

구조방정식 모형에서 정규성 가정 위배 시 ML의 대안 탐색*

신 은 경 김 수 영[†]

이화여자대학교 심리학과

구조방정식 모형을 추정할 때 일반적으로 사용되는 최대우도 방법은 자료가 정규분포를 따른다는 가정에 기반하고 있다. 그러나 심리학을 포함한 사회과학 분야에서 정규성 가정이 위배되는 사례가 빈번히 보고되고 있으며, 이러한 상황은 추정 결과에 편향을 초래하여 통계적 추론의 타당성을 저하시킨다. 이에 정규성 가정이 위배된 상황에서도 신뢰할 수 있는 결과를 주는 여러 대안적인 방법들이 탐색되어 왔으나, 방법별 수행도가 연구마다 일관적이지 않아 적절한 추정 방법을 선택하기 위한 기준이 명확하지 않은 상황이다. 따라서 본 연구는 정규성 가정 위배 시 발생하는 문제에 대응할 수 있는 대안적인 방법을 정리하고, 방법별 수행도를 비교·제시하기 위해 지난 30여 년간의 관련 연구를 통합하여 연구자들이 실질적으로 참고할 수 있는 지침을 제안하고자 한다. 먼저, 최대우도 방법에서 정규성 가정의 의미와 가정 위배가 추정 결과에 미치는 영향을 설명한다. 다음으로, 정규성 가정이 위배된 상황에서 활용 가능한 다양한 방법들을 소개하고, 이를 방법이 비정규성에 대응하는 원리를 논의한다. 나아가, 기존 연구들을 체계적으로 탐색한 후 연구 결과를 조건별로 분류하고, 이를 표와 그림으로 시각화하여 각 방법의 수행도를 비교하고 논의한다. 마지막으로, 위에서 논의된 내용을 종합한 가이드라인을 제공하면서 본 연구의 의의와 한계에 관해 논한다.

주요어 : 정규성, MLM, MLR, 부트스트랩, 베이지안

* 이 논문은 2024년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2024S1A5C3A03046593).

[†] 교신저자: 김수영, 이화여자대학교 심리학과, 서울시 서대문구 이화여대길 52

Tel: 02-3277-3792, E-mail: suyoung.kim@ewha.ac.kr

Copyright © 2025, The Korean Psychological Association. This is an open-access article distributed under the terms of the Creative Commons Attribution –NonCommercial Licenses(<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

구조방정식 모형은 관찰되는 지표변수로부터 잠재변수를 측정하고 측정된 잠재변수 간의 관계를 추정하는 분석 방법으로(Bollen, 1989; Jöreskog, 1969), 잠재적인 특성 변수를 주로 사용하는 교육학 및 심리학 등의 사회과학 분야에서 활발하게 사용되고 있다(Kline, 2016). 이러한 구조방정식 모형을 추정하고자 할 때 일반적으로 최대우도(maximum likelihood, ML) 추정을 이용한다. 최대우도 방법은 모형과 자료가 주어졌을 때 자료의 발생확률을 최대화시키는 모수, 즉 연구자의 자료를 가장 가능성 있게 만드는 모수를 추정하는 것을 목적으로 한다(김수영, 2016). 최대우도 방법을 이용하면 추정 결과인 모수 추정치는 점근적으로 불편향적이고 일관적이며 효율적인 특성을 갖게 된다(Bollen, 1989; Jöreskog, 1969). 이러한 특성은 최대우도 방법의 주요 가정 중 하나인 정규성 가정이 전제되어야 확보될 수 있는데(Browne, 1984; Browne & Shapiro, 1988), 사회과학 분야에서 사용되는 변수들은 많은 경우에 정규분포를 따르지 않기에 정규성 가정을 완전히 만족하기 어렵다(Micceri, 1989; Wang & Wang, 2020). 정규성 가정이 위배된 조건에서 최대우도 방법을 이용할 경우 추정 결과에 편향 등의 문제가 발생하고, 위배 정도가 심각해질수록 편향의 정도가 커지는 것으로 알려져 있다(Curran et al., 1996). 이러한 문제에 대응하기 위하여 기존의 최대우도 추정에 수정을 가하는 방법이나 부트스트랩(bootstrap)과 같은 대안적인 방법들이 제안되었다(Bollen & Stine, 1992; Efron, 1979; Satorra & Bentler, 1994). 이와 같은 여러 방법들이 제안되었음에도 불구하고, 각 방법의 강건한 수행 정도는 가정의 위배 정도와 시뮬레이션의 진행 조건에 따라 다르게 나타났다(Curran et al.,

1996; Yu, 2002). 이처럼 연구에 따라 일관적이지 않은 수행도(performance)를 보이기 때문에 연구자들이 일부의 단편적인 결과만을 참고하는 것만으로는 적절한 추정 방법이 무엇인지 판단하기 어렵다. 이에 본 연구는 정규성 가정이 위배된 상황에서 사용할 수 있는 대안적인 추정 방법들을 논의하여, 연구자의 주어진 상황에 따라 적절한 추정 방법을 선택할 수 있도록 정보를 제공하고자 한다. 이를 위하여 비정규성 조건에서 다양한 추정 방법의 수행도를 살펴본 해외 연구들을 탐색한 후, 가정의 위배 정도에 따른 수행 결과를 체계적으로 정리하여 살펴본다.

모형의 추정은 연구자가 설정한 모형을 분석하는 과정의 핵심적인 단계 중 하나이다. 이 단계에서는 연구자가 설정한 모형의 모수를 추정하는데, 관찰된 공분산 행렬과 추정된 공분산 행렬 간의 차이를 최소화하는 모수를 찾는 것이 일반적이다(Wang & Wang, 2020). 이와 같은 추정의 과정에서 적절한 모수를 찾기 위해서는 어떤 방법을 선택하느냐가 매우 중요한데, 이는 추정 방법의 선택에 따라 모수 및 표준오차 추정치와 전체적인 모형 적합도 지수 결과가 달라지기 때문이다(Lei & Wu, 2012). 일반적으로 구조방정식의 영역에서는 최대우도 방법이 가장 흔하게 사용되며(Hoyle, 2000; Wang & Wang, 2020), 모수를 안정적으로 추정하고 적합도를 평가하기 위한 검정통계량을 제공한다는 점에서 오랜 기간 사용되어 왔다(Bollen, 1989). 최대우도 방법을 사용하기에 앞서, 안정적인 결과 산출을 위해 연구자는 추정 방법이 전제하고 있는 가정을 만족하는지 확인해야 한다(Finney & DiStefano, 2013). 최대우도 방법의 핵심적인 가정 중 하나는 연속형 내생변수의 분포가 정규분포를

따른다는 것으로, 정규성 가정이 위배된 경우에는 추정 결과에서 편향의 문제를 보이는 것으로 알려져 있다(Bollen, 1989). 정규성 가정의 위배 정도가 커질수록 모형 적합도 검정을 위한 χ^2 검정통계량과 표준오차 추정치가 편향되는 결과를 보이며(Chou et al., 1991), 이러한 편향은 실수로 영가설을 기각하는 1종오류의 증가 문제로 이어지게 된다(Curran et al., 1996).

정규성 가정 위배 시 발생하는 문제에 대응하기 위해 여러 통계적인 접근이 제안되었다. 먼저, 정규성 가정이 기반된 방법에 수정을 가하는 접근으로 최대우도 추정을 수정하는 방법에는 Satorra와 Bentler(1994)가 제안한 MLM(ML with a mean adjustment) 및 MLMV(ML with a mean- and variance-adjustment)와 MLR(ML with robust standard errors) 등이 있다(Muthén & Muthén, 1998 - 2017; Wang & Wang, 2020). 특히, MLM과 MLR은 과대추정된 χ^2 검정통계량이 χ^2 분포에 잘 근사하도록 평균 조정을 하고 과소추정된 표준오차에 대한 교정을 제시하는 방법이라는 공통점이 있으며(Savalei, 2010), 정규성 가정이 위배된 조건에서 수행이 뛰어나다고 알려져 있다(Chou et al., 1991; Curran et al., 1996; Grønneberg & Foldnes, 2019; Hu et al., 1992; Lai, 2018). 이에 본 연구는 두 가지 방법을 수정된 최대우도(corrected ML) 방법으로 지칭하여 함께 다룬다.

다음으로 분포에 대한 가정을 하지 않는 부트스트랩이 있다. 부트스트랩은 연구자의 표본을 거짓 표집단(pseudo population)으로 가정하고, 이를 기반으로 생성한 경험적 표집분포를 추론에 이용하는 재표집(resampling) 방법이다(Hancock & Liu, 2012). 연구자의 표본이 표집단과 같은 형태의 점수 분포를 가져야 한다

는 가정 외에 다른 가정이 없으므로(김수영, 2016; Kline, 2016), 비정규성 조건에서 사용이 용이하다는 장점이 있다(Nevitt & Hancock, 2001). 다만, 부트스트랩은 표집 방법이라는 특성상 추정 방법의 고유한 정의에 부합하기 보다는 가설 검정의 대안적인 접근법으로 간주될 수 있다. 그러나 부트스트랩이 Mplus 등의 통계 소프트웨어에서 별도의 추정 방법(estimator)으로 구현되어 있으며, 특히 비정규성 조건 하에서 다른 방법과 구별되는 통계적 추론 결과를 산출한다는 점을 고려하여, 본 연구는 이를 추정 방법으로 다루어 논의한다.

