Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask

Dawn Iacobucci

Department of Marketing, Owen Graduate School of Management, Vanderbilt University, 401 21st Avenue South, Nashville, TN 37202, USA

Available online 9 October 2009

Abstract

This article is intended to serve as a primer for structural equations models for the behavioral researcher. The technique is not mysterious—it is a natural extension of factor analysis and regression. The measurement part of a structural equations model is essentially a confirmatory factor analysis, and the structural part of the model is like a regression but vastly more flexible in the types of theoretical models that may be tested. The models and notation are introduced and the syntax is provided to replicate the analyses in the paper. Part II of this article will appear in the next issue of the *Journal of Consumer Psychology*, and it covers advanced issues, including fit indices, sample size, moderators, longitudinal data, mediation, and so forth.
© 2009 Society for Consumer Psychology. Published by Elsevier Inc. All rights reserved.

Structural equations modeling (SEM) is an important tool for marketing researchers. Our top journals contain numerous articles with SEM analyses of data as varied as B2B relationship marketing studies (e.g., Anderson and Narus, 1990; Selnes and Sallis, 2003; Steenkamp and Baumgartner, 2000) or B2C customer satisfaction surveys (e.g., Luo and Bhattacharya, 2006). Survey data from studies such as these are a natural fit for SEM, but when SEM is considered as a natural extension of regression (to be demonstrated shortly), it should be clear that SEM is a statistical tool, orthogonal to the substantive domain of data on which it is implemented. Thus, SEM may also be applied to secondary data bases (e.g., behavioral CRM or loyalty program data) or data collected in experimental settings (e.g., measures of memory, attitudes, and intentions), etc. For example, it is not unusual for an experimental consumer psychologist to inquire about mediation in an attempt to clarify the theoretical processes by which some phenomenon occurs (e.g., Iacobucci, 2008). If mediation clarifies the conceptual picture somewhat, with the insertion of just one new construct—the mediator—imagine how much richer the theorizing might be if researchers tried to formulate and test even more complex nomological networks.

The *Journal of Consumer Psychology* presents this article to encourage more frequent and knowledgeable use of SEMs. We begin with an overview of the SEM approach: (1) we discuss the motivation underlying SEMs and a perennial cautionary concern regarding the interpretation of causality, (2) we review the source models of factor analysis and path analysis, and (3) we combine these to develop the full SEM model, and describe the entirety of parameters that may be estimated. Software syntax is introduced to enhance the likelihood of readers' trial and adoption.

Motivation for SEMs

Multiple regression can be thought of as a simple SEM: a set of p predictor variables, $X = \{X_1, X_2, ..., X_p\}$ is used to predict a single dependent variable, Y, i.e., $\{X_1, X_2, ..., X_p\} \rightarrow Y$. As tremendously useful as OLS regression is as an analytical tool, there are nevertheless many classes of models that are not estimable via regression. For example, to examine a mediation chain, e.g., $X \rightarrow M \rightarrow Y$, one might fit two sequential regressions, $X \rightarrow M$ and $M \rightarrow Y$ to approximate the chain, but statistical theory dictates that the simultaneous fitting of both paths is both more parsimonious and will yield better results (e.g., more precise estimates, as indicated by smaller standard errors, and less bias, as each effect is estimated while partially out the other effects). An alternative model that would be problematic in regression would be a nonrecursive relation, e.g., of the form $Y_1 \leftarrow \rightarrow Y_2$. These paths, or structural links among theoretical constructs of interest, are not easily accommodated by regression but are fit easily in SEMs.

E-mail address: dawn.iacobucci@owen.vanderbilt.edu.

SEM's second advantage is the ability to deal with measurement error. Regression models do not parse measurement error separately from the error attributable to a model's lack of fit—both are lumped into $1 - R^2$. In SEM, the path analysis part of the model is distinguishable from the factor analysis part of the model, and hence, their error variances are separable also. This distinction is important for two reasons: first, it allows finer diagnostics for model improvement (e.g., whether lack of fit is due to poor measures or model misspecification). In addition, it reduces problems with certain kinds of multicollinearity: e.g., imagine a construct (X_1) measured by multiple indicators (e.g., X_{1a}, X_{1b}, X_{1c}). If these variables are truly all indicators of the same constructs, they would presumably by highly intercorrelated. If that is the case, including all of them, X_{1a}, X_{1b}, X_{1c}, in a regression would create multicollinearity problems even before considering whether any other predictors (X_2, …, X_p) are correlated with X_1. In SEM, the multiple indicators are represented by a factor, the measurement is properly represented, and that problem is averted.

