The Report Committee for Xuan Xiao
Certifies that this is the approved version of the following report:

Structural Equation Modeling Compared with Ordinary Least Squares in Simulations and Life Insurers' Data

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

____
Thomas W. Sager

____
S. Natasha Beretvas

Structural Equation Modeling Compared with Ordinary Least Squares in Simulations and Life Insurers' Data

by

Xuan Xiao, B.S.

Report

Presented to the Faculty of the GraduateSchool of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2013

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Thomas Sager for his dedicated supervision, insightful comments, ongoing feedback, and extended support and patience throughout the writing process of this report. I would also thank Prof. Beretvas for her kindly help and useful advice.

Additional thanks go to professors and graduate students of Division of Scientific Computation & Statistics at the University of Texas at Austin for their help and encouragement during my graduate training.

Abstract

Structural Equation Modeling Compared with Ordinary Least Squares in Simulations and Life Insurers' Data

Xuan Xiao, M.S.Stat.
The University of Texas at Austin, 2013

Supervisor: Thomas W. Sager

Structural equation model (SEM) is a general approach to analyze multivariate data. It is a relatively comprehensive model and combines useful characteristics from many statistical approaches, thus enjoys a variety of advantages when dealing complex relationships. This report gives a brief introduction to SEM, focusing especially the comparison of SEM and OLS regression. A simple tutorial of how to apply SEM is also included with the introduction and comparison. SEM can be roughly seen as OLS regression added with features such as simultaneous estimation, latent factors and autocorrelation. Therefore, SEM enjoys a variety of advantages over OLS regression. However, it is not always the case that SEM will be the optimal choice. The biggest concern is the complexity of SEM, for simpler model will be preferable for researchers when the fitness is similar. Simulations on two datasets, one requires special features of SEM and one satisfies assumptions of OLS regression, are applied to illustrate the choice between SEM and OLS regression. A study using data from US life insurers in the year 1994 serves as a further illustration. The conclusion is when special features of SEM is required, SEM fits better and will be the better choice, while when OLS regression assumptions are satisfied, SEM and OLS regression will fit equally well, considering the complexity of SEM, OLS regression will be the better choice.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Structural equation model (SEM) is a general approach to analyze multivariate data. More specifically it can be viewed as a series of statistical methods allowing complex relationships between independent variables and dependent variables. SEM is relatively comprehensive and combines the characteristics of several analysis approaches. Multiple regression, path analysis, factor analysis, time series analysis, analysis of covariance, etc. can all be seen as special cases of SEM. In SEM, there are two variable specification methods. One way is to categorize variables as exogenous and endogenous, and the other is to categorize variables as manifest and latent. The relationship between a manifest variable and a latent variable is called loading, factor loading, or factor coefficient, while the relationship between two latent variables is called path, path loading, or path coefficient. Since SEM enjoys an increasing popularity, many software programs have developed tools to assist analyzing SEM, such as LISREL, EQS, Mplus, Amos, MX, SYSTAT, STATA, R, and SAS. This report employs SAS to analyze both SEM and OLS regression.

Compared to OLS regression, SEM enjoys a variety of advantages, such as the ability to analyze simultaneously, the ability to include latent variables, the ability to analyze time series data, more flexible assumptions, ability to test non-normal data, testing models with large number of equations as a whole and obtain global fit measures, ability to model mediating variables rather than additive models, ability to model error terms, and graphical modeling, etc.

However, it is not always optimal to choose SEM over OLS regression. The biggest concern is the complexity of SEM. Many would say that the choice depends upon the appropriateness of the model forms. If the fitting ability is similar, researchers will prefer the simpler models. In this sense, when the data are not so complex and satisfy the assumption of OLS regression, SEM may be overkill and OLS regression will be the better choice.

To illustrate the comparison of SEM and OLS regression, simulations on two datasets with randomly generated data and a case with real data are applied. In the first simulation case, we generate a set of random variables with an underlying structure, and then simulate predicted variables with fixed coefficients. We simulate the two predicted variables to explain each other, thus the model needs to be estimated simultaneously. When fitting this model, SEM provides estimated coefficients much closer to the "true" coefficients we set, which means SEM fits better than OLS regression. Then in the second simulation case, we generate another set of random variables which are mutually independent, then simulate predictor variables with fixed coefficients. In this simulation, the two predictor variables do not explain each other. Such a model meets the assumptions of OLS regression. When fitting this model, SEM provides almost the same estimated coefficients as OLS regression, and both of the coefficients are very close to the "true" coefficients we set, which means in the situation where OLS regression assumptions are satisfied, both SEM and OLS regression fit well. Considering the complexity of SEM, the optimal choice will be OLS regression.

In this report, we also use the data from annual statement of US life insurers in the year 1994 to further illustrate the comparison of SEM and OLS regression. The model here is to study how the risks of a life insurer affect its capital structure decision. In order keep the model simple and clear, we only include product risks which align with product underwriting, one of the major activities of life insurers. Because different products should have different risks, we focus on the risks of two life insurance products: health insurance and annuities, which represent relatively higher and lower risk respectively. Since risks cannot be observed directly, proxies are used to represent them. Here we use proportions of premium for each product line to the total premium as the risk proxies. However, in theory, risk proxy can be viewed as a mixture of “true" risk and other variables. In order to separate the pure effect of "true" risk, the feature of latent variables is applied, and we call this pure risk a "risk factor". Also, when separating the effects of risk factors from corresponding risk proxies, simultaneous estimation is needed. The model uses capital ratio of insurers to represent the capital structures, and three control variables are included, which are size of the insurer, the governance structure (whether the insurer is a mutual insurance company or a stock company), and group membership (whether the insurer is affiliated with a group of companies. In SEM, capital ratio is regressed on the two risk factors and the three control variables, while in OLS regression, capital ratio is regressed on the two risk proxies and also the three control variables. The results show although the overall fit of both models is not especially close, the R-square of SEM is higher than the R-square of OLS regression, which indicates a better fit of

SEM. Since SEM is very complex in nature, R-square alone is not a perfect index to indicate better fit. Moreover, although the purpose of comparing SEM and OLS regression is served, if we want to explain an actual economic phenomenon or reach useful theoretical conclusions, the SEM model here needs further improvements.

