STRUCTURAL EQUATION MODELING
2026, VOL. 33, NO. 2, 268–280
ISSN: 1070-5511 | E-ISSN: 1532-8007

From Path Diagrams to Causal Graphs: A Structural Causal Perspective on Cross-Sectional Multilevel Models

Moritz Ketzer[a,b], Christian Gische[a], and Manuel C. Voelkle[a,c]

[a]Humboldt-Universität zu Berlin; [b]Max Planck UCL Centre for Computational Psychiatry and Ageing Research; [c]Albert-Ludwigs-Universität Freiburg

ABSTRACT

Causal graphs provide a rigorous framework for encoding causal assumptions. Yet, their integration with multilevel models remains limited. We review common path diagram conventions in multilevel modeling and show, through a bivariate regression example, that these function as statistical model visualizations rather than causal graphs. We formalize parametric cross-sectional multilevel models as parametric structural causal models with linear causal effects (conditional on possible moderators) and Gaussian error terms. We then translate them into acyclic directed mixed graphs. We illustrate this framework using a well-known empirical example, the High School and Beyond study. This work provides a systematic bridge between cross-sectional parametric multilevel models and modern causal graph theory.

KEYWORDS

Causal graphs; directed acyclic graphs; multilevel models; structural causal models; unobserved heterogeneity

Path diagrams have a long history and are widely used in structural equation modeling (SEM) to visualize model structure, assumptions, and scientific results (Bollen, 1989; Mulaik, 2009). In recent years, *structural causal models* (SCMs) and their graphical representations, *directed acyclic graphs* (DAGs), have emerged as a formal framework to guide model specification, to reason about sources of bias, and to assess whether causal effects can be identified (Pearl, 2009; Pearl & Mackenzie, 2018). In linear, single-level models, the framework of path diagrams and DAGs largely aligns, allowing researchers to quickly leverage the formal machinery of causal graphs (Kline, 2023). However, in multilevel models, this alignment breaks down. Psychologists have developed a rich and diverse set of path diagraming conventions to visualize multilevel models (Curran & Bauer, 2007; Mehta & Neale, 2005; Muthén & Muthén, 2017; Skrondal & Rabe-Hesketh, 2004). However, none of these satisfy the formal requirements of DAGs and SCMs. This lack of a formal graphical framework for multilevel models is a serious limitation and hampers the proliferation and integration of causal graphs into many applied fields.

Multilevel modeling frameworks were developed to address violations of i.i.d. assumptions, arising frequently in real-world data due to dependencies of observations within clustered units or over time (Skrondal & Rabe-Hesketh, 2004; Snijders & Bosker, 2012). In contrast, SCMs are formulated for i.i.d. variables within a single unit of analysis (Peters et al., 2017). Because of this, the additional assumptions in multilevel models cannot be straightforwardly represented with single-level DAGs. Despite these divergences, causal inference in multilevel models has a long history (Hamaker & Muthén, 2020; Kim & Frees, 2006, 2007; Skrondal & Rabe-Hesketh, 2004), particularly in longitudinal and panel data studies common in econometrics (Chamberlain, 1982; Hsiao, 2022; Mundlak, 1978; Wooldridge, 2010). The recently emerging literature on causal graphs in multilevel models is still sparse and mainly formulated in the general nonparametric setting (Guo, Zhang, et al., 2024; e.g., Guo, Tóth, et al., 2024; Imai & Kim, 2019; Weinstein & Blei, 2024; but see Gische et al., 2021). In this paper, we contribute to this literature by constructing DAGs for *parametric cross-sectional* multilevel models to complement the existing nonparametric approaches. We restrict our focus to hierarchical structures, where repeated observations are nested within higher-level units but measured at a single time point. However, the arguments and the approach we propose are intended to lay the groundwork for future extensions to dynamic and panel data models, where temporal dependencies introduce additional challenges to causal representation and identification.

To establish a foundation, we first introduce the formalism and framework of SCMs and causal graphs. We then review a representative set of the most common path diagraming conventions in multilevel modeling in the context of a bivariate linear regression example. We demonstrate

and discuss the limitations of these conventions and emphasize how their shortcomings emerge from the distinction between statistical and causal graphs. Building on insights from the multilevel structural equation modeling (ML-SEM) literature, we conceptualize the bivariate multilevel model example as an SEM in wide format, treat it formally as an SCM, and propose a systematic approach for constructing its corresponding causal graph. Finally, we apply this approach to discuss the graphical representation of a widely used empirical example in multilevel modeling: the High School and Beyond data set.

1. Graph-Based Causal Models

In principle, SCMs can be conceived as nonparametric extensions of SEMs: they retain the use of structural equations but relax assumptions about linearity, additivity, and distributional form (Kline, 2023; Pearl, 2009). Each variable in an SCM is defined by a structural equation that expresses it as a function of its direct causes (or "parents") and an independent error term. The system of equations is paired with a graphical representation (typically a DAG) in which each directed edge encodes a direct causal relationship. These graphs make the model's assumptions explicit: the absence of an edge implies a conditional independence that can be read off using a graphical criterion known as *d-separation*.

A central advantage of SCMs is that they provide a formal language for expressing and analyzing interventions. The *do-operator*, written as $do(X = x)$, represents a hypothetical manipulation that sets a variable X to a value x, breaking any incoming edges to X in the graph. This allows us to distinguish between observational and interventional distributions. The distribution $P(Y|do(X = x))$ expresses the distribution of an outcome Y under an intervention on X, and summary statistics derived from this distribution define causal estimands (Gische & Voelkle, 2022). Two common estimands are the *average treatment effect* (ATE) and the *conditional average treatment effect* (CATE). The ATE compares the expected value of an outcome under two different treatment levels x and x' and applies generally to any scale of outcome or treatment:

$$\mathrm{ATE}_{x\to x'} = \mathbb{E}\left[Y|do(X = x')\right] - \mathbb{E}[Y|do(X = x)]. \tag{1}$$

The CATE conditions on a set of observed covariates:

$$\begin{aligned}\mathrm{CATE}_{x\to x'} = {} & \mathbb{E}\left[Y|do(X = x'), \mathbf{Z} = \mathbf{z}\right] \\ & - \mathbb{E}[Y|do(X = x), \mathbf{Z} = \mathbf{z}].\end{aligned} \tag{2}$$

These estimands allow us to formulate precise causal questions and evaluate them using graphical and statistical tools.

By extension, acyclic directed mixed graphs (ADMGs) include both directed and bidirected edges (Richardson, 2003). This addition reflects the structure of structural equation models, where correlated error terms correspond to bidirected edges between their associated variables (Spirtes et al., 1998). The corresponding separation criterion, *m-separation*, generalizes d-separation because bidirected edges simply introduce extra arrowheads, effectively creating more opportunities for collider structures. Intuitively, a bidirected edge $(X \leftrightarrow Y)$ can be seen as shorthand for a latent common cause $(X \leftarrow U \rightarrow Y)$, and m-separation reasons about the former just as d-separation does in the latter. The direct correspondence between SEMs and ADMGs is particularly relevant for our purposes, because multilevel models can be interpreted as systems of structural equations (Bauer, 2003; Curran, 2003; Mehta & Neale, 2005; Skrondal & Rabe-Hesketh, 2004), and therefore as ADMGs. Adopting this representation provides a direct bridge between the statistical formulation of multilevel models and their causal interpretation within the SCM framework. Throughout the remainder of this paper, we use ADMGs as the graphical representation of parametric multilevel SCMs.

2. Multilevel Models

Multilevel models (Skrondal & Rabe-Hesketh, 2004; Snijders & Bosker, 2012), also known as hierarchical models (Gelman & Hill, 2006; Raudenbush & Bryk, 2002), mixed models (McCulloch & Searle, 2001; Twisk, 2019), or variance component models (Searle et al., 2009), are statistical models accounting for data that are organized at more than one level of analysis. Multilevel data may be cross-sectional (units nested within groups at one time), longitudinal (repeated observations nested within units), or a combination of both. In what follows, we focus on cross-sectional models with two hierarchically nested levels, such as students (level-1) within classrooms (level-2). To demonstrate how representative multilevel path diagraming conventions violate the framework of causal graphs, we introduce a basic multilevel model without level-2 variables as a *running example*. Throughout this paper, we use the parameters μ for means, β for level-1 slopes, γ for level-2 slopes, and η for varying parameters.[1] Moreover, we use lowercase letters for measured and uppercase letters for random variables. For a given outcome Y, the model is specified as:

$$y_{ij} = \eta_{0j} + \eta_{1j}x_{ij} + e_{ij} \tag{3}$$

where y_{ij} is the outcome for unit i in cluster j, x_{ij} is a unit-level predictor, η_{0j} and η_{1j} are the intercept and slope coefficients varying over cluster j, and e_{ij} is the unit-level residual term. The varying intercept and slope are defined via *auxiliary* equations:

$$\begin{aligned}\eta_{0j} &= \mu + u_{0j} \\ \eta_{1j} &= \beta + u_{1j}\end{aligned} \tag{4}$$

where μ is the fixed intercept, β is the fixed slope, referred to collectively as fixed effects, and u_{0j}, u_{1j} represent the unexplained variation of the parameters, referred to collectively as random effects. Substituting these into the original model yields:

$$y_{ij} = (\mu + u_{0j}) + (\beta + u_{1j})x_{ij} + e_{ij} \tag{5}$$

[1] Using the parameter η makes the perspective of varying parameters as latent random variables explicit (Bauer, 2003; Curran, 2003; Mehta & Neale, 2005; Skrondal & Rabe-Hesketh, 2004).

which can be equivalently written as:

$$y_{ij} = \mu + \beta x_{ij} + x_{ij}u_{1j} + u_{0j} + e_{ij} \tag{6}$$

The following assumptions about the distributions of the errors are typically made in multilevel models:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{u_0} & \\ \psi_{u_0u_1} & \psi_{u_1} \end{bmatrix}\right) \tag{7}$$
$$\begin{bmatrix} e_{ij} \end{bmatrix} \sim N(0, \sigma^2)$$

We follow the common assumptions in the statistical literature: the level-1 residual e_{ij} is assumed to be independently and identically distributed (i.i.d.) following a normal distribution with mean 0 and variance σ^2; the random effects u_{0j} and u_{1j} are assumed to be i.i.d. following a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{\Psi}$.

