FREQUENTIST MODEL AVERAGING IN STRUCTURE EQUATION MODEL WITH ORDINAL DATA

Shaobo Jin

UPPSALA UNIVERSITY

In practice, it is common that a best fitting structural equation model (SEM) is selected from a set of candidate SEMs and inference is conducted conditional on the selected model. Such post-selection inference ignores the model selection uncertainty and yields too optimistic inference. Using the largest candidate model avoids model selection uncertainty but introduces a large variation. Jin and Ankargren (Psychometrika 84:84–104, 2019) proposed to use frequentist model averaging in SEM with continuous data as a compromise between model selection and the full model. They assumed that the true values of the parameters depend on $n^{-1/2}$ with n being the sample size, which is known as a local asymptotic framework. This paper shows that their results are not directly applicable to SEM with ordinal data. To address this issue, we prove consistency and asymptotic normality of the polychoric correlation estimators under the local asymptotic framework. Then, we propose a new frequentist model averaging estimator and a valid confidence interval that are suitable for ordinal data. Goodness-of-fit test statistics for the model averaging estimator are also derived.

Key words: mean squared error, confidence interval, goodness-of-fit test, model selection uncertainty, local asymptotic framework, pseudo maximum likelihood.

Structural equation models (SEMs) with ordinal data are widely used in social and behavioral sciences. As many other statistical models, a common practice is to choose an optimal model from a number of candidate models according to some criteria. A long-standing critique of model selection is that the post-selection inference is often conducted in such a way that model selection was never present. Because of the randomness in the data, the model selection step is also stochastic and ignoring such uncertainty yields too optimal inference. The reader is directed to Preacher and Merkle (2012) and Lubke et al. (2017) for discussions on the consequences of ignoring model selection uncertainty within SEM.

A well-known remedy to acknowledge the contribution of all candidate models is Bayesian model averaging. Instead of purely relying on the optimal model, Bayesian model averaging combines different candidate models together by a weighted average (Hoeting et al., 1999; Madigan & Raftery, 1994). In the past decades, the interest on frequentist model averaging (FMA) has grown exponentially in the statistics literature. The reader is directed to Fletcher (2018) for a long list of FMA related references.

The main purpose of the paper is to generalize the FMA principle to SEMs with ordinal data. Jin and Ankargren (2019) have applied the FMA technique to SEMs with continuous data, using the likelihood-based FMA machinery developed by Hjort and Claeskens (2003a). Similar to FMA in the other context, they showed that FMA tends to produce a mean squared error (MSE) that is lower than that of the full model if the population parameter value is small and is lower than that of model selection if the population value is moderate. Hence, the FMA estimator is a robust compromise between model selection and the full model in the SEM context. Since ordinal data are often encountered in practice, it is of interest to extend the FMA technique for SEM with

continuous data to SEM with ordinal data. As we shall see in the later sections, some results in Jin and Ankargren (2019) need to be revised for ordinal data.

The rest of the paper is organized as follows. First, the results in Jin and Ankargren (2019) are briefly reviewed. Second, the necessary modifications for the FMA estimator in ordinal SEM are presented. Third, a simulation study is conducted to investigate the small sample properties of the FMA technique. Fourth, FMA is applied to an empirical example as an illustration. A discussion ends the paper.

1. Background

Consider the SEM

$$
\begin{aligned}
x^* &= \Lambda_x \xi + \delta_x, \\
y^* &= \Lambda_y \eta + \delta_y, \\
\eta &= B\eta + \Gamma\xi + \varepsilon,
\end{aligned}
$$

where x^* ($p_x \times 1$) and y^* ($p_y \times 1$) are the vectors of continuous indicators, ξ and η are the latent variables, δ_x, δ_y, and ε are the error terms. Here, Λ_x and Λ_y are the loading matrices, B contains the linear effects among η, and Γ contains linear effects of ξ on η. The joint distribution of x^* and y^* is assumed to be multivariate normal. Throughout the paper, Σ is used to denote the model-implied covariance of the SEM, σ the vector of unique entries in Σ, and β the vector of all free parameters needed to determine Σ (i.e., free parameters in Λ_x, Λ_y, B, Γ, var(ξ), var(δ_x), var(δ_y), cov(δ_x, δ_y^T), and var(ε)).

1.1. Brief Review of Jin and Ankargren (2019)

Jin and Ankargren (2019) investigated FMA in SEM where x^* and y^* are observed. In order to clarify the modifications needed for ordinal SEM, their results are briefly reviewed in this subsection. Suppose that there exist a number of candidate SEMs, indexed by s, in which a full model that nests all other models and a narrow model that is nested in all other models are well-defined. The vector β is partitioned into $\beta^T = (\theta^T, \gamma^T)$, where θ is contained in all candidate models and $\gamma = \gamma_0$ is known in the narrow model. For candidate model s, the elements in γ to be estimated are $\gamma_s = \pi_s \gamma$, where π_s is a selection matrix. The candidate models are fitted by maximum likelihood, i.e., minimizing $F_{ML}(\beta) = n \log |\Sigma| + n\text{tr}\{S\Sigma^{-1}\} - n \log |S| - n(p_x + p_y)$, where n is the sample size, S is the sample covariance matrix, and tr {} is the matrix trace of the enclosed matrix. Suppose that the parameter vector of interest is $\mu = \mu(\theta, \gamma)$, which is continuously differentiable in θ and γ. The FMA estimator of μ is $\bar{\mu}(c) = \sum_s c_s \hat{\mu}_s$, where $\hat{\mu}_s$ is the estimator of μ in the candidate model s and the model weight vector $c = \{c_s\}$ lies in the unit simplex $\{c_s : \sum_s c_s = 1, \ 0 \le c_s \le 1\}$.

Jin and Ankargren (2019) assumed that $\beta_{true} = (\theta_0^T, \gamma_0^T + \delta^T / \sqrt{n})^T$ is the true value of β, where θ_0 is the true value of θ, $\gamma_0 + \delta / \sqrt{n}$ is the true value of γ, and δ is the local parameter. Hence, the true values of σ and μ are $\sigma_{true} = \sigma(\beta_{true})$ and $\mu_{true} = \mu(\beta_{true})$, respectively. The framework that the true value is drifted in a $n^{-1/2}$ neighborhood is known as the local asymptotic framework. In contrast, the standard asymptotic framework refers to the case where all true values are free of n. The local asymptotic framework is a popular choice to study FMA. The reader is directed to Hjort and Claeskens (2003b) for the reasons of using the local asymptotic framework.

Suppose that $n^{-1} \partial^2 F_{ML} \left(\boldsymbol{\beta}_0 \right) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$ converges in probability to $2 \boldsymbol{J}_{full}$, where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\theta}_0^T, \boldsymbol{\gamma}_0^T \right)^T$ and $\boldsymbol{J}_{full}$ can be partitioned to

$$\boldsymbol{J}_{full} = \begin{pmatrix} \boldsymbol{J}_{\theta\theta} & \boldsymbol{J}_{\theta\gamma} \\ \boldsymbol{J}_{\theta\gamma}^T & \boldsymbol{J}_{\gamma\gamma} \end{pmatrix}.$$

