Chapter 8

Models with Structural and Measurement Components

Good people are good because they've
come to wisdom through failure.

—William Saroyan

Overview

- Characteristics of SR models
- Analysis of SR models
- Estimation of SR models
- Equivalent SR models
- Analyzing single indicators

Characteristics of SR models

- Structural regression (SR) models—also called hybrid or LISREL models—can be viewed as syntheses of path and measurement models

- They are the most general of all the types of structural equation models considered to this point:

    1. As in path analysis, the specification of a SR model allows tests of hypotheses about patterns of casual effects

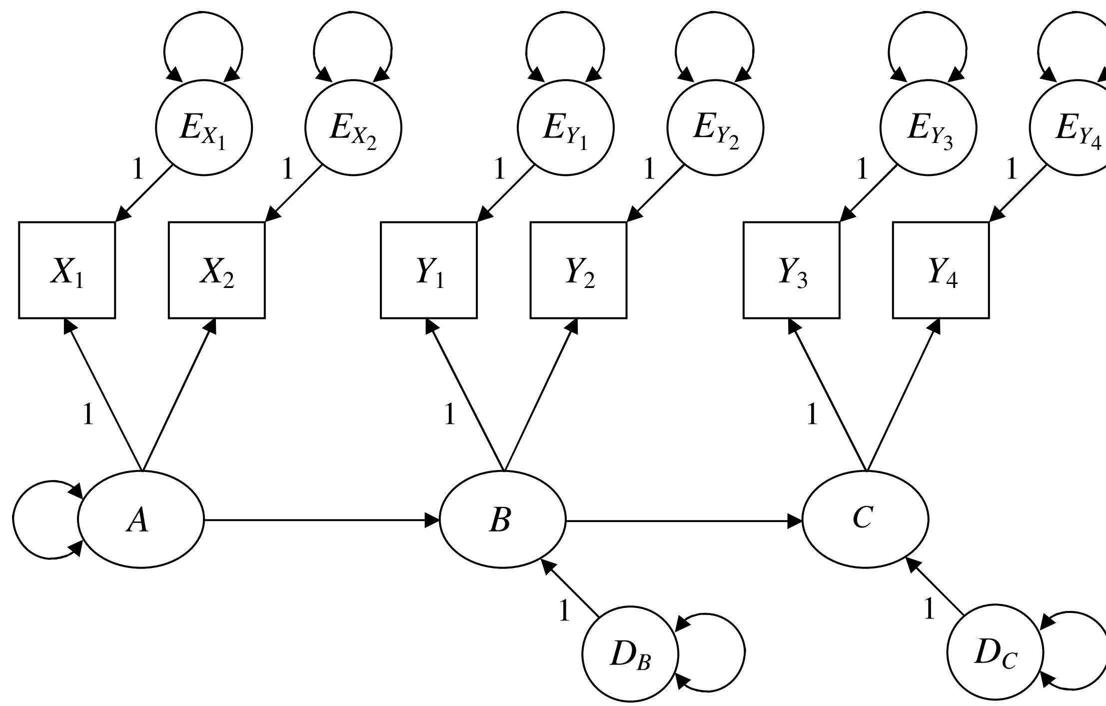
    2. Unlike path models, though, these effects can involve latent variables

    3. This is because a SR model also incorporates a measurement model, just as in confirmatory factor analysis (CFA)

Characteristics of SR models

- Example of a "fully latent" SR model (Figure 8.1(b))
- The structural model is $A \rightarrow B \rightarrow C$, and the measurement model concerns the predicted relation between the indicators and these latent variables:

(b) SR Model



Analysis of SR models

- The analysis of SR models can generally be decomposed into two parts:

    1. The specification of the structural component of a SR model follows the same basic rationale as in path analysis

    2. The specification of the measurement component of a SR model requires consideration of basically the same issues as in CFA

Analysis of SR models

- Evaluation of whether a SR model is identified and its subsequent estimation should also both be conducted separately for each part, measurement and structural
- There is a common theme to identification and analysis:

A valid measurement model is needed before the structural component of a SR model can be evaluated

Analysis of SR models

- This discussion assumes a “fully latent” SR model

- Any SR model must satisfy the same two necessary requirements for identification as any other kind of structural equation model: Every latent variable has a scale and $df_M \geq 0$

Analysis of SR models

- Parameters of SR models are counted as follows:

