STRUCTURAL EQUATION MODELING
2025, VOL. 32, NO. 3
ISSN: 1070-5511 | E-ISSN: 1532-8007

The Impact of Non-Normality on Causal Effect Estimates in Structural Equation Models with Ordered Categorical Variables

Benedikt Lugauer and Jana Holtmann

Leipzig University

ABSTRACT

Multilevel structural equation models (SEM) are widely used for estimating causal effects in social and behavioral sciences. However, the impact of non-normality on these estimates, particularly with ordered categorical variables, remains unclear. This study investigates how non-normality affects causal effect estimation in path and multigroup SEMs with ordered categorical variables. Using the Vine-to-Anything approach, we simulated multivariate distributions by independently varying marginals and copulas. We compared maximum likelihood (ML), robust ML, and diagonally weighted least squares (DWLS) estimators across various non-normal conditions in three simulation studies. Results show that bias in effect estimates critically depends on the combination of copulas and marginal distributions in treatment and control groups. DWLS generally outperformed ML, especially for average treatment effects and interaction effects. However, all methods were sensitive to distributional misspecifications. These findings highlight the importance of considering distributional assumptions when using ordered categorical data in causal inference studies.

KEYWORDS

Causal effects; copulas; multigroup structural equation modeling; non-normality; ordered categorical data

Assessing the effect of a treatment on an outcome is a fundamental goal of empirical research across various scientific disciplines including psychology, medicine, epidemiology, political science, and economics (Hernán & Robins, 2020). The average treatment effect (ATE) measures the impact of a treatment on an outcome and is central to uncovering causal relationships. Sometimes a more detailed understanding is needed, as the ATE may not capture the subtleties of different effects in different subpopulations or under different conditions. Conditional effects shed light on how treatment effects vary across subpopulations, while decomposing the total effect into direct and indirect (mediated) components can shed light on the mechanisms by which the treatment influences the outcome (Mayer et al., 2014).

Popular regression-based frameworks for estimating average, conditional, direct, and indirect treatment effects include Analysis of Covariance (ANCOVA, Pohl et al., 2009) and multigroup Structural Equation Modeling (SEM, Mayer et al., 2016; Sengewald & Mayer, 2024). One advantage of using a multigroup SEM approach is that both the outcome and the covariates can be modeled as latent variables, thereby accounting for measurement error (Mayer et al., 2016). However, standard SEM estimation approaches assume that the variables follow a multivariate normal distribution. This assumption can be problematic in empirical applications where variables are often non-normal. In addition, in the social and behavioral sciences, handling categorical variables is particularly crucial due to the prevalence of questionnaire items with ordered categorical response formats. Furthermore, the distributions of primary covariates in fields such as psychology, sociology or medicine (Boulton & Williford, 2018; Delucchi & Bostrom, 2004; Herschtal, 2023) tend to differ between treatment and control groups and tend to be skewed in one or both of these groups. For instance, depressive symptoms as measured by common depression inventories typically show skewed distributions (e.g., Fried et al., 2016; Tomitaka et al., 2015).

Research has shown that non-normality in data can lead to biased parameter estimates in SEM, particularly with ordered categorical variables. The level of bias depends on the degree of non-normality present and the estimator selected. For example, Flora and Curran (2004) found that ordinal SEM, when based on polychoric correlations that take into account the categorical nature of the variables, remains relatively robust even in the presence of moderate non-normality. Subsequent studies have consistently shown that Diagonally Weighted Least Squares (DWLS), based on polychoric correlations, tends to outperform Maximum Likelihood (ML) across various conditions with ordered categorical variables (Bandalos, 2014; Jin et al., 2016; Li, 2016b; Rhemtulla et al., 2012).

However, most studies investigating the effects of non-normality in ordinal data have used the Vale-Maurelli (VM) simulation method (Vale & Maurelli, 1983), which has recently been shown to generate ordered categorical data that is numerically equivalent to simulating from a multivariate normal vector (Grønneberg & Foldnes, 2019). Foldnes and Grønneberg (2019) introduced a new copula-based simulation methodology that ensures that the random variables' distributions are non-normal. Using this new approach, they show that the SEM parameter estimates and model fit indices are more sensitive to distributional misspecification than previous studies suggested (Foldnes & Grønneberg, 2019, 2022; Grønneberg & Foldnes, 2024). In addition, an understanding of the marginal distributions of each ordered categorical item is essential, as these affect the overall analysis and interpretation in factor models or SEM (Grønneberg & Foldnes, 2024).

The primary goal of this study is to investigate how non-normality affects the estimation of causal effects in multigroup SEMs with ordered categorical variables. Specifically, we aim to understand the extent to which different distributional misspecifications, through manipulation of both marginals and copulas, impact the bias in estimated causal effects. Given that the VM method does not generate truly non-normal data and is therefore unsuitable for studying the effects of non-normality in ordered categorical SEM, we utilize the Vine-to-Anything (VITA) algorithm (Grønneberg et al., 2022), which allows for the independent manipulation of marginals and bivariate copulas to simulate a wide range of multivariate distributions.

In an extensive simulation study, we assess the effect of varying multivariate (non-normal) distributions generated via VITA on the estimation of average, direct, and indirect treatment effect parameters in multigroup SEMs under different estimators and model specifications. We find that the degree of bias in effect estimates is significantly influenced by the interplay between multivariate distributions (copulas) and marginal distributions within treatment and control groups. These findings have important implications for researchers who seek to accurately estimate causal effects under various distributional conditions. Our results contribute to the ongoing discussion surrounding the robustness of estimation methods in the presence of non-normal data. Ultimately, this study aims to provide researchers with a better understanding of the limitations and strengths of different estimation methods in multigroup SEM and treatment effect estimation.

In the following, we provide a short overview of the definition of the investigated treatment effects, multigroup SEMs, previous results on the effect of non-normality on SEM estimation, as well as the vine copula methodology, particularly the VITA method (Grønneberg et al., 2022), before presenting the simulation study and its implications.

1. Stochastic Theory of Causal Effects

We rely on the stochastic theory of causal effects (Steyer et al., 2014) to define average, conditional, direct, and indirect effects, as it allows for the incorporation of latent variables into a causal theory within SEM (Steyer, 2005).

The stochastic theory of causal effects defines theoretical quantities using a random experiment. To illustrate this, consider a simplified random experiment involving a mediation process. Let Y be an outcome variable and X a treatment variable that can assume values 0 or 1. Additionally, consider multivariate vectors of covariates $\boldsymbol{Z}$ and $\boldsymbol{M}$, where $Y, \boldsymbol{Z}$, and $\boldsymbol{M}$ follow any distribution $\mathcal{D}$ within each treatment category X. Temporal ordering is crucial for establishing causal relationships: $\boldsymbol{Z}$ is observed prior or simultaneously to the treatment X, while $\boldsymbol{M}$ is observed post-treatment, creating a causal pathway. The total-effect true outcome variable τ_x is then defined as $\tau_x := \mathbb{E}(Y|X = x, C_X)$, where C_X includes all covariates prior or simultaneous to X, including a baseline measure Y^{pre} of the outcome variable's initial state (Mayer et al., 2014). The direct-effect true outcome variable τ_{x, t_M} is defined as $\tau_{x, t_M} := \mathbb{E}(Y|X = x, C_{X, t_M})$, where C_{X, t_M} includes covariates up to the sampling time of $\boldsymbol{M}$. The atomic total-effect variable is $\delta_{10} = \tau_1 - \tau_0$, and the atomic direct-effect variable is $\delta_{10, t_M} = \tau_{1, t_M} - \tau_{0, t_M}$. The atomic indirect-effect variable is then simply $\delta_{10} - \delta_{10, t_M}$.

