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

October 30, 2014

This supplementary appendix contains the following details, which have been 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 conducted to obtain the empirical formula in (23), 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 α from $\boldsymbol{\psi}_\alpha$. For a vector $\boldsymbol{x}$ and a function $f(\boldsymbol{x})$, let $\nabla_{\boldsymbol{x}^k} f(\boldsymbol{x})$ denote its k-th derivative with respect to $\boldsymbol{x}$, which can be a multidimensional array. Observe that, for any finite k and for a neighborhood $\mathcal{N}$ of $\boldsymbol{\psi}^*$, we obtain

$$\begin{aligned} &E||\nabla_{\boldsymbol{\psi}^k} g(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\psi}^*, \alpha)/g(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\psi}^*, \alpha)||^2 < \infty, \\ &E|| \sup_{\boldsymbol{\psi} \in \Theta_{\boldsymbol{\psi}} \cap \mathcal{N}} \nabla_{\boldsymbol{\psi}^k} \ln g(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\psi}, \alpha)||^2 < \infty, \end{aligned} \tag{28}$$

*Address for correspondence: Hiroyuki Kasahara, Vancouver School of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC V6T 1Z1, Canada.

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

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

$$
\begin{aligned}
&\text{for } k = 1, 2, 3 \text{ and } \ell = 0, 1, \ldots, \ \nabla_{\lambda_\mu^k \boldsymbol{\eta}^\ell} L_n(\boldsymbol{\psi}^*, \alpha) = 0; \\
&\text{for } k = 4, 5, 6, 7, \ \nabla_{\lambda_\mu^k} L_n(\boldsymbol{\psi}^*, \alpha) = O_p(n^{1/2}); \\
&\text{for } \ell = 0, 1, \ldots, \ \nabla_{\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} \boldsymbol{\eta}^\ell} L_n(\boldsymbol{\psi}^*, \alpha) = 0; \\
&\text{for } k = 2, 3, \ \nabla_{\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^k} L_n(\boldsymbol{\psi}^*, \alpha) = O_p(n^{1/2}); \\
&\nabla_{\lambda_\mu \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^2} L_n(\boldsymbol{\psi}^*, \alpha) = O_p(n^{1/2}); \\
&\text{for } k = 1, \ldots, 4, \ \nabla_{\lambda_\mu^k \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}} L_n(\boldsymbol{\psi}^*, \alpha) = O_p(n^{1/2}).
\end{aligned} \tag{29}
$$

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

$$
L_n(\boldsymbol{\psi}, \alpha) - L_n(\boldsymbol{\psi}^*, \alpha) =
$$

$$
\nabla_{\boldsymbol{\eta}} L_n^* (\boldsymbol{\eta} - \boldsymbol{\eta}^*) + \frac{1}{2!} (\boldsymbol{\eta} - \boldsymbol{\eta}^*)^\top \nabla_{\boldsymbol{\eta}\boldsymbol{\eta}^\top} L_n^* (\boldsymbol{\eta} - \boldsymbol{\eta}^*) \tag{30}
$$

$$
+ \frac{1}{2!} \left\{ 2 \nabla_{\lambda_\mu \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^\top} L_n^* \lambda_\mu \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} + \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^\top \nabla_{\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^\top} L_n^* \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} \right\} \tag{31}
$$

$$
+ \frac{1}{3!} \left\{ 3 \sum_{i=0}^q \sum_{j=0}^q \nabla_{\lambda_i \lambda_j \boldsymbol{\eta}^\top} L_n^* \lambda_i \lambda_j (\boldsymbol{\eta} - \boldsymbol{\eta}^*) + 6 \sum_{i=0}^q \nabla_{\lambda_\mu \lambda_i \boldsymbol{\eta}^\top} L_n^* \lambda_\mu \lambda_i (\boldsymbol{\eta} - \boldsymbol{\eta}^*) \right\} \tag{32}
$$

$$
+ \frac{1}{4!} \left\{ \nabla_{\lambda_\mu^4} L_n^* \lambda_\mu^4 + \sum_{i=0}^q \sum_{j=0}^q \sum_{k=0}^q \sum_{\ell=0}^q \nabla_{\lambda_i \lambda_j \lambda_k \lambda_\ell} L_n^* \lambda_i \lambda_j \lambda_k \lambda_\ell \right. \tag{33}
$$

$$
\left. + 4 \sum_{i=1}^q \sum_{j=1}^q \sum_{k=1}^q \nabla_{\lambda_i \lambda_j \lambda_k \lambda_\mu} L_n^* \lambda_i \lambda_j \lambda_k \lambda_\mu + 6 \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^\top \nabla_{\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}^\top \lambda_\mu^2} L_n^* \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} \lambda_\mu^2 \right\} \tag{34}
$$

$$
+ \frac{5}{5!} \nabla_{\lambda_\mu^4 \boldsymbol{\eta}^\top} L_n^* \lambda_\mu^4 (\boldsymbol{\eta} - \boldsymbol{\eta}^*) + \frac{6}{6!} \nabla_{\lambda_\mu^5 \boldsymbol{\lambda}_{\sigma\beta}} L_n^* \lambda_\mu^5 \boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}} + \frac{1}{8!} \nabla_{\lambda_\mu^8} L_n^* \lambda_\mu^8 + R_{1n}(\boldsymbol{\psi}, \alpha), \tag{35}
$$

where ∇L_n^* denotes the derivative of $L_n(\boldsymbol{\psi}, \alpha)$ evaluated at $(\boldsymbol{\psi}^*, \alpha)$, and

$$
\begin{aligned}
&R_{1n}(\boldsymbol{\psi}, \alpha) = \\
&O_p(n^{1/2})(||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^3 + \lambda_\mu^2||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}|| + \lambda_\mu||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^2) + O_p(n)||\dot{\boldsymbol{\eta}}||^3 + O_p(n)(\lambda_\mu + ||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||)||\dot{\boldsymbol{\eta}}||^2 \quad (36)\\
&+ O_p(n^{1/2})\lambda_\mu^3||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}|| + \sum_{(p,q,r)\in D(4)} O_p(n)\lambda_\mu^p||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^q||\dot{\boldsymbol{\eta}}||^r \quad (37)\\
&+ O_p(n^{1/2})\lambda_\mu^5 + O_p(n^{1/2})\lambda_\mu^4||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}|| + \sum_{(p,q,r)\in D(5)} O_p(n)\lambda_\mu^p||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^q||\dot{\boldsymbol{\eta}}||^r \quad (38)\\
&+ O_p(n^{1/2})\lambda_\mu^6 + \sum_{(p,q,r)\in D(6)} O_p(n)\lambda_\mu^p||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^q||\dot{\boldsymbol{\eta}}||^r \quad (39)\\
&+ O_p(n^{1/2})\lambda_\mu^7 + \sum_{(p,q,r)\in D(7)} O_p(n)\lambda_\mu^p||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^q||\dot{\boldsymbol{\eta}}||^r \quad (40)\\
&+ \sum_{(p,q,r)\in D(8)} O_p(n)\lambda_\mu^p||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^q||\dot{\boldsymbol{\eta}}||^r + \sum_{(p,q,r)\in D(9)} O_p(n)\lambda_\mu^p||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^q||\dot{\boldsymbol{\eta}}||^r, \quad (41)
\end{aligned}
$$

with $\dot{\boldsymbol{\eta}} := \boldsymbol{\eta} - \boldsymbol{\eta}^*$ and sets $D(4)$–$D(9)$ being defined as

$$
\begin{aligned}
D(4) &:= \{(p,q,r) : p+q+r = 4,\ r \neq 0, (p,q,r) \neq (3,0,1)\},\\
D(5) &:= \{(p,q,r) : p+q+r = 5,\ (p,q,r) \neq (5,0,0), (4,1,0), (4,0,1)\},\\
D(6) &:= \{(p,q,r) : p+q+r = 6,\ (p,q,r) \neq (6,0,0), (5,1,0)\},\\
D(7) &:= \{(p,q,r) : p+q+r = 7,\ (p,q,r) \neq (7,0,0)\},\\
D(8) &:= \{(p,q,r) : p+q+r = 8,\ (p,q,r) \neq (8,0,0)\},\\
D(9) &:= \{(p,q,r) : p+q+r = 9\}.
\end{aligned}
$$

We prove part (a) by showing that the terms in (30)–(35) are written as

$$
\boldsymbol{t}_n(\boldsymbol{\psi},\alpha)^\top \boldsymbol{S}_n - (1/2)\boldsymbol{t}_n(\boldsymbol{\psi},\alpha)^\top \boldsymbol{\mathcal{I}}_n \boldsymbol{t}_n(\boldsymbol{\psi},\alpha) + [O(||\boldsymbol{\psi}-\boldsymbol{\psi}^*||) + o(1)]O_p((1+||\boldsymbol{t}_n(\boldsymbol{\psi},\alpha)||)^2), \quad (42)
$$

Henceforth, we suppress $(\boldsymbol{\psi}, \alpha)$ from $\boldsymbol{t}_n(\boldsymbol{\psi}, \alpha)$. The first term in (30), the terms in (31), and the first term in (33) are written as $\boldsymbol{t}_n^\top \boldsymbol{S}_n$ because $\nabla_{\lambda_\mu\lambda_\sigma} L_n^* \lambda_\mu\lambda_\sigma = \alpha(1-\alpha)\sum_{i=1}^n H_i^{3*} 6\lambda_\mu\lambda_\sigma$, $\nabla_{\lambda_\mu\boldsymbol{\lambda}_{\boldsymbol{\beta}}^\top} L_n^* \lambda_\mu \boldsymbol{\lambda}_{\boldsymbol{\beta}} = \alpha(1-\alpha)\sum_{i=1}^n H_i^{2*} X_i^\top 2\lambda_\mu \boldsymbol{\lambda}_{\boldsymbol{\beta}}$, $[(1/2!)\nabla_{\lambda_\sigma^2} L_n^* \lambda_\sigma^2 + (1/4!)\nabla_{\lambda_\mu^4} L_n^* \lambda_\mu^4] = \alpha(1-\alpha)\sum_{i=1}^n H_i^{4*}[12\lambda_\sigma^2 + b(\alpha)\lambda_\mu^4]$, $\nabla_{\lambda_\sigma\boldsymbol{\lambda}_{\boldsymbol{\beta}}^\top} L_n^* \lambda_\sigma \boldsymbol{\lambda}_{\boldsymbol{\beta}} = \alpha(1-\alpha)\sum_{i=1}^n H_i^{3*} X_i^\top 6\lambda_\sigma \boldsymbol{\lambda}_\beta$, and $(1/2!)\boldsymbol{\lambda}_{\boldsymbol{\beta}}^\top \nabla_{\boldsymbol{\lambda}_{\boldsymbol{\beta}}\boldsymbol{\lambda}_{\boldsymbol{\beta}}^\top} L_n^* \boldsymbol{\lambda}_{\boldsymbol{\beta}} = \alpha(1-\alpha)\sum_{j=1}^q \sum_{k=1}^q \lambda_j\lambda_k \sum_{i=1}^n H_i^{2*} X_{ji} X_{ki} = \boldsymbol{v}_{\boldsymbol{\beta}}(\boldsymbol{\lambda}_{\boldsymbol{\beta}})^\top \boldsymbol{s}_{\boldsymbol{\lambda}_{\boldsymbol{\beta}} i}$, in view of Propositions A and C(c)(f)(g). The other terms in (30)–(35) except for $R_{1n}(\boldsymbol{\psi}, \alpha)$ are written as $-(1/2)\boldsymbol{t}_n^\top \boldsymbol{\mathcal{I}}_n \boldsymbol{t}_n + O_p(||\boldsymbol{\psi}-\boldsymbol{\psi}^*||||\boldsymbol{t}_n||^2) + O_p(n^{-1/2}||\boldsymbol{t}_n||^2)$ from a tedious but straightforward calculation in conjunction with Propositions A, B, and C.

