}$. Then, a similar argument as above gives $B_n(\hat{t}_{\lambda n}^2) \geq B_n(\hat{t}_{\lambda n}^1) + o_p(1)$.

For $B_n(\hat{t}_{\lambda n}^1)$ and $B_n(\hat{t}_{\lambda n}^2)$, observe that the parameter space of $n^{-1/2}\hat{t}_{\lambda n}^1$ and $n^{-1/2}\hat{t}_{\lambda n}^2$ are locally approximated by the cone $\Lambda_\lambda^1$ and $\Lambda_\lambda^2$, respectively, from Lemma 3 of A99 because Assumption 5* of A99 is satisfied with $B_T = n^{1/2}$. Therefore,

\[
(B_n(\hat{t}_{\lambda n}^1), B_n(\hat{t}_{\lambda n}^2)) \to_d ((\hat{t}_{\lambda n}^1)^\top I_{\lambda n}\hat{t}_{\lambda n}^1, (\hat{t}_{\lambda n}^2)^\top I_{\lambda n}\hat{t}_{\lambda n}^2), \tag{48}
\]

follows from Theorem 3(c) of A99 because Assumption 2 of A99 holds trivially for $B_n(t_{\lambda n})$, Assumption 3 of A99 is satisfied by Proposition 2(b)(c), and Assumption 4 of A99 is satisfied by part (a). Because $\max\{B_n(\hat{t}_{\lambda n}^1), B_n(\hat{t}_{\lambda n}^2)\} \geq \max\{B_n(\hat{t}_{\lambda n}^1), B_n(\hat{t}_{\lambda n}^2)\}$ from the definition of $\hat{\psi}$, we have $\max\{B_n(\hat{t}_{\lambda n}^1), B_n(\hat{t}_{\lambda n}^2)\} = \max\{B_n(\hat{t}_{\lambda n}^1), B_n(\hat{t}_{\lambda n}^2)\} + o_p(1)$, and part (c) follows from (46) and (48). □
A.4 Proof of Proposition 4

Using $X_0$, reparameterize the model so that it does not have a constant term. Then, by expanding $L_n(\psi_\alpha, \alpha) - L_n(\psi_\alpha^*, \alpha)$ as in the proof of Proposition 2 with a suitably defined $\psi_\alpha$, we obtain the same expansion as (14) with (re-)defining $x^0 := (x_0, x_1, \ldots, x_q)^\top$, $\beta^0 := (\beta_0, \beta_1, \ldots, \beta_q)^\top$, $\eta := (\gamma^T, \nu_0^T, \nu_\sigma)^T$, $\lambda_\beta^0 := (\lambda_0, \lambda_1, \ldots, \lambda_q)^T$, $\lambda^0 := ((\lambda_\beta^0)^\top, \lambda_\sigma)^T$, and $\lambda_\sigma^0 := (\lambda_0^2, \ldots, \lambda_q^2, \ldots, \lambda_q^2)^T$.

$$s_{\eta}^0 := \begin{pmatrix} H_{i^*}^1 Z_i \\ H_{i^*}^1 X_i^0 \\ H_{i^*}^2 X_i^0 \\ H_{i^*}^2 U_2^0 \end{pmatrix}, \quad s_{\lambda}^0 := \begin{pmatrix} H_{i^*}^4 X_i^0 \\ H_{i^*}^2 X_i^0 \\ H_{i^*}^2 U_2^0 \end{pmatrix}, \quad t_n(\psi_\alpha, \alpha) := \begin{pmatrix} n^{1/2}(\eta - \eta^*) \\ n^{1/2} \alpha(1 - \alpha) 12 \lambda_\sigma^2 \\ n^{1/2} \alpha(1 - \alpha) 6 \lambda_\beta^0 \lambda_\sigma \\ n^{1/2} \alpha(1 - \alpha) \bar{\nu}_\beta(\lambda_\beta^0) \end{pmatrix},$$

where $\bar{U}_2 := (X_0^2, (\bar{U}_2^0)^\top)^\top$, $\bar{U}_2 := (X_q^2, \ldots, X_2 X_2, \ldots, 2X_0 X_2, 2X_1 X_2, \ldots, 2X_{q-1} X_q)^T$, and $\bar{\nu}_\beta(\lambda_\beta^0) := (\lambda_0^2, \lambda_1^2, \ldots, \lambda_q^2, \lambda_0, \lambda_1, \ldots, \lambda_{q-1}, \lambda_q)^T$.

Observe that, because $X_0^2 = 1 - X_q^2$, the term $H_{i^*}^2 \nu_\sigma + H_{i^*}^2 X_0^2 \lambda_0^2 + H_{i^*}^2 X_q^2 \lambda_q^2$ in the expansion can be written as $H_{i^*}^2 (\nu_\sigma + \lambda_q^2) + H_{i^*}^2 X_0^2 (\lambda_0^2 - \lambda_0^2)$. Further, $\lambda_0^2$ in $(\nu_\sigma + \lambda_q^2)$ can be absorbed into $\nu_\sigma$. In view of this, the asymptotic distribution of $LR_n(\epsilon_1)$ is characterized by the score vector $s_{\eta}^0 := ((s_{\eta}^0)^\top, (s_{\lambda}^0)^\top)^\top$, where

$$s_{\eta}^0 := \begin{pmatrix} H_{i^*}^1 Z_i \\ H_{i^*}^1 X_i^0 \\ H_{i^*}^2 X_i^0 \\ H_{i^*}^2 U_2^0 \end{pmatrix}, \quad s_{\lambda}^0 := \begin{pmatrix} H_{i^*}^4 X_i^0 \\ H_{i^*}^2 X_i^0 \\ H_{i^*}^2 U_2^0 \end{pmatrix}.$$

Define $v_\beta^0(\lambda_\beta^0) := (\lambda_0^2 - \lambda_0^2, \lambda_1^2, \ldots, \lambda_q^2, \lambda_0, \lambda_1, \ldots, \lambda_{q-1}, \lambda_q)^T$ and $\Lambda_\chi^0 := \{t_\chi = (t_{\mu^*}, (t_{\beta^0})^\top)^\top : t_{\mu^*} = 12 \lambda_\sigma, t_{\beta^0} = 6 \lambda_\beta^0 \lambda_\sigma, v_\beta^0(\lambda_\beta^0) \}$ for some $\chi^0$. Define $W_\chi^0$ and $\mathcal{I}_{\lambda, \eta}$ as $W_\lambda$ and $\mathcal{I}_{\lambda, \eta}$ that appear in (16) but using $\mathcal{I}^0 := E(s_{\eta}^0(s_{\lambda}^0)^\top)$ in place of $\mathcal{I}$. Repeating the proof of Proposition 3 with these definitions gives $LR_n(\epsilon_1) \rightarrow_d (t_{\lambda}^0, \mathcal{I}_{\lambda, \eta})^\top$ in place of $\mathcal{I}_{\lambda, \eta}$, where $t_{\chi}^0$ is defined as $t_{\chi}$ in (16) but using $(t_{\lambda}^0, \Lambda_\chi^0, W_\lambda^0)$ in place of $(t_{\lambda}, \Lambda_\lambda^2, W_\lambda)$.