We can define meaningful and estimable causal effects using true outcome variables: The Average Treatment Effect (ATE) represents the average of the atomic total-effect variable and is defined as the difference between the expected outcomes under the two treatment conditions: $\text{ATE}_{10} = \mathbb{E}(\delta_{10}) = \mathbb{E}(\tau_1) - \mathbb{E}(\tau_0)$. To identify the Average Direct Effect (ADE), we condition on post-treatment variables $\boldsymbol{M}$. This gives us $\text{ADE}_{10} = \mathbb{E}(\delta_{10, t_M}) = \mathbb{E}(\tau_{1, t_M}) - \mathbb{E}(\tau_{0, t_M})$, which represents the expected difference in outcomes directly attributable to the treatment, after accounting for the intermediate variable $\boldsymbol{M}$. The Average Indirect Effect (AIE) captures the effect of the treatment that is mediated through the post-treatment variable $\boldsymbol{M}$. It is computed as the difference between the ATE and the ADE: $\text{AIE}_{10} = \text{ATE}_{10} - \text{ADE}_{10}$. Furthermore, more granular causal effects can be examined by conditioning on specific covariates $\boldsymbol{W}$. This conditional effect is defined as $g_{10}(\boldsymbol{W}) = \mathbb{E}(Y|X = 1, \boldsymbol{W}) - \mathbb{E}(Y|X = 0, \boldsymbol{W})$, allowing for an understanding of how the treatment effect varies with different values of $\boldsymbol{W}$.

These causal effect quantities are estimable under specific causality conditions, such as the independent cause or regressively independent outcome condition (Mayer et al., 2014). A detailed discussion of these conditions is beyond the scope of this work, and readers are referred to Mayer et al. (2014) or Steyer et al. (2014) for a formal treatment of causal effect definitions and causality conditions. The gold standard to guarantee causal unbiasedness of the ATE can be achieved by a randomized controlled trial. However, when it comes to direct and indirect effects, the exclusion of confounding variables can result in biased estimates for both types of effects, even in randomized controlled trials (Mayer et al., 2014). The aforementioned causality conditions relate to covariates and can be satisfied through methods such as propensity score matching or instrumental variable methods and regression-based approaches (Angrist et al., 1996; Austin, 2011). A more detailed explanation of the stochastic theory of causal effects is given in OSM2.

1.1 Estimation Methods

Multigroup SEM has gained considerable traction over the past decade for the estimation of causal effects (Mayer et al., 2016; Sengewald & Mayer, 2024; Steyer, 2005). This methodology allows researchers to investigate measurement invariance of constructs across groups and explore complex relationships among observed and latent variables while taking group differences into account. Furthermore, multigroup SEM facilitates the calculation of both direct and indirect effects. A general formulation of a multi-group SEM model can be represented as follows:

$$\boldsymbol{Y} = \boldsymbol{\gamma}^{(k)} + \Lambda^{(k)}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{1}$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \Theta^{(k)}) \tag{2}$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha}^{(k)} + \mathbf{B}^{(k)}\boldsymbol{\eta} + \boldsymbol{\zeta} \tag{3}$$

$$\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \Psi^{(k)}) \tag{4}$$

In the measurement model (1), for a specific group denoted by k, $\boldsymbol{\gamma}^{(k)}$ is a vector of measurement intercepts, $\Lambda^{(k)}$ is the matrix of factor loadings, $\boldsymbol{Y}$ is a vector of observed variables, and $\boldsymbol{\varepsilon}$ are measurement error variables with mean zero. In the structural model (3), for a specific group denoted by k, $\mathbf{B}^{(k)}$ is a matrix of structural regression weights with zero diagonal elements, and $\boldsymbol{\zeta}$ is a vector of structural residuals.

Consider as a special case a multigroup SEM model with $k = 2$ treatment groups depicted in Figure 4 in the Online Supplementary Material 1 (OSM1). In this simplified example, we assume that latent covariate ξ and latent outcome η are measured by three observed indicators each, Z_i and Y_i, respectively, with group-invariant measurement models:

$$\boldsymbol{Y} = \gamma_Y + \Lambda_Y \eta + \boldsymbol{\varepsilon}_Y \tag{5}$$

$$\boldsymbol{Z} = \gamma_Z + \Lambda_Z \xi + \boldsymbol{\varepsilon}_Z \tag{6}$$

and the structural equation is given by:

$$\begin{pmatrix} \eta \\ \xi \end{pmatrix} = \begin{pmatrix} \alpha^{(k)} \\ \mu^{(k)} \end{pmatrix} + \begin{pmatrix} 0 & \beta^{(k)} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \eta \\ \xi \end{pmatrix} + \begin{pmatrix} \zeta_0 \\ \zeta_1 \end{pmatrix} \tag{7}$$

The ATE is computed based on the difference of the expected conditional expectations $\mathbb{E}(\eta|\xi, X = 0) = \alpha^{(0)} + \beta^{(0)}\xi$ and $\mathbb{E}(\eta|\xi, X = 1) = \alpha^{(1)} + \beta^{(1)}\xi$, given by:

$$\text{ATE}_{10} = \mathbb{E}[\mathbb{E}(\eta|\xi, X = 1) - \mathbb{E}(\eta|\xi, X = 0)] \tag{8}$$

Beyond coarse aggregates such as the ATE, one can compute more granular causal effect functions by conditioning on one or more specific covariates, with $\boldsymbol{W}$ representing a subset of covariates of interest. When using regression-based methods like multigroup SEM, granular causal effects, such as the covariate-dependent treatment effect, are determined by interaction effects between the covariates $\boldsymbol{W}$ and the treatment X on the outcome Y. Therefore, it is important to focus on accurately estimating interaction effects.

1.2. Non-Normality

A fundamental assumption of standard ML estimation of SEMs is that the observed variables exhibit multivariate normality (Jöreskog, 1969). However, when ML is applied under the assumption of normality to model ordered categorical variables, this can lead to biased estimates of parameters, standard errors (SEs), and model fit statistics. The categorical nature of these variables and their inherent asymmetry contribute to these biases (Micceri, 1989; Yuan & Hayashi, 2006). This is also true for Robust Maximum Likelihood (MLR), which yields identical parameter estimates as ML but estimates robust standard errors by using a sandwich estimator. Robust standard errors are adjusted to account for violations of the aforementioned normality assumption, providing more reliable standard errors by correcting for model misspecification such as heteroscedasticity (Kline, 2023).

It is widely recognized that the non-normality of continuous variables adversely affects ML parameter estimates and their SEs with small to moderate sample sizes (Olsson et al., 2000). Consequently, various estimation methods have been developed to mitigate the effects of non-normality, such as MLR and DWLS, with DWLS specifically designed for ordered categorical data (Browne, 1984; Muthén, 1984).