3. Common Visualizations of Multilevel Models in the Psychological Literature

To demonstrate that common path diagraming conventions in the psychological literature are incompatible with the strict semantics of causal graphs, we illustrate four widely used conventions using the running example introduced in the previous section. These include the Mplus convention as detailed by Muthén and Muthén (2017), the extended RAM notation described by Mehta and Neale (2005), the multilevel path diagrams proposed by Curran and Bauer (2007), and the path diagrams used in the generalized latent and mixed model convention outlined by Skrondal and Rabe-Hesketh (2004). While we believe that these conventions are representative, they do not constitute an exhaustive list, and there may be other conventions that differ in their adherence to graph-theoretic principles. Furthermore, constructing these diagrams based on the system of equations introduces certain ambiguities, as path diagrams are often inconsistently defined. Throughout this section, rectangles represent observed variables, circles represent latent variables, and triangles represent intercepts or the mean structure. Directed arrows represent direct effects, and bidirected arrows represent variances and covariances.

3.1. Mplus Diagrams

Figure 1, adapted from Mehta and Neale (2005), illustrates the Mplus convention (Muthén & Muthén, 2017). It originated from the early formulations of multilevel SEM, where separate structural equation models are specified for the within-level and between-level covariance structures (Mehta & Neale, 2005). The diagraming convention is widely adopted due to the popularity of Mplus and the clarity of its diagrammatic approach (see e.g., Preacher et al., 2010, 2016). In Figure 1, the horizontal dashed line denotes the division between levels: The subunit level (level-1 or "within" level) is drawn above the dashed line and the unit level (level 2 or "between" level) is drawn below the dashed line.

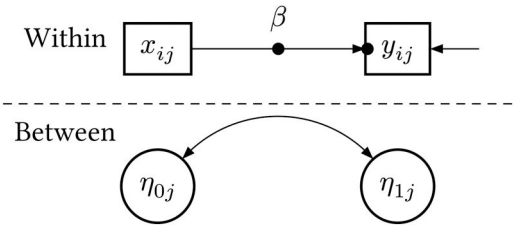

Figure 1. Mplus multilevel path diagram. *Note.* The diagram is divided into two panels by a dashed line, labeled 'Within' and 'Between' corresponding to level-1 and level-2 models respectively. Direct effects are represented by directed arrows, correlations by bidirected arrows, and varying parameters by black dots. The small arrow to the right of the outcome variable y_{ij} represents the residual.

While this is concise and intuitive, the drawback is that dependencies between levels are not explicitly represented, making hierarchical relationships ambiguous. The black dots represent varying parameters: a varying regression coefficient when positioned at the center and a varying intercept when placed at the tip of a directed arrow. These are special elements and have no direct analog in graph theory. In the "Between" panel, each varying parameter at the "Within" model is depicted as a latent variable, η_{0j} and η_{1j}, with a bidirected arrow indicating correlations between them. In the Mplus convention, indices or subscripts are usually omitted, (e.g., Muthén & Asparouhov, 2011), as levels are already indicated by the dashed bar. However, for reasons of clarity and consistency, we include indices explicitly.

3.2. Extended RAM Diagrams

Figure 2 displays the multilevel path diagram of our example model according to the Extended RAM convention and is adapted from Mehta and Neale (2005, p. 276). This diagrammatic convention emerged from efforts to create a general framework to specify multilevel models as SEMs, where both univariate and multivariate multilevel models can be cast as multivariate single-level models, referred to as "wide-format SEM".[2] Each row represents a single unit of analysis and each column represents the unit's repeated measures or nested observations (i.e., subunits, such as person i to N in Figure 2) as distinct variables. From this perspective, multilevel models can be understood as restricted measurement models (see also Bauer, 2003; Curran, 2003; Skrondal & Rabe-Hesketh, 2004), where varying parameters are estimated as factors with fixed loadings: loadings of ones for the varying intercept and loadings fixed to the observed variable X for the varying slope. This estimation method has notable consequences for visualizing multilevel models, because there is only a single model and therefore a single path diagram. Furthermore, by collapsing the multilevel model into a single-level framework, each nested observation is handled as an individual variable, making the graphical representation more intuitive but also more extensive.

In Figure 2, each subunit outcome variable is drawn explicitly, with three dots indicating the continuation of

[2]For the history and literature overview for this kind of perspective see Barendse and Rosseel (2020) and Barendse and Rosseel (2023).

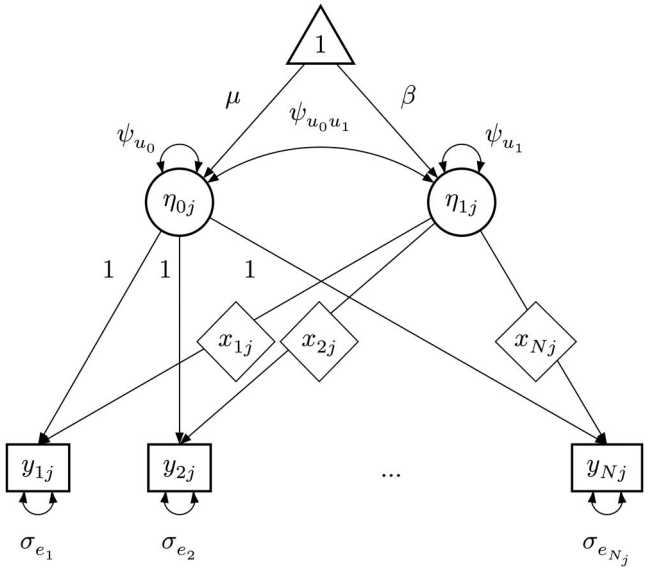

Figure 2. Extended RAM multilevel path diagram. *Note.* Diamonds represent definition variables. Direct effects are represented by directed arrows. Bidirected arrows represent covariances or correlations, when two distinct nodes are connected. Bidirected arrows going out and in to the same node represent (error) variances.

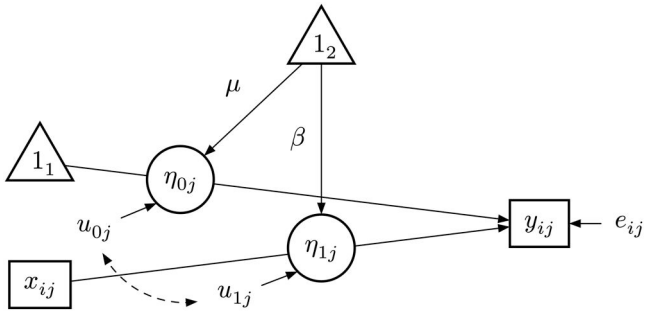

Figure 3. Curran and bauer multilevel path diagram. *Note.* Rectangles represent observed variables and circles represent random effects. Triangles represent intercepts, where the subscript indicates the corresponding level. Direct effects are represented by directed arrows. Bidirected arrows represent covariances or correlations between random effects. Residuals are made explicit at every level, that is, level-1 residuals e_{ij} and level-2 residuals of the varying effects u_{0j} and u_{1j}.

variables not shown in the diagram. The varying intercept, denoted by η_{0j}, and the varying slope, η_{1j}, are depicted as latent variables. The varying intercept is assigned unit loadings, while the random slope's loadings are fixed to the unit's observed values of X, represented by diamond shapes and referred to as definition variables. The mean of η_{0j} corresponds to the fixed intercept γ_0, while its variance, ψ_{u_0}, represents the variability of the varying intercept. Similarly, the mean of η_{1j} corresponds to the fixed slope β_1, and its variance, ψ_{u_1}, represents the variability of the varying slope. From a causal graph perspective, two issues arise: (1) fixed effects are modeled as latent variable means rather than path coefficients, whereas DAGs omit mean structures in the form of triangles and their paths as they are not relevant for causal analysis; and (2) definition variables, represented as diamond-shaped nodes, function as path coefficients rather than traditional variables.

3.3. Multilevel Path Diagrams According to Curran and Bauer

Figure 3 shows the multilevel path diagram of our example model, following the conventions established by Curran and Bauer and adapted from Curran and Bauer (2007, p. 287). The approach by Curran and Bauer (2007) aimed to improve upon the two previously reviewed path diagraming conventions and was developed to achieve two main objectives: to facilitate the derivation of underlying equations and to provide a clear and efficient means of communicating the model structure, assumptions, and empirical results. Importantly, the authors explicitly state that their focus is purposefully restricted to achieving the aforementioned goals for "the types of models most widely used in practice" [Curran and Bauer (2007), p. 284; i.e., linear multilevel models]. Since the varying parameters are explicitly represented as nodes, the path diagram more closely aligns with the multilevel notation rather than the mixed model notation. From a graphical modeling perspective, the issue is that nodes are placed on edges, and again, the mean structure is used to define fixed effects.