A.5 Proof of Proposition 5

For $h = 1, \ldots, m_0$, let $\mathcal{N}_h^* \subset \Theta_{m_0+1}(\epsilon_0)$ be a sufficiently small closed neighborhood of $\Theta_{1h}$, such that $(\theta_1, \sigma_1^2) < \cdots < (\theta_{h-1}, \sigma_{h-1}^2) < (\theta_h, \sigma_h^2), (\theta_{h+1}, \sigma_{h+1}^2) < (\theta_{h+2}, \sigma_{h+2}^2) < \cdots < (\theta_{m_0+1}, \sigma_{m_0+1}^2)$ and $\alpha_h, \alpha_{h+1} > 0$ hold and $\Gamma_{1h} \notin \mathcal{N}_h^*$ if $k \neq h$. For $\Theta_{m_0+1} \subset \mathcal{N}_h^*$, we introduce
the following one-to-one reparameterization, which is similar to (6):

\[
\delta_h := \alpha_h + \alpha_{h+1}, \quad \tau := \alpha_h / (\alpha_h + \alpha_{h+1}),
\]

\[
(\delta_1, \ldots, \delta_{h-1}, \delta_{h+1}, \ldots, \delta_{m_0-1})^\top := (\alpha_1, \ldots, \alpha_{h-1}, \alpha_{h+2}, \ldots, \alpha_{m_0})^\top,
\]

\[
\begin{pmatrix}
\theta_h \\
\theta_{h+1} \\
\sigma_h^2 \\
\sigma_{h+1}^2
\end{pmatrix} =
\begin{pmatrix}
\nu_\theta + (1 - \tau)\lambda_\theta \\
\nu_\theta - \tau\lambda_\theta \\
\nu_\sigma + (1 - \tau)(2\lambda_\sigma + C_1\lambda_\mu^2) \\
\nu_\sigma - \tau(2\lambda_\sigma + C_2\lambda_\mu^2)
\end{pmatrix},
\]

(49)

where \( \delta_{m_0} = 1 - \sum_{j=1}^{m_0-1} \delta_j \), \( C_1 = -(1/3)(1 + \tau) \), and \( C_2 = (1/3)(2 - \tau) \), and we suppress the dependence of \((\lambda_\theta, \nu_\theta, \lambda_\sigma, \nu_\sigma)\) on \( \tau \). With this reparameterization, the null restriction \((\theta_h, \sigma_h^2) = (\theta_{h+1}, \sigma_{h+1}^2)\) implied by \( H_{0,1h} \) holds if and only if \((\lambda_\theta, \lambda_\sigma) = (0, 0)\). Collect the reparameterized parameters except for \( \tau \) into one vector \( \psi^h_\tau \), and let \( \psi^{h*}_\tau \) denote its true value. Define the reparameterized density as

\[
g^h(y|x, z; \psi^h_\tau, \tau) := \delta_h \left[ \tau f (y|x, z; \gamma, \nu_\theta + (1 - \tau)\lambda_\theta, \nu_\sigma + (1 - \tau)(2\lambda_\sigma + C_1\lambda_\mu^2)) \right. \\
\left. + (1 - \tau)f (y|x, z; \gamma, \nu_\theta - \tau\lambda_\theta, \nu_\sigma - \tau(2\lambda_\sigma + C_2\lambda_\mu^2)) \right] \\
+ \sum_{j=1}^{h-1} \delta_j f (y|x, z; \gamma, \theta_j, \sigma_j^2) + \sum_{j=h+1}^{m_0} \delta_j f (y|x, z; \gamma, \theta_j, \sigma_j^2) + \sum_{j=h+1}^{m_0} \delta_j f (y|x, z; \gamma, \theta_j, \sigma_j^2).
\]

Define the local MLE of \( \psi^h_\tau \) by

\[
\hat{\psi}^h_\tau := \arg\max_{\psi^h_\tau \in \mathcal{N}_h^*} L^h_n(\psi^h_\tau, \tau),
\]

(50)

where \( L^h_n(\psi^h_\tau, \tau) := \sum_i^n \ln[g^h(Y_i|X_i, Z_i; \psi^h_\tau, \tau)] \). Because \( \psi^{h*}_\tau \) is the only parameter value in \( \mathcal{N}_h^* \) that generates true density, \( \hat{\psi}^h_\tau - \psi^{h*}_\tau \to p_\tau(1) \) follows from Proposition D. Define the LRT statistic for testing \( H_{0,1h} \) as \( LR_{n,1h}(\epsilon_\tau) := \max_{\tau \in [\epsilon_\tau, 1-\epsilon_\tau]} 2\{L^h_n(\hat{\psi}^h_\tau, \tau) - L_{0,n}(\hat{\theta}_{m_0})\} \) for some \( \epsilon_\tau \in (0, 1/2) \).

In view of Proposition D, the stated result holds if

\[
(LR_{n,11}(\epsilon_\tau), \ldots, LR_{n,1m_0}(\epsilon_\tau))^\top \to_d (v_1, \ldots, v_{m_0})^\top
\]

(51)

for any \( \epsilon_\tau \in (0, 1/2) \), where \( v_h = \max\{ (t_{\lambda,h}^1)^\top I_{\lambda,\eta} t_{\lambda,h}^1, (t_{\lambda,h}^2)^\top I_{\lambda,\eta} t_{\lambda,h}^2 \} \). We proceed to show (51). Observe that, as in (10), the first, second, and third derivatives of \( \ln[g^h(y|x, z; \psi^h_\tau, \tau)] \) w.r.t. \( \lambda_\theta \) and its first derivative w.r.t. \( \lambda_\sigma \) become zero when evaluated at \( \psi^h_\tau = \psi^{h*}_\tau \). Consec-
quently, \( L_n^h(\psi^h_\tau, \tau) - L_n(\psi^h_\tau, \tau) \) admits the same expansion (14) as \( L_n(\psi_\alpha, \alpha) - L_n(\psi^*_\alpha, \alpha) \) by replacing \((t_n(\psi_\alpha, \alpha), S_n, I_n, R_n(\psi_\alpha, \alpha))\) in (11)–(14) with \((t_{n, h}(\psi^h_\tau, \tau), S_{n, h}, I^h_n, R_{n, h}(\psi^h_\tau, \tau))\), where \((S_{n, h}, I^h_n)\) is defined similarly to \((S_n, I_n)\) but using \((s_{n, h}, s^h_\lambda)\) in place of \((s_n, s^h_\lambda)\).