  The total number of (a) variances and covariances (i.e., unanalyzed associations) of exogenous variables (measurement errors, disturbances, and exogenous factors), and (b) direct effects on endogenous variables (factor loadings of indicators, direct effects on endogenous factors from other factors) equal the number of parameters

Analysis of SR models

- It is generally assumed that the

    1. exogenous factors are uncorrelated with the disturbances of the endogenous factors

    2. factors (exogenous or endogenous) and the measurement errors are independent

Analysis of SR models

- Disturbances and measurement errors in SR models are usually assigned a scale through unit loading identification (ULI) constraints that fix the unstandardized residual path coefficients to 1.0

- Exogenous factors can be scaled either by imposing a

  1. ULI constraint where the loading of one indicator per factor (that of the reference variable) is fixed to 1.0 (i.e., the factor is unstandardized) or a

  2. unit variance identification (UVI) constraint where the factor variance is fixed to equal 1.0 (i.e., the factor is standardized)

Analysis of SR models

- However, most SEM computer programs allow only the first method mentioned (i.e., a ULI constraint) for scaling endogenous factors
- This is because the variances of endogenous variables are not generally considered model parameters
- This also implies that endogenous factors are unstandardized in perhaps most analyses

Analysis of SR models

- When a SR model is analyzed in a single sample, the choice between scaling an exogenous factor with either ULI or UVI constraints combined with the use of ULI constraints only to scale endogenous factors usually makes no difference

- An exception for SR models where some factors have only two indicators is when the identification constraints interact with equality constraints imposed on either factor loadings or path coefficients

- This is constraint interaction, which for SR models is considered later

Analysis of SR models

- An additional requirement for identification reflects the view that the analysis of a SR model is essentially a path analysis conducted with estimated variances and covariances among the factors
- Thus, it must be possible to derive unique estimates of the factor variances and covariances before specific causal effects among them can be estimated
- That is, in order for the structural portion of a SR model to be identified, its measurement portion must be identified

Analysis of SR models

- Bollen (1989) described the requirements for identification just described as the *two-step rule*
- The steps to evaluate it are as follows:

  1. Respecify the SR model as a CFA model with all possible unanalyzed associations among the factors, then evaluate this model against the requirements for identification outlined in the previous chapter

Analysis of SR models

- The steps to evaluate the two-step rule are:

  2. View the structural portion of the SR model as a path model

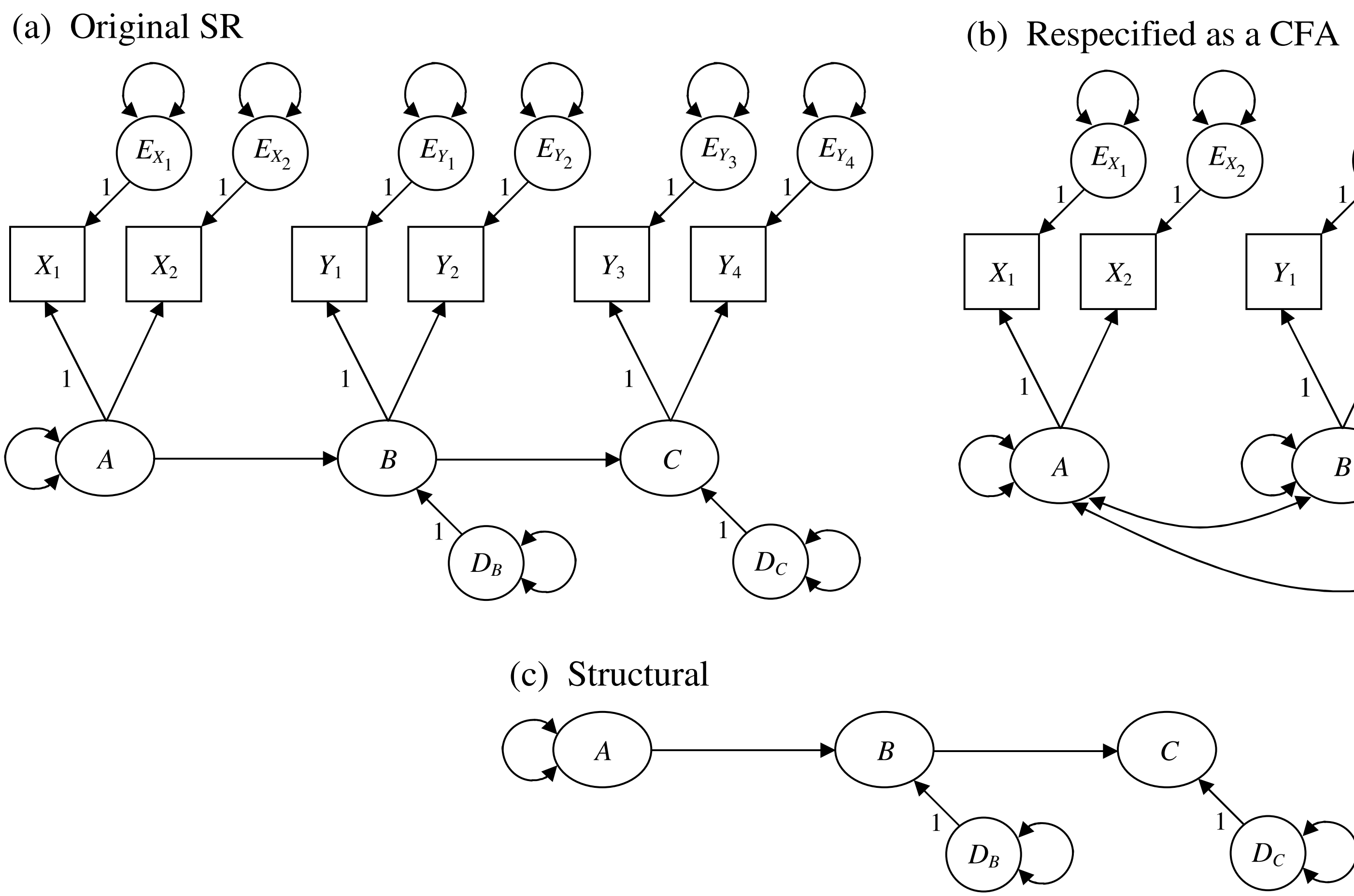
     a. If it is recursive, then the structural model is identified

