Taylor & Francis
Taylor & Francis Group

Structural Equation Models: A Review with Applications to Environmental Epidemiology
Author(s): Brisa N. Sánchez, Esben Budtz-Jørgensen, Louise M. Ryan and Howard Hu
Source: *Journal of the American Statistical Association*, Vol. 100, No. 472 (Dec., 2005), pp. 1443-1455
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/27590684
Accessed: 22-04-2020 14:09 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/27590684?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Taylor & Francis, Ltd., American Statistical Association are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

Structural Equation Models: A Review With Applications to Environmental Epidemiology

Brisa N. SÁNCHEZ, Esben BUDTZ-JØRGENSEN, Louise M. RYAN, and Howard HU

Structural equation models (SEMs) have been discussed extensively in the psychometrics and quantitative behavioral sciences literature. However, many statisticians and researchers in other areas of application are relatively unfamiliar with their implementation. Here we review some of the SEM literature and describe basic methods, using examples from environmental epidemiology. We make connections to recent work on latent variable models for multivariate outcomes and to measurement error methods, and discuss advantages and disadvantages of SEMs compared with traditional regressions. We give a detailed example in which two models fit the same data well, yet one is physiologically implausible. This underscores the critical role of subject matter knowledge in the successful implementation of SEMs. A brief discussion on open research areas is included.

KEY WORDS: Correlated outcomes; Latent variable; LISREL model; Measurement error; Multiple testing; Multivariate data.

1. INTRODUCTION

Structural equation models (SEMs) are a flexible class of models that allow complex modeling of multivariate data and multiple, closely related predictors. For example, through the use of latent variables, the SEM framework allows joint modeling of the classical measurement error model for predictors and random-effects models for multivariate outcomes. Although SEMs are commonly thought to necessarily involve latent variables, it is also common to see versions involving only observable variables (e.g., simultaneous equations models in econometrics; Amemiya 1985). Although some authors refer to SEMs as models for multivariate normal data (Bollen 1989), we view the term as defining a broader class of models.

The theory and statistical properties of SEMs are well developed but are scattered throughout several fields of research, particularly education, psychology, sociology, and econometrics (Muthén 1984; Browne and Arminger 1995; Yuan and Bentler 2000). SEM theory has also appeared in mainstream statistical journals under the terms *mean and covariance structures* and *latent variable models* (Jöreskog 1970; Sammel and Ryan 1996; Yuan and Bentler 1997; Bandeen-Roche, Miglioretti, Zeger, and Rathouz 1997; Lee and Shi 2001). An exhaustive review of the existing literature is outside the scope of the present article. Here our aims are to familiarize the reader with the SEM approach and to point to several areas where further statistical development would be useful.

Although there are several ways to specify SEMs (e.g., Jöreskog 1973; Bentler and Weeks 1980), we chose Muthén's (2002) formulation because it clearly distinguishes fixed covariates versus variables that need distributional assumptions. In his formulation, Muthén described SEMs as two-stage models. In addition to allowing relationships between latent variables, the two-stage formulation can easily allow additional covariate adjustment compared with multivariate regression by modeling covariate effects on the exposure as well as their effect on the outcome.

In Section 2 we discuss motivating examples and introduce *path diagrams*, which are an essential tool for structural equation modeling; these diagrams started with Wright's (1921) work on path analysis and follow some of the diagram conventions described by Pearl (1995) and Cox and Wermuth (1996). In Section 3 we introduce SEMs for the case where dependent variables are continuous. In this simplified setting, we discuss fitting procedures, inference, and identifiability issues, and establish that not all observed variables necessitate distributional assumptions as is commonly believed. We discuss misspecification, robustness, and goodness of fit in Section 4. In Section 5 we briefly review two extensions of SEMs, namely incorporating nonnormal dependent variables and modeling nonlinear relationships between latent variables. In Section 6 we make connections to other latent variable methods, including latent class models. Alternate model formulations and a brief note on software are included in Section 7. In Section 8 we present a detailed example and make cautionary points regarding the implementation of SEMs. We conclude with a discussion on the strengths and weaknesses of SEMs and explore areas where further work would be useful.

2. MOTIVATING EXAMPLES AND PATH DIAGRAMS

We use two examples from environmental epidemiology to introduce drawing conventions for path diagrams. Environmental epidemiology researchers are beginning to use SEMs to alleviate modeling challenges in this context; for example, SEMs can easily incorporate measurement error models for exposure measurements and model multiple outcomes (e.g., components of a clinical syndrome) simultaneously.

Path diagrams use various symbols (Fig. 1) to represent model assumptions graphically (Bollen 1989; Lohelin 1992). Variable names are drawn inside boxes or ovals, depending on whether variables are observed or latent. For observed variables, line types (solid or dashed), distinguish between dependent and independent variables. Relationships between variables are described by use of directed (single-direction) and double-headed arrows. Similar to the arrows in the directed acyclical graphs of Pearl (1995) and Cox and Wermuth

Brisa N. Sánchez is a Doctoral Candidate, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02139 (E-mail: *bsanchez@hsph.harvard.edu*). Esben Budtz-Jørgensen is Assistant Professor, Department of Biostatistics, Institute of Public Health, University of Copenhagen, Denmark-N2200 Copenhagen. Louise M. Ryan is Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02139. Sánchez is supported by a Howard Hughes Predoctoral Fellowship in Biological Sciences. This work was also supported in part by the Biostatistics Facility Core of the Kresge Center for Environmental Health (NIEHS ES00002), the Environmental Statistics Core of the Superfund Hazardous Substances Basic Research Program (NIEHS ES99001), NIH grants CA48061, ES07821, and P42 ES05947 Project 1, and ACC grant 2843. The authors thank James H. Ware and the referees for helpful comments.

© 2005 American Statistical Association
Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Review Article
DOI 10.1198/016214505000001005

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

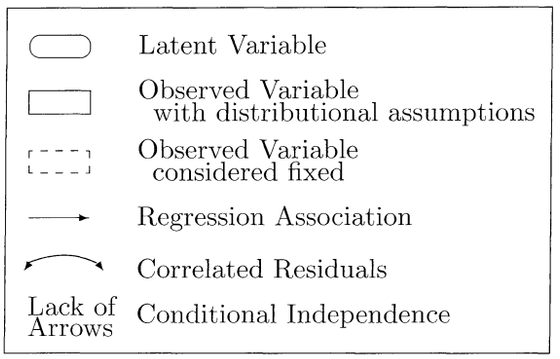

Figure 1. Path Diagram Drawing Conventions.

(1996), directed arrows in path diagrams represent causal relationships among variables. In path diagrams, however, arrows further characterize the form of the relationships as regression associations. Double-headed arrows represent nonzero correlations, for example, residual correlation between two variables left unexplained by common predictors. An absence of directed arrows between two variables means that there are no direct influences between them, whereas lack of a double-headed arrow between two variables implies that the variables are conditionally independent given the predictors in the model. Although we formally introduce the model formulation in Section 3, in this section we use heuristic notation to show how path diagrams translate into model equations.

2.1 Relationships Between Lead Biomarkers

Chuang et al. (2001) analyzed the relationships between lead levels in maternal whole blood, maternal tibia and patella bone, air, and umbilical cord blood using data collected in Mexico City on a cohort of 623 mother–baby pairs. One purpose of this investigation was to explain the variability in cord blood lead while considering all other lead measurements as explanatory variables. A primary goal was to compare the relative effect of each explanatory lead measurement on cord blood lead. In bivariate analyses, each of the lead measurements was significantly associated with cord blood lead (results not shown). Yet it is not clear whether the bivariate regression results were significant due to high correlation with whole blood lead, which, according to physiological explanations, is an intermediate variable. Multivariate analyses including whole blood lead in addition to one of the other explanatory lead measurements resulted in collinearity problems.

The proposed relationships among all lead biomarkers are described by the path diagram in Figure 2 (Chuang et al. 2001). The regression associations implied by the figure can be represented by

$$Y_1 = \alpha_1 + \gamma_{11}Z_1 + \gamma_{12}Z_2 + \gamma_{13}Z_3 + \epsilon_1,$$

$$Y_2 = \alpha_2 + \gamma_{21}Z_1 + \gamma_{22}Z_2 + \gamma_{23}Z_3 + \epsilon_2,$$

$$Y_3 = \alpha_3 + \beta_{31}Y_1 + \beta_{32}Y_2 + \gamma_{34}Z_4 + \gamma_{35}Z_5 + \epsilon_3,$$

and

$$Y_4 = \alpha_4 + \beta_{43}Y_3 + \epsilon_4.$$

As conveyed by the line type used for the variables in the figure, the Z's in these equations can be considered fixed covariates, but Y variables need at least some distributional assumptions to draw inference on the regression coefficients, for instance, a multivariate normal distribution on the errors ϵ.

2.2 Health Effects of Prenatal Exposure to Methylmercury

Grandjean et al. (1997) analyzed the effects of in utero methylmercury exposure on neurodevelopment using multivariate analysis. They used data collected on a cohort of newborn children recruited in the Faroe Islands during 1986–1987 and followed prospectively. Available exposure measurements were the methylmercury concentration in umbilical cord blood (BHg) and the mother's hair (HHg), and the mother's average weekly consumption of pilot whale meals (Whale) during pregnancy. Outcome data included a battery of neuropsychological tests completed on each child at age 7 years. Scores from the following tests were considered in the analysis: Neurobehavioral Examination System Finger Tapping (FT1, FT2, FT3), Hand–Eye Coordination (HEC), Wechsler Intelligence Scale for Children "Revised Digit Spans" (DS), California Verbal Learning Test under four different conditions (CVLT1–CVLT4), and the Boston Naming Test under two conditions (BNT1, BNT2). Many covariates were collected, including child's sex and age, mother's intelligence and socioeconomic status, and child's access to computers, among others.