We complete the proof of part (a) by showing that $R_{1n}(\boldsymbol{\psi}, \alpha)$ satisfies the order in (42). Note that

$$12\lambda_\sigma^3 = \lambda_\sigma[b(\alpha)\lambda_\mu^4 + 12\lambda_\sigma^2] - (\lambda_\mu^3 b(\alpha))\lambda_\mu\lambda_\sigma = O(n^{-1/2}||\boldsymbol{\lambda}||||\boldsymbol{t}_{\boldsymbol{\lambda}n}||). \tag{43}$$

Therefore, the term with λ_σ^3 in the first term in (36) is $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||)$. The other terms in (36) are either $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||)$ or $O_p((||\boldsymbol{\lambda}|| + ||\dot{\boldsymbol{\eta}}||)||\boldsymbol{t}_n||^2)$.

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

Note that $\lambda_\mu^5 = (\lambda_\mu/b(\alpha))[b(\alpha)\lambda_\mu^4 + 12\lambda_\sigma^2] - (12\lambda_\sigma/b(\alpha))\lambda_\mu\lambda_\sigma = O(n^{-1/2}||\boldsymbol{\lambda}||||\boldsymbol{t}_{\boldsymbol{\lambda}n}||)$. Therefore, the first term in (38) is $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||)$ 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 $\lambda_\mu^3||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^2$, $\lambda_\mu^2||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^3$, $\lambda_\mu||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^4$, and $||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^5$. The term with λ_σ^5 is $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||^2)$ because $12\lambda_\sigma^5 = \lambda_\sigma^3[b(\alpha)\lambda_\mu^4 + 12\lambda_\sigma^2] - (\lambda_\mu^3\lambda_\sigma^2 b(\alpha))\lambda_\mu\lambda_\sigma = O(n^{-1}||\boldsymbol{\lambda}||||\boldsymbol{t}_{\boldsymbol{\lambda}n}||^2)$ while the other terms in $D(5)$ with $r = 0$ are $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||^2)$ because, for example, the terms with $\lambda_\mu^3||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^2$ and $\lambda_\mu^2||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^3$ are $O_p(||\boldsymbol{\lambda}||n||\lambda_\mu\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||^2) = O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||^2)$. The terms in $D(5)$ with $r = 1$ are $O_p(||\boldsymbol{\psi} - \boldsymbol{\psi}^*||||\boldsymbol{t}_n||^2)$ from a simple calculation. Of the terms in $D(6)$ in (39) with $r = 0, 1$, the one with $\lambda_\mu^5||\dot{\boldsymbol{\eta}}||$ is $O_p(||\boldsymbol{\psi} - \boldsymbol{\psi}^*||||\boldsymbol{t}_n||^2)$ because $\lambda_\mu^5 = O(n^{-1/2}||\boldsymbol{\lambda}||||\boldsymbol{t}_{\boldsymbol{\lambda}n}||)$, 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 $\lambda_\mu^6||\boldsymbol{\lambda}_{\sigma\boldsymbol{\beta}}||$ are $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||^2)$ because $\lambda_\mu^5 = O(n^{-1/2}||\boldsymbol{\lambda}||||\boldsymbol{t}_{\boldsymbol{\lambda}n}||)$, 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 λ_μ^9 is $O_p(||\boldsymbol{\lambda}||||\boldsymbol{t}_n||^2)$ because $\lambda_\mu^9 = (\lambda_\mu^5/b(\alpha))[b(\alpha)\lambda_\mu^4 + 12\lambda_\sigma^2] - 12(\lambda_\mu^3/b(\alpha))(\lambda_\mu\lambda_\sigma)^2 = \lambda_\mu^5 O(n^{-1/2}||\boldsymbol{t}_{\boldsymbol{\lambda}n}||) + \lambda_\mu^3 O(n^{-1}||\boldsymbol{t}_{\boldsymbol{\lambda}n}||^2) = O(n^{-1}||\boldsymbol{\lambda}||||\boldsymbol{t}_{\boldsymbol{\lambda}n}||^2)$, 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 $\boldsymbol{\mathcal{I}}$ holds under Assumption 2(b) because the off-diagonal elements of $\boldsymbol{\mathcal{I}} = E[\boldsymbol{s}_i\boldsymbol{s}_i^\top]$ that involve the interaction terms of H_i^{j*} and H_i^{k*} are zero for $j \neq k$ by the property of Hermite polynomials. $\square$

A.3 Proof of Proposition 3

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

The proof of part (a) closely follows the proof of Theorem 1 of Andrews (1999) (A99, hereafter). Let $\boldsymbol{T}_n := \boldsymbol{\mathcal{I}}_n^{1/2}\hat{\boldsymbol{t}}_n$. Then, in view of (13), we have

$$
\begin{aligned}
o_p(1) &\leq L_n(\hat{\boldsymbol{\psi}}, \alpha) - L_n(\boldsymbol{\psi}^*, \alpha) \\
&= \boldsymbol{T}_n' \boldsymbol{\mathcal{I}}_n^{-1/2} \boldsymbol{S}_n - \frac{1}{2}||\boldsymbol{T}_n||^2 + R_n(\hat{\boldsymbol{\psi}}, \alpha) \\
&= O_p(||\boldsymbol{T}_n||) - \frac{1}{2}||\boldsymbol{T}_n||^2 + (1 + ||\boldsymbol{\mathcal{I}}_n^{-1/2}\boldsymbol{T}_n||)^2 o_p(1) \\
&= ||\boldsymbol{T}_n|| O_p(1) - \frac{1}{2}||\boldsymbol{T}_n||^2 + o_p(||\boldsymbol{T}_n||) + o_p(||\boldsymbol{T}_n||^2) + o_p(1),
\end{aligned}
$$

where the third equality holds because $\boldsymbol{\mathcal{I}}_n^{-1/2}\boldsymbol{S}_n = O_p(1)$ and $R_n(\hat{\boldsymbol{\psi}}, \alpha) = o_p((1+||\boldsymbol{\mathcal{I}}_n^{-1/2}\boldsymbol{T}_n||)^2)$ from Propositions 1 and 2. Rearranging this equation yields $||\boldsymbol{T}_n||^2 \leq 2||\boldsymbol{T}_n||O_p(1) + o_p(1)$. Denote the $O_p(1)$ term by ς_n. Then, $(||\boldsymbol{T}_n|| - \varsigma_n)^2 \leq \varsigma_n^2 + o_p(1) = O_p(1)$; taking its square root gives $||\boldsymbol{T}_n|| \leq O_p(1)$. In conjunction with $\boldsymbol{\mathcal{I}}_n \rightarrow_p \boldsymbol{\mathcal{I}}$, we obtain $\hat{\boldsymbol{t}}_n = O_p(1)$, and part (a) follows.

For parts (b) and (c), define

$$
\boldsymbol{W}_n := \boldsymbol{\mathcal{I}}^{-1}\boldsymbol{S}_n = \begin{bmatrix} \boldsymbol{W}_{\boldsymbol{\eta}n} \\ \boldsymbol{W}_{\boldsymbol{\lambda}n} \end{bmatrix}, \quad
\begin{aligned}
\boldsymbol{W}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n} &:= \boldsymbol{W}_{\eta n} - E[\boldsymbol{W}_{\boldsymbol{\eta}n}\boldsymbol{W}_{\boldsymbol{\lambda}n}^\top]\mathrm{Var}[\boldsymbol{W}_{\boldsymbol{\lambda}n}]^{-1}\boldsymbol{W}_{\boldsymbol{\lambda}n}, \\
\boldsymbol{t}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n} &:= \boldsymbol{t}_{\boldsymbol{\eta}n} - E[\boldsymbol{W}_{\boldsymbol{\eta}n}\boldsymbol{W}_{\boldsymbol{\lambda}n}^\top]\mathrm{Var}[\boldsymbol{W}_{\boldsymbol{\lambda}n}]^{-1}\boldsymbol{t}_{\boldsymbol{\lambda}n}.
\end{aligned}
$$

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

$$
\begin{aligned}
2[L_n(\boldsymbol{\psi}, \alpha) - L_n(\boldsymbol{\psi}^*, \alpha)] &= \boldsymbol{W}_n^\top \boldsymbol{\mathcal{I}} \boldsymbol{W}_n - (\boldsymbol{t}_n - \boldsymbol{W}_n)^\top \boldsymbol{\mathcal{I}} (\boldsymbol{t}_n - \boldsymbol{W}_n) + o_p(1) \\
&= A_n(\boldsymbol{t}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n}) + B_n(\boldsymbol{t}_{\boldsymbol{\lambda}n}) + o_p(1),
\end{aligned} \tag{44}
$$

where

$$
\begin{aligned}
A_n(\boldsymbol{t}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n}) &= \boldsymbol{W}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}} \boldsymbol{W}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n} - (\boldsymbol{t}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n} - \boldsymbol{W}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}} (\boldsymbol{t}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n} - \boldsymbol{W}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n}), \\
B_n(\boldsymbol{t}_{\boldsymbol{\lambda}n}) &= \boldsymbol{W}_{\boldsymbol{\lambda}n}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{W}_{\boldsymbol{\lambda}n} - (\boldsymbol{t}_{\boldsymbol{\lambda}n} - \boldsymbol{W}_{\boldsymbol{\lambda}n})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} (\boldsymbol{t}_{\boldsymbol{\lambda}n} - \boldsymbol{W}_{\boldsymbol{\lambda}n}).
\end{aligned}
$$

Note that $\boldsymbol{W}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n} = \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\boldsymbol{\eta}n}$, $\nabla_{\boldsymbol{\eta}} l(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\psi}^*, \alpha)$ equals the score of the one-component model as shown in (8), and the set of admissible values of $\hat{\boldsymbol{t}}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n}$ approaches $\mathbb{R}^{p+q+2}$. Therefore,

$A_n(\hat{\boldsymbol{t}}_{\boldsymbol{\eta}.\boldsymbol{\lambda}n}) = 2[L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0, \hat{\sigma}_0^2) - L_{0,n}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*, \sigma^{2*})] + o_p(1)$, and it follows from (44) that

$$2[L_n(\hat{\boldsymbol{\psi}}, \alpha) - L_{0,n}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}_0, \hat{\sigma}_0^2)] = B_n(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}n}) + o_p(1). \tag{45}$$

When no conditioning variable $\boldsymbol{X}$ is present, $B_n(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}n}) \to_d \chi^2(2)$ because the set of admissible values of $(\hat{t}_{\mu\sigma n}, \hat{t}_{\mu^4 n})^\top$ approaches $\mathbb{R}^2$, and part (b) follows.