"Causal" models?

To establish evidence for causal relationships, philosophers of science largely converge on the criteria of Hume and Mill, requiring: (1) concomitant variation (i.e., if X causes Y, then X and Y should be correlated), (2) sequential ordering, and (3) elimination of rival explanations. Many social science logicians uphold experimentation—with its mechanism of manipulation and measurement—as the optimal means of establishing causality (cf., Baumrind, 1983; Holland, 1986; Rubin, 2005).

Unfortunately (from the perspective of clear causal attribution), most data analyzed via SEMs are not derived from experiments but are correlational. It is well known that a correlation between X and Y can arise from a variety of models: $X \rightarrow Y$, $Y \rightarrow X$, $X \leftarrow F \rightarrow Y$, to name a few. Yet as is evidenced in published SEM articles, most researchers find irresistible the temptation to interpret SEM results as causal relationships.

This problem is not unique to SEM; the same issue arises in multiple regression. In some applications of regressions, the researchers are hoping to establish causality, whereas in other studies, the researchers simply have the goal of determining whether knowledge of X helps predict Y. The researcher pursuing causality can only do so convincingly in the presence of certain logical operators, such as control over X or elimination of all alternative explanations for Y. Thus, the issue of causality on correlational data is not new, but simply more daunting in SEM, where there are usually more Y's being modeled. Accordingly, while fitting SEMs has been referred to as "causal modeling," many statisticians are conservative regarding the use of the term "causal" (or even "affects" or "impacts"), preferring instead to conclude, "X helps predict Y."

Finally, something that should not surprise the readers of this journal, it is true that establishing causality in SEM is best evaluated in the presence of strong theory. Given the complexities of SEM, experts recommend that users have not only a strong, theoretically supported focal model, but also several nontrivial competing models to fit and test comparatively.

Illustration data

To make the model introductions more concrete, we will work with an example. Table 1 contains a survey of 15 items assessing customers' perceptions of Hewlett-Packard printers. The structure of the survey is such that the first three items are intended to tap quality, the next three measure price perceptions, then three items each tapping customers' perceptions of value, satisfaction, and behavioral intentions. The covariances among the data representing 100 respondents are presented in Table 2 (to facilitate new users attempting replicate the results in this article). We will use these data to illustrate the models.

The measurement model: confirmatory factor analysis

SEM models are comprised of a measurement model, which relates the variables to the constructs, and a structural path model, which relates the constructs to other constructs. We will begin with the first, the measurement model, which is basically

Table 1
Customer quality survey.

	Strongly disagree						Strongly agree
1. The quality of the HP printer I bought is excellent.	1	2	3	4	5	6	7
2. HP printers are known to be highly reliable.	1	2	3	4	5	6	7
3. I'm sure my HP printer will last a long time.	1	2	3	4	5	6	7
4. The HP printer was reasonably priced.	1	2	3	4	5	6	7
5. HP sets fair prices for its products.	1	2	3	4	5	6	7
6. The HP printers are no more expensive than others.	1	2	3	4	5	6	7
7. I feel like I got good value for this purchase.	1	2	3	4	5	6	7
8. The quality of the printer is worth its cost.	1	2	3	4	5	6	7
9. I could tell my boss this purchase was good value.	1	2	3	4	5	6	7
10. I am very satisfied with my newly purchase HP printer.	1	2	3	4	5	6	7
11. My printer is better than I expected it would be.	1	2	3	4	5	6	7
12. I have no regrets about having bought this printer.	1	2	3	4	5	6	7
13. I would buy another HP if I had to buy another printer.	1	2	3	4	5	6	7
14. I would buy other HP products.	1	2	3	4	5	6	7
15. I would tell my friends and coworkers to buy HPs.	1	2	3	4	5	6	7

Table 2
Covariances among quality data ($n = 100$).