3.4. Generalized Latent Variable Diagrams

Figure 4 shows the multilevel path diagram of our example model according to the convention of Skrondal and Rabe-Hesketh (2004). This diagrammatic convention emerged from efforts to unify SEM (i.e., measurement *and* structural models) and multilevel models. A unique feature that appears in no other path diagraming convention is the use of rectangular boxes that represent repeating structures. These are similar to plate notation from hierarchical Bayesian modeling, which we will come back to later in the manuscript. However, the authors did not name them as such; instead, they are referred to as "nested frames" (Rabe-Hesketh et al., 2004, p. 176) or "box" (Skrondal & Rabe-Hesketh, 2004, e.g., Section 1.2 or Section 4.2.4).

However, the generalized latent variable diagrams are not without their own idiosyncrasies and are closely tied to the generalized linear latent and mixed models framework (GLLAMM) (see Figure 2, which illustrates Equation 19 in Rabe-Hesketh et al., 2004). The most striking incompatibility with causal graphs is the representation of varying slopes by edges pointing to edges (see Figure 3.8 in Skrondal & Rabe-Hesketh, 2004; or Figure 3 in Rabe-Hesketh et al., 2004, p. 182)—a common convention to represent interactions in path diagrams (Kim & Jung, 2025). Indeed, Rabe-Hesketh et al. (2004) explicitly state that a "random coefficient is represented by a latent variable having an arrow pointing from it to the path representing the corresponding fixed coefficient, a convention previously used in Pickles et al. (1998)" (p. 182).

4. Statistical Versus Causal Diagrams

The diagraming conventions reviewed in the previous section diverge in several ways from DAG and SCM principles.

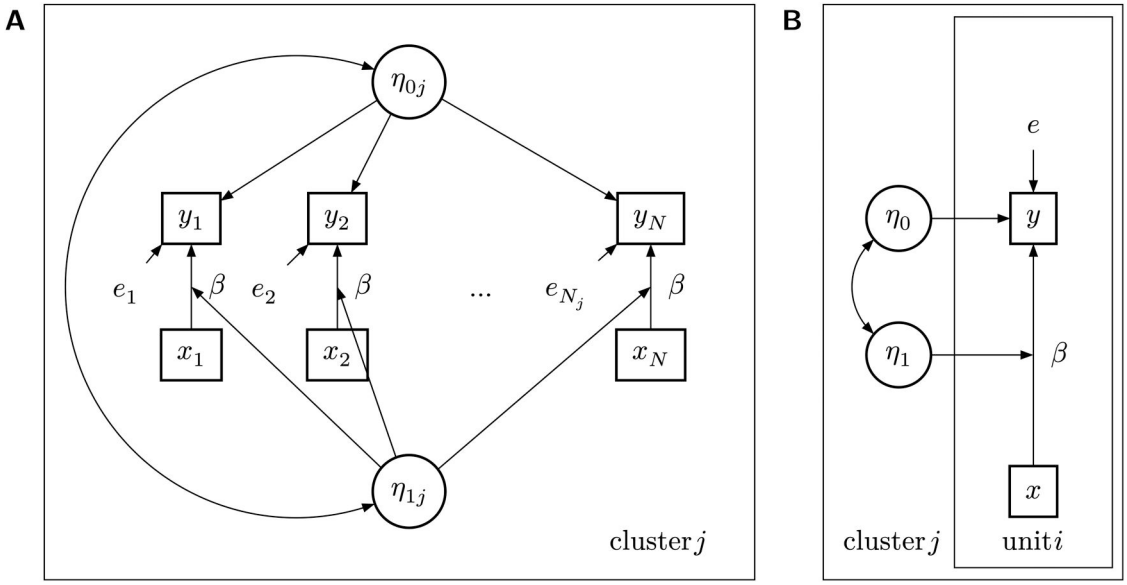

Figure 4. Generalized latent variable diagram. *Note.* Rectangles represent observed variables and circles represent latent variables. Direct effects are represented by directed arrows. Bidirected arrows represent covariances or correlations between random effects. Surrounding boxes represent repeating structures.

First, all of these diagrams violate core graphical principles: they separate levels into disconnected subgraphs (Muthén & Muthén, 2017), place points (Muthén & Muthén, 2017), nodes (Curran & Bauer, 2007), or variables on edges (Mehta & Neale, 2005), and use edges pointing to edges (Skrondal & Rabe-Hesketh, 2004). Second, they reflect the conventions of statistical modeling and diverge from core principles of SCMs. For example, assumptions about the independence of the residuals are often the default, such that causal assumptions are built into the diagram by construction, rather than being interrogable features. Furthermore, covariates are often treated as fixed inputs, instead of random structural variables.

Most importantly, the interpretation of auxiliary equations (such as those in Equation (4)) as structural equations results in conceptual ambiguities. By treating varying parameters simultaneously as deterministic endogenous variables, population-level structural coefficients, and exogenous error terms, the boundaries between distinct SCM concepts become blurred. If parameters are considered deterministic endogenous variables, this implies the possibility of direct intervention upon them—a notion that remains theoretically unclear and practically ambiguous.[3] Furthermore, structural coefficients, such as slopes, become conflated with mean structures. For example, Figures 2 and 3 incorporate the causal effect β within the mean structure of the varying slopes η_{1j}. This is conceptually problematic from the perspective of causal graphs, which exclude mean structures by definition. Finally, varying parameters as latent variables may or may not be exogenous causes: unconditional models (such as Equation (4)) of centered covariates correspond directly to exogenous causes, yet, in conditional models the level-2 exogenous error is represented explicitly as its own (latent exogenous) node (see Skrondal & Rabe-Hesketh, 2022, p. 818, Figure 5). To resolve these tensions, we now recast multilevel models as Multilevel Structural Causal Models (ML-SCMs) and show how a ADMG can be systematically constructed.

5. Multilevel Structural Causal Models

The SCM framework operates with a strict formal vocabulary: *endogenous deterministic variables* are determined within the system and *exogenous random variables* represent external sources of variation (i.e., unobserved causes that are not themselves explained by the model). Each deterministic endogenous variable is generated as a *deterministic function* or *structural equation* f of other endogenous variables and exogenous variables. The randomness in the system comes entirely from the probability distribution P over the exogenous variables. In many practical applications, both the causal functions f_θ and the distribution over the exogenous variables P_θ are assumed to depend on a set of *exogenous population-level parameters* θ. Throughout this paper, we assume all causal functions are linear (conditional on possible moderator variables) and all exogenous variables are generated by a multivariate Gaussian distribution. However, these assumptions can be relaxed.

Standard SCMs assume a flat data-generating process: observations are assumed to be i.i.d. and the causal system is represented at the level of a single unit. In other words, there is only one level or indexing variable, typically i, along which variables vary. In contrast, multilevel models accommodate causal structures that vary along multiple levels, each introducing dependencies that violate the i.i.d. assumption. A natural assumption in such settings is that lower-level units are conditionally independent given their shared cluster-level context. In other words, observations are i.i.d. within clusters once we condition on latent group-level variables representing unobserved causes common to that

[3]Changes of parameters due to interventions can be conceptualized by the CATE, such that moderators of the causal effect are the target of intervention. This logic holds irrespective of the level of the variable.

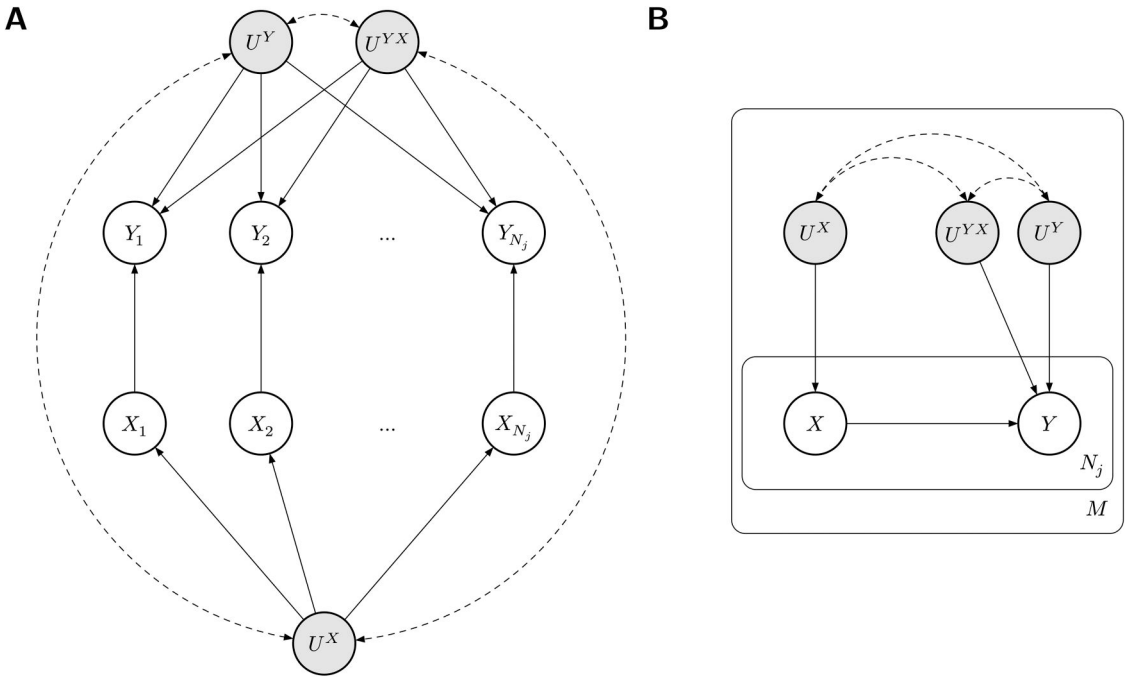

Figure 5. Multilevel hierarchical DAG. *Note.* Circles represent variables and gray shading represents unobserved variables. Unique errors are omitted and arrows are attached to their children. Shared errors are included. Bidirected arrows are drawn with a dashed line for better visibility. The graph is drawn at the cluster level j. Panel A depicts the fully expanded DAG. Panel B depicts the same graph using plate notation. Paths are unlabeled to retain their generality, but all error paths could be labeled with ones as usual, the directed arrow(s) with β and the bidirected arrows with the entries in the covariance matrices. Plate notation implies that repeated edges are identical.

cluster.[4] For example, in a model of students nested within schools, individual outcomes are not independent in the population but become independent once we condition on the (unobserved) school-specific factors that affect all students in that school.