Let $\boldsymbol{M}$ and $\boldsymbol{N}$ be the random variables such that

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

For fixed weights, Jin and Ankargren (2019) derived that

$$\sqrt{n} \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true} \right) = \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{M} + \boldsymbol{W} \left\{ \boldsymbol{\delta} - \left(\sum_s c_s \boldsymbol{K}^{(s)} \right) \boldsymbol{K}^{-1} \boldsymbol{D} \right\} + O_P \left(n^{-1/2} \right), \quad (1)$$

where $\boldsymbol{\mu}_0 = \boldsymbol{\mu} \left(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 \right)$, $\boldsymbol{W} = \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{J}_{\theta\gamma} - \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\gamma}^T}$, $\boldsymbol{K}^{-1} = \boldsymbol{J}_{\gamma\gamma} - \boldsymbol{J}_{\theta\gamma}^T \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{J}_{\theta\gamma}$, $\boldsymbol{K}_s = \left(\boldsymbol{\pi}_s \boldsymbol{K}^{-1} \boldsymbol{\pi}_s^T \right)^{-1}$, $\boldsymbol{K}^{(s)} = \boldsymbol{\pi}_s^T \boldsymbol{K}_s \boldsymbol{\pi}_s$, and $\boldsymbol{D} = \boldsymbol{\delta} - \boldsymbol{K} \boldsymbol{J}_{\theta\gamma}^T \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{M} + \boldsymbol{K} \boldsymbol{N}$ [1]. To estimate $\boldsymbol{c}$, the limit of $n E \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true} \right)^T \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true} \right)$ is minimized, which is equivalent to minimizing

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

subject to the unit simplex, where $\boldsymbol{\Delta}_1 = -\boldsymbol{W} \boldsymbol{\delta} \boldsymbol{\delta}^T \boldsymbol{K}^{-1}$, and $\boldsymbol{\Delta}_2 = \boldsymbol{K}^{-1} + \boldsymbol{K}^{-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \boldsymbol{K}^{-1}$.

In the spirit of Hjort and Claeskens (2003a) and Liu (2015), Jin and Ankargren (2019) also proposed a confidence interval for μ_i, the ith entry in $\boldsymbol{\mu}$, which is given by

$$\left[\bar{\mu}_i - \frac{\hat{u}_i}{\sqrt{n}} - z_{1-\alpha/2} \frac{\hat{\kappa}_i}{\sqrt{n}}, \quad \bar{\mu}_i - \frac{\hat{u}_i}{\sqrt{n}} + z_{1-\alpha/2} \frac{\hat{\kappa}_i}{\sqrt{n}} \right], \quad (3)$$

where $\hat{u}_i$ is the ith entry of the vector $\hat{\boldsymbol{W}} \left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}} \right)$ with $\tilde{\boldsymbol{\delta}} = \sum_s \hat{c}_s \hat{\boldsymbol{K}}^{(s)} \hat{\boldsymbol{K}}^{-1} \hat{\boldsymbol{\delta}}$, $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, and κ_i is (i, i)th entry of the covariance matrix of $\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{M} - \boldsymbol{W} \boldsymbol{D}$ with $\hat{\kappa}_i$ being its estimator.

1.2. Ordinal Data Model

In the ordinal SEM, the ordinal counterparts $\boldsymbol{x}$ and $\boldsymbol{y}$ are observed, which are obtained by discretizing $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$. In the ordinal data model, the diagonal entries in $\boldsymbol{\Sigma}$ are assumed to be 1. Accordingly, $\boldsymbol{\Sigma}$ is a polychoric correlation matrix. In practice, a multi-step procedure is commonly used to fit an ordinal SEM. First, the polychoric correlation matrix and its asymptotic covariance matrix are estimated. Second, the least squares fit function $F_{LS} \left(\boldsymbol{\beta} \right) = n \left(\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma} \left(\boldsymbol{\beta} \right) \right)^T \hat{\boldsymbol{V}} \left(\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma} \left(\boldsymbol{\beta} \right) \right)$ is minimized, where $\boldsymbol{\rho}$ is the vector of unique polychoric correlation coefficients and $\boldsymbol{V}$ is some weight matrix.

---

[1] For the purpose of presentation, the notations in Jin and Ankargren (2019) are modified. $\boldsymbol{J}_{full}$, $\boldsymbol{M}$, and $\boldsymbol{N}$ here corresponds to $\boldsymbol{J}_{full}/4$, $\boldsymbol{M}/2$, and $\boldsymbol{N}/2$ in Jin and Ankargren (2019).

In the current study, we will extend Jin and Ankargren (2019) to SEM with ordinal data. Since both F_{ML} and F_{LS} can be viewed as distance functions between ρ and σ, many results from Jin and Ankargren (2019) still hold. However, as we shall explain in the next section, some modifications are also needed, due to the choice of V.

2. Frequentist Model Averaging Estimator

2.1. Polychoric Correlation Estimation

Throughout the paper, we estimate the polychoric correlation coefficients 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. Estimation of the polychoric correlation matrix and its asymptotic covariance matrix has been extensively studied (e.g., Jin & Yang-Wallentin, 2017; Jöreskog, 1994; Monroe, 2017; Muthén, 1984) within the standard asymptotic framework. For example, Jöreskog (1994) showed that

$$\sqrt{n} \left(\hat{\rho} - \sigma_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \boldsymbol{\Upsilon} \right), \tag{4}$$

where $\sigma_0 = \sigma \left(\beta_0 \right)$ and $\boldsymbol{\Upsilon}$ is the asymptotic covariance matrix.

We still assume that the thresholds are not locally drifted. However, the true values of the polychoric correlation coefficients depend on n. To the best of our knowledge, none of the above mentioned studies on polychoric correlations are conducted under the local asymptotic framework. Further, the polychoric correlation estimator from the two-step procedure is a pseudo-maximum likelihood estimator (Gong & Samaniego, 1981), making the results in Hjort and Claeskens (2003a) not directly applicable. For these reasons, we establish consistency and asymptotic normality of the polychoric correlation estimators in this subsection. For ease of presentation, all regularity conditions and mathematical proofs are placed in the online appendix.

Theorem 1. *Under the regularity conditions stated in the online appendix, $\hat{\rho} \xrightarrow{p} \sigma_0$.*

Theorem 1 shows that $\hat{\rho}$ remains a consistent estimator of σ_0 under the local asymptotic framework. The following theorem shows asymptotic normality.

Theorem 2. *Under the regularity conditions stated in the online appendix, $\sqrt{n} \left(\hat{\rho} - \sigma_{true} \right) \xrightarrow{d} N \left(\mathbf{0}, \boldsymbol{\Upsilon} \right)$, where $\boldsymbol{\Upsilon}$ is the same as the covariance matrix in* (4).

Theorem 2 shows that the estimator of the asymptotic covariance matrix under the standard asymptotic framework is also valid under the local asymptotic framework. The implication is that the estimated asymptotic covariance matrix can simply be extracted from the standard SEM packages. Theorem 2 also shows that the mean of the limiting distribution of $\sqrt{n} \left(\hat{\rho} - \sigma_0 \right)$ is nonzero under the local asymptotic framework. In contrast, the mean of the limiting distribution of $\sqrt{n} \left(\hat{\rho} - \sigma_0 \right)$ is zero under the standard asymptotic framework.

2.2. Frequentist Model Averaging: From Continuous Data to Ordinal Data

Since both F_{LS} and F_{ML} can be viewed as distance functions between ρ and σ, the expansion (1) still holds if F_{ML} is replaced by F_{LS} when obtaining J_{full}, M, and N. However, Jin and Ankargren (2019) derived the weight estimation criterion (2) and the confidence interval (3) under the assumption that M and D are independent, which holds for observed multivariate

normal data. Regarding ordinal SEM, it is shown in the appendix that the joint distribution of $\boldsymbol{M}$ and $\boldsymbol{N}$ is multivariate normal with mean $\boldsymbol{0}$ and covariance matrix

$$\boldsymbol{H} \overset{\text{def}}{=} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \right)^T \boldsymbol{V} \boldsymbol{\Upsilon} \boldsymbol{V} \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T}. \tag{5}$$

Consequently,

$$\text{cov}\left(\boldsymbol{M}, \boldsymbol{D}^T \right) = \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right)^T \boldsymbol{V} \boldsymbol{\Upsilon} \boldsymbol{V} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \right) \boldsymbol{K} - \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right)^T \boldsymbol{V} \boldsymbol{\Upsilon} \boldsymbol{V} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right) \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{J}_{\theta\gamma} \boldsymbol{K}, \tag{6}$$

which depends on the choice of $\boldsymbol{V}$. Commonly used $\boldsymbol{V}$ includes $\boldsymbol{V} = \boldsymbol{I}$ in unweighted least squares (ULS; Muthén, 1978), the inverse of diagonal elements of $\boldsymbol{\Upsilon}$ in diagonally weighted least squares (DWLS; Muthén et al., 1997), and $\boldsymbol{V} = \boldsymbol{\Upsilon}^{-1}$ in weighted least squares (WLS; Browne, 1984). If WLS is used, then $\text{cov}\left(\boldsymbol{M}, \boldsymbol{D}^T \right) = \boldsymbol{0}$ and the results in Jin and Ankargren (2019) remain applicable. However, if ULS or DWLS is used, $\boldsymbol{M}$ and $\boldsymbol{D}$ are not necessarily independent. Consequently, modifications are needed.