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

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

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

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

Define $\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^1 := \arg\max_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \tilde{\Lambda}_{\boldsymbol{\lambda}}^1} B_n(\boldsymbol{t}_{\boldsymbol{\lambda}})$, where $\tilde{\Lambda}_{\boldsymbol{\lambda}}^1 := \boldsymbol{t}_{\boldsymbol{\lambda}n}(\Theta_{\boldsymbol{\psi}}(\epsilon_\alpha), \alpha) \cap \{\boldsymbol{t}_{\boldsymbol{\beta}\sigma n} = 0, \boldsymbol{t}_{\boldsymbol{\beta}^2 n} = 0, t_{\mu^4 n} = n^{1/2}\alpha(1-\alpha)b(\alpha)\lambda_\mu^4\}$. Then, we have $B_n(\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^1) \geq B_n(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^1) + o_p(1)$ from the definition of $\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^1$, definition of $B_n(\boldsymbol{t}_{\boldsymbol{\lambda}n})$, and from (47). Note that $\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^2$ satisfies $\hat{t}_{\mu^4 n}^2 = n^{1/2}\alpha(1-\alpha)12(\hat{\lambda}_\sigma^2)^2 + o_p(1)$ because $|\lambda_\mu| < n^{-1/8}(\ln n)^{-1}$ if $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}}^2$. Define $\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^2 := \arg\max_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \tilde{\Lambda}_{\boldsymbol{\lambda}}^2} B_n(\boldsymbol{t}_{\boldsymbol{\lambda}})$, where $\tilde{\Lambda}_{\boldsymbol{\lambda}}^2 := \boldsymbol{t}_{\boldsymbol{\lambda}n}(\Theta_{\boldsymbol{\psi}}(\epsilon_\alpha), \alpha) \cap \{t_{\mu^4 n} = n^{1/2}\alpha(1-\alpha)12\lambda_\sigma^2\}$. Then, a similar argument as above gives $B_n(\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^2) \geq B_n(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^2) + o_p(1)$.

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

$$(B_n(\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^1), B_n(\tilde{\boldsymbol{t}}_{\boldsymbol{\lambda}n}^2)) \to_d ((\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^1)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\eta}} \hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^1, (\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^2)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\eta}} \hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^2) \tag{48}$$

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

A.4 Proof of Proposition 4

Define a $d \times q_\lambda$ matrix $\tilde{\boldsymbol{B}} := [\mathbf{0}_{d\times 2}\ \boldsymbol{B}_x\ \mathbf{0}_{d\times q}\ \boldsymbol{B}_v]$, where $\boldsymbol{B}_x$ denotes the first q columns of $\boldsymbol{B}$, and $\boldsymbol{B}_v$ denotes the last $q(q+1)/2$ columns of $\boldsymbol{B}$. Then, we have, for any $k = 1, 2, \ldots,$

$$\tilde{\boldsymbol{B}} \boldsymbol{s}_{\boldsymbol{\lambda} i} = 0 \quad \text{and} \quad \nabla_{\boldsymbol{\psi}_\alpha^k} \tilde{\boldsymbol{B}} \boldsymbol{s}_{\boldsymbol{\lambda} i} = 0. \tag{49}$$

Collect the basis of the orthogonal complement of the row space of $\tilde{\boldsymbol{B}}$ into a $(q_\lambda - d) \times q_\lambda$ matrix $\tilde{\boldsymbol{B}}^\perp$; then, $\tilde{\boldsymbol{B}}^\perp$ satisfies $\tilde{\boldsymbol{B}}^\perp \tilde{\boldsymbol{B}}^\top = 0$ and $\tilde{\boldsymbol{B}}^\perp (\tilde{\boldsymbol{B}}^\perp)^\top = \boldsymbol{I}_{q_\lambda - d}$. Define a $(\dim(\boldsymbol{s}_i) - d) \times \dim(\boldsymbol{s}_i)$ matrix $\boldsymbol{Q}$ and $\dim(\boldsymbol{s}_i) \times \dim(\boldsymbol{s}_i)$ matrix $\tilde{\boldsymbol{Q}}$ as

$$\boldsymbol{Q} := \begin{pmatrix} \boldsymbol{I}_{2+q+p} & 0 \\ 0 & \tilde{\boldsymbol{B}}^\perp \end{pmatrix}, \quad \tilde{\boldsymbol{Q}} := \begin{pmatrix} & \boldsymbol{Q} \\ 0 & \tilde{\boldsymbol{B}} \end{pmatrix}, \tag{50}$$

Then $\tilde{\boldsymbol{Q}}$ satisfies

$$\tilde{\boldsymbol{Q}}^{-1} = \begin{pmatrix} \boldsymbol{Q}^\top & \begin{matrix} 0 \\ \tilde{\boldsymbol{B}}^\top (\tilde{\boldsymbol{B}} \tilde{\boldsymbol{B}}^\top)^{-1} \end{matrix} \end{pmatrix}.$$

Rewrite (13) as

$$\begin{aligned} L_n(\boldsymbol{\psi}_\alpha, \alpha) - L_n(\boldsymbol{\psi}_\alpha^*, \alpha) &= (\boldsymbol{t}_n)^\top \tilde{\boldsymbol{Q}}^{-1} \tilde{\boldsymbol{Q}} \boldsymbol{S}_n - \frac{1}{2} (\boldsymbol{t}_n)^\top \tilde{\boldsymbol{Q}}^{-1} \tilde{\boldsymbol{Q}} \boldsymbol{\mathcal{I}}_n \tilde{\boldsymbol{Q}}^\top (\tilde{\boldsymbol{Q}^\top})^{-1} \boldsymbol{t}_n + R_n(\boldsymbol{\psi}_\alpha, \alpha) \\ &= (\boldsymbol{Q} \boldsymbol{t}_n)^\top \boldsymbol{Q} \boldsymbol{S}_n - \frac{1}{2} (\boldsymbol{Q} \boldsymbol{t}_n)^\top (\boldsymbol{Q} \boldsymbol{\mathcal{I}}_n \boldsymbol{Q}^\top) \boldsymbol{Q} \boldsymbol{t}_n + R_n(\boldsymbol{\psi}_\alpha, \alpha), \end{aligned}$$

where the second equality follows from $\tilde{\boldsymbol{Q}} \boldsymbol{S}_n = \begin{pmatrix} \boldsymbol{Q}\boldsymbol{S}_n \\ 0 \end{pmatrix}$, $(\tilde{\boldsymbol{Q}}^{-1})^\top \boldsymbol{t}_n = \begin{pmatrix} \boldsymbol{Q}\boldsymbol{t}_n \\ * \end{pmatrix}$, and $\tilde{\boldsymbol{Q}} \boldsymbol{\mathcal{I}}_n \tilde{\boldsymbol{Q}}^\top = \begin{pmatrix} \boldsymbol{Q}\boldsymbol{\mathcal{I}}_n\boldsymbol{Q}^\top & 0 \\ 0 & 0 \end{pmatrix}$. Further, in view of (49), $R_n(\boldsymbol{\psi}_\alpha, \alpha)$ satisfies $\limsup_{n\to\infty} \Pr(\sup_{\boldsymbol{\psi}_\alpha \in \Theta_{\boldsymbol{\psi}_\alpha} : ||\boldsymbol{\psi}_\alpha - \boldsymbol{\psi}_\alpha^*|| \le \kappa} |R_n(\boldsymbol{\psi}_\alpha, \alpha)| > \delta(1 + ||\boldsymbol{Q}\boldsymbol{t}_n||)^2) \to 0$ as $\kappa \to 0$.

Observe that

$$\boldsymbol{Q}\boldsymbol{S}_n = n^{-1/2} \sum_{i=1}^n \begin{pmatrix} \boldsymbol{s}_{\boldsymbol{\eta} i} \\ \tilde{\boldsymbol{B}}^\perp \boldsymbol{s}_{\boldsymbol{\lambda} i} \end{pmatrix}, \quad \boldsymbol{Q}\boldsymbol{t}_n = \begin{pmatrix} \boldsymbol{t}_{\boldsymbol{\eta} n} \\ \tilde{\boldsymbol{B}}^\perp \boldsymbol{t}_{\boldsymbol{\lambda} n} \end{pmatrix}.$$

Define $q_\lambda - d$ vectors $\boldsymbol{S}_{\boldsymbol{\lambda}}^B$ and $\boldsymbol{W}_{\boldsymbol{\lambda}}^B$, and a $(q_\lambda - d) \times (q_\lambda - d)$ matrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}}^B$ in the same manner as $\boldsymbol{S}_{\boldsymbol{\lambda}}$, $\boldsymbol{W}_{\boldsymbol{\lambda}}$, and $\boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}}$, respectively, but using $\boldsymbol{Q}\boldsymbol{S}$ and $\boldsymbol{W}^B := (\boldsymbol{Q}\boldsymbol{\mathcal{I}}\boldsymbol{Q}^\top)^{-1}\boldsymbol{Q}\boldsymbol{S}$ in place of $\boldsymbol{S}$ and $\boldsymbol{W}$. For $j = 1, 2$, define $\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^{Bj}$ by $r_{\boldsymbol{\lambda}}^B(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^{Bj}) = \inf_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \Lambda_{\boldsymbol{\lambda}}^j} r_{\boldsymbol{\lambda}}^B(\boldsymbol{t}_{\boldsymbol{\lambda}})$, where $r_{\boldsymbol{\lambda}}^B(\boldsymbol{t}_{\boldsymbol{\lambda}}) := (\tilde{\boldsymbol{B}}^\perp \boldsymbol{t}_{\boldsymbol{\lambda}} - \boldsymbol{W}_{\boldsymbol{\lambda}}^B)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}}^B (\tilde{\boldsymbol{B}}^\perp \boldsymbol{t}_{\boldsymbol{\lambda}} - \boldsymbol{W}_{\boldsymbol{\lambda}}^B)$. Then, the stated result follows from repeating the proof of Proposition 3. □