We now recast the multilevel regression defined in Equations (3), (4), (5), and (7) as an ML-SCM. In this formulation, covariates are decomposed into level-1 and level-2 components—a strategy well-established in the ML-SEM literature (Asparouhov & Muthén, 2019; Hoffman & Walters, 2022; Lüdtke et al., 2008), though a similar formulation also appears in econometrics (Wooldridge, 2010, p. 383). For the sake of readability and to accommodate more complex models later on, we revise the typical multilevel regression convention of using numeric subscripts to differentiate unobserved level-2 exogenous variables (e.g., $U^{Y\mu}$ instead of u_{0j} and U^{YX} instead of u_{1j}). Observed and unobserved random variables are written in upper-case Latin variables, while parameters are written in Greek. The following equations recast the bivariate multilevel regression again in three parts: the level-1 equations with varying parameters, the auxiliary equations decomposing heterogeneous parameters into a structured collection of level-2 (observed and unobserved) causes and constant population-level parameters, and the resulting structural equations obtained by substituting the latter into the former. The level-1 equations describe the data-generating process using varying parameters:

$$\begin{aligned} X_{ij} &= \eta_j^{X\mu} + E_{ij}^X \\ Y_{ij} &= \eta_j^{Y\mu} + \eta_j^{YX} X_{ij} + E_{ij}^Y \end{aligned} \tag{8}$$

where $\eta_j^{X\mu}$ is the varying mean of X_{ij}, $\eta_j^{Y\mu}$ the varying intercept of Y_{ij} (equal to the group mean of Y only when X is centered), and η_j^{YX} is the varying causal effect. The varying parameters can be expressed by the following *auxiliary equations*:

$$\begin{aligned} \eta_j^{X\mu} &= \mu^X + U_j^{X\mu} \\ \eta_j^{Y\mu} &= \mu^Y + U_j^{Y\mu} \\ \eta_j^{YX} &= \beta^{YX} + U_j^{YX} \end{aligned} \tag{9}$$

By substituting Equation (9) into Equation (8) we obtain the mixed *structural* equations:

$$\begin{aligned} X_{ij} &= \mu^X + U_j^{X\mu} + E_{ij}^X \\ Y_{ij} &= (\mu^Y + U_j^{Y\mu}) + (\beta^{YX} + U_j^{YX}) X_{ij} + E_{ij}^Y \\ &= \mu^Y + \beta^{YX} X_{ij} + U_j^{YX} X_{ij} + U_j^{Y\mu} + E_{ij}^Y \end{aligned} \tag{10}$$

In the mixed equation, we can clearly distinguish between *endogenous deterministic variables* (X, Y), *exogenous random variables* (U_j, E_{ij}), and *exogenous population-level causal parameters* (μ^X, μ^Y, and β^{YX}). The grand mean of X is μ^X (i.e., the marginal expectation $\mathbb{E}[X]$). If X is grand-mean centered, μ^Y corresponds to the grand mean of Y; otherwise it is the expected value of Y when $X = 0$. The ATE of $X_{ij} \to Y_{ij}$ is β^{YX}. The terms E_{ij} and U_j capture the variation due to exogenous causes that are *unique* to each

[4] More generally, this idea connects to the concept of (conditional) exchangeability, which provides a nonparametric representation for causal inference theory in cross-sectional, longitudinal (Saarela et al., 2023), hierarchical (Guo, Tóth, et al., 2024; Guo, Zhang, et al., 2024; Weinstein & Blei, 2024) and panel data (Athey & Imbens, 2025). A discussion of this literature is beyond the scope of this paper. However, note that our approach can be interpreted as a parametric and frequentist perspective on this literature.

level-1 unit and are *shared* by all level-1 units in a given level-2 unit, respectively. In other words, E_{ij} represents the collection of all unobserved causes at level-1 and U_j represents all unobserved causes at level-2. For a general causal theory of errors, see van Bork et al. (2024); for a causal perspective on errors in multilevel models, see Skrondal and Rabe-Hesketh (2004), Bates et al. (2014), Skrondal and Rabe-Hesketh (2022), and Kim and Frees (2006).

The exogenous errors are assumed to be generated by the following multivariate normal distributions:

$$\begin{bmatrix} U_j^{X\mu} \\ U_j^{Y\mu} \\ U_j^{YX} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{U^{X\mu}} & & \\ \psi_{U^{Y\mu}U^{X\mu}} & \psi_{U^{Y\mu}} & \\ \psi_{U^{YX}U^{X\mu}} & \psi_{U^{YX}U^{Y\mu}} & \psi_{U^{YX}} \end{bmatrix} \right) \\ \begin{bmatrix} E_{ij}^X \\ E_{ij}^Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{E^X} & \\ \sigma_{E^YE^X}\sigma^2_{E^Y} & \end{bmatrix} \right) \tag{11}$$

Let $\mathbf{U}_j = [U_j^{X\mu}, U_j^{Y\mu}, U_j^{YX}]$, such that $\mathbf{U}_j \sim N(\mathbf{0}, \boldsymbol{\Psi})$ and $\mathbf{E}_{ij} = [E_{ij}^X, E_{ij}^Y]$, such that $\mathbf{E}_{ij} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, resulting in:

$$\begin{bmatrix} \mathbf{U}_j \\ \mathbf{E}_{ij} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Psi} & \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix} \right) \tag{12}$$

The diagonal elements of Ψ_U (i.e., $\psi_{U^{X\mu}}$, $\psi_{U^{Y\mu}}$, $\psi_{U^{YX}}$) represent the variances of each exogenous variable at level-2. The off-diagonal elements are the covariances between exogenous variables. The covariances of exogenous causes between different variables (i.e., $\psi_{U^{X\mu}U^{Y\mu}}$ and $\psi_{U^{YX}U^{X\mu}}$) represent unexplained covariation due to shared unobserved causes (i.e., they represent unobserved confounding at level-2). In contrast, $\psi_{U^{YX}U^{Y\mu}}$ does not represent unobserved confounding but rather captures the extent to which unobserved causes jointly influence the error terms belonging to the same variable. Equivalently, $\sigma^2_{E^X}$ and $\sigma^2_{E^Y}$ represent the variances of the exogenous causes at level-1 and $\sigma_{E^YE^X}$ represents the covariance due to shared unobserved causes at level-1. The zero on the off-diagonal in Equation (12) encodes the assumption that the errors $\mathbf{U}_j$ and $\mathbf{E}_{ij}$ are independent. If we do not explicitly model the level-2 variation—that is, if we marginalize the distribution of the group-level exogenous variables—all of that variation and covariation is absorbed by the exogenous variables at level-1. In this case, the level-1 errors would capture not just unique and unidentifiable but also structured and identifiable variation, such that we would lose the ability to adjust for unobserved confounders at level-2.

6. From Multilevel Structural Equations to Causal Graphs

Given the structural equations of the ML-SCM defined in Equation (10), we can construct the associated causal graph. By definition, the causal graph of an SCM is determined by the functional dependencies: each endogenous variable receives directed edges from all variables appearing on the right-hand side of its structural equation. The mixed equations clearly separate the endogenous variables, exogenous variables, and exogenous population-level parameters, allowing us to directly identify the parent–child relationships that define the graph. Together with the joint distribution over the exogenous variables, we can construct an ADMG that includes both directed and bidirected edges.