마지막으로 최대우도 방법과 같은 빈도주의(frequentist) 접근의 한계에 대한 대안으로서 베이지안(Bayesian) 방법이 제안되었다(Levy & Mislevy, 2016). 베이지안은 빈도주의 접근과 달리 모수를 변수로 가정하고 모수의 분포를 추정하는 접근법으로(이지윤, 김수영, 2021), 이 분포는 자료의 특성을 반영하는 우도함수(likelihood function)와 모수에 대한 연구자의 믿음을 나타내는 사전분포(prior distribution)의 결합을 통해 형성된다(Gelman et al., 2013). 이때 자료에 대해서 특정한 분포를 따라야 한다는 가정을 요구하지 않으므로, 다양한 형태의 자료를 이용할 수 있다는 유연성이 주요 특징으로 강조된다(Kaplan, 2023). 특히, 자료의 비정규성은 추정 과정에 큰 영향을 미치지 않는 것으로 알려져 있으며(Depaoli, 2021), 이로 인해 정규성 가정이 충족되지 않는 상황에서의 적용 가능성성이 탐색되어 왔다(Lee & Song, 2004; Lüdtke et al., 2021).

대안적인 추정 방법이 제안됨에 따라 해당 방법들의 수행도를 살펴본 연구가 지난 30여 년간 활발하게 진행되었다(Chou et al., 1991; Curran et al., 1996; Falk, 2018; Maydeu-Olivares,

2017; Nevitt & Hancock, 2001). 예를 들어, Curran 등(1996)은 정규성 가정이 위배된 조건에서 최대우도 방법과 대안적인 방법의 수행도를 살펴보고자 하였으며, 강건한 방법으로는 분포적 가정을 완화하여 구체적인 분포 형태를 요구하지 않는 ADF(asymptotically distribution-free; Browne, 1984)와 MLM의 수행도를 비교하는 연구를 진행하였다. Nevitt과 Hancock(2001)은 Curran 등(1996)과 동일하게 연구 조건을 설정하였지만, 추정 방법의 종류를 다르게 하여 MLM과 부트스트랩 방법 간의 수행 비교에 초점을 두었다. 이러한 선행 연구들을 통해 대안적인 방법들의 전반적인 수행 양상을 파악할 수 있지만, 각각의 연구들이 진행한 시뮬레이션 결과는 일관적이지 않았다. 예를 들어, MLM의 수행도를 살펴본 일부 연구들은 다소 상반된 결과를 보였는데, 정규성 가정을 심각하게 위배한 조건일 때는 1종오류를 제대로 통제하지 못한다는 연구 결과(Yu, 2002)와 표본크기만 충분하다면 위배 조건과 상관없이 1종오류를 제대로 통제함을 보인 연구 결과(Nevitt & Hancock, 2001)도 있었다. 또한, 비정규성 조건 하에서 부트스트랩의 수행도는 여러 조건에 따라 그 결과가 다르게 나타났는데, Grønneberg와 Foldnes(2019)는 부트스트랩의 수행이 충분하지 않은 표본크기 조건에서도 우수하다고 보고한 반면, Ferraz 등(2022)은 모형의 크기가 증가된 조건, 즉 변수의 수가 많은 조건에서는 부트스트랩이 보수적으로 기능함을 보였다. 이는 정규성 가정이 위배된 조건에서 수행 결과를 살펴보고자 할 때, 해당 시뮬레이션이 진행된 조건 또한 고려해야 함을 의미한다. 다시 말해, 선행 연구들은 자료 생성의 조건(예, 비정규성의 정도, 표본크기 등)이나 추정 방법의 종류를 서로

다르게 설정하였다는 점에서 이들이 내린 결론으로부터 일반적인 결과를 이끌어 내는 것이 어렵다는 한계가 있다. 이로 인해 연구자는 자신의 연구 조건과 자료 특성에 맞는 적절한 추정 방법을 선택하는 데 어려움을 겪을 수 있으므로, 정규성 가정 위배 조건에서 수행 결과를 직관적으로 파악하기 위한 체계적인 정리가 필요하다.

본 연구는 구조방정식 모형에서 정규성 가정이 위배되었을 경우에 사용할 수 있는 대안적인 추정 방법을 소개하고, 연구자들이 적절한 방법을 선택하는 데 참고할 수 있는 가이드라인을 제공한다. 정규성 가정이 위배된 조건에서 사용할 수 있는 추정 방법들이 제안됨에 따라 해당 방법들의 수행도에 대한 논의가 오랜 기간 진행되었음에도 불구하고, 각 방법들의 수행 결과가 일관되지 않고 시뮬레이션 조건도 서로 다르게 설정되어 어떤 추정 방법을 이용해야 할지 선택하는 것이 쉽지 않다. 이에 본 연구는 본격적인 논의에 앞서, 구조방정식 모형에서 일반적으로 사용되는 최대우도 방법과 정규성 가정을 간단히 소개하고, 정규성 가정이 만족되지 않을 때 나타나는 문제와 이에 대응 가능한 대안적 추정 방법들을 논의한다. 이어서 지난 수십 년간의 연구들을 탐색한 후 해당 연구들이 공통적으로 설정한 기준, 즉 정규성 가정 위배 정도, 표본크기 등의 기준에 따라 방법 간의 수행 결과를 분류한다. 분류한 결과를 시각화하여 수정된 최대우도, 부트스트랩, 베이지안 방법들의 수행도를 비교한다. 각 방법의 수행도는 먼저 1종오류 및 검정력 측면에서 비교하고, 상대편향(relative bias)을 확인하여 여러 조건에서 추정 방법의 수행에 대해 통합적으로 논의한다. 이후, 각 측면의 수행도 결과에 대한 논의를 종

합하여 추정 방법의 선택을 위한 가이드라인을 제공함으로써 연구자들의 적절한 의사결정을 돋는다.

최대우도 방법에서 이용하는 합치함수는 다음과 같다(Bollen, 1989; Jöreskog, 1969).

$$f_{ML} = \log|\Sigma(\hat{\theta})| + \text{tr}(S\Sigma^{-1}(\hat{\theta})) - \log|S| - p \quad (1)$$

최대우도 방법의 정규성 가정

구조방정식 모형을 추정하고자 할 때, 일반적으로 이용되는 최대우도 방법에 대해 간략하게 서술하고 최대우도 방법의 사용에 있어 전제가 되는 정규성 가정에 대해 논의한다. 이어서 이러한 정규성 가정이 위배된 상황에서 최대우도 방법을 이용할 경우, 모형 및 모수의 검정 단계에서 발생할 수 있는 문제점을 다룬다. 이때 정규성 가정이 위배된 경우는 정규성 가정을 만족하지 못하는 연속형 변수 이거나 이분형 자료(binary data)와 같이 본질적으로 정규성 가정이 위배된 범주형 자료(categorical data)인 경우로 나눌 수 있는데(Yu, 2002), 본 연구는 내생변수가 연속형인 구조방정식 모형을 중심으로 추정 방법에 대한 논의를 진행한다.

최대우도 방법

최대우도 방법은 구조방정식 모형의 추정에 가장 많이 사용되는 방법으로(Wang & Wang, 2020), 모형이 주어진 상태에서 자료의 발생확률을 최대화시키는 모수를 추정하는 것을 목적으로 한다. 이때 모수 추정의 핵심은 연구자의 자료인 표본 공분산 행렬과 설정한 모형을 통해 추정된 공분산 행렬의 차이를 최소화하는 것에 있다(Bollen, 1989). 합치함수(fitting function)를 이용하여 표본 공분산 행렬과 추정된 공분산 행렬의 차이를 수식화할 수 있는데,

위 식에서 $\Sigma(\hat{\theta})$ 은 연구자가 설정한 모형에서 추정된 공분산 행렬이고, S 는 표본 공분산 행렬이며, p 는 관찰변수의 개수이다. 이를 통해 변수 p 개로 구성된 모형에서의 공분산 행렬 $\Sigma(\hat{\theta})$ 과 자료에서의 공분산 행렬 S 간의 차이를 확인할 수 있으며, 행렬 간의 차이가 작다는 것은 연구자가 설정한 모형이 자료와 잘 부합함을 의미하게 된다. 또한, 합치함수가 최소화된 모수 추정치는 모수에 대한 일관적인 특성을 갖게 되고(Bollen, 1989; Browne, 1984), 결과적으로 최대우도 방법을 통해 산출되는 모수 추정치는 일정 가정을 만족한다면 점근적으로 효율적이고 불편향적인 특성을 갖는다(Bollen, 1989).

