

Frequentist Model Averaging in Structure Equation Model With Ordinal Data:
Supplementary Material

Frequentist Model Averaging in Structure Equation Model With Ordinal Data:
Supplementary Material

1 Introduction

This supplementary material contains mathematical proofs of the main results in the paper. The reader is directed to the main text for the explanations of the notations.

Suppose that there exists a $q \times 1$ vector of underlying continuous random variables $\mathbf{x}^*$ that follows a multivariate normal distribution with variances 1. Let $\mathbf{x}$ be the vector of discrete counterparts of $\mathbf{x}^*$, x_g the g th entry of $\mathbf{x}$, and m_g the number of possible outcomes that x_g can take. Then, $x_g = a$ if $\tau_{g,a-1} < x_g^* \leq \tau_{g,a}$, where x_g^* is the g th underlying continuous random variable, $\tau_{g,0} \equiv -\infty$, and $\tau_{g,m_g} \equiv \infty$. The vector of unknown thresholds for x_g is denoted by $\boldsymbol{\tau}_g$. The probability of observing $\mathbf{x} = \begin{pmatrix} a_1 & a_2 & \cdots & a_q \end{pmatrix}$ is

$$\pi_{a_1 a_2 \cdots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) = \int_{\tau_{1,a_1-1}}^{\tau_{1,a_1}} \int_{\tau_{2,a_2-1}}^{\tau_{2,a_2}} \cdots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(t_1, \dots, t_q; \boldsymbol{\rho}) dt_q \cdots dt_1,$$

where $\boldsymbol{\tau} = (\boldsymbol{\tau}_1^T, \boldsymbol{\tau}_2^T \cdots \boldsymbol{\tau}_q^T)^T$ is the vector of all thresholds, $\boldsymbol{\rho}$ is the vector of all unique polychoric correlation coefficients, and $\phi_q(\cdot; \boldsymbol{\rho})$ is the density function of a q -dimensional normal random variable with means 0, variances 1, and correlation coefficients $\boldsymbol{\rho}$. Let $\boldsymbol{\tau}_{gh}^T = (\boldsymbol{\tau}_g^T, \boldsymbol{\tau}_h^T)$. The probability of observing $x_g = a$ and $x_h = b$ is

$$\pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh}, \rho_{gh}) = \int_{\tau_{g,a-1}}^{\tau_{g,a}} \int_{\tau_{h,b-1}}^{\tau_{h,b}} \phi_2(t_1, t_2; \rho_{gh}) dt_2 dt_1, \quad (1)$$

where ρ_{gh} is the polychoric correlation coefficient between x_g and x_h ; and $\phi_2(\cdot; \rho)$ is the density function of a bivariate normal random variable with means 0, variances 1, and correlation ρ . Likewise, we can let $\pi_{a_g a_h a_s a_t}^{(ghst)}(\boldsymbol{\tau}_{ghst}, \boldsymbol{\rho}_{ghst})$ be the probability of observing $\begin{pmatrix} x_g & x_h & x_s & x_t \end{pmatrix} = \begin{pmatrix} a_g & a_h & a_s & a_t \end{pmatrix}$ and $\pi_{a_g a_h a_s}^{(ghs)}(\boldsymbol{\tau}_{ghs}, \boldsymbol{\rho}_{ghs})$ be the probability of

observing $\begin{pmatrix} x_g & x_h & x_s \end{pmatrix} = \begin{pmatrix} a_g & a_h & a_s \end{pmatrix}$, where

$$\begin{aligned}\boldsymbol{\tau}_{ghst} &= \begin{pmatrix} \boldsymbol{\tau}_g^T & \boldsymbol{\tau}_h^T & \boldsymbol{\tau}_s^T & \boldsymbol{\tau}_t^T \end{pmatrix}^T, \\ \boldsymbol{\rho}_{ghst} &= \begin{pmatrix} \rho_{gh} & \rho_{gs} & \rho_{gt} & \rho_{hs} & \rho_{ht} & \rho_{st} \end{pmatrix}^T, \\ \boldsymbol{\tau}_{ghs} &= \begin{pmatrix} \boldsymbol{\tau}_g^T & \boldsymbol{\tau}_h^T & \boldsymbol{\tau}_s^T \end{pmatrix}^T, \\ \boldsymbol{\rho}_{ghs} &= \begin{pmatrix} \rho_{gh} & \rho_{gs} & \rho_{hs} \end{pmatrix}^T.\end{aligned}$$

Throughout the paper, we estimate the polychoric correlation coefficient using the two-step procedure of Olsson (1979). First, the thresholds are estimated from the univariate standard normal distribution. Second, the correlation coefficient is estimated conditional on the estimated thresholds. The multinomial log-likelihood function in the second step is

$$\ell_{gh}(\boldsymbol{\tau}_{gh}, \boldsymbol{\rho}_{gh}) = \sum_{i=1}^n \ell_{gh,i}(\boldsymbol{\tau}_{gh}, \boldsymbol{\rho}_{gh}),$$

where n is the sample size, $\boldsymbol{\tau}_{gh}$ is fixed to the first-step estimates $\hat{\boldsymbol{\tau}}_{gh}$, and

$$\begin{aligned}\ell_{gh,i}(\boldsymbol{\tau}_{gh}, \boldsymbol{\rho}_{gh}) &= \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} I_{gh,i}(a, b) \log \pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh}, \boldsymbol{\rho}_{gh}), \\ I_{gh,i}(a, b) &= \begin{cases} 1, & \text{if we observe } x_{g,i} = a \text{ and } x_{h,i} = b, \\ 0, & \text{otherwise.} \end{cases}\end{aligned}$$

Here, $\{\ell_{gh,i}(\boldsymbol{\tau}_{gh}, \boldsymbol{\rho}_{gh})\}$ are mutually independent.

2 Polychoric Correlation Estimation

2.1 Standard Asymptotic Framework

Under the standard asymptotic framework, we assume that the true values of the polychoric correlations do not depend on the sample size n . Most SEM related studies lie within the standard asymptotic framework. Suppose that the true value of $\boldsymbol{\rho}$ is $\boldsymbol{\rho}_0$. Under some conditions, Jöreskog (1994) showed that

$$\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Upsilon}), \quad (2)$$

where Υ is the asymptotic covariance matrix. The expression of Υ can be obtained from Jöreskog (1994).

2.2 Local Asymptotic Framework

Under the local asymptotic framework, we assume that the true values of some parameters depend on the sample size n . Our paper assumes that the true value of γ is $\gamma_0 + \delta/\sqrt{n}$. Hence, the true model-implied σ is of the form

$$\sigma_{true} = \sigma(\beta_{true}) = \sigma(\theta_0, \gamma_0 + \delta/\sqrt{n}).$$

Suppose that σ_{true} can be expanded into

$$\sigma_{true} = \sigma_0 + \delta^*/\sqrt{n} + O(n^{-1}), \quad (3)$$

where $\sigma_0 = \sigma(\theta_0, \gamma_0)$ and

$$\delta^* = \frac{\partial \sigma(\theta_0, \gamma_0)}{\partial \gamma^T} \delta.$$

Then, for any pair (g, h) , the polychoric correlation coefficient between x_g and x_h , denoted by ρ_{gh} , is also locally drifted. Its true value satisfies

$$\rho_{gh,true} = \rho_{gh,0} + \delta_{gh}^*/\sqrt{n} + O(n^{-1})$$

for some local parameter δ_{gh}^* . However, the thresholds τ_g and τ_h are not locally drifted.

Consequently, the true probability of observing $x_g = a$ and $x_h = b$ becomes

$\pi_{ab}^{(gh)}(\tau_{gh,0}, \rho_{gh,true})$, where the subscript 0 is used to denote the true value of the thresholds.

2.3 Regularity Conditions

Let $\bar{O}^{(\rho)}$ and $\bar{O}_g^{(\tau)}$ be closed balls that contain ρ_0 and $\tau_{g,0}$ as interior points, respectively. Suppose that the polychoric correlation matrix with the correlation coefficients ρ_0 is a positive definite matrix; and that all $\pi_{a_1 a_2 \dots a_q}(\tau_0, \rho_0) > 0$. Because the polychoric correlation matrix is a continuous function of ρ and $\pi_{a_1 a_2 \dots a_q}(\tau, \rho)$ is a

continuous function of $(\boldsymbol{\tau}, \boldsymbol{\rho})$, we can choose $\bar{O}^{(\rho)}$ and $\bar{O}_g^{(\tau)}$ such that the polychoric correlation matrix remains positive definite and $\pi_{a_1 a_2 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ is bounded below from zero for all $(\boldsymbol{\tau}, \boldsymbol{\rho}) \in \bar{O}^{(\tau)} \times \bar{O}^{(\rho)}$, where $\bar{O}^{(\tau)} = \bar{O}_1^{(\tau)} \times \bar{O}_2^{(\tau)} \times \dots \times \bar{O}_q^{(\tau)}$ and $\times$ is the Cartesian product. Recall that a compact set is closed and bounded in the Euclidean space. Then by the Tychonoff theorem (e.g., Richmond, 2020, Theorem 6.2.5), $\bar{O}^{(\tau)}$ and $\bar{O}^{(\tau)} \times \bar{O}^{(\rho)}$ are also closed and bounded. By the extreme value theorem (e.g., Rudin, 1976, Theorem 4.16), any continuous function will attain its minimum and maximum on $\bar{O}^{(\tau)} \times \bar{O}^{(\rho)}$. Thus, we can find open balls $O^{(\tau)} \subset \bar{O}^{(\tau)}$ and $O^{(\rho)} \subset \bar{O}^{(\rho)}$ such that $\boldsymbol{\tau}_0$ is an interior point of $O^{(\tau)}$, $\boldsymbol{\rho}_0$ is an interior point of $O^{(\rho)}$, and the continuous functions that are bounded on $\bar{O}^{(\tau)} \times \bar{O}^{(\rho)}$ are still bounded on $O^{(\tau)} \times O^{(\rho)}$.

Before we proceed to the main results and their proofs, we need to establish some properties regarding the partial derivatives of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ for $(\boldsymbol{\tau}, \boldsymbol{\rho}) \in O^{(\tau)} \times O^{(\rho)}$.

Result 1. *For any finite dimension $q \geq 2$, $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$, the first-order partial derivatives of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$, the second-order partial derivatives of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$, and the second-order partial derivatives of $\partial \log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\rho}$ are all bounded on $O^{(\tau)} \times O^{(\rho)}$.*

Proof. The boundness of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ holds since $\pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) > 0$ is bounded on the open set $O^{(\tau)} \times O^{(\rho)}$ by construction. Regarding the first-order and higher-order partial derivatives of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$, $\pi_{a_1 a_2 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ is bounded below from zero for all $(\boldsymbol{\tau}, \boldsymbol{\rho}) \in O^{(\tau)} \times O^{(\rho)}$. Then we only need to show that the derivatives of $\pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ are bounded. Since $\pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ involves a multiple integral, the partial derivatives of $\pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ is also a multiple integral. For some constants d_k , some vectors $\mathbf{a}_k$ and some

matrices $\mathbf{A}_k$, we have

$$\begin{aligned}
& \left| \int_{\tau_{1,a_1-1}}^{\tau_{1,a_1}} \cdots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(\mathbf{t}; \boldsymbol{\rho}) \prod_k \left(d_k + \mathbf{a}_k^T \mathbf{t} + \mathbf{t}^T \mathbf{A}_k \mathbf{t} \right) dt_q \cdots dt_1 \right| \\
& \leq \int_{\tau_{1,a_1-1}}^{\tau_{1,a_1}} \cdots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(\mathbf{t}; \boldsymbol{\rho}) \prod_k \left| d_k + \mathbf{a}_k^T \mathbf{t} + \mathbf{t}^T \mathbf{A}_k \mathbf{t} \right| dt_q \cdots dt_1 \\
& \leq E \left| \prod_k \left(d_k + \mathbf{a}_k^T \mathbf{t} + \mathbf{t}^T \mathbf{A}_k \mathbf{t} \right) \right|, \tag{4}
\end{aligned}$$

where $\mathbf{t} = (t_1, \dots, t_q)^T$. That is, the above integral is bounded by a linear combination of absolute moments of a multivariate normal distribution. Kamat (1953) have proposed an approach to compute those absolute moments. Then, (4) is bounded on $O^{(\tau)} \times O^{(\rho)}$, if d_k , $\mathbf{a}_k$ and $\mathbf{A}_k$ are all continuous functions of $(\boldsymbol{\tau}, \boldsymbol{\rho})$. Thus, it suffices to show that the integrand needed to compute the partial derivatives are $\phi_q(\mathbf{t}; \boldsymbol{\rho})$ multiplied by polynomials of $\mathbf{t}$.

Let ρ_g be the g th entry of $\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}$ be a correlation matrix with off-diagonal elements $\boldsymbol{\rho}$. That is, a random variable with density function $\phi_q(\mathbf{t}; \boldsymbol{\rho})$ has the correlation matrix $\boldsymbol{\Sigma}$. Let σ_{ij} be the (i, j) th entry of $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma}^{(g)} = \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \rho_g} \boldsymbol{\Sigma}^{-1},$$

and $\sigma_{ij}^{(g)}$ be the (i, j) th entry of $\boldsymbol{\Sigma}^{(g)}$. For any g ,

$$\frac{\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \rho_g} = \int_{\tau_{1,a_1-1}}^{\tau_{1,a_1}} \cdots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(\mathbf{t}; \boldsymbol{\rho}) \frac{\partial \log \phi_q(\mathbf{t}; \boldsymbol{\rho})}{\partial \rho_g} dt_q \cdots dt_1,$$

where

$$\frac{\partial \log \phi_q(\mathbf{t}; \boldsymbol{\rho})}{\partial \rho_g} = -\frac{1}{2} \text{tr} \{ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(g)} \} + \frac{1}{2} \mathbf{t}^T \boldsymbol{\Sigma}^{(g)} \mathbf{t}.$$

Thus, $\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \rho_g$ is bounded on $O^{(\tau)} \times O^{(\rho)}$. Regarding the first-order partial

derivative with respect to $\boldsymbol{\tau}$, note that

$$\begin{aligned} \left| \frac{\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \tau_{1,a}} \right| &= \int_{\tau_{2,a_2-1}}^{\tau_{2,a_2}} \dots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(\tau_{1,a}, t_2, \dots, t_q; \boldsymbol{\rho}) dt_q \dots dt_2 \\ &\leq \phi_1(\tau_{1,a}) \leq \frac{1}{\sqrt{2\pi}}, \text{ if } \tau_{1,a} = \tau_{1,a_1} \text{ or } \tau_{1,a_1-1}, \\ \left| \frac{\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \tau_{1,a}} \right| &= 0, \text{ if } \tau_{1,a} \neq \tau_{1,a_1} \text{ and } \tau_{1,a} \neq \tau_{1,a_1-1}, \end{aligned}$$

where $\phi_1(\cdot)$ is the density of a univariate standard normal random variable. Therefore, the first-order partial derivatives of $\pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ with respect to $\boldsymbol{\tau}_1$ are bounded. By symmetry in the sense that the order of the integral can be interchanged, the first-order partial derivatives of $\pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ with respect to $\boldsymbol{\tau}$ are bounded. Consequently, the first-order partial derivatives of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$ are bounded on $O^{(\tau)} \times O^{(\rho)}$.

Now consider the first-order partial derivatives of $\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\rho}$. Note that

$$\frac{\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \rho_g \partial \rho_h} = \int_{\tau_{1,a_1-1}}^{\tau_{1,a_1}} \dots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \frac{\partial^2 \phi_q(\mathbf{t}; \boldsymbol{\rho})}{\partial \rho_g \partial \rho_h} dt_q \dots dt_1,$$

where

$$\begin{aligned} \frac{\partial^2 \phi_q(\mathbf{t}; \boldsymbol{\rho})}{\partial \rho_g \partial \rho_h} &= \phi_q(\mathbf{t}; \boldsymbol{\rho}) \left(\frac{1}{4} \text{tr} \{ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(g)} \} \text{tr} \{ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(h)} \} - \frac{1}{2} \text{tr} \left\{ \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \rho_g \partial \rho_h} \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(g)} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(h)} \right\} \right) \\ &\quad + \frac{1}{4} \phi_q(\mathbf{t}; \boldsymbol{\rho}) \mathbf{t}^T \left(2 \frac{\partial \boldsymbol{\Sigma}^{(g)}}{\partial \rho_h} - \text{tr} \{ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(g)} \} \boldsymbol{\Sigma}^{(h)} - \text{tr} \{ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(h)} \} \boldsymbol{\Sigma}^{(g)} \right) \mathbf{t} \\ &\quad + \frac{1}{4} \phi_q(\mathbf{t}; \boldsymbol{\rho}) \left(\mathbf{t}^T \boldsymbol{\Sigma}^{(h)} \mathbf{t} \right) \left(\mathbf{t}^T \boldsymbol{\Sigma}^{(g)} \mathbf{t} \right). \end{aligned}$$

Hence, $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \rho_g \partial \rho_h$ is bounded on $O^{(\tau)} \times O^{(\rho)}$. Regarding the first-order partial derivatives of $\partial \log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\rho}$ with respect to $\boldsymbol{\tau}$, we still use the derivatives with respect to $\boldsymbol{\tau}_1$ as an example. It is easy to see that $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a} \partial \rho_h = 0$ if $a \neq a_1 - 1$ and $a \neq a_1$. If $a = a_1 - 1$ or a_1 , then

$$\left| \frac{\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \tau_{1,a} \partial \rho_h} \right| = \left| \int_{\tau_{2,a_2-1}}^{\tau_{2,a_2}} \dots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \left[\frac{1}{2} \phi_q(\tilde{\mathbf{t}}_a; \boldsymbol{\rho}) \left(\tilde{\mathbf{t}}_a^T \boldsymbol{\Sigma}^{(g)} \tilde{\mathbf{t}}_a \right) - \frac{1}{2} \phi_q(\tilde{\mathbf{t}}_a; \boldsymbol{\rho}) \text{tr} \{ \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{(g)} \} \right] dt_q \dots dt_2 \right|,$$

where $\tilde{\mathbf{t}}_a = \begin{pmatrix} \tau_{1,a} & t_2 & \cdots & t_q \end{pmatrix}^T$. Hence, $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a} \partial \rho_h$ is bounded on $O^{(\tau)} \times O^{(\rho)}$ if $a = a_{1-1}$ or a_1 . Regarding the first-order partial derivatives of $\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a}$ with respect to $\boldsymbol{\tau}$, it is easy to see that $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a_1} \partial \tau_{1,a} = 0$ if $a \neq a_1$. If $a = a_1$,

$$\frac{\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \tau_{1,a_1}^2} = \int_{\tau_{2,a_2-1}}^{\tau_{2,a_2}} \cdots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(\tau_{1,a_1}, t_2, \dots, t_q; \boldsymbol{\rho}) \left(-\mathbf{e}_1^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{t}}_{a_1} \right) dt_q \cdots dt_2,$$

where $\mathbf{e}_i$ is a vector with i th entry being 1 and other entries being 0. Hence,

$\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a_1}^2$ is bounded if $a = a_1$. Further, $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a_1} \partial \tau_{2,a} = 0$, if $a \neq a_2 - 1$ and $a \neq a_2$. If $a = a_2 - 1$ or a_2 , then

$$\left| \frac{\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})}{\partial \tau_{1,a_1} \partial \tau_{2,a}} \right| = \left| \int_{\tau_{3,a_3-1}}^{\tau_{3,a_3}} \cdots \int_{\tau_{q,a_q-1}}^{\tau_{q,a_q}} \phi_q(\tau_{1,a_1}, \tau_{2,a}, t_3, \dots, t_q; \boldsymbol{\rho}) dt_q \cdots dt_3 \right| \leq \phi_2(\tau_{1,a_1}, \tau_{2,a}; \rho_{12}).$$

Therefore, by symmetry, every entry in $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T$ are bounded on $O^{(\tau)} \times O^{(\rho)}$.