To facilitate the presentation, we reiterate the expanded form of Equation (10):

$$\begin{aligned} X_{ij} &= \mu^X + U_j^{X\mu} + E_{ij}^X \\ Y_{ij} &= \mu^Y + \beta^{YX}X_{ij} + U_j^{YX}X_{ij} + U_j^{Y\mu} + E_{ij}^Y \end{aligned} \tag{13}$$

To identify the parent-child relationships, we can write each variable as a function of its unique exogenous and endogenous causes, omitting exogenous population-level parameters:

$$\begin{aligned} X_{ij} &= f_X(U_j^{X\mu}, E_{ij}^X) \\ Y_{ij} &= f_Y(X_{ij}, U_j^{Y\mu}, U_j^{YX}, E_{ij}^Y) \end{aligned} \tag{14}$$

After fixing a given cluster j, we can expand the system of equations across all level-1 units $i = 1, \ldots, N_j$, where N_j denotes the number of observations within cluster j. Since j is fixed, we drop the subscript to simplify notation.

$$\begin{aligned} X_1 &= \mu^X + U^{X\mu} + E_1^X \\ Y_1 &= \mu^Y + \beta^{YX}X_1 + U^{YX}X_1 + U^{Y\mu} + E_1^Y \\ X_2 &= \mu^X + U^{X\mu} + E_2^X \\ Y_2 &= \mu^Y + \beta^{YX}X_2 + U^{YX}X_2 + U^{Y\mu} + E_2^Y \\ &\vdots \\ X_{N_j} &= \mu^X + U^{X\mu} + E_{N_j}^X \\ Y_{N_j} &= \mu^Y + \beta^{YX}X_{N_j} + U^{YX}X_{N_j} + U^{Y\mu} + E_{N_j}^Y \end{aligned} \tag{15}$$

In the expanded equations, we can see how repeated observations at level 1 (within a fixed cluster) are structurally determined by both unique and shared exogenous variables. Specifically, each unit-specific endogenous variable X_i depends on a unit-specific exogenous variable E_i^X and a shared exogenous variable $U^{X\mu}$ that enters the equations additively. Similarly, each Y_i depends on the unit-specific endogenous variable X_i, a shared additive exogenous variable $U^{Y\mu}$, a shared multiplicative (effect-modifying) exogenous variable U^{YX}, and a unit-specific exogenous variable E_i^Y. Notably, the population-level causal effect parameter β^{YX} is identical across all units i and j. Any deviation from this parameter is captured explicitly by the exogenous (unobserved) moderator variable U^{YX}. The mean/intercept parameters (μ^X, μ^Y) have no implications for the causal network—they describe baseline expectations rather than causal relationships—and are conventionally omitted from causal graphs.

The repetition across equations reflects a core feature of multilevel models: shared causes appear in multiple equations, connecting repeated observations and creating dependencies between them. The distributional assumptions for the exogenous variables defined in Equations (10), (11), and (12) determine the presence of bidirected edges in the graph. Specifically, any nonzero covariance between two exogenous variables implies shared latent causes and is represented graphically as a bidirected edge, reflecting unobserved confounding. For instance, if $\mathrm{Cov}(U^{Y\mu}, U^{X\mu}) \neq 0$, we draw a bidirected edge connecting those variables. In this

example, we assume that $\text{Cov}(E_i^X, E_i^Y) = \sigma_{E^Y E^X} = 0$ (i.e., no unobserved confounding of $X \rightarrow Y$ at level-1). Typically, causal graphs omit all exogenous variables, resulting in bidirected edges directly connecting the influenced endogenous variables. However, because multilevel models explicitly distinguish between shared and unique exogenous causes, we adopt a hybrid graphical convention: we omit unique exogenous variables but explicitly retain the shared exogenous variables in the graph. This choice reflects a fundamental difference in identifiability and estimability: shared exogenous variables (e.g., $U^{Y\mu}$, $U^{X\mu}$), and their covariance matrix, can be identified and estimated (with parametric assumptions) from repeated observations, while unique exogenous terms remain unidentifiable and unestimable and are omitted accordingly. Explicitly modeling shared exogenous variables and their covariance matrix (represented graphically as bidirected edges) enables the statistical adjustment for level-2 unobserved confounders (under the assumption that the functional form and parametric assumptions are correct).

The resulting DAG is illustrated in Figure 5. Panel A in Figure 5 presents the fully expanded DAG, corresponding to the expanded set of equations in Equation (15), explicitly drawing all level-1 units within a given level-2 unit. Panel B in Figure 5 presents the same DAG using plate notation. The level-1 variables are enclosed in a plate indexed by i, reflecting their replication within each level-2 unit j, whose group-level variables lie outside the plate. More generally, a plate represents a set of repeated, (conditionally) independent elements indexed by the label in its lower right corner. All variables within the plate are replicated once for each level of the indexing unit (e.g., for each level-1 unit $i = 1, \ldots, N_j$), reflecting a multistage sampling relationship (Jensen et al., 2020). When plates are nested, they express a hierarchical structure: for each instance of the outer index (e.g., group j), there are multiple instances of the inner index (e.g., observations i). The outer plate encodes the assumption that the highest units in the hierarchy (here level-2) are independent from each other, while the inner plates encode the assumption that the lower levels (here just level-1) are independent conditional on all level-2 exogenous (and possibly endogenous) variables.

Because the graph belongs to the class of ADMGs, we can reason with graph-based tools—do-operator, do-calculus,[5] and graphical criteria—to define and reason about causal quantities. The do-operator corresponds to intervening on X at value x, yielding the interventional distribution (Gische & Voelkle, 2022). In graphical terms, this corresponds to removing all incoming arrows to X (Pearl, 2009). The corresponding interventional expectation $\mathbb{E}[Y|do(X = x)]$ is the expected outcome in the population if X were fixed to x for every person. In a single-level model, this expectation is taken over all individual-level exogenous causes E^Y. In multilevel models, however, the function for Y contains both unit-specific and cluster-specific exogenous causes. The post-intervention expectation in multilevel models must therefore average over both unique and shared causes of Y (see Guo, Zhang, et al., 2024; Skrondal & Rabe-Hesketh, 2022).

Formally, with a Gaussian random-intercept and random-slope the post-intervention expectation can be defined as:

$$\begin{aligned}\mathbb{E}[Y|\text{do}(X = x)] &= \int \mathbb{E}\left[Y|x, \mathbf{U}^Y\right] p(\mathbf{U}^Y) d\mathbf{U}^Y \\ &= \mu^Y + \beta^{YX} x,\end{aligned} \tag{16}$$

where $\mathbf{U}^Y = [U^{Y\mu}, U^{YX}]$. Because all exogenous causes have mean zero, the interventional expectation reproduces the fixed part of the mixed model in Equation (10). The population-level ATE is then defined as a contrast between two such expectations $\text{ATE}_{x \rightarrow x'} = \mathbb{E}[Y|do(X = x')] - \mathbb{E}[Y|do(X = x)]$. For a continuous outcome and treatment, this represents a dose-response relationship: the expected change in Y when moving from the interventional level $\text{do}(X = x)$ to $\text{do}(X = x')$ (see Equation (1)). In a linear-Gaussian model with mean zero exogenous causes, the $\text{ATE}_{x \rightarrow x'} = \beta^{YX}(x' - x)$, where β^{YX} is the population-level linear slope (i.e., the causal effect per one-unit change in x). The intercept μ^Y cancels out, as it is the same across both treatment conditions. Cluster-specific interventional distributions and causal effects can be defined by conditioning on cluster-specific exogenous variables; see Equation (2) and Gische et al. (2021). Note that the estimand in a multilevel model (the population-level ATE) corresponds to the average of each cluster-specific slope. However, the quantity estimated via regression may diverge from the population-level ATE, as we see in the next chapter (see also Bates et al., 2014; Hamaker, 2024; Imai & Kim, 2019). One might also define the estimand by forming the contrasts of cluster-specific interventional distributions (i.e., CATEs), and then average across each cluster-specific causal effect. While both estimands are equivalent under the Gaussian (identity) link function, they diverge for nonlinear link functions.

Because unobserved confounders can enter via unique *and* shared errors, the do-operator in Equation (16) implies that arrows from E^X and $U^{X\mu}$ into X are deleted. If direct intervention is not possible, we need to account for the following possible backdoor paths[6] between X and Y in Figure 5:

$$X \leftarrow U^{X\mu} \leftrightarrow U^{Y\mu} \rightarrow Y \tag{17}$$

and

$$X \leftarrow U^{X\mu} \leftrightarrow U^{YX} \rightarrow Y \tag{18}$$

To identify the causal effect $X \rightarrow Y$ we need to find a minimal adjustment set, which is the smallest set of variables that blocks all back-door paths—no variable in it can be omitted without losing that property. In our example, the graph shows that the minimal adjustment sets are U^X and U^Y, U^{YX}, each of which blocks all back-door paths from X to Y. Because causal graphs abstract away the sample size,

[5]Note that the do-calculus is *not* complete for hierarchical multilevel models. Instead, specific causal structures, such as the hierarchical instrumental variables graph, require parametric identification. However, this is beyond the scope of this paper; but see Weinstein and Blei (2024).

[6]Note that the backdoor paths can be traced equivalently in panel **A** or **B** due to the implied equality restrictions.

both of these adjustment sets can be intuitively interpreted as the requirement of accounting for the multilevel structure in some way. However, real studies have to deal with sample size limitations of the number of units in the clusters (N_j) and/or the number of clusters M. For a more in-depth discussion of possible statistical adjustment techniques (such as cluster-mean centering, including the cluster mean or multivariate techniques) we refer the reader to Wooldridge (2010), Bates et al. (2014), Hamaker and Muthén (2020), and Skrondal and Rabe-Hesketh (2022).