A.5 Proof of Proposition 5

For $h = 1, \ldots, m_0$, let $\mathcal{N}_h^* \subset \Theta_{\boldsymbol{\vartheta}_{m_0+1}}(\epsilon_\sigma)$ be a sufficiently small closed neighborhood of Υ_{1h}^*, such that $(\boldsymbol{\theta_1}, \sigma_1^2) < \cdots < (\boldsymbol{\theta}_{h-1}, \sigma_{h-1}^2) < (\boldsymbol{\theta}_h, \sigma_h^2), (\boldsymbol{\theta}_{h+1}, \sigma_{h+1}^2) < (\boldsymbol{\theta}_{h+2}, \sigma_{h+2}^2) < \cdots < (\boldsymbol{\theta}_{m_0+1}, \sigma_{m_0+1}^2)$ and $\alpha_h, \alpha_{h+1} > 0$ hold and $\Upsilon_{1k}^* \notin \mathcal{N}_h^*$ if $k \neq h$. For $\boldsymbol{\vartheta}_{m_0+1} \in \mathcal{N}_h^*$, we introduce the following one-to-one reparameterization, which is similar to (5):

$$\begin{aligned}
&\delta_h := \alpha_h + \alpha_{h+1}, \quad \tau := \alpha_h/(\alpha_h + \alpha_{h+1}), \\
&(\delta_1, \ldots, \delta_{h-1}, \delta_{h+1} \ldots, \delta_{m_0-1})^\top := (\alpha_1, \ldots, \alpha_{h-1}, \alpha_{h+2}, \ldots, \alpha_{m_0})^\top, \\
&\begin{pmatrix} \boldsymbol{\theta}_h \\ \boldsymbol{\theta}_{h+1} \\ \sigma_h^2 \\ \sigma_{h+1}^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\nu_\theta} + (1-\tau)\boldsymbol{\lambda}_\theta \\ \boldsymbol{\nu_\theta} - \tau\boldsymbol{\lambda}_\theta \\ \nu_\sigma + (1-\tau)(2\lambda_\sigma + C_1\lambda_\mu^2) \\ \nu_\sigma - \tau(2\lambda_\sigma + C_2\lambda_\mu^2) \end{pmatrix},
\end{aligned} \tag{51}$$

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 $(\boldsymbol{\lambda_\theta}, \boldsymbol{\nu_\theta}, \lambda_\sigma, \nu_\sigma)$ on τ. With this reparameterization, the null restriction $(\boldsymbol{\theta}_h, \sigma_h^2) = (\boldsymbol{\theta}_{h+1}, \sigma_{h+1}^2)$ implied by $H_{0,1h}$ holds if and only if $(\boldsymbol{\lambda_\theta}, \lambda_\sigma) = (0, 0)$. Collect the reparameterized parameters except for τ into one vector $\boldsymbol{\psi}_\tau^h$, and let $\boldsymbol{\psi}_\tau^{h*}$ denote its true value. Define the reparameterized density as

$$\begin{aligned}
g^h(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\psi}_\tau^h, \tau) := \delta_h \big[&\tau f\left(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\gamma}, \boldsymbol{\nu_\theta} + (1-\tau)\boldsymbol{\lambda_\theta}, \nu_\sigma + (1-\tau)(2\lambda_\sigma + C_1\lambda_\mu^2)\right) \\
&+ (1-\tau) f\left(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\gamma}, \boldsymbol{\nu_\theta} - \tau\boldsymbol{\lambda_\theta}, \nu_\sigma - \tau(2\lambda_\sigma + C_2\lambda_\mu^2)\right)\big] \\
&+ \sum_{j=1}^{h-1} \delta_j f(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\gamma}, \boldsymbol{\theta}_j, \sigma_j^2) + \sum_{j=h+1}^{m_0} \delta_j f(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\gamma}, \boldsymbol{\theta}_{j+1}, \sigma_{j+1}^2).
\end{aligned}$$

Define the local MLE of $\boldsymbol{\psi}_\tau^h$ by

$$\hat{\boldsymbol{\psi}}_\tau^h := \arg\max_{\boldsymbol{\psi}_\tau^h \in \mathcal{N}_h^*} L_n^h(\boldsymbol{\psi}_\tau^h, \tau), \tag{52}$$

where $L_n^h(\boldsymbol{\psi}_\tau^h, \tau) := \sum_{i=1}^n \ln[g^h(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\psi}_\tau^h, \tau)]$. Because $\boldsymbol{\psi}_\tau^{h*}$ is the only parameter value in $\mathcal{N}_h^*$ that generates true density, $\hat{\boldsymbol{\psi}}_\tau^h - \boldsymbol{\psi}_\tau^{h*} = o_p(1)$ follows from Proposition D. Define the LRT statistic for testing $H_{0,1h}$ as $LR_{n,1h}(\epsilon_\tau) := \max_{\tau \in [\epsilon_\tau, 1-\epsilon_\tau]} 2\{L_n^h(\hat{\boldsymbol{\psi}}_\tau^h, \tau) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{m_0})\}$ for some $\epsilon_\tau \in (0, 1/2)$.

In view of Proposition D, the stated result holds if

$$(LR_{n,11}(\epsilon_\tau), \ldots, LR_{n,1m_0}(\epsilon_\tau))^\top \to_d (v_1, \ldots, v_{m_0})^\top \tag{53}$$

for any $\epsilon_\tau \in (0, 1/2)$, where $v_h = \max\{(\hat{\boldsymbol{t}}^1_{\boldsymbol{\lambda},h})^\top \boldsymbol{\mathcal{I}}^h_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \hat{\boldsymbol{t}}^1_{\boldsymbol{\lambda},h}, (\hat{\boldsymbol{t}}^2_{\boldsymbol{\lambda},h})^\top \boldsymbol{\mathcal{I}}^h_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \hat{\boldsymbol{t}}^2_{\boldsymbol{\lambda},h}\}$. We proceed to show (53). Observe that as in (9), the first, second, and third derivatives of $\ln[g^h(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\psi}^h_\tau, \tau)]$ w.r.t. $\boldsymbol{\lambda_\theta}$ and its first derivative w.r.t. λ_σ become zero when evaluated at $\boldsymbol{\psi}^h_\tau = \boldsymbol{\psi}^{h*}_\tau$. Consequently, $L^h_n(\boldsymbol{\psi}^h_\tau, \tau) - L^h_n(\boldsymbol{\psi}^{h*}_\tau, \tau)$ admits the same expansion (13) as $L_n(\boldsymbol{\psi}_\alpha, \alpha) - L_n(\boldsymbol{\psi}^*_\alpha, \alpha)$ by replacing $(\boldsymbol{t}_n(\boldsymbol{\psi}_\alpha, \alpha), \boldsymbol{S}_n, \boldsymbol{\mathcal{I}}_n, R_n(\boldsymbol{\psi}_\alpha, \alpha))$ in (10)–(13) with $(\boldsymbol{t}_{n,h}(\boldsymbol{\psi}^h_\tau, \tau), \boldsymbol{S}_{n,h}, \boldsymbol{\mathcal{I}}^h_n, R^h_n(\boldsymbol{\psi}^h_\tau, \tau))$, where $(\boldsymbol{S}_{n,h}, \boldsymbol{\mathcal{I}}^h_n)$ is defined in the same manner as $(\boldsymbol{S}_n, \boldsymbol{\mathcal{I}}_n)$ but using $(\tilde{\boldsymbol{s}}_{\boldsymbol{\eta}i}, \boldsymbol{s}^h_{\boldsymbol{\lambda}i})$ in place of $(\boldsymbol{s}_{\boldsymbol{\eta}i}, \boldsymbol{s}_{\boldsymbol{\lambda}i})$. Applying the proof of Proposition 2(b)(c), we have $\boldsymbol{S}_{n,h} \to_d \boldsymbol{S}_h \sim N(0, \boldsymbol{\mathcal{I}}^h)$ and $\boldsymbol{\mathcal{I}}^h_n \to_p \boldsymbol{\mathcal{I}}^h := E[\boldsymbol{S}_{n,h}\boldsymbol{S}^\top_{n,h}]$. Define $\boldsymbol{W}_{n,h}$ in the same manner as $\boldsymbol{W}_n$ but using $(\boldsymbol{S}_n, \boldsymbol{\mathcal{I}})$ in place of $(\boldsymbol{S}_{n,h}, \boldsymbol{\mathcal{I}}^h)$ in the proof of Proposition 3. Then, (53) follows from the application of the proofs of Propositions 2 and 3 for each local MLE by replacing $(\boldsymbol{W}_n, \hat{\boldsymbol{t}}^1_{\boldsymbol{\lambda}}, \hat{\boldsymbol{t}}^2_{\boldsymbol{\lambda}}, \boldsymbol{\mathcal{I}}_{\lambda.\boldsymbol{\eta}})$ with $(\boldsymbol{W}_{n,h}, \hat{\boldsymbol{t}}^1_{\boldsymbol{\lambda},h}, \hat{\boldsymbol{t}}^2_{\boldsymbol{\lambda},h}, \boldsymbol{\mathcal{I}}^h_{\lambda.\boldsymbol{\eta}})$, and collecting the results while noting that $(\boldsymbol{S}^\top_{n,1}, \ldots, \boldsymbol{S}^\top_{n,m_0})^\top \to_d (\boldsymbol{S}^\top_1, \ldots, \boldsymbol{S}^\top_{m_0})^\top$. □

A.6 Proof of Proposition 6

Under $H_{0,2h}$, we obtain for $\boldsymbol{\vartheta}_{m_0+1} \in \Upsilon^*_{2h}$,

$$
\begin{aligned}
&E\left[\{\nabla_{\alpha_h} \ln f_{m_0+1}(Y_i; \boldsymbol{\vartheta}_{m_0+1})\}^2\right] \\
&= \int \frac{\{f(y; \mu_h, \sigma^2_h) - f(y; \mu^*_{m_0}, \sigma^{2*}_{m_0})\}^2}{\sum_{j=1}^{m_0} \alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j)} dy \\
&= \int \frac{\{f(y; \mu_h, \sigma^2_h)\}^2}{\sum_{j=1}^{m_0} \alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j)} dy + \int \frac{\{f(y; \mu^*_{m_0}, \sigma^{2*}_{m_0})\}^2}{\sum_{j=1}^{m_0} \alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j)} dy - 2\int \frac{f(y; \mu_h, \sigma^2_h) f(y; \mu^*_{m_0}, \sigma^{2*}_{m_0})}{\sum_{j=1}^{m_0} \alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j)} dy.
\end{aligned}
\tag{54}
$$

The latter two terms on the right-hand side of (54) are bounded because $f(y; \mu^*_{m_0}, \sigma^{2*}_{m_0}) / \sum_{j=1}^{m_0} \alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j) \le (1/\alpha^*_{m_0})$ for any y, and $f(y; \mu, \sigma^2)$ integrates to one. Thus, the left-hand side of (54) is infinity if and only if the first term on the right-hand side of (54) is infinity.

