结构方程模型的理论、应用以及若干误区

谢益辉

ABSTRACT

Early in 1970s, the theory of *Structural Equation Model* (SEM) was brought forward by Jöreskog, K. G (1973) and Duncan, O. D. (1975), etc. Some thirty years later, SEM also becomes popular in China due to its advantages in the applications of many areas such as sociology, psychology, etc, however, there are few interpretations about this theory from a view of mathematical statistics, which directly leads to its widely applications without in-depth research for its theoretical background. This paper mainly deals with the theories of SEM, including theoretical presuppositions and assumptions, methods for estimating parameters in SEM and interpretations for the output of SEM.

关键词：结构方程模型；潜变量；最大似然估计；多元正态分布；LISREL；AMOS

早在二十世纪六七十年代，结构方程模型（Structural Equation Model[1]）在国外已经初见端倪，尤其是在社会学和心理学发展的驱动下[2]，由于传统的统计方法难以适应其发展，因而一些社会学家们摆脱传统的统计模型的桎梏，摒弃以往统计学在完美而苛刻的假设条件下推导出来的模型，开始寻找更适合处理社会科学中复杂关系的模型。

在世纪之交，JASA（*Journal of the American Statistical Association*）刊登了一系列短文，概括了统计学中的某一领域在即将过去的一个世纪所取得的进展；其中 Adrian. E. Raftery[3]介绍了统计学在社会学中 50 年的应用历程（参见[1]），该文描述了统计方法在社会科学研究领域中的不断更新进步，不乏结构方程模型的产生和发展；Blau 和 Duncan（1967）的有广泛影响力的著作《美国的职业结构》（The American Occupational Structure）堪称开因果关系模型（Causal Model）以及路径分析模型（Path Analysis）应用之先河（参见[2]），而后来的瑞典统计学家、心理测量学家 Jöreskog, K. G（1973）进一步引入潜变量（Latent Variable）的概念，结构方程模型至此已初具雏形[4]；接下来经过数位社会学家和统计学家的延伸和拓展，如今结构方程模型理论已经比较成熟，并能适用于各种不同的数据情况，例如分类数据（Categorical Data）、纵贯数据（Longitudinal Data）和多层次数据（Multilevel Data）等等。然而国内对于结构方程模型的理论研究偏少，更多在于其应用，所以往往容易忽视理论前提和假设，造成模型的滥用；从国内诸多文献中不难发现，研究者们对于 SEM 大多都持推崇态度，所介绍的多为 SEM 的优点；所以本文从数理统计角度入手，解释 SEM 的理论框架，其目的也在于澄清一些应用方面的误区。

---

1 简记为“SEM”，下文中不再另加说明。

2 为何 SEM 经过几十年的时间才在中国开始流行，我认为与中国国内学科发展的重点和趋势是有必然联系的；六七十年代，国外的社会学和心理学等学科都面临着测量和因果关系这两个难题，可以说，“潜变量”这个概念的提出，直接催生了 SEM 的发展；相比之下，国内对于人文社会科学一直以来都不够重视，至于该领域的定量研究方法，更显得薄弱。

3 华盛顿大学统计学和社会学教授，统计学和社会科学研究中心主任。

4 Bollen（1989）称：潜变量之间关系的度量最早起源于 Sewall Wright 在 1916 年的著作（参见[15]）

一、结构方程模型（SEM）概述

表格 1　　结构方程模型中的符号说明

符号	含义
N	观测样本量
m	内生潜变量个数
n	外生潜变量个数
p	内生观测变量数（因变量）
q	外生观测变量数（自变量）
$\underset{(m\times 1)}{\boldsymbol{\eta}_i}$	第 i 个内生潜变量
$\underset{(n\times 1)}{\boldsymbol{\xi}_i}$	第 i 个外生潜变量
$\underset{(m\times 1)}{\boldsymbol{\zeta}_i}$	结构模型的第 i 个随机扰动项（误差项）
$\underset{(m\times m)}{\mathbf{B}}$	内生潜变量之间的结构系数矩阵
$\underset{(m\times n)}{\mathbf{\Gamma}}$	内生潜变量与外生潜变量之间的结构系数矩阵[5]
$\underset{(p\times 1)}{\boldsymbol{y}_i}$	第 i 个内生观测变量
$\underset{(q\times 1)}{\boldsymbol{x}_i}$	第 i 个外生观测变量
$\underset{(p\times 1)}{\boldsymbol{\varepsilon}_i}$	内生变量的第 i 个测量误差项
$\underset{(q\times 1)}{\boldsymbol{\delta}_i}$	外生变量的第 i 个测量误差项
$\left.\begin{matrix}\underset{(p\times m)}{\mathbf{\Lambda}_y} \\ \underset{(q\times n)}{\mathbf{\Lambda}_x}\end{matrix}\right\}$	观测变量对潜变量的因子载荷阵——Λ_y 表示 Y 对 η 的因子载荷阵，Λ_x 表示 X 对 ξ 的因子载荷阵
$\underset{(n\times n)}{\mathbf{\Phi}}$	外生潜变量 ξ 的协差阵
$\underset{(m\times m)}{\mathbf{\Psi}}$	结构模型误差项 ζ 的协差阵
$\left.\begin{matrix}\underset{(p\times p)}{\mathbf{\Theta}_\varepsilon} \\ \underset{(q\times q)}{\mathbf{\Theta}_\delta}\end{matrix}\right\}$	测量误差项的协差阵
$\underset{((p+q)\times(p+q))}{\mathbf{\Sigma}}$	所有观测变量的协差阵

[5] 有研究者称这两个矩阵为因子载荷阵（参见[8]），我认为不妥，所谓“因子载荷阵”，应该是在因子分析中的特定称谓，结构模型中并不涉及到因子分析。

一般结构方程模型包括两部分模型，即测量模型和结构模型，由三个矩阵方程式组成，而模型的识别则是模型是否能求出唯一解的关键。

1. 模型的形式及假定

测量模型（Measurement Model）：反映观测变量与潜变量之间的线性关系，分别包括内生的和外生的观测变量与潜变量[6]；测量模型由下面二式给出：

$$\boldsymbol{x} = \Lambda_{\mathrm{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \quad -------------(1)$$

$$\boldsymbol{y} = \Lambda_{\mathrm{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad -------------(2)$$

结构模型（Structural Model）：反映外生潜变量和内生潜变量之间的（因果）关系；由下式给出：

$$\boldsymbol{\eta} = \mathrm{B}\boldsymbol{\eta} + \Gamma\boldsymbol{\xi} + \boldsymbol{\zeta} \quad -----------(3)$$