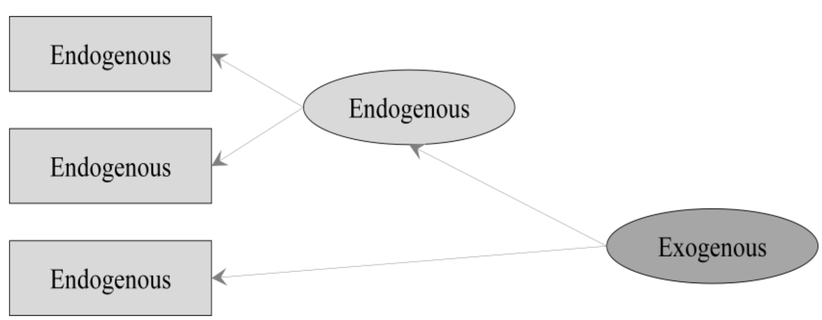
STRUCTURAL EQUATION MODELING

Structural equation modeling (SEM) is a general approach to multivariate data analysis. One definition of SEM is: "a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions", which was articulated by Sewall Wright (1921), TrygveHaavelmo (1943) and Herbert A. Simon (1953), and formally defined by Judea Pearl (2000) using a calculus of counterfactuals. In some circumstances, SEM is also known by other names including covariance structural analysis, latent variable modeling, or causal modeling.

SEM is a series of statistical methods allowing complex relationships between one or more independent variables and one or more dependent variables, some of which can be hypothesized or unobservable. That is, SEM can be thought of as a hybrid between regression and some form of factor analysis. In fact, SEM is a general model combining characteristics of multiple regression, path analysis, factor analysis, time series analysis, analysis of covariance, etc. Therefore, these procedures can be seen as special cases of SEM. For instance, ordinary least-square (OLS) regression can roughly be seen as SEM without simultaneity, latent factors, and autocorrelation(the comparison of SEM and OLS will be discussed later in this report); path analysis can be seen as SEM with only manifest variables thus without the part of measurement model.

In SEM, there are two variable specification methods. The first method is to distinguish variables to be exogenous variables and endogenous variables. Exogenous variables are variables that other variables regress on, while never regress on other variables, and which means that exogenous variables are only used to predict others but never be predicted by others. In a directed graph of the model, as shown in Figure 1, an exogenous variable is recognizable as an oval from which arrows only emanate, where the emanating arrows denote which variables that exogenous variable predicts. Endogenous variables are variables that regress on other variables, even if other variables also regress on them. In a directed graph, as shown in Figure 1, an endogenous variable is recognizable as any variable that receives an arrow. If comparing with regression, exogenous variables are similar to independent variables, whereas endogenous variables are similar to dependent variables.

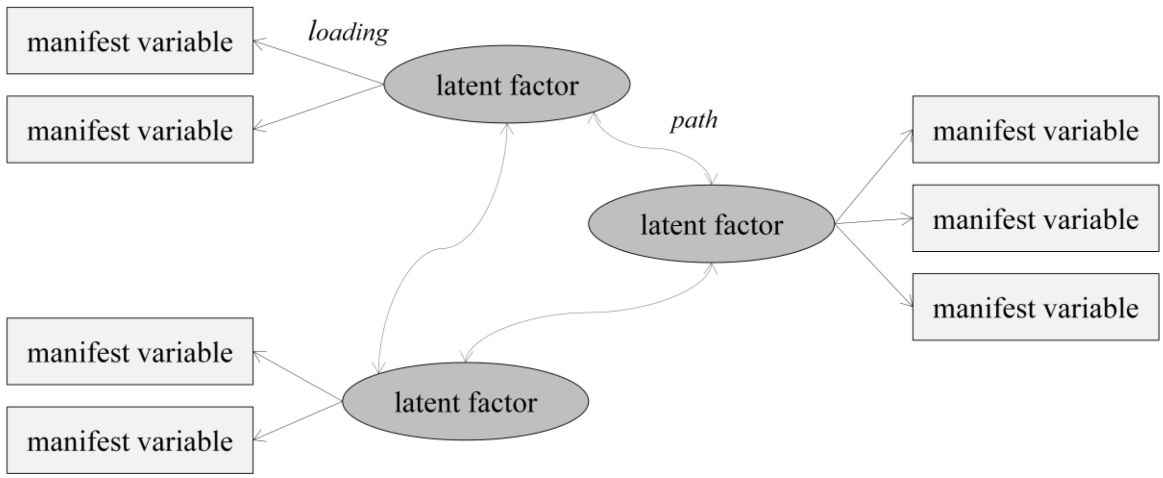
Figure 1: Endogenous and exogenous variables in SEM.



One of the advantages SEM takes from factor analysis is the ability to include unobserved factors, which are referred to as latent variables or latent factors. Latent variables are not directly observed but are rather inferred through a theoretical model from other variables that are observed. In SEM, the directly observed variables are referred to as manifest variables or indicator variables. This is the second method to distinguish variables by their roles. In some circumstances, latent variables correspond to aspects of reality which could be measured in principle, but may not be measured for practical reasons. In other circumstances, latent variables correspond to abstract concepts, like living standard, business confidence, or happiness.

In SEM graph, as shown in Figure 2, squares or rectangles denote manifest variables; circles or ellipses denotes latent variables; curve, double-headed arrows denote bi-directional relationships (often referred to as correlations or covariance),and straight, single-headed arrows denote causal relationships. There are two types of causal relationships shown in most SEM graphs: loading (also referred to as factor loading or factor coefficient), which is assumed causal relationship between a latent factor and its indicator or manifest variables, and path (also referred to as path loading or path coefficient), which is hypothesized causal relationship between two latent factors.

Figure 2: Manifest and latent variables in SEM.



The two main components of SEM are the structural model and the measurement model. The structural model shows the potential causal dependencies between endogenous and exogenous variables, involving all variables in the model, both manifest and latent; while the measurement model shows the relations between latent variables and their indicators. Path analysis can be viewed as SEM with only the structural part, while factor analysis can be viewed as SEM with only the measurement part.

Many software programs are available to analyzing SEM, including:

LISREL, one of the earliest SEM program and perhaps the most frequently referenced program in SEM articles. Software web site: http://www.ssicentral.com/.

EQS, provides many general statistics functions including SEM. Software web site: http://www.mvsoft. com/.

Mplus, has base program to analyze single-level models and add-on modules to analyze multilevel models and models with latent categorical variables. Software web site: http://www.statmodel.com/.

Amos, is distributed with SPSS. Software web site: http://www.smallwaters.com or http://www.spss.com/ amos/.

MX, software web site: http://www.vcu.edu/mx/.

SYSTAT, the RAMONS module. Software web site: http://www.systat.com/.

STATA, software web site: http://www.stata.com/

R, has packages for SEM such as sem, lavaan. Software web site: http://www.r-project.org/.

In this report, I use SAS with the CALIS procedure to analyze SEM. SAS, originally Statistical Analysis System, is an integrated system of software products provided by SAS Institute Inc.(http://www.sas.com/).

COMPARISON OF SEM AND OLS REGRESSION

As stated in previous paragraphs, SEM can roughly be seen as OLS regression, to which may be added simultaneity, latent factors and autocorrelation. Therefore, SEM has several advantages over OLS regression and enjoys wider applications in a variety of research situations than OLS regression.