q1	q2	q3	c1	c2	c3	v1	v2	v3	cs1	cs2	cs3	r1	r2	r3
2.813														
2.264	1.973													
1.961	1.670	1.528												
-0.018	-0.047	-0.097	3.060											
-0.054	-0.068	-0.118	2.755	2.576										
-0.071	-0.073	-0.103	2.481	2.297	2.142									
1.601	1.282	1.161	-1.436	-1.351	-1.252	2.692								
1.427	1.126	1.025	-1.304	-1.196	-1.127	2.341	2.16							
1.242	0.970	0.899	-1.162	-1.091	-1.020	1.970	1.788	1.616						
1.656	1.415	1.273	-0.672	-0.591	-0.484	1.081	0.976	0.808	2.694					
1.333	1.182	1.061	-0.531	-0.480	-0.373	0.859	0.764	0.636	2.230	1.987				
1.232	1.102	0.980	-0.490	-0.450	-0.363	0.788	0.713	0.606	1.957	1.715	1.583			
0.912	0.867	0.787	-1.144	-1.057	-0.875	0.441	0.394	0.384	1.479	1.317	1.176	2.831		
0.764	0.763	0.692	-1.065	-0.981	-0.822	0.397	0.361	0.343	1.254	1.113	1.032	2.237	1.933	
0.736	0.696	0.636	-0.894	-0.822	-0.690	0.425	0.380	0.354	1.147	1.006	0.925	1.951	1.646	1.530

a confirmatory factor analysis. It is written:

$$x = \Lambda_x \xi + \delta \qquad (1)$$

with terms defined as follows:

- x is a vector of p measured variables
- Λ_x is the $p \times r$ matrix of factor loadings of the x's on the factors, ξ
- ξ is a vector of the r factors
- δ is vector of measurement errors
- Φ is not shown in Eq. (1), but it will contain the covariances among the factors, ξ.

Fig. 1 illustrates the confirmatory factor analysis (CFA) measurement model for these data. The survey items are grouped three per construct. Thus, the first factor, ξ_1, is reflected in the first three variables, x_1, x_2, x_3, etc. It is convention to depict factors (or "constructs" or "latent variables") in ovals and measured variables in boxes. Arrows originate from the factor and point to (or "give rise to" or "are reflected in") the variables. Measurement errors, the δ's, also contribute to the resultant data.

Fig. 2 depicts the corresponding form of the factor loadings matrix, Λ_x, posited by this confirmatory factor model. Unlike exploratory factor analysis where all factor loadings are estimated and examined for their relative size (e.g., whether the loadings exceed a threshold such as 0.30), in confirmatory factor analysis, the researcher specifies which variables are hypothesized to load on which factors, and only those loadings are estimated. Most loadings are hypothesized to be zero, specifying that the column factor is not reflected in that row measured variable, and the model fixes them to zero.
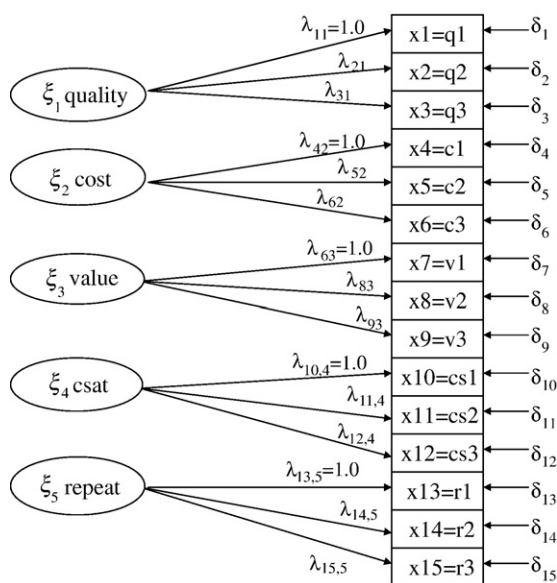

Fig. 1. Confirmatory factor analysis.

Fig. 2. Confirmatory factor loadings (arrow goes from column j to row i).

Table 3
Φ factor intercorrelations from CFA.

	Quality	Price	Value	Satisfaction	Intentions
Quality	1.00				
Price	-0.03	1.00			
Value	0.58	-0.53	1.00		
Satisfaction	0.62	-0.22	0.39	1.00	
Intentions	0.39	-0.43	0.19	0.59	1.00

Note that one loading per factor is fixed to one, to set the scale for each factor and assist the overall model estimation. The CFA model allows for inferential testing of two sorts: first, the significance of each of the factor loadings, and second, the overall fit of the model.