This completes the proof for the boundness of the second-order partial derivatives of $\log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho})$.

At last consider the second-order partial derivatives of $\partial \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\rho}$. From $\partial^2 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \rho_g \partial \rho_h$ derived above, the integrand needed in $\partial^3 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \rho_g \partial \rho_h \partial \rho_i$ is $\phi_q(\mathbf{t}; \boldsymbol{\rho})$ multiplied by a polynomial of $\mathbf{t}$ up to the sixth order. From $\partial^2 \pi_{a_1 a_2 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a} \partial \rho_h$ derived above, the integrand needed in $\partial^3 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \tau_{1,a} \partial \rho_g \partial \rho_h$ is $\phi_q(\mathbf{t}; \boldsymbol{\rho})$ multiplied by a polynomial of $\mathbf{t}$ up to the fourth order. From $\partial^2 \pi_{a_1 a_2 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T$ derived above, the integrands needed in $\partial^3 \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T \partial \rho_g$ is also a multivariate normal density multiplied by a polynomial of $\mathbf{t}$ up to the fourth order. Therefore, the second-order partial derivatives of $\partial \log \pi_{a_1 \dots a_q}(\boldsymbol{\tau}, \boldsymbol{\rho}) / \partial \boldsymbol{\rho}$ are bounded on $O^{(\tau)} \times O^{(\rho)}$.

□

In order to show the main results, we need the following regularity conditions.

A1 For any pair (g, h) , $\rho_{gh, true} = \rho_{gh, 0} + \delta_{gh}^* / \sqrt{n} + O(n^{-1})$ for some local parameter δ_{gh}^* .

A2 For any $(\tau_{gh}, \rho_{gh}) \neq (\tau_{gh,0}, \rho_{gh,0})$, $P(\ell_{gh,i}(\tau_{gh}, \rho_{gh}) = \ell_{gh,i}(\tau_{gh,0}, \rho_{gh,0})) < 1$.

A3 For any pair (g, h) , $\partial \ell_{gh}(\tau_{gh}, \rho_{gh}) / \partial \rho_{gh} = 0$ always has a unique solution for any $\tau_{gh} \in O_g^{(\tau)} \times O_h^{(\tau)}$.

A4 For any pair (g, h) , the information

$$E_0 \left(\frac{\partial^2 \ell_{gh,i}(\tau_{gh,0}, \rho_{gh,0})}{\partial \rho^2} \right) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \pi_{ab}(\tau_{gh,0}, \rho_{gh,0}) \frac{\partial^2 \log \pi_{ab}(\tau_{gh,0}, \rho_{gh,0})}{\partial \rho^2} \neq 0.$$

Assumptions A2 and A3 ensure the identifiability. Assumption A4 ensures that the reciprocal of the Fisher information is well-defined. Assumption A1 means that the true values of the polychoric correlation coefficients are locally drifted in a $n^{-1/2}$ neighborhood. This is a mild assumption when the true values of the structural equation model in the main text are locally drifted. The model-implied covariance matrix of the structural equation model in the main text is

$$\begin{pmatrix} \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} [\Gamma \text{var}(\boldsymbol{\xi}) \Gamma^T + \text{var}(\boldsymbol{\varepsilon})] \left[(\mathbf{I} - \mathbf{B})^{-1} \right]^T \Lambda_y^T & \Lambda_y (\mathbf{I} - \mathbf{B})^{-1} \Gamma \text{var}(\boldsymbol{\xi}) \Lambda_x^T \\ \Lambda_x \text{var}(\boldsymbol{\xi}) \Gamma^T \left[(\mathbf{I} - \mathbf{B})^{-1} \right]^T \Lambda_y^T & \Lambda_x \text{var}(\boldsymbol{\xi}) \Lambda_x^T \end{pmatrix} + \text{var} \begin{pmatrix} \boldsymbol{\delta}_x \\ \boldsymbol{\delta}_y \end{pmatrix}, \quad (5)$$

provided that $\mathbf{I} - \mathbf{B}$ is invertible. Let $\boldsymbol{\sigma}$ be the vector of all unique elements in (5) and $\boldsymbol{\beta}$ be the vector of free parameters in Λ_x , Λ_y , $\mathbf{B}$, Γ , $\text{var}(\boldsymbol{\xi})$, $\text{var}(\boldsymbol{\delta}_x)$, $\text{var}(\boldsymbol{\delta}_y)$, $\text{cov}(\boldsymbol{\delta}_x, \boldsymbol{\delta}_y^T)$, and $\text{var}(\boldsymbol{\varepsilon})$. In order to link (5) to Assumption A1, we need the following mild assumptions on the covariance matrix (5).

A5 The covariance matrix (5) evaluated at $\boldsymbol{\beta}_0$ is positive definite. $\boldsymbol{\sigma}(\boldsymbol{\beta})$ is a twice continuously differentiable function of $\boldsymbol{\beta}$ on a neighborhood of $\boldsymbol{\beta}_0$. The partial derivatives $\partial \boldsymbol{\sigma}(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta}$ and $\partial^2 \boldsymbol{\sigma}(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ are finite.

By positive definiteness and the continuity in Assumption A5, we can find open balls $O^{(\theta)}$ and $O^{(\gamma)}$ such that $\boldsymbol{\theta}_0$ is an interior point of $O^{(\theta)}$, $\boldsymbol{\gamma}_0$ is an interior point of $O^{(\gamma)}$; (5) is positive definite for all $(\boldsymbol{\theta}, \boldsymbol{\gamma}) \in O^{(\theta)} \times O^{(\gamma)}$; and $\boldsymbol{\sigma}(\boldsymbol{\theta}, \boldsymbol{\gamma})$, $\partial \boldsymbol{\sigma}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$ and $\partial^2 \boldsymbol{\sigma}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ are bounded on $O^{(\theta)} \times O^{(\gamma)}$. If the structural equation model in the main text is the true

data generation process, then $\boldsymbol{\rho} = \boldsymbol{\sigma}(\boldsymbol{\theta}, \boldsymbol{\gamma})$. By the continuity, we can let $O^{(\rho)}$ be an open ball that is nested in the image of $O^{(\theta)} \times O^{(\gamma)}$ under $\boldsymbol{\sigma}(\boldsymbol{\theta}, \boldsymbol{\gamma})$, and then Result 1 holds.

Further, Assumption A5 implies that $\partial \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}$ and $\partial^2 \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T$ are continuous functions of $\boldsymbol{\gamma} \in O^{(\gamma)}$. For a sufficiently large n , $\boldsymbol{\gamma}_0 + \boldsymbol{\delta} / \sqrt{n}$ is also an interior point of $O^{(\gamma)}$. Then, the Taylor's theorem (Rudin, 1976, pp 110 and pp 243) yields

$$\boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 + \boldsymbol{\delta} / \sqrt{n}) = \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0) + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}^T} \boldsymbol{\delta} / \sqrt{n} + \frac{1}{2} \boldsymbol{\delta}^T \frac{\partial^2 \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \tilde{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \boldsymbol{\delta} / n, \quad (6)$$

where $\tilde{\boldsymbol{\gamma}}$ lies between $\boldsymbol{\gamma}_0 + \boldsymbol{\delta} / \sqrt{n}$ and $\boldsymbol{\gamma}_0$. Since $\partial \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}$ and $\partial^2 \boldsymbol{\sigma}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T$ are bounded on $O^{(\gamma)}$, the third term in the right-hand side of (6) is $O(n^{-1})$. Thus, Assumption A1 holds under Assumption A5.

Using Assumption A1 and Result 1, we can expand $\pi_{a_g a_h a_s a_t}^{(ghst)}(\boldsymbol{\tau}_{ghst}, \boldsymbol{\rho}_{ghst})$ that will be used later. For notational simplicity, we will drop the indices from $\pi_{a_g a_h a_s a_t}^{(ghst)}(\boldsymbol{\tau}_{ghst}, \boldsymbol{\rho}_{ghst})$ and express it as $\pi(\boldsymbol{\tau}, \boldsymbol{\rho})$ in this subsection, whenever no confusion arises. The partial derivatives shown in Result 1 imply that $\log \pi(\boldsymbol{\tau}, \boldsymbol{\rho})$ has continuous second-order partial derivatives on $O^{(\tau)} \times O^{(\rho)}$. Because of Assumption A1, $\boldsymbol{\rho}_{true}$ is an interior point of $O^{(\rho)}$ for a sufficiently large n . Hence, the Taylor's theorem (e.g., Rudin, 1976, pp 110 and pp 243) yields

$$\begin{aligned} \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_{true}) &= \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0) + \frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} (\boldsymbol{\rho}_{true} - \boldsymbol{\rho}_0) \\ &\quad + \frac{1}{2} (\boldsymbol{\rho}_{true} - \boldsymbol{\rho}_0)^T \frac{\partial^2 \log \pi(\boldsymbol{\tau}_0, \tilde{\boldsymbol{\rho}})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^T} (\boldsymbol{\rho}_{true} - \boldsymbol{\rho}_0) \\ &= \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0) + \frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} (\boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1})) \\ &\quad + \frac{1}{2} (\boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1}))^T \frac{\partial^2 \log \pi(\boldsymbol{\tau}_0, \tilde{\boldsymbol{\rho}})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^T} (\boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1})) \\ &= \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0) + \frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1}), \end{aligned} \quad (7)$$

where $\boldsymbol{\rho}_0$ is the vector that collects all $\rho_{gh,0}$, and $\tilde{\boldsymbol{\rho}}$ lies between $\boldsymbol{\rho}_0$, and $\boldsymbol{\rho}_{true}$ and $\boldsymbol{\delta}^*$ is the vector that collects all δ_{gh}^* . The remainder in (7) is $O(n^{-1})$ since the partial derivatives are bounded on $O^{(\tau)} \times O^{(\rho)}$ by Result 1 such that $\frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} O(n^{-1}) = O(n^{-1})$ and

$\frac{\partial^2 \log \pi(\boldsymbol{\tau}_0, \bar{\boldsymbol{\rho}})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} = O(n^{-1/2})$. By taking $\exp(\cdot)$ on both sides of (7), the expansion of $\log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_{true})$ also gives the representation

$$\pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_{true}) = \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0) \exp \left\{ \frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1}) \right\}.$$

Using the Taylor series $\exp(x) = 1 + x + \exp(\tilde{x})x^2/2$ for some $0 < \tilde{x} < x$, we further obtain

$$\pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_{true}) = \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0) \left\{ 1 + \frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} + R(\boldsymbol{\delta}^*) \right\}, \quad (8)$$

where the non-random remainder

$$R(\boldsymbol{\delta}^*) = O(n^{-1}) + \frac{\exp(\tilde{x})}{2} \left[\frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1}) \right]^2$$

and $\tilde{x}$ lies between 0 and $\frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} + O(n^{-1})$. Since the partial derivatives are bounded by Result 1, $\frac{\partial \log \pi(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}^T} \boldsymbol{\delta}^* / \sqrt{n} = O(n^{-1/2})$ and $\tilde{x} = O(n^{-1/2})$. Hence, $R(\boldsymbol{\delta}^*) = O(n^{-1})$.

When the dimension is two, similar steps yield

$$\begin{aligned} \pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,true}) &= \pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0}) \exp \left\{ \frac{\partial \log \pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0})}{\partial \rho} \delta_{gh}^* / \sqrt{n} + O(n^{-1}) \right\} \\ &= \pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0}) \left[1 + \frac{\partial \log \pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0})}{\partial \rho} \delta_{gh}^* / \sqrt{n} + R_{ab}(\delta_{gh}^*) \right] \end{aligned} \quad (9)$$

for some non-random remainder $R_{ab}(\delta_{gh}^*) = O(n^{-1})$.

2.4 Consistency and Asymptotic Normality

Throughout the paper, we use $E_{true}(\cdot)$ to denote the expectation under the true distribution and $E_0(\cdot)$ to denote the expectation under the distribution with ρ_0 as the parameter value. For example,

$$\begin{aligned} E_0[\ell_{gh,i}(\boldsymbol{\tau}_{gh,0}, \rho_{gh})] &= \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \pi_{ab}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0}) \log \pi_{ab}(\boldsymbol{\tau}_{gh,0}, \rho_{gh}), \\ E_{true}[\ell_{gh,i}(\boldsymbol{\tau}_{gh,0}, \rho_{gh})] &= \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \pi_{ab}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,true}) \log \pi_{ab}(\boldsymbol{\tau}_{gh,0}, \rho_{gh}). \end{aligned}$$

The variances $\text{var}_{true}(\cdot)$ and $\text{var}_0(\cdot)$ are defined similarly.

When the two-step procedure is used to estimate the polychoric correlation coefficient, the thresholds τ_g are estimated from the univariate standard normal distributions. Hence, $\hat{\tau}_g$ is a still consistent estimator of τ_g and $\sqrt{n}(\hat{\tau}_g - \tau_g)$ is asymptotically normal, for any g . In the second step, $\ell_{gh}(\hat{\tau}_{gh}, \rho_{gh})$ is maximized with respect to ρ_{gh} , a pseudo-maximum likelihood setting. Hence, we will extend Gong and Samaniego (1981) from the standard asymptotic framework to the local asymptotic framework.

Before we present the main results and their proofs, two well-known theorems are presented here, which will be repeatedly used.

1. Law of Large Numbers (LLN; e.g., Jiang, 2010, Theorem 6.4). Let $X_{n,i}$, $i = 1, \dots, n$, $n = 1, 2, \dots$, be a triangular array of random variables such that for each n , the $X_{n,i}$'s are independent. Then, $\sum_{i=1}^n X_{n,i} \xrightarrow{p} 0$ as $n \rightarrow \infty$ if and only if

$$\begin{aligned} \sum_{i=1}^n P(|X_{n,i}| > \epsilon) &\rightarrow 0, \\ \sum_{i=1}^n E\{X_{n,i} I(|X_{n,i}| \leq \kappa)\} &\rightarrow 0, \\ \sum_{i=1}^n \text{var}\{X_{n,i} I(|X_{n,i}| \leq \kappa)\} &\rightarrow 0, \end{aligned}$$

for every $\epsilon > 0$ and some $\kappa > 0$.

2. Lindeberg-Feller Central Limit Theorem (e.g., van der Vaart, 1998, Proposition 2.27). For each n , let $Y_{n,1}, \dots, Y_{n,k_n}$ be independent random vectors with finite variances such that

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} E\left[\|Y_{n,i}\|^2 I\{\|Y_{n,i}\| \geq \epsilon\}\right] &= 0, \quad \text{for every } \epsilon > 0 \\ \sum_{i=1}^{k_n} \text{cov}(Y_{n,i}) &\rightarrow \Sigma. \end{aligned}$$

Then the sequence $\sum_{i=1}^{k_n} [Y_{n,i} - E(Y_{n,i})]$ converges in distribution to a normal $N(0, \Sigma)$ distribution.

Theorem 1. *Consider any pair (g, h) . For $\varepsilon > 0$, let $A_n(\varepsilon)$ be the event that, for any pairs (g, h) , there exists a root $\hat{\rho}_{gh}$ of the equation $\partial \ell_{gh}(\hat{\tau}_{gh}, \rho) / \partial \rho = 0$ for which*

$|\hat{\rho}_{gh} - \rho_{gh,0}| < \varepsilon$. Suppose that Assumptions A1 and A2 hold. Then for any $\varepsilon > 0$, $P\{A_n(\varepsilon)\} \rightarrow 1$ as $n \rightarrow \infty$, under the local asymptotic framework. If Assumption A3 also holds, then $\hat{\rho} \xrightarrow{p} \rho_0$, under the local asymptotic framework.

