

A Stagewise Approach to Structural Equation Modeling

Bora Lee^a · Changsoo Park^{a,1}

^aDepartment of Statistics, Chung-Ang University

(Received November 10, 2014; Revised December 18, 2014; Accepted January 7, 2015)

Abstract

Structural equation modeling (SEM) is a widely used in social sciences such as education, business administration, and psychology. In SEM, the latent variable score is the estimate of the latent variable which cannot be observed directly. This study uses stagewise structural equation modeling(stagewise SEM; SSEM) by partitioning the whole model into several stages. The traditional estimation method minimizes the discrepancy function using the variance-covariance of all observed variables. This method can lead to inappropriate situations where exogenous latent variables may be affected by endogenous latent variables. The SSEM approach can avoid such situations and reduce the complexity of the whole SEM in estimating parameters.

Keywords: Stagewise SEM, latent variable score, exogenous latent variable, endogenous latent variable.

1. 서론

최근 구조식 모형은 사회과학, 자연과학 분야 등 다양하게 활용되고 있다. 구조식 모형에서 잠재변수의 추정치 중 하나인 잠재변수 점수를 계산할 때, 외생잠재변수가 내생잠재변수의 관측변수의 영향을 받을 수 있다. 이러한 관계가 현실적으로 이치에 맞지 않는 경우가 있는데, 이 연구에서는 이러한 경우를 해결하기 위하여 구조식 모형과 잠재변수 점수에 대하여 기본적인 이론을 정리한 후, 단계별로 구조식 모형의 해를 구하는 단계별 구조식 모형을 제시하려 한다.

1.1. 구조식 모형(Structural Equation Modelling; SEM)

사회과학 분야는 자연과학 분야와는 달리, 실험을 설계할 때 완벽하게 독립변수를 통제하는 것이 불가능한 경우가 많다. 이에 따라 인과관계를 갖는 모형과 가설을 통계적으로 검증함으로써 추론을 이끌어내는 구조식 모형이 활발하게 이용되고 있다.

구조식 모형은 관측변수에 의해 잠재변수를 정의하고, 정의된 잠재변수들간의 상호관계, 측정오차 등을 동시에 고려하여 모형을 구축하는 기법이다. 구조식 모형은 크게 측정모형과 구조모형으로 구분되는데, 관측변수에 의해 잠재변수를 정의하는 측정모형은 확증적 요인분석과 동일한 모형을 가지며, 잠재변수들 간의 상호관계를 나타내는 구조모형은 회귀분석과 유사한 형태를 띤다. 즉, 구조식 모형은 확증적 요인분석과 회귀분석의 혼합형태라고 할 수 있다 (Lei와 Wu, 2007).

This research was supported by Chung-Ang University Research Scholarship Grants in 2012-2014.

¹Corresponding author: Department of Statistics, Chung-Ang University, Seoul 156-756, Korea.

E-mail: cspark@cau.ac.kr

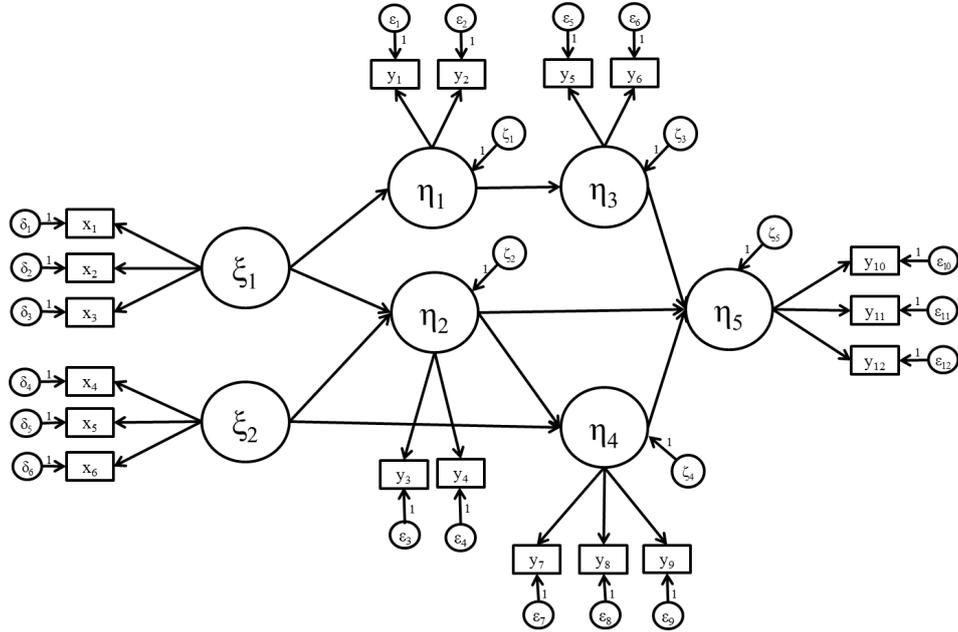


Figure 1.1. Example of a path diagram

구조식 모형을 이해하는 유용한 방법으로 경로도(path diagram)가 있다. 이것은 모든 변수와 오차항, 변수 간의 상호관계를 도식화한 형태로 모형의 구조를 한 눈에 파악하기 쉽다. 경로도의 구성요소를 살펴보면 화살표는 변수에서 변수로의 직접효과를 나타내며, 쌍방향 화살표는 두 변수 간의 상관관계를 나타낸다. 원에 나타난 변수는 잠재변수이며, 외생잠재변수와 내생잠재변수로 구분된다. 외생잠재변수는 어떤 변수에서도 화살표를 받지 않는 잠재변수이며 ξ 로 표기되고, 내생잠재변수는 하나 이상의 잠재변수로부터 화살표를 받는 잠재변수로서 η 로 표기된다. 내생잠재변수가 외생잠재변수에 의해 설명되지 못하는 부분을 구조오차로 정의하고 ζ 로 나타내며, 외생잠재변수의 분산-공분산은 ϕ , 구조오차의 분산-공분산은 ψ 로 표기한다. 외생잠재변수 ξ_j 에서 내생잠재변수 η_i 로의 직접효과를 나타내는 계수는 γ_{ij} 로, 내생잠재변수 η_j 에서 내생잠재변수 η_i 로의 직접효과를 나타내는 계수는 β_{ij} 로 정의한다. 각 계수의 첫 번째 아래첨자는 종속변수, 두 번째 아래첨자는 독립변수를 나타낸다.

Figure 1.1은 경로도의 한 가지 예이다. 사각형에 나타난 변수는 관측변수이며, 외생잠재변수의 관측변수는 x 로, 내생잠재변수의 관측변수는 y 로 구분된다. 관측변수 x 에 대한 측정오차는 δ , 관측변수 y 에 대한 측정오차는 ϵ 로 표기하며, 이 측정오차들의 분산-공분산은 각각 θ_δ , θ_ϵ 으로 정의한다. 외생잠재변수 ξ_j 에서 관측변수 x_i 로의 직접효과를 나타내는 계수는 λ_{xij} 로, 내생잠재변수 η_j 에서 관측변수 y_i 로의 직접효과를 나타내는 계수는 λ_{yij} 로 표기한다. 각 계수의 두 번째 아래첨자는 종속변수, 세 번째 아래첨자는 독립변수를 의미한다 (Jöreskog와 Sörbom, 1999). 이 때, 모형이 식별되기 위해서는 각 잠재변수마다 관측변수로의 직접효과를 나타내는 계수 중 하나를 1로 고정시켜야 한다.