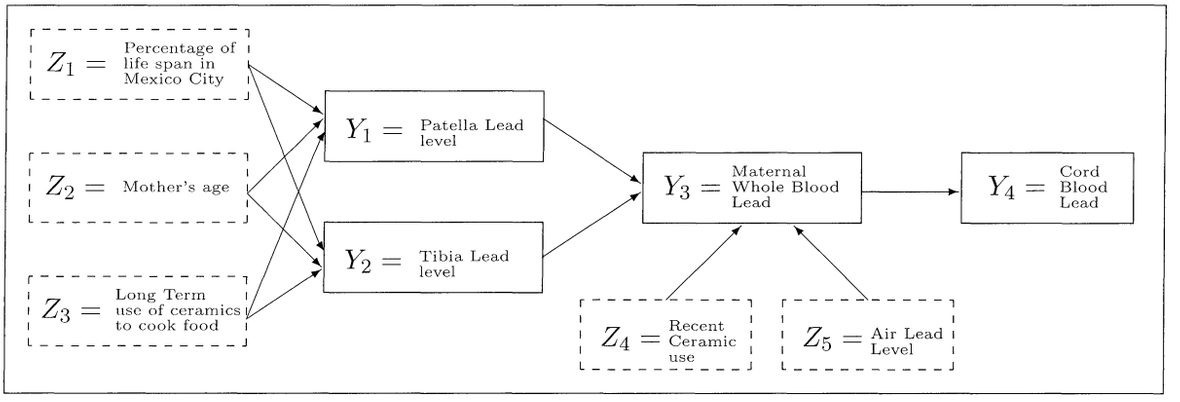

Figure 2. Interrelations Between Lead Biomarkers. Bone lead levels are affected by mother's characteristics and long-term ceramic use. Venous blood lead level depends on bone lead levels, recent ceramic use, and air lead levels. Cord blood lead is regressed on mother's blood lead level.

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

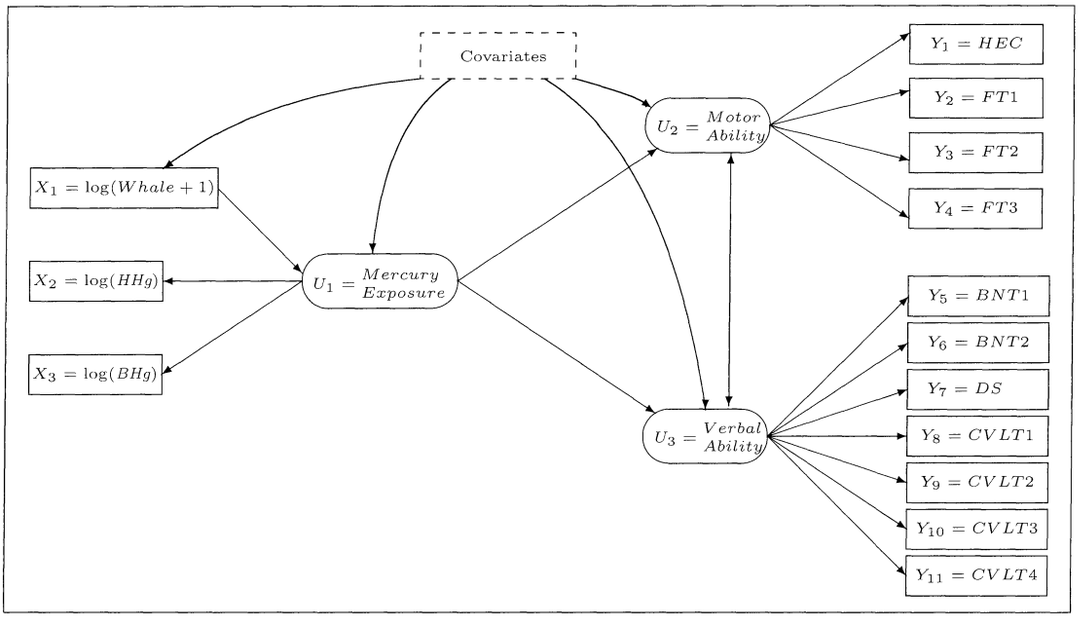

Figure 3. Methylmercury Exposure and Neurobehavioral Functions. Hair mercury (HHg) and child's cord blood methylmercury (BHg) are indicators of latent methylmercury exposure which is associated with neurobehavioral functions after covariate adjustment. Neurobehavioral functions are measured by a battery of neurological tests.

Given the 11 outcome measures and 3 predictors, standard multiple regression would require drawing inference on more than 30 exposure effect parameters.

Budtz-Jørgensen, Keiding, Grandjean, Weihe, and White (2002, 2003) argued for using SEMs as a parsimonious alternative to describe the relationship between methylmercury and neurodevelopment and to correct for measurement error in the exposure variables. Figure 3 describes the proposed relationships between all variables for this example; it introduces methylmercury exposure, motor function, and verbal function as latent variables (indicated by ovals). The latent exposure variable is assumed to be measured with error by HHg and BHg and is influenced by Whale. Latent exposure is assumed to affect latent verbal and motor abilities. Test scores reflect latent motor and verbal brain functions, and are assumed to be conditionally independent given these functions. The double-headed arrow connecting motor and verbal function represents residual correlation between motor function and verbal function not explained by the predictors considered. Suppressing subject index, the relationships in Figure 3 can be represented by the following:

Exposure indicators

$$X_1 = \nu_{x_1} + \boldsymbol{\gamma}_4^t \mathbf{Z} + \epsilon_{x_1},$$
$$X_2 = \nu_{x_2} + \lambda_{21} U_1 + \epsilon_{x_2},$$
$$X_3 = \nu_{x_3} + \lambda_{31} U_1 + \epsilon_{x_3}.$$

Latent exposure

$$U_1 = \alpha_1 + \beta_{14} X_1 + \boldsymbol{\gamma}_1^t \mathbf{Z} + \zeta_1.$$

Latent outcomes

$$U_2 = \alpha_2 + \beta_{21} U_1 + \boldsymbol{\gamma}_2^t \mathbf{Z} + \zeta_2,$$
$$U_3 = \alpha_3 + \beta_{31} U_1 + \boldsymbol{\gamma}_3^t \mathbf{Z} + \zeta_3.$$

Motor function indicators

$$Y_1 = \nu_{y_1} + \lambda_{12} U_2 + \epsilon_{y_1},$$
$$Y_2 = \nu_{y_2} + \lambda_{22} U_2 + \epsilon_{y_2},$$
$$Y_3 = \nu_{y_3} + \lambda_{32} U_2 + \epsilon_{y_3},$$
$$Y_4 = \nu_{y_4} + \lambda_{42} U_2 + \epsilon_{y_4}.$$

Verbal function indicators

$$Y_5 = \nu_{y_5} + \lambda_{53} U_3 + \epsilon_{y_5},$$
$$Y_6 = \nu_{y_6} + \lambda_{63} U_3 + \epsilon_{y_6},$$
$$Y_7 = \nu_{y_7} + \lambda_{73} U_3 + \epsilon_{y_7},$$
$$Y_8 = \nu_{y_8} + \lambda_{83} U_3 + \epsilon_{y_8},$$
$$Y_9 = \nu_{y_9} + \lambda_{93} U_3 + \epsilon_{y_9},$$
$$Y_{10} = \nu_{y_{10}} + \lambda_{10,3} U_3 + \epsilon_{y_{10}},$$
$$Y_{11} = \nu_{y_{11}} + \lambda_{11,3} U_3 + \epsilon_{y_{11}}.$$

In this model, the effect of methylmercury can be described by its overall effect on latent motor ability (β_{21}) and on latent verbal ability (β_{31}). Note that in addition to addressing multiple testing concerns associated with regression analyses, the model also corrects for measurement error bias in the exposure coefficients.

We deviate from the traditional path diagrams described in the social sciences (Bollen 1989) in two ways. First, we omit the inclusion of variables explicitly representing unexplained residual variation and random error. In traditional diagrams, Figure 2 would have additional arrows pointing toward tibia and patella bone lead, whole blood lead, and cord blood lead to indicate that those variables are not perfectly explained by their predictors. Second, we adopt the use of dashed-line boxes to distinguish between variables with and without distributional assumptions.

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

3. STRUCTURAL EQUATION MODELS, LINEAR CASE

In the general model formulation, we consider settings where predictor or outcome variables are not directly observed, but are instead measured with error using multiple instruments. For the ith of n independent units, let $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{im})^T$ represent m surrogate predictors for one or more latent predictors of interest, let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iq})^T$ represent q multiple surrogates for latent outcome(s) of interest, and let $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{ip})^T$ represent p covariates. Finally, let $\mathbf{U}_i$ be an ℓ-dimensional random vector that represents all latent variables.

The first stage of the model, called the *measurement part* in the social sciences, relates dependent variables ($\mathbf{X}_i$ and $\mathbf{Y}_i$) to the latent variables and fixed covariates by

$$\begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{U}_i + \mathbf{K}\mathbf{Z}_i + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\boldsymbol{\nu}_{(q+m)\times 1}$, $\boldsymbol{\Lambda}_{(q+m)\times \ell}$, and $\mathbf{K}_{(q+m)\times p}$ are parameter matrices. The vector of random errors $\boldsymbol{\epsilon}_i$ is multivariate normal with $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $\operatorname{cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$ and captures the deviation of the dependent variables from their conditional mean, including measurement error.

The second stage of the model, elsewhere called the *structural part*, defines a linear structure between the latent variables,

$$U_{ig} = \alpha_g + \sum_{h \neq g} \beta_{gh} U_{ih} + \sum_j \gamma_{gj} Z_{ij} + \zeta_{gi}, \tag{2}$$

where U_{ig} is the gth element of $\mathbf{U}_i$; that is, a latent variable, U_{ig} can be influenced by the covariates and other latent variables except itself. In matrix form, (2) is

$$\mathbf{U}_i = \boldsymbol{\alpha} + \mathbf{B}\mathbf{U}_i + \boldsymbol{\Gamma}\mathbf{Z}_i + \boldsymbol{\zeta}_i, \tag{3}$$

where $\boldsymbol{\Gamma}_{\ell\times p} = \{\gamma_{gj}\}$ and $\mathbf{B}_{\ell\times\ell} = \{\beta_{gh}\}$ (with $\beta_{gg} = 0$ for all g). A restriction on $\mathbf{B}$ is that $(\mathbf{I} - \mathbf{B})$ must be invertible; otherwise, (3) could lead to an inconsistent system of equations. The vector of residuals $\boldsymbol{\zeta}_i$ is independent of $\boldsymbol{\epsilon}_i$, and is normally distributed with $E(\boldsymbol{\zeta}_i) = \mathbf{0}$ and $\operatorname{cov}(\boldsymbol{\zeta}_i) = \boldsymbol{\Psi}$.

We deviate slightly from the notation of Muthén (2002) by breaking the dependent data vector into two pieces ($\mathbf{X}_i$ and $\mathbf{Y}_i$) to represent predictors and outcomes. Our choice of notation emphasizes that the model explicitly accommodates predictors measured with error. However, it may not be straightforward to decide whether a given variable should be considered as an X or as a Y, especially when there are intermediate variables. Once the modeling approach is understood, the choice of notation for the dependent data vector is of secondary importance.