Applying the proof of Proposition 2(b,c), we have \( S_{n, h} \rightarrow_d S_h \sim N(0, I^h) \) and \( I^h_n \rightarrow_p I^h := E[S_{n, h} S^\top_{n, h}] \). Define \( W_{n, h} \) similarly to \( W_n \) but using \((S_n, I_n)\) in place of \((S_{n, h}, I^h_n)\) in the proof of Proposition 3. Then, (51) follows from applying the proofs of Propositions 2 and 3 for each local MLE by replacing \((W_n, \hat{t}^1_\lambda, \hat{t}^2_\lambda, I\lambda, \eta)\) with \((W_{n, h, h}, \hat{t}^1_{\lambda, h}, \hat{t}^2_{\lambda, h}, I^h_{\lambda, \eta})\), and collecting the results while noting that \((S^\top_{n, 1}, \ldots, S^\top_{n, m_0}) \rightarrow_d (S^\top_1, \ldots, S^\top_{m_0})\). □

### A.6 Proof of Proposition 6

Under \( H_{0, 2h} \), we obtain, for \( \vartheta_{m_0+1} \in \mathcal{Y}_{2h}^* \),

\[
E \left[ \nabla_{\vartheta_h} \ln f_{m_0+1}(Y_i; \vartheta_{m_0+1}) \right]^2 \leq \int \frac{\{f(y; \mu_h^*, \sigma^2_h)\}^2}{\sum_{j=1}^{m_0} \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*)} dy + \frac{1}{m_0} \max_j \left\{ \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*) \right\}^2 \leq \frac{1}{m_0} \max_j \left\{ \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*) \right\}.
\]

The latter two terms on the right-hand side of (52) are bounded because

\[
f(y; \mu_{m_0}^*, \sigma^2_{m_0}^*) / \sum_{j=1}^{m_0} \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*) \leq (1/\alpha^*_m) \text{ for any } y, \text{ and } f(y; \mu, \sigma^2) \text{ integrates to one.}
\]

Thus, the left-hand side of (52) is infinity if and only if the first term on the right-hand side of (52) is infinity.

Because \( \max_j a_j \leq \sum_{j=1}^{m_0} a_j \leq m_0 \max_j a_j \) holds for general \( \{a_j\}_{j=1}^{m_0} \), we obtain

\[
\frac{1}{m_0} \max_j \left\{ \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*) \right\}^2 \leq \frac{1}{m_0} \max_j \left\{ \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*) \right\}^2 \leq \frac{1}{m_0} \max_j \left\{ \alpha_j^* f(y; \mu_j^*, \sigma^2_j^*) \right\}.
\]

Without loss of generality, assume that \( \sigma^2_{m_0}^* = \max \{\sigma^2_1^*, \ldots, \sigma^2_{m_0}^*\} \) and the maximum is unique. Then, there exists \( M \in (0, \infty) \), such that \( \max_j \{\alpha_j^* f(y; \mu_j^*, \sigma^2_j^*)\} = \alpha_{m_0}^* f(y; \mu_{m_0}^*, \sigma^2_{m_0}^*) \) when \( |y| \geq M \). Note that

\[
\frac{\{f(y; \mu_h, \sigma^2_h)\}^2}{f(y; \mu_{m_0}^*, \sigma^2_{m_0}^*)} = \frac{\sigma^2_{m_0}^*}{(2\pi)^{1/2}\sigma_h^*} \exp \left\{ -\frac{1}{2\sigma_h^*} (y - \mu_h)^2 + \frac{1}{2\sigma^2_{m_0}^*} (y - \mu_{m_0}^*)^2 \right\}.
\]

The stated result follows because the integral of this over \( |y| \geq M \) is finite if \( \sigma^2_h / \sigma^2_{m_0}^* < 2 \) and infinite if \( \sigma^2_h / \sigma^2_{m_0}^* > 2 \) while, when \( \sigma^2_h = 2\sigma^2_{m_0}^* \), it is finite if \( \mu_h = \mu_{m_0}^* \) and infinite if \( \mu_h \neq \mu_{m_0}^* \). □
A.7 Proof of Proposition 7

We suppress $(\tau_0)$ from $\varphi^{h(k)}(\tau_0)$ and $\tau^k(\tau_0)$. For $j = 1, 2$, let $\omega_{n,h}^j$ be the sample counterpart of $(\hat{\theta}^j_{\omega,h})^T \mathbf{I}_{\lambda,h} \hat{t}^j_{\omega,h}$ in Proposition 5 such that the local LRT statistic satisfies $2[L_n(\hat{\psi}_N, \tau) - L_{0,n}(\hat{\theta}_{m_0})] = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$, where $\hat{\psi}_N$ is the local MLE defined in (50).

Observe that, from Assumption 6 and $|x| \leq 1 + |x|^6$, we have $\tilde{p}_n(\sigma_j^2) - \hat{p}_n(\sigma_j^2) = o_p(n^{1/6}(\sigma_j^2 - \sigma_j^2) = o_p(1 + n^{1/2}(\sigma_j^2 - \sigma_j^2)^3) = o_p(1 + n^{1/2}(\lambda_j^3 + \lambda_j^6)$). Therefore, in view of (43) in the proof of Proposition 2, for any $\varphi^h$ whose corresponding $t_{n,h}(\psi^h_N)$ is $O_p(1)$, we have

$$PL_n^h(\varphi^h, \tau) - PL_n^h(\varphi^{h*}, \tau) = L_n^h(\varphi^h, \tau) - L_n^h(\varphi^{h*}, \tau) + o_p(1).$$  (53)

First, we show $EM_n^{h(1)} = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$. Given $\tau_0$, $\varphi^{h(1)}$ maximizes $PL_n^h(\varphi^h, \tau_0)$ under the constraint $\varsigma \in \hat{\Omega}_h$. Because $\varphi^{h*}$ is the only value of $\varphi^h$ that yields the true density if $\varsigma \in \hat{\Omega}_h$ and $\tau \in (0, 1)$, $\varphi^{h(1)}$ equals a reparameterized penalized local MLE in the neighborhood of $\varphi^{h*}$. Hence, $PL_n^h(\varphi^{h(1)}, \tau_0) \geq PL_n(\theta_{m_0+1}) + o_p(1)$ holds, and Proposition E gives $\varphi^{h(1)} - \varphi^{h*} = o_p(1)$. It follows from applying the argument in the proof of Propositions 3 and 5 that $t_{n,h}(\psi^h)$ corresponding $\varphi^{h(1)}$ is $O_p(1)$ and that $2[PL_n^h(\varphi^{h(1)}, \tau_0) - PL_{0,n}(\hat{\theta}_{m_0})] = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$. Therefore, $EM_n^{h(1)} = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$ follows from (53).

We proceed to show that $EM_n^{h(K)} = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$. Because $\varphi^{h(1)} - \varphi^{h*} = o_p(1)$ and $\tau^{(1)} - \tau_0 = 0$, it follows from Propositions E and F and induction that $\varphi^{h(K)} - \varphi^{h*} = o_p(1)$ and $\tau^{(K)} - \tau_0 = o_p(1)$ for any finite $K$. Because a generalized EM step never decreases the likelihood value (Dempster et al., 1977) and $\tau^{(1)} = \tau_0$, we have $PL_n^h(\varphi^{h(K)}, \tau^{(K)}) \geq PL_n^h(\varphi^{h(1)}, \tau_0)$. Let $\varphi^h$ be the maximizer of $PL_n^h(\varphi^h, \tau^{(K)})$ in an arbitrary small closed neighborhood of $\varphi^{h*}$, then we have $PL_n^h(\varphi^{h}, \tau^{(K)}) \geq PL_n^h(\varphi^{h(K)}, \tau^{(K)}) + o_p(1)$ from the consistency of $\varphi^{h(K)}$. Thus, $2[PL_n^h(\varphi^{h(K)}, \tau^{(K)}) - L_{0,n}(\hat{\theta}_{m_0})] = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$ holds because both $2[PL_n^h(\varphi^{h}, \tau^{(K)}) - L_{0,n}(\hat{\theta}_{m_0})]$ and $2[PL_n^h(\varphi^{h(1)}, \tau_0) - L_{0,n}(\hat{\theta}_{m_0})]$ can be written as $\max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$. Further, because $PL_n^h(\varphi^{h(K)}, \tau^{(K)}) \geq PL_n^h(\varphi^{h(1)}, \tau_0) \geq PL_n(\theta_{m_0+1}) + o_p(1)$, applying the proof of Proposition 3(a) to $(\varphi^{h(K)}, \tau^{(K)})$ gives that $t_{n,h}(\psi^h)$ corresponding $\varphi^{h(K)}$ is $O_p(1)$, and $EM_n^{h(K)} = \max\{\omega_{1,n,h}^1, \omega_{2,n,h}^2\} + o_p(1)$ holds for all $h$ from (53). The stated result follows from the definition of $EM_n^{K}$. □