Proof of Theorem 1. The proof by and large follows Theorem 2.1 in Gong and Samaniego (1981), but to the local asymptotic framework. Let $A_{gh,n}(\varepsilon)$ be the event that there exists a root $\hat{\rho}_{gh}$ of the equation $\partial \ell_{gh}(\hat{\tau}_{gh}, \rho) / \partial \rho = 0$ for which $|\hat{\rho}_{gh} - \rho_{gh,0}| < \varepsilon$. Suppose that, for any $\varepsilon_1 > 0$ and any $\varepsilon_2 > 0$, there exists a N_{gh} such that

$$P\{A_{gh,n}(\varepsilon_1)\} > 1 - 3\varepsilon_2, \quad (10)$$

for all $n > N_{gh}$. Let $N^* = \max_{(g,h)} N_{gh}$. Then,

$$P\{A_n(\varepsilon_1)\} = P\{A_{12,n}(\varepsilon_1) \cap A_{13,n}(\varepsilon_1) \cap \cdots \cap A_{q-1,q,n}(\varepsilon_1)\} > 1 - \frac{q(q-1)}{2} 3\varepsilon_2,$$

for all $n > N^*$. This means that $P\{A_n(\varepsilon_1)\} \rightarrow 1$ as $n \rightarrow \infty$. Further, under the assumption A3 that $\partial \ell_{gh}(\hat{\tau}_{gh}, \rho) / \partial \rho = 0$ always has a unique root for any (g, h) , we reach

$$P\left\{\sqrt{\sum_{g < h} (\hat{\rho}_{gh} - \rho_{gh,0})^2} < \sqrt{\frac{q(q-1)}{2}} \varepsilon_1\right\} > 1 - \frac{q(q-1)}{2} 3\varepsilon_2.$$

In other words, $\hat{\rho} \xrightarrow{p} \rho_0$. Thus, in the rest of the proof, we will show that (10) holds.

For notational simplicity, we drop the index g and h in the proof if no confusion arises. Let $O^{(\tau)}$ and $O^{(\rho)}$ be the open sets as stated in section 2.3. That is, τ_0 and ρ_0 are interior points of $O^{(\tau)}$ and $O^{(\rho)}$; and $\inf_{(\tau, \rho) \in O^{(\tau)} \times O^{(\rho)}} \pi_{ab}(\tau, \rho) > 0$, for any a and b . By LLN,

$$\frac{1}{n} \ell(\tau_0, \rho) - E_{true}[\ell_i(\tau_0, \rho)] \xrightarrow{p} 0, \quad (11)$$

for any fixed $\rho \in O^{(\rho)}$, provided that

$$\sum_{i=1}^n P(|\ell_i(\tau_0, \rho) - E_{true}[\ell_i(\tau_0, \rho)]| > n\varepsilon) \rightarrow 0, \quad (12)$$

$$\sum_{i=1}^n E_{true} \left\{ \frac{\ell_i(\tau_0, \rho) - E_{true}[\ell_i(\tau_0, \rho)]}{n} I(|\ell_i(\tau_0, \rho) - E_{true}[\ell_i(\tau_0, \rho)]| \leq n\kappa) \right\} \rightarrow 0, \quad (13)$$

$$\sum_{i=1}^n \text{var}_{true} \left\{ \frac{\ell_i(\tau_0, \rho) - E_{true}[\ell_i(\tau_0, \rho)]}{n} I(|\ell_i(\tau_0, \rho) - E_{true}[\ell_i(\tau_0, \rho)]| \leq n\kappa) \right\} \rightarrow 0, \quad (14)$$

for any $\varepsilon > 0$ and some $\kappa > 0$; and the expectations and variances are finite.

By Result 1, $\log \pi_{ab}(\boldsymbol{\tau}_0, \rho)$ is bounded on $O^{(\rho)}$ for any (a, b) . Then, $\ell_i(\boldsymbol{\tau}_0, \rho)$,

$$\begin{aligned} E_0[\ell_i(\boldsymbol{\tau}_0, \rho)] &= \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \log \pi_{ab}(\boldsymbol{\tau}_0, \rho), \\ E_0[\ell_i^2(\boldsymbol{\tau}_0, \rho)] &= \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) [\log \pi_{ab}(\boldsymbol{\tau}_0, \rho)]^2, \end{aligned}$$

and $\text{var}_0(\ell_i(\boldsymbol{\tau}_0, \rho))$ are all bounded on $O^{(\rho)}$. By Assumption A1, equation (9) holds if

$\rho_{true} \in O^{(\rho)}$. Hence, we have

$$\begin{aligned} E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)] &= \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_{true}) \log \pi_{ab}(\boldsymbol{\tau}_0, \rho) \\ &= E_0[\ell_i(\boldsymbol{\tau}_0, \rho)] + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \log \pi_{ab}(\boldsymbol{\tau}_0, \rho) \delta^* / \sqrt{n} \\ &\quad + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \log \pi_{ab}(\boldsymbol{\tau}_0, \rho) R_{ab}(\delta^*), \\ \text{var}_{true}(\ell_i(\boldsymbol{\tau}_0, \rho)) &= E_{true}[\ell_i^2(\boldsymbol{\tau}_0, \rho)] - E_{true}^2[\ell_i(\boldsymbol{\tau}_0, \rho)] \\ &= E_0[\ell_i^2(\boldsymbol{\tau}_0, \rho)] + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \log^2 \pi_{ab}(\boldsymbol{\tau}_0, \rho) \delta^* / \sqrt{n} \\ &\quad + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \log^2 \pi_{ab}(\boldsymbol{\tau}_0, \rho) R_{ab}(\delta^*) - E_{true}^2[\ell_i(\boldsymbol{\tau}_0, \rho)]. \end{aligned}$$

By Result 1, $\sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \log \pi_{ab}(\boldsymbol{\tau}_0, \rho)$, $\sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \log^2 \pi_{ab}(\boldsymbol{\tau}_0, \rho)$,

$$\begin{aligned} &\sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \log \pi_{ab}(\boldsymbol{\tau}_0, \rho), \\ \text{and} \quad &\sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \log^2 \pi_{ab}(\boldsymbol{\tau}_0, \rho), \end{aligned}$$

are all bounded on $O^{(\rho)}$. Since $R_{ab}(\delta^*) = O(n^{-1})$ in (9), we have

$$\begin{aligned} E_{true}(\ell_i(\boldsymbol{\tau}_0, \rho)) &= E_0[\ell_i(\boldsymbol{\tau}_0, \rho)] + O(n^{-1/2}), \\ \text{var}_{true}(\ell_i(\boldsymbol{\tau}_0, \rho)) &= \text{var}_0(\ell_i(\boldsymbol{\tau}_0, \rho)) + O(n^{-1/2}), \end{aligned}$$

and they are both finite for any fixed $\rho \in O^{(\rho)}$. Since both $\ell_i(\boldsymbol{\tau}_0, \rho)$ and $E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]$ are bounded, $|\ell_i(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]|$ is bounded on $\rho \in O^{(\rho)}$. Hence,

$|\ell_i(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]| < n\varepsilon$ for a sufficiently large n . Consequently, (12) holds.

Likewise, for a sufficiently large n , $|\ell_i(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]| \leq n\kappa$ for fixed κ . Thus, (13) and (14) hold, since

$$\sum_{i=1}^n E_{true} \left\{ \frac{\ell_i(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]}{n} \right\} = 0$$

and

$$\begin{aligned} \sum_{i=1}^n \text{var}_{true} \left\{ \frac{\ell_i(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]}{n} \right\} &= \frac{\text{var}_{true} \{\ell_i(\boldsymbol{\tau}_0, \rho)\}}{n} \\ &= \frac{\text{var}_0(\ell_i(\boldsymbol{\tau}_0, \rho)) + O(n^{-1/2})}{n} \rightarrow 0. \end{aligned}$$

Therefore, the law of large numbers (11) is applicable and

$$P \left(\left| \frac{1}{n} \ell(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)] \right| > \epsilon \right) \rightarrow 0,$$

for any $\epsilon > 0$, as $n \rightarrow \infty$. It further yields

$$\frac{1}{n} \ell(\boldsymbol{\tau}_0, \rho) \xrightarrow{p} E_0[\ell_i(\boldsymbol{\tau}_0, \rho)],$$

since the local drift in the expectation vanishes as $n \rightarrow \infty$.

From section 2.3 we know that $\ell(\boldsymbol{\tau}, \rho)$ has continuous partial derivatives with respect to $(\boldsymbol{\tau}, \rho)$ on the open set $O^{(\tau)} \times O^{(\rho)}$. Hence, for a sufficiently large n , the consistency of $\hat{\boldsymbol{\tau}}$ and the Taylor's theorem yields

$$\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) = \frac{1}{n} \ell(\boldsymbol{\tau}_0, \rho) + \frac{1}{n} \frac{\partial \ell(\tilde{\boldsymbol{\tau}}, \rho)}{\partial \boldsymbol{\tau}^T} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + o_P(1),$$

where $\tilde{\boldsymbol{\tau}}$ lies between $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}_0$. By Result 1, $|\partial \log \pi_{ab}(\boldsymbol{\tau}, \rho) / \partial \boldsymbol{\tau}|$ is bounded on $O^{(\tau)} \times O^{(\rho)}$. Hence, for all $(\boldsymbol{\tau}, \rho) \in O^{(\tau)} \times O^{(\rho)}$,

$$\left| \frac{\partial \ell_i(\boldsymbol{\tau}, \rho)}{\partial \boldsymbol{\tau}} \right| = \sum_{a,b} \left| I_i(a, b) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}, \rho)}{\partial \boldsymbol{\tau}} \right|$$

is also bounded. Since $\hat{\boldsymbol{\tau}}$ is still a consistent estimator of $\boldsymbol{\tau}_0$,

$P \left\{ (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)^T (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) < \epsilon \right\} \rightarrow 1$ as $n \rightarrow \infty$ for any $\epsilon > 0$. For a sufficiently small ϵ , the

open ball $\{\boldsymbol{\tau}; (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)^T (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) < \epsilon\}$ is a subset of $O^{(\tau)}$. Hence, $P(\hat{\boldsymbol{\tau}} \in O^{(\tau)}) \rightarrow 1$.

Equivalently,

$$P((\hat{\boldsymbol{\tau}}, \rho) \in O^{(\tau)} \times O^{(\rho)}) \rightarrow 1$$

for any fixed $\rho \in O^{(\rho)}$, as $n \rightarrow \infty$. Then $n^{-1} \partial \ell(\tilde{\boldsymbol{\tau}}, \rho) / \partial \boldsymbol{\tau}$ is bounded in probability. Hence,

$$P\left(\left|\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) - \frac{1}{n} \ell(\boldsymbol{\tau}_0, \rho)\right| > \epsilon\right) \rightarrow 0.$$

Above all,

$$\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)] \xrightarrow{p} 0, \quad (15)$$

for $\rho \in O^{(\rho)}$, since for any $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]\right| > 2\epsilon\right) &\leq P\left(\left|\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) - \frac{1}{n} \ell(\boldsymbol{\tau}_0, \rho)\right| > \epsilon\right) \\ &\quad + P\left(\left|\frac{1}{n} \ell(\boldsymbol{\tau}_0, \rho) - E_{true}[\ell_i(\boldsymbol{\tau}_0, \rho)]\right| > \epsilon\right), \end{aligned}$$

which converges to 0 as $n \rightarrow \infty$. It further yields

$$\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) \xrightarrow{p} E_0[\ell_i(\boldsymbol{\tau}_0, \rho)],$$

since the local drift in the expectation vanishes as $n \rightarrow \infty$.

The rest of the proof mimics the proof of Theorem 2.1 in Gong and Samaniego (1981). By (15),

$$\left[\frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho) - \frac{1}{n} \ell(\hat{\boldsymbol{\tau}}, \rho_0)\right] - \{E_0[\ell_i(\boldsymbol{\tau}_0, \rho)] - E_0[\ell_i(\boldsymbol{\tau}_0, \rho_0)]\} \xrightarrow{p} 0.$$

By Jensen's inequality

$$E_0[\ell_i(\boldsymbol{\tau}_0, \rho)] - E_0[\ell_i(\boldsymbol{\tau}_0, \rho_0)] \leq \log E_0 \frac{\prod_{a=1}^{m_x} \prod_{b=1}^{m_y} [\pi_{ab}(\boldsymbol{\tau}_0, \rho)]^{I_i(a,b)}}{\prod_{a=1}^{m_x} \prod_{b=1}^{m_y} [\pi_{ab}(\boldsymbol{\tau}_0, \rho_0)]^{I_i(a,b)}} = 0.$$

By Assumption A2, the equality is attained only if $\rho = \rho_0$. Hence, $n^{-1} [\ell(\hat{\boldsymbol{\tau}}, \rho) - \ell(\hat{\boldsymbol{\tau}}, \rho_0)]$ converges in probability to a negative number for $\rho \neq \rho_0$. Consequently, for any small

$\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, there exists a N such that

$$1 - \varepsilon_2 < P(\hat{\boldsymbol{\tau}} \in O^{(\tau)}), \quad (16)$$

$$1 - \varepsilon_2 < P(\ell(\hat{\boldsymbol{\tau}}, \rho_0 + \varepsilon_1) < \ell(\hat{\boldsymbol{\tau}}, \rho_0)), \quad (17)$$

$$1 - \varepsilon_2 < P(\ell(\hat{\boldsymbol{\tau}}, \rho_0 - \varepsilon_1) < \ell(\hat{\boldsymbol{\tau}}, \rho_0)), \quad (18)$$

for all $n > N$. Here (16) holds since $\hat{\boldsymbol{\tau}}$ is a consistent estimator of $\boldsymbol{\tau}_0$; (17) and (18) hold because of the convergence of $n^{-1}[\ell(\hat{\boldsymbol{\tau}}, \rho) - \ell(\hat{\boldsymbol{\tau}}, \rho_0)]$ that we just showed. Suppose that the event

$$\{\hat{\boldsymbol{\tau}} \in O^{(\tau)}, \ell(\hat{\boldsymbol{\tau}}, \rho_0 + \varepsilon_1) < \ell(\hat{\boldsymbol{\tau}}, \rho_0), \text{ and } \ell(\hat{\boldsymbol{\tau}}, \rho_0 - \varepsilon_1) < \ell(\hat{\boldsymbol{\tau}}, \rho_0)\} \quad (19)$$

holds. Note that, for a given $\hat{\boldsymbol{\tau}}$, $\ell(\hat{\boldsymbol{\tau}}, \rho)$ is a continuous function of ρ on the closed and bounded interval $[\rho_0 - \varepsilon_1, \rho_0 + \varepsilon_1]$. By the extreme value theorem (Rudin, 1976, Theorem 4.16), $\ell(\hat{\boldsymbol{\tau}}, \rho)$ must attain a maximum. Since the values of $\ell(\hat{\boldsymbol{\tau}}, \rho)$ at boundaries are lower than the value at an interior point ($\rho = \rho_0$) by event (19), the maximum of $\ell(\hat{\boldsymbol{\tau}}, \rho)$ on the closed interval $[\rho_0 - \varepsilon_1, \rho_0 + \varepsilon_1]$ must be attained at an interior point. This means that, if event (19) holds, $\ell(\hat{\boldsymbol{\tau}}, \rho)$ has a local maximum $\hat{\rho}$ on the open interval $(\rho_0 - \varepsilon_1, \rho_0 + \varepsilon_1)$. Since $\ell(\hat{\boldsymbol{\tau}}, \rho)$ is differentiable with respect to ρ , the local maximum $\hat{\rho}$ satisfies $\partial \ell(\hat{\boldsymbol{\tau}}, \rho) / \partial \rho = 0$. By inequalities (16), (17), and (18), the event (19) holds with probability greater than $1 - 3\varepsilon_2$, which completes the proof of (10).

□

Theorem 2. *Suppose that Assumptions A1, A2, A3, and A4 hold. Then*

$\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_{true}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Upsilon})$, under the local asymptotic framework, where $\boldsymbol{\Upsilon}$ is the same as the asymptotic covariance matrix in (2).

Proof of Theorem 2. For ease of presentation, we first prove the asymptotic normality of the polychoric correlation coefficient estimator between two variables. Second, we extend the proof to the multivariate case. Since the univariate distribution is not locally drifted, the expansion for the thresholds in Jöreskog (1994) is still applicable. However, ρ is locally

drifted by Assumption A1. Hence, the expansion for ρ in Jöreskog (1994) is not directly applicable.