模型假定：

（1）误差项的期望为零：$E(\boldsymbol{\delta}) = \mathbf{0}$，$E(\boldsymbol{\varepsilon}) = \mathbf{0}$，$E(\boldsymbol{\zeta}) = \mathbf{0}$

（2）误差项与潜变量之间互不相关：$E(\boldsymbol{\xi}\boldsymbol{\delta}') = \mathbf{0}$，$E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') = \mathbf{0}$，$E(\boldsymbol{\eta}\boldsymbol{\zeta}') = \mathbf{0}$，$E(\boldsymbol{\xi}\boldsymbol{\zeta}') = \mathbf{0}$，$E(\boldsymbol{\eta}\boldsymbol{\delta}') = \mathbf{0}$，$E(\boldsymbol{\xi}\boldsymbol{\varepsilon}') = \mathbf{0}$

（3）误差项之间互不相关：$E(\boldsymbol{\delta}\boldsymbol{\varepsilon}') = \mathbf{0}$

（4）B 的对角线元素全为零，$(\boldsymbol{I} - \mathrm{B})^{-1}$ 存在

2. 模型识别条件

识别工作主要是考虑模型中每一个自由参数能否由观测数据求得惟一解作为估计。对于某一个自由参数，如果不能将这一参数以样本方差协方差的代数函数表达，这个参数就不能识别。同样的原则适用于更复杂的结构方程模型。要是一个未知参数至少可以由观测变量的方差协方差矩阵（一般用 S 表示）中的一个或多个元素的代数函数来表达，就称这个参数可识别。对于结构方程模型，并没有一套简单的充要条件作为参数识别手段。然而，有一个简单的必要条件可以用来辅助判断模型是否可识别，即：数据点的数目不能少于自由参数的数目；数据点的数目就是观测变量的方差和协方差的数目[7]。它等于 $(p+q)(p+q+1)/2$，其中 p 与 q 的含义参见表格 1。自由参数的数目指待定的因子载荷、路径系数、潜变量和误差项各自方差的总数。要是数据点比自由参数多，这一模型即为过度识别；如果数据点比自由参数少，这一模型就是不能识别的，其参数也无法估计。因为，未知项多于已知项时，估计便不可能进行。

此外，在测量模型中，每个潜变量可能对应几个观测变量，那么通常将其中一个观测变量的系数设为 1，这相当于给了潜变量一个度量尺度；而对于只对应一个观测变量的潜变量，

---

[6] 观测变量（Observed Variable）即可从试验或调查中直接观测得到的变量，又可称为显变量（Manifest Variable）；潜变量（Latent Variable）即不可直接观测的变量；内生变量（Endogenous Variable）是由模型系统内其它变量所决定的，常为因变量，而外生变量（Exogenous Variable）则是由模型以外的因素决定，对模型系统产生影响，但不受模型系统影响，常为自变量。

[7] 协差阵 S 为对称阵，只有对角线元素及其上三角（或下三角）矩阵元素提供信息

我们将相应的误差方差设为零。

二、模型参数估计方法

一般说来，SEM 中有八个带估计的参数矩阵：$\boldsymbol{\Lambda}_x$，$\boldsymbol{\Lambda}_y$，$\boldsymbol{\Gamma}$，$\mathbf{B}$，$\boldsymbol{\Phi}$，$\boldsymbol{\Psi}$，$\boldsymbol{\Theta}_\delta$和$\boldsymbol{\Theta}_\varepsilon$，其中前四个包含载荷和结构系数的矩阵常常为关注的重点。

1. 模型拟合准则

结构方程模型对数据拟合好坏的准则在于比较所有观测样本协差阵S与其理论协差阵Σ的差异大小，其中，所有观测变量（令其为$\boldsymbol{z}=(\boldsymbol{y}', \boldsymbol{x}')'$）的理论协差阵推导如下（其中令$\boldsymbol{A}=(\boldsymbol{I}-\mathrm{B})^{-1}$）：

$$\boldsymbol{\eta}=\mathrm{B}\boldsymbol{\eta}+\Gamma\boldsymbol{\xi}+\boldsymbol{\zeta}\Rightarrow\boldsymbol{\eta}=(\boldsymbol{I}-\mathrm{B})^{-1}(\Gamma\boldsymbol{\xi}+\boldsymbol{\zeta})=\boldsymbol{A}(\Gamma\boldsymbol{\xi}+\boldsymbol{\zeta})$$

$$\left.\begin{aligned}
\mathrm{Var}(\boldsymbol{x})&=\mathrm{Var}(\Lambda_x\boldsymbol{\xi}+\boldsymbol{\delta})\\
&=\mathrm{Var}(\Lambda_x\boldsymbol{\xi})+\mathrm{Cov}(\Lambda_x\boldsymbol{\xi},\ \boldsymbol{\delta})+\mathrm{Var}(\boldsymbol{\delta})\\
&=\Lambda_x\boldsymbol{\Phi}\Lambda_x^{'}+\boldsymbol{\Phi}_\delta\\
\mathrm{Var}(\boldsymbol{y})&=\mathrm{Var}(\Lambda_y\boldsymbol{\eta}+\boldsymbol{\varepsilon})\\
&=\mathrm{Var}(\Lambda_y\boldsymbol{\eta})+\mathrm{Cov}(\Lambda_y\boldsymbol{\eta},\ \boldsymbol{\varepsilon})+\mathrm{Var}(\boldsymbol{\varepsilon})\\
&=\Lambda_y\mathrm{Var}[\boldsymbol{A}(\Gamma\boldsymbol{\xi}+\boldsymbol{\zeta})]\Lambda_y^{'}+\boldsymbol{\Phi}_\varepsilon\\
&=\Lambda_y\boldsymbol{A}(\boldsymbol{\Gamma\Phi\Gamma}'+\boldsymbol{\Psi})\boldsymbol{A}'\Lambda_y^{'}+\boldsymbol{\Phi}_\varepsilon\\
\mathrm{Cov}(\boldsymbol{x},\ \boldsymbol{y})&=\mathrm{Cov}(\Lambda_x\boldsymbol{\xi}+\boldsymbol{\delta},\ \Lambda_y\boldsymbol{\eta}+\boldsymbol{\varepsilon})\\
&=\mathrm{Cov}(\Lambda_x\boldsymbol{\xi},\ \Lambda_y\boldsymbol{\eta})+\mathrm{Cov}(\Lambda_x\boldsymbol{\xi},\ \boldsymbol{\varepsilon})+\mathrm{Cov}(\boldsymbol{\delta},\ \Lambda_y\boldsymbol{\eta})+\mathrm{Cov}(\boldsymbol{\delta},\ \boldsymbol{\varepsilon})\\
&=\Lambda_x\boldsymbol{\Phi\Gamma}'\boldsymbol{A}'\Lambda_y^{'}
\end{aligned}\right\}\Rightarrow$$