2.3. Weight Estimation

It is shown in the online appendix that, in the context of ordinal SEM, minimizing the limit of $nE \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true} \right)^T \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_{true} \right)$ is equivalent to minimizing $Q \left(\boldsymbol{c} \right)$ given in (2), but with modified $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ as

$$\boldsymbol{\Delta}_1 = \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \left(\text{var}\left(\boldsymbol{M} \right) \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{J}_{\theta\gamma} - \text{cov}\left(\boldsymbol{M}, \boldsymbol{N}^T \right) \right) - \boldsymbol{W} \boldsymbol{\delta} \boldsymbol{\delta}^T \boldsymbol{K}^{-1},$$

$$\boldsymbol{\Delta}_2 = \boldsymbol{J}_{\theta\gamma}^T \boldsymbol{J}_{\theta\theta}^{-1} \text{var}\left(\boldsymbol{M} \right) \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{J}_{\theta\gamma} + \text{var}\left(\boldsymbol{N} \right) - 2 \boldsymbol{J}_{\theta\gamma}^T \boldsymbol{J}_{\theta\theta}^{-1} \text{cov}\left(\boldsymbol{M}, \boldsymbol{N}^T \right) + \boldsymbol{K}^{-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \boldsymbol{K}^{-1},$$

where the covariance matrix of $\boldsymbol{M}$ and $\boldsymbol{N}$ is shown in (5).

In practice, the unknown population values (e.g., $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$, $\boldsymbol{K}^{(s)}$, and $\boldsymbol{W}$) are replaced by their estimators from the candidate models, yielding the estimator $\hat{Q} \left(\boldsymbol{c} \right)$ of $Q \left(\boldsymbol{c} \right)$. Then, $\hat{\boldsymbol{c}}$, the estimator of $\boldsymbol{c}$, is obtained by minimizing $\hat{Q} \left(\boldsymbol{c} \right)$. Similar to Jin and Ankargren (2019), the unknown population values can be consistently estimated from the full model, except $\boldsymbol{\delta}$. Hjort and Claeskens (2003a) and Liu (2015) showed that $\boldsymbol{\delta}$ can only be asymptotically unbiasedly estimated and suggested to estimate it from the full model as $\hat{\boldsymbol{\delta}} = \sqrt{n} \left(\hat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0 \right)$, where $\hat{\boldsymbol{\gamma}}_{full}$ is the estimator of $\boldsymbol{\gamma}$ from the full model. It is an unbiased estimator of $\boldsymbol{\delta}$ but a biased estimator of $\boldsymbol{\delta} \boldsymbol{\delta}^T$. Jin and Ankargren (2019) showed that, for SEM with continuous data, $\hat{\boldsymbol{\delta}}$ remains an unbiased estimator of $\boldsymbol{\delta}$ and $\hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}^T - \boldsymbol{K}$ is an unbiased estimator of $\boldsymbol{\delta} \boldsymbol{\delta}^T$. It is shown in the online appendix that, for SEM with ordinal data, $\hat{\boldsymbol{\delta}}$ is still an unbiased estimator of $\boldsymbol{\delta}$, but an unbiased estimator of $\boldsymbol{\delta} \boldsymbol{\delta}^T$ becomes $\hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}^T - \hat{\boldsymbol{G}} \hat{\boldsymbol{H}} \hat{\boldsymbol{G}}^T$, where $\boldsymbol{G} = \left(-\boldsymbol{K} \boldsymbol{J}_{\theta\gamma}^T \boldsymbol{J}_{\theta\theta}^{-1} \ \boldsymbol{K} \right)$.

2.4. Model Averaging Confidence Interval

In the spirit of Hjort and Claeskens (2003a) and Liu (2015), we conjecture that there is joint convergence in distribution of $\hat{\boldsymbol{c}}$ and all $\hat{\boldsymbol{\mu}}_s$ such that

$$\sqrt{n} \left(\bar{\boldsymbol{\mu}} \left(\hat{\boldsymbol{c}} \right) - \boldsymbol{\mu}_{true} \right) - \hat{\boldsymbol{W}} \left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}} \right) \overset{d}{\to} \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{M} + \boldsymbol{W} \left(\boldsymbol{\delta} - \boldsymbol{D} \right). \tag{7}$$

The reader is directed to the online appendix for a heuristic proof of the joint convergence. From (7), the FMA confidence interval for μ_i is still of the form (3) for SEM with ordinal data. In Jin and Ankargren (2019),

$$\mathrm{cov}\left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{M} - \boldsymbol{W}\boldsymbol{D}\right) = \frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T}\right)^T + \boldsymbol{W}\boldsymbol{K}\boldsymbol{W}^T.$$

In an ordinal SEM, the covariance matrix of $\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} \boldsymbol{M} - \boldsymbol{W}\boldsymbol{D}$ should be computed using (6), due to nonzero $\mathrm{cov}(\boldsymbol{M}, \boldsymbol{D}^T)$. Similar to weight estimation, the confidence interval in Jin and Ankargren (2019) remains applicable if WLS is used.

If (7) holds, the interval (3) that accounts for $\mathrm{cov}(\boldsymbol{M}, \boldsymbol{D}^T)$ attains the nominal level asymptotically. Another valid confidence interval is the one from the full model, given by $\left[\hat{\mu}_{i,full} - z_{1-\alpha/2}\hat{\kappa}_i/\sqrt{n}, \ \hat{\mu}_{i,full} + z_{1-\alpha/2}\hat{\kappa}_i/\sqrt{n}\right]$, where $\hat{\mu}_{i,full}$ is the estimator of μ_i from the full model. Various studies (e.g., Ankargren & Jin, 2018; Kabaila & Leeb, 2006; Wang & Zhou, 2013) have shown that the FMA confidence intervals of the form (3) can be asymptotically equivalent to the confidence interval from the full model. In the likelihood context, Wang and Zhou (2013) proved that

$$\bar{\boldsymbol{\mu}}\left(\hat{\boldsymbol{c}}\right) - \boldsymbol{W}\left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}\right)/\sqrt{n} = \hat{\boldsymbol{\mu}}_{full} + o_{\mathrm{P}}\left(n^{-1/2}\right), \tag{8}$$

and showed that the equivalence holds for all FMA confidence intervals suggested by Hjort and Claeskens (2003a). Since least squares are used in SEM with ordinal data, their results cannot be directly applied here. Nevertheless, it is shown in the online appendix that (8) still holds in ordinal SEM. Hence, the FMA confidence interval (3) remains asymptotically equivalent to the full model interval. The implication is that the small sample realizations of (3) may be different from the full model interval, but they will be similar to each other when the sample size is large.