Because $\max_j a_j \le \sum_{j=1}^{m_0} a_j \le m_0 \max_j a_j$ holds for general $\{a_j\}_{j=1}^{m_0}$, we obtain

$$
\frac{1}{m_0} \frac{\{f(y; \mu_h, \sigma^2_h)\}^2}{\max_j\{\alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j)\}} \le \frac{\{f(y; \mu_h, \sigma^2_h)\}^2}{\sum_{j=1}^{m_0} \alpha^*_j f(y; \sigma^{2*}_j)} \le \frac{\{f(y; \mu_h, \sigma^2_h)\}^2}{\max_j\{\alpha^*_j f(y; \mu^*_j, \sigma^{2*}_j)\}}.
$$

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

$$\frac{\{f(y;\mu_h,\sigma_h^2)\}^2}{f(y;\mu_{m_0}^*,\sigma_{m_0}^{2*})} = \frac{\sigma_{m_0}^*}{(2\pi)^{1/2}\sigma_h^2} \exp\left\{-\frac{1}{\sigma_h^2}(y-\mu_h)^2 + \frac{1}{2\sigma_{m_0}^{2*}}(y-\mu_{m_0}^*)^2\right\}.$$

The stated result follows because the integral of this over $|y| \geq M$ is finite if $\sigma_h^2/\sigma_{m_0}^{2*} < 2$ and infinite if $\sigma_h^2/\sigma_{m_0}^{2*} > 2$, while when $\sigma_h^2 = 2\sigma_{m_0}^{2*}$, it is finite if $\mu_h = \mu_{m_0}^*$ and infinite if $\mu_h \neq \mu_{m_0}^*$. □

A.7 Proof of Proposition 7

For $j = 1, 2$, let $\omega_{n,h}^j$ be the sample counterpart of $(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda},h}^j)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}}^h \hat{\boldsymbol{t}}_{\boldsymbol{\lambda},h}^j$ in Proposition 5 such that the local LRT statistic satisfies $2[L_n^h(\hat{\boldsymbol{\psi}}_\tau^h, \tau) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{m_0})] = \max\{\omega_{n,h}^1, \omega_{n,h}^2\} + o_p(1)$, where $\hat{\boldsymbol{\psi}}_\tau^h$ is the local MLE defined in (52).

For $\tau \in (0,1)$, define $\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau) := \{\boldsymbol{\vartheta}_{m_0+1} \in \Upsilon_{1h}^* : \alpha_h/(\alpha_h + \alpha_{h+1}) = \tau\}$, which gives the true density. Observe that from Assumption 6 and $|x| \leq 1 + |x|^3$, we have $p_n(\sigma_j^2) - p_n(\sigma_j^{2*}) = o_p(n^{1/6})(\sigma_j^2 - \sigma_j^{*2}) = o_p(1 + n^{1/2}(\sigma_j^2 - \sigma_j^{*2})^3) = o_p(1 + n^{1/2}(|\lambda_\sigma|^3 + \lambda_\mu^6))$. Therefore, in view of (43) in the proof of Proposition 2, for any $\boldsymbol{\vartheta}_{m_0+1}$ with $\alpha_h/(\alpha_h + \alpha_{h+1}) = \tau \in (0,1)$ and whose corresponding $\boldsymbol{t}_{n,h}(\boldsymbol{\psi}_\tau^h)$ is $O_p(1)$, we have

$$PL_n(\boldsymbol{\vartheta}_{m_0+1}) - PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau)) = L_n(\boldsymbol{\vartheta}_{m_0+1}) - L_n(\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau)) + o_p(1). \tag{55}$$

First, we show $\mathrm{EM}_n^{h(1)} = \max\{\omega_{n,h}^1, \omega_{n,h}^2\} + o_p(1)$. Because $\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0)$ is the only value of $\boldsymbol{\vartheta}_{m_0+1}$ that yields the true density if $\boldsymbol{\varsigma} \in \Omega_h^*$ and $\alpha_h/(\alpha_h + \alpha_{h+1}) = \tau_0$, $\boldsymbol{\vartheta}_{m_0+1}^{h(1)}(\tau_0)$ equals a reparameterized penalized local MLE in the neighborhood of $\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0)$. Hence, $PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h(1)}(\tau_0)) \geq PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0)) + o_p(1)$ holds, and Proposition E gives $\boldsymbol{\vartheta}_{m_0+1}^{h(1)}(\tau_0) - \boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0) = o_p(1)$. It follows from applying the argument in the proof of Propositions 3 and 5 that $\boldsymbol{t}_{n,h}(\boldsymbol{\psi}_\tau^h)$ corresponding to $\boldsymbol{\vartheta}_{m_0+1}^{h(1)}(\tau_0)$ is $O_p(1)$. Therefore, $\mathrm{EM}_n^{h(1)} = \max\{\omega_{n,h}^1, \omega_{n,h}^2\} + o_p(1)$ follows from (55).

We proceed to show that $\mathrm{EM}_n^{h(K)} = \max\{\omega_{n,h}^1, \omega_{n,h}^2\} + o_p(1)$. Because a generalized EM step never decreases the likelihood value (Dempster et al., 1977), we have $PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h(K)}(\tau_0)) \geq PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h(1)}(\tau_0))$. Therefore, it follows from Propositions E and F and induction that $\boldsymbol{\vartheta}_{m_0+1}^{h(K)}(\tau_0) - \boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0) = o_p(1)$ for any finite K. Let $\tilde{\boldsymbol{\vartheta}}_{m_0+1}^h(\tau^{(K)})$ be the maximizer of $PL_n(\boldsymbol{\vartheta}_{m_0+1})$ under the constraint $\alpha_h/(\alpha_h + \alpha_{h+1}) = \tau^{(K)}$ in an arbitrary small closed neighborhood of $\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau^{(K)})$; then, we have $PL_n(\tilde{\boldsymbol{\vartheta}}_{m_0+1}^h(\tau^{(K)})) \geq PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h(K)}(\tau_0)) + o_p(1)$ from the consistency of $\boldsymbol{\vartheta}_{m_0+1}^{h(K)}(\tau_0)$. Thus, $2[PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h(K)}(\tau_0)) - L_{0,n}(\tilde{\boldsymbol{\vartheta}}_{m_0})] = \max\{\omega_{n,h}^1, \omega_{n,h}^2\} + o_p(1)$ holds because both $2[PL_n(\tilde{\boldsymbol{\vartheta}}_{m_0+1}^h(\tau^{(K)})) - L_{0,n}(\tilde{\boldsymbol{\vartheta}}_{m_0})]$ and $2[PL_n(\boldsymbol{\vartheta}_{m_0+1}^{h(1)}(\tau_0)) - L_{0,n}(\tilde{\boldsymbol{\vartheta}}_{m_0})]$ can be written as $\max\{\omega^1_{n,h}, \omega^2_{n,h}\} + o_p(1)$. Further, because $PL_n(\boldsymbol{\vartheta}^{h(K)}_{m_0+1}(\tau_0)) \geq PL_n(\boldsymbol{\vartheta}^{h(1)}_{m_0+1}(\tau_0)) \geq PL_n(\boldsymbol{\vartheta}^{h*}_{m_0+1}(\tau_0)) + o_p(1)$, applying the proof of Proposition 3(a) to $\boldsymbol{\vartheta}^{h(K)}_{m_0+1}(\tau_0)$ gives that $\boldsymbol{t}_{n,h}(\boldsymbol{\psi}^h_\tau)$ corresponding to $\boldsymbol{\vartheta}^{h(K)}_{m_0+1}(\tau_0)$ is $O_p(1)$, and that $\mathrm{EM}^{h(K)}_n = \max\{\omega^1_{n,h}, \omega^2_{n,h}\} + o_p(1)$ holds for all h from (55). The stated result follows from the definition of $\mathrm{EM}^{(K)}_n$. □

A.8 Proof of Proposition 8

Let $\boldsymbol{\psi}_n$ be the value of $\boldsymbol{\psi}_\alpha$ under $H^n_{(\alpha^*,\boldsymbol{\Delta})}$ and define $V_n = L_n(\boldsymbol{\psi}_n, \alpha^*) - L_{0,n}(\boldsymbol{\gamma}^*, \boldsymbol{\theta}^*, \sigma^{2*})$. Under the null distribution, we have $(LR_n(\epsilon_1), V_n) \to_d (\max\{\sup_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \Lambda^1_{\boldsymbol{\lambda}}} Q(\boldsymbol{t}_{\boldsymbol{\lambda}}), \sup_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \Lambda^2_{\boldsymbol{\lambda}}} Q(\boldsymbol{t}_{\boldsymbol{\lambda}})\}, V)$, where $Q(\boldsymbol{t}_{\boldsymbol{\lambda}}) = 2\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{W}_{\boldsymbol{\lambda}} - \boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}}$ and $V = \boldsymbol{\Delta}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{W}_{\boldsymbol{\lambda}} - (1/2)\boldsymbol{\Delta}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{\Delta}$. From Le Cam's third lemma, the limiting distribution of $LR_n(\epsilon_1)$ under $H^n_{(\alpha^*,\boldsymbol{\Delta})}$ can be determined by the joint null distribution of $(Q(\boldsymbol{t}_{\boldsymbol{\lambda}}), V)$ given by

$$\begin{pmatrix} Q(\boldsymbol{t}_{\boldsymbol{\lambda}}) \\ V \end{pmatrix} \sim N\left(\begin{pmatrix} -\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}} \\ -(1/2)\boldsymbol{\Delta}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{\Delta} \end{pmatrix}, \begin{pmatrix} 4\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}} & 2\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{\Delta} \\ 2\boldsymbol{\Delta}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}} & \boldsymbol{\Delta}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{\Delta} \end{pmatrix} \right).$$