7. Empirical Example: High School and Beyond (HSB)

We turn to an empirical example: the High School and Beyond (HSB) study (Raudenbush & Bryk, 2002), capturing nationally representative data on students (level-1) nested within schools (level-2) in the United States. To facilitate the presentation, we only consider three variables: student socioeconomic status (SES) X_{ij}, student academic achievement Y_{ij}, and school type (private vs. public) W_j. The empirical example serves two purposes. First, it illustrates an often overlooked backdoor path and form of school-level heterogeneity: schools may differ not only in their average socioeconomic status (SES) and achievement, and in how strongly SES influences achievement, but also in the diversity of SES among students (i.e., *school-specific heteroskedasticity*). Second, it extends our discussion to include level-2 variables. We assume that higher SES causes higher achievement ($X_{ij} \rightarrow Y_{ij}$), as students with more financial and social resources typically experience better academic outcomes (Morgan et al., 2009). School type influences achievement ($W_j \rightarrow Y_{ij}$), reflecting peer effects, funding differences, and curriculum quality (Aikens & Barbarin, 2008). We assume that SES is not a cause of school type itself ($X_{ij} \not\rightarrow W_j$). Instead, we assume that the correlation $X_{ij} \leftrightarrow W_j$ arises from selection bias ($U_j^{X\mu} \leftrightarrow W_j$), since families with greater resources on average are more able to afford private education (i.e., pupils self-select into school type). In the remainder of the section, we will specify the SCM as a multivariate location-scale model,[7] which allows us to extend our discussion to an additional backdoor path via cluster-specific heteroskedasticity of the treatment.

First, we specify equations with varying parameters:

$$\begin{aligned} X_{ij} &= \eta_j^{X\mu} + \exp(\eta_j^{X\sigma})E_{ij}^X \\ Y_{ij} &= \eta_j^{Y\mu} + \eta_j^{YX}X_{ij} + E_{ij}^Y \\ W_j &= \mu^W + U_j^W \end{aligned} \tag{19}$$

These equations encode the assumption that schools differ in their (1) mean level of SES (X) and achievement (Y), (2) causal effect of SES on achievement, and (3) school-specific variance of SES. For the binary level-2 variable W_j, we could use a logit link, specifically, $\Pr(W_j = 1) = \text{logit}^{-1}(\mu^W + U_j^W)$. Note that U_j^W denotes the unique exogenous cause of the level-2 variable W, which can be understood as representing unobserved (institutional or policy) causes of school type. It is unique to each school-level observation and therefore unidentifiable and unestimable. The newly introduced scale or random variance part of the model is $\eta_j^{X\sigma}$. The exponential function constrains the random variance parameter to be positive. Note that the parameter σ^X lives on the log-scale because of the exponential link and as such does **not** equal the average variance of X across schools.

Second, we define the auxiliary equations of the varying parameters across schools:

$$\begin{aligned} \eta_j^{X\mu} &= \mu^X + U_j^{X\mu} \\ \eta_j^{X\sigma} &= \sigma^X + U_j^{X\sigma} \\ \eta_j^{Y\mu} &= \mu^Y + \gamma^{YW}W_j + U_j^{Y\mu} \\ \eta_j^{YX} &= \beta^{YX} + \gamma^{YXW}W_j + U_j^{YX} \end{aligned} \tag{20}$$

The auxiliary equations encode that all variation of the cluster-specific means and variances of SES is due to unobserved causes ($U_j^{X\mu}$, $U_j^{X\sigma}$), since we assumed that school type is not a direct cross-level cause of SES. However, we assume that school type is an observed cause of both the mean level of achievement and the effect heterogeneity of SES on achievement (Lee & Bryk, 1989). Note that we did not introduce a random variance term for Y because this keeps the model more readable and $U_j^{Y\sigma}$ does not imply an additional backdoor path for *directed* causal effects (for an algebraic treatment of the undirected correlation coefficient, see Hamaker, 2023).

Substituting these into Equation (19) yields the mixed structural equations:

$$\begin{aligned} X_{ij} &= \mu^X + U_j^{X\mu} + \exp(\sigma^X + U_j^{X\sigma})E_{ij}^X \\ Y_{ij} &= (\mu^Y + \gamma^{YW}W_j + U_j^{Y\mu}) + (\beta^{YX} + \gamma^{YXW}W_j + U_j^{YX})X_{ij} + E_{ij}^Y \\ &= \mu^Y + \beta^{YX}X_{ij} + \gamma^{YW}W_j + \gamma^{YXW}W_jX_{ij} + U_j^{YX}X_{ij} + U_j^{Y\mu} + E_{ij}^Y \\ W_j &= \mu^W + U_j^W \end{aligned} \tag{21}$$

The probability distributions over the exogenous variables are:

$$\begin{bmatrix} U_j^{X\mu} \\ U_j^{Y\mu} \\ U_j^{YX} \\ U_j^W \\ U_j^{X\sigma} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{U^{X\mu}} & & & & \\ \psi_{U^{Y\mu}U^{X\mu}}\psi_{U^{Y\mu}} & & & & \\ \psi_{U^{YX}U^{X\mu}}\psi_{U^{YX}U^{Y\mu}}\psi_{U^{YX}} & & & & \\ \psi_{U^WU^{X\mu}}\psi_{U^WU^{Y\mu}}\psi_{U^WU^{YX}}\psi_{U^W} & & & & \\ \psi_{U^{X\sigma}U^{X\mu}}\psi_{U^{X\sigma}U^{Y\mu}}\psi_{U^{X\sigma}U^{YX}}\psi_{U^{X\sigma}U^W}\psi_{U^{X\sigma}} & & & & \end{bmatrix} \right)$$

$$\begin{bmatrix} E_{ij}^X \\ E_{ij}^Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ \sigma_{E^YE^X} & \sigma_{E^Y}^2 \end{bmatrix} \right) \tag{22}$$

[7] Usually location-scale models are specified just for the outcome variable (Feng & Hancock, 2024; Hedeker et al., 2008; McNeish, 2023). For location-scale models in the context of SCMs see Immer et al. (2022). However, in this paper we will specify it for the treatment variable, in order to give a structural and graphical interpretation to a series of findings on the inconsistency of statistical adjustment for cluster-level unobserved confounding (Bates et al., 2014; Brumback et al., 2015; Chernozhukov et al., 2013; Imai & Kim, 2019; Wooldridge, 2005, 2010). Note that the inconsistency can be statistically adjusted for, but necessitates specific techniques (e.g., Bates et al., 2014) or a multivariate location-scale model, where the dependence between $U_j^{X\sigma}$ and U_j^{YX} can be explicitly modeled and thus, statistically adjusted for.

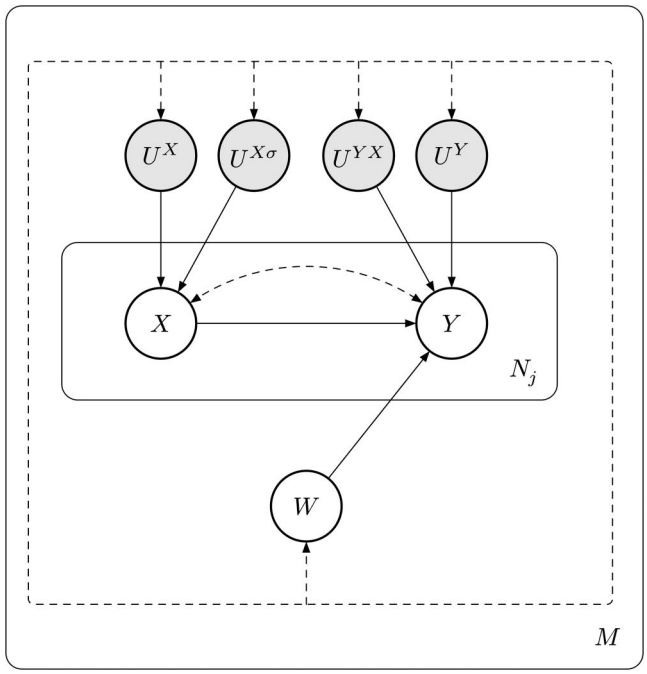

Figure 6. Graph of the high school and beyond ML-SCM. *Note.* The graph is drawn at the school level. The plate around the entire graph encodes the assumption that the graph is identical for each school *j* and the schools are unconditionally independent from each other. The rectangular bidirected arrow reflects potential correlations among all school-level exogenous variables due to shared unobserved causes. Unique errors, at level-1 and level-2, are omitted and their arrows are attached to their only and unique child.

where $\sigma^2_{E^X}$ is set to one, because it is scaled by the random heteroskedastic term $U_j^{X\sigma}$. Finally, the joint probability distribution P_θ over the exogenous variables is:

$$\begin{bmatrix} \mathbf{U}_j \\ \mathbf{E}_{ij} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{\Psi} & \\ \mathbf{0} & \mathbf{\Sigma} \end{bmatrix} \right) \tag{23}$$

The SCM graph with plate notation is shown in Figure 6. Because the HSB study is observational, we assume unobserved confounders at level-1 and level-2. As discussed earlier, only identifiable shared exogenous variables are explicitly shown in the graph; bidirected edges from marginalized unique errors are attached directly to their observed children. For example, the dashed arrow between X_{ij} (student SES) and Y_{ij} (student achievement) represents the correlation σ^{XY} between the unique exogenous causes E_{ij}^X and E_{ij}^Y. This correlation is nonzero in the presence of shared unobserved causes (such as family involvement, parental financial resources, and education). Because these unique exogenous causes cannot be identified and are therefore not shown in the graph, we cannot statistically separate the confounding effect via the backdoor path

$$X_{ij} \leftarrow E_{ij}^X \leftrightarrow E_{ij}^Y \rightarrow Y_{ij} \tag{24}$$

from the direct causal effect $X_{ij} \rightarrow Y_{ij}$. A similar logic applies to causal effects between levels. The backdoor path

$$W_j \leftarrow U_j^W \leftrightarrow U_j^{Y\mu} \rightarrow Y_{ij} \tag{25}$$

is open when the correlation $\psi_{U^W U^{Y\mu}}$ between the unobserved school-level causes U_j^W and $U_j^{Y\mu}$ is nonzero—encoded via the dashed rectangle. For example, especially motivated and achieving students or parents with high achievement standards might self-select into private schools, leading to a correlation between average achievement levels and school types. As each school contributes only one observation of W_j, the associated error term U_j^W is not identifiable and therefore omitted from the graph. Because of this, we cannot statistically separate the correlation $\psi_{U^W U^{Y\mu}}$ that parametrizes $U^W \leftrightarrow U^{Y\mu}$ from the direct causal effect γ^{YW} that parametrizes $W_j \rightarrow Y_{ij}$.