2.5. Goodness-of-fit Test

Consider the case where $\boldsymbol{\mu} = \boldsymbol{\beta}$, i.e., we want all parameters (expect the thresholds) to be accurately estimated. In practice, it is often of interest to test the overall fit of a hypothesized model. The full model goodness-of-fit test can be interpreted as testing whether the full model decomposition of the population covariance matrix evaluated at the full model estimator fits the data well. Since FMA aims to combine different models, the population covariance matrix is generally decomposed according to the full model, but evaluated at the FMA estimator. Hence, it is important to test whether the full model evaluated at the FMA estimators fits the data well (Jin & Ankargren, 2019). For this reason, a goodness-of-fit test for model averaged SEM with ordinal data is proposed here.

A residual-based test statistic is

$$T_{FMA} = \left(\sqrt{n}\left(\hat{\boldsymbol{\rho}} - \bar{\boldsymbol{\sigma}}\right) + \frac{\partial \boldsymbol{\sigma}\left(\hat{\boldsymbol{\beta}}_{full}\right)}{\partial \boldsymbol{\beta}^T}\hat{\boldsymbol{W}}\left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}\right)\right)^T \hat{\boldsymbol{V}} \left(\sqrt{n}\left(\hat{\boldsymbol{\rho}} - \bar{\boldsymbol{\sigma}}\right) + \frac{\partial \boldsymbol{\sigma}\left(\hat{\boldsymbol{\beta}}_{full}\right)}{\partial \boldsymbol{\beta}^T}\hat{\boldsymbol{W}}\left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}\right)\right),$$

where $\bar{\boldsymbol{\sigma}} = \boldsymbol{\sigma}\left(\bar{\boldsymbol{\mu}}\left(\hat{\boldsymbol{c}}\right)\right)$. The asymptotic property of T_{FMA} is shown in following theorem.

Theorem 3. *Let* $\boldsymbol{\mu} = \boldsymbol{\beta}$. *Suppose that the regularity conditions in the online appendix hold. Then,* $T_{FMA} = T_{full} + o_P(1)$, *where* $T_{full} = n\left(\hat{\boldsymbol{\rho}} - \hat{\boldsymbol{\sigma}}_{full}\right)^T \hat{\boldsymbol{V}}\left(\hat{\boldsymbol{\rho}} - \hat{\boldsymbol{\sigma}}_{full}\right)$ *is the test statistic for the full model.*

The assumption that $\boldsymbol{\mu} = \boldsymbol{\beta}$ plays an important role in the proof of the theorem. It is certainly the case that some applications may be only interested in a subset of $\boldsymbol{\beta}$. In such a case, Theorem 3 is not guaranteed to be applicable. A different test statistic is proposed by Jin and Ankargren (2019) for SEM with continuous data. Both test statistics adjust the fit function (F_{ML} or F_{LS}), but with different adjustments. The adjustment $\frac{\partial \boldsymbol{\sigma}_{full}}{\partial \boldsymbol{\beta}^T} \boldsymbol{W} \left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}} \right)$ is used in T_{FMA}, since $\boldsymbol{W} \left(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}} \right)$ is also the adjustment term in the FMA confidence interval (3).

Since T_{FMA} is asymptotically equivalent to T_{full}, we can define the Satorra and Bentler (1994) mean-scaled statistic and mean-variance adjusted statistic as

$$T_{FMA-SB-m} = \frac{r}{\text{tr}\left(\hat{\boldsymbol{\Xi}}_{FMA} \right)} T_{FMA} \text{ and } T_{FMA-SB-mv} = \frac{\text{tr}\left(\hat{\boldsymbol{\Xi}}_{FMA} \right)}{\text{tr}\left(\hat{\boldsymbol{\Xi}}^2_{FMA} \right)} T_{FMA},$$

respectively, where r is the difference between the number of unique polychoric correlation coefficients and the number of parameters and

$$\boldsymbol{\Xi}_{FMA} = \text{tr}\left\{ \left(\boldsymbol{I} - \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \boldsymbol{L} \boldsymbol{V} \right)^T \boldsymbol{V} \left(\boldsymbol{I} - \frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\beta}^T} \boldsymbol{L} \boldsymbol{V} \right) \boldsymbol{\Upsilon} \right\}$$

with

$$\boldsymbol{L} = \left(\frac{\partial \boldsymbol{\mu}_0}{\partial \boldsymbol{\theta}^T} \boldsymbol{J}_{\theta\theta}^{-1} + \boldsymbol{W} \boldsymbol{K} \boldsymbol{J}_{\theta\gamma}^T \boldsymbol{J}_{\theta\theta}^{-1} \right) \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\theta}^T} \right)^T - \boldsymbol{W} \boldsymbol{K} \left(\frac{\partial \boldsymbol{\sigma}_0}{\partial \boldsymbol{\gamma}^T} \right)^T.$$

$T_{FMA-SB-m}$ can be approximated by a Chi-square distribution with r degrees of freedom and $T_{FMA-SB-mv}$ can be approximated by a Chi-square distribution with $\left[\text{tr}\left(\hat{\boldsymbol{\Xi}}_{FMA} \right) \right]^2 / \text{tr}\left(\hat{\boldsymbol{\Xi}}^2_{FMA} \right)$ degrees of freedom. The fit indices such as robust RMSEA, CFI, and TLI (Brosseau-Liard & Savalei, 2014; Brosseau-Liard et al., 2012) can also be defined accordingly.

3. Simulation Study

In this section, a simulation study is conducted to compare the finite sample properties of FMA with the full model estimation and model selection.

3.1. Simulation Design

The population model is a four-factor SEM, where

$$\boldsymbol{\Lambda}_x = \boldsymbol{\Lambda}_y = \begin{pmatrix} 1.0 & 0.95 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0.95 & 0.9 \end{pmatrix}^T,$$

$$\text{var}(\boldsymbol{\xi}) = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 0.0 & 0.0 \\ b_{21} & 0.0 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} 0.5 & \gamma_{12} \\ \gamma_{21} & 0.45 \end{pmatrix},$$

Under the local asymptotic framework, we let $\gamma_{12} = \gamma_0 + \delta/\sqrt{n}$, $\gamma_{21} = \gamma_0 + 0.75 \cdot \delta/\sqrt{n}$, and $b_{21} = \gamma_0 + 0.5 \cdot \delta/\sqrt{n}$, where $\gamma_0 = 0$ and $\delta = 150^{1/2}\zeta$, with ζ being seven equidistant values

between 0 and 0.30 at the step size 0.05. The covariance matrices of $\boldsymbol{\delta}_x$ and $\boldsymbol{\delta}_y$ are set to be diagonal, of which the diagonal elements make the marginal variances of $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ be 1. The distribution of $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ is assumed to follow a multivariate normal distribution. These population models are similar to those of Jin and Ankargren (2019), with minor changes of the true values. $\mathrm{var}(\boldsymbol{\varepsilon})$ is set to be a diagonal matrix, and its diagonal elements are 0.45 and 0.5 if the narrow model is the data generation process ($\zeta = 0$). For other values of ζ, the diagonal elements are chosen such that the reliability of the measurement models remains the same as the $\zeta = 0$.

We fix the number of categories to be five and consider two sets of threshold values. In the first set, 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). In the second set, the probabilities of belonging to each category are 0.52, 0.15, 0.13, 0.11, and 0.09, which is the extreme asymmetry setting in Rhemtulla et al. (2012). We also take the sample sizes that are used in Rhemtulla et al. (2012), i.e., $n = 150$, 350, and $n = 600$. The number of replications is set to 10, 000.