DWLS effectively minimizes the discrepancies between observed and predicted polychoric correlation matrices by converting a latent continuous random vector, $\boldsymbol{Y}^*$, into an observed ordered categorical random vector, $\boldsymbol{Y}$, through a discretization process. Each random variable $Y \in \boldsymbol{Y}$ corresponds to a random variable in the continuous latent variable $\boldsymbol{Y}^*$ and assumes one of the discrete values $y_1, \ldots, y_K$ based on the relationship: $Y = y_k$ if $\kappa_{k-1} < Y^* \leq \kappa_k$, where the κ values serve as thresholds satisfying $\kappa_{j-1} \leq \kappa_j$. For distributions that are continuous and unbounded, the limits of these thresholds are $\kappa_0 = -\infty$ and $\kappa_K = \infty$. The marginals of $\boldsymbol{Y}^*$ are commonly assumed to be standard normal, $Y^* \sim N(0, 1)$ (Muthén & Asparouhov, 2002).

Previous research has investigated the consequences of violations of the normality assumption on the accuracy and the precision of model parameter estimation with various estimators including ML, MLR and DWLS. For categorical variables, the literature suggests that parameter estimates exhibit bias to varying degrees, contingent on both the degree of non-normality and the chosen estimator. Beaudcel and Herzberg (2006) demonstrated that, absent of model misspecification, DWLS surpasses ML on ordered categorical data. Categorical DWLS generally provides parameter estimates least impacted by non-normality (Bandalos, 2014; Li, 2016b). The results of comparing DWLS with MLR are mixed. Generally, under moderate to high non-normality, DWLS seems to provide less parameter bias than ML. However, in terms of SE estimates of the factor loadings and the interfactor correlations, MLR outperformed DWLS, especially with higher degrees of non-normality and smaller sample sizes (Bandalos, 2014; Li, 2016a, 2016b). Rhemtulla et al. (2012) show that the disparity between ML and DWLS in terms of parameter and SE bias becomes trivial when categorical variables have more than five categories. They recommend

DWLS with data having fewer than five categories and ML for five or more categories, especially with smaller sample sizes and approximately symmetric category thresholds.

Most studies that investigated the impact of non-normality used the VM simulation method (Grønneberg & Foldnes, 2019; Vale & Maurelli, 1983). The VM method is based on transforming the data vectors generated from a multivariate normal distribution by use of third degree polynomials and subsequently applying the discretization process. Grønneberg and Foldnes (2019) demonstrated that ordered categorical data produced via the VM method is numerically akin to simulating data from a multivariate normal vector. This discovery spurred a series of new investigations that reexamined the effect of non-normality. These investigations utilized a novel copula-based simulation methodology, ensuring that the random variables' distributions remained non-normal (Foldnes & Grønneberg, 2019, 2022; Grønneberg & Foldnes, 2024). The findings from these studies indicate that both SEM parameter estimates and the robustness of model fit indices are considerably more susceptible to distributional misspecification than previously believed. Specifically, non-normality can result in biased SEM estimates and elevated Type I error rates (Foldnes & Grønneberg, 2022).

1.3. Vine Copula Simulation Methodology

In this simulation study, we adopt a vine copula-based approach, specifically utilizing the VITA algorithm as implemented in the `covsim` package (Grønneberg et al., 2022). VITA utilizes vine copulas to generate multivariate distributions derived from a predefined covariance matrix, based on the provided marginal distributions and (conditional) bivariate copulas.

Central to VITA is the concept of a copula, a mathematical function that 'couples' or 'links' univariate marginal distributions to form a joint multivariate distribution. More formally, the copula function C maps the unit hypercube $[0,1]^d$, where d is the number of dimensions, to the interval [0, 1]. Each point in this d-dimensional hypercube corresponds to a specific set of variable outcomes. A copula is then defined as follows (Sklar, 1959):

$$C(u_1, u_2, \ldots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d),$$

where U_i, for $i = 1, \ldots, d$, are uniformly distributed random variables on the interval [0, 1], and $u_i \in [0,1]$ are their corresponding realizations. The copula function C describes the cumulative probability of each combination of variable outcomes occurring.

A central property of copulas is Sklar's theorem (Sklar, 1959), which states that for any multivariate joint distribution function F with continuous marginal distribution functions $F_1, F_2, \ldots, F_d$, there exists a unique copula C such that:

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)).$$

The theorem guarantees the existence and uniqueness of a copula for any multivariate distribution with continuous marginal distributions. This means that independent of how complex the dependence structure in the data may be, there is a copula that can perfectly capture it while separating it from the marginal distributions of each variable. This copula can belong to one of several copula families, which can represent a wide range of dependence structures, from linear to nonlinear relationships, and from tail independence to dependence. Each family has unique properties to handle variations in the type and strength of dependence among variables, allowing researchers to choose the most appropriate copula for their study.

Copulas from the elliptical copula family are often used because they include both Gaussian (normal) and t copulas. While normal copulas do not allow for tail dependencies, t copulas allow for symmetric tail dependencies by jointly capturing extreme values in both tails of the distribution, making them a valuable alternative in cases where tail behavior is an important consideration. The dependencies, both linear and nonlinear, are captured by copula parameters that determine the strength and type of dependence structure they represent. In the case of a normal copula, the dependence structure is captured by the ρ parameter. The bivariate normal copula is given by (Genest & Rivest, 1993; Hofert et al., 2018):

$$C_\rho(u_1, u_2) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \tag{9}$$

where Φ denotes the standard normal cumulative distribution function (CDF), Φ^{-1} is its inverse, and ρ is the correlation between the variables. Note that the bivariate normal copula $C_\rho(u_1, u_2)$ models the dependence structure between two random variables independent of the marginal distributions. The restriction of tail independence implies that extreme events in one variable do not correspond to extreme events in another. In an empirical context, we may find this limitation of normal copulas problematic because extreme events often co-occur, which can significantly affect parameter estimation. A copula that better captures tail dependencies in such data is the Clayton copula. It belongs to the family of Archimedean copulas (Genest & MacKay, 1986) and is specifically designed to model lower tail dependence, which refers to the tendency for low values in one variable to coincide with low values in another. Another commonly used copula is the Joe copula which can capture both upper and lower tail dependencies (Joe, 1993). Although copulas provide more flexibility in modeling joint probability distributions, their performance tends to degrade as the number of variables increases, making it more difficult to capture complex dependency structures. In addition, traditional copula-based methods do not allow for different dependency structures between different variables.

To address these limitations, vine copulas offer a more adaptable solution. Vine copulas decompose multivariate dependencies into simpler, bivariate components, each represented by a bivariate copula. This decomposition can be conceptualized as a sequence of nested trees, collectively known as a vine (Bedford & Cooke, 2001), where each tree reflects a distinct level of conditional dependence among the random variables. Formally, in a vine with p variables, the structure is represented as a series of connected trees $V = T_1, T_2, \ldots, T_{p-1}$, where the connections (or edges) in tree T_j become the nodes in the subsequent tree T_{j+1}. A specific type of vine structure, known as R-vines (regular vines), organizes the dependence structure among variables into this sequence of trees. In a R-vine, each tree captures a different level of conditional dependencies, with the edges representing bivariate copulas that describe the relationships between pairs of variables, which may be conditioned on other variables.