The first advantage is that, in SEM, a variable can be both predicted and explanatory, which means a variable can appear on the left side of one equation and on the right side(s) of one or more equation(s) in the same model, thus multiple interacting equations can be modeled simultaneously, whereas in OLS regression, a variable can only either be response or explanatory, thus if theoretical study requires a variable to play both roles, it has to be in separated models and cannot estimated at the same time. This is the simultaneous feature SEM has over OLS regression.

The second advantage is the allowance of including unobserved latent variables in SEM. Latent variables are useful tools in a variety of analyses. For instance, they can be used to present variables that cannot be observed or measured directly, or to reduce the number of variables by concentrating information of several variables into less conceptual variables, as in factor analysis. Also, by having multiple indicators per latent factor, the measurement error may reduce.

The third advantage is that SEM is able to deal with time series data. OLS regression has relatively strict assumptions: (1) observations are identical and independent distributed, (2) no perfect multicollinearity, i.e. independent variables are linear independent, (3) expected values of error terms conditioned on independent variables equal zero, (4) homoscedasticity, i.e. variance of error terms on condition of independent variables equal zero. Such assumptions restrict the applications of OLS regression, for if any of the assumptions is invalidated, the estimation will be inaccurate. Time series data often has time-dependent correlations which invalids OLS regression assumption that error terms are independent. If applying OLS regression to autoregressive data, the estimated error terms are often too small and the probability of type I error often increase.

Furthermore, there are more advantages of SEM compared to OLS regression, including more flexible assumptions, particularly allowing interpretation even when multicollinearity exists; the ability to test non-normal data, incomplete data which do not meet OLS regression assumptions; the availability of testing a complex model involving a large number of equations as a whole and obtain global fit measures which provides a summary evaluation of the overall model; the ability to model mediating variables rather than be restricted to additive model as in OLS regression where the dependent is a function of the sum of independents' effects; the ability to model error terms; the attraction of graphical modeling interface.

However, SEM also has disadvantages. One of the biggest concerns is its complexity, for researchers prefer simpler models when the fitting ability is similar. Another concern is requirement of large sample size. Over the years, a number of simulation studies have assessed the influence of variations in sample size of SEM analyses (Boomsma&Hoogland 2001, Hoogland&Boomsma 1998), and there is no recommended minimum sample size broadly applicable in all contexts (Tomarken& Waller, 2005). Several general rules suggest the lowest reasonable sample size is 200 for relatively simple models, and for more complicated models, larger sample size will be recommended. In fact, SEM is powerful when done with adequately large sample size, and the larger, the better. Also, SEM like most statistical models, are typically approximations of reality (Browne&Cudeck 1993, Cudeck & Henley 1991, MacCallum 2003, MacCallum & Austin 2000, Meehl & Waller 2002). One of ways SEM approximates is by omitting variables, but such omissions will lead to misleading measurements and/or cause structures and biased estimates. Although such problem is not unique to SEM, it is a well-known criticism of SEM (Cliff 1983, Freedman 1987).

In summary, although SEM has various advantages over OLS regression, it also has disadvantages and it is not always the case that applying SEM is preferred to OLS regression. If the data satisfies the assumptions of OLS regression well enough, OLS regression should be used rather than SEM. Even though in this case, SEM will produce basically the same result, using a much more complex model will be overkill.

SIMULATIONS TO ILLUSTRATE COMPARISON

Simple simulations on two datasets are used here to illustrate the comparisons made in previous part. The idea of these simulations is we have already known all the variables and the relationships among them, i.e. the "true" coefficients of equations are known, then fitting the model with either SEM or OLS regression, the closer the estimated coefficients to the "true" coefficients, the better a model fits.

The first simulation is designed to have variables with underlying structures and some predictor variables need to be predicted by others as well, thus a simultaneous feature is needed. The purpose is to illustrate when variables in the data set have more complicated relationships and the model requires more sophisticated form, SEM will provides coefficients closer to the "true" ones, thus fits better than OLS regression.

The second simulation is designed to have variables satisfying the assumptions of OLS regression and does not need simultaneous feature. The purpose is to illustrate when OLS regression is suitable, SEM will provide the same estimated coefficients, and thus also fits well. However, if the fitting abilities are the same, a much simpler model, such as OLS regression, is a better choice.

Simulation One:

Here we use three random generated variables, x_1, x_2, x_3, which have an underlying structure described as below, and two random generated variables r_1, r_2, which are independent and acting as error terms. Each of these five variables has 500 observations. We program it in SAS and the code is[1]:

```
do i=1 to 500;
x1=i + normal(12345678);
x2=500 + i + normal(12345678);
x3=1000 + i + normal(12345678);
r1=normal(12345678);
r2=normal(12345678);
```

Then simulate y_1, y_2 under following structure (here y_1 and y_2 appear on both the left and right sides of the equations), in this way, the "true" coefficients are known:

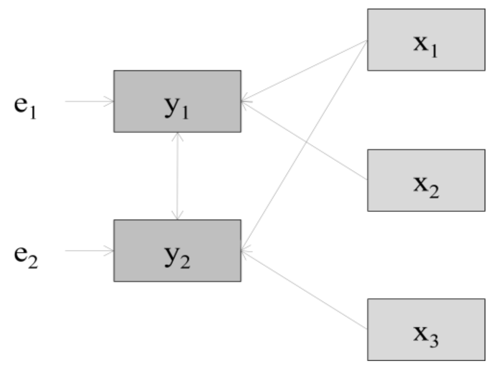
$$y_1 = x_1 + 2*x_2 + y_2 + r_1$$

$$y_2 = 3*x_1 + 4*x_3 + 2*y_1 + r_2$$

The structural model can be graphically presented as in Figure 3.

1normal( ) is the SAS function for generating random N(0, 1) observations with a given seed as argument.

Figure 3: Structural model of simulation one.



A mathematically equivalent reduced form model is:

$$y_1 = -(4*x_1 + 2*x_2 + 4*x_3 + r_1 + r_2)$$

$$y_2 = -(5*x_1 + 4*x_2 + 4*x_3 + 2*r_1 + r_2)$$

Fitting with SEM using PROC CALIS in SAS, the estimated structural model for non-standardized data is:

$$y_1 = 1.0499*x_1 + 1.9166*x_2 + 0.9974*y_2 + 1.0000*e_1$$

$$y_2 = 2.7934*x_1 + 3.7667*x_3 + 1.9560*y_1 + 1.0000*e_2$$

Fitting with OLS regression using PROC REG with separate models for each equation, the result is:

$$y_1 = 0.96770*x_1 + 1.60976*x_2 + 0.96748*y_2$$

$$y_2 = 2.19927*x_1 + 3.19942*x_3 + 1.83988*y_1$$

Comparing the two results, the estimated coefficients from SEM (1.0499, 1.9166, 0.9974; 2.7934, 3.7667, 1.9560) are much closer than the estimated coefficients from

OLS regression (0.9677, 1.60976, 0.96748; 2.19927, 3.19942, 1.83988) to the "true" coefficients in original designed structure model (1, 2, 1; 3, 4, 2), therefore, SEM is proved to fit better in this case.