A.8 Proof of Proposition 8

Let $\psi_\alpha$ be the value of $\psi_\alpha$ under $H_{(\psi_\alpha, \Delta)}$ and define $V_n = L_n(\psi_\alpha, \alpha^*) - L_{0,n}(\gamma^*, \theta^*, \sigma^{2*})$. Under the null distribution, we have $(LR_n(\epsilon_1), V_n) \rightarrow_d (\max\{\sup_{t_{\lambda} \in \Lambda^1} Q(t_{\lambda}), \sup_{t_{\lambda} \in \Lambda^2} Q(t_{\lambda})\}, V)$, where $Q(t_\lambda) = 2t^T_{\lambda} \mathbf{I}_{\lambda,n} W_{\lambda} - t^T_{\lambda} \mathbf{I}_{\lambda,n} t_{\lambda}$ and $V = \Delta^T \mathbf{I}_{\lambda,n} W_{\lambda} - (1/2) \Delta^T \mathbf{I}_{\lambda,n} \Delta$. From Le
Cam’s third lemma, the limiting distribution of $LR_n(\epsilon_1)$ under $H_{(\alpha^*, \Delta)}^n$ can be determined by the joint null distribution of $(Q(t_\lambda), V)$ given by
\[
\begin{pmatrix}
Q(t_\lambda) \\
V
\end{pmatrix} \sim N \left( \begin{pmatrix}
-t_\lambda^\top \mathcal{I}_{\lambda, \eta} t_\lambda \\
-(1/2)\Delta^\top \mathcal{I}_{\lambda, \eta} \Delta
\end{pmatrix}, \begin{pmatrix}
4t_\lambda^\top \mathcal{I}_{\lambda, \eta} t_\lambda & 2t_\lambda^\top \mathcal{I}_{\lambda, \eta} \Delta \\
2\Delta^\top \mathcal{I}_{\lambda, \eta} t_\lambda & \Delta^\top \mathcal{I}_{\lambda, \eta} \Delta
\end{pmatrix} \right).
\]

Applying Le Cam’s third lemma, we obtain the limiting distribution of $LR_n(\epsilon_1)$ under $H_{(\alpha^*, \Delta)}^n$ as $\max\{\sup_{t_\lambda \in \Lambda_{\lambda}^1} Q_\Delta(t_\lambda), \sup_{t_\lambda \in \Lambda_{\lambda}^2} Q_\Delta(t_\lambda)\}$, where $Q_\Delta(t_\lambda) \sim N(2t_\lambda^\top \mathcal{I}_{\lambda, \eta} \Delta - t_\lambda^\top \mathcal{I}_{\lambda, \eta} t_\lambda, 4t_\lambda^\top \mathcal{I}_{\lambda, \eta} t_\lambda)$. Because $W_\lambda \sim N(0, \mathcal{I}_{\lambda, \eta}^{-1})$, writing $Q_\Delta(t_\lambda) = 2t_\lambda^\top \mathcal{I}_{\lambda, \eta} (W_\lambda + \Delta) - t_\lambda^\top \mathcal{I}_{\lambda, \eta} t_\lambda = (W_\lambda + \Delta) \mathcal{I}_{\lambda, \eta} (W_\lambda + \Delta) - \{t_\lambda - (W_\lambda + \Delta)\}^\top \mathcal{I}_{\lambda, \eta} \{t_\lambda - (W_\lambda + \Delta)\}$ and using $(t_{\lambda, \Delta}^j)\mathcal{I}_{\lambda, \eta} (t_{\lambda, \Delta}^j - (W_\lambda + \Delta)) = 0$ for $j = 1, 2$ gives the stated limiting distribution of $LR_{1,n}(\epsilon_1)$ under $H_{(\alpha^*, \Delta)}^n$. The limiting distribution of $EM_n$ follows because $EM_n = LR_{1,n}(\epsilon_1) + o_p(1)$ from the proof of Proposition 7. □

### B Auxiliary results and their proofs

**Proposition A.** Let $\phi(x) := (2\pi)^{-1/2} \exp(-x^2/2)$ denote the density of $N(0, 1)$, and let $H^n(x)$ denote the Hermite polynomial of order $n$ ($H^0(x) = 1, H^1(x) = x, H^2(x) = x^2 - 1, H^3(x) = x^3 - 3x, H^4(x) = x^4 - 6x^2 + 3$). Then, the following holds for any nonnegative integer $k$ and $\ell$:
\[
\nabla^k \phi \left( \frac{x - \mu}{\sigma} \right) = \left( \frac{1}{\sigma} \right)^{k+2\ell+1} H^{k+2\ell} \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right).
\]

**Proof** The stated result holds trivially when $k = \ell = 0$. Suppose the stated result holds when $k + 2\ell = n$. First, differentiating $(1/\sigma)^{n+1} H^n((x - \mu)/\sigma) \phi((x - \mu)/\sigma)$ with respect to $\mu$ and using the relation $H^{n+1}(x) = x H^n(x) - \nabla H^n(x)$ give
\[
\nabla^k \left( \left( \frac{1}{\sigma} \right)^{n+1} H^n \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right) \right) = \left( \frac{1}{\sigma} \right)^{n+2} H^{n+1} \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right).
\]
\[\text{ (54)}\]
Second, differentiating \((1/\sigma)^{n+1}H^n((x - \mu)/\sigma)\phi((x - \mu)/\sigma)\) with respect to \(\sigma^2\) gives

\[
\nabla_{\sigma^2} \left[ \left( \frac{1}{\sigma} \right)^{n+1} H^n \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right) \right] \\
= \left( \frac{\partial}{\partial \sigma^2} \frac{1}{\sigma^{n+1}} \right) H^n \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right) \\
+ \left( \frac{1}{\sigma} \right)^{n+1} \left[ \nabla H^n \left( \frac{x - \mu}{\sigma} \right) + H^n \left( \frac{x - \mu}{\sigma} \right) \left( -\frac{x - \mu}{\sigma} \right) \right] \phi \left( \frac{x - \mu}{\sigma} \right) \left( -\frac{x - \mu}{2\sigma^3} \right) \\
= -\frac{n+1}{2} \frac{1}{\sigma^{n+3}} H^n \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right) + \frac{1}{2} \left( \frac{1}{\sigma} \right)^{n+3} H^{n+1} \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right) \left( \frac{x - \mu}{\sigma} \right) \\
= \frac{1}{2} \left( \frac{1}{\sigma} \right)^{n+3} H^{n+2} \left( \frac{x - \mu}{\sigma} \right) \phi \left( \frac{x - \mu}{\sigma} \right),
\]

where the third equality follows from the relation \(H^{n+1}(x) = xH^n(x) - \nabla H^n(x)\), and the last equality follows from the relation \(H^{n+2}(x) = xH^{n+1}(x) - (n+1)H^n(x)\). Using the chain rule, we obtain \(\nabla_{\sigma^2/2}[\left(1/\sigma\right)^{n+1}H^n((x - \mu)/\sigma)\phi((x - \mu)/\sigma)] = \left(1/\sigma\right)^{n+3}H^{n+2}((x - \mu)/\sigma)\phi((x - \mu)/\sigma)\), and the stated result follows from this and (54).