The VITA algorithm employs R-vine copulas to construct multivariate distributions that align with a user-specified covariance matrix Σ. The process begins by specifying the marginal distributions for each variable and selecting an R-vine structure that defines the dependencies among them. This structure is represented by a sequence of trees. VITA then calibrates the bivariate copulas associated with each edge in these trees, starting with the first tree, which contains the unconditional bivariate copulas between pairs of variables. For each copula, VITA adjusts the copula parameter so that the covariance of the resulting bivariate distribution matches the target covariance. After calibrating the bivariate copulas in the first tree, the algorithm proceeds to the subsequent trees, where the edges represent conditional dependencies, conditioned on other variables. The copula parameters are calibrated to match the corresponding elements of the covariance matrix. This process continues sequentially through each tree in the vine. The final outcome is a multivariate distribution that accurately reflects the desired covariance matrix Σ, constructed by combining the calibrated bivariate copulas according to the R-vine structure.

In summary, the VITA method (Grønneberg & Foldnes, 2017) uses vine copulas to construct a multivariate distribution by coupling prespecified marginal distributions and bivariate copulas given a target covariance matrix. The goal is to calibrate each pairwise bivariate distribution, and their dependence parameters, based on the provided target covariance matrix and marginals. These calibrated distributions are then combined to form the full vine distribution. VITA is described in detail in Grønneberg et al. (2022). An example illustrating the VITA method can be found in OSM1.

2. Methods

We investigated the impact of non-normality on causal effect estimation in multigroup SEMs in three separate simulation studies. Each study features distinct model specifications but shares the goal of evaluating ML, MLR, and DWLS estimators to estimate causal effects under different simulation parameters. These parameters include sample size, types of copulas, marginal distributions, number of categories, symmetry types (skewness) of the discretization process, and varying model parameters. A key aspect of our investigation focused on the interplay of different copulas and marginal distributions between treatment groups. The code used for the simulation studies is available at https://osf.io/47wup/

2.1. Evaluation Criteria and Data Generation

Across studies, we assessed estimator performance with relative bias and coverage. The relative bias depicts the average relative bias (ARB) across replications in a simulation condition. Coverage is the proportion of times the 95% confidence interval (CI) contained the true population value. For interpretation, an ARB value of less than 5% was considered to indicate trivial bias (termed unbiased in the following), between 5% and 10% as moderate bias, and greater than 10% as substantial bias, following the guidelines outlined by Curran et al. (1996). We deemed a 95% coverage rate between 92% and 98% acceptable.

In each of the simulation conditions, we generated 200 replications. The simulations were performed on resources provided by the Leipzig University Computing Center. The datasets were generated in the R computing environment (R Core Team, 2023), utilizing the `covsim` package (Grønneberg et al., 2022) for vine copula-based simulations and the `lavaan` (Rosseel, 2012) and `effectliteR` packages (Mayer et al., 2016) for causal inference in multigroup SEM.

2.2. Model Specification

Both Study 1 and Study 2 implemented multigroup SEMs with two treatment groups, including a covariate, a dependent variable, and an interaction effect between the treatment variable and the covariate. In Study 1, the dependent variable Y is a continuous variable which is modeled as a function of a latent covariate ξ and the dichotomous treatment variable X. The latent covariate ξ is measured by three ordered categorical indicators. In contrast, in Study 2, the dependent variable η is a latent variable that is measured by three categorical indicators. The covariate Z is a manifest continuous variable and the treatment variable X remains dichotomous. To ensure adequate filling of all bins under conditions of marginal misspecification, the size of the treatment effect was decreased (see Table 1). For Study 3, we adopted a mediation model as described in Mayer et al. (2014) to examine the effect of non-normality in the estimation of average total, direct, and indirect effects. This model represents a scenario where baseline measures of both the mediator M^{pre} and outcome Y^{pre} are correlated with their respective post-treatment counterparts. Study 3 A focused on the latent mediator M, while Study 3 B focused on the latent variable M^{pre}. Both variables were measured by three ordered categorical indicators. This setup allowed us to examine whether and how the latent baseline measure or mediator, captured by ordered categorical indicators, transmit bias due to the discretization process to the effect estimates.

In this randomized experimental design, the treatment variable X is uncorrelated with both Y^{pre} and M^{pre}, a design that guarantees an unbiased assessment of the treatment effect. The path diagrams for all three models are depicted in Figure 1. We selected sample sizes in each of the three simulation studies to align with scenarios typically encountered in empirical research, that is, small to medium sample sizes of $N = 100$ and $N = 250$ (per group). Factor loadings

Table 1. Simulation parameters.

	N_x	λ	Thresholds (# Bins)	C_x	$f_{Z^*;X_x}(Z^*)$	$\beta_{\xi 1}^{(0)}$	$\beta_{\xi 1}^{(1)}$	$\beta_{\xi X}$
Study 1	100	0.8	3 × symmetric (3, 4, 5)	$\mathcal{N}$	$\mathcal{N}(\mu_{X=0} \in \{0\}, \mu_{X=1} \in \{0\})$	0.3	0.5	0.2
	250	0.9	2 × left-skewed (4, 5)	C	$\text{Exp}\,(\lambda_{X=0} \in \{1\}, \lambda_{X=1} \in \{0.9, 1, 1.1\})$		0.7	0.4
			2 × right-skewed (4, 5)	J				

	N_x	λ	Thresholds (# Bins)	C_x	$f_{Y^*;X_x}(Y^*)$	$\beta_{Z1}^{(0)}$	$\beta_{Z1}^{(1)}$	β_{ZX}
Study 2	100	0.8	3 × symmetric (3, 4, 5)	$\mathcal{N}$	$\mathcal{N}(\mu_{X=0} \in \{0\}, \mu_{X=1} \in \{0.3\})$	0.3	0.5	0.2
	250	0.9	2 × left-skewed (4, 5)	C	$\text{Exp}\,(\lambda_{X=0} \in \{1.2\}, \lambda_{X=1} \in \{1.1, 1.2, 1.3\}))$		0.7	0.4
			2 × right-skewed (4, 5)	J				

	N_x	λ	Thresholds (# Bins)	C_x	$f_{Z^*;X_x}(Z^*)$	$\rho_{\text{Mpre;Ypre}}$	$\beta_{\text{Mpre;1}}^{(0)}$	$\beta_{\text{Mpre;1}}^{(1)}$	$\beta_{\text{Ypre;1}}^{(0)}$	$\beta_{\text{Ypre;1}}^{(1)}$	$\beta_{\text{M;1}}^{(0)}$	$\beta_{\text{M;1}}^{(1)}$
Study 3	100	0.8	3 × symmetric (3, 4, 5)	$\mathcal{N}$	$\mathcal{N}(\mu_{X=0} \in \{0\},$	0.3	0.8	0.8	0.75	0.75	0.5	0.5
	250	0.9	2 × left-skewed (4, 5)	C	$\mu_{X=1} \in \{0.3\})$							
			2 × right-skewed (4, 5)	J	$\text{Exp}\,(\lambda_{X=0} \in \{1.2\},$							
					$\lambda_{X=1} \in \{1.1, 1.2, 1.3\}))$							