Sometimes a nonsignificant loading is indicative of a variable that would cross-load on another factor (if the CFA had allowed additional estimations), and sometimes a nonsignificant loading simply means the variable is a poor measure (e.g., a confusing or ambiguous item). In either case, the variable is a candidate to be dropped from further analyses. It is good form to verify that the measurement model fits reasonably well before proceeding to examine the path structure model (cf., Anderson and Gerbing, 1988; Anderson, Gerbing and Hunter, 1987; Gerbing and Anderson 1988; Browne et al., 2002).

For our data, the model fits nicely[1]: $X^2_{80} = 96.53$ ($p = 0.1$), CFI = 0.99, SRMR = 0.019 (the indices and criteria are explained later), and all the factor loadings were significant (ranging from 0.97 to 1.00). The factor intercorrelations in Table 3 indicate positive relationships between quality and value, satisfaction and intentions, and negative relationships between price, value and intentions.

The structural model: path analysis

We have seen that the factor analytical model of SEM relates the variables to the constructs. The other part of SEM is the structural model, or "path model," which relates the constructs to other constructs. This model is usually of greatest theoretical interest.

In a path model, there is a distinction between exogenous and endogenous variables. "Exogenous" variables, x's, and constructs, ξ's, are what the readers of this journal would commonly call the "independent variables." In a path diagram, no arrows point to these factors; i.e., no other factors in the model are thought to give rise to an exogenous construct. In contrast, "endogenous" variables and constructs are the "dependent variables," y's, and constructs, η's, which are predicted by other constructs in the model.
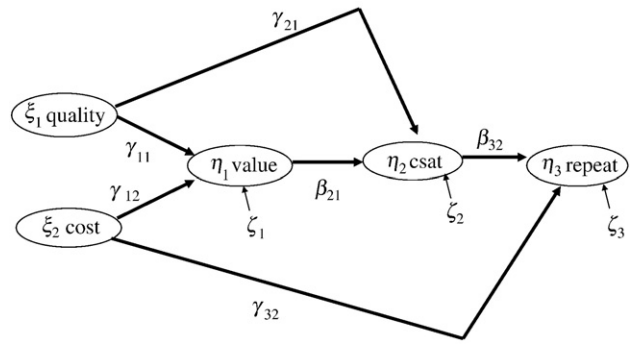

Fig. 3. Path model.

The path model of structural coefficients is as follows:

$$\eta = B\eta + \Gamma\xi + \zeta, \tag{2}$$

with terms defined as follows:

- η is a vector of endogenous ("dependent") factors
- B is a matrix of coefficients of the η's on other η's (part of the structural relationships)
- Γ is a matrix of coefficients of the ξ's on the η's (also part of the structural relationships)
- ξ is a vector of the independent latent variables, exogenous constructs (i.e., predictor factors)
- ζ is a vector of equation errors (random disturbances) trying to predict the endogenous constructs η (predictions inaccuracies).

Note that the model contains structural path parameters (γ and β), and structural prediction errors (ζ), but it has no depiction of measured variables or factor loadings. The constructs (ξ's and η's) are thought to be perfectly mirrored in their measures (x's and y's), so the factor loadings matrices (Λ_y and Λ_x) are identity matrices to quite literally depict the one-to-one mapping between construct and variable, and the matrices of measurement errors (for ε and δ) are zero. This special case will become clearer in contrast when we combine the measurement model with the path model in the complete SEM model, which we do shortly.

A path model for the HP survey data is represented in Fig. 3. This model posits that quality and cost together contribute to perceptions of value. Quality also contributes to satisfaction, and cost affects whether there is an intention to repurchase. Value contributes to satisfaction, which in turn contributes to intentions.[2] Fig. 4 contains the corresponding Γ and B matrices for this model.

For purposes of illustration, we have modeled the data matrix in Table 3.[3] This structural model does not fit the data particularly

---

[1] The Appendix contains the Lisrel syntax to run every model fit in this article. Strictly speaking, these indices should have been computed on the covariance matrix.

[2] Each link has the support of the marketing literature, but the reader whose favorite links are not represented is encouraged to take the data and fit a competing model.

[3] Alternatively we could have computed the correlations (actually, ideally the covariances) among five scales, each created as the mean of relevant measures (e.g., "quality"=mean of q_1, q_2, q_3), or by using a single item from each set to represent the construct (e.g., q_1, c_1, etc.).