3.1 Parameter Interpretation and Continued Examples

The parameters $\mathbf{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Gamma}$, and $\mathbf{K}$ represent the effects of latent variables and covariates. The matrix $\mathbf{B}$ describes the relationships between latent variables and usually contains the main parameters of interest. An element of $\boldsymbol{\Lambda}$ (e.g., $\lambda_{\ell j}$) is the effect of the latent variables (U_{ij}) on a dependent variable ($X_{i\ell}$ or $Y_{i\ell}$) and is called a factor loading (Bollen 1989). The matrix $\boldsymbol{\Gamma}$ contains the effects of $\mathbf{Z}_i$ on the latent variables; the matrix $\mathbf{K}$ contains covariate effects on particular indicators ($X_{i\ell}$ or $Y_{i\ell}$) that extend the covariate effects mediated through the latent variable [i.e., *item bias* (Beck 1982); also see Fig. 4]. When latent variables are used, the interpretations of $\mathbf{B}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Gamma}$ depend on the scale of the latent variables, which, due to identifiability issues described in Section 3.2, must be defined before the analysis.

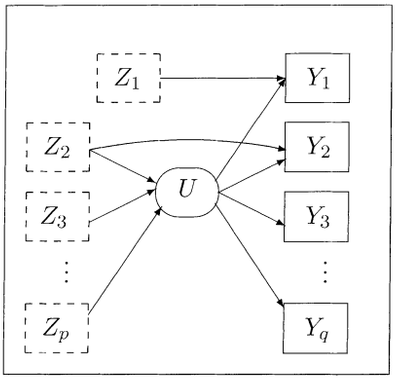

Figure 4. Covariate Effects. Z_1 affects only the particular measurement Y_1 and Z_2 affects the particular measurement Y_2 as well as the latent variable. The effects of $Z_3, \ldots, Z_p$ on the observed outcomes $\mathbf{Y}$ are mediated through the latent variable.

Example With Observed Variables Only. The model equations (1)–(3) for the lead biomarkers example in Figure 2 reduce to

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \mathbf{B}\mathbf{Y}_i + \boldsymbol{\Gamma}\mathbf{Z}_i + \boldsymbol{\zeta}_i. \tag{4}$$

The necessary restrictions to view this model as a special case of the general formulation are to omit $\mathbf{X}_i$, and to set $\boldsymbol{\nu} = \mathbf{0}$, $\boldsymbol{\Lambda} = \mathbf{I}$, $\boldsymbol{\Sigma} = \mathbf{0}$, and $\mathbf{K} = \mathbf{0}$. These restrictions imply that $\mathbf{Y}_i = \mathbf{U}_i$, which means that the dependent variables ($\mathbf{Y}_i$) are measured without error (i.e., $\boldsymbol{\epsilon}_i = \mathbf{0}$ for all i). The coefficient matrices in (4) are

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 \\ 0 & 0 & \beta_{43} & 0 \end{pmatrix}$$

and

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & 0 & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{34} & \gamma_{35} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Table 1 contains coefficients that compare the contribution from various lead measurements on whole blood lead and on cord blood lead, in both natural (columns 1 and 4) and standardized (columns 3 and 6) scales. The effect of recent ceramic use on cord blood lead is $\gamma_{34}\beta_{43}$; that is, the effect is decomposed into the effect of ceramic use on whole blood lead γ_{34} and the effect of whole blood lead on cord blood lead β_{43}. *Standardized effect* estimates (Table 1, columns labeled "Std") are interpreted as changes in standard deviation units in the outcome due to a one standard deviation increase in the predictor. Comparing effects with standardized coefficients is particularly useful when the scales of the predictor variables are disparate, for example, the scale of bone versus blood lead. Although standardized coefficients are not particular to SEMs, standardized coefficients are very popular in SEMs because of the use of *standardized latent variables*.

Although the parameters $\mathbf{B}$ and $\boldsymbol{\Gamma}$ can be estimated by separately fitting each of the multivariate regressions in Section 2.1, their estimation is more efficient through SEMs. Furthermore, because parameter covariances are a byproduct of the SEM estimation procedure, they can be used to estimate (δ-method)

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

Table 1. Relationships Between Lead Biomarkers

	Effect on whole blood lead			Effect on cord blood lead		
Lead measurement	*Coefficient*	*SE*	*Std**	*Coefficient*	*SE*	*Std**
Patella bone lead	$\widehat{\beta}_{31} = .0050$	.0012	.17	$\widehat{\beta}_{31}\widehat{\beta}_{43} = .0044$	.0011	.14
Tibia bone lead	$\widehat{\beta}_{32} = .0026$	.0019	.08	$\widehat{\beta}_{32}\widehat{\beta}_{43} = .0022$	.0017	.06
Recent ceramic use	$\widehat{\gamma}_{34} = .048$	.010	.20	$\widehat{\gamma}_{34}\widehat{\beta}_{43} = .042$	.0089	.16
Air lead	$\widehat{\gamma}_{35} = 1.09$	1.04	.05	$\widehat{\gamma}_{35}\widehat{\beta}_{43} = .95$	.91	.04
Whole blood lead				$\widehat{\beta}_{43} = .88$	.027	.82

*Standardized effect, that is, expressed as change in standard deviation units per one standard deviation increase in predictor.

variances of products (e.g., $\gamma_{34}\beta_{43}$). Some software packages calculate products of parameters and their variances on request.

Example With Latent Variables. The model proposed by Budtz-Jørgensen et al. (2002) (Fig. 3) defines whale meat consumption (X_1) as a variable that depends on covariates and as a predictor of latent methylmercury exposure (U_1). However, note that in (1) and (3), the latent variables ($\mathbf{U}_i$) can be influenced only by the latent variables and fixed covariates ($\mathbf{Z}_i$), not by the dependent variables ($\mathbf{X}_i$ or $\mathbf{Y}_i$). To formally allow for an effect of whale meat consumption on latent methylmercury exposure, it is necessary to introduce a new latent variable, U_4, which equals whale meat consumption (i.e., $X_1 = U_4$) and is assumed to affect latent exposure (i.e., $U_1 = \alpha_1 + \beta_{14}U_4 + \boldsymbol{\gamma}_1^T\mathbf{Z} + \zeta_1$).

The equations (Sec. 2.2) describing this example are not identifiable. Thus, in addition to predefining the scale of the latent variables, their location must also be defined before the analysis. A widely used convention is to set the location of the latent variables to 0, $\boldsymbol{\alpha} = \mathbf{0}$ (i.e., the latent variable is centered). The scale of the latent variables can be measured in standardized units, or it can be set to the scale of an observed variable by fixing the appropriate parameter in $\boldsymbol{\Lambda}$ to unity. For example, by setting $\lambda_{21} = 1$ (from $X_2 = \nu_{x_2} + \lambda_{21}U_1 + \epsilon_{x_2}$), a unit change in the latent mercury exposure corresponds to a unit change in ($\log_{10}$-transformed) cord blood mercury. We choose the second approach to scale the latent variables in this example. Given the identifiability restrictions, the parameter matrices to be estimated include

$$\boldsymbol{\Lambda}^T = \begin{pmatrix} 0 & 1 & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda_{12} & 1 & \lambda_{32} & \lambda_{42} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{53} & 1 & \lambda_{73} & \cdots & \lambda_{11,3} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & \beta_{14} \\ \beta_{21} & 0 & 0 & 0 \\ \beta_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{11} & 0 & 0 & 0 \\ \cdot & \psi_{22} & \psi_{23} & 0 \\ \cdot & \cdot & \psi_{33} & 0 \\ \cdot & \cdot & \cdot & \psi_{44} \end{pmatrix}.$$

The pattern of 0's in $\boldsymbol{\Lambda}$ arises because each dependent variable is assumed to be an indicator of only one latent variable. Assuming that indicators are reflective of only one latent variable is typical in SEMs, although it is not required. The matrix $\boldsymbol{\Psi}$ contains the covariances between the errors in the second stage of the model; its nonzero off-diagonal element, ψ_{23}, represents residual correlation between motor and verbal functions.

Given the parameter restrictions used to scale the latent variables (i.e., $\lambda_{21} = 1$, $\lambda_{22} = 1$, and $\lambda_{63} = 1$), we can conclude that the effect of a 10-fold increase in the (true, unobserved) cord blood mercury concentration leads to a 1.02-point (SE = .49) reduction in FT1, holding all covariates constant. The same increase in cord blood mercury concentration implies an average decrease of 1.63 points (SE = .52) points in BNT2 (Budtz-Jørgensen et al. 2002). Hence effect estimates are interpreted similar to multiple regression coefficients. Although the interpretation of the parameters in $\mathbf{B}$ depends on which factor loadings are set to unity, the statistical significance of the likelihood ratio test for the exposure–outcome associations does not depend on the choice of reference indicators.

3.2 Identifiability

SEMs are particularly susceptible to identifiability problems. According to classical definitions, given a parameter space $\boldsymbol{\Theta}$, model parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which index a family of distributions $F_{\boldsymbol{\theta}}(\cdot)$, are identifiable if for any two parameter values $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, the distributions $F_{\boldsymbol{\theta}_1}(\cdot) = F_{\boldsymbol{\theta}_2}(\cdot)$ if and only if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ (Lehmann and Casella 1998). In other words, the model is identifiable when the mapping $\boldsymbol{\theta} \to F_{\boldsymbol{\theta}}$ is one to one.

In the case when $F_{\boldsymbol{\theta}}(\cdot)$ is the multivariate normal distribution, which is completely identified by its mean $\boldsymbol{\mu}(\boldsymbol{\theta})$ and variance $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the model is not identifiable when there exists $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ such that $\boldsymbol{\mu}(\boldsymbol{\theta}_1) = \boldsymbol{\mu}(\boldsymbol{\theta}_2)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$. In the multivariate normal setting, a proof of identifiability proceeds by setting the nonredundant model moments equal to the corresponding population moments and solves for the model parameters. The location and scale of the latent variables are not identifiable, however, and parameter restrictions need to be made, as has been discussed previously.