경로도를 통해 전체 모형에서 특정 잠재변수로의 다양한 경로를 확인할 수 있는데, 하나의 외생잠재변수에서 하나의 내생잠재변수로 가는 경로는 하나의 화살표로 나타나는 경로가 있고, 둘 또는 그 이상의 화살표로 나타나는 경로가 있다. 이 때 하나의 화살표로 나타나는 경로를 직접경로라 하고 그렇지 못한 경우를 간접경로라 한다. 또한 직접경로에 의한 내생잠재변수는 직접내생잠재변수라 하고, 간접경로에 의

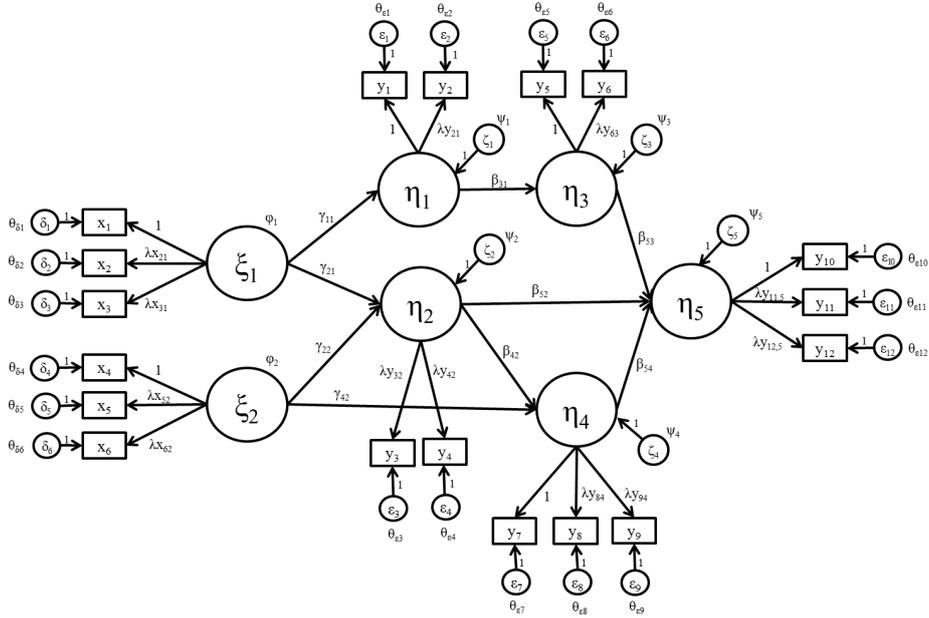


Figure 1.2. Example of a path diagram with parameters

한 내생잠재변수를 간접내생잠재변수라 한다. 따라서 외생잠재변수에서 내생잠재변수로 가는 경로 중에 직접경로와 간접경로가 공존하면, 내생잠재변수는 직접내생잠재변수도 되고 간접내생잠재변수도 될 수 있다. 예를 들면, Figure 1.1에서 ξ_2 에서 η_4 로 가는 경로 중에 ξ_2 에서 직접 η_4 로 가는 화살표는 직접 경로에 해당하고 ξ_2 에서 η_2 를 거쳐 η_4 로 가는 두 개의 화살표로 나타나는 경로는 간접경로에 해당한다.

총 n 개의 외생잠재변수 $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ 와 m 개의 내생잠재변수 $\eta = (\eta_1, \eta_2, \dots, \eta_m)^T$ 로 구성된 구조모형이 있다면 구조식은 다음과 같은 선형식으로 표현된다.

$$\eta = B\eta + \Gamma\xi + \zeta, \tag{1.1}$$

여기서 $B_{m \times m}$ 과 $\Gamma_{m \times n}$ 은 계수행렬이고, $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_m)^T$ 는 구조오차를 나타낸다. 이 때, B 는 대각원소가 0(zero)이고, $I - B$ 는 정칙(non-singular)행렬이어야 한다. 외생잠재변수 $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ 의 관측변수 $x = (x_1, x_2, \dots, x_q)^T$ 와 내생잠재변수 $\eta = (\eta_1, \eta_2, \dots, \eta_m)^T$ 의 관측변수 $y = (y_1, y_2, \dots, y_p)^T$ 에 대한 측정모형은 다음과 같은 선형회귀식으로 표현된다.

$$\begin{aligned} y &= \Lambda_y \eta + \epsilon, \\ x &= \Lambda_x \xi + \delta, \end{aligned} \tag{1.2}$$

여기서 $\Lambda_{y \times m}$ 과 $\Lambda_{x \times n}$ 은 계수행렬이고, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)^T$ 와 $\delta = (\delta_1, \delta_2, \dots, \delta_q)^T$ 는 각각 y 와 x 의 측정오차를 나타낸다. 이 때 외생잠재변수 $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ 의 분산-공분산행렬은 Φ , 구조오차 $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_m)^T$ 의 분산-공분산행렬은 $\Psi_{m \times m}$, y 의 측정오차 $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)^T$ 의 분산-공분산행렬은 Θ_ϵ , x 의 측정오차 $\delta = (\delta_1, \delta_2, \dots, \delta_q)^T$ 의 분산-공분산행렬은 Θ_δ 로 표기한다.

Figure 1.2는 Figure 1.1의 경로도 상에 계수 및 분산을 기호로 표기한 것이다. Figure 1.2에 나타난 2개의 외생잠재변수 $\xi = (\xi_1, \xi_2)^T$ 와 5개의 내생잠재변수 $\eta = (\eta_1, \eta_2, \dots, \eta_5)^T$ 로 구성된 구조식과 측정식

을 표현하면 다음과 같다. 모형 식별을 위하여 각 잠재변수마다 관측변수로의 직접효과를 나타내는 계수 중 임의로 첫 번째 계수를 1로 고정하였다.

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{31} & 0 & 0 & 0 & 0 \\ 0 & \beta_{42} & 0 & 0 & 0 \\ 0 & \beta_{52} & \beta_{53} & \beta_{54} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \\ 0 & 0 \\ 0 & \gamma_{42} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \end{bmatrix}, \quad (1.3)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} \lambda_{y_{11}} (= 1) & 0 & 0 & 0 & 0 & 0 \\ \lambda_{y_{21}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{y_{32}} (= 1) & 0 & 0 & 0 & 0 \\ 0 & \lambda_{y_{42}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{y_{53}} (= 1) & 0 & 0 & 0 \\ 0 & 0 & \lambda_{y_{63}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{y_{74}} (= 1) & 0 & 0 \\ 0 & 0 & 0 & \lambda_{y_{84}} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{y_{94}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{y_{10,5}} (= 1) & 0 \\ 0 & 0 & 0 & 0 & \lambda_{y_{11,5}} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{y_{12,5}} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \end{bmatrix}, \quad (1.4)$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{x_{11}} (= 1) & 0 \\ \lambda_{x_{21}} & 0 \\ \lambda_{x_{31}} & 0 \\ 0 & \lambda_{y_{12}} (= 1) \\ 0 & \lambda_{y_{22}} \\ 0 & \lambda_{y_{32}} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}. \quad (1.5)$$

이 때 잠재변수 및 오차들의 분산-공분산행렬과 계수행렬 $\Phi, \Psi, \Theta_\delta, \Theta_\varepsilon, \mathbf{B}, \Gamma, \Lambda_x, \Lambda_y$ 을 구조식 모형에서 추정해야 할 모수 θ 라고 정의하면, 측정변수 \mathbf{x}, \mathbf{y} 들의 분산-공분산행렬은 모수 θ 에 의해 다음과 같이 표현된다.