Γ relates ξ's to η's (column → row)
B relates η's to other η's (column → row)

$$\Gamma = \begin{array}{c} \\ \eta_1 = value \\ \eta_2 = csat \\ \eta_3 = repeat \end{array} \begin{array}{cc} \xi_1 = quality & \xi_2 = cost \\ \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & 0 \\ 0 & \gamma_{32} \end{pmatrix} \end{array}$$

$$B = \begin{array}{c} \\ \eta_1 = value \\ \eta_2 = csat \\ \eta_3 = repeat \end{array} \begin{array}{ccc} \eta_1 & \eta_2 & \eta_3 \\ \begin{pmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \end{pmatrix} \end{array}$$

Fig. 4. Structural matrices corresponding to path model in Fig. 3.

well: $X^2{}_3 = 34.28$ ($p = .00$), CFI = 0.83, SRMR = 0.076. The results were as follows: satisfaction → intentions (0.52*), quality → value (0.56∗), quality → satisfaction (0.59*), price → value (-0.51*), price → intention (-0.32*), (*p < .05), and value → satisfaction (0.05, n.s.). Thus, several paths were significant, supporting the hypotheses, but one was not, which is a diagnostic clue that the model may be mis-specified and the theory needs re-thinking.

The full SEM model: the measurement and structural models combined

Let us now combine the two models. A full SEM model is comprised of a measurement model, which relates the variables to the constructs, and a structural model, which relates the constructs to other constructs. The factor analytic models for both the exogenous and endogenous variables are written:

$$x + \Lambda_x \xi + \delta \tag{1a}$$

$$y = \Lambda_y \eta + \varepsilon \tag{1b}$$

The path model of structural coefficients is as before:

$$\eta = B\eta + \Gamma\xi + \zeta, \tag{2}$$

with terms defined as follows:

- x is a vector of independent, or exogenous, variables
- Λ_x is the matrix of exogenous factor loadings of x on ξ
- ξ is a vector of the independent latent variables, exogenous constructs
- δ is vector of measurement errors in x
- y is a vector of observed variable indicators of dependent latent endogenous constructs
- Λ_y is the matrix of endogenous factor loadings of y on η
- η is a vector of latent dependent, or endogenous constructs
- ε is vector of measurement errors in y
- Φ is not shown in Eq. (1a), but it will contain the correlations among the ξ's
- Γ is a matrix of coefficients of the ξ's on the η's (the structural relationships)
- B is a matrix of coefficients of the η's on η's (also part of the structural relationships)
- ζ is a vector of random disturbances as before, with covariance matrix Ψ.

Thus, the researcher attempting to model the relationships among the constructs (ξ's and η's), and variables (x and y), must operate eight matrices: Λ_x, Λ_y, θ_ε, θ_δ, Γ, B, Φ, and Ψ.

Fitting the model in Figs. 5 and 6, we see this model fits the data reasonably well: $X^2{}_{86} = 262.64$ ($p = .00$), CFI = 0.94, SRMR = 0.080. The results were as follows: satisfaction → intentions (0.48*), quality → value (0.56*), quality → satisfaction (0.59*), price → value (−0.50*), price → intention (−0. 29*), (*p < .05), and value → satisfaction (0.06, n.s.). The factor loadings were all significant and large (0.90 and higher). The results of this combined model are consistent with the component models—the structural paths that were significant
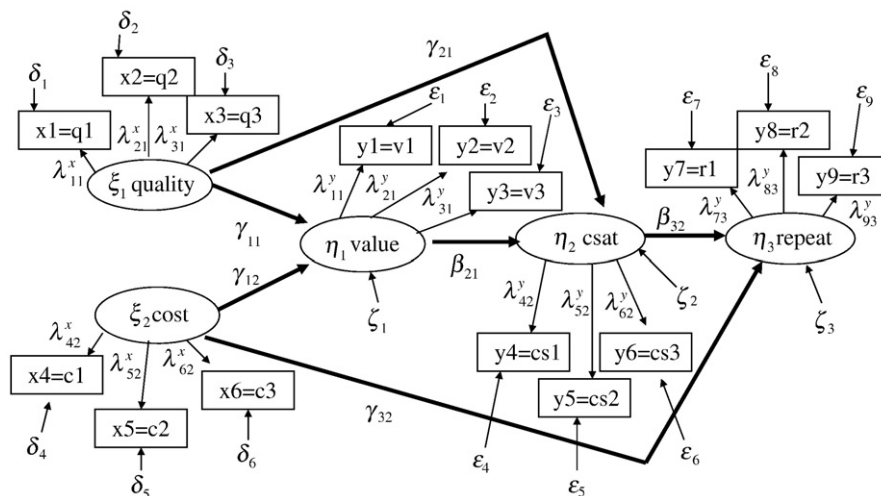

Fig. 5. Full Lisrel model.

before, are so here also, and the measurement model is very similar as well.