최대우도 방법의 장점은 모수를 안정적으로 추정할 뿐 아니라 이를 이용하여 모형 적합도를 평가하기 위한 검정통계량을 산출할 수 있다는 것에 있다(Bollen, 1989). 모형 적합도 검정을 위해서 우도비(Likelihood-Ratio, LR) χ^2 검정통계량을 이용하는데, 이는 아래와 같이 정의된다.

$$\chi^2 = nf_{ML} \quad (2)$$

통계 프로그램마다 제공하는 χ^2 검정통계량은 다를 수 있으나 일반적으로 χ^2 검정통계량은 식 2 혹은 $(n-1)f_{ML}$ 이 사용된다. 최대우도 방법은 모형이 자료에 완벽하게 부합한다는 영가설을 검정하는 χ^2 검정을 이용하며,

검정통계량의 표집분포가 χ^2 분포에 근사한다는 점근 이론을 바탕으로 한다. 이때 점근 이론이 성립하기 위해서는 관찰변수의 왜도와 첨도가 0에 가까우면서 표본크기가 충분히 크다는 가정이 만족되어야 한다(Bollen, 1989).

정규성 가정

최대우도 방법은 변수가 다변량 정규분포를 따른다는 가정하에서 추정의 안정성이 확보되므로 정규분포 기반 방법이라고도 불린다 (Yuan & Gomer, 2021). 이때 다변량 정규성 가정이란 이용하고자 하는 각 변수의 단변량 분포가 정규분포를 따르며, 임의의 두 변수 간 결합분포가 이변량 정규분포를 따른다는 것을 의미한다(Kline, 2016). 이러한 다변량 정규성은 필요조건 중 하나인 각 변수의 단변량 정규성을 통해 어느 정도 확인할 수 있으며, 일반적으로 왜도 및 첨도의 절대값 크기로 가정이 만족되었는지 판단한다(Curran et al., 1996; Kline, 2016). 단변량 분포의 왜도 및 첨도 기준에 대해서 Curran 등(1996)은 절대값이 각각 2와 7을 넘으면 문제가 발생한다고 하였으며, Kline(2016)은 절대값이 각각 3과 10을 넘어가는 경우 심각한 문제가 발생한다고 보았다. 이러한 논의에서 볼 수 있듯이, 정규성 가정의 위배 정도에 대한 합의된 기준이 존재하지 않으므로, 연구자들은 자료의 특성과 연구 상황에 적합한 기준을 선택하는 것이 일반적이다. 본 연구에서도 특정 기준값에 따른 수행도 변화를 살펴보기 위해 단변량 왜도 및 첨도의 절대값 크기를 정규성 가정 위배의 기준으로 설정하였다.

정규성 가정 위배 시의 문제점

정규성 가정이 확보되지 않은 상황에서 최대우도 방법을 이용한다면 모두 추정치 결과는 비교적 편향되지 않지만(Chou et al., 1991; Finch et al., 1997), 이를 토대로 수행하는 통계적 검정에서는 두 가지 주요 문제가 발생한다. 첫째, 모형 적합도 검정에 이용되는 χ^2 검정통계량이 과대추정됨에 따라 모형과 자료 간의 완벽적합 가설을 과도하게 기각하게 된다. 즉, 정규성 가정이 위배되었을 경우 χ^2 검정통계량이 χ^2 분포를 따르지 않는 문제가 발생하며 (Ferraz et al., 2022), 이는 실수로 영가설을 기각하는 1종오류를 높이는 통계적인 문제로 이어지는 것이다(Curran et al., 1996). 뿐만 아니라, 편향된 χ^2 검정통계량 문제는 근사적인 적합도 지수의 결과에도 영향을 미치는데, 특히 χ^2 값을 기반으로 추정되는 CFI, RMSEA 등에도 편향의 문제가 발생하는 것으로 알려져 있다(Gao et al., 2020; Yu, 2002). 둘째, 개별 모두의 유의성 검정 시 이용되는 표준오차가 과소추정되는 문제가 발생한다(Chou et al., 1991). 모두의 유의성을 검정하기 위한 통계량은 비표준화 모두 추정치 $\hat{\theta}$ 을 $\hat{\theta}$ 의 표준오차 $SE_{\hat{\theta}}$ 으로 나눈 값이다. 일반적으로 표준오차는 기대 정보행렬의 역함수와 표본크기를 이용하여 구하는데, 만약 정규성 가정이 위배되면 이러한 정보행렬에 편향이 발생한다(Yuan & Hayashi, 2006). 이로 인해 검정통계량의 계산에 편향이 생기고, 결과적으로 통계적인 검정에 영향을 미치게 된다. 즉, 개별모수가 유의하지 않음에도 유의하다는 결론을 내리는 거짓유의성(false positive) 문제가 발생할 수 있다.

대안적인 추정 방법 소개

정규성 가정 위배 시 대안적으로 사용할 수 있는 추정 방법은 여러 문헌들에 의해 소개되었다(Asparouhov & Muthén, 2005, 2010; Beran & Srivastava, 1985; Bollen & Stine, 1992; Gelman et al., 2013; Satorra & Bentler, 1994; Yuan & Bentler, 2000). 그중 본 연구에서는 대표적으로 이용되는 수정된 최대우도 방법과 부트스트랩 및 베이지안 추정을 다룬다. 세 개의 대안적 추정 방법에 대한 간단한 소개와 각 방법이 정규성 가정 위배에 대응하는 방식에 관해 논의한다. 특히 앞서 소개한 일반적인 최대우도 방법과의 차이를 중심으로 각 방법이 모형 및 모두 검정에 발생하는 문제에 대해 어떻게 접근하는지 살펴본다.

수정된 최대우도 방법

비정규성 문제를 해결하기 위해 χ^2 검정통계량과 표준오차를 재보정하는 수정된 최대우도 방법이 제안되었다. 일반적으로 가장 많이 사용되는 수정된 최대우도 방법은 MLM과 MLR이다(Wang & Wang, 2020). MLM은 Satorra 와 Bentler(1994)가 제안한 방법으로서 평균 보정된(mean scaled) χ^2 검정통계량과 강건한 표준오차를 제공하며, MLR은 점근적으로 Yuan-Bentler 검정통계량에 근사하는 검정통계량과 강건한 표준오차를 제공하는 방법이다(Asparouhov & Muthén, 2005; Muthén & Muthén, 1998-2017). 두 방법 모두 χ^2 분포를 따르도록 평균 보정된 χ^2 검정통계량을 제공하고(Yuan & Bentler, 2000), 샌드위치 타입의 공분산 행렬(sandwich-type covariance matrix; Bentler, 1983;

Browne, 1984)을 이용하여 강건한 표준오차를 산출한다는 공통점이 있는(Yuan & Hayashi, 2006) 반면, 검정통계량과 표준오차를 계산하는 데 있어 다른 종류의 정보행렬을 이용한다는 차이가 있다(Savalei, 2010).¹⁾ 이에 따라 MLM과 MLR은 점근적으로는 동일한 χ^2 검정통계량과 표준오차 결과를 보이지만, 표본크기가 작을 때는 서로 다른 결과를 산출하는 것으로 알려져 있다(Maydeu-Olivares, 2017).²⁾ 또한, 두 추정 방법의 결과를 비교했을 때, MLR은 점근적으로 비정규성과 모형 설정 오류가 있는 모형에도 강건하다는 점에서 MLM보다 더 일반적인 형태로 볼 수 있다(Lai, 2018). 앞서 다룬 평균 보정뿐 아니라 평균과 분산 모두 보정(mean-variance adjustment)된 χ^2 검정통계량과 강건한 표준오차를 제공하는 MLMV의 수행 역시 우수하다고 보고된 바 있으나(Ferraz et al., 2022), 다양한 조건에서 해당 방법의 수행도에 대한 연구는 아직 충분하지 않다(Lei & Wu, 2012).

먼저, 비정규성 조건 하에서 과대추정된 χ^2 검정통계량은 다음과 같이 척도 보정계수(scaling correction factor) c 를 이용해서 수정된다.

-
- 1) MLM은 기대 정보행렬을 기본으로 이용하며, MLR은 관찰 정보행렬을 기본으로 이용한다 (Maydeu-Olivares, 2017).
 - 2) 일반적으로 구조방정식 모형 추정 시 정보행렬의 종류와 상관없이 일관성 있는 표준오차 결과가 산출된다는 점에서 정보행렬의 구별에 큰 차이를 두지 않지만(Yuan & Hayashi, 2006), 표본크기가 충분히 크지 않은 조건에서 비정규성이 탐지된 경우, 이러한 정보행렬 간의 차이가 모형 및 모두 검정의 통계적인 결과에 영향을 미치는 것으로 알려져 있다(Savalei, 2010).

$$scaled \ \chi^2 = \frac{\chi^2_{ML}}{c} \quad (3)$$

위에서 보정계수 c 는 변수들의 다변량 첨도 크기를 반영하는 값으로(Hu et al., 1992), 비정규성으로 인해 c 가 1보다 큰 값을 갖게 됨으로써 과대추정된 χ^2 검정통계량의 문제를 상쇄시킬 수 있다. 즉, 검정통계량에 보정을 가함으로써 평균이 점진적으로 χ^2 분포의 평균과 같아지며, 보정된 검정통계량으로 모형 적합도를 판단할 수 있게 된다(Bollen, 1989).