Note. Crossed Simulation Conditions. Each row represents a set of conditions under which the simulations were conducted, with variations in sample sizes (N_x), regression coefficients ($\beta_{\xi 1}^{(x)}$ for Study 1, $\beta_{Z1}^{(x)}$ for Study 2) which determine the size of the interaction effect ($\beta_{\xi X}$ and β_{ZX}), factor loadings (λ), number of threshold types (symmetric, left-skewed, right-skewed), and the number of bins. The column labeled C_x specifies the types of copula distributions used to model the dependency structures among the indicators of the latent variables in the control and treatment groups for each study. Here, $\mathcal{N}$ represents the Normal copula, C represents the Clayton copula, and J represents the Joe copula. The column labeled $f_{Z \cdot X_x}(Z)$ or $f_{Y \cdot X_x}(Y)$ specifies the marginal distributions of the indicators for the latent variables (ξ or η) in the control and treatment groups for each study. For instance, in Study 1, either normal or exponential distributions were used with different parameters for the control ($X = 0$) and treatment groups ($X = 1$). For Studies 1 and 3, the ATE is set to 1, while the ATE for Study 2 can be derived from the expectations of the latent variables' distributions in the control and treatment group. Additionally, the table lists constants across all conditions for Study 3, including the correlation between baseline measures of the mediator and outcome ($\rho_{\text{Mpre;Ypre}}$) in the control and treatment group, regression coefficients for baseline measures of the mediator ($\beta_{\text{Mpre;1}}^{(0)}$ and $\beta_{\text{Mpre;1}}^{(1)}$) and the outcome ($\beta_{\text{Ypre;1}}^{(0)}$ and $\beta_{\text{Ypre;1}}^{(1)}$), and the effect of the mediator on the outcome ($\beta_{\text{M;1}}^{(0)}$ and $\beta_{\text{M;1}}^{(1)}$).

were set to either 0.8 or 0.9 to test the robustness of the model under varying indicator reliability. In Studies 1 and 2, the interaction effect was controlled by varying the regression coefficient of the covariate in the treatment group, resulting in an interaction effect of 0.2 or 0.4. All relevant simulation conditions and model parameters are provided in Table 1

2.3. Distributional Specification

In this study, we investigated the effects of different marginal distributions, copulas, and thresholds on the discretization of latent variables. We used the same thresholds for both treatment and control groups. However, the underlying latent distributions differed between groups due to treatment effect or marginal misspecification, which involved using different parameters within the same distribution type (e.g., different means in normal distributions or different rate parameters in exponential distributions).

Because the same thresholds were applied to latent distributions with differing parameters, the observed categorical distributions differed between groups. Specifically, the observed distributions are influenced by:

1. Latent Distributions: In our simulation conditions, both the treatment and control groups had latent variables following the same type of distribution—either normal or exponential. However, the parameters of these distributions could differ between groups. For example, the mean of the normal distribution or the rate parameter of the exponential distribution might vary between the treatment and control groups, leading to differences in the shape or location of the latent distributions and, consequently, in the observed categorical distributions after discretization.
2. Thresholds: We employed two types of thresholds in the discretization process:
   - Symmetric Thresholds: Thresholds set at quantiles that divide a normal distribution into categories with equal probabilities (e.g., z-scores at −1, 0, 1). When applied to a normal latent distribution, this results in an observed categorical variable with a symmetric distribution.
   - Skewed Thresholds: Thresholds set at quantiles that create categories with unequal probabilities when applied to a normal latent distribution, resulting in left-skewed or right-skewed observed distributions.

If symmetric thresholds are applied to normal latent distributions with different means (due to treatment effects), the observed distributions may differ in symmetry or skewness. Similarly, applying the same symmetric thresholds to exponential latent distributions with different rate parameters will result in observed distributions that differ between groups in terms of skewness and spread. This interplay between latent distributions (and their parameters) and thresholds has a significant impact on the observed categorical distributions, as shown in Figures 2 and 3 of OSM1. These figures show how different configurations-including varying the number of bins (3, 4, and 5), types of latent distributions (normal or exponential), and distribution parameters-affect the observed results.

Moreover, we assessed the impact of different dependency structures by incorporating various copulas in a fully crossed design. The bivariate normal copula served as our baseline. Utilizing the VITA approach, we constructed vine distributions that integrated Clayton or Joe bivariate copulas for each treatment group. Specifically, the copulas linking the indicators of latent variables to the outcome variable were varied among Clayton, Joe, or Gaussian types. The terms "Joe-Gaussian" or "Gaussian-Gaussian" refer to the specific combinations of copulas used in the control and

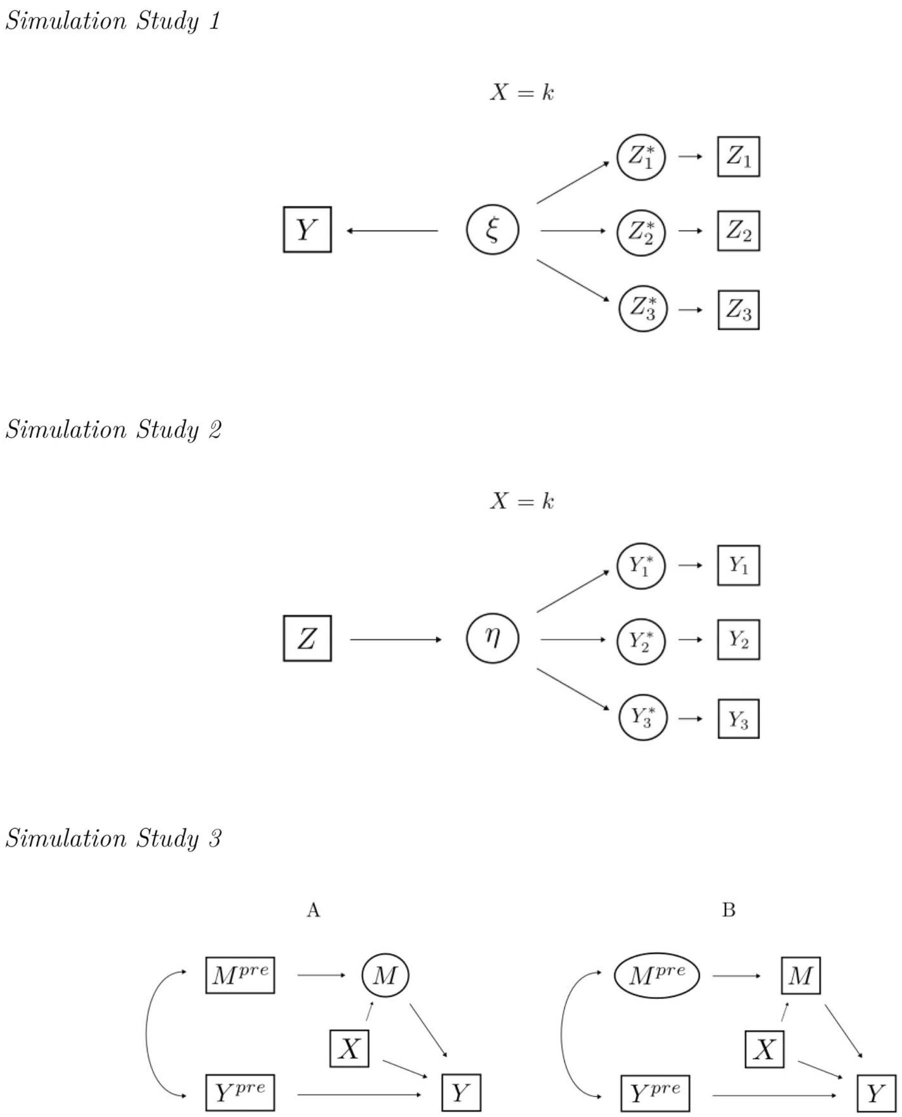

Figure 1. Model descriptions. This figure illustrates the multigroup SEMs used in all three simulation studies. In Study 1, a latent covariate interacts with a treatment variable *X*, which can take value *k* indicating different treatment groups. Study 2 features a latent dependent variable influenced by both a manifest covariate and the treatment variable *X*. Study 3 employs a path model to examine average total, direct, and indirect effects, where *X* determines the treatment condition affecting both the mediator *M* and the outcome variable *Y*, with *M* and a baseline measure M^{pre} modeled as latent variables with indicators derived from a discretization process.