     b. If it is nonrecursive, then evaluate the structural model against the requirements for identification outlined in the next chapter

- SR models that satisfy both parts of the two-step rule are identified (assuming the two necessary conditions are met)

Analysis of SR models

- Example: SR model of Figure 8.2(a) respecified as a CFA measurement model (Figure 8.2(b)) and as a structural model (Figure 8.2(c))—this model is identified:



Analysis of SR models

- It is not always possible to determine the identification status of every SR model using the two-step rule—some examples:

    1. The measurement portion of a SR model expressed as a CFA model is not standard due to the presence of correlated measurement errors or indicators that load on two or more factors (i.e., multidimensional measurement)

    2. The structural model is nonrecursive such that there is no easily-applied sufficient condition to determine whether it is identified

Analysis of SR models

- If either the measurement or structural portions of a SR model are "none of the above" such that their identification cannot be clearly established, the two-step rule may be too strict
- This is also true for "partially latent" SR models with single indicators (considered later)

Analysis of SR models

- There are two general approaches to testing SR models:

  1. Two-step modeling (e.g., Anderson & Gerbing, 1988)

  2. Four-step modeling (e.g., Mulaik & Millsap, 2000)

- Both methods

  1. generally require a fully latent SR model

  2. deal with the problem of how to locate the source of specification error in a SR model with poor fit to the data

- The second goal is accomplished by separating the analysis of the measurement model from that of the structural model

Analysis of SR models

- *Two-step modeling* parallels the two-step rule for identification
- In this approach, a SR model is first respecified as a CFA measurement model (e.g., Figure 8.2(b))
- The CFA model is then analyzed in order to determine whether it fits the data
- If the fit of this CFA model is poor, then

    1. not only may the researcher's hypotheses about measurement be wrong but
    2. the fit of the original SR model to the data may be even worse if its structural model is overidentified

Analysis of SR models

- The first part of two-step modeling thus involves finding an acceptable CFA measurement model
- If the initial measurement model is rejected, then the suggestions discussed in the previous chapter for respecification can be followed
- Given an acceptable measurement model, the second stage of two-step modeling is to compare the fits of the original SR model and those with different structural models to one another and to the fit of the CFA model with the chi-square difference test (assumes hierarchical models)

Analysis of SR models

- *Four-step modeling* is basically an extension of two-step modeling
- It is intended to even more precisely diagnose specification error in the measurement model than two-step modeling
- In four-step modeling, the researcher specifies and tests a sequence of at up to four hierarchical models
- In order for these nested models to be identified, each factor in the original SR model should have at least four indicators

Analysis of SR models

- As in two-step modeling, if the fit of a model in four-step modeling with fewer constraints is poor, then a model with even more constraints should not even be considered

- The least restrictive model specified at the first step is a basically an exploratory common factor analysis model where

  1. all indicators are specified to load on each factor, and

  2. the number of factors is the same as that in the original SR model

Analysis of SR models

- This least restrictive model should be analyzed with the same method of estimation, such as maximum likelihood (ML), as used to analyze the final SR model
- This first step is intended to test the provisional correctness of the hypothesis regarding the number of factors, but it cannot confirm that the correct number of factors has been specified (e.g., Hayduk & Gleser, 2000).