The marginal probabilities of observing $x_g = a$ and $x_h = b$ are

$$\pi_{g,a}(\boldsymbol{\tau}_g) = \int_{\tau_{g,a-1}}^{\tau_{g,a}} \phi_1(t) dt, \quad \pi_{h,b}(\boldsymbol{\tau}_h) = \int_{\tau_{h,b-1}}^{\tau_{h,b}} \phi_1(t) dt,$$

respectively, where $\phi_1(\cdot)$ is the density function of a standard normal random variable. To estimate $\boldsymbol{\tau}_g$, the univariate log-likelihood scaled by n^{-1} is

$$n^{-1} \sum_{i=1}^n \sum_{a=1}^{m_g} I\{x_{g,i} = a\} \log \pi_{g,a}(\boldsymbol{\tau}_g), \quad (20)$$

where $\tau_{g,0} \equiv -\infty$ and $\tau_{g,m_g} \equiv \infty$. Denote the observed relative frequency matrix by $\mathbf{P}^{(gh)}$ ($m_g \times m_h$). Let the operator $\text{diag}(\cdot)$ construct a diagonal matrix using the enclosed vector as diagonal elements. Jöreskog (1994) showed that

$$\sqrt{n}(\hat{\boldsymbol{\tau}}_g - \boldsymbol{\tau}_{g,0}) = \left(\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} \right)^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n} \mathbf{P}^{(gh)} \mathbf{1}_g + o_P(1), \quad (21)$$

where $\mathbf{B}_{g,0} = \mathbf{B}_g(\boldsymbol{\tau}_{g,0})$ with

$$\mathbf{B}_g(\boldsymbol{\tau}_g) = \begin{pmatrix} \phi_1(\tau_{g,1}) & 0 & \cdots & 0 \\ -\phi_1(\tau_{g,1}) & \phi_1(\tau_{g,2}) & \cdots & 0 \\ 0 & -\phi_1(\tau_{g,2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_1(\tau_{g,m_g-1}) \\ 0 & 0 & \cdots & -\phi_1(\tau_{g,m_g-1}) \end{pmatrix},$$

$\mathbf{D}_{g,0} = \mathbf{D}_g(\boldsymbol{\pi}_{g,0})$ with $\mathbf{D}_g(\boldsymbol{\pi}_g) = \text{diag}(\boldsymbol{\pi}_g)$ and $\boldsymbol{\pi}_g = \left(\pi_{g,1} \ \cdots \ \pi_{g,m_g} \right)^T$, and $\mathbf{1}_g$ is a $m_h \times 1$ column vector of ones. It is worth mentioning that $\sqrt{n} \mathbf{P}^{(gh)} \mathbf{1}_g$ in (21) diverges as $n \rightarrow \infty$, but $\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n} \mathbf{P}^{(gh)} \mathbf{1}_g$ does not. Let $\boldsymbol{\pi}_0^{(gh)}$ be the $m_g \times m_h$ matrix with entries $\pi_{ab}^{(gh)}(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0})$. It is easy to see that $\mathbf{D}_{g,0}^{-1} \boldsymbol{\pi}_0^{(gh)} \mathbf{1}_g = \mathbf{1}_g$ and $\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \boldsymbol{\pi}_0^{(gh)} \mathbf{1}_g = \mathbf{B}_{g,0}^T \mathbf{1}_g = \mathbf{0}$. Hence, we get $\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n} \mathbf{P}^{(gh)} \mathbf{1}_g = \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n} \left(\mathbf{P}^{(gh)} - \boldsymbol{\pi}_0^{(gh)} \right) \mathbf{1}_g$. Then, an equivalent

expression of (21) is

$$\sqrt{n}(\hat{\tau}_g - \tau_{g,0}) = \left(\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} \right)^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n} \left(\mathbf{P}^{(gh)} - \boldsymbol{\pi}_0^{(gh)} \right) \mathbf{1}_g + o_P(1). \quad (22)$$

In fact, $\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n} \mathbf{P}^{(gh)} \mathbf{1}_g$ originates from the first-order derivative of (20) given by a vector with entries

$$\left(\frac{n^{-1} \sum_{i=1}^n I\{x_{g,i} = a\}}{\pi_{g,a}(\boldsymbol{\tau}_g)} - \frac{n^{-1} \sum_{i=1}^n I\{x_{g,i} = a+1\}}{\pi_{g,a+1}(\boldsymbol{\tau}_g)} \right) \phi_1(\tau_{g,a}).$$

Likewise, $\hat{\tau}_h$ satisfies

$$\sqrt{n}(\hat{\tau}_h - \tau_{h,0}) = \left(\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \right)^{-1} \mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \sqrt{n} \left(\mathbf{P}^{(gh)} - \boldsymbol{\pi}_0^{(gh)} \right)^T \mathbf{1}_h + o_P(1). \quad (23)$$

For the asymptotic normality of the polychoric correlation coefficient estimator between two variables, we often suppress the index g and h for notational simplicity if it does not cause confusion. From section 2.3, it can be easily seen that $\ell(\boldsymbol{\tau}, \rho)$ has continuous third-order partial derivatives with respect to $(\boldsymbol{\tau}, \rho)$ on the open set $O^{(\tau)} \times O^{(\rho)}$. Hence, the Taylor's theorem yields

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \frac{\partial \ell(\hat{\boldsymbol{\tau}}, \hat{\rho})}{\partial \rho} \\ &= \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} + \frac{1}{2n} \frac{\partial^3 \ell(\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho^3} (\hat{\rho} - \rho_0) + \frac{1}{n} \frac{\partial^3 \ell(\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho^2 \partial \boldsymbol{\tau}^T} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \right] \sqrt{n} (\hat{\rho} - \rho_0) \\ &\quad + \frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} + \frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \sqrt{n} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + \frac{1}{2\sqrt{n}} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)^T \frac{\partial^2 \ell(\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0), \end{aligned}$$

where $\tilde{\rho}$ lies between $\hat{\rho}$ and ρ_0 , and $\tilde{\boldsymbol{\tau}}$ lies between $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}_0$. Consequently,

$$\sqrt{n}(\hat{\rho} - \rho_0) = - \frac{\frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} + \frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \sqrt{n} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + \frac{1}{2\sqrt{n}} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)^T \frac{\partial^2 \ell(\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)}{\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} + \frac{1}{2n} \frac{\partial^3 \ell(\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho^3} (\hat{\rho} - \rho_0) + \frac{1}{n} \frac{\partial^3 \ell(\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho^2 \partial \boldsymbol{\tau}^T} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)}, \quad (24)$$

provided that the denominator is nonzero. We will show later that the denominator converges in probability to a nonzero constant. Hence, with probability approaching 1, the denominator is nonzero.

We first consider the numerator of (24). From equations (22) and (23), we have

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} + \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] \sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \\
&= \frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} + \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] \left(\begin{aligned} & \left(\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} \right)^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0) \mathbf{1}_g + o_P(1) \\ & \left(\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \right)^{-1} \mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0)^T \mathbf{1}_h + o_P(1) \end{aligned} \right) \\
&= \sqrt{n} \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} P_{ab} \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} + \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right] \left(\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} \right)^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0) \mathbf{1}_g \\
&\quad + \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right] \left(\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \right)^{-1} \mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0)^T \mathbf{1}_h + \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] o_P(1) \\
&= \sqrt{n} \text{tr} \{ \mathbf{A}^T \mathbf{P} \} + \text{tr} \left\{ \mathbf{1}_g \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right] \left(\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} \right)^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0) \right\} \\
&\quad + \text{tr} \left\{ \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \left(\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \right)^{-1} \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right] \mathbf{1}_h^T \sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0) \right\} \\
&\quad + \left[\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] o_P(1), \tag{25}
\end{aligned}$$

where P_{ab} is the (a, b) th entry of $\mathbf{P}$ and the (a, b) th entry of $\mathbf{A}$ is $\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) / \partial \rho$. Note that

$$\text{tr} \{ \mathbf{A}^T \boldsymbol{\pi}_0 \} = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \pi_{ab,0} = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \frac{\partial \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} = 0.$$

Then, we can express $\sqrt{n} \text{tr} \{ \mathbf{A}^T \mathbf{P} \}$ in (25) as $\sqrt{n} \text{tr} \{ \mathbf{A}^T (\mathbf{P} - \boldsymbol{\pi}_0) \}$. Further, by Result 1,

$\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) / \partial \rho \partial \boldsymbol{\tau}$ is bounded on $O^{(\tau)} \times O^{(\rho)}$. Hence

$$\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}^T} = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} P_{ab} \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}^T}$$

is also bounded. Consequently

$$\left[\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] o_P(1) = o_P(1).$$

As a result, (25) becomes

$$\frac{1}{\sqrt{n}} \frac{\partial \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho} + \left[\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] \sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) = \sqrt{n} \text{tr} \{ \boldsymbol{\Lambda} (\mathbf{P} - \boldsymbol{\pi}_0) \} + o_P(1), \tag{26}$$

with $\text{tr} \{ \boldsymbol{\Lambda} \boldsymbol{\pi}_0 \} = \mathbf{0}$, where

$$\begin{aligned}
\boldsymbol{\Lambda} &= \mathbf{A}^T + \mathbf{1}_g \left[\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right] \left(\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} \right)^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \\
&\quad + \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \left(\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} \right)^{-1} \left[\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right] \mathbf{1}_h^T.
\end{aligned}$$

By LLN,

$$\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \xrightarrow{p} 0, \quad (27)$$

if and only if

$$\begin{aligned} \sum_{i=1}^n P \left(\left| \frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right| > n\epsilon \right) &\rightarrow 0, \\ \sum_{i=1}^n E_{true} \left\{ \frac{\frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)}{n} I \left(\left| \frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right| \leq n\kappa \right) \right\} &\rightarrow 0, \\ \sum_{i=1}^n \text{var}_{true} \left\{ \frac{\frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)}{n} I \left(\left| \frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right| \leq n\kappa \right) \right\} &\rightarrow 0, \end{aligned}$$

for every $\epsilon > 0$ and some $\kappa > 0$; and the expectations and variances are finite. From equation (9), we have

$$\begin{aligned} E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) &= \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_{true}) \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \\ &= E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \\ &\quad + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \delta^* / \sqrt{n} \\ &\quad + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} R_{ab}(\delta^*), \end{aligned} \quad (28)$$

$$\begin{aligned} \text{var}_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) &= E_{true} \left[\left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2 \right] - \left[E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right]^2 \\ &= E_0 \left[\left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2 \right] - \left[E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right]^2 \\ &\quad + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \left(\frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2 \delta^* / \sqrt{n} \\ &\quad + \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \left(\frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2 R_{ab}(\delta^*). \end{aligned} \quad (29)$$

By Result 1, the first- and second-order partial derivatives of $\log \pi_{ab}(\boldsymbol{\tau}, \rho)$ are bounded on

$O^{(\tau)} \times O^{(\rho)}$. Hence,

$$\begin{aligned} E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) &= \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2}, \\ E_0 \left[\left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2 \right] &= \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \left[\frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right]^2, \\ \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2}, \\ \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \left(\frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2, \end{aligned}$$

are bounded, so as the variance $\text{var}_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)$. Further, since $R_{ab}(\delta^*) = O(n^{-1})$ in (9),

$$\begin{aligned} \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} R_{ab}(\delta^*) &= O(n^{-1}), \\ \sum_{a,b} \pi_{ab}(\boldsymbol{\tau}_0, \rho_0) \left(\frac{\partial^2 \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)^2 R_{ab}(\delta^*) &= O(n^{-1}). \end{aligned}$$

Therefore,

$$\begin{aligned} E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) &= E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) + O(n^{-1/2}), \\ \text{var}_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) &= \text{var}_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) + O(n^{-1/2}), \end{aligned}$$

and they are both finite. Since both $\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0) / \partial \rho^2$ and $E_{true} [\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0) / \partial \rho^2]$ are finite and bounded,

$$\left| \frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right| < n\epsilon$$

for a sufficiently large n . Consequently, (27) holds since

$$\begin{aligned} \sum_{i=1}^n E_{true} \left\{ \frac{\frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)}{n} \right\} &= 0, \\ \sum_{i=1}^n \text{var}_{true} \left\{ \frac{\frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)}{n} \right\} &= \frac{1}{n} \text{var}_{true} \left(\frac{\partial^2 \ell_i(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho^2} \right) \rightarrow 0. \end{aligned}$$

Likewise, similar arguments yields

$$\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}} - E_{true} \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}} \right) \xrightarrow{p} \mathbf{0}. \quad (30)$$

Consequently, $\mathbf{\Lambda}$ converges in probability to

$$\begin{aligned} \mathbf{\Gamma} = & \mathbf{A}^T + \mathbf{1}_g E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right) (\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0})^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \\ & + \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} (\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0})^{-1} E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right) \mathbf{1}_h^T, \end{aligned} \quad (31)$$

which coincides with equation (16) in Jöreskog (1994), as if the local parameter is zero.

Further, by the Slutsky's theorem (Theorem 2.13, Jiang, 2010),

$$\begin{aligned} \sqrt{n} \text{tr} \{ \mathbf{\Lambda} (\mathbf{P} - \boldsymbol{\pi}_0) \} &= \sqrt{n} \text{tr} \{ \mathbf{\Gamma} (\mathbf{P} - \boldsymbol{\pi}_0) \} + \sqrt{n} \text{tr} \{ (\mathbf{\Lambda} - \mathbf{\Gamma}) (\mathbf{P} - \boldsymbol{\pi}_0) \} \\ &= \sqrt{n} \text{tr} \{ \mathbf{\Gamma} (\mathbf{P} - \boldsymbol{\pi}_0) \} + o_P(1), \end{aligned}$$

if $\sqrt{n}(\mathbf{P} - \boldsymbol{\pi}_0)$ converges in distribution to some random variable. Consequently, equation (26) becomes

$$\frac{1}{\sqrt{n}} \frac{\partial \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho} + \left[\frac{1}{n} \frac{\partial^2 \ell(\rho_0, \boldsymbol{\tau}_0)}{\partial \rho \partial \boldsymbol{\tau}^T} \right] \sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) = \sqrt{n} \text{tr} \{ \mathbf{\Gamma} (\mathbf{P} - \boldsymbol{\pi}_0) \} + o_P(1). \quad (32)$$

We now consider the last term in the numerator and the last term in the denominator of (24). Let the operator $\text{vec}(\cdot)$ vectorize the enclosed matrix by stacking columns on top of each other. In order to derive the asymptotic distribution of $\sqrt{n} \text{vec}(\mathbf{P} - \boldsymbol{\pi}_0)$, we assume that a random vector $\mathbf{z}_n$ follows a multinomial distribution with one trial and probabilities $\text{vec}(\boldsymbol{\pi}_{true})$ with $\boldsymbol{\pi}_{true} = \boldsymbol{\pi}(\boldsymbol{\tau}_0, \rho_{true})$, and that the observed relative frequency is $\text{vec}(\mathbf{P})$. Here the subscript n is used to emphasize that the distribution of $\mathbf{z}_n$ depends on n . We also define $\mathbf{z}_0$ as a multinomial distribution with probabilities $\text{vec}(\boldsymbol{\pi}_0)$ with $\boldsymbol{\pi}_0 = \boldsymbol{\pi}(\boldsymbol{\tau}_0, \rho_0)$. Because of (9), we have

$$E_{true}(\mathbf{z}_n) = \text{vec}(\boldsymbol{\pi}_{true}) = E_0(\mathbf{z}_0) + \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \delta^* / \sqrt{n} + O(n^{-1}). \quad (33)$$

Since $\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\} / \partial \rho$ is bounded by Result 1, $E_{true}(\mathbf{z}_n) = E_0(\mathbf{z}_0) + O(n^{-1/2})$. Note that

$$\begin{aligned} \text{var}_{true}(\mathbf{z}_n) &= E_{true}(\mathbf{z}_n \mathbf{z}_n^T) - E_{true}(\mathbf{z}_n) E_{true}(\mathbf{z}_n^T) \\ &= \text{diag} \{ \text{vec}(\boldsymbol{\pi}_{true}) \} - \text{vec}(\boldsymbol{\pi}_{true}) \text{vec}^T(\boldsymbol{\pi}_{true}). \end{aligned}$$

By (33), we get

$$\begin{aligned}
\text{var}_{true}(\mathbf{z}_n) &= \text{diag}\{\text{vec}(\boldsymbol{\pi}_{true})\} - \text{vec}(\boldsymbol{\pi}_{true}) \text{vec}^T(\boldsymbol{\pi}_{true}) \\
&= \text{diag}\{E_0(\mathbf{z}_0) + O(n^{-1/2})\} - [E_0(\mathbf{z}_0) + O(n^{-1/2})] [E_0(\mathbf{z}_0) + O(n^{-1/2})]^T \\
&= \text{var}_0(\mathbf{z}_0) + O(n^{-1/2}).
\end{aligned} \tag{34}$$

If we observe $x_g = a$ and $x_h = b$,

$$\left\| \frac{\mathbf{z}_n - E_{true}(\mathbf{z}_n)}{\sqrt{n}} \right\|^2 = \frac{\|\text{vec}(\boldsymbol{\pi}_{true})\|^2 - \pi_{ab}^2(\boldsymbol{\tau}_0, \rho_{true}) + (1 - \pi_{ab}(\boldsymbol{\tau}_0, \rho_{true}))^2}{n} \leq \frac{2}{n}.$$

Hence, the Lindeberg condition holds, since, for every $\epsilon > 0$,

$$\left\| \frac{\mathbf{z}_{n,i} - E_{true}(\mathbf{z}_{n,i})}{\sqrt{n}} \right\|^2 < \epsilon^2$$

and

$$\sum_{i=1}^n E \left[\left\| \frac{\mathbf{z}_{n,i} - E_{true}(\mathbf{z}_{n,i})}{\sqrt{n}} \right\|^2 I \left\{ \left\| \frac{\mathbf{z}_{n,i} - E_{true}(\mathbf{z}_{n,i})}{\sqrt{n}} \right\|^2 \geq \epsilon^2 \right\} \right] = 0,$$

for a sufficiently large n . The diagonal and off-diagonal entries in the covariance matrix of $\mathbf{z}_{n,i}$ are $\pi_a(\boldsymbol{\tau}_0, \rho_{true})[1 - \pi_a(\boldsymbol{\tau}_0, \rho_{true})]$ and $-\pi_a(\boldsymbol{\tau}_0, \rho_{true})\pi_b(\boldsymbol{\tau}_0, \rho_{true})$, respectively. Note that

$$\sum_{i=1}^n \text{var}_{true} \left(\frac{\mathbf{z}_{n,i} - E_{true}(\mathbf{z}_{n,i})}{\sqrt{n}} \right) = \text{var}_{true}(\mathbf{z}_{n,i}) \rightarrow \text{var}_0(\mathbf{z}_0)$$

by (34). Therefore, the central limit theorem implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{z}_{n,i} - E_{true}(\mathbf{z}_{n,i})] \xrightarrow{d} N(\mathbf{0}, \text{var}_0(\mathbf{z}_0)).$$