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{xx} \end{bmatrix} = \begin{bmatrix} \Lambda_y \mathbf{A} (\Gamma \Phi \Gamma^T + \Psi) \mathbf{A}^T \Lambda_y^T + \Theta_\varepsilon & \Lambda_y \mathbf{A} \Gamma \Phi \Lambda_x^T \\ \Lambda_x \Phi \Gamma^T \mathbf{A}^T \Lambda_y^T & \Lambda_x \Phi \Lambda_x^T + \Theta_\delta \end{bmatrix}, \quad (1.6)$$

여기서 $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ 이다.

관측변수 \mathbf{x}, \mathbf{y} 의 표본분산-공분산행렬을

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{yy} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{xx} \end{bmatrix} \quad (1.7)$$

라 할 때, \mathbf{S} 에 가장 가까운 Σ 가 되게 하는 모수 θ 의 값을 계산하여 θ 의 추정값인 $\hat{\theta}$ 으로 설정하고, 이 모수 추정치 $\hat{\theta} = (\hat{\Phi}, \hat{\Psi}, \hat{\Theta}_\delta, \hat{\Theta}_\varepsilon, \hat{\mathbf{B}}, \hat{\Gamma}, \hat{\Lambda}_x, \hat{\Lambda}_y)$ 를 구조식 모형의 해로 선택한다. 이 때 \mathbf{S} 와 Σ 의 가까운 정도를 측정하기 위하여 불일치 함수(discrepancy function)를 이용하는데, 정의된 불일치 함수에 따라 최대우도법(maximum likelihood; ML), 일반화 최소제곱법(generalized least squares; GLS), 비가중 최

소제곱법(unweighted least squares; ULS) 등이 있으며 각 방법에서 정의하는 불일치 함수는 다음과 같다 (Jöreskog, 1967, 1972).

$$\begin{aligned} F_{ML} &= \log |\Sigma| + \text{tr}(\mathbf{S}\Sigma^{-1}) - \log |\mathbf{S}| - (p + q), \\ F_{GLS} &= \frac{1}{2} \text{tr} [(\mathbf{I} - \mathbf{S}^{-1}\Sigma)^2], \\ F_{ULS} &= \frac{1}{2} \text{tr} [(\mathbf{S} - \Sigma)^2]. \end{aligned} \quad (1.8)$$

구조식 모형 분석에 특화된 소프트웨어 프로그램으로 LISREL, AMOS, EQS 등이 있으며, R에서도 OpenMix, sem, lavaan 등의 패키지를 설치할 경우 구조식 모형을 수행할 수 있다. SAS에서는 TCALIS와 CALIS 프로시저를 이용하여 구조식 모형 분석이 가능하다.

1.2. 잠재변수 점수(Latent Variable Score; LVS)

구조식 모형이 활발하게 이용되면서 점차 복잡한 형태의 모형 분석이 요구되고, 잠재변수 간의 관계를 심도 있게 해석하려는 시도도 증가하고 있다. 이러한 연구 중에는 잠재변수의 추정치인 잠재변수 점수(latent variable score)를 이용하는 방법이 있다. 잠재변수 점수는 직접적인 측정이 불가능한 잠재변수의 추정치를 수량화하여 나타낸 값으로 잠재변수 점수를 계산하는 방법으로는 Anderson과 Rubin (1956)이 제안한 방법, Lawley와 Maxwell (1962)이 제안한 방법, Bartholomew와 Knott (1999)이 제안한 방법 등 여러 가지가 있다. 이 중 Anderson과 Rubin이 제안한 방법은 가중 최소제곱법(weighted least squares; WLS)으로, 실제 잠재변수와 동일한 분산-공분산 행렬을 갖는 잠재변수 점수를 얻을 수 있다는 장점이 있으며, Jöreskog (2000)는 이 방법을 통해 잠재변수 점수를 얻을 수 있도록 LISREL 8.8 버전부터 프로시저를 제공하고 있다.

잠재변수 점수는 탐색적 요인 분석(explanatory factor analysis)의 요인 점수(factor score)와 유사한 형태이며, 이를 이용하여 잠재변수에 대한 각 케이스의 순위를 매기거나 여러 집단으로 분류도 가능하다. 또한 외부변수와 잠재변수 사이의 상관관계를 유추하거나 잠재변수 간의 비선형관계를 예측할 수 있고 오차항의 분포에 대한 연구에도 활용할 수 있다 (Jöreskog, 2000; Jöreskog 등, 2006).

잠재변수 점수를 구하는 과정을 요약하면 다음과 같다. 식 (1.1)의 구조모형과 식 (1.2)의 측정모형을 갖는 구조식 모형을 고려하자. 측정모형에 대한 식 (1.2)를 행렬을 이용해 아래와 같이 표현할 수 있다.

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \Lambda_y & \mathbf{O} \\ \mathbf{O} & \Lambda_x \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{bmatrix}. \quad (1.9)$$

이 때, $\mathbf{x}^* = (\mathbf{y}, \mathbf{x})^T$, $\mathbf{\Lambda} = \begin{bmatrix} \Lambda_y & \mathbf{O} \\ \mathbf{O} & \Lambda_x \end{bmatrix}$, $\boldsymbol{\xi}^* = (\boldsymbol{\eta}, \boldsymbol{\xi})^T$, $\boldsymbol{\delta}^* = (\boldsymbol{\varepsilon}, \boldsymbol{\delta})^T$ 라 하면 식 (1.9)는 다음과 같이 다시 나타낼 수 있다.

$$\mathbf{x}^* = \mathbf{\Lambda}\boldsymbol{\xi}^* + \boldsymbol{\delta}^*, \quad (1.10)$$

여기서 $E(\boldsymbol{\xi}^*) = \boldsymbol{\kappa}^*$, $\text{Cov}(\boldsymbol{\xi}^*) = \boldsymbol{\Phi}^*$, $\text{Cov}(\mathbf{x}^*) = \boldsymbol{\Theta}^*$ 일 때, $\boldsymbol{\Xi} = \boldsymbol{\xi}^* - \boldsymbol{\kappa}^*$, $\mathbf{X} = \mathbf{x}^* - \mathbf{\Lambda}\boldsymbol{\kappa}^*$, $\boldsymbol{\Delta} = \boldsymbol{\delta}^*$ 라 하면 식 (1.10)은 아래의 단순 선형회귀식으로 표현된다.