다음으로, 추정치의 표준오차가 과소추정되는 문제는 샌드위치 공분산 행렬을 이용함으로써 어느 정도 해결할 수 있다(Satorra & Bentler, 1994; Yuan & Hayashi, 2006). 샌드위치 공분산 행렬은 정보행렬의 역행렬, 즉 기존의 공분산 행렬을 가운데에 두고 관찰변수 벡터의 외적행렬(outer product matrix of the score vector)이 양쪽으로 감싸고 있는 형태의 행렬로 (Savalei, 2010), 이 행렬을 기반으로 수정된 표준오차를 강건한 표준오차 혹은 샌드위치 표준오차라고 하며(Huber, 1967; White, 1980), 이는 MLM과 MLR을 이용하여 구할 수 있다 (Maydeu-Olivares, 2017). MLM을 이용한 결과는 일반적으로 사용되는 샌드위치 공분산 행렬 기반의 표준오차로 지칭하고, MLR을 이용한 결과는 Huber-White 표준오차로 지칭하기도 한다(Huber, 1967; Lai, 2019; White, 1980).

부트스트랩 방법

부트스트랩은 모집단의 분포에 대해 특정 형태를 가정하는 모수적 접근과 달리, 연구자의 표본을 가상의 모집단으로 가정하여 경험적인 표집분포를 형성하고 이를 검정에 이용

하는 재표집 방법이다(Hancock & Liu, 2012). 가상의 모집단으로 가정된 연구자의 표본으로부터 무선적으로 복원 추출하여 표집한 새로운 표본을 부트스트랩 표본으로 명명한다. 이 과정을 B 번 반복하여 B 개의 부트스트랩 표본을 생성한 후, 각 B 개의 부트스트랩 표본으로부터 총 B 개의 추정치를 구한다. 이때 추정치는 일반적으로 최대우도 방법으로 산출되며, 각 부트스트랩 표본의 추정치로 형성된 경험적인 표집분포를 토대로 모집단의 모수에 대해 추론하게 된다.

부트스트랩은 추정치의 표집분포를 구할 때 이론적인 표집분포의 근사 가정에 의존하지 않고 경험적인 부트스트랩 표집분포를 이용한다(Nevitt & Hancock, 2001). 이에 어떠한 분포적 가정을 하지 않고도 비정규성과 같은 통계적인 문제를 다룰 수 있게 된다(Yung & Bentler, 1996). 이때 모형과 모수를 검정하기 위해 사용되는 부트스트랩으로는 비모수 부트스트랩(nonparametric bootstrap; Beran & Srivastava, 1985; Efron, 1979)과 Bollen-Stine의 모형 기반 부트스트랩(model-based bootstrap; Bollen & Stine, 1992)이 있다. 두 방법은 기본적으로 동일한 부트스트랩 절차를 따르지만, 가상의 모집단으로 가정되는 표본을 활용하는 방식에 차이가 있다(Maydeu-Olivares, 2017). 비모수 부트스트랩은 표본 그 자체를 이용하여 추정치의 표집분포를 근사하는 반면, Bollen-Stine 부트스트랩은 원래 표본의 분포를 유지하면서도 모형과 완벽적합이 되도록 변환하는 과정을 포함한다(Bollen & Stine, 1992; Hancock & Liu, 2012). 변환하지 않은 표본을 이용하는 비모수 부트스트랩의 분포는 비중심 χ^2 분포를 따르기 때문에, 모형 적합도 검정에서 결과가 왜곡되는 문제가 발생한다(Bollen &

Stine, 1992). 따라서 Bollen-Stine 부트스트랩은 완벽적합이 가정된 표본을 통해 모형 적합도를 검정하게 된다(Beran & Srivastava, 1985; Bollen & Stine, 1992). 변환된 표본을 가상의 표집단으로 가정하고 동일한 부트스트랩 절차가 진행되어(Falk, 2018), B 개의 부트스트랩 표본을 생성한 후 B 개의 표본 추정치에 대해서 부트스트랩 검정통계량(\tilde{T}_{ML})을 계산할 수 있다. 이에 \tilde{T}_{ML} 의 분포가 모형 적합도 검정을 위한 표집분포가 되고, 부트스트랩 p 값은 \tilde{T}_{ML} 이 표집분포 상에서 표본에 기반한 검정통계량(T_{ML})보다 더 극단적일 확률로 정의된다. 일반적으로 이 값이 0.05보다 작다면 모형 적합도가 나쁘다는 결론의 통계적 근거로 사용될 수 있다(Hancock & Liu, 2012).

이처럼 부트스트랩 절차를 이용하여 부트스트랩 표집분포를 형성하면, 이어서 모수의 통계적 유의성을 확인할 수 있다. 일반적으로 연구자의 표본을 통하여 구해진 경험적 분포의 분위수(quantile)를 이용하여 신뢰구간을 추정하고, 이 신뢰구간이 검정하고자 하는 값(일반적으로 0)을 포함하고 있지 않다면 영가설을 기각한다(Yung & Bentler, 1996). 또는, 해당 분포의 표준편차인 부트스트랩 표준오차를 추정하여 모수의 유의성에 대한 추론(예를 들어, z 검정)을 진행할 수도 있다(Hancock & Liu, 2012).

베이지안 방법

베이지안 접근과 빈도주의 접근은 통계적 추론에 대해 근본적으로 다른 관점을 보인다. 접근적 표집이론을 중심으로 추론을 진행하는 빈도주의와 달리 베이지안은 베이즈 정리를

이용하여 연구자의 자료를 모수에 대한 사전정보와 결합하여 추론을 진행한다(Depaoli, 2021). 즉, 베이즈 정리를 통해 주어진 모수에 대한 연구자의 지식을 반영한 사전분포를 설정하고, 자료가 가지고 있는 모수에 대한 정보를 우도함수로 나타내어, 사전분포와 우도함수의 결합분포인 사후분포(posterior distribution)를 추정한다. 그런데 이러한 사후분포에 대한 직접적인 추론은 분석적인 어려움으로 인해 가능하지 않다. 따라서 무작위 추출을 반복하는 표집 방법인 마코프 체인 몬테 카를로(Markov Chain Monte Carlo, MCMC) 알고리즘을 이용하여 추론을 진행해야 한다(Kaplan & Depaoli, 2012). MCMC 방법은 우도함수를 명확히 정의하기 어려운 경우에도 사후분포를 효과적으로 추정할 수 있다는 장점이 있다(Yuan & Gomer, 2021). 특히, MCMC 방법 중에서도 대표적으로 이용되는 갑스 샘플러(Gibbs sampler; Geman & Geman, 1984)는 자료의 특성을 반영하는 우도함수의 형태나 표본크기 조건에 크게 제약받지 않고 사용될 수 있다(Scheines et al., 1999)는 점에서 분석의 유연성을 보여준다. 즉, MCMC 방법은 자료에 대해 특정 분포 형태의 가정을 요구하지 않고 점근이론에 의존하지 않으며, 비정규성 자료를 다루는 데에도 유용할 수 있다.

사후분포를 이용하여 추론하는 베이지안은 빈도주의 접근과 비교하여 모수 추정과 모형 적합도 검정 방식에 차이가 있다(Muthén & Asparouhov, 2012). 먼저, 베이지안과 빈도주의 접근은 모수를 취급하는 관점의 차이로 인해 모수 추정 방법이 다르다. 모수를 상수로 가정함에 따라 모수에 대해 점 추정 혹은 구간 추정을 하는 최대우도 방법과 달리, 베이지안 추정은 모수를 확률변수로 가정하고 이에 확

률분포를 할당하여 모수의 불확실성을 설명한다(van de Schoot et al., 2021). 즉, 추정의 결과인 사후분포를 그 평균이나 중앙값 등을 통하여 요약하거나, 신용구간(credibility interval)을 구하여 불확실성을 확인할 수 있다(Muthén & Asparouhov, 2012). 나아가, 이러한 요약된 통계치의 편향 정도를 평가함으로써, 빈도주의 접근처럼 모수 추정의 정확도를 검증할 수 있다(Lüdtke et al., 2021; Xu, 2019).

다음으로, 모형 적합도 평가를 위해서는 사후 예측 확인 절차(posterior predictive checking, PPC)가 활용된다(Asparouhov & Muthén, 2010; Gelman et al., 1996). 이는 모형이 관찰된 자료와 얼마나 부합하는지 평가하는 과정이며, 모형을 기반으로 생성된 자료와 관찰된 자료가 유사한지 확인하는 것이다(Gelman et al., 2013). 이때, 사후 예측 p 값(posterior predictive p -value, ppp ; Gelman et al., 1996)을 이용하여 실제 자료에 대한 검정통계량이 사후 예측 분포(posterior predictive distribution)에서 극단적인 값인지 평가함으로써, 모형이 실제 자료를 잘 설명하는지 판단할 수 있다. 사후 예측 분포는 모형을 기반으로 생성된 사후 예측 자료(posterior predictive data)의 통계량의 분포로, 모형 적합도 검정을 위한 표집분포로 사용된다(Levy & Mislevy, 2016). ppp 값은 사후 예측 분포하에서 반복적으로 생성된 자료에 기반한 통계량이 관찰된 자료 기반의 통계량보다 더 클 확률로 정의되며, 모형 기반으로 생성된 자료가 관찰된 자료와 유사할 때 모형이 적합하다고 평가할 수 있다(Gelman et al., 2013).³⁾

3) 모형과 자료가 완벽하게 부합하는 경우 ppp 값은 0.5일 것이고, 0.5에서 떨어진 값일수록 모형과 자료가 적합하지 않다는 것을 의미하게 된다.