Since $n^{-1} \sum_{i=1}^n \mathbf{z}_{n,i} = \text{vec}(\mathbf{P})$, we have

$$\sqrt{n} \left(\text{vec}(\mathbf{P}) - \text{vec}(\boldsymbol{\pi}_0) - \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \delta^* / \sqrt{n} \right) \xrightarrow{d} N(\mathbf{0}, \text{var}_0(\mathbf{z}_0)). \tag{35}$$

Equivalently,

$$\sqrt{n} (\text{vec}(\mathbf{P}) - \text{vec}(\boldsymbol{\pi}_0)) \xrightarrow{d} N \left(\text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \delta^*, \text{var}_0(\mathbf{z}_0) \right). \tag{36}$$

Hence, the Slutsky's theorem (Theorem 2.13, Jiang, 2010) leads to

$$\begin{aligned} \sqrt{n} \text{tr} \{ \mathbf{\Gamma} (\mathbf{P} - \boldsymbol{\pi}_0) \} &= \text{vec}^T (\mathbf{\Gamma}) \sqrt{n} \text{vec} (\mathbf{P} - \boldsymbol{\pi}_0) \\ &\xrightarrow{d} N \left(\text{vec}^T (\mathbf{\Gamma}) \text{diag} (\text{vec} (\boldsymbol{\pi}_0)) \frac{\partial \log \{ \text{vec} (\boldsymbol{\pi}_0) \}}{\partial \rho} \delta^*, \text{vec}^T (\mathbf{\Gamma}) \text{var}_0 (\mathbf{z}_0) \text{vec} (\mathbf{\Gamma}) \right). \end{aligned} \quad (37)$$

Note that

$$\begin{aligned} \frac{\partial^3 \ell_i (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} &= \sum_{a=1}^{m_x} \sum_{b=1}^{m_y} I_i (a, b) \left[\frac{1}{\pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})} \frac{\partial^3 \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \right. \\ &\quad - \frac{1}{\pi_{ab}^2 (\tilde{\boldsymbol{\tau}}, \tilde{\rho})} \frac{\partial^2 \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau}} \frac{\partial \log \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \boldsymbol{\tau}^T} \\ &\quad \left. - \frac{\partial^2 \log \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \frac{\partial \log \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho} - \frac{\partial \log \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \boldsymbol{\tau}} \frac{\partial^2 \log \pi_{ab} (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau}^T} \right], \end{aligned}$$

where all partial derivatives are finite and bounded on $O^{(\tau)} \times O^{(\rho)}$ by Result 1. Then, $n^{-1} \partial^3 \ell (\tilde{\boldsymbol{\tau}}, \tilde{\rho}) / \partial \rho \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T$ is bounded in probability. Likewise, $n^{-1} \partial^3 \ell (\tilde{\boldsymbol{\tau}}, \tilde{\rho}) / \partial \rho^3$ and $n^{-1} \partial^3 \ell (\tilde{\boldsymbol{\tau}}, \tilde{\rho}) / \partial \rho^2 \partial \boldsymbol{\tau}$ are also bounded in probability. Consequently,

$$\sqrt{n} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)^T \left[\frac{1}{n} \frac{\partial^3 \ell (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho \partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^T} \right] (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \xrightarrow{p} 0, \quad (38)$$

$$\frac{1}{n} \frac{\partial^2 \ell (\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} + \frac{1}{2n} \frac{\partial^3 \ell (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho^3} (\hat{\rho} - \rho_0) + \frac{1}{n} \frac{\partial^3 \ell (\tilde{\boldsymbol{\tau}}, \tilde{\rho})}{\partial \rho^2 \partial \boldsymbol{\tau}^T} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \xrightarrow{p} E_0 \left(\frac{\partial^2 \ell_i (\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \quad (39)$$

since $\hat{\rho} - \rho_0 = o_P(1)$ by Theorem 1, $\sqrt{n} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) = O_P(1)$ from (22) and (23), and (27) holds. By equation (39) and $E_0 \left(\frac{\partial^2 \ell_i (\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \neq 0$ under Assumption A4, the denominator of (24) is nonzero with probability approaching 1. By equations (32) and (38), the numerator of (24) becomes

$$\sqrt{n} \text{tr} \{ \mathbf{\Gamma} (\mathbf{P} - \boldsymbol{\pi}_0) \} + o_P(1),$$

which converges in distribution to the normal distribution (37) by the Slutsky's theorem.

The denominator of (24) converges in probability to a constant by (39). Hence, by the corollary of the Slutsky's theorem in Ferguson (1996, P40), we obtain

$$\sqrt{n} (\hat{\rho} - \rho_0) = \frac{1}{-E_0 \left(\frac{\partial^2 \ell_i (\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)} \sqrt{n} \text{tr} \{ \mathbf{\Gamma} (\mathbf{P} - \boldsymbol{\pi}_0) \} + o_P(1).$$

Therefore,

$$\sqrt{n}(\hat{\rho} - \rho_0) \xrightarrow{d} N \left(\frac{\text{vec}^T(\Gamma) \text{diag}(\text{vec}(\boldsymbol{\pi}_0))}{-E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right)} \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \delta^*, \frac{\text{vec}^T(\Gamma) \text{var}_0(\mathbf{z}_0) \text{vec}(\Gamma)}{\left[E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right]^2} \right),$$

the asymptotic covariance matrix of which is in line with the result in Jöreskog (1994) for the standard asymptotic framework. Note that

$$\begin{aligned} & \text{vec}^T \left[\mathbf{1}_g E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right) (\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0})^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \right] \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \\ &= \text{vec}^T \left[\mathbf{1}_g E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right) (\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0})^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \right] \frac{\partial \text{vec}(\boldsymbol{\pi}_0)}{\partial \rho} \\ &= \text{tr} \left\{ \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0} (\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0})^{-1} E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right) \mathbf{1}_g^T \frac{\partial \boldsymbol{\pi}_0}{\partial \rho} \right\}, \end{aligned} \quad (40)$$

and

$$\begin{aligned} & \text{vec}^T \left[\mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} (\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0})^{-1} E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right) \mathbf{1}_h^T \right] \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \\ &= \text{vec}^T \left[\mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} (\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0})^{-1} E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right) \mathbf{1}_h^T \right] \frac{\partial \text{vec}(\boldsymbol{\pi}_0)}{\partial \rho} \\ &= \text{tr} \left\{ \mathbf{1}_h E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right) (\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0})^{-1} \mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \frac{\partial \boldsymbol{\pi}_0}{\partial \rho} \right\} \\ &= \text{tr} \left\{ E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right) (\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0})^{-1} \mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \frac{\partial \boldsymbol{\pi}_0}{\partial \rho} \mathbf{1}_h \right\}. \end{aligned} \quad (41)$$

Since

$$\mathbf{1}_g^T \frac{\partial \boldsymbol{\pi}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} = \mathbf{0} \quad \text{and} \quad \frac{\partial \boldsymbol{\pi}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \mathbf{1}_h = \mathbf{0},$$

we can see that (40) and (41) are both 0. Further,

$$\begin{aligned} \text{vec}^T(\mathbf{A}^T) \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} &= \text{vec}^T(\mathbf{A}^T) \frac{\partial \text{vec}(\boldsymbol{\pi}_0)}{\partial \rho} \\ &= \text{tr} \left\{ \frac{\partial \log(\boldsymbol{\pi}_0^T)}{\partial \rho} \frac{\partial \boldsymbol{\pi}_0}{\partial \rho} \right\} \\ &= -E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right), \end{aligned}$$

since

$$\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} = \sum_{a,b} \left[\frac{I_i(a,b)}{\pi_{ab}(\boldsymbol{\tau}_0, \rho_0)} \frac{\partial^2 \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} - \frac{I_i(a,b)}{\pi_{ab}^2(\boldsymbol{\tau}_0, \rho_0)} \left(\frac{\partial \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \right)^2 \right]$$

and

$$E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) = - \sum_{a,b} \left[\frac{1}{\pi_{ab}(\boldsymbol{\tau}_0, \rho_0)} \left(\frac{\partial \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \right)^2 \right] = - \sum_{a,b} \left[\frac{\partial \log \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \frac{\partial \pi_{ab}(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho} \right].$$

Therefore, using the expression of $\boldsymbol{\Gamma}$ given by (31), we get

$$\begin{aligned} & \text{vec}^T(\boldsymbol{\Gamma}) \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \\ = & \text{vec}^T(\mathbf{A}^T) \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \\ & + \text{vec}^T \left(\mathbf{1}_g E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_g^T} \right) (\mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \mathbf{B}_{g,0})^{-1} \mathbf{B}_{g,0}^T \mathbf{D}_{g,0}^{-1} \right) \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \\ & + \text{vec}^T \left(\mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0} (\mathbf{B}_{h,0}^T \mathbf{D}_{h,0}^{-1} \mathbf{B}_{h,0})^{-1} E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho \partial \boldsymbol{\tau}_h^T} \right) \mathbf{1}_h^T \right) \text{diag}(\text{vec}(\boldsymbol{\pi}_0)) \frac{\partial \log \{\text{vec}(\boldsymbol{\pi}_0)\}}{\partial \rho} \\ = & -E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right), \end{aligned}$$

which means that the mean of the asymptotic distribution of $\sqrt{n}(\hat{\rho} - \rho_0)$ is equivalent to

δ^* . In other words,

$$\sqrt{n}(\hat{\rho} - \rho_{true}) \xrightarrow{d} N \left(0, \frac{\text{vec}^T(\boldsymbol{\Gamma}) \text{var}_0(\mathbf{z}_0) \text{vec}(\boldsymbol{\Gamma})}{\left[E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0, \rho_0)}{\partial \rho^2} \right) \right]^2} \right).$$

The proof for the polychoric correlation coefficient estimator between two variables is completed.

Regarding the asymptotic distribution of more than one polychoric correlation estimators, the proof by and large follows from the one estimator case. Our proof above implies that

$$\sqrt{n}(\hat{\rho}_{gh} - \rho_{gh,0}) = \frac{1}{-E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_{gh,0}, \rho_{gh,0})}{\partial \rho^2} \right)} \sqrt{n} \text{tr} \left\{ \boldsymbol{\Gamma}^{(gh)} (\mathbf{P}^{(gh)} - \boldsymbol{\pi}_0^{(gh)}) \right\} + o_P(1) \quad (42)$$

holds for any pair of g th indicator and h th indicator, provided that all regularity conditions

hold. Here indices are added to indicate the pair of indicators. Thus,

$$\begin{aligned}
 \begin{bmatrix} \sqrt{n} (\hat{\rho}_{12} - \rho_{12,0}) \\ \vdots \\ \sqrt{n} (\hat{\rho}_{q-1,q} - \rho_{q-1,q,0}) \end{bmatrix} &= \begin{bmatrix} \frac{1}{-E_0 \left(\frac{\partial^2 \ell_i(\tau_{12,0}, \rho_{12,0})}{\partial \rho^2} \right)} \sqrt{n} \text{tr} \left\{ \mathbf{\Gamma}^{(12)} \left(\mathbf{P}^{(12)} - \boldsymbol{\pi}_0^{(12)} \right) \right\} \\ \vdots \\ \frac{1}{-E_0 \left(\frac{\partial^2 \ell_i(\tau_{q-1,q,0}, \rho_{q-1,q,0})}{\partial \rho^2} \right)} \sqrt{n} \text{tr} \left\{ \mathbf{\Gamma}^{(q-1,q)} \left(\mathbf{P}^{(q-1,q)} - \boldsymbol{\pi}_0^{(q-1,q)} \right) \right\} \end{bmatrix} + o_P(1) \\
 &= -\sqrt{n} \mathbf{H}^{-1} \mathbf{E} \begin{bmatrix} \text{vec} \left(\mathbf{P}^{(12)} - \boldsymbol{\pi}_0^{(12)} \right) \\ \text{vec} \left(\mathbf{P}^{(13)} - \boldsymbol{\pi}_0^{(13)} \right) \\ \vdots \\ \text{vec} \left(\mathbf{P}^{(q-1,q)} - \boldsymbol{\pi}_0^{(q-1,q)} \right) \end{bmatrix} + o_P(1), \tag{43}
 \end{aligned}$$

where

$$\begin{aligned}
 \mathbf{H} &= \text{diag} \left(E_0 \left(\frac{\partial^2 \ell_i(\tau_{12,0}, \rho_{12,0})}{\partial \rho^2} \right) \quad \dots \quad E_0 \left(\frac{\partial^2 \ell_i(\tau_{q-1,q,0}, \rho_{q-1,q,0})}{\partial \rho^2} \right) \right), \\
 \mathbf{E} &= \begin{bmatrix} \text{vec}^T \left(\mathbf{\Gamma}^{(12)} \right) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \text{vec}^T \left(\mathbf{\Gamma}^{(13)} \right) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \text{vec}^T \left(\mathbf{\Gamma}^{(q-1,q)} \right) \end{bmatrix}.
 \end{aligned}$$

Hence, the asymptotic distribution of the vector of all polychoric correlation estimators $\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_{true})$ depends on the joint distribution of

$$\mathbf{p} = \left(\text{vec}^T \left(\mathbf{P}^{(12)} - \boldsymbol{\pi}_0^{(12)} \right) \quad \text{vec}^T \left(\mathbf{P}^{(13)} - \boldsymbol{\pi}_0^{(13)} \right) \quad \dots \quad \text{vec}^T \left(\mathbf{P}^{(q-1,q)} - \boldsymbol{\pi}_0^{(q-1,q)} \right) \right)^T,$$

where $\mathbf{p}$ is a $\sum_{i=1}^q \sum_{j=i+1}^q m_i m_j \times 1$ vector. In the univariate case, we use $\boldsymbol{\pi}_{true}$ to denote $\boldsymbol{\pi}_{true}^{(gh)}$ and $\boldsymbol{\pi}_0$ to denote $\boldsymbol{\pi}_0^{(gh)}$. In the multivariate case, we redefine

$$\boldsymbol{\pi}_{true} = \left(\text{vec}^T \left(\boldsymbol{\pi}_{true}^{(12)} \right) \quad \text{vec}^T \left(\boldsymbol{\pi}_{true}^{(13)} \right) \quad \dots \quad \text{vec}^T \left(\boldsymbol{\pi}_{true}^{(q-1,q)} \right) \right)^T, \tag{44}$$

$$\boldsymbol{\pi}_0 = \left(\text{vec}^T \left(\boldsymbol{\pi}_0^{(12)} \right) \quad \text{vec}^T \left(\boldsymbol{\pi}_0^{(13)} \right) \quad \dots \quad \text{vec}^T \left(\boldsymbol{\pi}_0^{(q-1,q)} \right) \right)^T, \tag{45}$$

as a generalization of $\text{vec} \left(\boldsymbol{\pi}_{true}^{(gh)} \right)$ and $\text{vec} \left(\boldsymbol{\pi}_0^{(gh)} \right)$. Using the newly defined $\boldsymbol{\pi}_0$, we will first generalize (33) and (34) to multivariate cases. Let $\mathbf{y}_n^{(gh)}$ be the $m_g m_h \times 1$ vector with

entries $I \{x_g = a \text{ and } x_h = b\}$ for all possible combinations of a , and b . Let $\mathbf{y}_n$ be the $\sum_{i=1}^q \sum_{j=i+1}^q m_i m_j \times 1$ vector that stacks all $\mathbf{y}_n^{(gh)}$ on top of another for all possible pairs (g, h) . Without loss of generality, we assume that the entries in $\mathbf{y}_n$ are sorted such that $E_{true}(\mathbf{y}_n) = \boldsymbol{\pi}_{true}$. We also define $\mathbf{y}_0^{(gh)}$ and $\mathbf{y}_0$ as analogues to $\mathbf{y}_n^{(gh)}$ and $\mathbf{y}_n$, respectively, but $E_0(\mathbf{y}_0^{(gh)}) = \text{vec}(\boldsymbol{\pi}_0^{(gh)})$ and $E_0(\mathbf{y}_0) = \boldsymbol{\pi}_0$. $\mathbf{y}_n$ and $\mathbf{y}_0$ are simply the generalizations of $\mathbf{z}_n$ and $\mathbf{z}_0$ from the univariate case to the multivariate case. By stacking (33) for all pairs (g, h) from the univariate case, we get

$$E_{true}(\mathbf{y}_n) = E_0(\mathbf{y}_0) + \begin{pmatrix} \pi_{11,0}^{(12)} \frac{\partial \log \pi_{11}^{(12)}(\boldsymbol{\tau}_{12,0}, \boldsymbol{\rho}_{12,0})}{\partial \rho_{1,2}} \delta_{12}^* \\ \vdots \\ \pi_{m_1 m_2, 0}^{(12)} \frac{\partial \log \pi_{m_1 m_2}^{(12)}(\boldsymbol{\tau}_{12,0}, \boldsymbol{\rho}_{12,0})}{\partial \rho_{1,2}} \delta_{12}^* \\ \vdots \\ \pi_{m_{q-1} m_q, 0}^{(q-1, q)} \frac{\partial \log \pi_{m_{q-1} m_q}^{(q-1, q)}(\boldsymbol{\tau}_{q-1, q, 0}, \boldsymbol{\rho}_{q-1, q, 0})}{\partial \rho_{q-1, q}} \delta_{q-1, q}^* \end{pmatrix} / \sqrt{n} + O(n^{-1}), \quad (46)$$