Applying Le Cam's third lemma, we obtain the limiting distribution of $LR_n(\epsilon_1)$ under $H^n_{(\alpha^*,\boldsymbol{\Delta})}$ as $\max\{\sup_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \Lambda^1_{\boldsymbol{\lambda}}} Q_{\boldsymbol{\Delta}}(\boldsymbol{t}_{\boldsymbol{\lambda}}), \sup_{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \Lambda^2_{\boldsymbol{\lambda}}} Q_{\boldsymbol{\Delta}}(\boldsymbol{t}_{\boldsymbol{\lambda}})\}$, where $Q_{\boldsymbol{\Delta}}(\boldsymbol{t}_{\boldsymbol{\lambda}}) \sim N(2\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{\Delta} - \boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}}, 4\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}})$. Because $\boldsymbol{W}_{\boldsymbol{\lambda}} \sim N(0, \boldsymbol{\mathcal{I}}^{-1}_{\boldsymbol{\lambda}.\boldsymbol{\eta}})$, writing $Q_{\boldsymbol{\Delta}}(\boldsymbol{t}_{\boldsymbol{\lambda}}) = 2\boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} (\boldsymbol{W}_{\boldsymbol{\lambda}} + \boldsymbol{\Delta}) - \boldsymbol{t}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}} = (\boldsymbol{W}_{\boldsymbol{\lambda}} + \boldsymbol{\Delta}) \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} (\boldsymbol{W}_{\boldsymbol{\lambda}} + \boldsymbol{\Delta}) - \{\boldsymbol{t}_{\boldsymbol{\lambda}} - (\boldsymbol{W}_{\boldsymbol{\lambda}} + \boldsymbol{\Delta})\}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \{\boldsymbol{t}_{\boldsymbol{\lambda}} - (\boldsymbol{W}_{\boldsymbol{\lambda}} + \boldsymbol{\Delta})\}$ and using $(\hat{\boldsymbol{t}}^j_{\boldsymbol{\lambda},\boldsymbol{\Delta}})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}.\boldsymbol{\eta}} \{\hat{\boldsymbol{t}}^j_{\boldsymbol{\lambda},\boldsymbol{\Delta}} - (\boldsymbol{W}_{\boldsymbol{\lambda}} + \boldsymbol{\Delta})\} = 0$ for $j = 1, 2$ gives the stated limiting distribution of $LR_{1,n}(\epsilon_1)$ under $H^n_{(\alpha^*,\boldsymbol{\Delta})}$. The limiting distribution of EM_n follows because $EM_n = LR_{1,n}(\epsilon_1) + o_p(1)$ from the proof of Proposition 7. Part (c) follows from part (b). □

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 ℓ:*

$$\nabla_{\mu^k} \nabla_{(\sigma^2/2)^\ell} \left[\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma} \right) \right] = \left(\frac{1}{\sigma} \right)^{k+2\ell+1} H^{k+2\ell}\left(\frac{x-\mu}{\sigma} \right) \phi\left(\frac{x-\mu}{\sigma} \right).$$

Proof. The stated result holds trivially when $k = \ell = 0$. Suppose that 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

μ and using the relation $H^{n+1}(x) = xH^n(x) - \nabla H^n(x)$ give

$$\nabla_\mu \left[\left(\frac{1}{\sigma}\right)^{n+1} H^n\left(\frac{x-\mu}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right)\right] = \left(\frac{1}{\sigma}\right)^{n+2} H^{n+1}\left(\frac{x-\mu}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right). \tag{56}$$

Second, differentiating $(1/\sigma)^{n+1}H^n((x-\mu)/\sigma)\phi((x-\mu)/\sigma)$ with respect to σ^2 gives

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

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 $\nabla_{\sigma^2/2}[(1/\sigma)^{n+1}H^n((x-\mu)/\sigma)\phi((x-\mu)/\sigma)] = (1/\sigma)^{n+3}H^{n+2}((x-\mu)/\sigma)\phi((x-\mu)/\sigma)$, and the stated result follows from this and (56). □

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

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

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

$$\frac{d^q f(g(x))}{dx^q} = \sum_{(k_1,\dots,k_q)} \frac{q!}{k_1!\dots k_q!}\left(\frac{\partial^p f(g(x))}{\partial g^p}\right)\left(\frac{\partial g(x)}{\partial x}\right)^{k_1}\left(\frac{1}{2!}\frac{\partial^2 g(x)}{\partial x^2}\right)^{k_2}\cdots\left(\frac{1}{q!}\frac{\partial^q g(x)}{\partial x^q}\right)^{k_q}, \tag{57}$$

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

Proposition C. *Suppose that $g(y|\boldsymbol{x},\boldsymbol{z};\boldsymbol{\psi}_\alpha,\alpha)$ is given by (7), where $\boldsymbol{\psi}=(\boldsymbol{\eta}^\top,\lambda_\mu,\boldsymbol{\lambda_\beta},\lambda_\sigma)^\top$ and $\boldsymbol{\eta}=(\boldsymbol{\gamma}^\top,\nu_\mu,\boldsymbol{\nu_\beta},\nu_\sigma)^\top$. Let g^*, $\ln g^*$, ∇g^*, and $\nabla\ln g^*$ denote $g(y|\boldsymbol{x},\boldsymbol{z};\boldsymbol{\psi}_\alpha,\alpha)$, $\ln g(y|\boldsymbol{x},\boldsymbol{z};\boldsymbol{\psi}_\alpha,\alpha)$, and their derivatives evaluated at $(\boldsymbol{\psi}^*_\alpha,\alpha)$, respectively. Then, for $\lambda_i,\lambda_j,\lambda_k,\lambda_\ell\in\{\lambda_\sigma,\lambda_1,\dots,\lambda_q\}$,*

(a) *for $k=1,2,3$ and $\ell=0,1,\dots,$ $\nabla_{\lambda_\mu^k\boldsymbol{\eta}^\ell}g^*=0$ and $\nabla_{\lambda_\mu^k\boldsymbol{\eta}^\ell}\ln g^*=0$;*

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

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

(d) $\nabla_{\lambda_\mu^8}\ln g^*=\dfrac{\nabla_{\lambda_\mu^8}g^*}{g^*}-\dfrac{8!}{2}\left(\dfrac{\nabla_{\lambda_\mu^4}g^*}{4!g^*}\right)^2$ *and* $\nabla_{\lambda_\mu^4\boldsymbol{\eta}}\ln g^*=\dfrac{\nabla_{\lambda_\mu^4\boldsymbol{\eta}}g^*}{g^*}-\dfrac{\nabla_{\lambda_\mu^4}g^*}{g^*}\dfrac{\nabla_{\boldsymbol{\eta}}g^*}{g^*}$;

(e) *for $\ell=0,1,\dots,$ $\nabla_{\lambda_i\boldsymbol{\eta}^\ell}g^*=0$ and $\nabla_{\lambda_i\boldsymbol{\eta}^\ell}\ln g^*=0$;*

(f) $\nabla_{\lambda_i\lambda_j}\ln g^*=\dfrac{\nabla_{\lambda_i\lambda_j}g^*}{g^*}$, $\nabla_{\lambda_i\lambda_j\lambda_k}\ln g^*=\dfrac{\nabla_{\lambda_i\lambda_j\lambda_k}g^*}{g^*}$, $\nabla_{\lambda_\mu\lambda_i\lambda_j}\ln g^*=\dfrac{\nabla_{\lambda_\mu\lambda_i\lambda_j}g^*}{g^*}$;

(g) *for $k=1,\dots,4,$* $\nabla_{\lambda_\mu^k\lambda_i}\ln g^*=\dfrac{\nabla_{\lambda_\mu^k\lambda_i}g^*}{g^*}$;

(h) $\nabla_{\lambda_i\lambda_j\boldsymbol{\eta}}\ln g^*=\dfrac{\nabla_{\lambda_i\lambda_j\boldsymbol{\eta}}g^*}{g^*}-\dfrac{\nabla_{\lambda_i\lambda_j}g^*}{g^*}\dfrac{\nabla_{\boldsymbol{\eta}}g^*}{g^*}$, $\nabla_{\lambda_\mu\lambda_i\boldsymbol{\eta}}\ln g^*=\dfrac{\nabla_{\lambda_\mu\lambda_i\boldsymbol{\eta}}g^*}{g^*}-\dfrac{\nabla_{\lambda_\mu\lambda_i}g^*}{g^*}\dfrac{\nabla_{\boldsymbol{\eta}}g^*}{g^*}$;

(i) $\nabla_{\lambda_\mu\lambda_i\lambda_j\lambda_k}\ln g^*=\dfrac{\nabla_{\lambda_\mu\lambda_i\lambda_j\lambda_k}g^*}{g^*}-\dfrac{\nabla_{\lambda_\mu\lambda_i}g^*}{g^*}\dfrac{\nabla_{\lambda_j\lambda_k}g^*}{g^*}-\dfrac{\nabla_{\lambda_\mu\lambda_j}g^*}{g^*}\dfrac{\nabla_{\lambda_i\lambda_k}g^*}{g^*}-\dfrac{\nabla_{\lambda_\mu\lambda_k}g^*}{g^*}\dfrac{\nabla_{\lambda_i\lambda_j}g^*}{g^*}$;

(j) $\nabla_{\lambda_i\lambda_j\lambda_k\lambda_\ell}\ln g^*=\dfrac{\nabla_{\lambda_i\lambda_j\lambda_k\lambda_\ell}g^*}{g^*}-\dfrac{\nabla_{\lambda_i\lambda_j}g^*}{g^*}\dfrac{\nabla_{\lambda_k\lambda_\ell}g^*}{g^*}-\dfrac{\nabla_{\lambda_i\lambda_k}g^*}{g^*}\dfrac{\nabla_{\lambda_j\lambda_\ell}g^*}{g^*}-\dfrac{\nabla_{\lambda_i\lambda_\ell}g^*}{g^*}\dfrac{\nabla_{\lambda_j\lambda_k}g^*}{g^*}$;

(k) $\nabla_{\lambda_\mu^2\lambda_i\lambda_j}\ln g^*=\dfrac{\nabla_{\lambda_\mu^2\lambda_i\lambda_j}g^*}{g^*}-2\dfrac{\nabla_{\lambda_\mu\lambda_i}g^*}{g^*}\dfrac{\nabla_{\lambda_\mu\lambda_j}g^*}{g^*}$, $\nabla_{\lambda_\mu^5\lambda_i}\ln g^*=\dfrac{\nabla_{\lambda_\mu^5\lambda_i}g^*}{g^*}-5!\dfrac{\nabla_{\lambda_\mu^4}g^*}{4!g^*}\dfrac{\nabla_{\lambda_\mu\lambda_i}g^*}{g^*}$.