Here we show an example of proof of identifiability. Consider a simplified version of the relationships between surrogate predictors $\mathbf{X}_i = (X_{1i}, X_{2i}, X_{3i})^T$ and latent exposure in the methylmercury example,

$$X_{1i} = \nu_{x_1} + \epsilon_{x_{1i}},$$
$$X_{2i} = \nu_{x_2} + \lambda_{21}U_{1i} + \epsilon_{x_{2i}},$$
$$X_{3i} = \nu_{x_3} + \lambda_{31}U_{1i} + \epsilon_{x_{3i}},$$

and

$$U_{1i} = \alpha_1 + \beta_{41}X_{1i} + \zeta_{1i}.$$

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

The mean and covariance implied by the model are

$$E_{\boldsymbol{\theta}}(\mathbf{X}_i) = \begin{pmatrix} \nu_{x_1} \\ \alpha_1 + \nu_{x_2} + \lambda_{21}\beta_{14}\nu_{x_1} \\ \alpha_1 + \nu_{x_3} + \lambda_{31}\beta_{14}\nu_{x_1} \end{pmatrix}$$

and

$$\operatorname{var}_{\boldsymbol{\theta}}(\mathbf{X}_i) = \begin{pmatrix} \sigma_1^2 & \lambda_{21}\beta_{14}\sigma_1^2 & \lambda_{31}\beta_{14}\sigma_1^2 \\ \cdot & \lambda_{21}^2(\phi^2 + \beta_{14}^2\sigma_1^2) + \sigma_2^2 & \lambda_{31}\beta_{14}(\phi^2 + \beta_{14}^2\sigma_1^2) \\ \cdot & \cdot & \lambda_{31}^2(\phi^2 + \beta_{14}^2\sigma_1^2) + \sigma_3^2 \end{pmatrix},$$

which are indexed by 11 parameters, namely $\boldsymbol{\theta} = (\nu_{x_1}, \nu_{x_2}, \nu_{x_3}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \beta_{14}, \lambda_{21}, \lambda_{31}, \alpha, \phi^2)$. Let $E_{\text{pop}}(\mathbf{X}) = (\mu_1, \mu_2, \mu_3)^T$ and $\operatorname{var}_{\text{pop}}(\mathbf{X}) = \{v_{ij}\}$ represent the data's first and second population moments. Setting the nonredundant model moments equal to the corresponding population moments (e.g., $\lambda_{21}\beta_{14}\sigma_1^2 = v_{12}$), we have 9 equations and 11 unknowns. Clearly, the model is not identifiable without additional restrictions. We restrict the mean and scale of the latent variable by setting $\alpha_1 = 0$ and $\lambda_{31} = 1$. It is now straightforward to find the unique solution for the remaining model parameters in terms of the population's first- and second-order moments, and hence establish model identifiability, for example, $\nu_{x_1} = \mu_1$, $\sigma_1^2 = v_{11}$, and $\beta_{14} = \frac{v_{13}}{v_{11}}$.

In this simple example, the number of free parameters is equal to the number of equations, and the model is identical to the *unrestricted* model, which includes all possible means and covariances. A model is said to be *overidentified* if it is identified and defines a structure for the implied mean and covariance. For overidentified models, the number of free parameters is less than the number of equations, and these equations may not have a solution. For example, if we add one more surrogate predictor X_4 to the simple example (e.g., $X_{4i} = \nu_{x_4} + \lambda_{41}U_{1i} + \epsilon_{x_{4i}}$), then λ_{41} must equal *both* $\frac{v_{34}}{A}$ and $\frac{v_{24}v_{13}}{Av_{12}}$, where $A = v_{13}^2 + \frac{v_{23}v_{11}}{v_{13}} - \frac{v_{13}^2}{v_{11}}$. Thus covariance matrices in overidentified models must obey some constraints. In Section 4 we describe how the fit of overidentified models can be assessed based on the difference between the observed and the theoretical means and covariances.

When models involve many variables, algebraic proofs of identifiability can be difficult. To ease computations, the proof can be broken into two steps, showing that the first and second stages of the model (1) and (3) are identified separately. For instance, a two-step approach can be taken to show that the model parameters for the methylmercury example are identifiable. Further, conditions to show that each stage of the model is identifiable have been developed that may not always require intense algebraic calculations. For example, a sufficient condition to show that the measurement part is identifiable is having at least three indicators per latent variable, and a sufficient condition to show that the second stage of the model is identified is to show that the matrix $\mathbf{B}$ can be written as a triangular matrix (as in the lead example) and to assume that the regression errors $\boldsymbol{\zeta}_i$ are uncorrelated. There are many other sufficient conditions ensuring identifiability of each model stage (Bollen 1989).

Confusion about identifiability of SEMs may arise because the formulation of the first stage of the model (1) draws on factor analysis. With respect to identifiability, a main issue in factor analysis is the so-called *parameter rotation*, which refers to, for example, rotations of $\boldsymbol{\Lambda}$ that give rise to the same value of the covariance matrix for the dependent data (Jolliffe 1998). However, there are two types of factor analysis: exploratory and confirmatory. In exploratory factor analysis (EFA), the matrix $\boldsymbol{\Lambda}$ is unconstrained, and the number of latent variables is sometimes also estimated. Solutions for EFA are not unique, because any orthonormal rotation of the matrix $\boldsymbol{\Lambda}$ will give rise to the same covariance matrix implied by the model (Long 1987; Jolliffe 1998). In contrast, confirmatory factor analysis (CFA) draws from substantive knowledge to predefine the number of latent variables and fix some of the parameters in $\boldsymbol{\Lambda}$ to 0. The zeroed elements of $\boldsymbol{\Lambda}$ induce a pattern such that finding a rotation that preserves the pattern of 0's is difficult, therefore minimizing identifiability issues. Because the first stage of SEMs borrows from CFA, SEMs do not necessarily suffer from the identifiability problems that burden EFA.

3.3 Estimation and Inference

Here we discuss likelihood methods in detail and clarify some misconceptions regarding distributional assumptions on predictor variables. The likelihood function of the dependent data $\mathbf{X}_i$ and $\mathbf{Y}_i$, conditional on covariates $\mathbf{Z}_i$, is obtained by integrating the distribution of the latent variables,

$$\begin{aligned} L_{\text{cond}}(\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{Z}_i, \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \int f(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{Z}_i, \mathbf{U}_i, \boldsymbol{\theta}) f(\mathbf{U}_i | \mathbf{Z}_i, \boldsymbol{\theta})\, d\mathbf{U}_i, \end{aligned} \tag{5}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\nu}, \boldsymbol{\Lambda}, \mathbf{B}, \boldsymbol{\alpha}, \boldsymbol{\Gamma}, \mathbf{K}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\}$. The marginal moments of the observed dependent variables conditional on the covariates are

$$E\begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} \Big| \mathbf{Z}_i \Big) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + [\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} + \mathbf{K}]\mathbf{Z}_i \tag{6}$$

and

$$\operatorname{var}\begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} \Big| \mathbf{Z}_i \Big) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}[(\mathbf{I} - \mathbf{B})^{-1}]^T\boldsymbol{\Lambda}^T + \boldsymbol{\Sigma}. \tag{7}$$

Hence, under the assumption of multivariate normality of $\mathbf{X}_i$ and $\mathbf{Y}_i$ conditional on covariates $\mathbf{Z}_i$, the likelihood function (5) has a closed form, which can be written as

$$\begin{aligned} L_{\text{cond}}(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} & |\boldsymbol{\Omega}(\boldsymbol{\theta})|^{-1/2} \\ & \times \exp\left\{ \frac{1}{2}\left[(\mathbf{x}_i^T, \mathbf{y}_i^T)^T - \boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{z}_i\right]^T \right. \\ & \left. \times [\boldsymbol{\Omega}(\boldsymbol{\theta})]^{-1}\left[(\mathbf{x}_i^T, \mathbf{y}_i^T)^T - \boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}(\boldsymbol{\theta})\mathbf{z}_i\right] \right\}, \end{aligned} \tag{8}$$

where

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha},$$

$$\boldsymbol{\Upsilon}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} + \mathbf{K},$$

and

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}[(\mathbf{I} - \mathbf{B})^{-1}]^T\boldsymbol{\Lambda}^T + \boldsymbol{\Sigma}.$$

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

Note that the model imposes the variance structure in (7), but is not allowed to depend on the subject index i.

Estimation techniques derived in the social sciences have exploited the lack of dependence of the covariance structure on the subject index i (Browne and Arminger 1995; Bollen 1989). Further, by assuming that the covariates have expected value $E(\mathbf{Z}_i) = \boldsymbol{\mu}_z$ and covariance $\text{var}(\mathbf{Z}_i) = \boldsymbol{\Omega}_Z$, fitting procedures can be derived from the joint distribution of the observed data $\mathbf{V}_i = (\mathbf{X}_i^T, \mathbf{Y}_i^T, \mathbf{Z}_i^T)^T$. With the additional assumptions on the covariates, both moments of the observed data $\mathbf{V}_i$ are independent of subject index,

$$\boldsymbol{\mu}^*(\boldsymbol{\theta}) \equiv E(\mathbf{V}_i) = \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\mu}_z \\ \boldsymbol{\mu}_z \end{pmatrix}$$

and

$$\boldsymbol{\Omega}^*(\boldsymbol{\theta}) \equiv \text{var}(\mathbf{V}_i) = \begin{pmatrix} \boldsymbol{\Omega}(\boldsymbol{\theta}) + \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\Omega}_Z\boldsymbol{\Upsilon}(\boldsymbol{\theta})^T & \boldsymbol{\Upsilon}(\boldsymbol{\theta})\boldsymbol{\Omega}_Z \\ \boldsymbol{\Omega}_Z \end{pmatrix}.$$

Hence it is natural to estimate the model parameters by minimizing a distance function between the observed moments and the moments implied by the model, for example, the negative of the likelihood. Assuming multivariate normality of $\mathbf{V}_i$, such a likelihood is

$$L_{\text{full}}(\boldsymbol{\theta}) \propto \exp\Big\{\big(\boldsymbol{\mu}^*(\boldsymbol{\theta}) - \bar{\mathbf{v}}\big)^T \boldsymbol{\Omega}^*(\boldsymbol{\theta})^{-1}\big(\boldsymbol{\mu}^*(\boldsymbol{\theta}) - \bar{\mathbf{v}}\big) + \log|\boldsymbol{\Omega}^*(\boldsymbol{\theta})| + \text{tr}\big(\mathbf{S}\boldsymbol{\Omega}^*(\boldsymbol{\theta})^{-1}\big) - \log|\mathbf{S}|\Big\}, \quad (9)$$

where $\bar{\mathbf{v}} = (\bar{\mathbf{y}}^T, \bar{\mathbf{x}}^T, \bar{\mathbf{z}}^T)^T$ and $\mathbf{S}$ is the observed covariance matrix for $\mathbf{V}_i$. Versions of (9) are typically found in the social science literature. Standard numerical maximization techniques (e.g., Newton–Raphson, Fisher scoring) can be used to find maximum likelihood estimates.

Normality assumptions on the $\mathbf{Z}_i$ are often overemphasized; however, such assumptions are not necessary for making inference on $\boldsymbol{\theta}$ (Jöreskog and Goldberger 1975). Because the parameters $\boldsymbol{\mu}_Z$ and $\boldsymbol{\Omega}_Z$ are variation-independent of $\boldsymbol{\theta}$ and (9) can be factored into the product of the conditional likelihood (8) and the marginal distribution of $\mathbf{Z}_i$, by the ancillary principle, inferences on $\boldsymbol{\theta}$ based on the likelihood (9) are asymptotically equivalent to inference from the conditional likelihood (8). Standard testing for the significance of individual parameters can be done via likelihood ratio, score, and Wald tests.