where δ_{gh}^* is the local parameter in $\boldsymbol{\delta}^*$ corresponding to ρ_{gh} . As we showed in Result 1, the partial derivatives are bounded. Hence, (46) implies that $E_{true}(\mathbf{y}_n) = E_0(\mathbf{y}_0) + O(n^{-1/2})$. The entries in $\text{var}_{true}(\mathbf{y}_n)$ are either $\text{var}_{true}(\mathbf{y}_n^{(gh)})$ or $\text{cov}_{true}(\mathbf{y}_n^{(gh)}, \mathbf{y}_n^{(st)})$ for $g < h$ and $s < t$. Equation (34) is directly applicable to $\text{var}_{true}(\mathbf{y}_n^{(gh)})$, yielding $\text{var}_{true}(\mathbf{y}_n^{(gh)}) = \text{var}_0(\mathbf{y}_0^{(gh)}) + O(n^{-1/2})$. For $\text{cov}_{true}(\mathbf{y}_n^{(gh)}, \mathbf{y}_n^{(st)})$, note that the entries in $E_{true}[\mathbf{y}_n^{(gh)} (\mathbf{y}_n^{(st)})^T]$ are of the form $E_{true}[I \{x_g = a \text{ and } x_h = b\} I \{x_s = c \text{ and } x_t = d\}]$, which can be 0, or $\pi_{a_1 a_2, true}^{(g_1 g_2)}$, or $\pi_{a_1 a_2 a_3, true}^{(g_1 g_2 g_3)}$, or $\pi_{a_1 a_2 a_3 a_4, true}^{(g_1 g_2 g_3 g_4)}$. By (8), we get

$$\begin{aligned} & E_{true}[I \{x_g = a \text{ and } x_h = b\} I \{x_s = c \text{ and } x_t = d\}] \\ &= E_0[I \{x_g = a \text{ and } x_h = b\} I \{x_s = c \text{ and } x_t = d\}] + O(n^{-1/2}), \end{aligned}$$

and $E_{true}[\mathbf{y}_n^{(gh)} (\mathbf{y}_n^{(st)})^T] = E_0[\mathbf{y}_n^{(gh)} (\mathbf{y}_n^{(st)})^T] + O(n^{-1/2})$. Hence,

$$\begin{aligned} \text{cov}_{true}(\mathbf{y}_n^{(gh)}, \mathbf{y}_n^{(st)}) &= E_{true}[\mathbf{y}_n^{(gh)} (\mathbf{y}_n^{(st)})^T] - E_{true}(\mathbf{y}_n^{(gh)}) [E_{true}(\mathbf{y}_n^{(st)})]^T \\ &= E_0[\mathbf{y}_n^{(gh)} (\mathbf{y}_n^{(st)})^T] - E_0(\mathbf{y}_0^{(gh)}) [E_0(\mathbf{y}_0^{(st)})]^T + O(n^{-1/2}). \end{aligned}$$

Therefore,

$$\text{var}_{true}(\mathbf{y}_n) = \text{var}_0(\mathbf{y}_0) + O(n^{-1/2}), \quad (47)$$

which is a generalization of (34) to the multivariate case.

To apply the Lindeberg-Feller theorem, we consider $\mathbf{y}_n - E_{true}(\mathbf{y}_n)$, which has mean $\mathbf{0}$. Since $\left\| \mathbf{y}_n^{(gh)} - E_{true}(\mathbf{y}_n^{(gh)}) \right\|^2 \leq 2$ for any pair (g, h) , then

$$\left\| \frac{\mathbf{y}_{n,i} - E_{true}(\mathbf{y}_{n,i})}{\sqrt{n}} \right\|^2 \leq \frac{q(q-1)}{n}.$$

Hence, the Lindeberg condition holds since, for every $\epsilon > 0$, $\|\mathbf{y}_{n,i} - E_{true}(\mathbf{y}_{n,i})\|^2 < n\epsilon^2$ and

$$\sum_{i=1}^n E \left[\left\| \frac{\mathbf{y}_{n,i} - E_{true}(\mathbf{y}_{n,i})}{\sqrt{n}} \right\|^2 I \left\{ \left\| \frac{\mathbf{y}_{n,i} - E_{true}(\mathbf{y}_{n,i})}{\sqrt{n}} \right\| \geq \epsilon \right\} \right] = 0$$

for a sufficiently large n . Note that

$$\sum_{i=1}^n \text{var}_{true} \left(\frac{\mathbf{y}_{n,i} - E_{true}(\mathbf{y}_{n,i})}{\sqrt{n}} \right) = \text{var}_{true}(\mathbf{y}_{n,i}) \rightarrow \text{var}_0(\mathbf{y}_0)$$

by (47). Therefore, the central limit theorem implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{y}_{n,i} - E_{true}(\mathbf{y}_{n,i})] \xrightarrow{d} N(\mathbf{0}, \text{var}_0(\mathbf{y}_0)).$$

By (46), we further get

$$\sqrt{n} [\mathbf{p} - E_0(\mathbf{y}_0)] \xrightarrow{d} N \left(\begin{pmatrix} \pi_{11,0}^{(12)} \frac{\partial \log \pi_{11}^{(12)}(\boldsymbol{\tau}_{12,0}, \boldsymbol{\rho}_{12,0})}{\partial \rho_{1,2}} \delta_{12}^* \\ \vdots \\ \pi_{m_1 m_2, 0}^{(12)} \frac{\partial \log \pi_{m_1 m_2}^{(12)}(\boldsymbol{\tau}_{12,0}, \boldsymbol{\rho}_{12,0})}{\partial \rho_{1,2}} \delta_{12}^* \\ \vdots \\ \pi_{m_{q-1} m_q, 0}^{(q-1,q)} \frac{\partial \log \pi_{m_{q-1} m_q}^{(q-1,q)}(\boldsymbol{\tau}_{q-1,q,0}, \boldsymbol{\rho}_{q-1,q,0})}{\partial \rho_{q-1,q}} \delta_{q-1,q}^* \end{pmatrix}, \text{var}_0(\mathbf{y}_0) \right),$$

where the asymptotic covariance matrix $\text{var}_0(\mathbf{y}_0)$ is evaluated at $(\boldsymbol{\tau}_0, \boldsymbol{\rho}_0)$, rather than $(\boldsymbol{\tau}_0, \boldsymbol{\rho}_{true})$. Hence, (43) implies that

$$\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0) \xrightarrow{d} N \left(\mathbf{H}_0^{-1} \mathbf{E}_0 \begin{pmatrix} \pi_{11,0}^{(12)} \frac{\partial \log \pi_{11}^{(12)}(\boldsymbol{\tau}_{12,0}, \boldsymbol{\rho}_{12,0})}{\partial \rho_{1,2}} \delta_{12}^* \\ \vdots \\ \pi_{m_1 m_2, 0}^{(12)} \frac{\partial \log \pi_{m_1 m_2}^{(12)}(\boldsymbol{\tau}_{12,0}, \boldsymbol{\rho}_{12,0})}{\partial \rho_{1,2}} \delta_{12}^* \\ \vdots \\ \pi_{m_{q-1} m_q, 0}^{(q-1,q)} \frac{\partial \log \pi_{m_{q-1} m_q}^{(q-1,q)}(\boldsymbol{\tau}_{q-1,q,0}, \boldsymbol{\rho}_{q-1,q,0})}{\partial \rho_{q-1,q}} \delta_{q-1,q}^* \end{pmatrix}, \mathbf{H}_0^{-1} \mathbf{E}_0 \text{var}_0(\mathbf{y}_0) \mathbf{E}_0^T \mathbf{H}_0^{-1} \right).$$

As we have showed for the univariate case,

$$\frac{\text{vec}^T(\mathbf{\Gamma}^{(gh)}) \text{diag}(\text{vec}(\boldsymbol{\pi}_0^{(gh)})) \partial \log \left\{ \text{vec} \left[\boldsymbol{\pi}^{(gh)}(\boldsymbol{\tau}_0^{(gh)}, \rho_{gh,0}) \right] \right\}}{-E_0 \left(\frac{\partial^2 \ell_i(\boldsymbol{\tau}_0^{(gh)}, \rho_{gh,0})}{\partial \rho^2} \right) \partial \rho} \delta_{gh}^* = \delta_{gh}^*,$$

for all g and h . Thus,

$$\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0) \xrightarrow{d} N(\boldsymbol{\delta}^*, \mathbf{H}_0^{-1} \mathbf{E}_0 \text{var}_0(\mathbf{y}_0) \mathbf{H}_0^{-1} \mathbf{E}_0^T)$$

and

$$\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_{true}) \xrightarrow{d} N(\mathbf{0}, \mathbf{H}_0^{-1} \mathbf{E}_0 \text{var}_0(\mathbf{y}_0) \mathbf{H}_0^{-1} \mathbf{E}_0^T)$$

This means that the local asymptotic framework affects asymptotic mean but the asymptotic covariance matrix remains the same as the one under the standard asymptotic framework. The proof for the multivariate case is completed. □

2.5 Simulation Study

A simulation is conducted to investigate the property of $\hat{\boldsymbol{\rho}}$ under the local asymptotic framework. Theorem 1 means that $\hat{\boldsymbol{\rho}}$ is a consistent estimator of $\boldsymbol{\rho}_0$ also under the local asymptotic framework. Theorem 2 means that the estimated asymptotic covariance matrix from standard SEM package remains valid under the local asymptotic framework. However, $\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_{true}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Upsilon})$ does not necessarily imply $\lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_{true})] = \mathbf{0}$.

Consider the function $\mathbf{h}(\mathbf{x}) = \begin{pmatrix} h_1(x_1) & \cdots & h_p(x_p) \end{pmatrix}$, where $\mathbf{x}$ is a $p \times 1$ vector,

$$h_i(x_i) = \begin{cases} x_i, & |x_i| \leq M, \\ M, & |x_i| > M, \end{cases}$$

and M is a fixed constant. It is easy to see that $\mathbf{h}(\mathbf{x})$ is a continuous and bounded function. Then, for a sufficiently large M , $\lim_{n \rightarrow \infty} E[\mathbf{h}\{\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_{true})\}]$ is approximately $\mathbf{0}$ and $\lim_{n \rightarrow \infty} E[\mathbf{h}\{\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)\}]$ is approximately $\mathbf{h}(\boldsymbol{\delta}^*)$ by Jiang (2010, pp 45).

Continuous data are generated from the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\begin{pmatrix} 1 & 0.4 + \delta^* n^{-1/2} & 0.3 + \delta^* n^{-1/2} \\ 0.4 + \delta^* n^{-1/2} & 1 & 0.35 + \delta^* n^{-1/2} \\ 0.3 + \delta^* n^{-1/2} & 0.35 + \delta^* n^{-1/2} & 1 \end{pmatrix}.$$

Hence, there are three correlation coefficients and six estimates of asymptotic covariances/variances. We consider five different δ^* values, namely, $\delta^* = 0, 0.1, 0.2, 0.3$, and 0.4 . When $\delta^* = 0$, we are simply working with the standard asymptotic framework. The population thresholds are chosen such that the probabilities of belonging to each category are 0.24, 0.41, 0.22, 0.1, and 0.03, which is the moderate asymmetry setting in Rhemtulla et al. (2012). The sample size considered here is $n = 100, 200, 400, 1000, 5000$, and $10,000$. The number of replication is 10,000. We use the `lavaan` (Rosseel, 2012) package to estimate the polychoric correlation coefficient and its asymptotic covariance/variance.

Figure 1 illustrates the estimated density of $\hat{\rho} - \rho_0$ for each pair of indicators. If Theorem 1 holds, $\hat{\rho} - \rho_0$ will be sufficiently close to 0 as n increases. It is seen that the estimated density becomes more concentrated around 0 as n increases. This implies that $P(|\hat{\rho} - \rho_0| > \epsilon)$ may be sufficiently small for a small ϵ .

As mentioned above, if Theorem 2 holds, $\lim_{n \rightarrow \infty} E[h\{\sqrt{n}(\hat{\rho} - \rho_{true})\}] \approx 0$ and $\lim_{n \rightarrow \infty} E[h\{\sqrt{n}(\hat{\rho} - \rho_{true})\}] \approx \delta^*$ for all correlation coefficients if M is sufficiently large. In our simulation, the realized $\sqrt{n}(\hat{\rho} - \rho_{true})$ is always bounded between -10 and 10 . We can let $M \geq 10$ without changing the sample mean of $\sqrt{n}(\hat{\rho} - \rho_{true})$. Figure 2 illustrates the the bias of the polychoric correlation estimators, where the bias is defined by

$$\frac{1}{R} \sum_{r=1}^R h\{\sqrt{n}(\hat{\rho}_r - \rho_{true})\} \quad \text{or} \quad \frac{1}{R} \sum_{r=1}^R h\{\sqrt{n}(\hat{\rho}_r - \rho_0)\}$$

$\hat{\rho}_r$ is the estimator at iteration r , and R is the number of replications. As expected, it is seen that the $R^{-1} \sum_{r=1}^R h\{\sqrt{n}(\hat{\rho}_r - \rho_{true})\}$ remains low, whereas $R^{-1} \sum_{r=1}^R h\{\sqrt{n}(\hat{\rho}_r - \rho_0)\}$ is approximately δ^* .

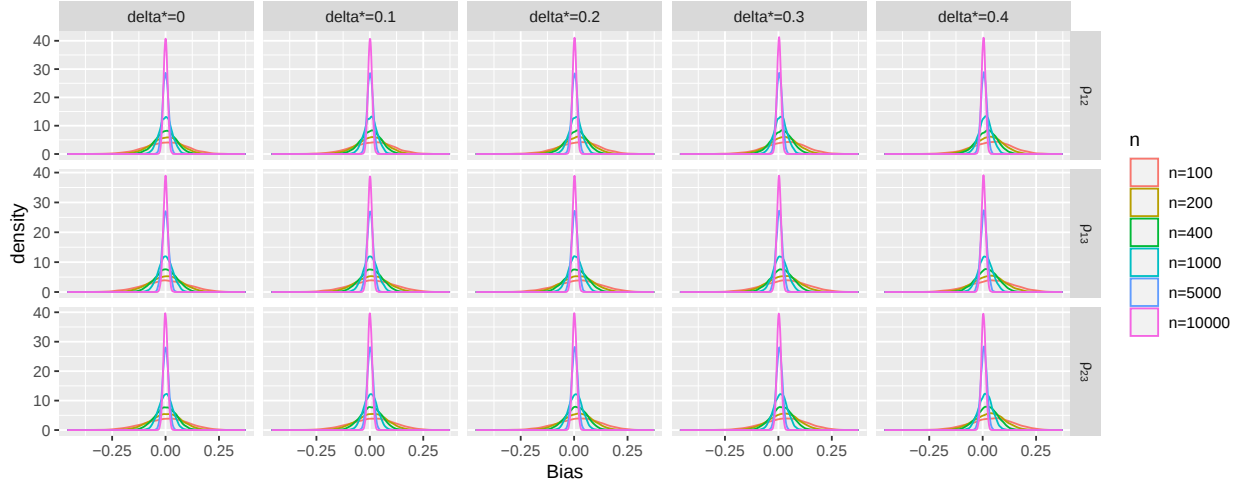


Figure 1. Estimated density of $\hat{\rho} - \rho_0$. Columns correspond to different δ^* values and rows correspond to different correlation coefficients.

Regarding the asymptotic covariance estimator, we do not know the true value. Nevertheless, if Theorem 2 holds, the estimated asymptotic covariance matrix from `lavaan` should be close to the sample covariance of the estimated correlation coefficients. Figure 3 illustrates the relative bias of asymptotic covariance/variance estimators of polychoric correlation estimators, where the relative bias is defined by

$$100 \cdot \frac{1}{R} \sum_{r=1}^R \frac{\hat{\Upsilon}_r - \tilde{\Upsilon}_0}{\tilde{\Upsilon}_0},$$

where $\hat{\Upsilon}_r$ is the estimate of Υ_0 at iteration r , and $\tilde{\Upsilon}_0$ is the sample covariance of the estimated correlation coefficients. It is seen from Figure 3 that the relative bias of the asymptotic covariance/variance estimators is generally small, as the sample size increases.

3 Frequentist Model Averaging

In the context of SEM with normally distributed continuous data, Jin and Ankargren (2019) derived the asymptotic distribution of $\sqrt{n} \begin{pmatrix} \hat{\theta}_s - \theta_0 \\ \hat{\gamma}_s - \gamma_{0,s} \end{pmatrix}$ in their equation (6) and the asymptotic distribution of $\sqrt{n}(\bar{\mu} - \mu_{true})$ in their equation (9). For the purpose of presentation, their results are placed in the following lemma.