Analysis of SR models

- The second step of four-step modeling is basically the same as the first step of two-step modeling: A CFA model is specified wherein the loadings (pattern coefficients) of indicators on certain factors are fixed to zero

- If the fit of the CFA model at the second step is acceptable, one goes on to test the original SR model; otherwise, the measurement model should be respecified

Analysis of SR models

- The third step involves testing the SR model with the same pattern of zero pattern coefficients as represented in the measurement model from the second step but where

  1. at least one unanalyzed association between two factors is respecified as a direct effect or reciprocal effect

  2. some of the factors are respecified as endogenous

Analysis of SR models

- The last step of four-step modeling involves tests of a priori hypotheses about parameters free from the outset of model testing
- These tests typically involve the imposition of zero or equality constraints, such as on pairs of reciprocal direct effects between two factors
- The third and fourth steps of four-step modeling are basically a more specific statement of activities that could fall under the second step of two-step modeling

Analysis of SR models

- The goal of both two-step modeling and four-step modeling is the same: to find a parsimonious structural model that still explains the data (estimated covariances among the factors)
- Both methods capitalize on chance variation when measurement and SR models are tested and respecified using data from the same sample
- The two-step method has the advantage of simplicity, and it does not require models with at least four indicators per factor

Analysis of SR models

- Both methods (two- and four-step) are better than one-step modeling where there is no separation of measurement issues from the estimation of causal effects among factors

- Neither method is a "gold standard" for testing SR models, but there is no such thing in SEM (e.g., Bentler, 2000; Bollen, 2000)

Estimation of SR models

- Interpretation of parameter estimates from the analysis of a SR model should not be difficult if one knows about path analysis and CFA

- Path coefficients are interpreted for SR models as regression coefficients for effects on endogenous variables from other variables presumed to directly cause them, just as for path models

- Total effects among the factors that make up the structural part of a SR model can be broken down into direct and indirect effects using the principles of effects decomposition from path analysis

Estimation of SR models

- Factors loadings are interpreted for SR models as regression coefficients for effects of factors on indicators, just as they for CFA models
- For CFA and SR models alike, a zero pattern coefficient does not imply a zero structural coefficient
- That is, indicators typically have nonzero model-implied correlations with factors other than the one(s) they are specified to measure
- Look out for Heywood cases in the solution, such as negative variance estimates, that suggest a problem with the data, specification, or identification status of the model

Estimation of SR models

- Some SEM computer programs print estimated squared multiple correlations (R^2_{smc}) for each endogenous variable

- This includes for SR models the indicators and factors with direct effects on them from other factors

- Values of R^2_{smc} are usually computed for indicators in the unstandardized solution as the one minus the ratio of the estimated measurement error variance over the observed variance of that indicator

Estimation of SR models

- Variances of endogenous variables are not generally considered model parameters, but they nevertheless have model-implied variances

- Therefore, values of R^2_{smc} are usually calculated for endogenous factors as one minus the ratio of the estimated disturbance variance over the predicted variance for that factor

Estimation of SR models

- Most SEM computer programs calculate a standardized solution for SR models in two steps:

    1. Find the unstandardized solution with ULI constraints for endogenous factors

    2. Transform this solution to standardized form

- Steiger (2002) noted that this method assumes that the ULI constraints function only to scale the endogenous variables (i.e., they do not affect overall model fit)

- This assumption is probably valid most of the time, but not when there is constraint interaction

Estimation of SR models

- Recall that constraint interaction for CFA models is indicated when the value of χ^2_D for the test of the equality-constrained loadings of indicators on different factors depends on how the factors are scaled

- Steiger (2002) shows that the same phenomenon can happen with SR models where some factors

    1. have only two indicators, and

    2. when estimates of direct effect on two or more *different* endogenous factors are constrained to be equal

Estimation of SR models

- Constraint interaction can also result in an incorrect standardized solution for a SR model if calculated in the way just described (i.e., in two steps)