4. ROBUSTNESS, MISSPECIFICATION, AND MODEL EVALUATION

Robust estimation approaches have been proposed that relax the assumption of multivariate normality, for example, various forms of least squares and pseudolikelihood. Least squares approaches minimize a weighted difference between the model mean and variance and the observed data moments. A general formula for the least squares objective function is

$$F_{\mathbf{W}}(\boldsymbol{\theta}) = [\mathbf{u} - \boldsymbol{\kappa}(\boldsymbol{\theta})]^T \mathbf{W}^{-1} [\mathbf{u} - \boldsymbol{\kappa}(\boldsymbol{\theta})], \quad (10)$$

where $\mathbf{u} = \text{vec}(\bar{\mathbf{v}}, \mathbf{S})$, $\boldsymbol{\kappa}(\boldsymbol{\theta}) = \text{vec}(\boldsymbol{\mu}^*(\boldsymbol{\theta}), \boldsymbol{\Omega}^*(\boldsymbol{\theta}))$, and $\mathbf{W}$ is a weight matrix. In the case of ordinary least squares, $\mathbf{W}$ is an identity matrix; in generalized least squares (GLS), $\mathbf{W}$ is an estimate of the variance of $\mathbf{u}$ derived under multivariate normality of the observed data (for a review, see Browne and Arminger 1995); and in the *asymptotically distribution-free* (ADF) procedure, $\mathbf{W}$ is the empirical variance of $\mathbf{u}$ (Browne 1984). Least squares estimates are consistent when the mean $\boldsymbol{\mu}^*(\boldsymbol{\theta})$ and variance $\boldsymbol{\Omega}^*(\boldsymbol{\theta})$ are correctly specified. However, standard errors for $\widehat{\boldsymbol{\theta}}$ are correct only when $\mathbf{W}$ converges to the variance of $\mathbf{u}$ (Browne 1984). Therefore, when multivariate normality does not hold, standard errors derived from GLS are not correct (Browne 1987), whereas (asymptotically) correct inferences can be drawn from ADF procedures because, under weak conditions, the empirical variance converges to the true variance of $\mathbf{u}$. The ADF estimator is asymptotically efficient among the class of estimators defined by (10) (Browne 1984). However, ADF is not widely used in practice, especially in small samples, because the weight matrix depends on highly variable fourth-order sample moments. Satorra and Bentler (1994) modified the ADF estimator to improve its performance in small samples. The pseudolikelihood approach uses the score equations for $\boldsymbol{\theta}$ derived under multivariate normality assumptions but computes robust, sandwich-type variance estimates (Arminger and Schoenberg 1989).

Because the model mean and covariance are highly structured, misspecification can easily occur. Given that the mean and covariance have common parameters, bias in the mean parameters can result if the covariance is misspecified. For instance, Sammel and Ryan (2002) studied the effects of covariance misspecification in a special case of SEMs and concluded that parameter estimates can be seriously biased under a correctly specified mean but incorrectly specified covariance structure, and that power can be greatly reduced. (For a review of other robustness studies, see Hoogland and Boomsma 1998.)

Methods for evaluating overall model fit for SEMs examine the differences between the proposed model and the unrestricted model. Assuming multivariate normality, a likelihood ratio test statistic can be used to assess significant lack of fit and asymptotically follows a chi-squared distribution. Analogous tests can be derived from the weighted least squares (WLS) fitting function (10), because under the model $(n-1)F_{\mathbf{W}}(\widehat{\boldsymbol{\theta}})$ follows a chi-squared distribution (Browne 1984). Model tests that are robust to multivariate normality but only require correctly specified mean and variance have been developed by appropriately choosing the weight matrix (Browne 1984; Sattora 1992). Numerous other measures of model fit have been proposed that quantify the degree of variance in the data accounted for by the model, or quantify the improvement of the current model over a "baseline" model; examples include the comparative fit index and the root mean squared residual, among many others (Hu and Bentler 1995; Jöreskog and Sörbom 1989). Because these indices have unknown distributions, guidelines for determining lack of fit are ad hoc, or at best have been developed through simulations. In essence, goodness-of-fit statistics measure empirical fit; however, it is possible to have different theoretical models that render very similar measures of model fit (MacCallum et al. 1993).

Measures of model fit are affected by sample size, the fitting procedure, and model misspecification. For example, at small sample sizes, fit indices are more likely to indicate a poor fit, whereas the chi-squared test has low power to reject incorrect models. At larger sample sizes, fit indices tend to be overoptimistic, whereas the chi-squared test will reject more frequently than expected, even when discrepancies between the model and the observed data are small (Hu and Bentler 1995). Olsson,

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

Foss, Troye, and Howell (2000) investigated the sensitivity of model fit measures to violations of multivariate normality and incorrect model specification (e.g., omitting variables) using WLS and maximum likelihood fitting procedures; they concluded that GLS and ADF result in overoptimistic fit indexes compared with maximum likelihood.

If a model fits poorly, then researchers should use substantive knowledge as a first step in to modifying the model, because data-driven modifications may have no substantive interpretation. As a secondary step, cautious use of available statistical tools, such as *modification indices* (MIs), can help improve model fit. MIs are essentially univariate score tests,

$$MI(\theta^*) = \left[\frac{\partial \log L(\boldsymbol{\theta}, \theta^*)}{\partial \theta^*} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \theta^*=0} \right]^2 \sim \chi^2_1, \qquad (11)$$

where $L(\boldsymbol{\theta}, \theta^*)$ is the likelihood, (5) or (9), augmented by one extra parameter θ^* that was originally fixed (Saris, Satorra, and Sörbom 1987). MIs can suggest which restricted parameters should be freed and can help identify where (including mean and variance parameters) the model is misspecified. However, repeatedly using available indices to respecify the model leads to multiple testing issues and increased Type II errors, that is, increased probability of retaining a model that is incorrect (Kaplan 1995).

To the extent that they affect the sample mean and variance, outliers can also influence model fit and parameter estimates. Standard techniques (e.g., graphical displays, Mahalanobis distance) can be used to identify outliers. Examining changes in model parameters and fit statistics when taking out one observation at a time and refitting the model can identify influential observations. Although the latter approach seems computationally expensive, refitting the model multiple times can be done relatively rapidly when using software specialized for SEMs (see Sec. 7).

5. MODEL EXTENSIONS

In Section 3 we described SEMs where all dependent variables are assumed to be continuous and relationships between latent variables are linear. In this section we describe two ways of extending linear SEMs. The topics discussed here draw from a large body of literature and are by no means exhaustive.

5.1 Noncontinuous Dependent Data

Methods to include categorical dependent variables ($\mathbf{X}_i$ and $\mathbf{Y}_i$) have been developed in both the social science literature and the traditional statistics literature (Muthén 1984; Reboussin and Liang 1998; Lee and Shi 2001). Muthén (1984) extended the first stage of the model (1) using so-called *threshold models*. Threshold models are based on the idea that categorical variables correspond to a discrete version of an underlying continuous variable. Take, for example, Y_{ij}, the jth component of $\mathbf{Y}_i$, and suppose that it is ordinal (e.g., three categories). Then a threshold model for Y_{ij} is

$$Y_{ij} = \begin{cases} 0 & \text{if } Y^*_{ij} \le \tau_1 \\ 1 & \text{if } \tau_1 < Y^*_{ij} \le \tau_2 \\ 2 & \text{if } Y^*_{ij} > \tau_2, \end{cases} \qquad (12)$$

where Y^*_{ij} is an underlying continuous variable and $\boldsymbol{\tau} = (\tau_1, \tau_2)$ are cutpoints in the distribution of Y^*_{ij} that define the categories of Y_{ij}. Threshold models can be defined similarly for other types of nonnormal data, for example, censored data (Browne and Arminger 1995). When threshold models are used, Y^*_{ij} replaces Y_{ij} in (1). Although including threshold models adds to the flexibility of the modeling framework, it complicates maximum likelihood fitting procedures, which need to integrate over the distribution of Y^*_{ij} to obtain the likelihood. Furthermore, limits of integration may depend on the thresholds, which in many cases also must be estimated. In this setting, Muthén (1984) presented a three-stage WLS fitting procedure for estimating model parameters, including thresholds. Reboussin and Liang (1998) provided an alternative estimating equations approach, and Shi and Lee (2000) proposed a Monte Carlo EM algorithm to produce maximum likelihood estimates.

Other authors have used generalized linear models (McCullagh and Nelder 1989) to include categorical responses (Dunson 2000; Sammel, Ryan, and Legler 1997; Bandeen-Roche et al. 1997). For instance, Sammel et al. (1997) allowed observed outcomes ($\mathbf{Y}_i$) from distributions in an exponential family. If, for example, Y_{ij} is binary, then, heuristically, $Y_{ij} = \nu_j + \boldsymbol{\lambda}_j^T \mathbf{U}_i + \mathbf{k}^T \mathbf{Z}_i + \epsilon_i$ in (1) can be replaced by $E(Y_{ij}) = \text{logit}(\nu_j + \boldsymbol{\lambda}_j^T \mathbf{U}_i + \mathbf{k}^T \mathbf{Z}_i)$. Given the added complexity of the model, the estimation approaches are restricted to cases where there is only one latent variable [also known as MIMIC models (Jöreskog and Goldberger 1975)] or where latent variables are affected by covariates but cannot influence each other (i.e., $\mathbf{B} = \mathbf{0}$). Most methods that use generalized linear links draw on conditional independence assumptions of the dependent data [i.e., $f(\mathbf{y}_i|\mathbf{U}_i) = \prod_j f(y_{ij}|\mathbf{U}_i)$] to derive estimating procedures.

5.2 Nonlinear Relationships Between Latent Variables

Wall and Amemiya (2000) developed two-stage estimators for the case of polynomial associations between latent variables. For example, if $\mathbf{U}_i = (U_{1i}, U_{2i}, U_{31})$, then relationships such as $U_{1i} = \alpha_1 + \beta_1 U_{2i} + \beta_2 U_{3i} + \beta_3 U_{2i} U_{3i} + \beta_4 U^2_{2i} + \beta_5 U^2_{3i} + \epsilon_{1i}$ are allowed. Other estimation approaches allow for more general nonlinear associations among latent variables, for example, Bayesian estimation (Arminger and Muthén 1998) and estimation via the EM algorithm (Lee and Zhu 2002).