Four candidate models are considered in the current study, corresponding to the free parameters in $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$. Model 1, the narrow model, assumes $b_{21} = \gamma_{12} = \gamma_{21} = 0$. Model 2 frees γ_{12} but assumes $b_{21} = \gamma_{21} = 0$. Model 3 freely estimates γ_{12} and γ_{21}, but assumes $b_{21} = 0$. Model 4, the full model, freely estimates b_{21}, γ_{12} and γ_{21}. Hence, only the full model is the true model if $\zeta \neq 0$. To estimate the local parameter $\boldsymbol{\delta}$, the unbiased estimator $\hat{\boldsymbol{\delta}} = \sqrt{n}\left(\hat{\boldsymbol{\gamma}}_{full} - \boldsymbol{\gamma}_0\right)$ is used for simplicity. The parameter of interest is defined to be the vector of free parameters in the full model, excluding the thresholds.

In order to examine the effects of the local asymptotic framework, data are also generated from the standard asymptotic framework, where $\gamma_{12} = \zeta$, $\gamma_{21} = 0.75\zeta$, and $b_{21} = 0.5\zeta$. The local asymptotic framework coincides with the standard asymptotic framework when $n = 150$ but is different when $n = 350$ and $n = 600$.

The methods considered here include FMA, full model, and model selection. Three implementations of FMA will be considered. The first implementation, denoted by FMAord, is the ordinal data FMA proposed in the current study by accounting for $\mathrm{cov}(\boldsymbol{M}, \boldsymbol{D}^T)$. The second implementation, denoted by FMAordcont, uses the weight estimation method and confidence interval from Jin and Ankargren (2019) (i.e., treating $\mathrm{cov}(\boldsymbol{M}, \boldsymbol{D}^T) = \boldsymbol{0}$), but ordinal data are still treated as ordinal. Comparing FMAord with FMAordcont allows us to examine the consequence of ignoring a nonzero $\mathrm{cov}(\boldsymbol{M}, \boldsymbol{D}^T)$. It is of interest to see whether ordinal data can be treated as continuous in the context of FMA, since it is common that the applied researchers treat ordinal data as continuous. Hence, the third implementation, denoted by FMAcont, treats ordinal data as continuous and directly uses the results from Jin and Ankargren (2019). Since least-squares are used, information criterion is not used for model selection. Rather, we start with whether the narrow model fits the data well by the robust RMSEA (Brosseau-Liard et al., 2012) from lavaan (Rosseel, 2012). In the current simulation, an RMSEA no higher than 0.05 is an indication of a good fit. If the narrow model does not fit the data well, Model 2 is under investigation. If Model 2 does not fit the data well, Model 3 is under investigation. If Model 3 does not fit the data well, the full model is chosen.

All candidate models are estimated in lavaan (Rosseel, 2012) using DWLS. The FMA estimator, confidence interval, and test statistics are programmed in R Core Team (2020). The code can be retrieved from the online appendix.

3.2. Simulation Results

It is likely to encounter non-convergence or Heywood cases when estimating the candidate models. The candidate models that are not converged or have non-positive definite covariance matrices are regarded as inadmissible. The inadmissible candidate models are removed from further analysis. If either the narrow model or full model are inadmissible, the corresponding FMA

replication is also considered as inadmissible. Further, outliers are encountered in the current simulation. For simplicity, replications with MSE values that are twice higher than the 99% sample percentile are considered as possible outliers and are removed from further analysis. When $n = 150$, at most 0.4% and 1.2% replications are removed if the thresholds are moderately asymmetric and extremely asymmetric, respectively. When $n = 600$, at most 0.02% replications are removed. Hence, the percentage of inadmissible solutions are not tabulated here.

To compare the finite sample performance of FMA with model selection and the full model, we compute the normalized MSE, defined as the average of MSE of some method relative to the infeasible minimum MSE across all candidate models, i.e.,

$$\frac{1}{R} \sum_{r=1}^{R} \frac{\text{MSE} \left(\hat{\boldsymbol{\mu}}_r - \boldsymbol{\mu}_{true} \right)^T \left(\hat{\boldsymbol{\mu}}_r - \boldsymbol{\mu}_{true} \right) \text{ at iteration } r}{\text{Minimum MSE of all admissible candidate models at iteration } r},$$

with R being the number of replications. We also compute the average of absolute bias $p^{-1} \sum_{i=1}^{p} |\text{median of } \{\hat{\mu}_{i,r} - \mu_{i,true}; \ r = 1, \cdots R\}|$, where $\hat{\mu}_{i,r}$ is the estimate of μ_i at iteration r, $\mu_{i,true}$ is the ith entry in $\boldsymbol{\mu}_{true}$, and p is the number of parameters. Both the normalized MSE and the averaged bias for moderately asymmetric thresholds are illustrated in Fig. 1. The pattern when $n = 600$ is similar to $n = 350$ and the pattern for extremely asymmetric thresholds is similar to that for moderately asymmetric thresholds. Hence, they are not reported here due to space limitation. The conclusions that we can draw from Fig. 1 are similar to those in Jin and Ankargren (2019). FMAord tends to yield a lower normalized MSE than the full model, especially when the parameter value is small. Model selection performs well with the lowest normalized MSE when the parameter value is low, whereas it performs the worst when the parameter value is large. This is due to the fact that RMSEA often picks the correct model if the narrow model is the true model, and that RMSEA often implies that a too simple model fits the data well enough. As expected, the cost of MSE reduction is the inclusion of bias. Nevertheless, the induced averaged bias is generally low (Fig. 1), comparing with model selection. It is also seen from Fig. 1 that FMAord tends to have a slightly lower normalized MSE but a higher absolute bias than FMAordcont. A lower normalized MSE is in line with our expectation, since FMAord aims to minimize the correctly derived asymptotic MSE. Further, FMAcont has a higher normalized MSE and a higher absolute bias than FMAord and FMAordcont. Results not presented here show that the normalized MSE and the absolute bias of FMAcont are even higher when the thresholds are extremely asymmetric.

To investigate the coverage probability, the probability of covering $\gamma_{11} = 0.5$ when the thresholds are moderately asymmetric is used as an illustration in Table 1. It is seen that the model selection interval is generally accurate when the narrow model is the true model, but is greatly undercovered when the full model is the true model. It is also seen that the FMAord interval performs similar to the full model interval, which is close to the nominal coverage probability 95%. These observations are also in line with the findings in Jin and Ankargren (2019). Further, FMAordcont yields a lower coverage probability than FMAord, suggesting that it is important to account for the correlation between $\boldsymbol{M}$ and $\boldsymbol{D}$ when constructing confidence intervals. It is interesting to see that the coverage probability of the FMAcont interval is close to 95% when the thresholds are moderately asymmetric. However, the coverage probability of the FMAcont interval tends to be lower than 95% when the thresholds are extremely asymmetric.

Regarding the goodness-of-fit tests, we only consider the full model, FMAord, and FMAordcont. FMAcont is not considered since the Chi-square test derived in Jin and Ankargren (2019) requires normally distributed data and no robust corrections are derived yet. Table 2 tabulates the empirical rejection rate of the mean-scaled statistic and the mean-and-variance adjusted statistic at the significance level 0.05 when the thresholds are moderately asymmetric. Results for extremely