대안적인 추정 방법의 수행도 비교

지금까지 비정규성을 가정한 다양한 시뮬레이션 연구를 통하여 대안적인 추정 방법들의 수행도가 평가되어 왔다. 여기서 주목할 점은 선행 연구들에서 각 방법의 수행도가 일관되지 않은 결과를 보였다는 것이다. 예를 들어, MLM의 수행도와 관련하여 Nevitt과 Hancock (2001)은 충분한 표본크기가 확보된 경우에는 정규성 가정 위배의 정도와 관계없이 1종오류를 적절히 통제할 수 있다고 보고한 반면, Maydeu-Olivares(2017) 및 Yu(2002)는 가정 위배 정도와 상관없이 1종오류 통제에 어려움이 있음을 지적하였다. 상반된 결과가 도출된 연구들을 자세히 보면, 자료 생성 조건(예, 자료 분포의 형태, 비정규성 정도, 표본크기 등)이 연구마다 상당히 달랐다는 것을 확인할 수 있었다. 이는 추정의 수행도를 평가한 연구 결과들을 비교하기 위해서는 다양한 시뮬레이션 조건을 충분히 반영한 포괄적인 논의가 요구된다는 것을 의미한다. 따라서 본 연구는 지난 수십 년간의 연구를 검토하여 공통된 기준을 설정하고, 그에 따라 각 추정 방법의 수행 결과를 체계적으로 분류한다. 이를 위해 정규성 가정 위배와 관련된 방법론 연구를 선별하고, 이들 연구를 바탕으로 각 방법이 특정 조건에서 보이는 수행 양상을 분석한다. 나아가, 연구의 주요 목적 중 하나인 조건별로 신뢰할 수 있는 결과를 도출하는 방법을 제안하기 위해, 추정 방법의 종류와 비정규성의 정도 등의 기준에 따라 연구 결과를 분류하고 이를 토대로 종합적인 결론을 제시한다. 분석에 이용된 연구의 시뮬레이션 설계는 요약되어 표 1에 제공되었으며, 각 연구에서 사용한 추정 방법의 종류는 본 논문의 주요 논의 범위에

표 1. 분석에 이용된 시뮬레이션 설계 요약

저자(연도)	추정 방법	모형		비정규성 수준		표본크기	반복 횟수	데이터 생성
		f	p	왜도	첨도			
Curran 등(1996)	ML, MLM	3	9	0	0	100	200	EQS (VM)
				2	7	200		
				3	21	500		
						1,000		
Ferraz 등(2022)	ML, BS	1 2 3	10 20 30	0 2 2	0 3 7	100 200 500 1,000	1,000	R (covism package)
Grønneberg & Foldnes(2019)	MLM, BS	3	11	0	0	100	2,000	R (VM)
				1	7	300		
				2	7	900		
Lai(2019)	ML, MLM, MLR, BS	6	18	-0.5 0 0.5 -1 0.5 1 -1.5 1.5 2	2~3 3~5 6~10	250 400 700 1,000	3,000	R (IG)
Liang & Yang(2016)	MLM, B	2 2	8 16	0 1.25 2	0 3.75 7	100 200 400	2,000	SAS (VM)
Maydeu-Olivares(2017)	ML, MLM, MLR	2 2	16 32	0 0 -2	0 2 3.18	200 500 1,000	1,000	Mplus
Nevitt & Hancock(2001)	ML, MLM, BS	3	9	0 2 3	0 7 21	100 200 500 1,000	200	GAUSS (VM)
Nevitt & Hancock(2004)	ML, MLM	7	21	0 0 3	0 6 21	129 258 645 1,290	2,000	GAUSS (VM)
Savalei(2010)	ML, MLM, MLR	2 4	16 16	0 2	0 7	100 200 300 400 500	1,000	EQS (VM)
Xu(2019)	MLR, B	3	9	0 1 2	0 3 7	200 500 200 1,000	200	R (VM)
Yu(2002)	ML, MLM	3	15	0 2 3	0 7 21	100 250 500 1,000	500	SAS (VM)

주. f는 요인(factor)의 개수, p는 지표변수(indicator)의 개수를 의미. VM은 Fleishman(1978; Vale & Maurelli, 1983) 방법, IG는 independent generator(IG) 변환법(Foldnes & Olsson, 2016)을 의미. 비정규성 수준이 범위로 작성된 경우 관찰된 표본 첨도를 의미. BS = 부트스트랩; B = 베이지안.

한정하여 제시되었다.

결과의 분류 기준

다양한 조건에서 수행된 연구 결과들은 크게 추정 방법의 종류, 정규성 가정의 위배 정도, 표본크기, 지표변수 개수의 네 가지 기준으로 분류될 수 있다. 먼저, 첫 번째 조건인 추정 방법의 종류는 앞서 소개한 수정된 최대 우도 방법(MLM, MLR), 부트스트랩(비모수 부트스트랩, Bollen-Stine 부트스트랩) 및 디폴트 사전분포가 지정된 베이지안을 고려하였다. 두 번째 조건인 정규성 가정의 위배 정도는 수용 가능한 기준에 대해서 제대로 된 합의가 부재하지만, 일반적으로 가장 많이 사용되는 왜도 2, 첨도 7인 조건(Curran et al., 1996)을 기반으로 구분하였다. 이에 따라 정규분포일 때(왜도=0, 첨도=0), 정규성 가정이 약간 위배되었을 때(0<왜도<2, 0<첨도<7), 정규성 가정이 중간 정도로 위배되었을 때(왜도=2, 첨도=7)로 나누었다. 본 연구의 정리 부분에서는 정규성 가정이 심각하게 위배된 조건은 다루지 않았는데, 이는 이러한 상황에서 단순히 추정 방법의 선택만으로는 해결할 수 없는 심각한 추정 문제가 발생할 수 있기 때문이다. 예를 들어, 비정규성이 심각한 변수가 포함된 구조방정식 모형을 분석한 김재철, 라종민(2018)은 분석 결과에서 분산 추정치가 음수로 나타나는 Heywood case와 같은 문제의 발생 가능성을 보고하였다. 이에 따라 정규성 가정의 위배 정도는 정규성 가정의 위배가 심각한 조건을 제외하고, 그 외의 세 가지 수준으로 구분하였다.

세 번째 표본크기 조건은 시뮬레이션 결과들을 바탕으로 표본크기가 작은 경우($n \leq 250$),

중간인 경우($250 < n \leq 500$), 큰 경우($n > 500$)의 세 수준으로 분류하였다. 마지막으로, 지표변수의 개수 조건이 고려되었는데, 이는 앞선 세 가지 기준과 달리 모형 적합도의 수행을 살펴볼 때만 적용된다. 적합도 검정통계량이 공분산 행렬의 크기에 의존하므로, 모형 적합도 결과의 평가에서 지표변수 개수 조건은 반드시 고려되어야 한다(Moshagen, 2012). 이에 따라 선행 연구를 바탕으로 지표변수 개수가 적은 경우(15개 미만)와 많은 경우(15개 이상)의 두 수준으로 구분되었다.

자료 분석의 준거

선행 연구들의 결론이 일관되지 않은 이유 중 하나는 각 연구에서 시뮬레이션 조건뿐만 아니라 평가 기준이 다르게 설정되었다는 점에 있다(Bandalos & Leite, 2013). 따라서 서로 다른 연구의 결과를 비교하기 위해 기준 결과의 서술 방식을 그대로 따르기보다는 비교 가능한 공통 지표를 설정하였다. 크게 모형 적합도와 모수 유의성, 두 가지 측면의 지표를 통해 각 추정 방법이 비정규성으로 인한 문제에 적절히 대응할 수 있는지 확인하였다.

먼저, 모형 적합도 측면에서 수행도는 시뮬레이션 기반의 연구들에서 공통적으로 보고되는 1종오류 발생률을 통해 판단되며, 이때 빈도주의와 베이지안은 평가 방식에서 차이를 보인다. 빈도주의 접근의 경우, 일반적으로 추정 방법의 모형 기각률을 바탕으로 1종오류 발생률을 확인한다. 1종오류는 모형이 자료와 적합함에도 불구하고 영가설을 기각해버리는 잘못된 의사결정을 내릴 확률을 의미하며, 일반적으로 Bradley(1978)가 제시한 유의수준 0.05에서의 준거(0.025~0.075)가 평가 기준으로 이

용되고 있다.⁴⁾ 경험적 1종오류 발생률이 이 구간에 속하는 경우, 1종오류가 통제되었다고 간주될 수 있다. 한편, 베이지안 접근의 경우에는 빈도주의 접근과 유사하게 모형 적합도를 판단하지만, p 값 대신 ppp 값을 사용한다는 차이가 있다. p 값과 달리 균일한 분포를 따르지 않는 ppp 값에 대해서는 통계적 판단이 가능한 이론적 기준값(cut-off)이 정해져 있지 않다(Cain & Zhang, 2018; Hjort et al., 2006). 그럼에도 불구하고 기준값으로 0.05나 0.10을 이용하는 것은 합리적이라고 제안되었고, 이를 바탕으로 ppp 값이 0.05 이하일 때 해당 모형이 자료와 적합하지 않다고 결론을 내릴 수 있다(Asparouhov & Muthén, 2010). 또한, 적합도를 보다 종합적으로 판단하기 위해 검정력을 함께 확인하는 경우도 있다(Curran et al., 1996; Liang & Yang, 2016; Nevitt & Hancock, 2001, 2004; Xu, 2019; Yu, 2002). 일반적으로 검정력은 1종오류의 통제 여부를 확인한 이후에 논의되는데, 이는 1종오류가 적절히 통제된 조건에서 신뢰할 수 있는 검정력 결과가 도출되기 때문이다(Yang et al., 2017).⁵⁾ 검정력은 자

4) 비정규성 조건에서 실제 1종오류 발생률이 유의수준보다 높거나 낮을 가능성이 있으므로, 단순히 하나의 1종오류 발생률이 아닌 적정 범위 기준의 사용이 권장된다(Ferraz et al., 2022; Maydeu-Olivares, 2017; Nevitt & Hancock, 2001, 2004).