6. CONNECTIONS TO OTHER LITERATURE

6.1 Measurement Error Models

Because SEMs provide a framework for handling multiple related exposure measurements, they are closely related, at least in philosophical terms, to statistical methods for measurement error. For example, SEMs are similar to the *structural approach* to errors in variables (Carroll 1998) because both assume a distribution for the unobserved variables. Similar to (5), the structural approach to measurement error integrates over the distribution of the unobserved, true predictor U (where X is the observed predictor and $X = U + \epsilon$) to obtain the likelihood function

$$f(Y|X) \propto \int f(Y|U) f(X|U) f(U)\, dU.$$

In settings where there are two error-prone measurements of an exposure and a univariate outcome, the SEM estimate for the

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

exposure effect is the same as the estimate from an instrumental variables approach to measurement error (Fuller 1987). Similar to measurement error methodology, SEMs can model measurement error in more than one predictor by including more than one latent predictor variable. Therefore, SEMs can easily correct for both the bias caused in the coefficients of covariates ($\mathbf{Z}$) due to measurement error in the predictor of interest (U) and the bias in the coefficient of the predictor of interest (U) caused by measurement error in a covariate (Z) (e.g., Budtz-Jørgensen et al. 2002). Furthermore, handling nonnormal outcomes has been important in both the measurement error and structural equation literature (Carroll, Ruppert, and Stefanski 1995; Muthén 1984; Yuan and Chan 2002).

There are important differences, however. The measurement error literature evolved from settings where a subset of the study population has data available on what can be considered a gold standard measurement instrument or where replicate data are available to estimate the measurement error variance, whereas in SEMs gold standard measurements or replicates need not be available. Further, in contrast to SEMs, which rely heavily on parametric methods such as multivariate normality, much of the measurement error literature focuses on development of semiparametric or functional approaches, including regression calibration and estimating equations–based approaches, which do not assume distributions for the latent variables (Carroll 1998). Although the SEM fitting procedures may need further development compared to the measurement error methods, SEMs provide a tool for handling the more complex problem where both the response and predictors are measured with error.

6.2 Linear Mixed-Effects Models

The statistical literature includes numerous examples where latent variables are used to explain correlations between observed data. Perhaps the most common examples are random-effects models for multivariate and repeated-measures data (Laird and Ware 1982), which can be considered special cases of SEMs. For example, consider the case of four repeated measurements of an outcome Y, measured at time points $t = 0, 1, 2, 3$. Suppose that we want to fit a longitudinal data model with a random slope and random coefficient for time, both of which are influenced by covariates $\mathbf{Z}_i$. In the notation of (1)–(3), the longitudinal model can be written as

$$\mathbf{Y}_i = \boldsymbol{\Lambda}\mathbf{U}_i + \boldsymbol{\epsilon}_i \qquad \text{and} \qquad \mathbf{U}_i = \boldsymbol{\Gamma}\mathbf{Z}_i + \boldsymbol{\zeta}_i,$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i4})^T$, and $\mathbf{U}_i = (U_{1i}, U_{2i})^T$ represents the random intercept and random slope for each of the n individuals. In this case, the matrix of coefficients $\boldsymbol{\Lambda}$ is the design matrix for time and is completely fixed a priori as

$$\boldsymbol{\Lambda}^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix},$$

$\mathbf{B} = \mathbf{0}$, and the variance–covariance matrix for the latent variables $\boldsymbol{\Psi}$ is assumed to have a nonzero off-diagonal term to represent the correlation between the random slope and intercept.

Further, SEMs can accommodate the case where multivariate outcomes reflecting a latent variable are measured repeatedly over time, similar to the approach of Roy and Lin (2000). In contrast with the models proposed by Laird and Ware (1982) and Roy and Lin (2000), classical fitting procedures for SEMs require the design matrix for time to be the same for every subject, such that the covariance matrix does not depend on subject index; modern software, however, relaxes this restriction (Muthén and Muthén 1998–2004). SEMs encompass special cases of models for correlated and longitudinal data and allows one to simultaneously correct for measurement error in the predictor variables.

6.3 Latent Class Models

Although there is a large literature on latent class models as a modeling framework on their own (Clogg 1995), we make reference to latent class models because they can be viewed as versions of SEMs with discrete latent variables. In latent class models, a population is divided into classes that are not directly observed. The latent class indicators ($\mathbf{X}_i$ and $\mathbf{Y}_i$) can be continuous or categorical. In the former, latent class models are also known as mixture models; in the latter, they can be viewed as a way to analyze multidimensional contingency tables (Clogg 1995).

Muthén (2002) discussed a modeling framework more general than (1)–(3) that unifies SEMs with latent class models; details on fitting procedures are found in Muthén and Shedden (1999) and Muthén et al. (2002). Other authors have explicitly developed model formulations and fitting procedures for latent class models in settings where latent variables are not allowed to influence each other ($\mathbf{B} = \mathbf{0}$) and where they are allowed to do so. In the first setting, Bandeen-Roche et al. (1997) developed methods for incorporating continuous covariates using a generalized linear link between the latent variable and the covariates; they assumed a fixed number of latent classes and that the dependent data, $\mathbf{Y}_i$, are categorical. Legler and Ryan (1997) allowed the latent variable, U, to have a Poisson distribution; they used a log-linear link between the latent variable and the covariates, and assumed that the observed outcomes, $\mathbf{Y}_i$, are normal conditional on the latent variable. In the setting where latent variables are allowed to influence each other, Dunson (2000) incorporated generalized linear links among the latent variables and in the measurement part of the model.

7. SOFTWARE AND ALTERNATE MODEL FORMULATIONS

Several software packages are available to fit SEMs. General purpose statistical software have specialized routines to fit linear, normal case SEMs, for example, *sem* in R (Fox 2004), and *proc calis* in SAS (Hatcher and Stepanski 1994). Others, such as *gllamm* in STATA (Rabe-Hesketh, Pickles, and Skrondal 2001) fit SEMs where generalized linear links are allowed in stage one of the model. Specialized software includes Mplus (Muthén and Muthén 1998–2004), which estimates parameters for the model specification discussed in this review. Other specialized software packages are based on alternative model formulations. We discuss two of these formulations here. A more detailed software comparison can be obtained from the first author.

7.1 LISREL and Jöreskog's Formulation

Jöreskog's formulation (Jöreskog and Sörbom 1989) for SEMs is equivalent to that presented in Section 3; this formulation is commonly known as the "LISREL model" because it

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

is implemented in a software program by that name (Jöreskog and Sörbom 1989). In the LISREL formulation, stage one of the model is

$$\mathbf{X}_i = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x \mathbf{U}_i^x + \boldsymbol{\epsilon}_i^x, \tag{13}$$

$$\mathbf{Y}_i = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \mathbf{U}_i^y + \boldsymbol{\epsilon}_i^y, \tag{14}$$

and stage two is

$$\mathbf{U}_i^y = \boldsymbol{\alpha}_y + \mathbf{B}_1 \mathbf{U}_i^y + \mathbf{B}_2 \mathbf{U}_i^x + \boldsymbol{\zeta}_i. \tag{15}$$

The latent variables are partitioned into two parts such that the $\mathbf{U}_i^x$'s do not depend on any other variables in the model, whereas the variables $\mathbf{U}_i^y$ can depend on latent variables $\mathbf{U}_i^x$ and $\mathbf{U}_i^y$. Parameter matrices, as well as the random vectors are separated into pieces accordingly. In this formulation, latent variables $\mathbf{U}_i^x$, and random errors $\boldsymbol{\epsilon}_i^x$ and $\boldsymbol{\epsilon}_i^y$ are assumed to be normally distributed. Notice that this formulation does not explicitly allow for fixed covariates $\mathbf{Z}_i$. However, the LISREL formulation can be shown to be equivalent to (1)–(3) by restricting parameters.

7.2 EQS, and the Bentler and Weeks Model

The software program EQS fits the Bentler and Weeks (1980) model, which generalized the LISREL formulation by explicitly allowing fixed covariates, and is more general than (1)–(3) because it explicitly allows observed variables to influence each other directly. In our notation, their model is

$$\begin{pmatrix} \mathbf{U}_i^y \\ \mathbf{Y}_i \\ \mathbf{X}_i \end{pmatrix} = \mathbf{B}^* \begin{pmatrix} \mathbf{U}_i^y \\ \mathbf{Y}_i \\ \mathbf{X}_i \end{pmatrix} + \boldsymbol{\Gamma}^* \begin{pmatrix} \mathbf{U}_i^x \\ \mathbf{W}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} + \mathbf{K}^* \mathbf{Z}_i, \tag{16}$$

where the latent variable vector $\mathbf{U}_i$ is again split into two pieces, and $\mathbf{W}_i$ is an additional vector of independent variables. In their model, Bentler and Weeks distinguish only variables based on whether they are dependent or independent, not on their observed status.

8. ADDITIONAL ANALYSIS ON THE LEAD EXAMPLE

In a follow-up to the Chuang et al. (2001) SEM analysis of the interrelationships between lead biomarkers, Gomaa et al. (2002) used traditional regression methods to analyze the relationships between lead biomarkers (cord blood, mother's whole blood, and tibia and patella bone lead) and neurodevelopment measured by BAYLEY's mental development index at age 24 months (MDI_{24}). In multiple regression analysis including only one biomarker at a time (and covariates such as mother's age and IQ score), Gomaa et al. (2002) found that both cord blood lead and patella lead levels were significant predictors of MDI_{24} (Table 2). Multivariate models that included all lead measurements simultaneously resulted in multicollinearity problems.

Table 2. Covariate-Adjusted Regressions of MDI on Each Lead Biomarker

Model	β	*p value*
1. Patella bone lead	−.16	.03
2. Tibia bone lead	−.10	.30
3. log(cord blood lead)	−4.94	.02
4. log(mother's blood lead)	−3.72	.12

In this section we reframe the analysis of Gomaa et al. (2002), using SEMs to raise a number of cautionary points when conducting and interpreting SEM analyses. Figure 5 presents two SEMs (A and B) that could be used to describe the relationships between the lead biomarkers and MDI scores. Model A assumes that a latent lead exposure affects MDI_{24} scores. The latent exposure is measured by patella and tibia bone lead concentrations, as well as cord blood and mother's whole blood lead. In Model A, the blood measurements are assumed to have correlated errors; the correlation was introduced as suggested by modification indices obtained from a poor-fitting model otherwise equal to Model A. In contrast, Model B posits the existence of two latent variables, one variable representing exposure integrated across gestation measured by patella and tibia bone lead concentrations and the other representing late gestational exposure measured by cord blood lead and mother's whole blood lead.