Simulation Two:

Here we use five random generated variables, x_1, x_2, x_3, x_4, x_5, which are independent from each other, and two generated random variables, r_1, r_2, which are independent and acting as error terms. In this way, OLS regression assumptions are met. Each of these seven variables has 500 observations. We program it in SAS and the code is:

```
do i=1 to 500;
x1=5+1*normal(123456);
x2=4+2*normal(123456);
x3=3+3*normal(123456);
x4=2+4*normal(123456);
x5=1+5*normal(123456);
r1=normal(123456);
r2=normal(123456);
```

Then simulate y_1, y_2 under following structure (here both y_1 and $y_2$appear on only one side of the equations), in this way, the "true" coefficients are known:

$$y_1 = x_1 + 2*x_3 + x_5 + r_1$$

$$y_2 = 2*x_1 + x_2 + 2*x_4 + r_2$$

Fitting with SEM using PROC CALIS in SAS, the estimated structural model for non-standardized data is:

$$y_1 = 0.9648*x_1 + 2.0060*x_3 + 1.0014*x_5 + 1.0000*e_1$$

$$y_2 = 1.9912*x_1 + 1.0033*x_2 + 1.9917*x_4 + 1.0000*e_2$$

Fitting with OLS regression using PROC REG with separate models for each equation, the result is:

$$y_1 = 0.99851*x_1 + 2.00889*x_3 + 1.00201*x_5$$

$$y_2 = 2.01160*x_1 + 1.00777*x_2 + 1.99255*x_4$$

The estimated coefficients from SEM (0.9648, 2.0060, 1.0014; 1.9912, 1.0033, 1.9917) are very similar to the estimated coefficients from OLS regression (0.99851, 2.00889, 1.00201; 2.01160, 1.00777, 1.99255). Both are very close to the "true" coefficients in original designed structure model (1, 2, 1; 2, 1, 2). Therefore, SEM and OLS regression fit equally well in this case. There is no need to apply a much more complex model as SEM, and OLS regression is the preferred choice here.

In summary, SEM, as a more general and powerful model, enjoys a variety of advantages over OLS regression, and can be applied in much wider research fields. However, although SEM can serve all the functions OLS regression can, it is not always the best choice and it cannot replace OLS regression. When it comes to simple situations where OLS regression can fit well, SEM will be overkill due to its complex nature. Also, SEM has a higher requirement for sample size than OLS regression, which is a disadvantage of SEM. It is more difficult to meet required sample size of SEM to avoid it losing power.

REAL DATA MODEL

The model provides a simple analysis of how capital structure decisions are made within the framework of certain risks enterprises are facing. For life insurers, two principal activities are asset investing and product underwriting. Therefore, one of the major categories of enterprise risks is product risk, which aligns with activity of product underwriting. Product risk arises from the nature and volume of products sold, thus different level of risks are associated with different product lines. Although a variety of products are underwritten by life insurers, health insurance contracts and annuity contracts are chosen in this model to represent relatively higher and lower risk respectively.

Risks cannot be observed directly. Thus in empirical studies, proxies are used in order to measure them. However, there are numerous risks aligned with product underwriting, many of which are not neatly distinct but overlapping to some degree. Thus listing all the product risks and analyzing the impact of each of them is neither practical nor necessary. In this sense, proxies for the risks could be viewed as a mixture of different underlying theoretical risks. Thus instead of analyzing separately, a factor analysis is applied here to view product risks as two unobserved risk factors, for health insurance products and annuity products respectively.

In the model, some important control variables are also included in order to represent elements of capital structure decision that are not from product risks for life insurer. In previous studies, size, organization form, and group membership are used and considered to be of the most importance with regards to capital structure decisions. Size is the measure for whether the scale of a life insurer is large or small, which has been generally recognized to be of significant effect on the capital structure decisions. Organization form is the measure for whether a life insurer is a mutual insurance company or a stock company, and under agency theory, risk taking is inversely related to the degree of separation of ownership from management, which implies that managers of mutual insurance companies will take less risk than those of stock companies. Group membership is a dummy variable for whether a life insurer is a member of an affiliated group or not.

The data is for US life insurers taken from the annual statement of life insurers filed with NAIC for the year 1994.

Capital structure decision is represented by capital ratio, which is calculated as the logarithm of the ratio of adjusted book value of capital to total firm invested asset. Adjusted book value of capital is the sum of capital and surplus, asset valuation reserve, voluntary investment reserve, dividends apportioned for payment, dividends not yet apportioned, and the life subsidiaries asset valuation reserve, voluntary investment reserve and dividend liability less property/casualty subsidiaries non-tabular discount. Previous studies argue that economic capital, such as market value of equity, would be a preferable measure to accounting data used here. But since most life insurers are not publicly traded, such data are not available.

The two product risk proxies are calculated as the logarithms of the proportions of insurers' premiums derived from health insurance lines and annuity lines respectively. In the case that an insurer does not write health insurance or annuities and the proportion is zero, the corresponding risk index is set to be a very small value which is less than any of the nonzero data in order to calculating logarithms. As discussed above, proxies here are mixtures of a number of underlying risks, so factor feature of SEM is applied here by constructing two unobservable factors, namely F^{ph} and F^{pan}. Instead of directly using calculated proxies as surrogates for two kinds of product risks, unobservable factors are constructed in this model in order to provide purer proxies than the measured variables.

As discussed above, three control variables are derived from the data set and presented as categorical variables. An insurers with total assets larger or equal to the median of sample total asset is considered to be large, which is indicated by size=1 in the SAS dataset, otherwise, the insurer will be considered to be small and indicated by size=0. In the SAS dataset, type=1 means an insurer is a stock insurer, and type=0 means a mutual insurance company. In similar way, group=1 means an insurer is affiliated with a group of companies, otherwise, group=0.

In order to fit in the model, the insurers with missing values for any of these variables, insurers with nonpositive capital ratio, negative proportions of health insurance or annuity premium are removed from data set. After adjustment, 796 insurers remained for analysis.

SEM is applied due to two main reasons: (1) product risks are estimated by unobservable factors instead of directly calculated proxies, (2) the relationship among capital ratio and two product risk factors, the relationship between product risk factors and product risk proxies are simultaneously estimated.