Proof. We prove part (a) for $\ell=0$ first. Suppress all arguments in $g(y|\boldsymbol{x},\boldsymbol{z};\boldsymbol{\psi}_\alpha,\alpha)$ and $f(y|\boldsymbol{x},\boldsymbol{z};\boldsymbol{\gamma},\boldsymbol{\theta},\sigma^2)$ but λ_μ, and rewrite as follows:

$$g(\lambda_\mu)=\alpha f((1-\alpha)\lambda_\mu,(1-\alpha)C_1\lambda_\mu^2)+(1-\alpha)f(-\alpha\lambda_\mu,-\alpha C_2\lambda_\mu^2). \quad (58)$$

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

$$\nabla_{\lambda_\mu^k}f(\lambda_\mu,h(\lambda_\mu))=(\nabla_{\lambda_\mu}+\nabla_u)^k f(\lambda_\mu,h(u))|_{u=\lambda_\mu}=\sum_{j=0}^{k}\binom{k}{j}\nabla_{\lambda_\mu^{k-j}u^j}f(\lambda_\mu,h(u))|_{u=\lambda_\mu}. \quad (59)$$

Because $\nabla_{u^j} u^2|_{u=0} = 0$ except for $j = 2$, it follows from Faà di Bruno's formula (57) that

$$\nabla_{u^j} f((1-\alpha)\lambda_\mu, (1-\alpha)C_1 u^2)|_{\lambda_\mu=u=0} = \begin{cases} 0 & \text{if } j = 1, 3, \\ 2(1-\alpha)C_1 \nabla_h f(0, h(0)) & \text{if } j = 2, \\ 12(1-\alpha)^2 C_1^2 \nabla_{h^2} f(0, h(0)) & \text{if } j = 4, \end{cases} \tag{60}$$

and a similar result holds for $\nabla_{\lambda_\mu^{k-j} u^j} f((1-\alpha)\lambda_\mu, (1-\alpha)C_1 u^2)$ and $\nabla_{\lambda_\mu^{k-j} u^j} f(-\alpha\lambda_\mu, -\alpha C_2 u^2)$.

Differentiating (58) and using (59) ($h(\lambda_\mu)$ corresponds to $(1-\alpha)C_1\lambda_\mu^2$ and $-\alpha C_2 \lambda_\mu^2$), (60), $C_1 - C_2 = -1$, $\nabla_{\mu^2} f(0,0) = 2\nabla_{\sigma^2} f(0,0)$, $\nabla_{\mu^3} f(0,0) = 2\nabla_{\mu\sigma^2} f(0,0)$, and $3((1-\alpha)C_1 + \alpha C_2) = 2\alpha - 1$, we obtain

$$\begin{aligned}
&\nabla_{\lambda_\mu} g(0) = 0, \\
&\nabla_{\lambda_\mu^2} g(0) = \alpha(1-\alpha)\nabla_{\mu^2} f(0,0) + 2\alpha(1-\alpha)(C_1 - C_2)\nabla_{\sigma^2} f(0,0) = 0, \\
&\nabla_{\lambda_\mu^3} g(0) = \alpha(1-\alpha)(1-2\alpha)\nabla_{\mu^3} f(0,0) + 3\alpha(1-\alpha)((1-\alpha)C_1 + \alpha C_2) 2\nabla_{\mu\sigma^2} f(0,0) = 0,
\end{aligned}$$

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

For the second results of part (a) and part (b), suppressing all arguments but λ_μ and $\boldsymbol{\eta}$ from $g(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\psi}_\alpha, \alpha)$ and applying Faà di Bruno's formula (57) to $\partial^q g(\lambda_\mu, \boldsymbol{\eta})/\partial\lambda_\mu^q$, we obtain

$$\begin{aligned}
\frac{\partial^q \ln g(\lambda_\mu, \boldsymbol{\eta})}{\partial \lambda_\mu^q} = &\sum_{(k_1,\ldots,k_q)} \frac{q!}{k_1! \ldots k_q!} \frac{(-1)^{p-1}(p-1)!}{g(\lambda_\mu, \boldsymbol{\eta})^p} \\
&\times \left(\frac{\partial g(\lambda_\mu, \boldsymbol{\eta})}{\partial \lambda_\mu}\right)^{k_1} \left(\frac{1}{2!}\frac{\partial^2 g(\lambda_\mu, \boldsymbol{\eta})}{\partial \lambda_\mu^2}\right)^{k_2} \cdots \left(\frac{1}{q!}\frac{\partial^q g(\lambda_\mu, \boldsymbol{\eta})}{\partial \lambda_\mu^q}\right)^{k_q},
\end{aligned} \tag{61}$$

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 $q = \sum_{i=1}^q i k_i$. For example, setting $q = 2$ gives $\nabla_{\lambda_\mu^2} \ln g(\lambda_\mu, \boldsymbol{\eta}) = \nabla_{\lambda_\mu^2} g(\lambda_\mu, \boldsymbol{\eta})/g(\lambda_\mu, \boldsymbol{\eta}) - (\nabla_{\lambda_\mu} g(\lambda_\mu, \boldsymbol{\eta})/g(\lambda_\mu, \boldsymbol{\eta}))^2$, where the sum is taken over $(k_1, k_2) = (2, 0)$ and $(0, 1)$. The second result of part (a) follows from evaluating (61) at $q = 1, 2, 3$, differentiating it ℓ times with respect to $\boldsymbol{\eta}$, and using the first result of part (a). Part (b) follows from evaluating (61) at $q = 4, 5, 6, 7$ and applying part (a).

For part (c), differentiating (58) and using (59), (60), 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

$$\begin{aligned}\nabla_{\lambda_\mu^4} g(0) &= \alpha(1-\alpha)[(1-\alpha)^3+\alpha^3]\nabla_{\mu^4} f(0,0) + 6\alpha(1-\alpha)((1-\alpha)^2 C_1 - \alpha^2 C_2) \\ &\quad \times 2\nabla_{\mu^2\sigma^2} f(0,0) + 12\alpha(1-\alpha)((1-\alpha)C_1^2 + \alpha C_2^2)\nabla_{\sigma^2\sigma^2} f(0,0) \\ &= \alpha(1-\alpha) b(\alpha) \nabla_{\mu^4} f(0,0),\end{aligned}$$

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

The first result of part (d) follows from evaluating (61) at $q = 8$ and using part (a). The second result of part (d) follows from differentiating (61) at $q = 4$ with respect to $\boldsymbol{\eta}$ and using part (a). A direct calculation gives part (e). Part (f) follows from (61) with $q = 2, 3$ and part (e). Part (g) follows from differentiating (61) at $q = 1, \ldots, 4$ with respect to λ_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 (61) at $q = 2$ and $q = 5$ with respect to $\lambda_i\lambda_j$ and λ_i, respectively, and using parts (a) and (e). □

Proposition D. *Suppose that* $\{Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\}$, $i = 1, \ldots, n$, *are* n *independent observations from density* $f_{m_0}(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\vartheta}_{m_0}^*)$, *Assumption 1 holds, and* $c \in (0, 1]$ *is chosen so that* $\min_{i,j}(\sigma_i^*/\sigma_j^*) \geq c$. *For any* $\boldsymbol{\vartheta}_m^n$ *satisfying* $\min_{i,j}(\sigma_i/\sigma_j) \geq c$ *and* $\sum_{i=1}^n f_m(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\vartheta}_m^n) \geq \sum_{i=1}^n f_{m_0}(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\vartheta}_{m_0}^*) + o_p(n)$ *for all* n, *we have* $\inf_{\boldsymbol{\vartheta}_m^* \in \Upsilon_m^*} ||\boldsymbol{\vartheta}_m^n - \boldsymbol{\vartheta}_m^*|| \to_p 0$, *where* $\Upsilon_m^* := \{\boldsymbol{\vartheta}_m : f_m(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\vartheta}_m) = f_{m_0}(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\vartheta}_{m_0}^*)$ *with probability one*$\}$.

Proof. Our proof closely follows the proof of Theorem 3.3 in Hathaway (1985). Because our model has additional free parameters β_js and $\boldsymbol{\gamma}$, we modify the proof of Hathaway (1985) to consider the joint density of $m_{qp} := m(p+q+1)+1$ observations instead of $m+1$ observations in Hathaway (1985, p. 798). The joint density function of m_{qp} observations is itself a mixture of $m^{m_{qp}}$ components, where each component is given by $\prod_{j=1}^{m_{qp}} P(y_j; \mu_{i_j} + \boldsymbol{x}_j^\top \boldsymbol{\beta}_{i_j} + \boldsymbol{z}_j^\top \boldsymbol{\gamma}, \sigma_{i_j})$ 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_{qp} observations. We verify Assumption 5 of Kiefer and Wolfowitz (1956) for the joint density function of m_{qp} observations by showing that

$$E\left[\prod_{j=1}^{m_{qp}} P(y_j; \mu_{i_j}^* + \boldsymbol{x}_j^\top \boldsymbol{\beta}_{i_j}^* + \boldsymbol{z}_j^\top \boldsymbol{\gamma}^*, \sigma_{i_j}^*)\right] > -\infty \tag{62}$$

for $\boldsymbol{\vartheta}_m^* \in \Upsilon_m^*$ and

$$E \sup_{\boldsymbol{\vartheta}_m \in \Theta_{\boldsymbol{\vartheta}_m}(c)} \ln \left[\prod_{j=1}^{m_{qp}} P(y_j; \mu_{i_j} + \boldsymbol{x}_j^\top \boldsymbol{\beta}_{i_j} + \boldsymbol{z}_j^\top \boldsymbol{\gamma}, \sigma_{i_j}) \right] < \infty, \tag{63}$$

which correspond to equations (3.1) and (3.2) in Hathaway (1985), respectively. (62) follows from the argument in the proof of Theorem 3.3 of Hathaway (1985). For (63), proceeding as in Hathaway (1985, pp. 798–799), we can show that $\sup_{\boldsymbol{\vartheta}_m \in \Theta_{\boldsymbol{\vartheta}_m}(c)} \ln \left[\prod_{j=1}^{m_{qp}} P(y_j; \mu_{i_j} + \boldsymbol{x}_j^\top \boldsymbol{\beta}_{i_j} + \boldsymbol{z}_j^\top \boldsymbol{\gamma}, \sigma_{i_j}) \right]$ is no greater than, for some $\ell \in \{1, \dots, m\}$ and $j_1, \dots, j_{p+q+2} \in \{1, \dots, m_{qp}\}$,

$$\sup_{\mu_\ell, \boldsymbol{\beta}_\ell, \sigma_\ell, \boldsymbol{\gamma}} \ln \left(\beta(\sigma_\ell) \prod_{k=1}^{p+q+2} P(y_{j_k}; \mu_\ell + \boldsymbol{x}_{j_k}^\top \boldsymbol{\beta}_\ell + \boldsymbol{z}_{j_k}^\top \boldsymbol{\gamma}, \sigma_\ell) \right), \tag{64}$$

where $\beta(\sigma_\ell) := (2\pi)^{(p+q+2-m_{qp})/2} (c\sigma_\ell)^{p+q+2-m_{qp}}$.