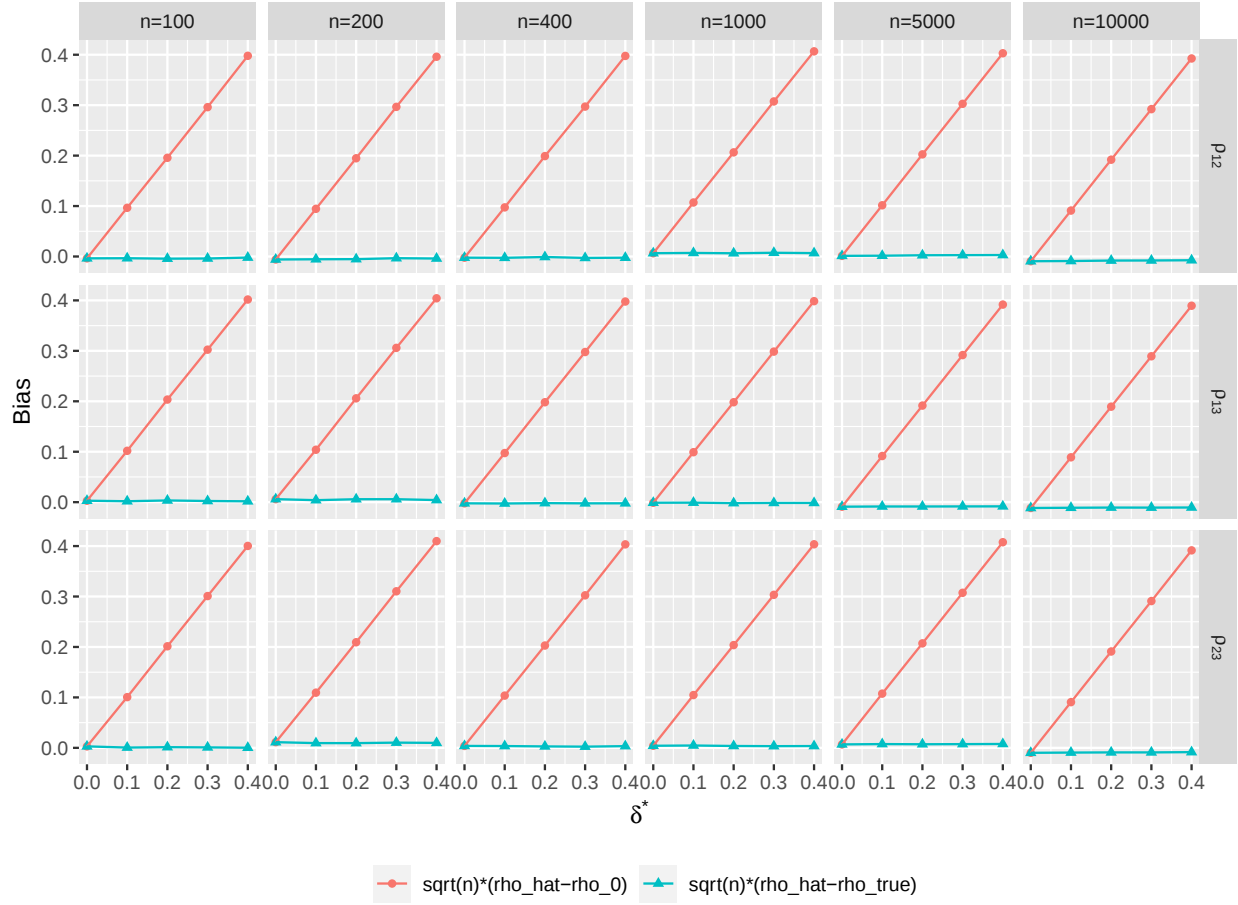


Figure 2. Bias of polychoric correlation estimators. Different colors correspond to the bias of estimating ρ_0 or ρ_{true} .

Lemma 1. Suppose that $\mathbf{J}_s$ is invertible and all third-order partial derivatives of the fit function $F(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_{0,s}, \boldsymbol{\gamma}_{0,s^c})$ in a neighborhood of $(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$ are dominated by functions with finite means. Then,

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\gamma}}_s - \boldsymbol{\gamma}_{0,s} \end{pmatrix} = \mathbf{J}_s^{-1} \left[\begin{pmatrix} \mathbf{J}_{\theta\gamma} \\ \boldsymbol{\pi}_s \mathbf{J}_{\gamma\gamma} \end{pmatrix} \boldsymbol{\delta} + \begin{pmatrix} \mathbf{M} \\ \boldsymbol{\pi}_s \mathbf{N} \end{pmatrix} \right] + O_P(n^{-1/2}). \quad (48)$$

and

$$\begin{aligned} \sqrt{n}(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true}) &= \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \sum_c c_s \sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_0) + \sum_c c_s \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\gamma}_s^T} \sqrt{n}(\hat{\boldsymbol{\gamma}}_s - \boldsymbol{\gamma}_{0,s}) - \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\gamma}^T} \boldsymbol{\delta} + O_P(n^{-1/2}) \\ &= \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \mathbf{M} + \mathbf{W} \left\{ \boldsymbol{\delta} - \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \mathbf{D} \right\} + O_P(n^{-1/2}), \end{aligned} \quad (49)$$

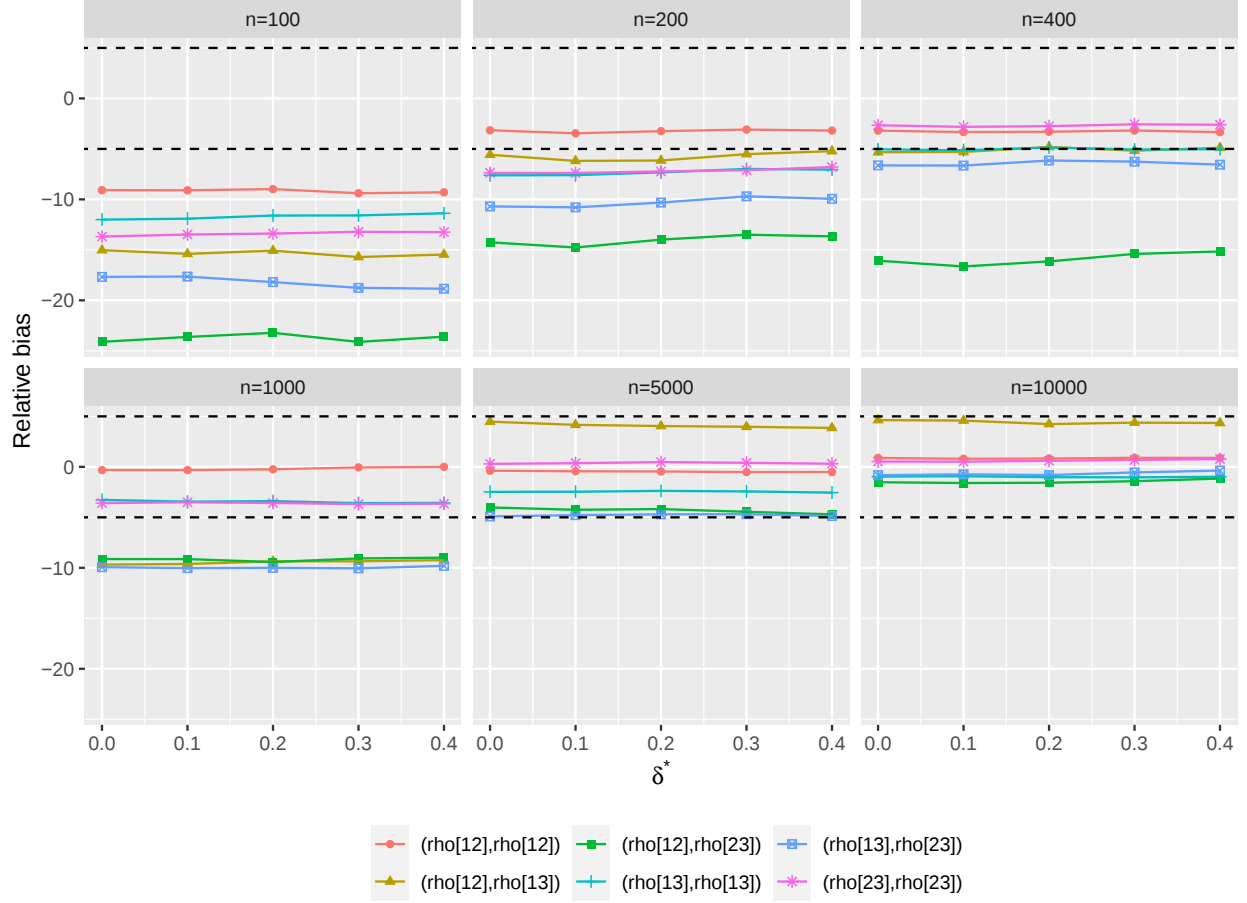


Figure 3. Relative bias of asymptotic covariance/variance estimators of polychoric correlation estimators. Different colors correspond to the asymptotic covariance between estimators of different correlation coefficients.

where $\mu_{true} = \mu(\theta_0, \gamma_0 + \delta/\sqrt{n})$, $\mu_0 = \mu(\theta_0, \gamma_0)$, $W = \frac{\partial \mu_0}{\partial \theta^T} J_{\theta\theta}^{-1} J_{\theta\gamma} - \frac{\partial \mu_0}{\partial \gamma^T}$,

$K^{-1} = J_{\gamma\gamma} - J_{\theta\gamma}^T J_{\theta\theta}^{-1} J_{\theta\gamma}$, $K_s = (\pi_s K^{-1} \pi_s^T)^{-1}$, $K^{(s)} = \pi_s^T K_s \pi_s$,

$D = \delta - K J_{\theta\gamma}^T J_{\theta\theta}^{-1} M + K N$, and

$$-\frac{1}{2\sqrt{n}} \frac{\partial F(\beta_0)}{\partial \beta} - \begin{pmatrix} J_{\theta\gamma} \\ J_{\gamma\gamma} \end{pmatrix} \delta \xrightarrow{d} \begin{pmatrix} M \\ N \end{pmatrix}.$$

(Jin and Ankargren, 2019).

The fit function F used in Jin and Ankargren (2019) is the maximum likelihood fit function that will be denoted by F_{ML} . When the observed indicators are ordinal, the least

squares fit function is commonly used, which will be denoted by F_{LS} . Since both F_{ML} and F_{LS} can be viewed as distance functions between $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$, Lemma 1 remains valid for ordinal SEM, if $F = F_{LS}$ and Assumptions A1 to A5 also hold. Hereafter, we assume all these assumptions hold.

We first derive the joint distribution of $\mathbf{M}$ and $\mathbf{N}$ under the assumptions in Theorem 2. Here $\mathbf{M}$ and $\mathbf{N}$ are defined by

$$-\frac{1}{2\sqrt{n}} \frac{\partial F_{LS}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} - \begin{pmatrix} \mathbf{J}_{\theta\gamma} \\ \mathbf{J}_{\gamma\gamma} \end{pmatrix} \boldsymbol{\delta} \xrightarrow{d} \begin{pmatrix} \mathbf{M} \\ \mathbf{N} \end{pmatrix},$$

simply replacing F by F_{LS} in all partial derivatives.

Proof of joint distribution of $\mathbf{M}$ and $\mathbf{N}$. Note that

$$\frac{\partial F_{LS}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2n \left(\frac{\partial \boldsymbol{\sigma}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right)^T \hat{\mathbf{V}} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\boldsymbol{\beta})),$$

where $\hat{\mathbf{V}}$ is a consistent estimator of $\mathbf{V}$. By Theorem 1, $\hat{\boldsymbol{\rho}}$ is a consistent estimator of $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(\boldsymbol{\beta}_0)$. Then,

$$\frac{1}{n} \frac{\partial^2 F(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \xrightarrow{p} 2\mathbf{J}_{full},$$

where

$$\mathbf{J}_{full} = \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right)^T \mathbf{V} \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T}.$$

Hence,

$$\begin{aligned} -\frac{1}{2\sqrt{n}} \frac{\partial F_{LS}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} - \begin{pmatrix} \mathbf{J}_{\theta\gamma} \\ \mathbf{J}_{\gamma\gamma} \end{pmatrix} \boldsymbol{\delta} &= \sqrt{n} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right)^T \hat{\mathbf{V}} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_0) - \begin{pmatrix} \mathbf{J}_{\theta\gamma} \\ \mathbf{J}_{\gamma\gamma} \end{pmatrix} \boldsymbol{\delta} \\ &= \begin{pmatrix} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right)^T \hat{\mathbf{V}} \sqrt{n} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_0 - \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \boldsymbol{\delta} / \sqrt{n}) \\ \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \right)^T \hat{\mathbf{V}} \sqrt{n} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_0 - \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \boldsymbol{\delta} / \sqrt{n}) \end{pmatrix} \\ &= \begin{pmatrix} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right)^T \mathbf{V} \sqrt{n} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_{true}) \\ \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \right)^T \mathbf{V} \sqrt{n} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_{true}) \end{pmatrix} + o_P(1), \end{aligned}$$

where the last equality holds from (3) and Assumption A5. By Theorem 2,

$$\begin{pmatrix} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right)^T \mathbf{V} \sqrt{n} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_{true}) \\ \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \right)^T \mathbf{V} \sqrt{n} (\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}_{true}) \end{pmatrix} \xrightarrow{d} N \left(\mathbf{0}, \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right)^T \mathbf{V} \boldsymbol{\Upsilon} \mathbf{V} \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right).$$

Hence, the joint distribution of $\mathbf{M}$ and $\mathbf{N}$ is multivariate normal with mean $\mathbf{0}$ and the covariance matrix

$$\left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right)^T \mathbf{V} \boldsymbol{\Upsilon} \mathbf{V} \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T}.$$

□

Derivation of the quadratic programming $Q(\mathbf{c})$. By (49), the mean of the limiting distribution of $\sqrt{n}(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true})$ is $\boldsymbol{\zeta} = \mathbf{W} \left\{ \mathbf{I} - \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \right\} \boldsymbol{\delta}$ and the corresponding covariance matrix is

$$\begin{aligned} \boldsymbol{\Omega} &= \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T \\ &\quad + \mathbf{W} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \text{var}(\mathbf{D}) \mathbf{K}^{-1} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{W}^T \\ &\quad - \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{cov}(\mathbf{M}, \mathbf{D}^T) \mathbf{K}^{-1} \left(\sum_s c_s \boldsymbol{\pi}_s^T \mathbf{K}_s \boldsymbol{\pi}_s \right) \mathbf{W}^T \\ &\quad - \mathbf{W} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \text{cov}(\mathbf{D}, \mathbf{M}^T) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T. \end{aligned}$$

The limit of $nE(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true})^T(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true})$ is then

$$\begin{aligned}
\text{tr} \left\{ \boldsymbol{\Omega} + \boldsymbol{\zeta} \boldsymbol{\zeta}^T \right\} &= \text{tr} \left\{ \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T \right\} \\
&\quad + \text{tr} \left\{ \mathbf{W} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \text{var}(\mathbf{D}) \mathbf{K}^{-1} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{W}^T \right\} \\
&\quad - \text{tr} \left\{ \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{cov}(\mathbf{M}, \mathbf{D}^T) \mathbf{K}^{-1} \left(\sum_s c_s \boldsymbol{\pi}_s^T \mathbf{K}_s \boldsymbol{\pi}_s \right) \mathbf{W}^T \right\} \\
&\quad - \text{tr} \left\{ \mathbf{W} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \text{cov}(\mathbf{D}, \mathbf{M}^T) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T \right\} \\
&\quad + \text{tr} \left\{ \mathbf{W} \left[\mathbf{I} - \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \right] \boldsymbol{\delta} \boldsymbol{\delta}^T \left[\mathbf{I} - \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \right]^T \mathbf{W}^T \right\} \\
&= \text{tr} \left\{ \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T \right\} \\
&\quad + \text{tr} \left\{ \mathbf{W} \left[\sum_s \sum_t c_s c_t \mathbf{K}^{(s)} \mathbf{K}^{-1} \text{var}(\mathbf{D}) \mathbf{K}^{-1} \mathbf{K}^{(t)} \right] \mathbf{W}^T \right\} \\
&\quad - 2 \text{tr} \left\{ \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{cov}(\mathbf{M}, \mathbf{D}^T) \mathbf{K}^{-1} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{W}^T \right\} \\
&\quad + \text{tr} \left\{ \mathbf{W} \left[\boldsymbol{\delta} \boldsymbol{\delta}^T - 2 \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \right] \mathbf{W}^T \right\} \\
&\quad + \text{tr} \left\{ \mathbf{W} \left(\sum_s c_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathbf{K}^{-1} \left(\sum_t c_t \mathbf{K}^{(t)} \right) \mathbf{W}^T \right\},
\end{aligned}$$

which can be simplified to

$$\begin{aligned}
&\text{tr} \left\{ \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T + \mathbf{W} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathbf{W}^T \right\} \\
&+ 2 \sum_s c_s \text{tr} \left\{ -\mathbf{W} \mathbf{K}^{(s)} \mathbf{K}^{-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathbf{W}^T - \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{cov}(\mathbf{M}, \mathbf{D}^T) \mathbf{K}^{-1} \mathbf{K}^{(s)} \mathbf{W}^T \right\} \\
&+ \sum_s \sum_t c_s c_t \text{tr} \left\{ \mathbf{W} \left[\mathbf{K}^{(s)} \mathbf{K}^{-1} (\text{var}(\mathbf{D}) + \boldsymbol{\delta} \boldsymbol{\delta}^T) \mathbf{K}^{-1} \mathbf{K}^{(t)} \right] \mathbf{W}^T \right\}.
\end{aligned}$$

Recall that $\mathbf{D} = \boldsymbol{\delta} - \mathbf{K} \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} \mathbf{M} + \mathbf{K} \mathbf{N}$, so that

$$\begin{aligned}
\text{cov}(\mathbf{M}, \mathbf{D}^T) &= \text{cov}(\mathbf{M}, \mathbf{N}^T) \mathbf{K} - \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} \mathbf{K} \\
\text{var}(\mathbf{D}) &= \mathbf{K} \text{var}(\mathbf{N}) \mathbf{K} + \mathbf{K} \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} \mathbf{K} \\
&\quad - \mathbf{K} \text{cov}(\mathbf{N}, \mathbf{M}^T) \mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} \mathbf{K} - \mathbf{K} \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} \text{cov}(\mathbf{M}, \mathbf{N}^T) \mathbf{K}.
\end{aligned}$$