Standard measures of model fit (Table 3, first column) indicate that Model A has no substantial lack of fit (e.g., $\chi^2_{df=28} = 31.3$; p value $= .30$). Modification indices did not identify any further misspecification, and no influential observations were found (from leave one-out analyses). We question the validity of Model A, however, because Gomaa et al. (2002) argued that blood lead measurements are more indicative of lead exposure within 1 month of delivery, whereas patella and tibia lead levels are indicative of exposure throughout pregnancy. This cumulative exposure, as measured by bone lead concentration, is a better proxy for exposure during the first trimester of gestation when the fetus is more susceptible to neurologic damage (Gomaa et al. 2002). Hence, Gomaa et al. (2002) argued that for fixed levels of late exposure (e.g., cord blood lead), bone lead levels have an independent effect on developmental outcome. Physiologically, there are reasons to believe that the lead biomarkers are not measuring the same latent variable as assumed by Model A. The assumption of only one latent variable is necessary for the correct interpretation of the exposure coefficient from Model A (not shown).

Model B more accurately captures the biological considerations just described. Fit indices for Model B (Table 3, second column) do not suggest significant lack of fit either. However, whereas Model B better describes the physiology of the problem, Model A fits marginally better according to model selection criteria [Akaike information criterion (AIC) and Bayes information criterion (BIC) in Table 3].

From Model B, the effect of integrated gestational exposure is a decrease of .25 point (SE = .19) in MDI_{24} per 1μg/g increase in patella bone lead concentration, and the effect of late exposure is a decrease of 4.69 points (SE = 4.65) in MDI_{24} per unit increase in ($\log_e$) cord blood lead. The analogous multivariate regression coefficients, −.13 (SE = .07) for patella lead and −4.23 (SE = 2.04) for cord blood lead, from a model that adjusts for both exposures simultaneously are attenuated. The relative attenuation is to be expected, because the SEM parameters are corrected for measurement error. Also note that the standard errors have increased, and the associations between exposure and predictor variables are no longer statistically significant.

Increases in the standard errors of multiple regression coefficients also occur when removing biases due to measurement error, that is, the "bias versus variance" trade-off (Carroll 1995).

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

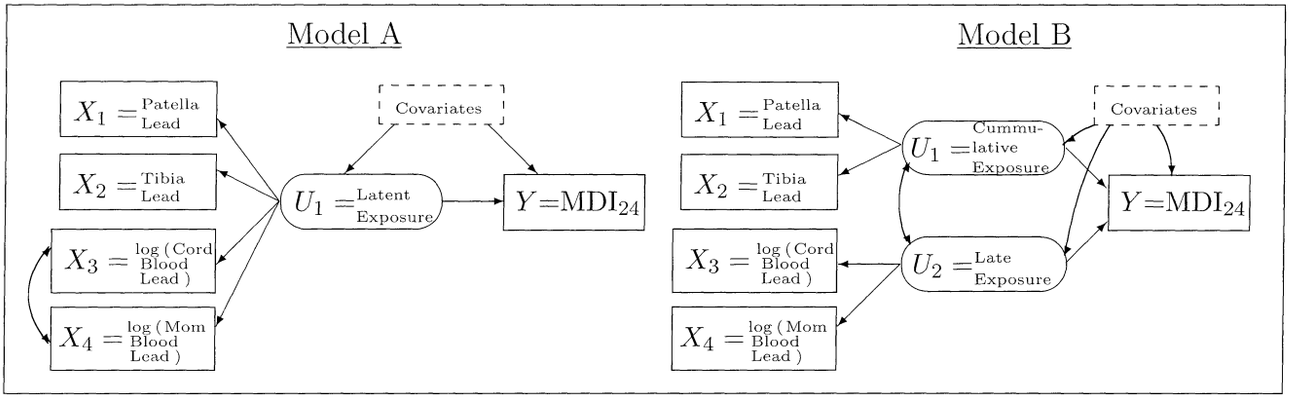

Figure 5. Alternate Models for Lead Data. (a) Model assuming that lead measurements are surrogates for the same underlying latent exposure. (b) Model considers separate effects of cord blood lead and bone lead.

Similarly, in SEMs, power may not be gained by combining information from several indicator variables of a common latent variable. For example, the statistically insignificant association between MDI_{24} and cumulative fetal lead exposure in Model B might be due to small correlations between the bone lead measurements and the latent exposure (.45 for patella and .31 for tibia). Low correlations imply that bone measurements assess integrated gestational exposure poorly (i.e., they have large measurement error). In this example, low correlations also support the idea that although both bone lead measurements assess chronic toxicity, they represent different types of exposure for the fetus due to their different resorption rates (Hu, Rabinowitz, and Smith 1998).

This example illustrates that although SEMs can be a powerful tool in modeling multivariate data, careful consideration of physiological mechanisms, including which observed variables are surrogates of latent variables, is key to successful SEM modeling and interpretation. We saw that while adding parameters to a poor-fitting model may improve its fit (Model A), the resulting model might not be substantively feasible; furthermore, including surrogate variables that poorly measure a given latent variable may compromise study power.

9. CONCLUSIONS

We have reviewed the theory and estimation for linear SEMs, briefly discussed model extensions, and drawn connections between SEMs and other latent variable models. In their most general formulation, SEMs can be viewed as a broad class of models that explain associations between observed variables by using latent variables, although there are examples in which all variables in the model are observed.

Table 3. Model Fit Indices

Index (criterion for "good fit")*	*Model A*	*Model B*
Chi-squared test, *p* value (>.05)	.30	.38
TLI (>.95)	.96	.98
RMSEA (<.05)	.02	.02
SRMR (<.05)	.03	.02
AIC (smaller is better)	10,834.0	10,840.9
BIC (smaller is better)	10,938.9	10,975.3

*From Hu and Bentler (1995).

The SEMs formulation is preferable to traditional regression models for several reasons. SEMs that include latent variables easily allow for measurement error modeling on more than one predictor, including confounders, and can accommodate parsimonious models for multivariate outcomes (e.g., the methylmercury example). Furthermore, SEMs are more general than random-effects models, because regression associations are permitted among latent variables, hence allowing regression associations between latent predictors (e.g., measured with error) and random effects. In settings where latent variables are not used (e.g., interrelationships of lead biomarkers), the SEM formulation is preferable multiple regression because parameter estimation is more efficient, and because decomposing effects of mediating variables (e.g., whole blood lead in the biomarkers example) can be easily done.

A limitation of SEMs is their sensitivity to model misspecification. Misspecification can easily occur because traditional approaches to fitting SEMs are highly parametric and distributional assumptions are imposed at both stages of the model. Although there are procedures that relax distributional assumptions (Browne 1984; Arminger and Schoenberg 1989), correct covariance specification is still required. Furthermore, incorrect covariance specification can lead to bias in the mean parameters (Sammel and Ryan 2002). Hence procedures that only require that the mean be correctly specified are of interest. Although the model presented by Lin et al. (2000) considered only cases where the latent variables do not influence each other ($\mathbf{B} = \mathbf{0}$), it could serve as groundwork for robust procedures to estimate the SEM parameters. Because much information about the regression parameters in $\mathbf{B}$ is contained in the covariance of the observed data, however, developing methods that relax correct covariance specification might be challenging in the general framework.

Given the flexibility of SEMs to adjust for covariates at both stages of the model, covariate adjustment merits careful consideration, probably more so than in multivariate regression. In regression analysis, adjusting for variables that are correlated with both the exposure and the outcome is needed to obtain unbiased (exposure) effect estimates, and power may be gained by also including variables that are correlated only with the outcome. In SEMs, the true (latent) exposure can be considered to depend on covariates. Hence power can also be gained

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

by including covariates that are correlated only with the exposure, because these help in estimating the latent variable. For instance, in the methylmercury example, the number of pilot whale meals was included as a predictor only for the exposure. Including this variable provides additional information about the latent exposure and thus reduces estimation uncertainty for the latent variable.

Because of shared features with other modeling approaches, SEMs might be erroneously confused with them. For example, SEMs use diagrams similar to kinetic models (Becka 2002); however, the arrows in SEMs are interpreted as regression associations, not as rates of change between "compartments." The possibility of including kinetic components in SEMs is an open research question. SEMs also share the name "structural" with the work of Robbins, Hernán, and Brumback (2000); although both modeling frameworks use diagrams and latent variables, the models are distinct. In SEMs the latent variables are never observed, but surrogate information is available for all subjects, whereas in the work of Robbins et al. (2000), the unobservable variables, so-called "counterfactuals," are observed in some of the subjects some of the time.

As with most statistical models, SEMs rely on extensive substantive knowledge to infer causality; that is, SEMs are a confirmatory tool rather than an exploratory tool. For example, SEMs by themselves do not discover directionality of arrows, nor can they tell us exactly how many latent variables there are. Although many statistical tools are available to guide the researcher with model building and model selection, it is ultimately up to the researcher to decide which of, for example, two equally good-fitting models is more substantively sound. A loose analogy of this challenge in SEM model selection can be made to the setting of multivariate regression, where regression parameters can be interpreted as exposure effects only after all relevant confounders have been adjusted for, and it is up to the substantive researcher to decide which variables are confounders. Furthermore, given the available number of fit indices and the predictable (but sometimes contradictory) ways in which fit indices are affected by sample size, fitting procedure, type of misspecification, and other factors, model selection based solely on statistics remains a difficult problem.

Although SEMs have been primarily developed and extensively discussed in the social science literature, they are applicable to other areas of research, including environmental epidemiology. In social science research, carefully designed and reliable and validated instruments are usually used to measure latent variables. This helps ensure that the instruments measure the latent variable reasonably well. Consequently, the power to detect relationships between latent variables increases. Further, in light of the example of Section 8, researchers are encouraged to think a priori (or through pilot studies) about which instruments (e.g., biomarkers) would better measure the latent variables of their interest.

Overall, SEMs are a flexible class of models that control for covariates at different stages of the model, alleviate multiple testing issues, and account for measurement error in the exposure and the outcome simultaneously, hence reducing parameter bias. There is an opportunity to investigate the specification of fixed covariates and to improve model diagnostics and estimation techniques, particularly to relax covariance assumptions.

[Received September 2004. Revised July 2005.]

REFERENCES

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

Arminger, G., and Muthén, B. O. (1998), "A Bayesian Approach to Nonlinear Latent Variable Models Using the Gibbs Sampler and the Metropolis–Hastings Algorithm," *Psychometrika*, 63, 271–300.

Arminger, G., and Schoenberg, R. J. (1989), "Pseudo–Maximum-Likelihood Estimation and a Test for Misspecification in Mean and Covariance Structure Models," *Psychometrika*, 54, 409–425.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., and Rathouz, P. J. (1997), "Latent Variable Regression for Multiple Discrete Outcomes," *Journal of the American Statistical Association*, 92, 1375–1386.