$$\mathbf{X} = \mathbf{\Lambda}\boldsymbol{\Xi} + \boldsymbol{\Delta}. \quad (1.11)$$

크기가 n 인 \mathbf{X} 의 관측표본 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ 을 얻었다고 하자. $(1/n) \sum_{i=1}^n \boldsymbol{\Xi}_i \boldsymbol{\Xi}_i^T = \boldsymbol{\Phi}^*$ 라는 제약조건 하에서 가중된 오차 제곱합인 $\sum_{i=1}^n (\mathbf{X}_i - \mathbf{\Lambda}\boldsymbol{\Xi}_i)^T \boldsymbol{\Theta}^{*-1} (\mathbf{X}_i - \mathbf{\Lambda}\boldsymbol{\Xi}_i)$ 을 최소화하는 $\boldsymbol{\Xi}$ 를 추정하면 다음과 같

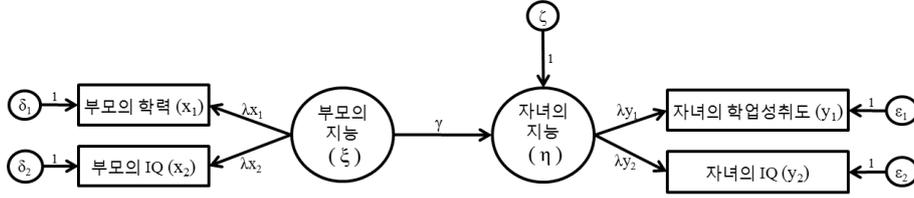


Figure 2.1. Example of an inappropriate case where the exogenous latent variable is affected by the endogenous latent variable

이 계산된다.

$$\hat{\Xi}_i = \left(\Lambda^T \Theta^{*-1} \Lambda + L \right)^{-1} \Lambda^T \Theta^{*-1} \mathbf{X}_i, \quad (1.12)$$

여기서 L 은 $\sum_{i=1}^n \Xi_i \Xi_i^T - n\Phi^*$ 에 해당하는 라그랑지 승수이다. 이후 상세한 계산과정을 생략하고 i 번째 개체의 잠재변수 점수를 정리해보면,

$$\hat{\xi}_i^* = \kappa^* + \left(\Lambda^T \Theta^{*-1} \Lambda + L \right)^{-1} \Lambda^T \Theta^{*-1} \mathbf{X}_i \quad (1.13)$$

와 같다 (Jöreskog, 2000).

실제 데이터를 이용하여 LISREL에서 식 (1.13)의 잠재변수 점수를 얻고자 할 때에는 프로그램 구문을 작성할 때 구조식 모형의 경로를 정의한 후에 다음의 명령을 추가해야 한다.

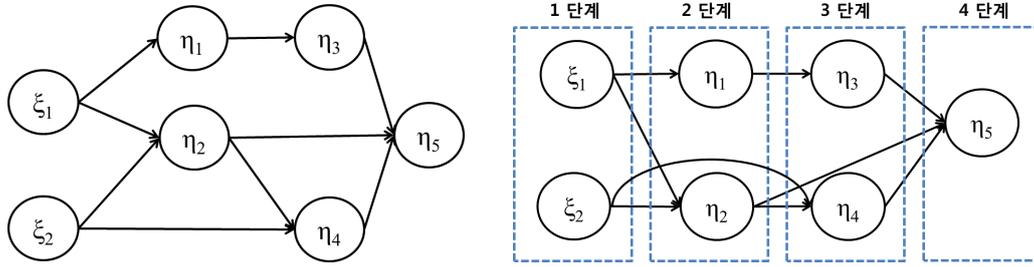
LSFfile <filename.lsf>,

여기서 <filename.lsf>는 데이터를 불러오는 명령에 사용된 LSF 데이터 파일명과 동일해야 한다. 이외의 LISREL 프로그램 구문에 대한 기본적인 설명은 LISREL 프로그램 제공업체인 Scientific Software International 홈페이지(<http://www.ssicentral.com/lisrel/examples.html>)를 참고하면 된다. 작성된 프로그램 구문을 실행하고 나면 데이터 파일이 있는 동일한 폴더에 'new'라는 단어가 추가된 <filenameew.lsf>라는 새로운 파일이 생성되는데, 이 파일에는 기존 데이터와 함께 잠재변수 점수가 마지막 열로 추가되어 있다.

2. 단계별 구조식 모형(Stagewise Structural Equation Modeling)

잠재변수 점수의 계산과정을 살펴보면 외생잠재변수 ξ 가 내생잠재변수의 관측변수 y 의 영향을 받을 수 있다. 이러한 관계가 현실적으로 타당하지 않은 경우가 있는데, Figure 2.1을 통해 그러한 예를 확인할 수 있다. Figure 2.1의 경로도는 부모의 학력과 IQ에 의해 관측된 부모의 지능이 자녀의 지능에 영향을 미치는 형태를 나타낸 구조식 모형이다. 이 때, 자녀의 지능을 관측하는 변수 중 하나인 자녀의 학업성취도가 부모의 학력과 IQ에 의해 이미 결정된 부모의 지능에 영향을 주는 것은 이치에 맞지 않다. 이를 수식으로 확인해보면 Figure 2.1의 경로도에 해당하는 식 (1.8)의 불일치 함수를 생각할 때, 이 불일치 함수가 S 와 Σ 의 영향을 받는데, S 는 x_1, x_2 뿐만 아니라 y_1, y_2 가 포함된 분산과 공분산으로 구성되어 불일치 함수를 최소화하는 γ 의 값은 어떤 형태로든 어떤 크기이든 y_1, y_2 에 의해 영향을 받게 된다.

이 연구는 구조식 전체 모형을 단계별(stagewise)로 분할하여 분석하는 방법을 통해 내생잠재변수의 관측변수가 외생잠재변수에 영향을 미치는 경우를 해결해보고자 하였다. 구조식 모형 전체를 한번에 분석하는 기존의 방법을 WSEM(Whole Structural Equation Modeling)이라 하고, 여기에서 제안하는 단



(a) A path diagram before their stages are assigned

(b) A path diagram after their stages are assigned

Figure 2.2. Example of stage-allocation using the path diagram

계별로 모형을 나누어 분석하는 방법을 SSEM(Stagewise Structural Equation Modeling)이라 명명하기로 한다.

전체를 한번에 분석하는 WSEM에 비하여, SSEM은 분석 단계를 여러 번 거쳐야 하지만 각 단계에서의 모수 추정은 용이하다. 또한 모형의 복잡성을 감소시켜 해가 구해지지 않았던 모형에서도 해를 도출할 수 있을 것으로 기대된다.

단계별 구조식 모형을 수행하는 과정은 두 가지로 나누어 볼 수 있다. 첫째는 모든 잠재변수에 단계를 할당하는 것이고, 두 번째는 할당된 단계별로 구조식 모형을 재설정하여 분석하는 것이다.

2.1. 잠재변수에 대한 단계 할당

잠재변수에 대한 단계 할당이란, 아래 절차에 의해 각 잠재변수마다 단계를 할당하는 것을 의미한다.