Note that $\prod_{k=1}^{p+q+2} P(y_{j_k}; \mu_\ell + \boldsymbol{x}_{j_k}^\top \boldsymbol{\beta}_\ell + \boldsymbol{z}_{j_k}^\top \boldsymbol{\gamma}, \sigma_\ell)$ is the likelihood function of a linear Gaussian model. Therefore, the maximized value of (64) equals $C - (m_{qp}/2) \ln S$, where C is a finite constant that depends only on m, p, and q; and S is the sum of squared residuals from regressing $\{y_{j_k}\}_{k=1}^{p+q+2}$ on $\{1, \boldsymbol{x}_{j_k}, \boldsymbol{z}_{j_k}\}_{k=1}^{p+q+2}$. 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 (64) is finite, and (63) holds. This verifies Assumption 5 of Kiefer and Wolfowitz (1956), and the stated consistency result under Assumption 1 follows. □

Given the parameter $\boldsymbol{\vartheta}_m$, write the distribution of $(\boldsymbol{\theta}, \sigma^2)$ associated with $\boldsymbol{\vartheta}_m$ as $G(\boldsymbol{\theta}, \sigma^2; \boldsymbol{\vartheta}_m) := \sum_{j=1}^m \alpha_j I\{(\boldsymbol{\theta}_j, \sigma_j^2) \leq (\boldsymbol{\theta}, \sigma^2)\}$, and let $G^*(\boldsymbol{\theta}, \sigma^2) := G(\boldsymbol{\theta}, \sigma^2; \boldsymbol{\vartheta}_{m_0}^*)$ denote the true mixing distribution. Let $\gamma_{(s)}$ denote the s-th element of $\boldsymbol{\gamma}$. Define the penalized log-likelihood function as $PL_n(\boldsymbol{\vartheta}_m) := \sum_{i=1}^n \ln \sum_{j=1}^m \alpha_j f(Y_i | \boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\gamma}, \boldsymbol{\theta}_j, \sigma_j^2) + \sum_{j=1}^m p_n(\sigma_j^2)$. The following proposition shows the consistency of the penalized MLE. It extends Theorem 5 of Chen et al. (2008) to accommodate a regressor.

Proposition E. *Suppose that Assumptions 1 and 5 hold. For any $\boldsymbol{\vartheta}_m^n$ satisfying $PL_n(\boldsymbol{\vartheta}_m^n) \geq PL_n(\boldsymbol{\vartheta}_{m_0}^*) + o_p(1)$ for all n, we have $\sum_{s=1}^p |\arctan \gamma_{(s)}^n - \arctan \gamma_{0(s)}^*| + \int \int_{\mathbb{R}^{q+1} \times \mathbb{R}^+} |G(\boldsymbol{\theta}, \sigma; \boldsymbol{\vartheta}_m^n) - G^*(\boldsymbol{\theta}, \sigma)| e^{-||\boldsymbol{\theta}|| - \sigma} d\boldsymbol{\theta} d\sigma \to_p 0$.*

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 MLE by showing that the penalty term $\sum_{j=1}^m p_n(\sigma_j^2)$ in effect places a positive constant lower bound on σ_j^2. Key results for establishing the existence of such a lower bound are 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 θ in CTZ). In a model with a covariate, Lemmas 1 and 2 of CTZ hold when we replace their θ, X_i, and $\sup_\theta$ with our $\mu + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{\gamma}$, Y_i, and $\sup_{\mu,\boldsymbol{x},\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{\gamma}}$, respectively. 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 $\sup_{\mu,\boldsymbol{x},\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{\gamma}} \sum_{i=1}^n I(|Y_i - \mu - \boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{\gamma}| < |\sigma \ln \sigma|)$. We can follow the proof of Theorem 4 of CTZ to set a lower bound on σ_j^2. Once a lower bound on σ_j^2 is set, the consistency is proven by resorting to Kiefer and Wolfowitz (1956) as CTZ do. The presence of a structural parameter $\boldsymbol{\gamma}$ has no effect because Kiefer and Wolfowitz (1956) accommodate a structural parameter. □

Proposition F. *Suppose that Assumptions 1, 4, and 5 hold. If* $\boldsymbol{\vartheta}_{m_0+1}^{h(k)}(\tau_0) - \boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0) = o_p(1)$*, then* $\alpha_h^{(k+1)}/[\alpha_h^{(k+1)} + \alpha_{h+1}^{(k+1)}] - \tau_0 = o_p(1)$.

Proof. We suppress (τ_0) from $\boldsymbol{\vartheta}_{m_0+1}^{h(k)}(\tau_0)$ and $\boldsymbol{\vartheta}_{m_0+1}^{h*}(\tau_0)$. The proof is similar to the proof of Lemma 3 of Li and Chen (2010). Let $f_i(\boldsymbol{\gamma}, \boldsymbol{\theta}, \sigma^2)$ and $f_i(\boldsymbol{\vartheta}_{m_0+1})$ denote $f(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\gamma}, \boldsymbol{\theta}, \sigma^2)$ and $f_{m_0+1}(Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\vartheta}_{m_0+1})$, respectively. Applying a Taylor expansion to $\alpha_h^{(k+1)} = n^{-1}\sum_{i=1}^n w_{ih}^{(k)}$ and using $\boldsymbol{\vartheta}_{m_0+1}^{h(k)} - \boldsymbol{\vartheta}_{m_0+1}^{h*} = o_p(1)$, we obtain

$$\alpha_h^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \frac{\alpha_h^{(k)} f_i(\boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}_h^{(k)}, \sigma_h^{2(k)})}{f_i(\boldsymbol{\vartheta}_{m_0+1}^{h(k)})} = \frac{1}{n}\sum_{i=1}^n \frac{\tau_0 \alpha_h^* f_i(\boldsymbol{\gamma}^*, \boldsymbol{\theta}_h^*, \sigma_h^{2*})}{f_i(\boldsymbol{\vartheta}_{m_0+1}^{h*})} + o_p(1) = \tau_0 \alpha_h^* + o_p(1),$$

where the last equality follows from $E[f_i(\boldsymbol{\gamma}^*, \boldsymbol{\theta}_h^*, \sigma_h^{2*})/f_i(\boldsymbol{\vartheta}_{m_0+1}^{h*})] = 1$ and the law of large numbers. A similar argument gives $\alpha_h^{(k+1)} = (1 - \tau_0)\alpha_h^* + o_p(1)$, and the stated result follows. □

C Computer experiments to obtain the empirical formula in (23)

The empirical formula in (23) 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 $1,000$ 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 $\hat{y} = -0.892 - 0.542 \ln(a/(2 - a)) - 0.236 \ln(\omega_{12}/(1 - \omega_{12})) - 55.06/n$ with $R^2 = 0.53$. Setting $\hat{y} = 0$ and adjusting the value of the multiplicative constant yields the first formula in (23). 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 (23).

D Additional results from empirical examples

Estimation results for 30 stocks in the Dow Jones Industrial Average are reported in Table 1. The modified EM test chooses a three-component model for 15 stocks, and a model with four or more components for 13 stocks. Of the 30 stocks, the AIC and BIC select a different number of components from the modified EM test for 17 and 15 stocks, respectively.

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

	p-value of modified EM test (in %)			Selection by		
Security	$H_0 : m = 1$	$H_0 : m = 2$	$H_0 : m = 3$	Modified EM	AIC	BIC
1. Allied Chemical Corp	0.0	0.0	0.0	4	4	3
2. Aluminum Co America	0.0	0.0	43.4	3	4	3
3. American Brands Inc	0.0	0.0	0.0	4	4	4
4. American Can Co	0.0	0.0	0.0	4	4	4
5. American Tel and Teleg	0.0	0.0	0.1	4	3	3
6. Bethlehem Steel Corp	0.0	0.0	0.0	4	4	4
7. Du Pont	0.0	0.0	20.9	3	4	3
8. Eastman Kodak Co	0.0	0.0	62.7	3	4	3
9. Exxon Corp	0.0	19.9	19.2	2	4	2
10. General Electric Co	0.0	0.0	99.9	3	4	2
11. General Foods Corp	0.0	0.0	0.0	4	4	4
12. General Motors Corp	0.0	0.1	67.9	3	4	2
13. Goodyear	0.0	10.6	18.2	2	4	3
14. Inco Ltd	0.0	0.0	2.5	3	4	3
15. Inter. Business Mach.	0.0	0.0	99.0	3	4	3
16. Inter. Harvester Co	0.0	0.0	0.0	4	4	3
17. Inter. Paper Co	0.0	0.0	7.1	3	3	3
18. Johns Manville Corp	0.0	0.0	34.3	3	4	4
19. Merck and Co. Inco	0.0	0.0	15.0	3	4	3
20. Minnesota Mng & Mfg	0.0	0.0	27.6	3	4	2
21. Owens Illinois Inco	0.0	0.0	10.7	3	4	3
22. Proctor & Gamble Co	0.0	0.0	24.9	3	4	3
23. Sears Roebuck & Co	0.0	0.0	0.0	4	4	3
24. Standard Oil Co Cal	0.0	0.0	0.2	4	4	2
25. Texaco Inc	0.0	0.0	0.0	4	4	3
26. Union Carbide Corp	0.0	0.0	91.5	3	4	4
27. United Aircraft Prod	0.0	0.0	0.0	4	4	3
28. United Sts Stl Corp	0.0	0.0	40.4	3	4	3
29. Westinghouse Elc Co	0.0	0.0	0.8	4	4	3
30. Woolworth F W Co	0.0	0.0	0.0	4	4	3

References

Andrews, D. W. K. (1999), "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

Chen, J. and Li, P. (2009), "Hypothesis Test for Normal Mixture Models: The EM Approach," *Annals of Statistics*, 37, 2523–2542.

Chen, J., Li, P., and Fu, Y. (2012), "Inference on the Order of a Normal Mixture," *Journal of the American Statistical Association*, 107, 1096–1105.

Chen, J., Tan, X., and Zhang, R. (2008), "Inference for Normal Mixtures in Mean and Variance," *Statistica Sinica*, 18, 443–465.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via EM Algorithm (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Hathaway, R. J. (1985), "A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions," *Annals of Statistics*, 13, 795–800.

Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887–906.

Li, P. and Chen, J. (2010), "Testing the Order of a Finite Mixture," *Journal of the American Statistical Association*, 105, 1084–1092.