(a) $n = 150$

(b) $n = 350$, local asymptotic
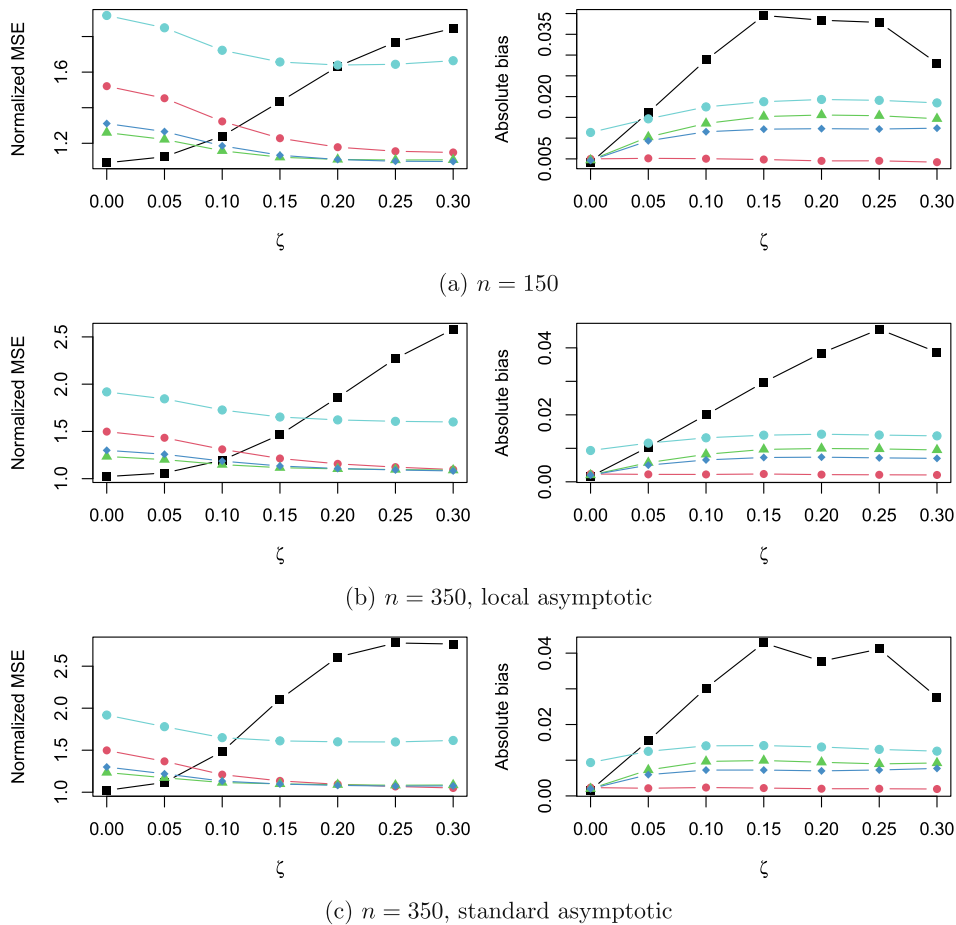

(c) $n = 350$, standard asymptotic

FIGURE 1.
Normalized mean squared error (MSE) and averaged absolute bias of model selection (black square), the full model (red dot), FMAord (green triangle), FMAordcont (blue diamond), and FMAcont (cyan dot) (Color figure online).

asymmetric thresholds show similar patterns. Hence, they are not reported here. Since the full model is correctly specified, we expect the empirical rejection rate to be approximately 0.05. It is seen that the FMA goodness-of-fit test statistic performs similar to the goodness-of-fit test statistic of the full model, which is in line with Theorem 3 that the goodness-of-fit test statistics are asymptotically equivalent. Comparing the mean-scaled statistic with the mean-and-variance adjusted statistic, the latter tends to have a better size. It is also seen that FMAord and FMAord-cont often yield similar empirical sizes, indicating that the effect of $\text{cov}(\boldsymbol{M}, \boldsymbol{D}^T)$ is minor when it comes to the goodness-of-fit tests.

4. Empirical Example

In this section, an empirical example is analyzed as an illustration. In order to study the supplier–customer relationship, Selnes and Sallis (2003) sampled 780 Scandinavian companies that have more than 50 employees. A total of 665 of them participated in the study. A total of 315 dyads in the sense of supplier and customer are identified. Selnes and Sallis (2003) used

TABLE 1.
Coverage probabilities of covering $\gamma_{11} = 0.5$ of different methods at the nominal level 95%.

Framework	n	Method	ζ						
			0.00	0.05	0.10	0.15	0.20	0.25	0.30
Moderately asymmetric thresholds									
Local	150	Model selection	0.92	0.89	0.83	0.75	0.72	0.76	0.81
		Full model	0.92	0.92	0.93	0.92	0.92	0.93	0.93
		FMAord	0.92	0.93	0.93	0.93	0.93	0.93	0.93
		FMAordcont	0.84	0.84	0.84	0.84	0.84	0.84	0.84
		FMAcont	0.93	0.93	0.93	0.93	0.94	0.94	0.94
	600	Model selection	0.94	0.92	0.84	0.72	0.55	0.38	0.30
		Full model	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		FMAord	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		FMAordcont	0.86	0.86	0.86	0.86	0.86	0.85	0.85
		FMAcont	0.93	0.93	0.93	0.93	0.94	0.94	0.94
Standard	600	Model selection	0.94	0.84	0.55	0.30	0.53	0.80	0.85
		Full model	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		FMAord	0.94	0.94	0.94	0.94	0.94	0.94	0.94
		FMAordcont	0.86	0.86	0.86	0.85	0.85	0.85	0.85
		FMAcont	0.93	0.93	0.94	0.94	0.94	0.94	0.93
Extremely asymmetric thresholds									
Local	150	Model selection	0.92	0.90	0.85	0.76	0.71	0.70	0.73
		Full model	0.93	0.93	0.93	0.93	0.93	0.93	0.93
		FMAord	0.93	0.93	0.93	0.93	0.93	0.93	0.93
		FMAordcont	0.86	0.86	0.85	0.85	0.85	0.85	0.86
		FMAcont	0.90	0.90	0.91	0.91	0.91	0.91	0.91
	600	Model selection	0.95	0.93	0.87	0.77	0.62	0.45	0.32
		Full model	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		FMAord	0.94	0.95	0.95	0.95	0.95	0.95	0.95
		FMAordcont	0.86	0.86	0.86	0.86	0.86	0.86	0.86
		FMAcont	0.88	0.88	0.88	0.89	0.89	0.89	0.90
Standard	600	Model selection	0.95	0.87	0.62	0.32	0.33	0.62	0.78
		Full model	0.95	0.95	0.95	0.95	0.95	0.95	0.95
		FMAord	0.94	0.95	0.95	0.95	0.95	0.95	0.94
		FMAordcont	0.86	0.86	0.86	0.86	0.86	0.86	0.86
		FMAcont	0.88	0.88	0.89	0.90	0.90	0.90	0.91

a subset of this data set to study how the learning capacity of supplier–customer relationship can be promoted by management. Recently, Sallis (2018) used another subset with 303 dyads to study the effect of relationship flexibility to relationship performance. Two five-factor models under consideration are shown in Fig. 2, which are simplified from Sallis (2018). The narrow model omits the paths from *Goal congruence* to *Relationship performance* and *Environmental uncertainty* to *Relationship performance*, whereas the full model also estimates such two paths. The sample size remained after deleting missing values is $n = 266$. All indicators are measured on the 7-point Likert scale. They are aggregated into a 3-point Likert scale that is close to the 3-category extreme asymmetry setting in Rhemtulla et al. (2012). The focus parameter is defined to be the vector of all free parameters needed for the model-implied covariance matrix.

Table 3 tabulates the estimated effects of *Goal congruence* and *Environmental uncertainty* on *Relationship performance*, *Coordination effort*, and *Flexibility*. Model selection chooses the nar-

TABLE 2.
Empirical rejection rate of the goodness-of-fit test statistics at the significance level 0.05, when the thresholds are moderately asymmetric.