$$\begin{aligned}
\boldsymbol{\Sigma}&=\mathrm{Var}(\boldsymbol{z})=\mathrm{Var}[(\boldsymbol{y}',\ \boldsymbol{x}')']\\
&=\begin{pmatrix}\mathrm{Cov}(\boldsymbol{y}',\ \boldsymbol{y}') & \mathrm{Cov}(\boldsymbol{y}',\ \boldsymbol{x}')\\ \mathrm{Cov}(\boldsymbol{x}',\ \boldsymbol{y}') & \mathrm{Cov}(\boldsymbol{x}',\ \boldsymbol{x}')\end{pmatrix}\\
&=\begin{pmatrix}\Lambda_y\boldsymbol{A}(\boldsymbol{\Gamma\Phi\Gamma}'+\boldsymbol{\Psi})\boldsymbol{A}'\Lambda_y^{'}+\boldsymbol{\Phi}_\varepsilon & \Lambda_y\boldsymbol{A\Gamma\Phi}\Lambda_x^{'}\\ \Lambda_x\boldsymbol{\Phi\Gamma}'\boldsymbol{A}'\Lambda_y^{'} & \Lambda_x\boldsymbol{\Phi}\Lambda_x^{'}+\boldsymbol{\Phi}_\delta\end{pmatrix}
\end{aligned}$$

2. 参数估计方法

SEM 参数估计方法有多种，例如常用的最大似然估计法（Maximum Likelihood）和广义最小二乘法（Generalized Least Squares），还有工具变量法（Instrumental Variables）、两阶段最小二乘法（Two-Stage Least Squares）、未加权的最小二乘法（Unweighted Least Squares）、广义加权最小二乘法（Generally Weighted Least Squares）、对角加权最小二乘法（Diagonally

Weighted Least Squares）等等。限于篇幅，本文仅对最大似然估计法的理论作较为详尽的介绍。

Browne（1984）对于 SEM 的参数估计给出了一个统一的拟合标准即差异函数 $\mathrm{F}(\boldsymbol{S},\ \Sigma)$（Discrepancy Functions），这个函数是用来度量观测样本协差阵与理论协差阵的差异性的。而 SEM 的参数估计，也是建立在最小化差异函数的基础之上的。Browne 提出差异函数应该有如下的性质：

1、$\mathrm{F}(\boldsymbol{S},\ \Sigma)\geq 0$ 2、$\boldsymbol{S}=\Sigma\Rightarrow \mathrm{F}(\boldsymbol{S},\ \Sigma)=0$ 3、$\mathrm{F}(\boldsymbol{S},\ \Sigma)$ 一致连续

（1）**最大似然估计**

这种估计方法是基于多元正态分布的[8]；对于服从多元正态分布 $N_m(\boldsymbol{\mu},\ \Sigma)$ ($|\Sigma|\neq 0$) 的随机向量 $\boldsymbol{z}$，其分布密度如下：

$$f_{\mathrm{m}}(\boldsymbol{z})=\frac{1}{(2\pi)^{m/2}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right\}$$

那么显然，对于 SEM 的 $p+q$ 个观测变量，似然函数为：

$$L(\boldsymbol{\mu},\ \Sigma)\triangleq\frac{1}{(2\pi)^{(p+q)N/2}|\Sigma|^{N/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})\right\}$$

两边取对数：

$$\ln L(\boldsymbol{\mu},\ \Sigma)=-\frac{1}{2}(p+q)N\ln(2\pi)+\frac{N}{2}\ln\left|\Sigma^{-1}\right|-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})$$

对于上式的最后一项（记作 C）作如下的变换[9]：

$$\begin{aligned}\mathrm{C}&=\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})=\mathrm{tr}[\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})]\\&=\mathrm{tr}[\frac{1}{2}\sum_{i=1}^{N}\Sigma^{-1}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})']=\frac{1}{2}\mathrm{tr}[\Sigma^{-1}\sum_{i=1}^{N}(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})(\boldsymbol{z}_{(i)}-\boldsymbol{\mu})']\\&\triangleq\frac{1}{2}\mathrm{tr}[\Sigma^{-1}\sum_{i=1}^{N}(\boldsymbol{z}_{(i)}-\bar{\boldsymbol{z}})(\boldsymbol{z}_{(i)}-\bar{\boldsymbol{z}})']=\frac{N}{2}\mathrm{tr}(\Sigma^{-1}S^{*})\triangleq\frac{N}{2}\mathrm{tr}(S\Sigma^{-1})\end{aligned}$$

这样的话，我们舍弃一些对于最大化过程没有影响的常数项，对数似然函数可以整理为：

$$\ln L(\boldsymbol{\mu},\ \Sigma)\triangleq\ln\left|\Sigma^{-1}\right|-\mathrm{tr}(\boldsymbol{S}\Sigma^{-1})$$

根据对数似然函数整理之后的差异函数为[10]：

$$\mathrm{F}=\mathrm{tr}(\boldsymbol{S}\Sigma^{-1})-(p+q)+\ln|\Sigma|-\ln|\boldsymbol{S}|$$

我们的目标就是将差异函数最小化，从而求出相应的参数值。

---

[8] 有研究者称 ML 估计是基于 Wishart 分布的，但未给出证明，我的证明参见附录 1

[9] 注意其中用到了两次近似：一次是用样本均值代替总体均值，因为样本均值是总体均值的无偏估计；另一次是用 S 代替 S^{*}，原因是在大样本条件下，S^{*} 与 S 的差异可以忽略。这个过程中也用到了矩阵的迹（trace）的性质。

[10] 整理的目的是为了符合前面提到的差异函数的性质

（2）其他估计方法

当前的估计方法总体说来可以根据数值计算方法分两类型：迭代型和非迭代型。前面提到的最大似然估计就属于迭代型的估计方法，对于迭代型的方法，每一种参数估计方法都对应着一个拟合函数，而参数的估计则是通过最小化差异函数得到，其中需要用到数值迭代；而非迭代型方法，顾名思义即不需要通过数值迭代方法既可求解。表格 2 给出了这些方法的概览。图 1 展示了 AMOS 软件中所使用的五种估计方法。