5) Bradley(1978)의 준거에 따르면, 경험적 1종오류 발생률이 7.5%를 초과한 경우 해당 검정통계량을 이용한 검정력은 신뢰할 수 없으며, 이러한 조건에서는 1종오류가 적절히 통제된 다른 검정통계량과의 비교가 적절하지 않게 된다. 반면, 모형 기각률이 2.5% 미만일 때는 검정통계량이 여전히 허용 가능한 검정력을 제공할 가능성이 있기에 후속 검정력 분석에 활용될 수 있다(Nevitt & Hancock, 2004).

료와 적합하지 않은 모형을 제대로 기각함으로써 올바른 의사결정을 내릴 확률을 뜻하며, 사회과학 연구에서는 최소 80% 이상 확보하는 것이 추천된다. 이때, 자료와 적합하지 않은 모형이란 완벽적합을 가정한 참 모형(true model)과 달리, 연구자가 요인 구조를 잘못 설정하거나(Liang & Yang, 2016) 0이 아닌(nonzero) 요인부하량을 의도적으로 생략하여 직접 설정한 모형(Curran et al., 1996; Nevitt & Hancock, 2001, 2004; Yu, 2002), 혹은 RMSEA 등의 모형 부적합 지표를 이용하여 부적합 정도를 특정 값으로 설정한 모형(Xu, 2019) 등을 의미한다.⁶⁾

다음으로, 모두 유의성 측면에서 수행의 정확성은 표준오차의 상대적 편향을 통해 확인할 수 있다. 앞서 언급한 바와 같이, 최대우도 방법은 비정규성 조건에서도 비교적 편향되지 않은 모수를 추정할 것으로 기대되기 때문에, 선행 연구들은 각 방법이 표준오차 추정치를 얼마나 정확하게 산출할 수 있는지에 초점을 두어 살펴보았다(Lai, 2019; Maydeu-Olivares, 2017; Nevitt & Hancock, 2001; Xu, 2019). 표준오차 추정치의 상대적 편향은 모집단 표준오차와 추정된 표준오차 간의 차이로 평가되는데, 이때 추정된 표준오차는 이론적 분포나 통계모형을 따르도록 생성된 자료 세트를 기반으로 추정된 모수 추정치에 기반한다(Bandalos & Leite, 2013). 상대적 편향 비율이

6) 검정력의 명확한 해석을 위해서는 효과크기 제시가 필수적이나, 본 연구에서 검토한 5개의 연구는 이를 수치적으로 명확히 보고하지 않아 모형 설정 오류의 정도에 대한 구체적인 근거를 제시할 수 없었다. 반면, Xu(2019)는 RMSEA를 통해 모형 부적합 정도를 설정하였으며, 본 연구는 RMSEA를 0.24로 설정한 조건에서의 검정력 결과를 이용하였다.

10% 미만일 경우 허용 가능한 수준으로 간주되며(Hoogland & Boomsma, 1998), 본 연구에서도 표준오차 추정치의 상대적 편향에 대한 평가 지표에 일관되게 10% 기준을 적용하여 살펴본다.

준거에 따른 결과 비교

일정 기준에 따라 분류된 수행 결과는 자료 분석의 준거에 따라 정리되었다. 각 조건별 세부 결과를 정리할 때, 모든 연구에서 1종오류, 검정력 및 표준오차 추정치의 수행도가 일괄적으로 검토되지 않았으므로 조건마다 포함되는 결과의 개수가 다를 수 있다. 동일한 조건에 단일 결과만 있는 경우 해당 결과를 그대로 제시하고, 여러 결과가 포함된 경우에는 대표적으로 최소값과 최대값을 제시하여 결과의 전반적인 경향을 보이고자 한다. 이후부터 정규성 가정의 위배 정도에 따른 수행 결과를 순차적으로 제시하고, 각 조건에서 신뢰할 수 있는 결과를 도출하는 추정 방법에 대해 종합적으로 논의한다.

1종오류

각 방법의 수행도 차이를 비교한 10편의 시뮬레이션 연구는 15개의 모형에 대해 다양한 조건에서의 결과를 보고하였다(Curran et al., 1996; Ferraz et al., 2022; Grønneberg & Foldnes, 2019; Liang & Yang, 2016; Maydeu-Olivares, 2017; Nevitt & Hancock, 2001, 2004; Savalei, 2010; Xu, 2019; Yu, 2002). 해당 연구들의 1종오류 발생률 결과를 세 가지 조건에 따라 분류한 결과가 표 2에 제공된다. 이때 수정된 최대우도 방법은 MLM과 MLR의 통합된 결과를 의미하며, 두 방법의 결과 간 유의한 차이

가 없었기 때문에 이를 구분하지 않고 제시하였다.⁷⁾ 추정 방법에 따른 1종오류 결과를 살펴보면, 대부분의 조건에서 부트스트랩이 가장 낮은 1종오류 발생률을 보였으며, 수정된 최대우도 방법과 베이지안이 그 뒤를 따랐다. 이러한 결과는 정규성 가정 위배 정도, 지표 변수 개수, 표본크기 조건에 따라 다르게 나타났으며, 각 조건에 따른 결과의 양상을 순차적으로 살펴본다.

먼저, 정규성 가정이 만족된 상황(왜도=0, 침도=0)에서 표본크기 조건이 1종오류 발생률에 미치는 영향은 크지 않았으나 지표변수 개수에 따라서는 1종오류 결과의 양상이 달라지는 것을 확인할 수 있었다. 지표변수가 적은 경우에는 표본크기 조건에 상관없이 수정된 최대우도 방법과 부트스트랩의 결과는 대부분 기준을 충족하고, 베이지안의 결과는 상대적으로 1종오류의 하한 기준보다 낮게 나타났다. 반면, 지표변수가 많은 경우에는 수정된 최대우도 방법과 부트스트랩 모두 기준을 충족하지 못한 결과들이 대부분이었으며, 특히 수정된 최대우도 방법에서 1종오류 상한 기준 이상의 결과, 부트스트랩에서는 하한 기준 이하의 결과가 관찰되었다. 한편, 베이지안 방법에서는 오히려 기준에 부합하는 결과가 상대적으로 더 많이 발견되었다. 덧붙여, 표본크기에 따른 변화 양상은 수정된 최대우도 방법에서만 확인할 수 있었다.

다음으로, 정규성 가정의 위배 정도가 0<왜

7) MLM과 MLR의 차이가 뚜렷하게 나타난 경우는 결측치 조건을 포함시킨 연구(Savalei, 2010)에서만 관찰되었다. 해당 연구에서는 비정규 데이터와 결측치 조건(MCAR)이 동시에 존재할 때 MLM이 전반적으로 1종오류를 적절히 통제하지 못하였다.

표 2. 시뮬레이션 조건에 따른 각 추정 방법의 1종오류 비교

비정규성 정도	지표변수 개수	표본 크기	추정 방법					
			Corrected ML		BS		B	
			min	max	min	max	min	max
왜도=0	15개 미만	S	0.050*	0.095	0.023	0.049	<0.025*	0.063
		M	0.050	0.065	0.036	0.050	<0.025*	0.021
		L	0.067	0.080	0.043	0.059	<0.025*	
	15개 이상	S	0.076	0.350	0.000	0.007	0.007	
		M	0.062	0.110	0.012	0.030	0.059	0.060
		L	0.046	0.090	0.021	0.044	-	
0<왜도<2	15개 미만	S	0.108	0.125	0.010	0.035	<0.025*	0.143
		M	0.070	0.100*	0.039	0.053	0.000	0.099
		L	0.050*	0.063	0.033	0.047	<0.025*	
	15개 이상	S	0.130	0.940	0.000	0.014	0.123	0.186
		M	0.070	0.380	0.013	0.021	0.338	0.477
		L	0.056	0.190	0.018	0.041	-	
0<왜도<7	15개 미만	S	0.055	0.175*	0.023	0.040	0.025*	0.272
		M	0.065	0.094	0.030	0.068	<0.025*	0.255
		L	0.040	0.076	0.045	0.059	<0.025*	
	15개 이상	S	0.017	0.516	0.000	0.009	0.454	0.649
		M	0.015	0.130	0.009	0.022	0.749	0.888
		L	0.064	0.020	0.027	-		

주. Bradley(1978)의 기준(2.5~7.5%)에 부합하는 시뮬레이션 결과는 굵은 글씨체로 표시. min은 해당 조건에서의 최소값을, max는 최대값을 의미. 표본크기 조건의 경우, S는 250 이하, M은 250 초과 500 이하, L은 500 초과인 경우를 의미. *를 표시한 경우, 시뮬레이션 결과가 그림으로 제시됨에 따라 해당 그림의 눈금을 기반으로 대략적인 값 기재. Corrected ML = 수정된 최대우도 방법; BS = 부트스트랩; B = 베이지안.