Framework	n	Method	ζ						
			0.00	0.05	0.10	0.15	0.20	0.25	0.30
Mean-scaled statistic									
Local	150	Full model	0.133	0.126	0.123	0.115	0.115	0.111	0.109
		FMAord	0.141	0.134	0.133	0.128	0.127	0.125	0.123
		FMAordcont	0.141	0.133	0.132	0.124	0.125	0.122	0.120
	350	Full model	0.103	0.101	0.099	0.099	0.095	0.092	0.088
		FMAord	0.106	0.104	0.102	0.105	0.101	0.098	0.094
		FMAordcont	0.106	0.104	0.101	0.103	0.099	0.096	0.092
	600	Full model	0.092	0.092	0.090	0.092	0.091	0.085	0.084
		FMAord	0.093	0.094	0.091	0.094	0.093	0.089	0.088
		FMAordcont	0.093	0.093	0.091	0.093	0.092	0.087	0.086
Standard	350	Full model	0.103	0.099	0.098	0.092	0.088	0.086	0.083
		FMAord	0.106	0.102	0.103	0.097	0.095	0.092	0.090
		FMAordcont	0.106	0.102	0.102	0.095	0.093	0.090	0.088
	600	Full model	0.092	0.090	0.091	0.084	0.081	0.079	0.078
		FMAord	0.093	0.091	0.093	0.088	0.085	0.083	0.082
		FMAordcont	0.093	0.091	0.092	0.086	0.084	0.081	0.081
Mean-and-variance adjusted statistic									
Local	150	Full model	0.068	0.065	0.061	0.060	0.059	0.057	0.057
		FMAord	0.069	0.066	0.064	0.063	0.062	0.061	0.062
		FMAordcont	0.068	0.066	0.062	0.061	0.060	0.060	0.060
	350	Full model	0.058	0.056	0.059	0.057	0.055	0.052	0.052
		FMAord	0.059	0.057	0.060	0.058	0.056	0.054	0.054
		FMAordcont	0.059	0.057	0.059	0.058	0.056	0.053	0.052
	600	Full model	0.054	0.054	0.054	0.054	0.052	0.051	0.050
		FMAord	0.055	0.054	0.055	0.055	0.053	0.053	0.051
		FMAordcont	0.054	0.054	0.054	0.054	0.053	0.052	0.051
Standard	350	Full model	0.058	0.056	0.056	0.053	0.049	0.048	0.048
		FMAord	0.059	0.057	0.058	0.054	0.053	0.051	0.050
		FMAordcont	0.059	0.057	0.057	0.054	0.050	0.050	0.049
	600	Full model	0.054	0.054	0.052	0.050	0.048	0.048	0.049
		FMAord	0.055	0.055	0.053	0.051	0.050	0.051	0.052
		FMAordcont	0.054	0.054	0.053	0.051	0.049	0.049	0.050

row model, since it yields satisfactory fit indices. Nevertheless, all FMA implementations assign non-ignorable weights to the full model, which can be seen from the estimates of the path *Goal congruence* to *Relationship performance*: 0.33 (=0.027/0.083) for FMAord, 0.42 (=0.034/0.083) for FMAordcont, and 0.70 (=0.067/0.083) for FMAcont. FMAord and FMAordcont produce similar point estimates, which can be largely different from those produced by FMAcont. This is in line with our observations from Fig. 1 in the simulation study that FMAcont can be much more biased than FMAord and FMAordcont. Despite similar point estimates, it is also seen that the FMAordcont intervals are generally nested in the FMAord interval, which may lead to a lower coverage probability that we observed in the simulation study. Further, the fit indices and the confidence intervals of the full model are similar to those of FMAord, which is in line with our theoretical results.
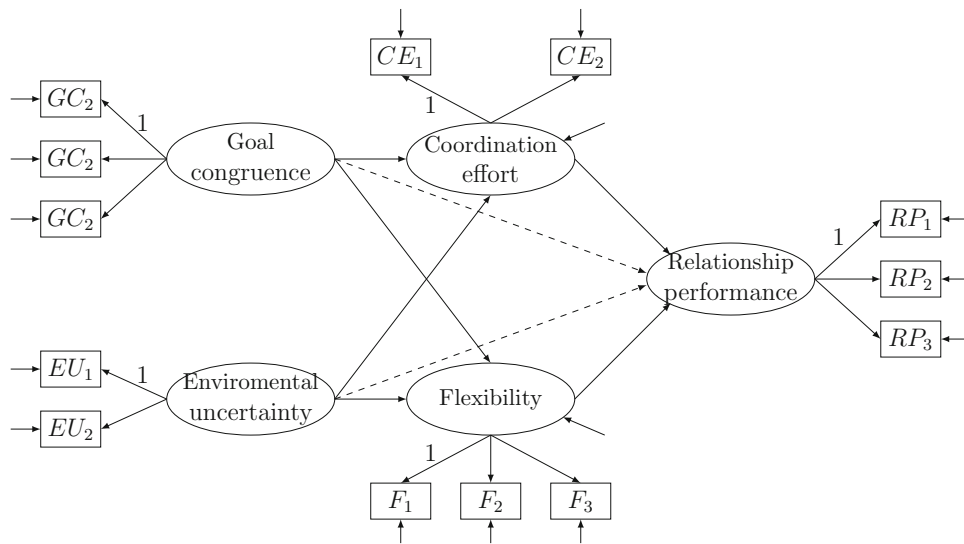
FIGURE 2.
Path diagonal of the seller example. The dashed lines are present in the full model but is omitted in the narrow model.

5. Conclusion and Discussion

In this study, FMA is generalized to the SEM with ordinal indicators. We showed that one assumption in Jin and Ankargren (2019), namely $\mathrm{cov}(\boldsymbol{M}, \boldsymbol{D}^T) = \boldsymbol{0}$, is violated in SEM with ordinal data. Hence, the results in Jin and Ankargren (2019) for SEM with continuous data need to be revised for SEM with ordinal data. To this end, we derived the correct criterion function for weight estimation, and the valid confidence interval for SEM with ordinal data. To evaluate the global fit, a mean-scaled test statistic and a mean-variance adjusted test statistic are proposed. In the simulation study, we showed that the ordinal data cannot always be treated as continuous, since FMAcont can yield much more biased estimators. We also showed that FMAordcont generally yields similar point estimators to FMAord. However, the FMAordcont interval can be undercovered. Hence, FMAord is still preferred.

Similar to FMA for SEM with continuous indicators, FMA does not uniformly dominate model selection nor the full model in our simulation. The same phenomenon has also been observed in various other models (e.g., Wan et al., 2014; Wang and Zou, 2012; Yang, 2003). This is a general issue for FMA, which is closely related with the combination puzzle (Claeskens et al., 2016) in the forecasting literature. The asymptotic MSE that we aim to minimize is derived under the assumption that the weights are fixed. However, the weights are generally random, if they are estimated from data. Since the uncertainty in the random weights is not accounted for when computing the asymptotic MSE, there is no guarantee that the FMA estimator will be dominating (Claeskens et al., 2016). Nevertheless, as Jin and Ankargren (2019) suggested in the context of SEM with continuous indicators, FMA is a robust compromise between model selection and the full model. Estimators followed by model selection can be unstable (Breiman, 1996) and neither the bias nor the MSE are necessarily bounded (Leeb & Pötscher, 2005). The full model estimator is often unstable (Hjort & Claeskens, 2003a) due to the presence of small parameters. FMA, on the other hand, tends to produce a robust MSE also for SEM with ordinal indicators. Future research is needed to provide guidelines on this matter.

The confidence interval considered in this paper is the Hjort and Claeskens (2003a) type, which is asymptotically equivalent to the full model confidence interval. Various other model-

TABLE 3.
Estimated effects of the latent exogenous latent variables on the endogenous latent variables of the supplier–customer relationship example.