以上提到的方法都是一些经典的估计方法，这些方法都有自己的优缺点，随着 SEM 多年来的发展，为了克服这些缺陷，也有一些研究者在估计方法上提出创新，比如文献[8]所提及的借助偏最小二乘法（PLS）的思想估计求解，不失为统计方法的综合妙用。

表格 2　　　　SEM 主要参数估计方法一览表

类型	方法	备注
迭代型	未加权的最小二乘法（ULS）	差异函数 $\mathrm{F}=\frac{1}{2}\mathrm{tr}[(S-\Sigma)^2]$
	广义最小二乘法（GLS）	差异函数 $\mathrm{F}=\frac{1}{2}\mathrm{tr}[(I-S^{-1}\Sigma)^2]$
	加权最小二乘法（WLS）[11]	差异函数 $\mathrm{F}=(s-\sigma)'W^{-1}(s-\sigma)$
	对角加权最小二乘法（DWLS）[12]	差异函数 $\mathrm{F}=\sum_{g=1}^{k}\sum_{h=1}^{k}(1/w_{gh})(s_{gh}-\sigma_{gh})^2$
非迭代型	工具变量法（IV）	
	两阶段最小二乘法（TSLS）	
	其他估计方法	例如[8]中的 PLS 方法

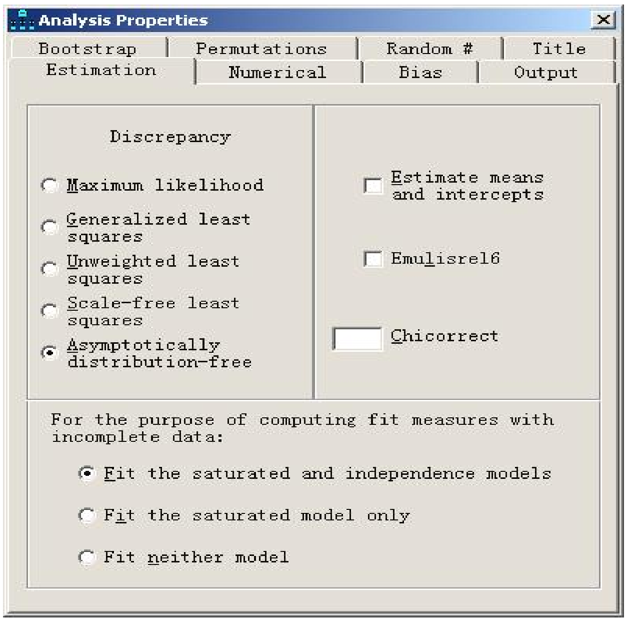

图 1 AMOS 中的参数估计方法

[11] 在 AMOS 软件中，此处的 WLS 估计法称作“*asymptotically distribution-free* estimation”（ADF），国内有研究者将其称为“渐进自由无干扰的加权最小二乘法”（参见[21]）

[12] 在 AMOS 软件中，此处的 DWLS 估计法称作“*‘scale free’ least squares* estimation (SLS)”；LISREL 软件则同样称为 DWLS

三、软件求解

目前比较流行的求解 SEM 的软件大致有两种[13]：AMOS 和 LISREL。前者原本是 SmallWaters 公司的产品，后来被 SPSS 公司收购，AMOS 是“Analysis of Moment Structure”的缩写，在 SEM 分析方面，它以图形界面见长，如图 2 所示，模型中的观测变量、潜变量、误差项都有不同的图形来表示，这些图形都可以从左边的工具栏中使用相应的工具在右边绘图区直接手工绘制，模型中的数据也可以通过菜单“File → Data Files → File Name”导入，在设定完参数的属性以及模型属性（包括标题、模型估计方法、模型输出结果选项、Bootstrap 选项等）之后，通过菜单“Model-Fit → Calculate Estimates”即可对模型求解；