Some schools may have a more socioeconomically diverse student body, reflected in between-school differences in the variance of SES. We represent this using a school-level exogenous variable, $U_j^{X\sigma}$, which captures all unobserved causes of school-specific heteroskedasticity in student SES and enters the structural model by scaling the residual variance. Because multiple students are observed per school, the effect of $U_j^{X\sigma}$ on the distribution of SES is identifiable, and the exogenous variable is therefore explicitly represented in the graph. As with the other school-level exogenous variables, $U_j^{X\sigma}$ is allowed to correlate with all other unobserved school-level causes—indicated by the rectangular bidirected arrow. Admission policies offer a plausible example: they may simultaneously affect SES diversity ($U_j^{X\sigma}$), average achievement ($U_j^{Y\mu}$), and the strength of the SES–achievement relationship (U_j^{YX}). Inclusive policies increase socioeconomic heterogeneity, while restrictive ones produce homogeneous student bodies and concentrate high-SES students. Because higher-SES students tend to perform better on average, these policies therefore also shift the school-level average of academic achievement. This creates a backdoor path:

$$X \leftarrow U^{X\sigma} \leftrightarrow U^{Y\mu} \rightarrow Y \tag{26}$$

Moreover, the SES–achievement relationship may itself depend on SES diversity: in homogeneous schools, the link may be weaker due to limited SES variation; in more diverse schools, it may be amplified by disparities in resources, support, or expectations. Alternatively, schools with greater diversity may invest in equity programs that mitigate SES effects, whereas more homogeneous schools may lack the perceived need for such efforts and divert resources elsewhere, for example toward enrichment programs that benefit already advantaged students. Either way, this implies another backdoor path:

$$X \leftarrow U^{X\sigma} \leftrightarrow U^{YX} \rightarrow Y \tag{27}$$

These examples show that the same unobserved factor (e.g., admission policy) may jointly influence $U_j^{X\sigma}$, $U_j^{Y\mu}$, and U_j^{YX}. Even if not all confounders affect all three, the possibility of shared unobserved causes between any pair can create these backdoor paths that bias the estimation of the causal effect $X_{ij} \rightarrow Y_{ij}$.

In psychology and educational research, it is common to center only the covariates, as this preserves the ability to estimate the effects of level-2 variables (Enders, 2013; Kreft et al., 1995). In contrast, in econometrics it is common to center both the covariates and the outcome (Bates et al., 2014; Wooldridge, 2010). While these approaches are

equivalent for adjusting for the sources of confounding discussed in the previous section, this equivalence breaks down when school-specific heteroskedasticity is present. When we group mean center Y, we cut the path $U^{Y\mu} \rightarrow Y$ such that the path $X \leftarrow U^{X\sigma} \leftrightarrow U^{Y\mu} \rightarrow Y$ is blocked. However, when we center just X, this path is *not* blocked, because centering X does not cut $U^{X\sigma} \rightarrow X$ and $U^{Y\mu} \rightarrow Y$ remains. Even when we center both X and Y, the paths $U^{YX} \rightarrow Y$ and $U^{X\sigma} \rightarrow X$ remain, such that the path $X \leftarrow U^{X\sigma} \leftrightarrow U^{YX} \rightarrow Y$ is *open* (see also Bates et al., 2014; Imai & Kim, 2019).

Furthermore, Consider the Backdoor Path

$$X \leftarrow U^{X\sigma} \leftrightarrow U_j^W \leftrightarrow U^{YX} \rightarrow Y \tag{28}$$

passing through U_j^W—a unique level-2 error and as such not visible in Figure 6, where the bidirected arrows connected to U_j^W are therefore attached to its child W. Due to the rectangular representation of all bidirected arrows in the graph it is easy to miss that $U^{X\sigma} \leftrightarrow W_j \leftrightarrow U^{YX}$ is a collider, and as such, a closed path—as long as we do not include W in the statistical model. However, if we include W in our model, we open this previously blocked path and consequently risk introducing collider bias into the causal effect estimation of $X \rightarrow Y$. Conversely, omitting W allows its variation to be captured by cluster-level noise terms, thus maintaining a closed path. This example highlights a broader principle in multilevel modeling: appropriate statistical adjustment can account for unobserved confounders without inadvertently creating collider or mediator bias (Jensen et al., 2020). In other words, if W were both a confounder and a collider, we could eliminate the bias through the shared exogenous causes without introducing collider bias.

8. Discussion

In this paper, we studied the integration of cross-sectional multilevel models with structural causal models and causal graphs. We first demonstrated that common path diagram conventions for linear Gaussian multilevel models do not adhere to the strict semantics of graph theory. Instead, these diagrams primarily serve as statistical visualizations, limiting researchers' ability to define precise causal estimands and to utilize modern graphical identification tools.

To bridge this gap, we explicitly showed how a standard linear Gaussian multilevel regression model can be reinterpreted as a structural causal model. By substituting auxiliary equations for varying parameters, each mixed model equation can be seen as a structural equation where each variable is explicitly defined by its endogenous and exogenous causal parents. This structural viewpoint of mixed model equations naturally leads to a decomposition of all variables into distinct level-specific exogenous components. We demonstrated this by fixing a given cluster in the multilevel SCM and showing that the resulting expanded equations directly map onto a causal graph, closely aligning with the wide-format ML-SEM perspective (Barendse & Rosseel, 2020, 2023; Hamaker & Muthén, 2020). Plate notation allows us to collapse structurally identical subgraphs into a compact representation that preserves the full causal semantics, allowing causal graphical criteria to be applied just as on the expanded graph.

A distinguishing feature of the proposed approach is that the causal graphs are formulated in wide format, representing each cluster as a collection of units rather than as a single node. This representation emerges naturally from a structural interpretation of the mixed model equations and makes the dependencies among cluster-specific means, slopes, and variances explicit. By contrast, treating clusters as a categorical variable and thus as a node—an approach common in conventional causal graphs—conceals the underlying network of confounding.

Although group-mean centering is widely advocated as the standard approach for addressing cluster-level confounding (Hamaker & Muthén, 2020), our graphical analysis demonstrated critical limitations of this practice. Even after group-mean centering both covariates and outcomes, backdoor paths due to cluster-specific heterogeneity in slopes and variances can remain open. This is because treatments in multilevel models have effectively two components that are relevant for backdoor paths: cluster-specific means *and* variances. Furthermore, we showed how adjusting for level-2 covariates correlated with these exogenous causes can introduce collider bias, creating new confounding pathways rather than eliminating them. Thus, our structural framework encourages researchers to carefully reflect on the dependency structures among random effects and systematically question whether commonly used adjustments, such as centering, sufficiently control for specific forms of heterogeneity.

Furthermore, by omitting unique errors and including shared errors, our graphical representation directly visualizes the confounding we can statistically adjust for and those that remain unadjustable. Specifically, unobserved confounding among level-1 components cannot be removed, while confounding among the level-2 parts of level-1 variables can be removed (Weinstein & Blei, 2024); cross-level confounding (between level-1 and level-2 components), however, remains unadjustable with purely statistical methods.

While we focused on linear Gaussian models to align closely with current multilevel modeling practices, our causal graphical framework generalizes to accommodate more complex models. Directed arrows and bidirected arrows can represent nonlinear dependencies (Brumback et al., 2010), where structural equality restrictions can still be nonparametrically encoded via plates. The distributions of the errors and outcomes may also be non-Gaussian. However, it is important to be aware of the issue of noncollapsibility for nonlinear link functions (Greenland et al., 1999; Kiefer et al., 2024; Skrondal & Rabe-Hesketh, 2022). Furthermore, varying intercepts, slopes, and residual variances can be understood nonparametrically as reflecting distinct features of the joint distribution: heterogeneity in (conditional) expectations, average treatment effects, and (conditional) variances across clusters. This separation of exogenous causes reveals distinct pathways through which

shared confounders may enter the model and threaten identification, enabling researchers to critically reflect on and justify their modeling decisions with greater precision.

This paper deliberately focused on cross-sectional models with a single hierarchical dimension as a foundational step in aligning structural causal models with multilevel modeling. A natural and necessary next step for future research is to extend the structural approach introduced here to longitudinal and cross-nested data. In longitudinal models, temporal ordering matters critically, since past observations directly influence future outcomes through (cross-)lagged causal pathways (Gische et al., 2021; Wooldridge, 2010). Heterogeneity in these additional paths introduces confounding pathways not covered in this paper.

By explicitly incorporating heterogeneity and its associated confounding paths into causal graphs, our structural framework provides behavioral scientists with a rigorous methodological foundation for addressing causal identification challenges inherent in multilevel data. This perspective enriches both methodological rigor and conceptual clarity, ultimately paving the way toward more nuanced and accurate causal theories in the social and behavioral sciences.