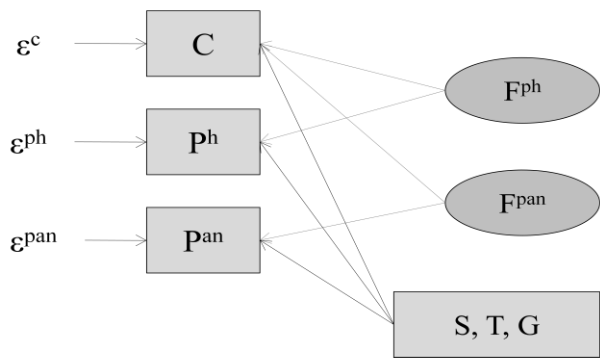
Specifically, the SEM model can be described both analytically and graphically. Analytically, the SEM model can be presented by a set of equations:

$$\begin{cases} C = \alpha_c * S + \beta_c * T + \gamma_c * G + \lambda_{ph} * F^{ph} + \lambda_{pan} * F^{pan} + \varepsilon^c & (1) \\ P^h = \alpha_{ph} * S + \beta_{ph} * T + \gamma_{ph} * G + F^{ph} + \varepsilon^{ph} & (2) \\ P^{an} = \alpha_{pan} * S + \beta_{pan} * T + \gamma_{pan} * G + F^{pan} + \varepsilon^{pan} & (3) \end{cases}$$

In this set of equations, C stands for capital ratio; P^h and P^{an} stand for health insurance and annuity risk proxies respectively; S, T and G are control variables standing for size, organization type and group membership, respectively; F^{ph} and F^{pan} stand for the unobservable factors for health and annuity risk. Equation (1) describes the relationship among capital ratio and risk factors for both health insurance and annuity, controlled by size, organization type and group membership, and the coefficients λ_{ph}, λ_{pan} describe how health insurance / annuity influence the capital structure. Equation (2), (3) describe the relationships between risk proxies and risk factors, which are directly observable risk proxies consist of underlying true risks (i.e. unobservable factors) and other variables. In Equations (2), (3), the coefficients of F^{ph} and F^{pan} are restricted to be 1.0, in order to provide identifiability of parameters without loss of generality. The three ε terms are residual error terms standing for the unexplained part of each equation.

Graphically, the SEM model can be presented in the form of a path diagram, shown in Figure 4.

Figure 4: SEM model for US insurers' risks and capital decision.



As discussed in part II Structural Equation Modeling, manifest variables (C, P^h, P^{an}, S, T, G) are presented in rectangles, whereas latent factors (F^{ph}, F^{pan}) are presented in ovals, error terms without enclosures. Straight, one-headed arrows stand for relationships, for instance, C receives straight arrows from two factors, S, T, G and error term ε^c, which means capital ratio is explained by a function with explanatory variables F^{ph}, F^{pan}, S, T, G, and error term ε^c. On the other hand, health insurance risk factor F^{ph} emanates two straight arrows to C and p^h, which means F^{ph} is used to explain both C and p^h in two equations in the model.

In SAS, the SEM model can be coded with procedure CALIS. CALIS is designed to analyze models in which hypothesized relationships among variables specified in terms of the variances and covariances of variables and fit to an observed covariance matrix.

For this model, the SAS code is written as:

```
PROC CALIS  DATA = real_data;
LINEQS
cap_ratio = alpha_c size + beta_c type + gamma_c group + lambda_phF_ph + lambda_panF_pan + e1,
p_health  =alpha_ph size + beta_ph type + gamma_ph group +  F_ph + e2,
p_annuity = alpha_pan size + beta_pan type + gamma_pan group +  F_pan + e3;
STD
e1-e3 = vare1-vare3,
size = 0.500315,
type = 0.322918,
group = 0.441651,
F_ph = 1,
F_pan = 1;
COV
F_ph F_pan = cov,
size type      = -0.03141,
size group    = 0.06972,
type group   = 0.01267,
```

```
F_ph size     = 0,
F_pan size   = 0,
F_ph type    = 0,
F_pantype  =0,
F_phgroup  =0,
F_pan group = 0;
RUN;
```

In the code, LINEQS is one of CALIS's model specification statements, which is to define the model type and specify the main model. The endogenous variables used on the left side can be manifest variables (with names that must be defined by the input data set) or latent variables (which must have names starting with F). The variables used on the right side can be manifest variables, latent variables (with names that must start with an F), or error variables (which must have names starting with an E or D). The coefficients to be estimated are indicated by names. If no name is used, the coefficient is constant, either equal to a specified number or, if no number is used, equal to 1. The equation here is directly from the SEM equations (1), (2), (3). The other main model specification statements in CALIS are RAM, FACTOR, and COSAN, which give slightly different statements from LINEQS.

STD statement specifies which variance to be estimated. Here e1-e3 = vare1-vare3 means we want to estimate the standard deviation of error terms e1, e2, e3, and name them as vare1, vare2, vare3. F_ph = 1, F_pan = 1 means we fix the standard deviation of latent factors to be 1, the reason for which is to set scale to latent factors. The standard deviations of size, type and group are also fixed at their sample standard deviations (calculated from the data set), the reason for which is in theory, when the sample size is large, the sample standard deviation is unbiased estimate of population standard deviation. Also, to prove fixing these standard deviations is statistically reasonable, another model with these three standard deviations not fixed but to be estimated is run, i.e. size = 0.500315, type = 0.322918, group = 0.441651, is replaced by: size = stdev_size, type = stdev_type, group = stdev_group, all the rest are the same, and the result shows the coefficients of cap ratio functions is very similar to the original model, only the R square for capital ratio is 0.7499, a little less than that from original model which is 0.8204.

COV statement specifies which covariance to be estimated. Here F_ph F_pan = cov means we want to estimate the covariance of the two latent factors F^{ph} and $F^{pan}$and name it as cov. And size type = -0.03141, size group = 0.06972, type group = 0.01267 means we fix the covariances between size, type and group at their sample covariance (calculated from the data set), the reason for which is in theory, when sample size is large, sample covariance is unbiased estimate of population covariance. Also, to prove fixing these covariances is statistically reasonable, another model with these covariances not fixed but to be estimated is run, i.e. size type = -0.03141, size group = 0.06972, type group = 0.01267, is replaced by: size type = covst, size group = covsg, type group = covtg, all the rest are the same, and the result shows the coefficients of cap ratio functions is very similar to the original model, only the R square for capital ratio is 0.8178, a little less than that from original model which is 0.8204. The covariances between factors and control variables are all fixed at 0, for in this model, size, type and group only serves as control variables and it assumes that the two latent factors underlie only the part of capital ratio, two product risk proxies, but not explained by the control variables.

Result:

The output of CALIS procedure includes both non-standardized and standardized results for linear equations. Table 1 displays the non-standardized estimates of coefficients and corresponding t values for variable capital_ratio.