The reader using this article as a primer, trying to replicate the results, will have noticed by now that the Lisrel outputs contain even more information that than which have been highlighted thus far. For example, Lisrel computes the "Squared Multiple Correlations for Structural Equations," and prints an R^2 for each eta (η, the endogenous, "dependent" constructs). In this model, value $R^2 = 0.58$, customer satisfaction $R^2 = 0.39$, and repeat inclination $R^2 = 0.33$. As R^2's go, these are not bad, but if there is room for improvement, we might focus on incorporating additional constructs to model customer satisfaction and repeat a little better.

I) Measurement model:

LX: factor loadings for exogenous constructs:

$$\Lambda_x = \begin{matrix} & & \text{quality} & \text{cost} \\ & & \xi_1 & \xi_2 \\ x1=q1 & \\ x2=q2 & \\ x3=q3 & \\ x4=c1 & \\ x5=c2 & \\ x6=c3 & \end{matrix} \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix}$$

LY: factor loadings for endogenous constructs:

$$\Lambda_Y = \begin{matrix} & & \text{value} & \text{csat} & \text{repeat} \\ & & \eta_1 & \eta_2 & \eta_3 \\ y1=v1 & \\ y2=v2 & \\ y3=v3 & \\ y4=cs1 & \\ y5=cs2 & \\ y6=cs3 & \\ y7=r1 & \\ y8=r2 & \\ y9=r3 & \end{matrix} \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{pmatrix}$$

TD: measurement errors on exogenous variables:

$$\Theta_\delta = \begin{matrix} & x1 & x2 & x3 & x4 & x5 & x6 \\ x1 \\ x2 \\ x3 \\ x4 \\ x5 \\ x6 \end{matrix} \begin{pmatrix} \delta_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta_6 \end{pmatrix}$$

TE: measurement errors on endogenous variables:

$$\Theta_\varepsilon = \begin{matrix} & y1 & y2 & y3 & y4 & y5 & y6 & y7 & y8 & y9 \\ y1 \\ y2 \\ y3 \\ y4 \\ y5 \\ y6 \\ y7 \\ y8 \\ y9 \end{matrix} \begin{pmatrix} \varepsilon_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \varepsilon_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \varepsilon_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_9 \end{pmatrix}$$

Fig. 6. Eight Lisrel matrices and notation.

II) Structural model:

GA: relates ξ's to η's:

$$\Gamma = \begin{matrix} & & \xi_1 = \text{quality} & \xi_2 = \text{cost} \\ \eta_1 = \text{value} \\ \eta_2 = \text{csat} \\ \eta_3 = \text{repeat} \end{matrix} \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & 0 \\ 0 & \gamma_{32} \end{pmatrix}$$

BE: relates η's to other η's:

$$B = \begin{matrix} & & \eta_1 & \eta_2 & \eta_3 \\ \eta_1 = \text{value} \\ \eta_2 = \text{csat} \\ \eta_3 = \text{repeat} \end{matrix} \begin{pmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \end{pmatrix}$$

PSI (prediction/modeling) errors on η's:

$$\begin{matrix} & \zeta_1 & \zeta_2 & \zeta_3 \\ \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{matrix} \begin{pmatrix} \psi_1^2 & 0 & 0 \\ 0 & \psi_2^2 & 0 \\ 0 & 0 & \psi_3^2 \end{pmatrix}$$

PHI (inter-correlations among the ξ's):

$$\begin{matrix} & \xi_1 & \xi_2 \\ \xi_1 \\ \xi_2 \end{matrix} \begin{pmatrix} \phi_1^2 & \phi_{12} \\ \phi_{21} & \phi_2^2 \end{pmatrix}$$

Fig. 6 (*continued*).