- 절차 (1-1) 모든 외생잠재변수 $\xi_i, (i = 1, 2, \dots, n)$ 은 1단계로 할당한다.
- 절차 (1-2) k 단계에 할당된 잠재변수에 대한 직접내생잠재변수 중에서 간접내생잠재변수가 되지 않는 내생잠재변수를 $(k + 1)$ 단계로 할당한다.
- 절차 (1-3) 모든 잠재변수에 대해 단계가 할당될 때까지 절차 (1-2)를 반복한다.

Figure 2.2(a)는 Figure 1.1의 경로도에서 잠재변수와 그 관계만을 따로 표현한 것이다. 위에 제시된 절차에 따라 Figure 2.2(a)의 잠재변수에 단계를 할당하면 Figure 2.2(b)가 된다. 절차 (1-1)에 의해 외생잠재변수 ξ_1, ξ_2 를 1단계로 할당하고, 절차 (1-2)에 의해 ξ_1, ξ_2 에 대한 직접내생잠재변수인 η_1, η_2, η_4 중에서 간접내생잠재변수가 되지 않는 내생잠재변수 η_1 과 η_2 를 2단계로 할당한다. 동일한 과정에 의해 η_3 과 η_4 는 3단계로, η_5 는 4단계로 할당한다.

2.2. 단계별 구조식 모형 설정

- 절차 (2-1) 1단계에 할당된 잠재변수와 그에 해당하는 관측변수로 확증적 요인분석을 실시하여 모수를 추정하고 잠재변수 점수를 추정하여, 이를 1단계 잠재변수 점수라 한다.
- 절차 (2-2) k 단계에 할당된 잠재변수와 그에 해당하는 관측변수, 그리고 $(k - 1)$ 단계 잠재변수 점수를 사용하여 k 단계 구조식 모형을 설정한다.
- 절차 (2-3) k 단계 구조식 모형에서 k 단계 잠재변수 점수를 추정한다.

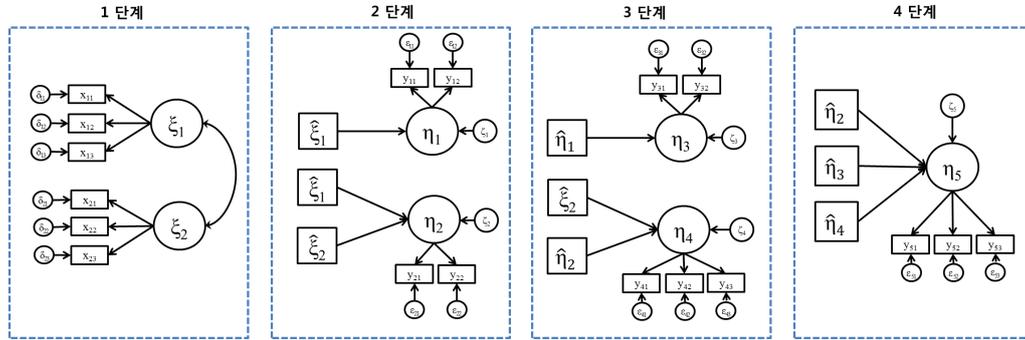


Figure 2.3. Example of designing a SSEM

절차 (2-4) 최종 단계까지 절차 (2-2)와 (2-3)을 반복한다.

Figure 2.3은 Figure 2.2(b)에서 단계가 할당된 잠재변수들에 대해 위에 제시된 절차에 따라 모형을 재 설정한 것이다. 1단계에 할당된 잠재변수 ξ_1, ξ_2 와 ξ_1 에 해당하는 관측변수 x_{11}, x_{12}, x_{13} 과 ξ_2 에 해당하는 관측변수 x_{21}, x_{22}, x_{23} 으로 확증적 요인분석을 실시하여 모수를 추정하고 잠재변수 점수 $\hat{\xi}_1, \hat{\xi}_2$ 를 추정하여 1단계 잠재변수 점수로 저장한다. 2단계에 할당된 잠재변수인 η_1, η_2 와 η_1 에 해당하는 관측변수 y_{11}, y_{12}, η_2 에 해당하는 관측변수 y_{21} , 그리고 1단계 잠재변수 점수 $\hat{\xi}_1, \hat{\xi}_2$ 를 사용하여 Figure 2.3의 두 번째 그림과 같은 2단계 구조식 모형을 설정하여 2단계 잠재변수 점수 $\hat{\eta}_1, \hat{\eta}_2$ 를 추정한다. 동일한 과정에 의해 마지막 4단계까지 구조식 모형을 분석한다.

3. 시뮬레이션 분석

이 절에서는 WSEM과 SSEM 두 가지 분석 방법을 통해 추정된 경로계수들과 분산, 잠재변수 점수를 비교하고자 한다. 이 때 SSEM을 수행할 때 반드시 필요한 값인 잠재변수 점수는 실제로는 관측 불가능한 값이므로, 시뮬레이션 데이터를 직접 생성하여 각 잠재변수의 참값을 구현하였고, 이를 통해 WSEM과 SSEM에서 추정된 모수를 참값과 비교하여 두 분석 방법을 평가하고자 하였다. 단계별 구조식 모형의 분석 및 잠재변수 점수 계산은 1.2절에서 설명한 방법을 이용하여 LISREL(9.1 버전, student edition)에서 수행하였다.

3.1. 시뮬레이션 데이터 생성

참 모형은 Figure 3.1과 같이 설정하였다. Figure 3.1에 나타난 모수의 참값을 수식으로 정리하면 식 (3.1)과 같다.

$$\begin{aligned}
 \lambda_{x1} &= \lambda_{y1} = \lambda_{y3} = 1, \\
 \lambda_{x2} &= 1.7, \quad \lambda_{x3} = \lambda_{y4} = 0.5, \quad \lambda_{y2} = 0.8 \\
 \theta_{\delta1} &= \theta_{\delta3} = 2, \quad \theta_{\delta2} = \theta_{\varepsilon1} = \theta_{\varepsilon4} = 4, \quad \theta_{\varepsilon2} = \theta_{\varepsilon3} = 3 \\
 \gamma_1 &= 1.3, \quad \gamma_2 = 0.5, \quad \beta = 0.7 \\
 \phi &= 2, \quad \psi_1 = 5, \quad \psi_2 = 3.5.
 \end{aligned} \tag{3.1}$$