**Proposition B.** Let \(h(x; \beta)\) be the density function of a random variable \(X\) with parameter \(\beta\). Then, \(E_{\beta'}[\nabla g_k h(x; \beta^*)/h(x; \beta^*)] = 0\) if \(h(x; \beta)\) is \(k\) times differentiable in \(\beta\) in a neighborhood of \(\beta^*\).

**Proof** The stated result follows from differentiating both sides of \(\int h(x; \beta)dx = 1\) \(k\) times with respect to \(\beta\) and evaluating at \(\beta^*\).

In the proof of the following proposition, we make extensive use of Faà di Bruno’s formula on derivatives of the composition of two functions. For a composite function \(f(g(x))\), Faà di Bruno’s formula is

\[
\frac{d^q f(g(x))}{dx^q} = \sum_{(k_1, \ldots, k_q)} \frac{q!}{k_1! \ldots k_q!} \left( \frac{\partial^q f(g(x))}{\partial g^p} \right) \left( \frac{\partial g(x)}{\partial x} \right)^{k_1} \left( \frac{1}{2!} \frac{\partial^2 g(x)}{\partial x^2} \right)^{k_2} \cdots \left( \frac{1}{q!} \frac{\partial^q g(x)}{\partial x^q} \right)^{k_q},
\]

(55)

where \(p = \sum_{i=1}^q k_i\), and the sum \(\sum_{(k_1, \ldots, k_q)}\) is taken over all possible combinations of \((k_1, \ldots, k_q)\) such that \(q = \sum_{i=1}^q ik_i\).

**Proposition C.** Suppose \(g(y|x, z; \psi_\alpha, \alpha)\) is given by (8), where \(\psi = (\eta^T, \lambda_\mu, \lambda_\beta, \lambda_\sigma)^T\) and \(\eta = (\gamma^T, \nu_\mu, \nu_\beta, \nu_\sigma)^T\). Let \(g^*, \ln g^*, \nabla g^*, \text{ and } \nabla \ln g^*\) denote \(g(y|x, z; \psi_\alpha, \alpha)\), \(\ln g(y|x, z; \psi_\alpha, \alpha)\), \(g^*(y|x, z; \psi_\alpha, \alpha)\), and \(\nabla \ln g^*(y|x, z; \psi_\alpha, \alpha)\).
and their derivatives evaluated at \((\psi^*_\alpha, \alpha)\). Then, for \(\lambda_i, \lambda_j, \lambda_k, \lambda_\ell \in \{\lambda_\sigma, \lambda_1, \ldots, \lambda_q\}\),

(a) for \(k = 1, 2, 3\) and \(\ell = 0, 1, \ldots\), \(\nabla_{\lambda_k^\mu} \ln g^* = \nabla_{\lambda_k^\mu} g^*/g^*\)

(b) for \(k = 4, 5, 6, 7\), \(\nabla_{\lambda_k^\mu} \ln g^* = \nabla_{\lambda_k^\mu} g^*/g^*\)

(c) \(\nabla_{\lambda_k^\mu} \ln g^* = \alpha(1 - \alpha)b(\alpha) \frac{\nabla^4 f(\gamma^*, \mu^*, \beta^*, \sigma^{2*})}{f(\gamma^*, \mu^*, \beta^*, \sigma^{2*})}\) with \(b(\alpha) := -(2/3)(\alpha^2 - \alpha + 1)\);

(d) \(\nabla_{\lambda_k^\mu} \ln g^* = \frac{8!}{2} \left(\frac{\nabla_{\lambda_k^\mu} g^*}{4!g^*}\right)^2\) and \(\nabla_{\lambda_k^\mu} \ln g^* = \frac{\nabla_{\lambda_k^\mu} g^*}{g^*} - \frac{\nabla_{\lambda_k^\mu} g^* \nabla \eta g^*}{g^*}\);

(e) for \(\ell = 0, 1, \ldots\), \(\nabla_{\lambda_\ell^\mu} \ln g^* = 0\) and \(\nabla_{\lambda_\ell^\mu} \ln g^* = 0\);

(f) \(\nabla_{\lambda_i \lambda_j} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j} g^*}{g^*}, \ \nabla_{\lambda_i \lambda_j \lambda_k} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j \lambda_k} g^*}{g^*}, \ \nabla_{\lambda_i \lambda_j} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j} g^*}{g^*}\);

(g) for \(k = 1, \ldots, 4\), \(\nabla_{\lambda_k^\mu} \ln g^* = \frac{\nabla_{\lambda_k^\mu} g^*}{g^*}\);

(h) \(\nabla_{\lambda_i \lambda_j \ell} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j \eta} g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} g^* \nabla \eta g^*}{g^*}\), \(\nabla_{\lambda_i \lambda_j \ell} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j} \eta g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} \eta g^* \nabla \eta g^*}{g^*}\);

(i) \(\nabla_{\lambda_i \lambda_j \lambda_k} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j \lambda_k} g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} g^* \nabla \lambda_k g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} \eta g^* \nabla \lambda_k g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} \lambda_k g^* \nabla \eta \lambda_k g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} \lambda_k g^* \nabla \lambda_k \eta g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j} \lambda_k g^* \nabla \lambda_k \lambda_k g^*}{g^*};\)

(j) \(\nabla_{\lambda_i \lambda_j \lambda_k \ell} \ln g^* = \frac{\nabla_{\lambda_i \lambda_j \lambda_k \eta} g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j \lambda_k} g^* \nabla \eta g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j \lambda_k} \eta g^* \nabla \eta g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j \lambda_k} \eta g^* \nabla \lambda_k \eta g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j \lambda_k} \lambda_k g^* \nabla \eta \lambda_k g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j \lambda_k} \lambda_k g^* \nabla \lambda_k \eta g^*}{g^*} - \frac{\nabla_{\lambda_i \lambda_j \lambda_k} \lambda_k g^* \nabla \lambda_k \lambda_k g^*}{g^*};\)

(k) \(\nabla_{\lambda^2_{i} \lambda_{j}} \ln g^* = \frac{\nabla_{\lambda^2_{i} \lambda_{j}} g^*}{g^*} - \frac{2 \nabla_{\lambda_i \lambda_j} g^* \nabla \lambda_i \lambda_j g^*}{g^*} \), \(\nabla_{\lambda^2_{i} \lambda_{j}} \ln g^* = \frac{\nabla_{\lambda^2_{i} \lambda_{j}} g^*}{g^*} - 5! \frac{\nabla_{\lambda^2_{i} \lambda_{j}} g^* \nabla \lambda_i \lambda_j g^*}{g^*}.