treatment groups, respectively. For instance, "Joe-Gaussian" indicates that the copula in the control group (where $X = 0$) is of type Joe, while the copula in the treatment group (where $X = 1$) is of type Gaussian.

3. Results

The results for conditions with a sample size of $N = 100$ are presented below. Similar findings were observed for $N = 250$ and are reported in OSM2. Values of all evaluation criteria per parameter type and condition are depicted in the figures in OSM2.

3.1. Study 1, Normal Margins

3.1.1. ATE

Across all conditions, ATE estimates were unbiased across conditions, with both DWLS and ML displaying relative biases within ±2.18%. All estimation methods show good coverage rates close to the nominal level (see Figures 1, 2, 9, 10 in OSM1).

3.1.2. Interaction Effect

The interaction effect under conditions with correctly specified margins showed differing estimator performances. DWLS generally outperformed ML and MLR, providing mostly unbiased estimates when copula structures within treatment groups were identical. Even under different copula configurations, DWLS exhibited lower bias, particularly in symmetric distributions, though its performance varied depending on the specific copula combination and skewness of the observed distributions (see Figures 4 and 12 in OSM2). ML and MLR, in contrast, showed biased estimates under almost all conditions for the interaction effect, with only a few exceptions where the bias was reduced but not eliminated (see Figures 3 and 4 in OSM1). This was particularly evident when the underlying latent distributions of the indicators were asymmetric due to skewed thresholds, which did not substantially alleviate the bias inherent in ML estimation. Regarding coverage, all estimation methods demonstrated rates close to nominal under some conditions, but none outperformed the others. The number of bins had a significant impact on the coverage rates for ML and MLR, showing their sensitivity to the discretization process.

3.2. Study 1, Distributional Misspecification in Margins

3.2.1. ATE

Under conditions of marginal misspecification, ML and DWLS exhibited less than 10% bias in estimating the ATE (see Figures 5, 6, 9, 10 in OSM1). ML and MLR showed a strong deterioration in coverage compared to conditions without marginal misspecification, while DWLS maintained good coverage rates.

3.2.2. Interaction Effect

The effectiveness of DWLS varied for the estimation of the interaction effect, showing acceptable results (under 10% bias) when the rate parameter was set to 1.1, but displaying bias when adjusted to 0.9 (see Figures 7 and 8 in OSM1). In scenarios where the copulas within treatment groups are identical, the DWLS estimator generally provides unbiased estimates of the interaction effect. However, when the copulas are different, while the DWLS estimator is able to maintain lower bias in symmetric distributions, its performance is contingent upon the specific copula combination and the skewness of the observed distributions.

ML and MLR demonstrated substantial bias across almost all conditions involving interaction effects. Even with adjustments in the rate parameter and the number of bins, which generally led to a more uniform distribution of observations across bins, ML's bias was only somewhat mitigated but not eliminated. The interaction effect was often underestimated, particularly when the control group used a Gaussian copula and the treatment group used a Clayton copula. This pattern of underestimation was observed consistently in combinations such as Joe-Gaussian and Joe-Clayton, whereas combinations such as Clayton-Gaussian, Gaussian-Joe, and Clayton-Joe exhibited overestimation of the interaction effect (see Figure 5 in OSM2).

3.3. Study 2, Normal Margins

3.3.1. ATE

DWLS displayed acceptable bias across conditions when estimating the ATE. This contrasted with ML, which showed substantial bias, particularly in symmetric conditions, with performance slightly improving in skewed distributions (see Figures 17, 18, 25, and 26 in OSM1). The effectiveness of DWLS in reducing bias was evident across both identical and different copulas. However, coverage rates for DWLS were below the desired level, rarely exceeding 80% in symmetrical distributions and approximately 70% in asymmetrical conditions.

3.3.2. Interaction Effect

Under conditions with identical copulas between treatment groups, DWLS generally provided unbiased estimates of the interaction effect, with exceptions noted in specific scenarios: for Clayton-Clayton copulas with left-skewed data, where it overestimated the effects, and in some cases with Joe-Joe copulas with left-skewed data, where it underestimated them. The coverage rates for DWLS in these settings were mostly satisfactory and significantly better than those achieved by ML and MLR (see Figure 19 in OSM1).

In the context of differing copulas between treatment groups, DWLS's estimates were largely unbiased, maintaining coverage rates within acceptable levels, for symmetrical distributions (see Figure 20 in OSM1). In left-skewed distributions, however, the direction and magnitude of bias depended on the specific copula combinations. Notably, in right-skewed distributions, DWLS predominantly showed unbiased estimates, except in scenarios involving Joe-Clayton copulas with an interaction effect parameter of 0.2. Furthermore, a slight variation was observed between the results at factor loadings of 0.8 and 0.9, suggesting that the factor loading value influenced the extent of bias in left- and right-skewed distributions.

3.4. Study 2, Distributional Misspecification in Margins

3.4.1. ATE

When considering the effects of marginal misspecification on the ATE, all estimators showed substantial bias, as observed in Figures 21, 22, 29, and 30 in OSM2. This bias was primarily driven by the exponential distribution's rate parameter in the treatment group, which directly influenced the magnitude of the ATE. Coverage rates were predominantly low across all methods and did not systematically vary with the copula configurations used.

3.4.2. Interaction Effect

For the interaction effect, the results were similarly influenced by marginal misspecification but exhibited less bias compared to the ATE estimates. DWLS showed variability in bias that depended on the rate parameter of the exponential distribution, with tendencies for the estimated interaction effect to be larger when the rate parameter was set at 1.1 (see Figure 23 in OSM1). This occurred despite the observed distributions not appearing markedly different across rate parameter settings. Coverage rates for DWLS were mostly adequate, aligning closer to nominal levels compared to ML and MLR.

In scenarios involving different copulas between treatment groups, the direction and magnitude of bias in DWLS estimates were dependent not just on the rate parameter but also on the specific copula combinations used. Interestingly, the estimated interaction effect was generally larger when the factor loading was 0.8 compared to 0.9, although there was no consistent pattern of over- or underestimation. For ML and MLR, the bias was again influenced by the number of bins, which affected the distribution of observations across the bins and thus the estimation of the interaction effect.