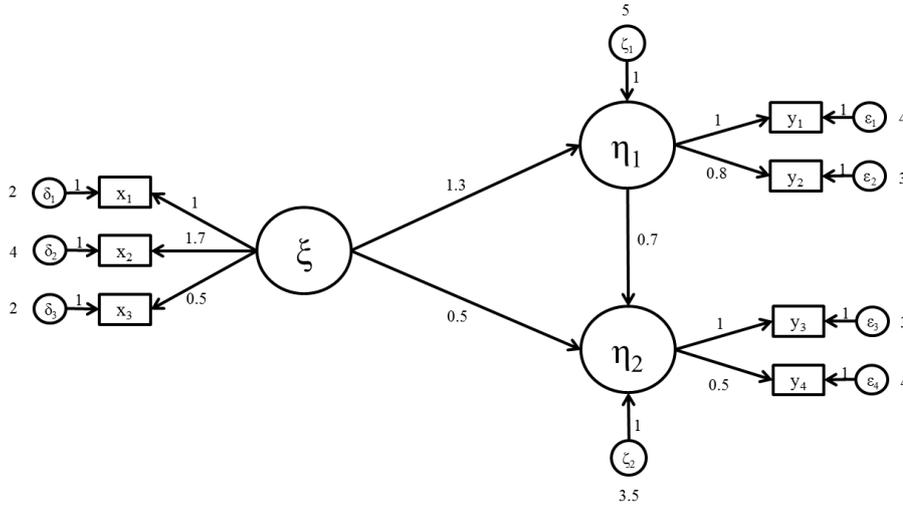


Figure 3.1. Path diagram of the SSEM used in simulation with given parameters

이 모형에서는 다음과 같이 7개의 관측변수 $\mathbf{x}_i^* = (y_{1i}, y_{2i}, y_{3i}, y_{4i}, x_{1i}, x_{2i}, x_{3i})^T$ 와 3개의 잠재변수 $\boldsymbol{\xi}_i^* = (\eta_{1i}, \eta_{2i}, \xi_i)^T$ 가 있으며 아래의 관측모형으로 표현된다.

$$\begin{aligned} y_{ti} &= \lambda_{yt}\eta_{1i} + \varepsilon_{ti}, & t = 1, 2, \\ y_{ti} &= \lambda_{yt}\eta_{2i} + \varepsilon_{ti}, & t = 3, 4, \\ x_{ki} &= \lambda_{xk}\xi_i + \delta_{ki}, & k = 1, 2, 3. \end{aligned} \quad (3.2)$$

이 때, ε_{ti} ($t = 1, 2, 3, 4$), δ_{ki} ($k = 1, 2, 3$)은 각각 독립적으로 $N(0, \theta_{\delta k}), N(0, \theta_{\varepsilon k})$ 을 따르며 $\boldsymbol{\xi}_i^*$ 와도 독립이다. 이 모형의 3개의 잠재변수 $\boldsymbol{\xi}_i^* = (\eta_{1i}, \eta_{2i}, \xi_i)^T$ 에 대한 구조모형은 다음과 같다.

$$\begin{aligned} \eta_{1i} &= \gamma_1\xi_i + \zeta_{1i}, \\ \eta_{2i} &= \beta\eta_{1i} + \gamma_2\xi_i + \zeta_{2i}, \end{aligned} \quad (3.3)$$

여기서 ξ_i 는 $N(0, \phi)$ 를 따르고, ζ_{ki} ($k = 1, 2$)은 각각 독립적으로 $N(0, \psi_k)$ 를 따르며 ξ_i 와 ζ_{ki} ($k = 1, 2$)는 독립이다.

표본크기에 따른 효율을 비교하기 위하여 $n = 50, 200, 1000, 2000$ 인 4개의 데이터셋을 구성하였다.

3.2. 결과

시뮬레이션 데이터를 생성할 때 설정한 모수의 참값과 추정치를 비교한 결과는 Table 3.1과 같다. 각 모수의 WSEM와 SSEM 추정치에 대하여 4개의 표본크기별 추정치 중 참값에 가장 근접한 값을 굵은 글씨로 표시하였다.

Table 3.1에 나타난 추정치들을 확인해본 결과, 두 가지 방법 모두 표본크기가 늘어남에 따라 추정치가 참값에 가까워지는 경향을 보였다. 표본크기가 작은 경우, WSEM에 비해 SSEM이 참값에 가까운 회귀계수 추정치가 도출되기도 하였다. 그러나 SSEM에 의한 외생잠재변수 ξ 의 분산 추정치와 오차 분산

Table 3.1. Estimates of parameters by WSEM and SSEM

		WSEM 추정치 (표준오차)				SSEM 추정치 (표준오차)				
		참값	$n = 50$	$n = 200$	$n = 1000$	$n = 2000$	$n = 50$	$n = 200$	$n = 1000$	$n = 2000$
회귀 계수	$\xi \rightarrow \eta_1$	1.3	1.462 (0.378)	1.584 (0.226)	1.377 (0.088)	1.335 (0.065)	1.196 (0.340)	1.233 (0.150)	1.209 (0.067)	1.160 (0.049)
	$\xi \rightarrow \eta_2$	0.5	0.149 (0.376)	0.865 (0.279)	0.396 (0.109)	0.376 (0.076)	0.400 (0.265)	0.479 (0.072)	0.510 (0.033)	0.560 (0.024)
	$\eta_1 \rightarrow \eta_2$	0.7	0.956 (0.198)	0.542 (0.114)	0.731 (0.058)	0.767 (0.040)	0.830 (0.131)	0.872 (0.154)	0.751 (0.067)	0.689 (0.048)
	$\xi \rightarrow x_2$	1.7	1.759 (0.316)	1.593 (0.198)	1.577 (0.087)	1.671 (0.069)	1.629 (0.334)	1.392 (0.318)	1.674 (0.141)	1.716 (0.109)
	$\xi \rightarrow x_3$	0.5	0.593 (0.132)	0.434 (0.090)	0.493 (0.382)	0.509 (0.030)	0.641 (0.137)	0.418 (0.101)	0.507 (0.043)	0.524 (0.033)
	$\eta_1 \rightarrow y_2$	0.8	0.766 (0.110)	0.758 (0.061)	0.779 (0.032)	0.785 (0.023)	0.858 (0.192)	0.761 (0.079)	0.790 (0.043)	0.784 (0.031)
	$\eta_2 \rightarrow y_4$	0.5	0.354 (0.072)	0.537 (0.070)	0.533 (0.029)	0.521 (0.021)	0.348 (0.075)	0.541 (0.073)	0.532 (0.031)	0.537 (0.022)
	분산	ϕ	2.0	2.985 (0.987)	1.910 (0.374)	2.233 (0.191)	2.103 (0.133)	3.084 (1.040)	2.171 (0.576)	2.104 (0.229)
η_1		5.0	8.046 (2.827)	5.324 (1.017)	4.582 (0.431)	4.858 (0.308)	8.481 (2.943)	6.778 (1.119)	5.620 (0.477)	5.870 (0.350)
η_2		3.5	1.867 (2.262)	2.544 (1.038)	2.511 (0.415)	3.258 (0.333)	4.953 (2.497)	3.850 (1.109)	4.367 (0.474)	4.811 (0.356)
x_1		2.0	2.027 (0.565)	1.733 (0.260)	1.850 (0.130)	1.903 (0.094)	1.928 (0.629)	1.472 (0.494)	1.979 (0.187)	1.969 (0.136)
오차 분산	x_2	4.0	2.743 (1.284)	3.723 (0.613)	4.219 (0.312)	4.251 (0.240)	3.797 (1.522)	4.360 (1.013)	3.877 (0.493)	4.121 (0.379)
	x_3	2.0	1.394 (0.321)	1.957 (0.206)	1.862 (0.091)	2.169 (0.074)	1.177 (0.312)	1.937 (0.211)	1.863 (0.094)	2.155 (0.076)
	y_1	4.0	7.768 (1.996)	2.906 (0.673)	3.708 (0.317)	3.840 (0.222)	9.304 (3.041)	2.948 (0.937)	3.828 (0.440)	3.833 (0.321)
	y_2	3.0	2.273 (0.865)	2.406 (0.423)	3.343 (0.222)	2.912 (0.147)	1.264 (1.786)	2.382 (0.567)	3.269 (0.292)	2.916 (0.205)
	y_3	3.0	3.227 (2.115)	4.813 (1.098)	3.643 (0.415)	2.949 (0.316)	2.941 (2.366)	4.885 (1.149)	3.636 (0.461)	3.250 (0.337)
	y_4	4.0	2.889 (0.630)	4.828 (0.561)	3.749 (0.200)	4.120 (0.154)	2.924 (0.647)	4.807 (0.569)	3.750 (0.208)	4.036 (0.158)