Table 1: Result of CALIS procedure.

	size	type	group	F_ph	F_pan
coefficient	-1.1248	0.0475	0.0723	0.1366	-0.0837
t value	-31.2172	1.0703	1.8893	12.7684	-9.4853

The coefficients describe the how capital_ratio is modeled to change when the value of either of the two factors changes controlled for all the rest variables. For instance, the estimated coefficient of health insurance risk factor, F_ph, equals to 0.1366, which means for one unit change of the value of the factor, capital ratio will change 0.1366 units, ceteris paribus. In similar ways, the effect of whether the insurer is large or small, its governance structure, group membership, as well as the annuity risk factor can be analyzed.

CALIS procedure provides a variety of model fitting indices, including absolute indices, parsimony indices and incremental indices, presented in the Fit Summary part of the output, as well as the R-squares for each equation in Squared Multiple Correlations part. Table 2 displays some of the information about model fitting.

Table 2: Fitting indices of CALIS procedure.

Pr> Chi-Square	GFI	RMSR	R-Square for capital_ratio
< .001	0.6953	1.8829	0.8204

Although Pr> Chi-Square is less than 0.001, which may be a good sign for model fitting, GFI and RMSR both indicated the fitting is not so good. However, the reason to build this model here is to illustrate how to use SEM and whether SEM is better than OLS regression facing particular situations, but not to build a preferred model to actually solve a real economic problem. Therefore, R-Square is used here, for it is provided in both models and then can be used to compare CALIS output with OLS regression (the comparison will be in the following OLS regression model part).

OLS Regression

Using the same data set as in SEM model. Since there is no latent factor feature available in OLS regression model, the proxies of health insurance and annuity are used directly in this model, and the linear equation for capital ratio is built as:

$$C = \alpha_0 + \alpha_1 * S + \alpha_2 * T + \alpha_3 * G + \alpha_4 * P^h + \alpha_5 * P^{an} + \varepsilon$$

The interaction terms of size, type and group are omitted from the equation because they are tested to be statistically insignificant. In SAS, procedure REG is used to estimate OLS regression model. The code will be:

```
PROC REG DATA=real_data;
MODEL cap_ratio = p_healthp_annuity  size  type  group;
RUN;
```

The estimated coefficients and responding t values for capital ratio variable is displayed in table 3.

Table 3: Result of REG procedure.

	intercept	size	type	group	p_health	p_annuity
coefficient	-1.86455	-0.68209	0.06027	0.09317	0.00492	-0.08434
t value	-21.25	-13.18	0.87	1.82	1.02	-15.07

The estimated coefficients are not exactly the same but comparable with those given by SEM. Both models show that capital ratio will increase as health insurance risk increases, while decrease as annuity risk increase. The direction of change for the three control variables is also consistent.

In order to compare the models, model fitting information is needed, as shown in table 4.

Table 4: Fitting indices of REG procedure.

F value	Pr> F	R-Square	Adj R-Sq
206.94	<.0001	0.5671	0.5643

F value shows the explanatory variables overall do have effects on capital_ratio, while the R-square and Adjusted R-square are only 0.5671 and 0.5643, which are less than the R-square from the SEM model, which is 0.8204. Despite the fact that neither of these two models fits very well, SEM does fit better than OLS regression, which could be evidence for the discussions that in certain situations, SEM will provide a better estimation than OLS regression does.

SUMMARY

This report gives a brief description of structural equation modeling (SEM), focusing mainly on the differences of SEM from ordinary least square (OLS) regression. Since SEM is a comprehensive model which combines characteristics of a variety of different modeling methods, we assume that it is able to deal with more complex situations than many other models, such as multiple regression, path analysis, factor analysis, time series analysis, analysis of covariance, which in fact can be viewed as special cases of SEM, especially when compared with OLS regression, which is a very simple model with relatively strict assumptions. Several advantages of SEM over OLS regression are introduced in the report, and three of them are discussed: the ability of a variable to be both predictor and response in the same model; the ability to include unobserved latent factors; and the ability to deal with time series data.

Although SEM enjoys a variety of advantages over OLS regression, it is not always the better choice. Since OLS regression can be viewed as one special case of SEM, SEM is able to deal with all the situations where OLS regression fits and provides the same results as OLS regression. However, SEM is a much more complex model, while in research, simpler models are always preferable. Therefore, applying SEM to simple situations is neither necessary nor preferable, and may be viewed as overkill.

Simulation cases on two datasets are used in the report to illustrate whether SEM or OLS regression is a preferable choice in different situations. In the first simulation case, data are random generated with an underlying structure, and variables are required

to be on both sides of the equations. In this case, the coefficients provided by SEM are much closer to the "true" coefficients we set than those by OLS regression. In the second simulation case, data are randomly generated to satisfy the assumptions of OLS regression. In this case, the coefficients provided by SEM are almost the same with those by OLS regression, and both are very close to the "true" coefficients we set, which means both SEM and OLS regression fit the model well. However, when multiple models all fit well, researchers will prefer the simpler ones, so incurring the complexity of SEM here is unnecessary and OLS regression is the better choice. In conclusion, SEM and OLS regression are modeling different situations. If simultaneous estimation or latent factors, for instance, are supposed to be involved, SEM may be appropriate while OLS regression is not. On the other hand, if there is only one equation and no latent factors, SEM is unnecessary and OLS regression is appropriate.

A further example is provided by a real data model. The model use information on US life insurers in the year 1994. We assume that the capital structure decisions made by life insurers are influenced by the risks aligned with their product underwriting activities. Those risks can be called product risks. In our model, we focus on two groups of product: health insurance and annuity, which are of relatively higher risk and lower risk respectively. Since risks cannot be observed directly, one way to represent the risks is to use proxies. In our data, such risk proxies are proportion of health insurance premium to total premium and proportion of annuity premium to total premium, which can be calculated from the data set. While proxies are actually mixture of pure " true" risks and other variables. The "true" risks cannot be measured directly, so applying the concept of latent factors can be a good way to represent them. In our model, we use two latent factors, which can be referred to as health insurance risk factor and annuity risk factor, to represent such purer risks implied by risk proxies. Here in the model, capital structure is presented by the insurer's capital ratio, and three control variables, size, type, group, are included for insurer' size, governance structure and group membership. CALIS procedure in SAS is used by SEM, where capital ratio is regressed on the two risk factors and the three control variables, while the two risk proxies are regressed on the corresponding risk factors as well as the three control variables respectively. REG procedure in SAS is used by OLS regression, where capital ratio is regressed directly on the two risk proxies and the three control variables. Because the purpose of building this model is to compare SEM and OLS regression, but not to explain real economic phenomena, some useful variables are omitted in order to make tour model not too complicated. The overall fitting of both SEM and OLS regression are not very good. However, the results can still serve our purpose to compare the relative fitness of SEM and OLS regression. We use the two R-squares of estimated linear equations of capital ratio. From the outputs, the R-square provided by SEM is 0.8204, while the R-square provided by OLS regression is only 0.5671, which concludes that SEM fits relatively better than OLS regression in our situation where latent factors is included and simultaneous estimation is needed.