Author Contributions

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Moritz Ketzer: conceptualization, writing, editing; Christian Gische: conceptualization, supervision, editing; Manuel C. Voelkle: conceptualization, supervision, editing

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 523691032.

References

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, *100*, 235–251. https://doi.org/10.1037/0022-0663.100.2.235

Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 119–142. https://doi.org/10.1080/10705511.2018.1511375

Athey, S., & Imbens, G. (2025, March 25). *Identification of average treatment effects in nonparametric panel models*. https://doi.org/10.48550/arXiv.2503.19873

Barendse, M. T., & Rosseel, Y. (2020). Multilevel modeling in the 'wide format' approach with discrete data: A solution for small cluster sizes. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*, 696–721. https://doi.org/10.1080/10705511.2019.1689366

Barendse, M. T., & Rosseel, Y. (2023). Multilevel SEM with random slopes in discrete data using the pairwise maximum likelihood. *The British Journal of Mathematical and Statistical Psychology*, *76*, 327–352. https://doi.org/10.1111/bmsp.12294

Bates, M. D., Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Handling correlations between covariates and random slopes in multilevel models. *Journal of Educational and Behavioral Statistics*, *39*, 524–549. https://doi.org/10.3102/1076998614559420

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 135–167. https://doi.org/10.3102/10769986028002135

Bollen, K. A. (1989). *Structural equations with latent variables*. https://doi.org/10.1002/9781118619179

Brumback, B. A., Dailey, A. B., Brumback, L. C., Livingston, M. D., & He, Z. (2010). Adjusting for confounding by cluster using generalized linear mixed models. *Statistics & Probability Letters*, *80*, 1650–1654. https://doi.org/10.1016/j.spl.2010.07.006

Brumback, B. A., Li, L., & Cai, Z. (2015). On the use of between–within models to adjust for confounding due to unmeasured cluster-level covariates. *Communications in Statistics - Simulation and Computation*, *46*, 3841–3854. https://doi.org/10.1080/03610918.2015.1024858

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, *18*, 5–46. https://doi.org/10.1016/0304-4076(82)90094-X

Chernozhukov, V., Fernandez-Val, I., Hahn, J., & Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, *81*, 535–580. https://doi.org/10.3982/ECTA8405

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529–569. https://doi.org/10.1207/s15327906mbr3804_5

Curran, P. J., & Bauer, D. J. (2007). Building path diagrams for multilevel models. *Psychological Methods*, *12*, 283–297. https://doi.org/10.1037/1082-989X.12.3.283

Enders, C. K. (2013). Centering predictors and contextual effects. In M. A. Scott, J. S. Simonoff, & B. D. Marx (Eds.), *The SAGE handbook of multilevel modeling* (pp. 89–108). SAGE Publications. https://doi.org/10.4135/9781446247600.n6

Feng, Y., & Hancock, G. R. (2024). A structural equation modeling approach for modeling variability as a latent variable. *Psychological Methods*, *29*, 262–286. https://doi.org/10.1037/met0000477

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gische, C., & Voelkle, M. C. (2022). Beyond the mean: A flexible framework for studying causal effects using linear models. *Psychometrika*, *87*, 868–901. https://doi.org/10.1007/s11336-021-09811-z

Gische, C., West, S. G., & Voelkle, M. C. (2021). Forecasting causal effects of interventions versus predicting future outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 475–492. https://doi.org/10.1080/10705511.2020.1780598

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, *14*, 29–46. https://doi.org/10.1214/ss/1009211805

Guo, S., Tóth, V., Schölkopf, B., & Huszár, F. (2024). *Causal de Finetti: On the identification of invariant causal structure in exchangeable data*. arXiv:2203.15756. https://doi.org/10.48550/arXiv.2203.15756

Guo, S., Zhang, C., Mohan, K., Huszár, F., & Schölkopf, B. (2024). *Do Finetti: On causal effects for exchangeable data*. *Advances in Neural Information Processing Systems, 37* (NeurIPS 2024). https://doi.org/10.52202/079017-4044

Hamaker, E. L. (2024). The curious case of the cross-sectional correlation. *Multivariate Behavioral Research*, *59*, 1111–1122. https://doi.org/10.1080/00273171.2022.2155930

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*, 365–379. https://doi.org/10.1037/met0000239

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*, 627–634. https://doi.org/10.1111/j.1541-0420.2007.00924.x

Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology*, *73*, 659–689. https://doi.org/10.1146/annurev-psych-020821-103525

Hsiao, C. (2022). *Analysis of panel data*. Cambridge University Press. https://books.google.com/books?hl=de&lr=&id=DHtrEAAAQBAJ&oi=fnd&pg=PR13&dq=(Hsiao,+2014)+panel&ots=QTdbsN0Du1&sig=oI6esApvgPJJQZwb2mqugEoJYNU

Imai, K., & Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science*, *63*, 467–490. https://doi.org/10.1111/ajps.12417

Immer, A., Schultheiss, C., Vogt, J. E., Schölkopf, B., Bühlmann, P., & Marx, A. (2022). *On the identifiability and estimation of causal location-scale noise models*. arXiv. https://doi.org/10.48550/ARXIV.2210.09054

Jensen, D., Burroni, J., & Rattigan, M. (2020). *Object conditioning for causal inference* (pp. 1072–1082). https://proceedings.mlr.press/v115/jensen20a.html

Kiefer, C., Lugauer, B., & Mayer, A. (2024). Definition and identification of causal ratio effects. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000711

Kim, J.-S., & Frees, E. W. (2006). Omitted variables in multilevel models. *Psychometrika*, *71*, 659–690. https://doi.org/10.1007/s11336-005-1283-0

Kim, J.-S., & Frees, E. W. (2007). Multilevel modeling with correlated effects. *Psychometrika*, *72*, 505–533. https://doi.org/10.1007/s11336-007-9008-1

Kim, Y., & Jung, G. (2025). Understanding linear interaction analysis with causal graphs. *The British Journal of Mathematical and Statistical Psychology*, *78*, 486–499. https://doi.org/10.1111/bmsp.12369

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). The Guilford Press.

Kreft, I. G., De Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1–21. https://doi.org/10.1207/s15327906mbr3001_1

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*, 172. https://doi.org/10.2307/2112866

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229. https://doi.org/10.1037/a0012869

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. John Wiley & Sons.

McNeish, D. (2023). A practical guide to selecting and blending approaches for clustered data: Clustered errors, multilevel models, and fixed-effect models. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000620

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*, 259–284. https://doi.org/10.1037/1082-989X.10.3.259

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2009). Risk factors for learning-related behavior problems at 24 months of age: Population-based estimates. *Journal of Abnormal Child Psychology*, *37*, 401–413. https://doi.org/10.1007/s10802-008-9279-8

Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. CRC Press.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica, 46*, 69–85. https://doi.org/10.2307/1913646

Muthén, B., & Asparouhov, T. (2011). *Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework*. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). Routledge. https://doi.org/10.4324/9780203848852-4

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide (eighth)*. Muthén & Muthén.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed., reprinted with corrections). Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect* (1st ed.). Basic Books.

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. The MIT Press.

Pickles, A., Pickering, K., Simonoff, E., Silberg, J., Meyer, J., & Maes, H. (1998). Genetic "clocks" and "soft" events: A twin model for pubertal development and other recalled sequences of developmental milestones, transitions, or ages at onset. *Behavior Genetics, 28*, 243–253. https://doi.org/10.1023/A:1021615228995

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, *21*, 189–205. https://doi.org/10.1037/met0000052

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233. https://doi.org/10.1037/a0020141

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190. https://doi.org/10.1007/BF02295939

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.

Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, *30*, 145–157. https://doi.org/10.1111/1467-9469.00323

Saarela, O., Stephens, D. A., & Moodie, E. E. M. (2023). The role of exchangeability in causal inference. *Statistical Science*, *38*, 369–385. https://doi.org/10.1214/22-STS879

Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components*. John Wiley & Sons. https://books.google.com/books?hl=de&lr=&id=mkCSOOzHMngC&oi=fnd&pg=PP2&dq=variance+component+analysis&ots=RbqmDqWKTE&sig=cSph7_NkQYphGcHI5KzrzT9CMvk

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC.

Skrondal, A., & Rabe-Hesketh, S. (2022). The role of conditional likelihoods in latent variable modeling. *Psychometrika*, *87*, 799–834. https://doi.org/10.1007/s11336-021-09816-8

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., & Glymour, C. (1998). Using path diagrams as a structural equation modeling tool. *Sociological Methods & Research*, *27*, 182–225. https://doi.org/10.1177/0049124198027002003

Twisk, J. W. R. (2019). *Applied mixed model analysis: A practical guide* (2nd ed.). Cambridge University Press. https://www.cambridge.org/core/product/identifier/9781108635660/type/book

van Bork, R., Rhemtulla, M., Sijtsma, K., & Borsboom, D. (2024). A causal theory of error scores. *Psychological Methods*, *29*, 807–826. https://doi.org/10.1037/met0000521

Weinstein, E. N., & Blei, D. M. (2024). *Hierarchical causal models*. https://doi.org/10.48550/arXiv.2401.05330

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, *87*, 385–390. https://doi.org/10.1162/0034653053970320

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press. https://books.google.com/books?hl=en&lr=&id=hSs3AgAAQBAJ&oi=fnd&pg=PP1&dq=Econometric+Analysis+of+Cross+Section+and+Panel+Data&ots=VZNQoCS-Nn&sig=4Dy779DWDgRul9BmLokDOpXEUzo