Similarly, Lisrel computes the "Squared Multiple Correlations for the Y-Variables" and "for the X-Variables." In this model, the R^2's were all extremely high (0.80 and up), but in general, when one is relatively much lower than the others, that R^2 is diagnostic information that the variable is a good candidate to drop from subsequent modeling.

Finally, it is important to share a word about "modification indices." Modification indices are computed for any place in a matrix (measurement or structural) where a parameter had not been included or estimated in the current model. A modification index is large if the model would fit better had that parameter been estimated. Lisrel also prints a guesstimate as to what the parameter estimate would be if it were estimated. These indices are only indicative of what is likely to happen if one parameter were changed at a time. Most Lisrel experts rue the day that modification indices were programmed in as a user option. They are from the statistics devil—they are seductive because all users seek better fit indices, but they result in models that are nonsensical, such as paths that are not supported by theory or literature, or variables cross-loaded on more than one factor, for example.

Conclusion

We hope this article has provided the reader a grasp of the fundamental SEM model. We close with a few suggestions. These comments are equally relevant to the researcher building and testing models as to the reviewer assessing a paper in which the authors had used SEM.

1. SEMs are not scary—they are natural progressions from factor analysis and regression.
2. As such, be careful not to overinterpret path coefficients as if they were causal, any more so than if the results had been obtained via regression.

SEM could be used more frequently among academics, and in industry wherever practitioners espouse conceptual models. For more information, read Anderson and Gerbing (e.g., 1988), Bagozzi (1980), Bentler and Dudgeon (1996), Bollen, Kenneth and Long (1993), Byrne (1998), Kline (2004), Raykov and Marcoulides (2000), and Schumacker and Lomax (2004).

Acknowledgments

Appendix I: LISREL commands for fitting SEM models

The LISREL software may be obtained at www.ssicentral.com. Other software packages include Eqs (www.mvsoft.com), and Amos (available through SPSS). Useful documentation is provided by Barbara Byrne (1998) for Lisrel (and other books on Eqs and Amos). We focus on Lisrel because it is the most comprehensive of the packages, and the matrix notation in the literature and SEM texts derive from Lisrel.

1) Confirmatory factor analysis syntax:

```
Confirmatory factor analysis example.
da ni=15 no=100 ma=cm
la
q1 q2 q3 c1 c2 c3 v1 v2 v3 cs1 cs2 cs3 r1 r2 r3
cm sy
2.813
2.264 1.973
1.961 1.670 1.528
etc.
… 1.530
mo nx=15 nk=5 lx=fu,fr td=di,fr ph=st,fr
pa lx
1 0 0 0 0
1 0 0 0 0
1 0 0 0 0
0 1 0 0 0
0 1 0 0 0
0 1 0 0 0
0 0 1 0 0
0 0 1 0 0
0 0 1 0 0
0 0 0 1 0
0 0 0 1 0
0 0 0 1 0
0 0 0 0 1
0 0 0 0 1
0 0 0 0 1
pd
out me=ml rs mi
```

Commentary: first line is a title, arbitrary but ends with a period. "da" stands for dataline, with number of input variables (ni=15), number of observations or sample size (no=100) and that the matrix to be analyzed is a covariance matrix (ma=cm). "la" says that the variable labels follow on the next line. "cm sy" says that a covariance matrix follows, in symmetric format (lower triangle of the matrix). "mo"=model, nx=number of $\times$ variables, nk=number of ksi (constructs). "lx" is the lambda x (factor loadings matrix) which begins as a "fu,fr" full matrix (15×5) with all elements free to be estimated (this status is modified shortly). "td" is theta delta and all its diagonal elements are free to be estimated. "ph" is phi, the factor intercorrelation matrix, it is "st" standardized to be a correlation matrix. "pa lx" tells Lisrel that the pattern for the lx matrix follows; 0's mean the loading is not to be estimated and is set by default to zero; 1's mean the loading is free to be estimated (to any value). "pd" draws a path diagram for the user (very cool). "out" requests the appropriate outputs. "me=ml" says the estimation method is maximum likelihood (the default). "rs" produces residuals and "mi" produces modification indices.