**Proof.** We prove part (a) for \(\ell = 0\) first. Suppress all the arguments in \(g(y|x, z; \psi^*_\alpha, \alpha)\) and \(f(y|x, z; \gamma, \theta, \sigma^2)\) but \(\lambda_\mu\), and write

\[
g(\lambda_\mu) = \alpha f((1 - \alpha)\lambda_\mu, (1 - \alpha)C_1\lambda_\mu + (1 - \alpha)f(-\alpha\lambda_\mu, -\alpha C_2\lambda_\mu). \tag{56}\]

Note that, for a composite function \(f(\lambda_\mu, h(\lambda_\mu))\), the following result holds:

\[
\nabla_{\lambda_\mu} f(\lambda_\mu, h(\lambda_\mu)) = (\nabla_{\lambda_\mu} + \nabla_u)^k f(\lambda_\mu, h(u))|_{u=\lambda_\mu} = \sum_{j=0}^{k} \binom{k}{j} \nabla_{\lambda_\mu}^{k-j} \nabla_u^j f(\lambda_\mu, h(u))|_{u=\lambda_\mu}. \tag{57}\]

Because \(\nabla_u u^2|_{u=0} = 0\) except for \(j = 2\), it follows from Faà di Bruno’s formula (55) that

\[
\nabla_u f((1 - \alpha)\lambda_\mu, (1 - \alpha)C_1 u^2)|_{\lambda_\mu=u=0} = \begin{cases} 0 & \text{if } j = 1, 3, \\
2(1 - \alpha)C_1 \nabla_h f(0, h(0)) & \text{if } j = 2, \\
12(1 - \alpha)^2 C_1^2 \nabla_h^2 f(0, h(0)) & \text{if } j = 4, \end{cases} \tag{58}\]
and a similar result holds for $\nabla_{\lambda'^{-1}u}f((1-\alpha)\lambda, (1-\alpha)C_1u^2)$ and $\nabla_{\lambda'^{-1}u}f(-\alpha\lambda, -\alpha C_2u^2)$.

Differentiating (56) and using (57) ($h(\lambda)$ corresponds to $(1-\alpha)C_1\lambda^2$ and $-\alpha C_2\lambda^2$), (58), $C_1 - C_2 = -1$, $\nabla_{\mu^2}f(0,0) = 2\nabla_{\sigma^2}f(0,0)$, $\nabla_{\mu^3}f(0,0) = 2\nabla_{\mu\sigma^2}f(0,0)$, and $3((1-\alpha)C_1 + \alpha C_2) = 2\alpha - 1$, we obtain

$$\nabla_{\lambda}(g(0), 0),$$

$$\nabla_{\lambda^2}(g(0)) = \alpha(1-\alpha)\nabla_{\mu^2}f(0,0) + 2\alpha(1-\alpha)(C_1 - C_2)\nabla_{\sigma^2}f(0,0) = 0,$$

$$\nabla_{\lambda^3}(g(0)) = \alpha(1-\alpha)(1 - 2\alpha)\nabla_{\mu^3}f(0,0) + 3\alpha(1-\alpha)((1-\alpha)C_1 + \alpha C_2)2\nabla_{\mu\sigma^2}f(0,0) = 0,$$

and the first result of part (a) for $\ell = 0$ follows. Repeating the same argument with $\nabla_{\eta^p}(g(\lambda, \eta))$ gives the first result of part (a) for $\ell \geq 1$.

For the second result of part (a) and part (b), suppressing the other arguments but $\lambda$ and $\eta$ from $g(y|x, z; \psi, \alpha)$ and applying Faà di Bruno’s formula (55) to $\partial^q g(\lambda, \eta)/\partial \lambda^q$, we obtain

\[
\frac{\partial^q \ln g(\lambda, \eta)}{\partial \lambda^q} = \sum_{(k_1, \ldots, k_q)} \frac{q!}{k_1! \cdots k_q!} \frac{(-1)^{p-1}(p-1)!}{g(\lambda, \eta)^p} \times \left( \frac{\partial g(\lambda, \eta)}{\partial \lambda^1} \right)^{k_1} \left( \frac{\partial^2 g(\lambda, \eta)}{\partial \lambda^2} \right)^{k_2} \cdots \left( \frac{\partial^q g(\lambda, \eta)}{\partial \lambda^q} \right)^{k_q}
\]  

(59)

where $p = \sum_{i=1}^q k_i$, and the sum $\sum_{(k_1, \ldots, k_q)}$ is taken over all possible combinations of $(k_1, \ldots, k_q)$ such that $q = \sum_{i=1}^q ik_i$. For example, setting $q = 2$ gives $\nabla_{\lambda^2} \ln g(\lambda, \eta) = \nabla_{\lambda^2} g(\lambda, \eta)/g(\lambda, \eta) - (\nabla_{\lambda}, g(\lambda, \eta)/g(\lambda, \eta))^2$, where the sum is taken over $(k_1, k_2) = (2, 0)$ and $(0, 1)$. The second result of part (a) follows from evaluating (59) at $q = 1, 2, 3$, differentiating it $\ell$ times with respect to $\eta$, and using the first result of part (a). Part (b) follows from evaluating (59) at $q = 4, 5, 6, 7$ and applying part (a).

For part (c), differentiating (56) and using (57), (58), and $\nabla_{\mu^4}f(0,0) = 2\nabla_{\mu^2\sigma^2}f(0,0) = 4\nabla_{\sigma^2\sigma^2}f(0,0)$ gives

$$\nabla_{\lambda}(g(0)) = \alpha(1-\alpha)[(1-\alpha)^3 + \alpha^3]\nabla_{\mu^3}f(0,0) + 6\alpha(1-\alpha)((1-\alpha)^2C_1 - \alpha^2C_2)$$

$$\times 2\nabla_{\mu^2\sigma^2}f(0,0) + 12\alpha(1-\alpha)((1-\alpha)C_1 + \alpha C_2)\nabla_{\sigma^2\sigma^2}f(0,0)$$

$$= \alpha(1-\alpha)b(\alpha)\nabla_{\mu^3}f(0,0),$$

with $b(\alpha) := -(2/3)(\alpha^2 - \alpha + 1) < 0$, and the stated result follows from applying Faà di Bruno’s formula in conjunction with part (a).

The first result of part (d) follows from evaluating (59) at $q = 8$ and using part (a). The
second result of part (d) follows from differentiating (59) at \( q = 4 \) with respect to \( \eta \) and using part (a). A direct calculation gives part (e). Part (f) follows from (59) with \( q = 2, 3 \) and part (e). Part (g) follows from differentiating (59) at \( q = 1, \ldots, 4 \) with respect to \( \lambda_i \) and applying parts (a) and (e). A direct calculation in conjunction with parts (a) and (e) gives parts (h)–(j). Part (k) follows from differentiating (59) at \( q = 2 \) and \( q = 5 \) with respect to \( \lambda_i \lambda_j \) and \( \lambda_i \), respectively, and using parts (a) and (e).

\[ \square \]

**Proposition D.** Suppose that \( \{Y_i, X_i, Z_i\}, i = 1, \ldots, n \), are \( n \) independent observations from the density \( f_m(y|x, z; \theta^*_{ma}) \), Assumption 1 holds, and \( c \in (0, 1] \) is chosen so that \( \min_{i,j}(\sigma_i^*/\sigma_j^*) \geq c \). For any \( \theta^*_m \) satisfying \( \min_{i,j}(\sigma_i/\sigma_j) \geq c \) and \( \sum_{i=1}^n f_m(Y_i|X_i, Z_i; \theta^*_m) \geq \sum_{i=1}^n f_m(Y_i|X_i, Z_i; \theta^*_m) + o_p(n) \) for all \( n \), we have \( \inf_{\sigma_m \in \Upsilon_m} ||\theta^*_m - \theta^*_m|| \to 0 \), where \( \Upsilon_m^* := \{\theta_m : f_m(y|x, z; \theta_m) = f_m(y|x, z; \theta^*_ma) \text{ with probability one}\} \).