APPENDIX

SAS code for the simulations on two datasets in this report is:

Simulation One:

```
data simu1;/* generating random data */
do i=1 to 500;
x1=i + normal(12345678);
x2=500 + i + normal(12345678);
x3=1000 + i + normal(12345678);
r1=normal(12345678);
r2=normal(12345678);
y1 = - (4*x1 + 2*x2 + 4*x3 + r1 + r2);
y2 = - (5*x1 + 4*x2 + 4*x3 + 2*r1 + r2);
output;
end;
run;

/* SEM for simulation one */
proc calis data=simu1;
lineqs y1=a1 x1 +a2 x2 + a3 y2 + e1,
       y2=b1 x1 +b2 x3 + b3 y1 + e2;
std e1=vare1, e2=vare2;
run;

/* OLS regression for simulation two */
proc reg data=simu;
model y1=x1 x2 y2 /NOINT;
model y2=x1 x3 y1/NOINT;
run;
quit;
```

Simulation Two:

```
data simu2; /* generating random data */
do i=1 to 500;
x1=5+1*normal(123456);
x2=4+2*normal(123456);
x3=3+3*normal(123456);
x4=2+4*normal(123456);
x5=1+5*normal(123456);
r1=normal(123456);
r2=normal(123456);
y1=x1       +2*x3       +x5+ r1;
y2=2*x1 +x2       +2*x4+ r2;
output;
end;
```

```
run;

/* SEM for simulation two */
proc calis data=simu2;
lineqs y1=a1 x1 +a2 x3 +a3 x5 +e1,
       y2=b1 x1 +b2 x2 +b3 x4 +e2;
std e1=vare1, e2=vare2;
run;

/* OLS regression for simulation two */
proc reg data=simu2;
model y1=x1 x3 x5/NOINT;
model y2=x1 x2 x4/NOINT;
run; quit;
```

Real data model:

The data for this report are taken from the annual statements of life insurers filed with the NAIC for the year 1994.

For capital structure, this report uses the ratio of adjusted book value of capital to total firm invested assets. The adjusted capital formula is the sum of Capital and Surplus, Asset Valuation Reserve (AVR), Voluntary Investment Reserve, Dividends Apportioned for Payment, Dividends not yet Apportioned, and the Life Subsidiaries AVR, Voluntary Investment Reserves and Dividend Liability less Property/Casualty Subsidiaries Non-Tabular Discount. (Source: page LR022 of the *1996 Life NAIC Risk Based Capital Report Including Overview and Instructions for Companies*.) For product risk proxies, this report uses the proportions of firm premium derived from health insurance and annuities, which represents higher and lower product risk respectively. For the size measure, this report uses the medium of total assets. The insurer with total asset larger than the medium is given size=1, otherwise size=0. For the type measure, the insurer which is a stock insurer is given type=1, and a mutual insurance company is given type=0. For the group measure, the insurer which is affiliated with a group of companies is given group=1, otherwise, group=0.

The insurers with missing values for any of the six variables, insurers with nonpositive capital ratio, negative proportion of health insurance or annuity premium are removed from the data set. After adjustment, 796 insurers remained in the data sat.

Then we take the logarithms of the capital ratio and two product risk proxies. Some insurers do not involve in either health insurance or annuity and thus the responding proportion of premium equals zero. In order to calculate the logarithm, we set a very small value which is less than any of the nonzero data before taking the logarithm. Specifically, for health insurance risk product, the minimum value is 3.40897E-07, and we replace the zeros with 3.40000E-07; for annuity risk product, the minimum value is 3.80630E-06, and we replace the zeros with 3.80000E-06.

Below is the full SAS code:

```
data logdata;
set data;
        cap_ratio= log(cap_ratio);
        p_health = log(p_health);
        p_annuity = log(p_annuity);
run;

/* SEM for real data model */
proc calis data=logdata;
lineqs
       cap_ratio = alpha_c size + beta_c type + gamma_c group +
lambda_ph F_ph + lambda_pan F_pan + e1,
         p_health  = alpha_ph size + beta_ph type + gamma_ph
group +  F_ph + e2,
         p_annuity = alpha_pan size + beta_pan type + gamma_pan
group +  F_pan + e3;
std e1-e3 = vare1-vare3,
```

```
    size = 0.500315,
    type = 0.322918,
    group = 0.441651,
    F_ph = 1,
    F_pan = 1;
cov F_ph F_pan = cov,
    size type = -0.03141,
    size group = 0.06972,
    type group = 0.01267,
    F_ph size = 0,
    F_pan size = 0,
    F_ph type = 0,
    F_pan type = 0,
    F_ph group = 0,
    F_pan group = 0;
run;

/* OLS regression for real data model */
proc reg data=logdata;
model cap_ratio = size  type group p_health p_annuity;
run; quit;
```

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.

Austin, James T., & Calderón, Robert. F. 1996. "Theoretical and technical contributions to structural equation modeling: An updated annotated bibliography, structural equation modeling." *A Multidisciplinary Journal* 3:2, 105-175.

Babbel, D. F., & A. M. Santomero, 1999. "An analysis of the financial risk management process used by life insurers."*Changes in the Life Insurance Industry: Efficiency, Technology and Risk Management*, Eds: Cummins, J. D. and A. M. Santomero, Boston: Kluwer Academy Publishers.

Baldwin, B. 1989. "A primer in the use and interpretation of structural equation models."*Measurement and Evaluation in Counseling and Development* **22**: 100-112.

Baranoff, E. G., Baranoff, D., & Sager, T. W. 2000. "Nonuniform regulatory treatment of broker distribution systems: An impact analysis for life insurers." *Journal of Insurance Regulation* **19(1)**: 94-129.

Baranoff,,Etti. G., Papadopoulos, Savas., & Sager, Thomas. W. 2007. "Capital and risk revisited: A structural equation model approach for life insurers." *The Journal of Risk and Insurance* **74(3)**: 653-681.

Baranoff, E. G., & Sager, T. W. 2002. "The relations among asset risk, product risk, and capital in the life insurance industry." *Journal of Banking & Finance* **26**: 1181-1197.

Baranoff, E. G., & Sager, T. W. 2003. "The relations among organizational and distribution forms and capital and asset risk structures in the life insurance industry." *The Journal of Risk and Insurance* **70(3)**: 375-400.

Baranoff, E. G., Sager, T. W., & Witt, R. C. 1999. "Industry segmentation and predictor motifs for solvency analysis of the life/health insurance industry."*Journal of Risk and Insurance* **66**: 99-123.