2) Path analysis syntax

```
Path analysis.
da ni=5 no=100 ma=cm
la
qual price value cs intent
cm sy
1.00
-0.03 1.00
0.58 -0.53 1.00
0.62 -0.22 0.39 1.00
0.39 -0.43 0.19 0.59 1.00
se
value cs intent qual price
mo  ny=3  ne=3  nx=2  nk=2  ly=id,fi  te=ze,fi  lx=id,fi
td=ze,fi ph=st be=fu,fr ga=fu,fr
pa be
0 0 0
1 0 0
0 1 0
pa ga
1 1
1 0
0 1
pd
out me=ml rs mi
```

Commentary: the new lines include "se" which can serve two functions: first, it can be used to select a subset of the variables in the "la" list, and second, it can be used to rearrange the variables, to put the endogenous variables first, which is a rule required by Lisrel. In the model statement, the factor loadings matrices, lx and ly, are fixed to be identity matrices (=id,fi), and their corresponding error matrices, td and te, are fixed to be zero matrices (=ze,fi). The patterns for beta and gamma follow— beta is $ne \times ne$, and gamma is $nk \times ne$.

3) Full SEM syntax

```
Big example woohoo.
da ni=15 no=100 ma=cm
la
q1 q2 q3 c1 c2 c3 v1 v2 v3 cs1 cs2 cs3 r1 r2 r3
cm sy
2.813
2.264 1.973
1.961 1.670 1.528
etc.
… 1.530
se
v1 v2 v3 cs1 cs2 cs3 r1 r2 r3 q1 q2 q3 c1 c2 c3
mo ny=9 ne=3 nx=6 nk=2 ly=fu,fr te=di,fr lx=fu,fr
td=di,fr be=fu,fr ga=fu,fr
pa ly
0 0 0
1 0 0
1 0 0
0 0 0
0 1 0
0 1 0
0 0 0
0 0 1
0 0 1
pa lx
0 0
1 0
1 0
0 0
0 1
0 1
pa be
0 0 0
1 0 0
0 1 0
pa ga
1 1
1 0
0 1
va 1.0 ly(1,1) ly(4,2) ly(7,3)
pd
out me=ml rs mi
```

Commentary: the new lines include "va 1.0" which fixes the three lambda y's to 1.0.

References

Anderson, James C., & Gerbing, David W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411−423.

Anderson, James C., Gerbing, David W., & Hunter, John E. (1987). On the assessment of unidimensional measurement: Internal and external consistency, and overall consistency criteria. *Journal of Marketing Research*, *24*, 432−437 (November).

Anderson, James C., & Narus, James A. (1990). A model of distributor firm and manufacturer firm working partnerships. *Journal of Marketing*, *54*, 42−58 (Jan.).

Bagozzi, Richard P. (1980). *Causal modeling in marketing.* NY: Wiley.

Baumrind, Diana (1983). Specious causal attributions in the social sciences. *Journal of Personality and Social Psychology*, *45*(6), 1289−1298.

Bentler, Peter M., & Dudgeon, Paul (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology*, *47*, 563−592.

Bollen, Kenneth A., & Scott Long, J. (Eds.). (1993). *Testing structural equation models* Newbury Park, CA: Sage.

Browne, Michael W., MacCallum, Robert C., Kim, Cheong-Tag, Andersen, Barbara L., & Glaser, Ronald (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*(4), 403−421.

Byrne, Barbara M. (1998). *Structural equation modeling with Lisrel, Prelis, and Simplis: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Gerbing, David W., & Anderson, James C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*(May), 186−192.

Holland, Paul W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945−960.

Iacobucci, Dawn (2008). *Mediation analysis.* Thousand Oaks, CA: Sage.

Kline, Rex B. (2004). *Principles and practice of structural equation modeling*, (2nd ed.) New York: Guildford.

Luo, Xueming, & Bhattacharya, C. B. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of Marketing*, *70*, 1−18 (Oct.).

Raykov, Tenko, & Marcoulides, George A. (2000). *A first course in structural equation modeling.* Mahwah, NJ: Erlbaum.

Rubin, Donald B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322−331.

Schumacker, Randall E., & Lomax, Richard G. (2004). *A beginner's guide to structural equation modeling*, 2nd ed. Erlbaum: Mahwah, NJ.

Selnes, Fred, & Sallis, James (2003). Promoting relationship learning. *Journal of Marketing*, *67*(July), 80−95.

Steenkamp, Jan-Benedict E. M., & Baumgartner, Hans (2000). On the use of structural equation models in marketing modeling. *International Journal of Research in Marketing*, *17*(June), 195−202.