도<2, 0<첨도<7인 상황에서도 정규성 가정이 만족된 상황과 유사하게 지표변수 조건에 따른 수행 차이가 확인되었다. 지표변수가 적은 경우에 부트스트랩은 표본크기와 관계없이 대 부분의 결과가 기준에 부합했으나, 베이지안에서는 기준을 충족하지 못하는 결과들이 주

로 관찰되었다. 수정된 최대우도 방법에서는 표본크기에 따라 기준을 충족하는 결과의 양상이 달라졌는데, 중간 이상의 표본크기에서 1종오류가 적절히 통제되는 것으로 나타났다. 반면, 지표변수가 많은 경우에는 모든 방법이 1종오류를 제대로 통제하지 못하는 경향

표 3. 시뮬레이션 조건에 따른 추정 방법의 검정력 비교

비정규성 정도	표본 크기	추정 방법					
		Corrected ML		BS		B	
		min	max	min	max	min	max
왜도=0 첨도=0	S	0.520	1.000	0.305	0.785	0.936	1.000
	M		1.000		1.000		1.000
	L		1.000		1.000		1.000
0<왜도<2 0<첨도<7	S	0.977	1.000			0.937	1.000
	M		1.000		-		1.000
	L		1.000				1.000
왜도=2 첨도=7	S	0.400	1.000	0.235	0.645	0.125*	0.999
	M	0.750	1.000		0.990	0.625*	1.000
	L		1.000		1.000		1.000

주. 검정력이 0.8 이상인 시뮬레이션 결과는 굵은 글씨체로 표시. min은 해당 조건에서의 최소값을, max는 최대값을 의미. 표본크기 조건의 경우, S는 250 이하, M은 250 초과 500 이하, L은 500 초과인 경우를 의미. *를 표시한 경우, 시뮬레이션 결과가 그림으로 제시됨에 따라 해당 그림의 눈금을 기반으로 대략적인 값 기재. Corrected ML = 수정된 최대우도 방법; BS = 부트스트랩; B = 베이지안.

을 보였다. 이는 특히 수정된 최대우도 방법에서 두드러졌으며, 표본크기가 작은 조건에서 최소값(0.130)과 최대값(0.940) 간의 차이가 크게 나타난 점을 예로 들 수 있다. 이때 최소값은 지표변수가 16개일 때, 최대값은 32개일 때의 결과로(Maydeu-Olivares, 2017), 동일한 조건임에도 지표변수가 많아질수록 1종오류 발생률이 높아질 수 있음을 시사한다. 이러한 결과는 표본크기와 지표변수 조건의 결합이 결과에 중요한 영향을 미칠 수 있다는 것을 보여주지만, 충분한 표본크기가 확보되면 이러한 경향은 완화되는 것으로 나타났다.

마지막으로, 정규성 가정의 위배 정도가 왜도=2, 첨도=7인 상황에서도 지표변수 조건에 따라 유의한 수준의 수행 차이가 나타났다. 지표변수가 적은 경우, 부트스트랩은 표본크

기에 상관없이 우수한 결과를 보인 반면, 수정된 최대우도 방법과 베이지안의 결과들은 일관되지 않은 경향을 보였다. 한편, 지표변수가 많은 경우에는 세 방법 모두 전반적으로 낮은 수행도를 보였으며, 특히 부트스트랩의 수행도가 크게 저하된 것을 확인할 수 있었다. 베이지안은 조건에 따른 변동성이 가장 크고 안정적인 수행도를 보이지 않았으나, 수정된 최대우도 방법은 표본크기가 일정 수준 이상인 조건에서 이러한 경향이 완화되는 것으로 나타났다.

요약하자면, 각 방법의 1종오류 발생률은 시뮬레이션 조건에 따라 차이를 보였다. 특히 지표변수 조건에 따라 유의한 수준의 차이가 나타났으며, 이러한 차이는 정규성 가정이 위배된 상황에서 더욱 두드러졌다. 이에 따라

정규성 가정이 위배된 상황에서 지표변수 조건에 따라 권장되는 방법도 달라졌는데, 지표 변수 개수가 적은 경우에는 표본크기와 상관 없이 부트스트랩의 사용이 추천되며, 지표변수 개수가 많은 경우에는 표본크기가 충분히 확보된 조건에서 수정된 최대우도 방법의 사용이 바람직할 것으로 판단된다. 한편, 베이지안은 가정의 위배 정도에 상관없이 1종오류를 적절히 통제하지 못하는 경우가 많고, 1종오류 발생률도 일관적이지 않은 것으로 나타났다. 다만, 분류에 이용된 선행 시뮬레이션 연구들이 서로 다른 프로그램을 사용했으며, 자료 세트의 수를 다르게 설정했다는 점에서 이러한 요인들이 결과에 영향을 미쳤을 가능성 있다.⁸⁾

검정력

시뮬레이션 연구에서는 1종오류 결과를 확인한 이후, 검정력의 측면에서 어떻게 수행되는지를 살펴보게 된다. 1종오류를 확인한 선행 연구 중에서 검정력까지 함께 검토한 연구는 충분하지 않았으며, 그중 지표변수 개수를 조건으로 설정한 연구도 거의 없었다. 이에 지표변수 조건을 분류 기준으로 이용하지 않고, 수행도를 확인한 일곱 개의 모형에 대한 결과를 정규성 가정 위배 정도와 표본크기 조건에 따라 표 3에 정리하였다(Curran et al., 1996; Liang & Yang, 2016; Nevitt & Hancock, 2001, 2004; Xu, 2019; Yu, 2002).

표의 내용을 살펴보면, 모든 방법이 대체로 기준 이상의 검정력을 보이는 것으로 나타났다. 정규성 가정이 만족된 경우, 수정된 최대

우도 방법과 부트스트랩은 표본크기에 따라 결과 차이를 보인 반면, 베이지안 방법은 표본크기와 관계없이 높은 검정력을 유지하는 경향을 보였다. 정규성 가정의 위배 정도가 $0 < \text{왜도} < 2$, $0 < \text{첨도} < 7$ 인 상황에서는 베이지안과 수정된 최대우도 방법 모두 표본크기와 상관없이 높은 검정력을 유지하였다. 부트스트랩의 경우, 검정력 측면에서 가정이 약간 위배된 상황에서의 수행을 확인한 연구가 없었으므로, 해당 방법의 성능은 확인할 수 없었다. 정규성 가정의 위배 정도가 $\text{왜도} = 2$, $\text{첨도} = 7$ 인 상황에서는 베이지안 방법과 수정된 최대우도 방법 모두 분류 조건에 따라 상이한 결과를 나타냈으며, 특히 베이지안을 이용한 결과는 표본크기가 작은 조건에서 일관되지 않은 경향을 보였다. 한편, 부트스트랩은 정규성 가정이 만족되었을 때와 유사하게, 표본크기가 일정 수준 이상 주어진다면 비교적 안정적이고 충분한 검정력이 확보되는 것으로 나타났다.

요약하자면, 검정력 측면에서 정규성 가정의 위배 정도와 표본크기 조건에 따라 수행 차이가 나타났다. 정규성 가정이 위배되고 표본크기가 일정 수준 이상일 때 베이지안과 수정된 최대우도 방법의 검정력 결과는 안정적으로 유지되었으며, 이러한 조건들에서 두 방법의 사용이 권장된다. 부트스트랩 방법도 표본크기가 충분한 경우 안정적으로 추정이 수행되지만, 검정력은 앞선 두 방법보다 약간 낮은 것으로 확인되었다. 따라서 동일한 조건에서 일정 이상의 표본크기가 확보된 경우, 수정된 최대우도 방법, 베이지안 그리고 부트스트랩 순으로 사용이 추천된다.

8) Liang과 Yang(2016)은 Mplus의 디플트 사전분포를 이용하고, Xu(2019)는 blavaan(R package)의 디플트 사전분포를 이용하였다.