Barrett, P. 2007. "Structural equation modeling: Adjudging model fit."*Personality and Individual Differences* **42(5)**: 815-24.

Bentler, P. M. 1983. "Some contributions to efficient statistics in structural models: specification and estimation of moment structures." *Psychometrika* **48**: 493-517.

Bentler, P. M. 1990. "Comparative fit indexes in structural models." *Psychological Bulletin* **107(2):** 238-46.

Bentler, P. M., & Bonett, D. G. 1980. "Significance tests and goodness of fit in the analysis of covariance structures."*Psychological Bulletin* **88**: 588-606.

Bentler, P. M., & Chou, C. 1987. "Practical issues in structural modeling."*Sociological Methods & Research* **16**: 78-117.

Bentler, P. M., & 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 & Sons, Inc.

Browne, M. W. 1987. "Robustness of statistical inference in factor analysis and related models."*Biometrika* **74**: 375-84.

Browne, M. W., & Shapiro, A. 1988. "Robustness of normal theory methods in the analysis of linear latent variatemodels."*British Journal of Mathematical and Statistical Psychology* **41:**193-208.

Cliff, N. 1983. "Some cautions concerning the application of causal modeling methods." *Multiv.Behav. Res.* **18**: 81-105.

Crowley, Susan L, & Fan, Xiao. 1997. "Structural equation modeling: basic concepts and applications in personality assessment research." *Journal of Personality Assessment* **68(3)**: 508-53.

Cummins, J. D., & Harrington, S. E. 1988. "The relationship between risk and return: evidence for property-liability insurance stocks." *Journal of Risk and Insurance* **55**: 15-31.

Drezner, Z., & Drezner, T. 1995. "On a binomial test of model fit."*Structural Equation Modeling* **2**:144-145.

Everitt, B. S. 1984. *An Introduction to Latent Variable Methods*. London: Chapman & Hall.

Fan, X., Thompson, B., & Wang, L. 1999. "Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes." *Structural Equation Modeling***6 (1)**: 56-83.

Fornell, C., & Yi, Y. 1992. "Assumptions of the two-step approach: reply to Anderson and Gerbing." *Sociological Methods & Research* **20**:334-339.

Haagen, K..Bartholomew., D. J., & Deistler, M. (Eds.). 1993. *Statistical modelling and latentvariables.* Amsterdam: North-Holland.

Hooper, D., Coughlan, J., & Mullen, M. 2008. "Structural equation modelling: guidelines for determining model fit electronic." *Journal of Business Research Methods* **6(1)**: 53-60.

Hoyle, R. H. (Ed.). 1995. *Structural equation modeling: concepts, issues, and applications*. Thousand Oaks, CA: Sage.

Hsu, L. M. 1995. "On Fraas and Newman's goodness of fittest for structural equation models."*Structural Equation Modeling* **2**: 152-154.

Hu, L. T., & Bentler, P. M. 1999. "Cutoff criteria for fit Indexes in covariance structure analysis: conventional criteria versus new alternatives." *Structural Equation Modeling* **6(1)**: 1-55.

Joreskog, K. G., & Sorbom, D. 1979.*Advances in Factor Analysis and Structural Equation Modeling*. Cambridge MA: Abt Books.

Kaplan, D. 1989. "Model modification in covariance structure analysis: application of the expected parameter change statistic." *Multivariate Behavioral Research* **24**:285-305.

Kelloway, E. K. 1995. "Structural equation modelling in perspective."*Journal of Organizational Behavior* **16**: 215-224.

Kenny, D. A., & McCoach, D. B. 2003. "Effect of the number of variables on measures of fit in structural equation modeling."*Structural Equation Modeling* **10(3)**: 333-51.

Koziol, J. A. 1987. "Review of 'goodness-of-fit techniques' by D'Agostino& Stephens." *Journal of Educational Statistics* **12**:412-416.

La Du, T. J., & Tanaka, J. S. 1989. "Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation modelling." *Journal of Applied Psychology* **74**: 625-636.

Lei, Pui - Wa.,& Wu, Qiong. 2007. "Introduction to structural equation modeling: issues and practical considerations." *Educational Measurement: Issues and Practice* **26**: 33-43.

Loehlin, J. C. 1987. *Latent Variable Models, An Introduction to Factor, Path, and Structural Analysis*. Hillsdale, NJ: L. Erlbaum Associates.

Martin, N. G., & Eaves, L. G. 1977."The genetical analysis of covariance structure."*Heredity* **138**: 79-95.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. 1996. "Power analysis and determination of sample size for covariance structure modeling." *Psychological Methods* **1**: 130-149.

McDonald, R., & Marsh, H. W. 1990."Choosing a multivariate model: noncentrality and goodness of fit."*Psychological Bulletin* **107**: 247-255.

McIntosh, C. 2006. "Rethinking fit assessment in structural equation modeling: a commentary and elaboration on Barrett (2007)." *Personality and Individual Differences* **42(5)**: 859-67.

McQuitty, S. 2004. "Statistical power and structural equation models in business research." *Journal of Business Research* **57(2)**: 175-83.

Miles, J., & Shevlin, M. 1998. "Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis." *Personality and Individual Differences* **25**: 85-90.

Muthén, B. 1983."Latent variable structural equation modelling with categorical data."*Journal of Econometrics* **22**:43-65.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. 1989."Evaluation of goodness-of-fit indices for structural equation models."*Psychological Bulletin* **105**: 430-445.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.ISBN0-521-77362-8.

Raykov, T., & Penev, S. 1995."On an approach to structural model-fit assessment."*Structural Equation Modeling* **2**:146-151.

Reisinger, Y., & Mavondo, F. 2006. "Structural equation modeling: critical issues and new developments." *Journal of Travel and Tourism Marketing* **21(4)**: 41-71.

Rigdon, E. E. 1995. "A necessary and sufficient identification rule for structural models estimated in practice." *Multivariate BehaviorResearch* **30**: 359-383.

Sclove, L. S. 1987. "Application of model-selection criteria to some problems in multivariate analysis."*Psychometrika* **52**: 333-343.

Simon, Herbert. 1953. "Causal ordering and identifiability." In Hood, W.C.; Koopmans, T.C. *Studies in Econometric Method*. New York: Wiley. pp. 49-74.

Thurstone, L. L. 1927. "A law of comparative judgment."*Psychological Review* **34**:273-286.

Tomarken, Andrew J., & Waller, Niels G. 2005. "Structural equation modeling: strengths, limitations, and misconceptions." *Annual Review Of Clinical Psychology* **Vol. 1**: 31-65.

Tucker, L. R., & Lewis, C. 1973. "A reliability coefficient for maximum likelihood factor analysis." *Psychometrika* **38**: 1-10.