3.5. Study 3 A, Normal Margins

Both the ATE and the ADE were unbiased across all estimators (see Figures 33, 34, 39, 40 in OSM1). The AIE remained unbiased in symmetrical distributions. However, in left- and right-skewed conditions, AIE biases varied with the copula combination. Notably, when factor loadings were increased to 0.9, DWLS showed substantially smaller bias compared to ML and MLR, which exhibited substantial bias. Coverage rates for DWLS were also consistently better than those for ML and MLR.

3.6. Study 3 A, Distributional Misspecification in Margins

The ATE remained unbiased even with marginal misspecification, showcasing the robustness of all estimators to such conditions. However, ADE showed biases influenced by both the rate parameter and the copula combination, with DWLS maintaining better coverage compared to ML and MLR. The AIE was strongly biased under marginal misspecification for all estimators. The extent of this bias was predominantly driven by the rate parameter, highlighting the sensitivity of AIE estimates to the underlying distributional assumptions.

3.7. Study 3 B, Normal Margins

In Study 3 B, both the ATE and the ADE were found to be unbiased across all conditions, consistent with the findings from Study 3 A. Contrary to the variability observed in Study 3 A, the AIE exhibited bias in all scenarios, influenced by the factor loading. At a factor loading of 0.8, the bias in AIE was noticeable but mostly manageable, with DWLS often maintaining acceptable coverage rates. However, when factor loadings increased to 0.9, the AIE exhibited substantial bias, which varied depending on the copula combination and rate parameter.

3.8. Study 3 B, Distributional Misspecification in Margins

The ATE remained generally unbiased under marginal misspecification, aligning with previous findings. The ADE also showed generally unbiased results. Under marginal misspecification, AIE was biased across all estimators, with the rate parameter predominantly influencing the direction and magnitude of the bias.

4. Discussion

This paper examines the impact of non-normality on causal effect estimates in path and multigroup SEMs with ordered categorical variables utilizing ML, MLR, and DWLS estimators. Conditions were systematically varied, including sample size, copula specifications, marginal distributions, skewness types, and model parameters. The results clarify how non-normality affects the performance of estimators, providing guidance to applied researchers when selecting statistical methodologies for causal inference.

For the model with a latent covariate (Study 1), our findings suggest that both ML and DWLS estimators show acceptable bias for estimating the ATE across different simulation conditions. Additionally, coverage rates for both methods were close to the nominal level, with DWLS leaning towards more conservative rates. As for the interaction effect, the results were more nuanced. DWLS exhibited mostly unbiased estimates of the interaction effect under identical copula distributions in the treatment groups or with symmetric observed marginal distributions, but its performance was significantly impacted by varying bivariate copula structures between treatment groups with skewed observed distributions. This effect was particularly evident in scenarios with skewed distributions, indicating that DWLS's ability to estimate interaction effects is sensitive to the particular copula configurations employed in the simulation. However, despite some reduction in bias under skewed conditions, ML estimation, traditionally known for its bias when applied to ordered categorical variables (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Li, 2016a; Rhemtulla et al., 2012), still exhibits substantial bias and does not achieve the more consistent performance observed with DWLS, especially in handling interaction effects. The bias observed in DWLS estimates under varying copula structures and skewed distributions reveals a sensitivity to distributional characteristics that are not testable. The impact of marginal misspecification was particularly evident in the ML method, where changes in the rate parameter of the underlying latent distribution and discretization process significantly influenced the bias observed. Additionally, the specific copula combinations led to underestimation while others led to overestimation.

For the model with a latent dependent variable (Study 2), the results revealed distinct patterns in estimator performance. For the ATE, DWLS displayed up to moderate bias, while ML exhibited substantial bias, indicating ML's heightened sensitivity to the distributional characteristics of the data. Interestingly, for the interaction effect estimation, even though the covariate was manifest, DWLS again showed up to moderate bias, whereas ML demonstrated substantial bias. This suggests that ML's limitations in handling ordered categorical variables extend to scenarios where covariates are manifest, potentially complicating analyses in practical applications. It was also observed that the bias shown by DWLS in estimating interaction effects happened in conditions of skewness with differing copula configurations. The results of Study 1 are supported by this finding, which emphasizes a consistent pattern of DWLS in scenarios with skewed distributions. In conditions of marginal misspecification, both DWLS and ML showed significant bias in estimating the ATE and interaction effects.

Both ML and DWLS demonstrated negligible bias and robust performance for ATE and ADE in the multigroup path model, achieving coverage rates close to nominal levels. However, challenges arose in the estimation of AIE, where substantial biases were observed, especially under varied copula structures and skewness conditions. The study found that the biases in ML estimates for AIE were amplified for the Joe-Joe and Clayton-Clayton copula configurations. It was found that increased factor loadings in Joe-Joe and Clayton-Clayton copula configurations aggravated bias in ML estimates of AIE, while DWLS maintained relative stability. In cases of marginal misspecification, all estimators consistently overestimated ADE and underestimated AIE.

In the realm of SEM with ordinal variables, several key studies have shaped our understanding of estimator robustness. Flora and Curran (2004) and Rhemtulla et al. (2012) concluded that ordinal SEM estimation, such as DWLS, is relatively robust against violations of underlying normality. However, the present study supports the findings of Foldnes and Grønneberg (2022), showing that ordered categorical SEM is indeed sensitive to underlying non-normality, resulting in biased causal parameter estimates and inadequate CI coverage rates. This suggests a need for a more cautious approach, including empirical assessments of the normality of the data and and expert knowledge. Furthermore, when neither the copula class nor the marginals are known, the problem becomes more challenging. In such cases, practitioners are advised to start their analysis with a test for the normality of the copula using a bootstrap test (Grønneberg & Foldnes, 2019). However, when the test indicates a non-normal copula and the copula is unknown, the options are limited. Foldnes and Grønneberg (2022) suggest that ML can be used if equally spaced thresholds are known, and DWLS can be employed when the response marginals are known conditional on the fact that the number bins is sufficiently large.

Summarizing our findings, we can conclude that:

1. DWLS provides an unbiased estimation of the ATE in scenarios where the dependent variable is manifest and the covariate is latent and the indicators of the covariate result from a discretization process.
2. Our findings stress the importance of considering skewness in the data and the underlying copula structures when modeling interaction effects in SEM. Researchers should explore the characteristics of the distributions of the data at hand and consider employing simulation studies or bootstrap methods to understand how these factors might influence their results.
3. The bias in interaction effect estimation raise important considerations for causal inference, especially in the context of regression methods estimating a conditional expectation $\mathbb{E}(Y|X,Z)$. Accurate estimation of these effects is critical, as it directly influences the understanding of conditional treatment effects. In cases where the covariate distributions differ between treatment groups, as often seen in quasi-experimental studies, biased interaction effect estimation could lead to incorrect interpretations of treatment efficacy in subpopulations.

For practitioners, our findings indicate that while no single estimator universally outperforms others under all conditions, DWLS consistently demonstrated robust performance across a range of scenarios in our studies. Therefore, we recommend DWLS as a generally reliable choice for estimating causal effects in settings involving ordered categorical data and complex dependency structures. Nevertheless, the selection of an estimator should be guided by the specific characteristics of the dataset, the research questions being addressed, and the practical constraints of the study, as certain conditions may favor alternative approaches.