**Proof.** Our proof closely follows the proof of Theorem 3.3 in Hathaway (1985). Because our model has additional free parameters \( \beta_j \)'s and \( \gamma \), we modify the proof of Hathaway (1985) to consider the joint density of \( m_{ap} := m(p+q+1) + 1 \) observations instead of \( m+1 \) observations in Hathaway (1985, p. 798). The joint density function of \( m_{ap} \) observations is itself a mixture of \( m_{ap} \) components, where each component is given by \( \prod_{j=1}^{m_{ap}} P(y_j; \mu_{ij} + x_j^\top \beta_{ij} + z_j^\top \gamma, \sigma_{ij}) \) for some choices \( i_j \in \{1, \ldots, m\}, j = 1, \ldots, m \), with the density of \( N(\mu, \sigma^2) \) denoted by \( P(y; \mu, \sigma) := (2\pi \sigma^2)^{-1/2} \exp(- (y - \mu)^2 / 2\sigma^2) \).

Assumptions 1, 2, and 3 of Kiefer and Wolfowitz (1956) are easily verified for the joint density of \( m_{ap} \) observations. We verify Assumption 5 of Kiefer and Wolfowitz (1956) for the joint density function of \( m_{ap} \) observations by showing that

\[
E \left[ \prod_{j=1}^{m_{ap}} P(y_j; \mu_{ij}^* + x_j^\top \beta_{ij}^* + z_j^\top \gamma^*, \sigma_{ij}^*) \right] > -\infty
\]

(60)

for \( \theta^*_m \in \Upsilon_m^* \) and

\[
E \sup_{\sigma_m \in \Theta_{\sigma_m(c)}} \ln \left[ \prod_{j=1}^{m_{ap}} P(y_j; \mu_{ij} + x_j^\top \beta_{ij} + z_j^\top \gamma, \sigma_{ij}) \right] < \infty,
\]

(61)

which correspond to equations (3.1) and (3.2) in Hathaway (1985), respectively. Equation (60) follows from the argument in the proof of Theorem 3.3 of Hathaway (1985). For equation (61), proceeding as in Hathaway (1985, pp. 798–799), we can show that

\[
\sup_{\sigma_m \in \Theta_{\sigma_m(c)}} \ln \left[ \prod_{j=1}^{m_{ap}} P(y_j; \mu_{ij} + x_j^\top \beta_{ij} + z_j^\top \gamma, \sigma_{ij}) \right]
\]

is no greater than, for some \( \ell \in \{1, \ldots, m\} \)
and \( j_1, \ldots, j_{p+q+2} \in \{1, \ldots, m_q \} \),

\[
\sup_{\mu_k, \beta_k \gamma_k} \ln \left( \beta(\sigma_k) \prod_{k=1}^{\beta+q+2} P(y_{jk}; \mu_k + x_{jk}^\top \beta_k + z_{jk}^\top \gamma_k, \sigma_k) \right),
\]

(62)

where \( \beta(\sigma_k) := (2\pi)^{(p+q+2-m_q)/2}(\sigma_k)^{p+q+2-m_q} \).

Note that \( \prod_{k=1}^{p+q+2} P(y_{jk}; \mu_k + x_{jk}^\top \beta_k + z_{jk}^\top \gamma_k, \sigma_k) \) is the likelihood function of a linear Gaussian model. Therefore, the maximized value of (62) equals \( C - (m_q/2) \ln S \), where \( C \) is a finite constant that depends only on \( m, p, q, \) and \( S \) is the sum of squared residuals from regressing \( \{y_{jk}\}_{k=1}^{p+q+2} \) on \( \{1, x_{jk}, z_{jk}\}_{k=1}^{p+q+2} \). Because we have one more observation than the number of parameters, the SSR is distributed as \( \chi^2(1) \). Since \( E \ln(\chi^2(1)) < \infty \), the expected value of (62) is finite, and (61) holds. This verifies Assumption 5 of Kiefer and Wolfowitz (1956), and the stated consistency result under Assumption 1 follows.

Given the parameter \( \theta_m \), write the distribution of \((\theta, \sigma^2)\) associated with \( \theta_m \) as \( G(\theta, \sigma^2; \theta_m) := \sum_{j=1}^{m} \alpha_j I\{(\theta_j, \sigma_j^2) \leq (\theta, \sigma^2)\} \), and let \( G^*(\theta, \sigma^2) := G(\theta, \sigma^2; \theta_m^*) \) denote the true mixing distribution. Let \( \gamma(s) \) denote the \( s \)th element of \( \gamma \). Define the penalized log-likelihood function with the penalty term \( \sum_{j=1}^{m} \tilde{p}_n(\sigma_j^2) \) as \( \text{PL}_n(\theta_m) := \sum_{i=1}^{n} \ln \sum_{j=1}^{m} \alpha_j f(Y_i|X_i; \gamma, \theta_j, \sigma_j^2) + \sum_{j=1}^{m} \tilde{p}_n(\sigma_j^2) \). The following proposition shows the consistency of the penalized maximum likelihood estimator. It extends Theorem 5 of Chen et al. (2008) to accommodate a regressor.

**Proposition E.** Suppose that Assumptions 1 and 5 hold. For any \( \theta_m^* \) satisfying \( \text{PL}_n(\theta_m^*) \geq \text{PL}_n(\theta_m^*) + o_p(1) \) for all \( n \), we have \( \sum_{i=1}^{n} |\arctan \gamma^{n}_i - \arctan \gamma^{*}_i| + \int_{\mathbb{R}^{q+1} \times \mathbb{R}^+} |G(\theta, \sigma; \theta_m^*) - G^*(\theta, \sigma)| e^{-||\theta|| - \sigma d\theta d\sigma} \rightarrow_p 0 \).

**Proof** Under Assumption 1(a), the stated result is an immediate consequence of Theorem 5 of Chen et al. (2008), henceforth CTZ.