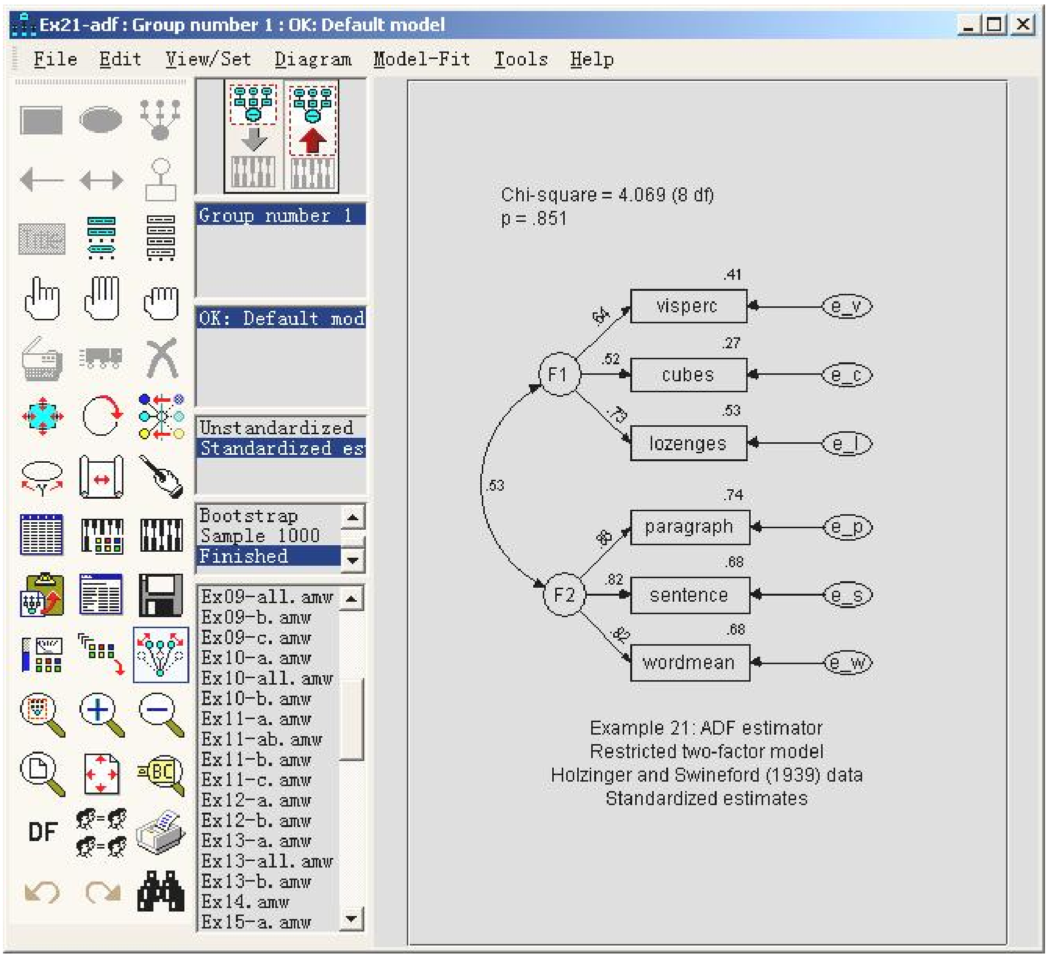

图 2 AMOS 软件的分析界面

而 LISREL（Linear Structural Relationship）一般是通过执行程序语句（Syntax）来求解计算的，观测变量、潜变量、协差阵等等在 LISREL 程序语句中都有相应的代码，如 LY、LX、BE、GA 等；在数据要求上 LISREL 与 AMOS 有所不同：AMOS 中可以直接使用原始数据（Raw Data）[14]，而 LISREL 软件要求输入的是原始数据的协差阵（Covariance Matrix）或者相关系数阵（Correlation Matrix），不能直接使用原始数据，通常使用 PRELIS[15]将原始数据读入然后转化为协差阵或者相关系数阵；在参数估计方法上，LISREL 可以使用前文所

[13] 此外还有 SAS 软件中的 CALIS 步、Mplus、EQS 等软件都可以用来求解 SEM

[14] 实际上，在进行计算求解时，AMOS 仍然需要首先把原始数据转化为协差阵或者相关系数阵。

[15] 取自 PRE-processor for LISrel 的缩写，意即在 LISREL 分析之前的处理工具。

提到的工具变量法和两阶段最小二乘法。

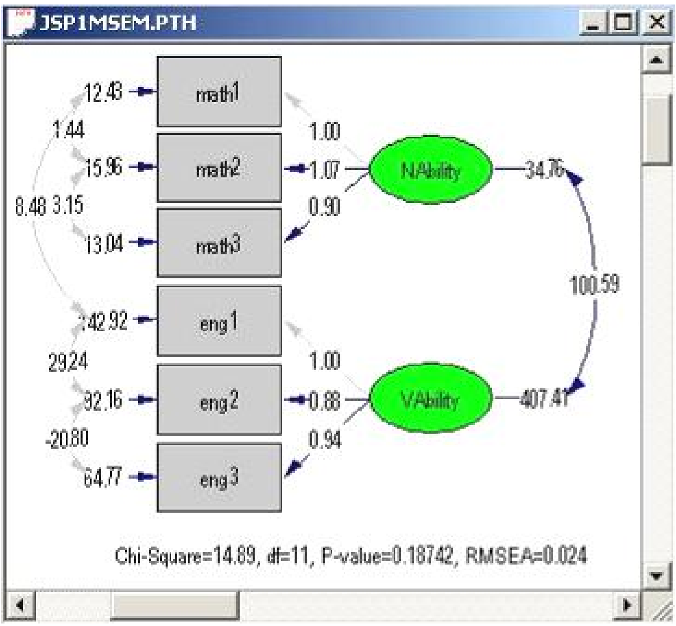

图 3 LISREL 软件的分析界面

SAS 软件从 6.04 版开始也提供了关于协方差结构分析的过程——CALIS（Covariance Analysis of Linear Structural Equations），该过程可广泛用于多变量线性回归、路径分析和因果关系模型、各种线性或非线性潜变量模型，它对于 SEM 的处理可能不如 AMOS 或 LISREL 那样能直观输出因果关系和路径图形（毕竟 SAS 不是专门的 SEM 处理工具），但是从程序语句上来说要比前二者简单一些。

Emply.sas *

```
proc calis data=Employ method=ml toteff;
lineqs x2 = f1 + e3,
       x4 = f2 + e4,
       x5 = f3 + e5,
       x6 = a16 f1 + a18 f3 + e6,
       y1 = f4 + e7,
       y2 = a20 f4 + e8,
       y3 = a21 f4 + e9,
       f4 = f1 + b2 f2 + b3 f3 + d1;
       std e3-e9 d1 = 8 * var:;
run;
```

图 4 SAS 的 CALIS 过程示例

四、模型结果解释与评价

在软件输出结果之后，我们可以通过一系列的指标来对 SEM 的各个组成部分进行评价，包括对测量模型、结构模型、整体模型以及模型参数设定的评价，检验模型和参数是否合理。这些指标如下所示：

1. 对测量模型的评价

根据经典的测量理论，单个测量Y的可靠性为：

$$\frac{\mathrm{Var}(T)}{\mathrm{Var}(T)+\mathrm{Var}(\varepsilon)}=1-\frac{\mathrm{Var}(\varepsilon)}{\mathrm{Var}(Y)}$$

其中$Y=T+\varepsilon$，T表示真实成分，ε表示测量误差。因此前面测量模型中第i个观测变量的可靠性为：

$$1-\hat{\theta}_{ii}^{2}/\hat{\sigma}_{ii}^{2}$$

在 LISREL 中称作平方多重相关系数，其中$\hat{\theta}_{ii}^{2}$为误差方差的估计量，$\hat{\sigma}_{ii}^{2}$为第i个观测变量的方差。整个测量模型的可靠性用下面的决定系数表示：

$$1-\left|\hat{\Theta}\right|/\left|\hat{\Sigma}\right|$$

决定系数一般在 0 到 1 之间，越接近于 1 表示测量指标的可靠性越好。

2. 对结构模型的评价

类似的，第i个结构方程式的多重相关系数以及全部结构模型的总决定系数分别为：

$$1-\widehat{\mathrm{Var}}(\boldsymbol{\zeta}_i)/\widehat{\mathrm{Var}}(\boldsymbol{\eta}_i)$$

$$1-\left|\hat{\boldsymbol{\Psi}}\right|/\left|\widehat{\mathrm{Var}}(\boldsymbol{\eta})\right|$$