Method	Path		Estimate
	From	To	
Model selection	Goal congruence	Rel. performance	-
		Coordination effort	0.471 (0.250, 0.691)
		Flexibility	0.349 (0.150, 0.548)
	Env. uncertainty	Rel. performance	-
		Coordination effort	0.518 (0.259, 0.777)
		Flexibility	0.467 (0.229, 0.706)
Full model	Goal congruence	Rel. performance	0.083 (−0.225, 0.391)
		Coordination effort	0.470 (0.240, 0.700)
		Flexibility	0.352 (0.147, 0.556)
	Env. uncertainty	Rel. performance	0.165 (−0.276, 0.605)
		Coordination effort	0.528 (0.253, 0.803)
		Flexibility	0.463 (0.221, 0.705)
FMAord	Goal congruence	Rel. performance	0.027 (−0.225, 0.391)
		Coordination effort	0.471 (0.237, 0.696)
		Flexibility	0.350 (0.145, 0.554)
	Env. uncertainty	Rel. performance	0.054 (−0.275, 0.605)
		Coordination effort	0.521 (0.261, 0.810)
		Flexibility	0.466 (0.227, 0.710)
FMAordcont	Goal congruence	Rel. performance	0.034 (−0.167, 0.333)
		Coordination effort	0.471 (0.288, 0.647)
		Flexibility	0.350 (0.206, 0.494)
	Env. uncertainty	Rel. performance	0.068 (−0.209, 0.538)
		Coordination effort	0.522 (0.295, 0.774)
		Flexibility	0.465 (0.274, 0.662)
FMAcont	Goal congruence	Rel. performance	0.067 (−0.079, 0.268)
		Coordination effort	0.456 (0.260, 0.649)
		Flexibility	0.270 (0.122, 0.414)
	Env. uncertainty	Rel. performance	0.131 (−0.056, 0.428)
		Coordination effort	0.471 (0.247, 0.710)
		Flexibility	0.387 (0.207, 0.571)

Env. uncertainty = Environmental uncertainty. Rel. performance = Relationship performance. The figures in the parentheses are the 95% confidence intervals. For the narrow model, CFI=0.997, TLI=0.997, RMSEA=0.030. For the full model, CFI=0.997, TLI=0.996, RMSEA=0.032. For FMAord, CFI=0.997, TLI=0.996, RMSEA=0.032. For FMAordcont, CFI=0.997, TLI=0.996, RMSEA=0.032. The reported fit indices are robust indices of Brosseau-Liard et al. (2012) and Brosseau-Liard and Savalei (2014).

averaging confidence interval have also emerged in the literature, such as (Fletcher and Dillingham (2011), Fletcher and Turek (2011), and Turek and Fletcher (2012)). Their properties have been investigated by Kabaila et al. (2016), Kabaila et al. (2017), and Kabaila (2018). Nevertheless, it is difficult to outperform the full model confidence interval (Kabaila et al., 2016). Wang and Zou (2012) suggested the use of the full model interval since it is computationally easy and the FMA interval does not offer a major improvement. Following these suggestions, both the full model interval and the FMA interval will offer valid inference in practice. Even if the point estimator is taken from one candidate model, Jin and Ankargren (2019) still suggested to use the

full model interval or the FMA interval to take the selection uncertainty into consideration. The same suggestion applies to the SEM with ordinal data.

One limitation of the proposed goodness-of-fit statistic is that the focus parameter $\boldsymbol{\mu}$ ought to be the vector of all free parameters in $\boldsymbol{\Sigma}$, an assumption needed for Theorem 3. If $\boldsymbol{\mu} \neq \boldsymbol{\beta}$ such as $\boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Gamma}$, we can still obtain valid FMA estimate $\bar{\boldsymbol{\mu}}$ and construct valid confidence intervals for $\boldsymbol{\mu}$. However, the proposed test statistic is not guaranteed to be asymptotically equivalent to the full model test statistic. Further studies will be devoted to the model evaluation for a general parameter vector of interest.

Acknowledgments

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Ankargren, S., & Jin, S. (2018). On the least squares model averaging interval estimator. *Communications in Statistics: Theory and Methods, 47*, 118–132.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics, 24*, 2350–2383.

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research, 49*, 460–470.

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research, 47*, 904–930.

Browne, M. W. (1984). Asymptotically distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Claeskens, G., Magnus, J., Vasnev, A. L., & Wang, W. (2016). The Forecast Combination Puzzle: A Simple Theoretical Explanation. *International Journal of Forecasting, 32*, 754–762.

Fletcher, D. (2018). *Model averaging*. Berlin, Germany: Springer.

Fletcher, D., & Dillingham, P. W. (2011). Model-averaged confidence intervals for factorial experiments. *Computational Statistics and Data Analysis, 55*, 3041–3048.

Fletcher, D., & Turek, D. (2011). Model-averaged profile likelihood intervals. *Journal of Agricultural, Biological and Environmental Statistics, 17*, 38–51.

Gong, G., & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics, 9*, 861–869.

Hjort, N. L., & Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association, 98*, 879–899.

Hjort, N. L., & Claeskens, G. (2003). Rejoinder. *Journal of the American Statistical Association, 98*, 938–945.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with discussion). *Statistical Science, 14*, 382–417.

Jin, S., & Ankargren, S. (2019). Frequentist model averaging in structural equation modeling. *Psychometrika, 84*, 84–104.

Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustnesss study of the polychoric correlation estimation. *Psychometrika, 82*, 67–85.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381–389.

Kabaila, P. (2018). On the minimum coverage probability of model averaged tail area confidence intervals. *The Canadian Journal of Statistics, 46*, 279–297.

Kabaila, P., & Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *Journal of the American Statistical Association, 101*, 619–629.

Kabaila, P., Welsh, A. H., & Abeysekera, W. (2016). Model-averaged confidence intervals. *Scandinavian Journal of Statistics, 43*, 35–48.

Kabaila, P., Welsh, A. H., & Mainzer, R. (2017). The performance of model averaged tail area confidence intervals. *Communications in Statistics - Theory and Methods, 46*, 10718–10732.

Leeb, H., & Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory, 21*(1), 21–59.

Liu, C.-A. (2015). Distribution theory of the least squares averaging estimator. *Journal of Econometrics, 186*, 142–159.

Lubke, G. H., Campbell, I., McArtor, D., Miller, P., Luningham, J., & van den Berg, S. M. (2017). Assissing model selection uncertainty using a bootstrap approach: an update. *Structural Equation Modeling: A Multidisciplinary Journal, 24,* 230–245.

Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association, 89*, 1535–1546.

Monroe, S. (2017). Contributions to estimation of polychoric correlation estimation. *Multivariate Behavioral Research, 53*, 247–266.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Muthén, B. (1984). A general structural equation model with dichotomous ordered categorical and continuous latent indicators. *Psychometrika, 49*, 115–132.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Retrieved from https://www.statmodel.com/download/Article$_075$.pdf. Accessed by 2013-09-12.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods, 17*, 1–14.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rhemtulla, M., Brosseau-Liard, P. E., & 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.

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36.

Sallis, J. (2018). *Relationship flexibility in buyer-seller partnerships*. Unpublished manuscript, Department of Business Studies, Uppsala University.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. V. Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA, US: SAGE Publications.

Selnes, F., & Sallis, J. (2003). Promoting relationship learning. *Journal of Marketing, 67*, 80–95.

Turek, D., & Fletcher, D. (2012). Model-averaged Wald confidence intervals. *Computational Statistics and Data Analysis, 56*, 2809–2815.

Wan, A. T. K., Zhang, X., & Wang, S. (2014). Frequentist model averaging for multinomial and ordered logit models. *International Journal of Forecasting, 30*, 118–128.

Wang, H., & Zhou, S. Z. F. (2013). Interval estimation by frequentist model averaging. *Communications in Statistics: Theory and Methods, 42*, 4342–4356.

Wang, H., & Zou, G. (2012). Model averaging for varying-coefficient partially linear measurement error models. *Electronic Journal of Statistics, 6*, 1017–1039.

Yang, Y. (2003). Regression with multiple candidate models: selecting or mixing? *Statistica Sinica, 13,* 783–809.