추정치들은 참값에 비해 과대 추정되는 경우가 많았다. 이는 모수를 추정하는 과정이 단계별로 분할됨에 따라 추정단계가 늘어나 발생하는 문제점으로 보인다.

표본크기에 상관없이 WSEM, SSEM 모두 회귀계수 추정치의 표준오차는 전반적으로 허용 가능한 크기로 나타났다. 그러나 표본크기가 작을 때, 오차 분산의 추정치에 대한 표준오차는 커지는 경우가 많았다. 특히, 표본크기가 50인 경우 SSEM에 의한 분산 추정치의 표준오차가 WSEM에 의한 분산추정치에 대한 표준오차보다 하나를 제외하고 모두 크게 나타났다. 그러나 표본크기가 커질수록 이러한 경향은 점점 줄어들어, 두 방법에 의한 추정치와 표준오차가 비슷해졌다.



Figure 4.1. Path diagram of the under-identified model in the 1st stage

전반적으로 비교해보면 전체 모형을 한 번에 고려하여 만든 데이터임에도 불구하고 SSEM을 적용하였을 때, WSEM에 비해 전혀 뒤쳐지지 않는 결과를 나타내었다. 모수 추정치의 절대적인 수치만 비교하였을 때에는 오히려 WSEM보다 참값에 더 근접한 결과를 보이기도 하였다. 이는 하나의 시뮬레이션 데이터에 대한 결과이므로 이것만으로 SSEM의 추정치가 WSEM의 추정치보다 더 낫다고 일반화할 수는 없으나, SSEM의 추정치도 잠재변수 간의 관계를 설명하는데 충분히 의미 있는 또 다른 추정치라고 판단된다.

4. 두 개 이하의 관측변수를 갖는 외생잠재변수

관측변수 x, y 의 표본분산-공분산행렬 S 를 이용하여 모수 θ 의 해를 추정할 때 모든 모수에 대하여 유일한 해가 구해지는 경우, 모형이 식별(identified)되었다고 정의하고 이 때의 구조식모형을 포화모형(saturated model) 또는 적정식별모형(just-identified model)이라고 말한다. 이와 달리 추정하고자 하는 모수의 수보다 표본분산-공분산행렬 S 의 정보의 수가 더 많은 경우에는 해가 유일하지 않고 하나 이상으로 나타나게 되는데 이를 과대식별(over-identified)되었다고 하며, 반대로 정보의 수가 모수의 수보다 적어 해를 구할 수 없는 경우를 과소식별(under-identified)되었다고 한다 (Hayduck, 1987). 모형이 식별되지 않은 경우에는 분석의 목적을 고려하여 모수에 제약을 부여함으로써, 추정해야 하는 모수의 수를 줄여 모형이 식별되도록 할 수 있다 (Hoyle, 1995).

SSEM은 모형의 복잡성을 감소시켜 해를 구하기 용이하다는 장점이 있으나, 1단계 구조식 모형에서 Figure 4.1의 경로도와 같이 ξ_i 의 관측변수가 두 개 이하인 경우는 추정하고자 하는 모수의 수에 비해 정보의 수가 부족하여 과소식별 문제가 발생하고 이로 인해 절차 (2-1)를 수행할 수 없게 된다. 과소식별 문제를 해결하기 위하여 관측변수가 하나일 때와 두 개일 때로 나누어 아래와 같이 모수에 제약을 부여한다.

관측변수가 하나인 경우의 외생잠재변수에 대한 측정모형을 수식으로 표현하면 아래와 같다.

$$x_1 = \xi + \varepsilon_1. \tag{4.1}$$

이 경우에는 관측변수 x_1 의 관측치를 1단계 잠재변수 점수 $\hat{\xi}$ 로 저장한 뒤, 절차 (2-2)를 수행한다. 이는 결과적으로 오차항 ε_1 의 오차 분산 $\theta_{\delta 1}$ 을 0(zero)으로 고정하는 것과 동일한 방법이다 (Hoyle, 1995).

관측변수가 두 개인 경우의 외생잠재변수에 대한 측정모형은 아래와 같이 표현되며, ε_1 과 ε_2 는 각각 독립적으로 $N(0, \theta_{\delta 1})$ 과 $N(0, \theta_{\delta 2})$ 를 따른다고 가정한다.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \xi + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}, \tag{4.2}$$

$$\begin{cases} x_1 = \xi + \varepsilon_1, \\ x_2 = \lambda\xi + \varepsilon_2. \end{cases} \tag{4.3}$$

이 때, 식 (4.3)에서 관측변수 x_1 과 외생잠재변수 ξ 에 대한 첫 번째 식은 다음과 같이 수정할 수 있다.

$$\xi = x_1 - \varepsilon_1. \quad (4.4)$$

이를 식 (4.3)에서 관측변수 x_2 에 대한 두 번째 식의 ξ 에 대입하면 다음과 같이 정리할 수 있다.

$$x_2 = \lambda(x_1 - \varepsilon_1) + \varepsilon_2 = \lambda x_1 + (\varepsilon_2 - \lambda \varepsilon_1). \quad (4.5)$$

이 때, $\varepsilon_2 - \lambda \varepsilon_1$ 을 새로운 오차항 ε^* 로 정의하면 식 (4.5)는 다음과 같이 정리되며 여기서 ε^* 은 $N(0, \lambda^2 \theta_{\varepsilon_1} + \theta_{\varepsilon_2})$ 를 따르는 것을 확인할 수 있다.

$$x_2 = \lambda x_1 + \varepsilon^*. \quad (4.6)$$

식 (4.6)에서 x_1 과 x_2 의 단순선형회귀분석을 수행하여 추정된 $\hat{\lambda}$ 을 식 (4.2)의 λ 의 값으로 대체한다. 이를 통해 추정해야 할 모수가 하나 줄어들어 모형 식별이 가능해지고 1단계 구조식 모형에 대하여 절차 (2-1)를 수행할 수 있다.