Additionally, the observed bias in interaction effects can have profound implications. Particularly in quasi-experimental contexts where treatment and control groups may differ on key characteristics, these biases can misrepresent the true effect of interventions (Kirchmann & Steyer, 2012). For example, if our simulated treatment is more effective for one person with a high score on a relevant covariate than another, biased interaction effects could either exaggerate or mask this difference. This is not merely a statistical concern but a substantive one, as it may lead to misinformed decisions about the efficacy and application of treatments. Our findings highlight the need for robust statistical techniques that can accommodate the complexities of real-world data and ensure accurate interpretation of treatment effects in diverse populations.

The simulation conditions in our study only cover some possible scenarios researchers might encounter. Future studies could explore a broader range of distributional scenarios, including different types of marginal misspecification and more complex copula structures. Furthermore, we focused on ML, MLR, and DWLS estimators due to their prevalence in SEM analysis. However, alternative approaches, such as Bayesian methods or nonparametric estimators, might offer different perspectives on non-normality and warrant exploration in future research. The impact of copula structures and marginal distributions on estimator performance emphasizes the need for SEM models that can incorporate copulas and marginal distributions into model estimation. While simulation studies provide controlled conditions to assess estimator performance, empirical validation using real-world data can offer additional insights into the practical implications of our findings. Future research could apply these estimators to datasets from various disciplines to evaluate their performance.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444–455. https://doi.org/10.1080/01621459.1996.10476902

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies.

Multivariate Behavioral Research, *46*, 399–424. https://doi.org/10.1080/00273171.2011.568786

Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: a Multidisciplinary Journal*, *21*, 102–116. https://doi.org/10.1080/10705511.2014.859510

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in cfa. *Structural Equation Modeling*, *13*, 186–203. https://doi.org/10.1207/s15328007sem1302_2

Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, *32*, 245–268. https://doi.org/10.1023/A:1016725902970

Boulton, A. J., & Williford, A. (2018). Analyzing skewed continuous outcomes with many zeros: A tutorial for social work and youth prevention science researchers. *Journal of the Society for Social Work and Research*, *9*, 721–740. https://doi.org/10.1086/701235

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *The British Journal of Mathematical and Statistical Psychology*, *37 (Pt 1)*, 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Delucchi, K. L., & Bostrom, A. (2004). Methods for analysis of skewed data distributions in psychiatric clinical studies: Working with many zero values. *The American Journal of Psychiatry*, *161*, 1159–1168. https://doi.org/10.1176/appi.ajp.161.7.1159

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Foldnes, N., & Grønneberg, S. (2019). On identification and non-normal simulation in ordinal covariance and item response models. *Psychometrika*, *84*, 1000–1017. https://doi.org/10.1007/s11336-019-09688-z

Foldnes, N., & Grønneberg, S. (2022). The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*, *27*, 541–567. https://doi.org/10.1037/met0000385

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time. or not? lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, *28*, 1354–1367. https://doi.org/10.1037/pas0000275

Genest, C., & MacKay, R. J. (1986). Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, *14*, 145–159. https://doi.org/10.2307/3314660

Genest, C., & Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, *88*, 1034–1043. https://doi.org/10.1080/01621459.1993.10476372

Grønneberg, S., & Foldnes, N. (2017). Covariance model simulation using regular vines. *Psychometrika*, *82*, 1035–1051. https://doi.org/10.1007/s11336-017-9569-6

Grønneberg, S., & Foldnes, N. (2019). A problem with discretizing vale–maurelli in simulation studies. *Psychometrika*, *84*, 554–561. https://doi.org/10.1007/s11336-019-09663-8

Grønneberg, S., & Foldnes, N. (2024). Factor analyzing ordinal items requires substantive knowledge of response marginals. *Psychological Methods*, *29*, 65–87. https://doi.org/10.1037/met0000495

Grønneberg, S., Foldnes, N., & Marcoulides, K. M. (2022). Covsim: An r package for simulating non-normal data for structural equation models using copulas. *Journal of Statistical Software*, *102*, 1–45. https://doi.org/10.18637/jss.v102.i03

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC.

Herschtal, A. (2023). The effect of dichotomization of skewed adjustment covariates in the analysis of clinical trials. *BMC Medical Research Methodology*, *23*, 60. https://doi.org/10.1186/s12874-023-01878-9

Hofert, M., Kojadinovic, I., Mächler, M., & Yan, J. (2018). *Elements of copula modeling with r*. Springer.

Jin, S., Luo, H., & Yang-Wallentin, F. (2016). A simulation study of polychoric instrumental variable estimation in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 680–694. https://doi.org/10.1080/10705511.2016.1189334

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, *46*, 262–282. https://doi.org/10.1006/jmva.1993.1061

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202. https://doi.org/10.1007/BF02289343

Kirchmann, H., & Steyer, R. (2012). Evaluation von behandlungseffekten in quasi-experimentellen studien. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, *62*, 149–151. https://doi.org/10.1055/s-0031-1299010

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.

Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, C.-H. (2016b). The performance of ml, dwls, and uls estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*, 369–387. https://doi.org/10.1037/met0000093

Mayer, A., Dietzfelbinger, L., Rosseel, Y., & Steyer, R. (2016). The effectliter approach for analyzing average and conditional effects. *Multivariate Behavioral Research*, *51*, 374–391. https://doi.org/10.1080/00273171.2016.1151334

Mayer, A., Thoemmes, F., Rose, N., Steyer, R., & West, S. G. (2014). Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research*, *49*, 425–442. https://doi.org/10.1080/00273171.2014.931797

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132. https://doi.org/10.1007/BF02294210

Muthén, B., & Asparouhov, T., (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in mplus. *Mplus Web Notes*, *4*, 1–22.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ml, gls, and wls estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*, 557–595. https://doi.org/10.1207/S15328007SEM0704_3

Pohl, S., Steiner, P. M., Eisermann, J., Soellner, R., & Cook, T. D. (2009). Unbiased causal inference from an observational study: Results of a within-study comparison. *Educational Evaluation and Policy Analysis*, *31*, 463–479. https://doi.org/10.3102/0162373709343964

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rhemtulla, M., Brosseau-Liard, P. É., & 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. https://doi.org/10.1037/a0029315

Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Sengewald, M.-A., & Mayer, A. (2024). Causal effect analysis in nonrandomized data with latent variables and categorical indicators:

The implementation and benefits of effectliter. *Psychological Methods*, *29*, 287–307. https://doi.org/10.1037/met0000489

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229–231.

Steyer, R. (2005). Analyzing individual and average causal effects via structural equation models. *Methodology*, *1*, 39–54. https://doi.org/10.1027/1614-1881.1.1.39

Steyer, R., Mayer, A., & Fiege, C. (2014). Causal inference on total, direct, and indirect effects. In *Encyclopedia of quality of life and well-being research* (pp. 606–630). Springer Netherlands.

Tomitaka, S., Kawasaki, Y., & Furukawa, T. (2015). Right tail of the distribution of depressive symptoms is stable and follows an exponential curve during middle adulthood. *PloS One*, *10*, e0114624. https://doi.org/10.1371/journal.pone.0114624

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471. https://doi.org/10.1007/BF02293687

Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *The British Journal of Mathematical and Statistical Psychology*, *59*, 397–417. https://doi.org/10.1348/000711005X85896