3. 对整个模型的评价

对于整个模型的拟合程度，可以用很多种拟合指数来描述[16]，以下只是简单介绍其中几种（比较全面地介绍可参见[27]）：

（1）Chi-方统计量

对于 ML、GLS 和 WLS，χ^2统计量就是$(N-1)$乘以拟合函数的最小值，对 ULS 和 DWLS 则要作一些修改（参见[26]）。χ^2值越小表示模型拟合越好；但这个统计量也需要一些假设条件，如观测变量服从多元正态分布、分析基于样本协差阵进行、样本量足够大等。不得不注意的是该统计量对于样本量和观测变量的正态性非常敏感，比如大的样本量将对应较大的χ^2值（相应的 P 值就小）从而容易得出拟合较差的结论。χ^2的自由度为：

$$df=\frac{1}{2}(p+q)(p+q+1)-t$$

其中 t 是待估计的参数的数目。

[16] AMOS 5 软件中，通过查看帮助，可以知道现已有 33 种拟合指数（Measures of Fit）

（2）拟合优度指数（GFI）

$$\text{GFI} = 1 - (s-\sigma)' W^{-1} (s-\sigma) / s' W^{-1} s$$

拟合优度越接近于 1 表示模型拟合越好。

（3）修正的拟合优度指数（AGFI）

$$\text{AGFI} = 1 - [(p+1)(p+q+1)/2df](1-\text{GFI})$$

同样 AGFI 越接近于 1 表示模型拟合越好，这两个指数与 χ^2 相比起来，对样本量和正态性并不敏感，但缺点是我们对其分布也并不了解。

（4）平均平方残差的平方根（RMR）

$$\text{RMR} = \sqrt{\frac{\sum_{i=1}^{p+q}\sum_{j=1}^{i}(s_{ij}-\sigma_{ij})^2}{\frac{1}{2}(p+q)(p+q+1)}}$$

RMR 所度量的是拟合残差的一种平均值，可以用来比较不同的模型对同一数据拟合好坏，越接近于零表示拟合越好。

4. 对参数设定的评价

在模型设定过程中，可能有些参数被设定为固定或者施加了约束条件，如果将这其中某个参数改为自由参数的话，χ^2 值会有所变化，这个变化值也就是所谓的模型修正指数（MI）；若 χ^2 值有明显的减小，那说明该参数设置并不合理，应该改为自由参数，以使模型更好地拟合，因此 MI 提供的信息实际上是修正模型的依据。

五、国内应用误区

对于 SEM，在国外学者中不乏有一些批评的声音，如 Freedman（1987）和 Sobel（1998）、Cliff[17]（1983），而国内则鲜以闻（尤其是大陆[18]，即使有，也难以真正从数理统计角度解释清楚原理），许多研究者对 SEM 推崇备至，甚至有人将之视为“多种统计方法之母”（参见[14]）。缺乏对理论的深入研究，就容易走入一些误区，不妨列举二三如下：

[17] Cliff 的批评是比较著名而深刻的，参见[29][31]
[18] 台湾的黄芳铭老师在其著作《结构方程式 理论与应用》最后提到了一些关于 SEM 的批评。

1. 做出“接受模型”的结论

正如本文的参考文献[13]明确指出：SEM models can never be accepted; they can only fail to be rejected. 这其实就是假设检验的思想：一般来说我们只能拒绝或不拒绝零假设，而不能接受零假设。纵观 SEM 的建立和求解、检验过程，其实就是一个假设检验（Hypothesis Test）的过程，在最开始我们给出零假设：总体协差阵与模型拟合协差阵相等，即 $\Sigma=\Sigma(\theta)$。基于这个假设，便可以推出近似的 χ^2 检验统计量[19]，从而判断模型总的拟合程度。即使是 χ^2 值较小（或者相应的 P 值较大），我们也不能轻易做出“接受模型”的结论[20]，因为我们并不知道是否还有其它更好的或者同等的模型（Alternative Models），也许那些模型才是客观上真实的模型。

2. 忽视假设前提

从 SEM 的参数估计方法上来看，事实上有一些假设前提是不能忽略的，比如，当我们使用最大似然估计时，应该首先检查观测变量是否满足服从多元正态分布和样本相互独立的要求，如果不能满足假设，则应该选择其他的估计方法。当然，随着 SEM 的发展，对于一些特殊情况，例如离散数据、非正态分布等，都出现了相应的解决办法。无论如何，任何一种方法都有其理论背景，在应用时应该加以注意。

3. SEM 似是而非的“优点”

不少研究者对于 SEM 的优点作了总结，给人一种“SEM 万能”的感觉。例如多位研究者对于 SEM 的误差项处理持赞赏态度，例如曲波等（2005）认为 SEM“容许自变量及因变量含测量误差”[21]（参见[9]），我认为充其量 SEM 只是提出了测量的概念，这是以往的统计方法很少提及的，这里的“测量误差”意义是否比其他统计方法如回归分析中的误差更为深刻或更有价值？从分析的角度来看并不是的；更让人匪夷所思的是曾武等（2001）认为 SEM 对误差的处理“较大增强统计效率”（参见[10]），因为测量模型和结构模型中都含有误差项，这是对误差的“二次分离”，相比起诸如回归分析等方法的误差“一次分离”，这种“二次分离”的方式显得更加精致。

在国外，备受争议的就是 SEM 中的因果关系（Causal Relationship）。关于因果关系的测定与检验，在统计学理论中一直以来就存有争议，其中较为著名的如 Granger 因果检验，其实都不能说是完美的理论（只能说是作为参考），因为从哲学意义上来说，因果关系的几乎是无法检验的（[28]）。

总体说来，SEM 引进到一些社会科学领域所带来的巨大变革主要原因仍在于本文前言中提到的它从理论上解决了测量问题和因果关系度量的难题，所以这两方面也是为人称道之处，然而，我认为，定量研究方法与社会科学领域毕竟是不能完全契合的，SEM 提出的仅仅是理论上的概念，这些概念不能被神化，具体实践上，SEM 仍然需要社会科学领域的正确理论指导，SEM 的结果，只能说是从统计学的角度对社会科学理论给予了一定的支持；至于实际情况究竟如何，真理究竟是什么样的，这些问题都不能完全依靠统计理论方法解决。

---

19 注意这个统计量的得出也必须有若干前提假设，前文中有所提及，具体可以参见[24]。

20 勉强一些可以说“临时接受模型”

21 很多研究者都提到了 SEM 的这一个优点，包括侯杰泰等著名的 SEM 研究者，但我并不赞同。

4. 对缺陷提出的表面可行的补救方法

SEM 应用中会暴露出模型的某些缺陷，举一例：对于统计研究来说，制约性最强的部分在于数据；SEM 中大部分估计方法都要求数据的多元正态性，这在实际应用中并不是很容易满足，据观察，大部分研究者对于这个问题的解决办法都只是简单的“正态化”（Normalize）；之所以说这个办法“表面可行”，是因为一般说来变量的正态化处理都是非线性的[22]，比如取平方根，以及著名的 Box-Cox 变换等等方法，非线性的致命之处不仅在于增加了模型的复杂性，而且将来解释起来将不如一般的线性模型那样清楚容易；而线性变换的正态化效果又很差，所以数据的非正态性，事实上不是那样简单的问题。