- If the analysis of standardized factors can be justified, then the method of constrained estimation can be used to

    1. test hypotheses of equal standardized path coefficients

    2. generate a correct standardized solution

- Constrained estimation of a SR model standardizes all factors, exogenous and endogenous

Equivalent SR models

- Just as in path analysis and CFA, it is often possible to generate equivalent versions of SR models
- An equivalent version of an SR model with a just-identified structural model is the measurement portion the model respecified as a CFA model
- If the structural portion of the SR model is just-identified, then any variation on it also generates an equivalent model

Equivalent SR models

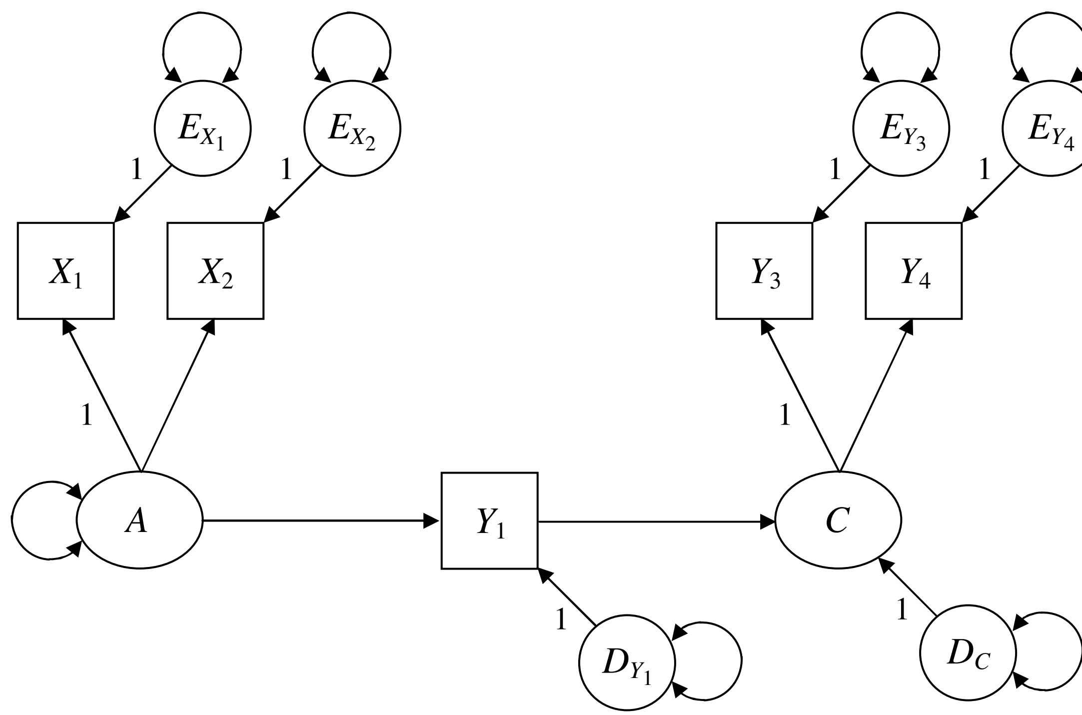
- If the structural portion of a SR model is overidentified, then the Lee-Herschberger replacing rules can be applied to generate an equivalent version of it
- Given no change in the measurement model, alternative SR models with equivalent structural models will fit the same data equally well
- Holding the structural model constant, it may also be possible to generate equivalent versions of the measurement model using Herschberger's reversed indicator rule

Analyzing single indicators

- There are times when a researcher has only one measure of some construct
- Scores from a single indicator are quite unlikely to have no measurement error, that is, to be both perfectly reliable and valid

Analyzing single indicators

- There is an alternative to representing a single indicator in the structural part of a SR model as one would in path analysis (i.e., without a measurement error term)—for example:



Analyzing single indicators

- For the model with a single indicator just shown, measurement error in Y_1 is reflected in its disturbance term
- A method to take account of measurement error in a single indicator is described next
- It requires an a priori estimate of the proportion of variance of the single indicator that is due to measurement error (.10, .20, etc.)
- This estimate may be based on the researcher's experience with the measure or on results reported in the research literature

Analyzing single indicators

- Recall that

  1. one minus a reliability coefficient, $1 - r_{XX}$, estimates the proportion of observed variance due to random error, which is only one source of measurement error

  2. specific types of reliability coefficients usually estimate only one kind of random error