Hence,

$$\begin{aligned} \text{tr} \{ \boldsymbol{\Omega} + \boldsymbol{\zeta} \boldsymbol{\zeta}^T \} &= \text{tr} \left\{ \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \text{var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \right)^T + \mathbf{W} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathbf{W}^T \right\} \\ &\quad + 2 \sum_s c_s \text{tr} \{ \boldsymbol{\Delta}_1 \mathbf{K}^{(s)} \mathbf{W}^T \} \\ &\quad + \sum_s \sum_t c_s c_t \text{tr} \{ \mathbf{W} [\mathbf{K}^{(s)} \boldsymbol{\Delta}_2 \mathbf{K}^{(t)}] \mathbf{W}^T \}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\Delta}_1 &= \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\theta\theta}^{-1} \left(\text{Var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} - \text{Cov}(\mathbf{M}, \mathbf{N}^T) \right) - \mathbf{W} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathbf{K}^{-1}, \\ \boldsymbol{\Delta}_2 &= \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} \text{Var}(\mathbf{M}) \mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} + \text{Var}(\mathbf{N}) - 2 \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} \text{Cov}(\mathbf{M}, \mathbf{N}^T) + \mathbf{K}^{-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \mathbf{K}^{-1}. \end{aligned}$$

The first term in $\text{tr} \{ \boldsymbol{\Omega} + \boldsymbol{\zeta} \boldsymbol{\zeta}^T \}$ does not depend on the weights. Hence, minimizing $\text{tr} \{ \boldsymbol{\Omega} + \boldsymbol{\zeta} \boldsymbol{\zeta}^T \}$ is equivalent to minimizing

$$Q(\mathbf{c}) = \sum_s c_s \text{tr} \{ \boldsymbol{\Delta}_1 \mathbf{K}^{(s)} \mathbf{W}^T \} + \frac{1}{2} \sum_s \sum_t c_s c_t \text{tr} \{ \mathbf{W} \mathbf{K}^{(s)} \boldsymbol{\Delta}_2 \mathbf{K}^{(t)} \mathbf{W}^T \}. \quad (50)$$

The proof is completed. □

Unbiased estimator of $\boldsymbol{\delta}$ and $\boldsymbol{\delta} \boldsymbol{\delta}^T$. Using the inverse of a 2×2 block matrix,

$$\mathbf{J}_s^{-1} = \begin{pmatrix} \mathbf{J}_{\theta\theta}^{-1} + \mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} \boldsymbol{\pi}_s^T \mathbf{K}_s \boldsymbol{\pi}_s \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} & -\mathbf{J}_{\theta\theta}^{-1} \mathbf{J}_{\theta\gamma} \boldsymbol{\pi}_s^T \mathbf{K}_s \\ -\mathbf{K}_s \boldsymbol{\pi}_s \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} & \mathbf{K}_s \end{pmatrix}.$$

Then, equation (48) implies that

$$\hat{\boldsymbol{\delta}} = \sqrt{n} (\hat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0) \xrightarrow{d} \boldsymbol{\delta} - \mathbf{K} \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} \mathbf{M} + \mathbf{K} \mathbf{N} \sim N(\boldsymbol{\delta}, \mathbf{G} \mathbf{H} \mathbf{G}^T),$$

where $\mathbf{G} = \begin{pmatrix} -\mathbf{K} \mathbf{J}_{\theta\gamma}^T \mathbf{J}_{\theta\theta}^{-1} & \mathbf{K} \end{pmatrix}$ and

$$\mathbf{H} = \text{var} \begin{pmatrix} \mathbf{M} \\ \mathbf{N} \end{pmatrix} = \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right)^T \mathbf{V} \boldsymbol{\Upsilon} \mathbf{V} \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T}.$$

Hence, the mean of the asymptotic distribution of $\hat{\boldsymbol{\delta}}$ is still $\boldsymbol{\delta}$. The mean of the asymptotic distribution of $\hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}^T$ is $\boldsymbol{\delta} \boldsymbol{\delta}^T + \mathbf{G} \mathbf{H} \mathbf{G}^T$. An asymptotically unbiased estimator of $\boldsymbol{\delta} \boldsymbol{\delta}^T$ is then $\hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}^T - \hat{\mathbf{G}} \hat{\mathbf{H}} \hat{\mathbf{G}}^T$. □

Heuristic proof of the joint convergence. In practice, $\mathbf{c}$ is estimated by maximizing

$$\hat{Q}(\mathbf{c}) = \sum_s c_s \text{tr} \left\{ \hat{\Delta}_1 \hat{\mathbf{K}}^{(s)} \hat{\mathbf{W}}^T \right\} + \frac{1}{2} \sum_s \sum_t c_s c_t \text{tr} \left\{ \hat{\mathbf{W}} \hat{\mathbf{K}}^{(s)} \hat{\Delta}_2 \hat{\mathbf{K}}^{(t)} \hat{\mathbf{W}}^T \right\},$$

subject to the unit simplex, where $\hat{\mathbf{W}}$ and $\hat{\mathbf{K}}^{(s)}$ are consistent estimators of $\mathbf{W}$ and $\mathbf{K}^{(s)}$ respectively. Since $\hat{\boldsymbol{\delta}} = \sqrt{n}(\hat{\gamma}_{full} - \gamma_0) \xrightarrow{d} \mathbf{D}$, $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are not consistent estimators of Δ_1 and Δ_2 . They are only asymptotically unbiased. Hence, $\hat{Q}(\mathbf{c}) \xrightarrow{d} Q^*(\mathbf{c})$ for some $Q^*(\mathbf{c})$, of which the distribution depends on the joint distribution of $\mathbf{M}$ and $\mathbf{N}$. If the quadratic programming is positive definite, then $\mathbf{c}^*$, the minimizer of $Q^*(\mathbf{c})$, is unique. In the context of linear regression, Liu (2015) showed that

$$\sqrt{n}(\bar{\boldsymbol{\mu}}(\hat{\mathbf{c}}) - \boldsymbol{\mu}_{true}) \xrightarrow{d} \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \mathbf{M} + \mathbf{W} \left\{ \boldsymbol{\delta} - \left(\sum_s c_s^* \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \mathbf{D} \right\}, \quad (51)$$

where $\hat{\mathbf{c}}$ is the minimizer of $\hat{Q}(\mathbf{c})$ and $\mathbf{c}^*$ is the minimizer of $Q^*(\mathbf{c})$. Suppose that $\hat{\mathbf{c}} \xrightarrow{d} \mathbf{c}^*$. Since the distribution of $Q^*(\mathbf{c})$ depends on the joint distribution of $\mathbf{M}$ and $\mathbf{N}$, the distribution of $\mathbf{c}^*$ also depends on $\mathbf{M}$ and $\mathbf{N}$. Recall that the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\mu}}_s - \boldsymbol{\mu}_{true})$ also depends on the joint distribution of $\mathbf{M}$ and $\mathbf{N}$ by Lemma 1. Therefore, there is joint convergence in the distribution of $\hat{\mathbf{c}}$ and $\sqrt{n}(\hat{\boldsymbol{\mu}}_s - \boldsymbol{\mu}_{true})$, and (51) holds.

A rigorous proof of $\hat{\mathbf{c}} \xrightarrow{d} \mathbf{c}^*$ should be based on the argmax continuous mapping theorem (e.g., Van der Vaart, and Wellner, 1996, Theorem 3.2.2). Intuitively speaking, since $\hat{\mathbf{c}}$ is the minimizer of $\hat{Q}(\mathbf{c})$, the distribution of $\hat{\mathbf{c}}$ depends on the distribution of $\hat{\boldsymbol{\delta}}$. Hence, the asymptotic distribution of $\hat{\mathbf{c}}$ depends on the joint distribution of $\mathbf{M}$ and $\mathbf{N}$. If the quadratic programming is positive definite, then its minimizer $\hat{\mathbf{c}}$ is unique and $\hat{\mathbf{c}} = O_P(1)$. As the minimizer, $\hat{Q}(\hat{\mathbf{c}}) \leq \inf_{\mathbf{c}} \hat{Q}(\mathbf{c}) + o_P(1)$. Hence, the assumption $\hat{\mathbf{c}} \xrightarrow{d} \mathbf{c}^*$ is likely to hold.

□

Proof of the equivalence of confidence intervals. Equation (49) indicates that

$$\sqrt{n}(\hat{\boldsymbol{\mu}}_{full} - \boldsymbol{\mu}_{true}) = \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \mathbf{M} + \mathbf{W}(\boldsymbol{\delta} - \mathbf{D}) + O_P(n^{-1/2}).$$

Hence, the full model confidence interval for μ_i is

$$\left[\hat{\mu}_{i,full} - z_{1-\alpha/2} \frac{\hat{\kappa}_i}{\sqrt{n}}, \quad \hat{\mu}_{i,full} + z_{1-\alpha/2} \frac{\hat{\kappa}_i}{\sqrt{n}} \right],$$

where $\hat{\mu}_{i,full}$ is the i th entry of $\hat{\boldsymbol{\mu}}_{full}$, $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, and κ_i is (i, i) th entry of the covariance matrix of $\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \mathbf{M} - \mathbf{W} \mathbf{D}$ with $\hat{\kappa}_i$ being its estimator. Contrasting it with equation (6) in the main text, it suffices to show that

$$\bar{\boldsymbol{\mu}}(\hat{\mathbf{c}}) - \hat{\mathbf{W}} \left[\mathbf{I} - \left(\sum_s \hat{c}_s \hat{\mathbf{K}}^{(s)} \right) \hat{\mathbf{K}}^{-1} \right] \hat{\boldsymbol{\delta}}/\sqrt{n} = \hat{\boldsymbol{\mu}}_{full} + o_P(n^{-1/2}). \quad (52)$$

Since $\hat{\mathbf{W}}$, $\hat{\mathbf{K}}$ and $\hat{\mathbf{K}}^{(s)}$ are consistent estimators of $\mathbf{W}$, $\mathbf{K}$ and $\mathbf{K}^{(s)}$ respectively, we get $\hat{\mathbf{W}} = \mathbf{W} + o_P(1)$ and $\hat{\mathbf{K}}^{(s)} \hat{\mathbf{K}}^{-1} = \mathbf{K}^{(s)} \mathbf{K}^{-1} + o_P(1)$. Since the entries of $\hat{\mathbf{c}}$ are bounded between 0 and 1, $\hat{\mathbf{c}} o_P(1) = o_P(1)$. Hence,

$$\begin{aligned} \hat{\mathbf{W}} \left[\mathbf{I} - \left(\sum_s \hat{c}_s \hat{\mathbf{K}}^{(s)} \hat{\mathbf{K}}^{-1} \right) \right] \hat{\boldsymbol{\delta}}/\sqrt{n} &= (\mathbf{W} + o_P(1)) \left[\mathbf{I} - \sum_s \hat{c}_s \mathbf{K}^{(s)} \mathbf{K}^{-1} + \sum_s \hat{c}_s o_P(1) \right] \hat{\boldsymbol{\delta}}/\sqrt{n} \\ &= \mathbf{W} \left[\mathbf{I} - \sum_s \hat{c}_s \mathbf{K}^{(s)} \mathbf{K}^{-1} \right] \hat{\boldsymbol{\delta}}/\sqrt{n} + o_P(n^{-1/2}), \end{aligned} \quad (53)$$

where the last equality holds since $\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathbf{D}$ and $o_P(1) \hat{\boldsymbol{\delta}} \xrightarrow{p} 0$. In the likelihood context, Wang and Zhou (2013) showed that

$$\bar{\boldsymbol{\mu}}(\hat{\mathbf{c}}) - \mathbf{W} \left[\mathbf{I} - \left(\sum_s \hat{c}_s \mathbf{K}^{(s)} \right) \mathbf{K}^{-1} \right] \hat{\boldsymbol{\delta}}/\sqrt{n} = \hat{\boldsymbol{\mu}}_{full} + o_P(n^{-1/2}). \quad (54)$$

Even though they only considered the likelihood context, it turns out that their proof is directly applicable to the context of SEM with ordinal data, which means that (54) also holds in ordinal SEM. Hence, (53) and (54) lead to (52). □

Theorem 3. *Let the focus parameter be $\boldsymbol{\mu} = \boldsymbol{\beta}$. Support that the assumptions in Lemma 1 hold, that the Assumptions A1 to A5 hold, and that the joint convergence in distribution (51) holds. Then, $T_{FMA} = T_{full} + o_P(1)$, where $T_{full} = n(\hat{\boldsymbol{\rho}} - \hat{\boldsymbol{\sigma}}_{full})^T \hat{\mathbf{V}}(\hat{\boldsymbol{\rho}} - \hat{\boldsymbol{\sigma}}_{full})$ is the test statistic for the full model.*

Proof of Theorem 3. Since both T_{FMA} and T_{full} are quadratic forms, it suffices to show that

$$\sqrt{n}(\hat{\boldsymbol{\rho}} - \bar{\boldsymbol{\sigma}}) + \frac{\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\beta}}_{full})}{\partial \boldsymbol{\beta}^T} \hat{\mathbf{W}} \left(\mathbf{I} - \sum_s \hat{c}_s \hat{\mathbf{K}}^{(s)} \hat{\mathbf{K}}^{-1} \right) \hat{\boldsymbol{\delta}} = \sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\hat{\boldsymbol{\mu}}_{full})) + o_P(1).$$

Because of (53), we get

$$\begin{aligned} & \frac{\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\beta}}_{full})}{\partial \boldsymbol{\beta}^T} \hat{\mathbf{W}} \left(\mathbf{I} - \sum_s \hat{c}_s \hat{\mathbf{K}}^{(s)} \hat{\mathbf{K}}^{-1} \right) \hat{\boldsymbol{\delta}} \\ &= \left(\frac{\partial \boldsymbol{\sigma}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} + o_P(1) \right) \left[\mathbf{W} \left(\mathbf{I} - \sum_s \hat{c}_s \mathbf{K}^{(s)} \mathbf{K}^{-1} \right) \hat{\boldsymbol{\delta}} + o_P(1) \right] \\ &= \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \mathbf{W} \left(\mathbf{I} - \sum_s \hat{c}_s \mathbf{K}^{(s)} \mathbf{K}^{-1} \right) \hat{\boldsymbol{\delta}} + o_P(1), \end{aligned} \quad (55)$$

where the second equality holds since $\hat{\mathbf{c}} o_P(1) = o_P(1)$, $\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathbf{D}$ and $o_P(1) \hat{\boldsymbol{\delta}} \xrightarrow{p} 0$. Note that

$$\bar{\boldsymbol{\sigma}} = \boldsymbol{\sigma}_0 + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\mu}_0)}{\partial \boldsymbol{\mu}^T} [\bar{\boldsymbol{\mu}}(\hat{\mathbf{c}}) - \boldsymbol{\mu}_0] + O_P(n^{-1})$$

by (51), and

$$\boldsymbol{\sigma}(\hat{\boldsymbol{\mu}}_{full}) = \boldsymbol{\sigma}_0 + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\mu}_0)}{\partial \boldsymbol{\mu}^T} (\hat{\boldsymbol{\mu}}_{full} - \boldsymbol{\mu}_0) + O_P(n^{-1})$$

by (49). Then

$$\bar{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\hat{\boldsymbol{\mu}}_{full}) + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\mu}_0)}{\partial \boldsymbol{\mu}^T} [\bar{\boldsymbol{\mu}}(\hat{\mathbf{c}}) - \hat{\boldsymbol{\mu}}_{full}] + O_P(n^{-1}). \quad (56)$$

Hence, using (55), (56), and (54), we further obtain

$$\begin{aligned} & \sqrt{n}(\hat{\boldsymbol{\rho}} - \bar{\boldsymbol{\sigma}}) + \frac{\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\beta}}_{full})}{\partial \boldsymbol{\beta}^T} \hat{\mathbf{W}} \left(\mathbf{I} - \sum_s \hat{c}_s \hat{\mathbf{K}}^{(s)} \hat{\mathbf{K}}^{-1} \right) \hat{\boldsymbol{\delta}} \\ &= \sqrt{n}(\hat{\boldsymbol{\rho}} - \bar{\boldsymbol{\sigma}}) + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \mathbf{W} \left(\mathbf{I} - \sum_s \hat{c}_s \mathbf{K}^{(s)} \mathbf{K}^{-1} \right) \hat{\boldsymbol{\delta}} + o_P(1) \\ &= \sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\hat{\boldsymbol{\mu}}_{full})) - \sqrt{n} \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\mu}_0)}{\partial \boldsymbol{\mu}^T} [\bar{\boldsymbol{\mu}}(\hat{\mathbf{c}}) - \hat{\boldsymbol{\mu}}_{full}] + \frac{\partial \boldsymbol{\sigma}(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \mathbf{W} \left(\mathbf{I} - \sum_s \hat{c}_s \mathbf{K}^{(s)} \mathbf{K}^{-1} \right) \hat{\boldsymbol{\delta}} + o_P(1) \\ &= \sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\hat{\boldsymbol{\mu}}_{full})) + o_P(1), \end{aligned}$$

which completes the proof. $\square$

4 References

- Ferguson, T. S. (1996). *A course in large sample theory*. Chapman & Hall, London, UK.
- Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, 9:861–869.
- Jiang, J. (2010). *Large sample techniques for statistics*. Springer, New York, NY.
- Jin, S. and Ankargren, S. (2019). Frequentist model averaging in structural equation modeling. *Psychometrika*, 84:84–104.
- Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, 59:381–389.
- Kamat, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications. *Biometrika*, 40:20–34.
- Liu, C.-A. (2015). Distribution theory of the least squares averaging estimator. *Journal of Econometrics*, 186:142–159.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44:443–460.
- Rhemtulla, M., Brosseau-Liard, P. E., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical sem estimation methods under suboptimal conditions. *Psychological Methods*, 17:354–373.
- Richmond, T. (2020). *General topology*. De Gruyter, Berlin, Germany.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48:1–36.

- Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, New York, NY, 3rd edition.
- Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, NY.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.
- Wang, H. and Zhou, S. Z. F. (2013). Interval estimation by frequentist model averaging. *Communications in Statistics: Theory and Methods*, 42:4342–4356.