We show that their results hold under Assumption 1(b). CTZ prove the consistency of the penalized maximum likelihood estimator by showing that the penalty term \( \sum_{j=1}^{m} \tilde{p}_n(\sigma_j^2) \) in effect places a positive constant lower bound on \( \sigma_j^2 \). A key result for establishing the existence of such a lower bound is Lemmas 1 and 2 and equations (2.2) and (2.3) in CTZ that set an upper bound on the number of observations falling in a small neighborhood of a given value of the location parameter (denoted by \( \theta \) in CTZ). In a model with a covariate, Lemmas 1 and 2 of CTZ hold when we replace their \( \theta, X_i \), and sup\( \theta \) with our \( \mu + x^\top \beta + z^\top \gamma, Y_i \), and sup\( \mu, x, \beta, z, \gamma \). Hence, equations (2.2) and (2.3) in CTZ hold when their sup\( \theta \sum_{i=1}^{n} I(|X_i - \theta| < \sigma \ln \sigma) \) is replaced with sup\( \mu, x, \beta, z, \gamma \sum_{i=1}^{n} I(|Y_i - \mu - x^\top \beta - z^\top \gamma| < \sigma \ln \sigma) \), and we can follow the proof of Theorems 4 of CTZ to set a lower bound on \( \sigma_j^2 \). Once a lower bound on
\( \sigma^2 \) is set, the consistency is proven by resorting to Kiefer and Wolfowitz (1956) as CTZ do. The presence of a structural parameter \( \gamma \) has no effect because Kiefer and Wolfowitz (1956) accommodate a structural parameter. \( \Box \)

**Proposition F.** Suppose that Assumptions 1, 4, and 5 hold. If \( \varphi^{h(k)}(\tau_0) - \varphi^{*h} = o_p(1) \) and \( \tau^{(k)}(\tau_0) - \tau_0 = o_p(1) \), then \( \tau^{(k+1)}(\tau_0) - \tau_0 = o_p(1) \).

**Proof.** We suppress \( (\tau_0) \) from \( \varphi^{h(k)}(\tau_0) \) and \( \tau^{(k)}(\tau_0) \). The proof is similar to the proof of Lemma 3 of Li and Chen (2010). Let \( f_i(\gamma, \theta, \sigma^2) \) and \( g^h_i(\varphi^h, \tau) \) denote \( f(Y_i|X_i, Z_i; \gamma, \theta, \sigma^2) \) and \( g^h(Y_i|X_i, Z_i; \varphi^h, \tau) \), respectively. Applying a Taylor expansion to \( (1/n) \sum_{i=1}^n w_{ih}^{(k)} \) and using \( \varphi^{h(k)} - \varphi^{*h} = o_p(1) \), \( \tau^{(k)}(\tau_0) - \tau_0 = o_p(1) \), we obtain

\[
\frac{1}{n} \sum_{i=1}^n w_{ih}^{(k)} = \frac{1}{n} \sum_{i=1}^n \frac{\tau^{(k)} \delta_h^{(k)} f_i(\gamma^{(k)}, \theta_h^{(k)}, \sigma_h^{2(k)})}{g^h_i(\varphi^{h(k)}, \tau^{(k)})} = \frac{1}{n} \sum_{i=1}^n \frac{\tau_0 \alpha_h^{*} f_i(\gamma^{*}, \theta_h^{*}, \sigma_h^{2*})}{g^h_i(\varphi^{*h}, \tau_0)} + o_p(1) = \tau_0 \alpha_h^{*} + o_p(1),
\]

where the last equality follows from \( E[f_i(\gamma^{*}, \theta_h^{*}, \sigma_h^{2*})/g^h_i(\varphi^{*h}, \tau_0)] = 1 \) and the law of large numbers. A similar argument gives \( (1/n) \sum_{i=1}^n w_{ih+1}^{(k)} = (1 - \tau_0) \alpha_h^{*} + o_p(1) \), and the stated result follows from (23). \( \Box \)
C Computer experiments to obtain the empirical formula in (27)

The empirical formula in (27) is obtained through computer experiments that are similar to those of Chen and Li (2009) and Chen et al. (2012). We set \( K = 2 \). For \( m_0 = 2 \), we computed the simulated Type I errors at the 5% nominal level with 1000 repetitions across different parameter settings. We employed three levels for the sample size \( n \): 100, 300, 500; five levels for \( a \): 0.4, 0.6, 0.8, 1.0, 1.2; two levels for the mixing proportions: \((\alpha_1, \alpha_2) = (0.25, 0.75), (0.5, 0.5)\); three levels for the component means: \((\mu_1, \mu_2) = (-1.5, 1.5), (-2, 2), (-2.5, 2.5)\); and two levels for the component variances: \((\sigma_1, \sigma_2) = (1, 1), (1.5, 0.75)\). There are \( 3 \times 5 \times 2 \times 3 \times 2 = 180 \) experiments. Let \( y = \ln(\hat{p}/(0.1 - \hat{p})) \), where \( \hat{p} \)'s are simulated Type I errors. Then, we regress \( y \) on constant, \( \ln(a/(2 - a)) \), \( \ln(\omega_{12}/(1 - \omega_{12})) \), and \( 1/n \). The fitted model based on 180 observations is \( \hat{y} = -0.045 - 0.564 \ln(a/(2 - a)) - 0.035 \ln(\omega_{12}/(1 - \omega_{12})) - 116.15/n \) with \( R^2 = 0.80 \). Setting \( \hat{y} = 0 \) yields the first formula in (27). For \( m_0 = 3 \), we employ three levels for the sample size \( n \) and five levels for \( a \), as in \( m_0 = 2 \); one level for the mixing proportions: \((\alpha_1, \alpha_2, \alpha_3) = (0.33, 0.33, 0.34)\); six levels for the component means: \((\mu_1, \mu_2, \mu_3) = (-4, 0, 4), (-4, 0, 5), (-5, 0, 5), (-4, 0, 6), (-5, 0, 6), (-6, 0, 6)\); and two levels for the component variances: \((\sigma_1, \sigma_2, \sigma_3) = (1, 1, 1), (0.75, 1.5, 0.75)\). Using these 180 experiments and a similar calculation to \( m_0 = 2 \), we obtain the second formula in (27).
## Additional results from empirical examples

### Table 1: Estimation results for 30 stocks in the Dow Jones Industrial Average

<table>
<thead>
<tr>
<th>Security</th>
<th>( p )-value of modified EM test (%)</th>
<th>Selection by</th>
<th>Modified EM</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.Allied Chemical Corp</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2.Aluminum Co America</td>
<td>0.0 0.0 47.6</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3.American Brands Inc</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4.American Can Co</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5.American Tel and Teleg</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>6.Bethlehem Steel Corp</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>7.Du Pont</td>
<td>0.0 0.1 22.8</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8.Eastman Kodak Co</td>
<td>0.0 0.0 84.1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9.Exxon Corp</td>
<td>0.0 0.0 73.0</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10.General Electric Co</td>
<td>0.0 0.0 100.0</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11.General Foods Corp</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>12.General Motors Corp</td>
<td>0.0 0.0 80.9</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>13.Goodyear</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>14.Inco Ltd</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>15.Inter. Business Match.</td>
<td>0.0 0.0 99.1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>16.Inter. Harvester Co.</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>17.Inter. Paper Co.</td>
<td>0.0 0.0 8.4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>18.Johns Manville Corp</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>19.Merck and Co.Inc</td>
<td>0.0 0.0 19.2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>20.Minnesota Mng &amp; Mfg</td>
<td>0.0 0.0 28.1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>21.Owens Illinois Inc</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>22.Proctor &amp; Gamble Co</td>
<td>0.0 0.0 25.4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>23.Sears Roebuck &amp; Co</td>
<td>0.0 0.0 99.1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>24.Standard Oil Co Cal</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>25.Texaco Inc</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>26.Union Carbide Corp</td>
<td>0.0 0.0 94.3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>27.United Aircraft Prod</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>28.United Sts Stl Corp</td>
<td>0.0 0.0 41.9</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>29.Westinghouse Elec Co</td>
<td>0.0 0.0 0.9</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>30.Woolworth F W Co</td>
<td>0.0 0.0 0.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
References