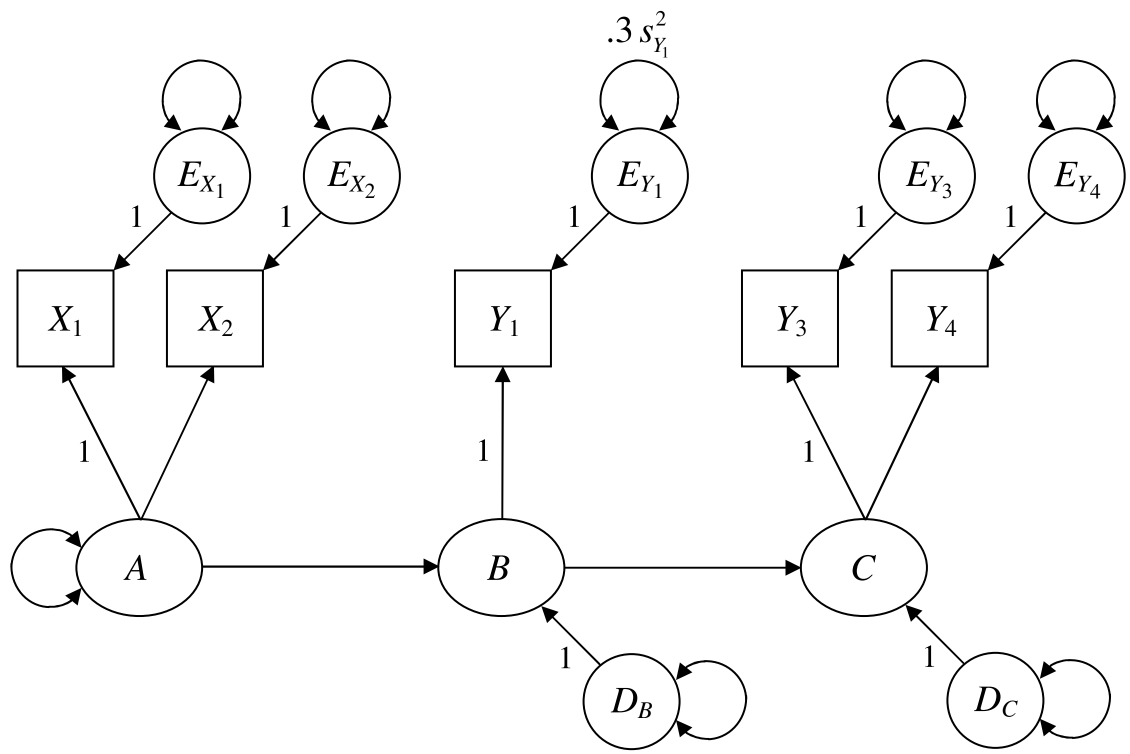
- Accordingly, the quantity $1 - r_{XX}$ probably *underestimates* the proportion of measurement error variance in a single indicator

Analyzing single indicators

- Suppose that variable Y_1 is the only indicator of an endogenous construct *B* and that the researcher estimates that 30% of Y_1's variance is due to measurement error

- Given this estimate, it is possible to specify a SR model like this one (Figure 8.5(b)):

(b) Single Indicator of an Endogenous Construct



Analyzing single indicators

- In the model just shown, the variance of the measurement error term for Y_1 is fixed to equal .30 times the observed variance, $.3\ s^2_{Y_1}$
- Because factor *B* must have a scale in order for the model to be identified, the unstandardized loading of Y_1 on *B* is fixed to equal 1.0
- With the specification of a residual term for Y_1, the direct effect on factor *B* is estimated controlling for measurement error in its single indicator Y_1

Analyzing single indicators

- Some more information about the method just described for taking account of measurement error in a single indicator:

    1. Specification of the measurement error variance for a single indicator as a free parameter may result in an identification problem

    2. If the researcher is uncertain about his or her estimate of the proportion of measurement error variance for a single indicator, then the model can be analyzed with a range of estimates

- See Bedeian, Day, and Kelloway (1997) for additional discussion

References

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

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement*, *57*, 785-799.

Bentler, P. M. (2000). Rites, wrongs, and gold in model testing. *Structural Equation Modeling*, *7*, 82-91.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K. A. (2000). Modeling strategies: In search of the holy grail. *Structural Equation Modeling*, *7*, 74-81.

Hayduk, L. A., & Glaser, D. N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, *7*, 1-35.

Muliak, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, *7*, 36-73.

Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods*, *7*, 210-227.