六、小结

SEM 的理论和应用在国外已经都趋于成熟，而在中国国内只是在应用上有所发展，而缺乏相关的理论研究，这是导致应用不当的直接原因，当然，客观上来说，很多应用 SEM 的人并非数学或统计专业，对理论背景不了解是可以理解的。本文以“纠错”为目的，对 SEM 的数理统计理论背景以及部分证明给出了介绍，以提醒应用者，SEM 并不是万能工具，一方面有理论前提假设的限制，另一方面在结果解释上也有若干应注意的地方。

当然，定量方法在很多社会科学领域确实有其难处，不能纯粹从理论角度苛刻地进行批判，应该承认，SEM 在这些领域开辟了很大的研究空间，这其中 Jöreskog 以及很多国外研究者对 SEM 的发展与推广做出了不可磨灭的贡献，我们国内的研究者们也应该在清楚研究 SEM 理论的前提下，结合学科发展实际情况，使 SEM 成为真正有效的研究工具。

致谢 本文中的部分数理统计公式推导在我推导遇到困难时得到了加拿大 McMaster 大学 John Fox 教授的悉心指引，还有中国传媒大学柯惠新教授、瑞典 Dalarna 大学 Kenneth Natanaelsson 老师、澳大利亚 Victoria 大学 Tran Van Hoa 教授、复旦大学黄荣贵同学、澳大利亚 Monash 大学的 Liang Zhuo 同学、丹麦的 Yangguoyi Ou 同学、浙江大学姚涛同学都曾给我热心帮助，在此一并表示衷心感谢。

[22] 可能这一点很少有人注意到。

附录

1. Wishart 分布密度

Wishart 分布由 Wishart 在 1928 年推导出来，与一元统计中的 χ^2 分布有着同样重要的作用。在引进 Wishart 分布密度之前，我们需要了解矩阵的分布的定义：对于一个矩阵 $X_{n\times m}=(x_{ij})$，其分布就是将其列向量一个接一个地组成一个长向量的分布，当 X 是对称阵的时候，其分布就是向量 $x=(x_{11},\cdots,x_{1n},x_{22},\cdots x_{2n},\cdots x_{n-1,n-1},x_{n-1,n},x_{n,n})'$ 的分布。

假定 $y_{(1)},\cdots,y_{(n)}$ 相互独立，$y_{(i)}\sim N_m(\mu_i,\ V),\ i=1,\cdots,n$，记

$$Y=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix} \triangleq \begin{bmatrix} y'_{(1)} \\ y'_{(2)} \\ \\ y'_{(n)} \end{bmatrix}$$

$$M=E(Y)=\begin{bmatrix} \mu_{11} & \cdots & \mu_{1m} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nm} \end{bmatrix}$$

定义 $A=\sum_{i=1}^{n} y_{(i)}y'_{(i)}=Y'Y$ 的分布为非中心 Wishart 分布，记作 $A\sim W_m(n,\ V,\ \tau)$，其中 $\tau=M'M$，若 $\tau=0$ 则称为中心 Wishart 分布，记作 $A\sim W_m(n,\ V)$。当 $n>m$ 时，$W_m(n,\ V)$ 的分布密度为：

$$f(A)=\begin{cases} \dfrac{|A|^{\frac{1}{2}(n-m-1)}\exp\left\{-\dfrac{1}{2}\mathrm{tr}V^{-1}A\right\}}{2^{nm/2}\pi^{m(m-1)/4}|\ V\ |^{n/2}\prod_{i=1}^{m}\Gamma\left(\dfrac{n-i+1}{2}\right)}, & \text{if } A \text{ is positive definite} \\ 0, & \text{else} \end{cases}$$

下面推导基于 Wishart 分布的最大似然估计。首先对于 SEM 中的观测变量 $\boldsymbol{z}=(\boldsymbol{y}',\ \boldsymbol{x}')'$ 我们假设服从多元正态分布：

$$\boldsymbol{z}_{(i)}=\left(z_{i1},\ z_{i2},\ \cdots,\ z_{i,p+q}\right)'\sim N_{p+q}(\boldsymbol{\mu},\ \boldsymbol{\Sigma})\quad i=1,\ 2,\ \ldots,N$$

根据 Wishart 分布密度不难推导样本协差阵[23] $\boldsymbol{S}=\dfrac{1}{N}\sum_{i=1}^{n}(\boldsymbol{z}_{(i)}-\bar{\boldsymbol{z}})(\boldsymbol{z}'_{(i)}-\bar{\boldsymbol{z}})'$ 的分布密度

[23] 同前面基于多元正态分布的推导一样，这里用 S 代替 S^*

为：

$$f(\boldsymbol{S})=\begin{cases} N\dfrac{\left|N\boldsymbol{S}\right|^{\frac{1}{2}(N-p+q-1)}\exp\left\{-\dfrac{N}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right\}}{2^{N(p+q)/2}\pi^{(p+q)(p+q-1)/4}\left|\boldsymbol{\Sigma}\right|^{N/2}\prod_{i=1}^{p+q}\Gamma\left(\dfrac{N-i+1}{2}\right)}, & \text{if } \boldsymbol{S} \text{ is positive definite} \\ 0, & \text{else} \end{cases} \qquad (m=p+q)$$

注意此处多了一个条件，即要求样本协差阵是正定阵（Positive Definite Matrix），这也是 LISREL 等软件需要检验样本协差阵是否正定的原因。最大似然估计方法所用的对数似然函数也就是对上面的密度函数取对数，后续的处理方法与前面雷同。

2. 用线性方法处理变量的结果

此处我们使用 S-Plus 软件产生服从 $\chi^2(5)$ 的数据，然后使用标准化的方法（这是一种线性处理），从图 5 中可以看出，这种处理对数据分布并没有实质性的影响。程序代码如下：

```
> chi <- rchisq(1000, 5)
> plot(density(chi), type="l", xlab="Chi", ylab="")
> plot(density((chi-mean(chi)) / stdev(chi)), type="l",
     xlab="Standardized Chi", ylab="")
```