Beck, R. (ed.) (1982), *Handbook of Methods for Detecting Test Bias*, Baltimore: Johns Hopkins University Press.

Becka, M. (2002), "Compartmental Analysis," in *Encyclopedia of Environmetrics*, eds. A. H. El-Shaarawi and W. W. Piegorsch, New York: Wiley, pp. 371–381.

Bentler, P. M. (1992), *EQS Structural Equations Program Manual*, Cork, Ireland: BMDP Statistical Software.

Bentler, P. M., and Weeks, D. G. (1980), "Linear Structural Equations With Latent Variables," *Psychometrika*, 45, 289–308.

Bollen, K. A. (1989), *Structural Equations With Latent Variables*, New York: Wiley.

Browne, M. W. (1984), "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.

——— (1987), "Robustness of Statistical Inference in Factor Analysis and Related Models," *Biometrika*, 74, 375–384.

Browne, M. W., and Arminger, G. (1995), "Specification and Estimation of Mean and Covariance Structures," in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, eds. G. Arminger, C. C. Clogg, and M. E. Sobel, New York: Plenum Press, pp. 185–249.

Budtz-Jørgensen, E., Keiding, N., Grandjean, P., Weihe, P., and White, R. F. (2002), "Estimation of Health Effects of Prenatal Mercury Exposure Using Structural Equation Models," *Environmental Health: A Global Access Science Source*, 1, 2.

——— (2003), "Statistical Methods for the Evaluation of Health Effects of Prenatal Mercury Exposure," *Environmetrics*, 14, 105–120.

Carroll, R. J. (1998), "Measurement Error in Epidemiologic Studies," in *Encyclopedia of Biostatistics*, New York: Wiley, pp. 2491–2519.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, New York: Chapman & Hall.

Chuang, H. Y., Schwartz, J., Gonzales-Cossio, T., Lugo, M. C., Palazuelos, E., Aro, A., Hu, H., and Hernandez-Avila, M. (2001), "Interrelations of Lead Levels in Bone, Venous Blood, and Umbilical Cord Blood With Exogenous Lead Exposure Through Maternal Plasma Lead in Peripartum Women," *Environmental Health Perspectives*, 109, 527–532.

Clogg, C. C. (1995), "Latent Class Models," in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, eds. G. Arminger, C. C. Clogg, and M. E. Sobel, New York: Plenum Press, pp. 311–359.

Cox, D. R., and Wermuth, N. (1996), *Multivariate Dependencies*, New York: Chapman & Hall.

Dunson, D. B. (2000), "Bayesian Latent Variable Models for Clustered Mixed Outcomes," *Journal of the Royal Statistical Society*, Ser. B, 62, 355–366.

Fox, J. (2004), "The sem Package," Technical Report 2004/05, McMaster University, Dept. Sociology, *http://www.socsci.mcmaster.ca/jfox/*.

Fuller, W. (1987), *Measurement Error Models*, New York: Wiley.

Gomaa, A., Hu, H., Bellinger, D., Schwartz, J., Tsaih, S. W., Gonzalez-Cossio, T., Schnaas, L., Peterson, K., Aro, A., and Hernandez-Avila, M. (2002), "Maternal Bone Lead as an Independent Risk Factor for Fetal Neurotoxicity: A Prospective Study," *Pediatrics*, 110, 110–118.

Grandjean, P., Weihe, P., White, R. F., Debes, F., Araki, S., Yokoyama, K., Murata, K., Sorensen, N., Dahl, R., and Jørgensen, P. J. (1997), "Cognitive Deficit in 7-Year-Old Children With Prenatal Exposure to Methylmercury," *Neurotoxicology and Teratology*, 19, 417–428.

Hatcher, L., and Stepanski, E. J. (1994), *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*, Cary, NC: SAS Institute.

Hoogland, J. J., and Boomsma, A. (1998), "Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis," *Sociological Methods and Research*, 26, 329–367.

Hu, L. T., and Bentler, P. M. (1995), "Evaluating Model Fit," in *Structural Equation Modeling*, ed. R. H. Hoyle, London: Sage, pp. 76–99.

Hu, H., Rabinowitz, M., and Smith, D. (1998), "Bone Lead as a Biologic Marker in Epidemiologic Studies of Chronic Toxicity: Conceptual Paradigms," *Environmental Health Perspectives*, 106, 1–8.

Jolliffe, I. T. (1998), "Factor Analysis, Overview," in *Encyclopedia of Biostatistics*, New York: Wiley, pp. 1474–1482.

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms

Jöreskog, K. G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika*, 57, 239–251.

——— (1973), "A General Method for Estimating Linear Structural Equation Systems," in *Structural Equation Models in the Social Sciences*, eds. A. S. Goldergand and O. D. Duncan, New York: Seminar Press, pp. 85–112.

Jöreskog, K. G., and Goldberger, A. S. (1975), "Estimation of a Model With Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 70, 631–639.

Jöreskog, K. G., and Sörbom, D. (1989), *Lisrel 7: A Guide to the Program and Applications*, Chicago: SPSS.

Kaplan, D. (1995), "Statistical Power in Structural Equation Modelling," in *Structural Equation Modeling*, ed. R. H. Hoyle, London: Sage, pp. 100–117.

Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Lee, S. Y., and Shi, J. Q. (2001), "Maximum Likelihood Estimation of Two-Level Latent Variable Models With Mixed Continuous and Polytomous Data," *Biometrics*, 57, 787–794.

Lee, S. Y., and Zhu, H. T. (2002), "Maximum Likelihood Estimation of Nonlinear Structural Equation Models," *Psychometrika*, 67, 189–210.

Legler, J. M., and Ryan, L. M. (1997), "Latent Variable Models for Teratogenesis Using Multiple Binary Outcomes," *Journal of the American Statistical Association*, 92, 13–20.

Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation* (2nd ed.), New York: Springer-Verlag.

Lin, X. H., Ryan, L. M., Sammel, M., Zhang, D. W., Padungtod, C., and Xu, X. P. (2000), "A Scaled Linear Mixed Model for Multiple Outcomes," *Biometrics*, 56, 593–601.

Lohelin, J. C. (1992), *Latent Variable Models: An Introduction to Factor, Path and Structural Analysis*, Hillsdale, NJ: Lawrence Erlbaum.

Long, J. S. (1987), *Confirmatory Factor Analysis*, Beverly Hills, CA: Sage.

MacCallum, R. C., Wegner, D. T. et al. (1993), "The Problem of Equivalent Models in Applications of Covariance Structure Analysis," *Psychological Bulletin*, 114, 185–199.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

Muthén, B. (1984), "A General Structural Equation Model With Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators," *Psychometrika*, 49, 115–132.

——— (2002), "Beyond SEM: General Latent Variable Modeling," *Behaviormetrika*, 29, 81–117.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S. G., Carlin, J. B., and Liao, J. (2002), "General Growth Mixture Modeling for Randomized Preventive Interventions," *Biostatistics*, 3, 459–475.

Muthén, B. O., and Muthén, L. K. (1998–2004), *Mplus User's Guide* (3rd ed.), Los Angeles: Muthén and Muthén.

Muthén, B., and Shedden, K. (1999), "Finite Mixture Modeling With Mixture Outcomes Using the EM Algorithm," *Biometrics*, 55, 463–469.

Olsson, U. H., Foss, T., Troye, S. V., and 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*, 7, 557–595.

Palta, M., and Lin, C. Y. (1999), "Latent Variables, Measurement Error and Methods for Analysing Longitudinal Binary and Ordinal Data," *Statistics in Medicine*, 18, 385–396.

Pearl, J. (1995), "Causal Diagrams for Empirical Research," *Biometrika*, 82, 669–688.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2001), "GLLAMM Manual," technical report, King's College, Institute of Psychiatry, Dept. Biostatistics and Computing.

Reboussin, B. A., and Liang, K. Y. (1998), "An Estimating Equations Approach for the Liscomp Model," *Psychometrika*, 63, 165–182.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000), "Marginal Structural Models and Causal Inference in Epidemiology," *Epidemiology*, 11, 550–560.

Roy, J., and Lin, X. H. (2000), "Latent Variable Models for Longitudinal Data With Multiple Continuous Outcomes," *Biometrics*, 56, 1047–1054.

Sammel, M. D., and Ryan, L. M. (1996), "Latent Variable Models With Fixed Effects," *Biometrics*, 52, 650–663.

——— (2002), "Effects of Covariance Misspecification in a Latent Variable Model for Multiple Outcomes," *Statistica Sinica*, 12, 1207–1222.

Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997), "Latent Variable Models for Mixed Discrete and Continuous Outcomes," *Journal of the Royal Statistical Society*, Ser. B, 59, 667–678.

Saris, W. E., Satorra, A., and Sörbom, D. (1987), "Detection of Specification Errors in Linear Structural Equation Models," in *Sociological Methodology*, ed. C. C. Clogg, Washington, DC: American Sociological Association, pp. 105–130.

Satorra, A. (1992), "Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures," *Sociological Methodology*, 22, 249–278.

Satorra, A., and Bentler, P. M. (1994), "Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis," in *Latent Variable Analysis: Applications to Developmental Research*, eds. C. Clogg and A. von Eye, Newbury Park, CA: Sage, pp. 399–419.

Shapiro, A. (1984), "A Note on the Consistency of Estimators in the Analysis of Moment Structures," *British Journal of Mathematical and Statistical Psychology*, 37, 84–88.

Shi, J. Q., and Lee, S. Y. (2000), "Latent Variable Models With Mixed Continuous and Polytomous Data," *Journal of the Royal Statistical Society*, Ser. B, 62, 77–87.

Wall, M. M., and Amemiya, Y. (2000), "Estimation for Polynomial Structural Equation Models," *Journal of the American Statistical Association*, 95, 929–940.

Wright, S. (1921), "Correlation and Causation," *Journal of Agricultural Research*, 20, 557–585.

Yuan, K. H., and Bentler, P. M. (1997), "Mean and Covariance Structure Analysis: Theoretical and Practical Improvements," *Journal of the American Statistical Association*, 92, 767–774.

——— (2000), "Robust Mean and Covariance Structure Analysis Through Iteratively Reweighted Least Squares," *Psychometrika*, 65, 43–58.

Yuan, K. H., and Chan, W. (2002), "Fitting Structural Equation Models Using Estimating Equations: A Model-Segregation Approach," *British Journal of Mathematical and Statistical Psychology*, 55, 41–62.

This content downloaded from 158.111.236.60 on Wed, 22 Apr 2020 14:09:20 UTC
All use subject to https://about.jstor.org/terms