또한, 2단계 이상의 구조식 모형에서 잠재변수의 관측변수가 하나인 경우, 이전 단계의 잠재변수 점수를 외생잠재변수로 간주하여 해가 구해지지 않고 추가적인 제약조건을 부여할 것이 요구된다. 이 때에는 1단계에서 관측변수가 하나인 잠재변수에 대해 수행한 것과 동일한 방법으로, 그 관측변수의 측정값을 잠재변수 점수로 취급하여 단순선형회귀분석을 통해 계수를 추정한다. 그러나 하나의 관측변수로 잠재변수가 구성되는 것은 드문 경우이므로, 이 논문에서 제시하는 단계별 구조식 모형의 절차는 관측변수가 두 개 이상인 잠재변수에 대해 수행하는 것으로 국한한다.

5. 결론

잠재변수의 추정치인 잠재변수 점수는 구조식 모형 연구에 다양하게 활용될 수 있다. 그러나 전체 구조를 고려하여 산출되는 잠재변수 점수의 계산 과정을 수식을 통해 확인해보면 현실적으로 타당하지 않은 경우가 있다. 이 연구에서 제안된 단계별 구조식 모형 분석방법을 통해 잠재변수 점수를 순차적으로 계산하고 모수를 추정하는 방식으로 이러한 경우를 해결해 보고자 하였다.

모형을 한번에 분석하는 기존의 방법인 WSEM은 모든 변수의 전체 분산을 한꺼번에 고려하기 때문에 일반적인 상식에 위배되는 경로계수나 분산이 도출되는 경우가 있는데 단계별 구조식 모형을 통해 이러한 가능성을 낮출 수 있으며, 분석과정에서 생성되는 잠재변수 점수를 통해 추가적인 분석이 용이하다.

논문에 제시된 시뮬레이션을 통해 단계별 구조식 모형에서 계산된 모수 추정치들이 충분히 참값과 유사한 값을 가지고, 경우에 따라 기존의 WSEM에 의한 모수 추정치보다 참값에 더 근사함을 확인할 수 있었다. 따라서, 단계별 구조식 모형을 통해 도출된 모수 추정치는 WSEM에서 해가 도출되지 않는 경우의 대안이 될 수도 있지만, 그 값 자체로도 충분히 의미 있는 추정치라고 볼 수 있겠다. 다만, 모수의 추정 과정이 단계별로 분할됨에 따라 외생잠재변수의 분산 추정치와 오차 분산 추정치들이 참값에 비해 과대 추정되는 경우가 많았다. SEM에 의한 분산 추정치의 안정화를 위하여 추가적으로 다양한 방안을 모색할 예정이다.

단계별 구조식 모형은 전체 모형을 단계별로 분할하고, 분할된 모형을 새로 그려야 하는 번거로움이 있을 수 있다. 이런 경우 AMOS와 같은 그래픽 인터페이스의 구조식 모형 프로그램을 사용하기 보다는 LISREL이나 R 등의 구문 형태의 프로그램을 이용하면 간단하게 코드를 수정하는 것만으로 단계별 모형 구성 및 분석이 용이하다.

구조식 모형의 해를 구하는 방법에는 1.1절에 제시된 최대우도법, 일반화 최소제곱법, 비가중 최소제곱법 외에도 모수를 확률변수로 취급하는 베이지안 추정(Bayesian estimation)이 있다 (Lee, 2007). 구조식 모형에서 베이지안 추정을 할 수 있는 프로그램으로는 AMOS, WinBUGS 등이 있고, WinBUGS를 R에서 구현할 수 있는 프로그램으로 R2WinBUGS가 있다. 이 연구에서는 최대우도법을 이용하여 모형을 추정하였으나, 최소제곱법이나 베이지안 방법 등 다른 추정법을 사용하는 경우에도 동일한 과정을 통해 단계별 구조식 모형을 분석할 수 있다.

이 연구에서 제안된 방법이 구조식 모형의 연구에서 해를 용이하게 구하고 잠재변수 간의 관계를 규명하는데 한 가지 대안이 되기를 기대한다.

References

- Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis, *In Proceedings of the Third Berkeley Symposium*, **5**, 111–150.
- Bartholomew, D. and Knott, M. (1999). *Latent Variable Models and Factor Analysis : Kendall's Library of Statistics 7*, 2nd ed, John Wiley & Sons, London.
- Hayduck, L. A. (1987). *Structural Equation Modeling with LISREL : Essentials and Advances*, Johns Hopkins University Press, Baltimore and London.
- Hoyle, R. H. (1995). *Structural Equation Modeling : Concepts, Issues and Applications*, SAGE Publications, Thousand Oak.
- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis, *Psychometrika*, **32**, 443–482.
- Jöreskog, K. G. (1972). Factor analysis by generalized least squares, *Psychometrika*, **37**, 243–260.
- Jöreskog, K. G. (2000). Latent variable scores and their uses, available at <http://www.ssicentral.com/lisrel/advancedtopics.html>.
- Jöreskog, K. G. and Sörbom, D. (1999). *LISREL 8 User's Reference Guide*, Scientific Software International.
- Jöreskog, K. G., Sorbom, D. and Wallentin, F. Y. (2006). Latent variable scores and observational residuals, available at <http://www.ssicentral.com/lisrel/advancedtopics.html>.
- Lawley, D. N. and Maxwell, A. E. (1962). Factor analysis as a statistical method, *Journal of the Royal Statistical Society*, **12**, 209–229.
- Lee, S. (2007). *Structural Equation Modeling : A Bayesian Approach*, John Wiley & Sons.
- Lei, P. and Wu, Q. (2007). Introduction to structural equation modeling : issues and practical considerations, *Educational Measurement : Issues and Practice*, **26**, 33–43.

구조식 모형에 대한 단계적 접근

이보라^a · 박창순^{a,1}

^a중앙대학교 응용통계학과

(2014년 11월 10일 접수, 2014년 12월 18일 수정, 2015년 1월 7일 채택)

요약

최근 교육학, 경영학, 심리학 등 사회과학 뿐만 아니라 공정관리, 생물정보학 등 자연과학에서도 널리 사용되고 있는 구조식 모형(structural equation modeling)에서 잠재변수점수(latent variable score)는 직접 측정이 불가능한 잠재변수를 수량화한 추정치이다. 이 연구에서는 구조식 모형을 단계(stage)별로 분할하여 분석하는 단계별 구조식 모형(stagewise SEM; SSEM)을 제안하였다. 기존 방법은 모든 관측변수의 분산-공분산을 한꺼번에 고려하므로 독립변수인 외생잠재변수(exogenous latent variable)가 종속변수인 내생잠재변수(endogenous latent variable)에 의해 영향을 받는, 논리적으로 타당하지 않은 경우가 있다. 단계별 구조식 모형은 이런 문제점을 해결할 뿐만 아니라 모형의 복잡성을 낮추어 쉽게 해를 찾을 수 있으며, 분석과정에서 생성되는 잠재변수점수로 추가 분석도 용이하다.

주요용어: 단계별 구조식 모형, 잠재변수 점수, 외생잠재변수, 내생잠재변수.

이 논문은 2012-2014년도 중앙대학교 연구장학기금 지원에 의한 것임.

¹교신저자 : (156-756) 서울특별시 동작구 흑석로 84, 중앙대학교 응용통계학과. E-mail: cspark@cau.ac.kr