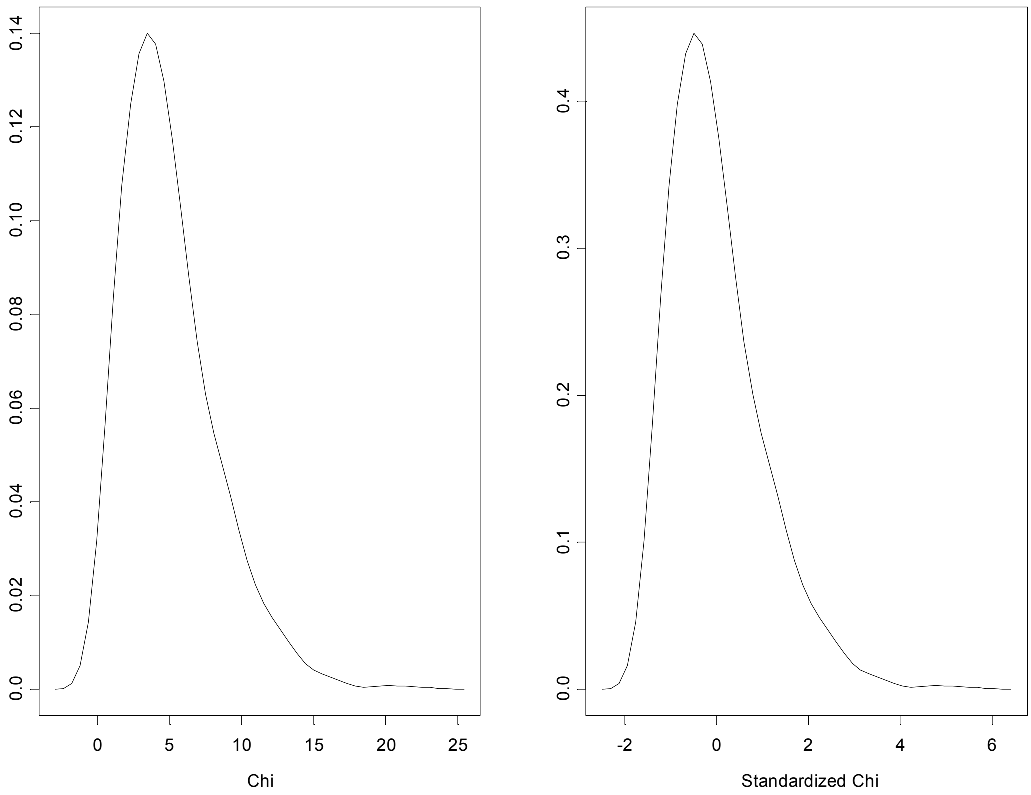

图 5 标准化处理对数据分布的影响效果

参考文献

[1] Adrian. E. Raftery. Statistics in Sociology, 1950-2000. *Journal of the American Statistical Association*, Vol. 95, No. 450 (Jun., 2000), 654-661

[2] Blau, P. M. & O. D. Duncan. 1967. *The American Occupational Structure*. New York: Wiley.

[3] Jöreskog, K. G. 1973. A General Method for Estimating a Linear Structural Equation System. In *Structural Equation Models in the Social Sciences*, ed. A. S. Goldberger & O. D. Duncan. New York: Seminar Press pp. 85—112.

[4] John Fox. January 2002. *Structural Equation Models: Appendix to An R and S-PLUS Companion to Applied Regression*. Source: http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-sems.pdf

[5] 张尧庭、方开泰，《多元统计分析引论》，科学出版社，北京，1982.6

[6] 娄峥嵘，浅析结构方程建模的基本步骤，《生产力研究》，NO.6.2005

[7] 田晓明、傅珏生，结构方程模型的统计方法及比较，苏州大学学报（自然科学版），第 21 卷第 4 期，2005 年 10 月

[8] 朱利平、刘莉，线性结构方程参数估计的一种简单方法，《应用概率统计》，第二十一卷第二期，2005 年 5 月

[9] 曲波、郭海强、任继萍、孙高，结构方程模型及其应用，《中国卫生统计》，第 22 卷第 6 期，2005 年 12 月

[10] 曾武、黄子杰，线性结构方程模型的原理及其实际应用，福建医科大学学报（社会科学版），第 2 卷第 1 期，2001 年 6 月

[11] 侯杰泰、温忠麟、成子娟，《结构方程模型及其应用》，教育科学出版社，2004

[12] 李子奈，《计量经济学》，高等教育出版社，北京，2003

[13] *Structural Equation Modeling using AMOS: An Introduction*. 来自互联网（2006-05-03 可访问）：http://www.utexas.edu/cc/stat/tutorials/amos/index.html

[14] 李伟明，心理计量学的长足进步[J]，心理科学，1998，21（6）：528

[15] Bollen, K.A., 1989. *Structural Equations with Latent Variables*. Wiley, New York.

[16] Anderson TW. *An Introduction to Multivariate Statistical Analysis*. 2nd Edition. Wiley. 1984

[17] William H. Greene. *Econometric Analysis* (5th Edition). Pearson Education. 2002

[18] *Multilevel Structural Equation Modeling*, from the technical documents of LISREL 8.7 for Windows Online Help

[19] Amos 5 Reference Guide

[20] John Fox, 2006. *An Introduction to Structural Equation Modelling*. Lecture Notes

[21] 柯惠新、祝建华、孙江华，《传播统计学》，北京广播学院出版社，北京，2003

[22] *LISREL 8 and PRELIS2: Getting Started*. 来自互联网（2006-05-05 可访问）：http://www.utexas.edu/its/rc/tutorials/stat/lisrel/lisrel8.html

[23] 陈智、孙高、王泓，介绍 SAS 中的一个新过程，中国卫生统计 1996 年第 13 卷第 5 期

[24] 柯惠新，因果关系模型的估计与检验（III），数理统计与管理，1991 年 03 期

[25] *Tutorial: Get Running with Amos Graphics*, Amos 4 User's Guide

[26] BROWN E. Asymptotically distribution-free methods for the analysis of covariance structures[J] .*British Journal of Mathematical and Statistical Psychology*, 1984, 37, 62 - 83.

[27] 温忠麟、侯杰泰、马什赫伯特，结构方程模型检验:拟合指数与卡方准则，心理学报 2004, 36 (2): 186～194

[28] 曹永福，格兰杰因果性检验评述，《世界经济统计研究》2005 年第 2 期

[29] 黄芳铭，《结构方程式 理论与应用》，中国税务出版社，北京，2005

[30] Richard A. Johnson, Dean W. Wichern 著，陆璇、葛余博、赵衡秀、叶俊译，《实用多元统计分析》

（Applied Multivariate Statistical Analysis），清华大学出版社，北京，2001

[31] Cliff N., Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*. 1983

[32] 李贤平，《概率论基础》，高等教育出版社，北京，1997

[33] 北京大学数学系几何与代数教研室代数小组，《高等代数》，高等教育出版社，北京，1988

[34] 茆诗松等，《数理统计》，华东师大出版社，1990