표준오차 편향

전체적인 모형 적합도를 확인한 후에는 모수 유의성 검정에 사용되는 표준오차 추정치의 정확성을 검토하는 단계로 이어진다. 각 방법의 수행도를 비교하기 위해, 요인부하량 표준오차의 상대적 편향을 이용하며, 상대적 편향은 특정 요인부하량의 표준오차에 대해 산출되거나(Nevitt & Hancock, 2001; Xu, 2019), 모든 요인부하량 표준오차에 대한 평균값으로 산출된다(Lai, 2019; Maydeu-Olivares, 2017). 이때 기존 연구들은 표준오차의 상대적 편향을 평가하는 데 세 가지 수식을 사용하였다(Hoogland & Boomsma, 1998; Lai, 2019; Maydeu-Olivares, 2017; Nevitt & Hancock, 2001). 각 연구에서 도출된 결과는 서로 다른 수식으로 계산되었으므로, 결과들을 동일한 기준에서 일괄적으로 비교하기는 어렵다. 이에 본 연구에서는 Forero와 Maydeu-Olivares(2009)의 방법을 변형하여 이용한다. Forero와 Maydeu-Olivares(2009)는 여러 모수에 대해 표준오차의 상대적 편향을 10% 미만, 10%~20% 등으로 평가하여, 각 기준을 충족하는 비율을 통해 추정 방법의 성능을 평가하였다. 본 연구에서는 각 수식으로 계산된 요인부하량 표준오차의 상대적 편향 결과가 기준(10%)에 부합하는 비율로써 결과를 제시한다. 예를 들어, 100%라는 결과는 해당 방법으로 산출된 표준오차의 상대적 편향 결과가 모두 10% 미만임을 의미한다. 이를 통해 각 방법에서 산출된 표준오차 상대적 편향의 경향을 파악하고, 절대적인 수치보다는 상대적인 패턴을 중점적으로 살펴본다. 또한, 표준오차의 경우에 MLM과 MLR 간의 수행 차이가 명확하게 나타나 이를 구분하여 제시하였다. 표준오차 추정치의 정확도를 확인한 네 개의 연구(Lai, 2019;

Maydeu-Olivares, 2017; Nevitt & Hancock, 2001;

Xu, 2019)를 정리한 결과는 그림 1과 같다.

그림 1의 (a)는 정규성 가정이 만족된 상황, (b)는 정규성 가정이 약간 위배된 상황, (c)는 정규성 가정이 중간 정도로 위배된 상황에서의 결과를 보여준다. 그림에서 보이는 바와 같이, (a)에서는 모든 방법이 우수한 것을 확인할 수 있었다. 한편, 정규성 가정이 위배된 상황에서는 방법 간의 수행 차이가 나타났는데, 먼저 (b)에서는 표본크기 조건에 따라 방법별로 두 가지 주요 패턴이 관찰되었다. 첫 번째는 MLM과 MLR에서 나타난 것으로 표본크기가 증가함에 따라 수행도가 향상되는 경향이 있었으며, 두 번째는 부트스트랩과 베이지안에서 기준에 부합한 결과의 비율이 감소하다가 증가하는 패턴을 확인할 수 있었다. 이러한 두 패턴의 차이는 각 방법이 점근 이론에 의존하는지 여부로 설명될 수 있으며, 이는 표본크기에 따른 수행도 차이에 방법의 이론적 특성이 중요한 영향을 미친다는 점을 시사한다. 다음으로, (c)에서는 앞선 (b)와 다르게 부트스트랩의 수행이 다른 방법들에 비해 우수한 것으로 확인된 반면, 나머지 방법들은 일관되게 저조한 수행도가 관찰되었다. 흥미롭게도, MLM은 (b)의 결과와 달리 MLR과 유의한 차이를 보이는 것으로 나타났다. 이는 해당 범주에 분류된 선행 연구(Lai, 2019)가 현실적인 상황을 반영하기 위해 시뮬레이션의 기본조건으로 모형 부적합(misfit) 조건을 포함했기 때문으로 추측된다.⁹⁾ 그 결과, 정규성 가

9) MacCallum(2003)에 따르면, 경험적 분석 모형은 참(true) 모형의 근사치에 불과하며, 약간의 부적합은 종종 발생할 수 있다. 이에 따라 Lai(2019)는 현실적인 상황을 반영하기 위해 RMSEA를 모형 부적합의 지표로 기본조건에 포함시켜 분석을 진

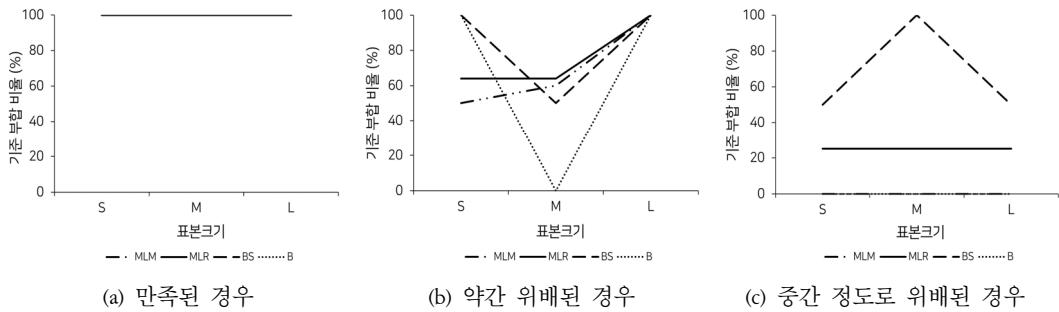


그림 1. 정규성 가정 위배 정도에 따른 요인부하 추정치의 표준오차 상대적 편향 연구 결과 중 편향이 10% 미만인 결과의 비율. 표본크기 조건의 경우, S는 250 이하, M은 250 초과 500 이하, L은 500 초과인 경우를 의미. BS = 부트스트랩; B = 베이지안.

정의 위배와 부적합 조건이 결합된 조건에서 는 MLM의 수행이 현저히 저하되었고, MLM과 MLR 간 유의한 차이가 발견되었다. 이는 MLR 이 MLM에 비해 비정규성과 설정 오류가 있는 모형에도 접근적으로 강건하다는 기존 연구 (Lai, 2018)와 동일한 결과이다.

요약하자면, 표준오차의 상대적 편향 측면에서 각 방법의 수행도는 정규성 가정 위배 정도와 표본크기 조건에 따라 달라지는 경향을 보였다. 정규성 가정이 위배된 상황에서 부트스트랩은 표본크기 조건과 무관하게 상대적으로 우수한 수행도를 보이는 것으로 나타났다. 한편, MLM과 MLR은 정규성이 약간 위배된 상황에서는 유사한 결과를 보였으나, 다른 조건이 결합되었을 때에는 수행도에 차이가 관찰되었다. 특히 모형 설정의 오류나 결측치가 있을 때, 대체로 MLR이 더 강건한 수행도를 보이는 경향을 확인할 수 있었다 (Lai, 2018, 2019; Savalei, 2010). 반면, 베이지안의 결과는 표본크기와 관계없이 정규성 가정의 위

행했다. 이때 Cudeck과 Browne(1992)의 절차에 따라 RMSEA를 0.06으로 설정하여 시뮬레이션을 수행하였다.

배 조건에 대해 전반적으로 강건하지 않은 것으로 나타났다. 이는 비정규성 문제에 대응하는 데 일반적인 최대우도 방법보다 나은 성능을 보일 수 있지만, 대안적인 방법으로 제안된 수정된 최대우도 방법 및 부트스트랩과 비교할 때는 덜 효과적일 수 있음을 시사한다.

제안 및 논의

심리학을 비롯한 사회과학 분야에서 사용되는 자료는 정규성 가정을 만족하지 않는 경우가 많아, 이를 보완하기 위한 다양한 추정 방법이 제안되어 왔다. 그러나 이러한 방법의 수행도는 연구마다 조건이 다르게 설정됨에 따라 산발적으로 논의되고 있으며, 이에 따라 각 방법의 구체적인 수행도에 대해서는 아직 까지 명확한 합의에 이르지 못한 상황이다. 따라서 본 연구는 정규성 가정이 위배되었을 때 대안적으로 사용할 수 있는 방법들을 소개하고 어떤 원리로 대응되는지 설명함과 동시에, 기존의 수행도 연구 결과를 종합적으로 정리하여 추정 방법을 선택하는 데 연구자가

실질적으로 참고할 수 있는 가이드라인을 제안하는 것을 목적으로 하였다.

연구자들에게 실질적인 도움을 주고자 하는 본 연구의 목적을 위해 그림 2에는 대안적인 방법의 사용을 고려할 때 참고할 수 있는 가이드라인을 간략하게 나타낸 차트가 제공되어 있다. 굵은 글씨체는 1종오류 측면에서, 기울임체는 검정력 측면에서, 밑줄은 표준오차 편향의 측면에서 추천되는 방법을 나타낸다. 그림에 따라 왜도와 첨도의 절대값 크기로 자료의 정규성 가정의 만족 여부를 확인한 이후, 위배된 경우에는 그 정도에 따라 추정 방법의 선택에 대한 도움을 얻을 수 있다. 먼저, 정규성 가정이 만족된 경우, 대부분의 방법이 비교적 우수한 성능을 보여 어떤 방법을 사용하더라도 큰 차이는 없을 것으로 보인다. 최대우도 방법에 기반하는 MLM과 MLR의 경우에도 최대우도 방법과 다른 수식을 이용하지만, 수행도에 있어 통계적으로 유의한 차이는 보

고되지 않았다(Maydeu-Olivares, 2017). 다만, 결측치 등 다른 조건이 결합되어 있을 경우, 정보행렬의 차이로 인해 이를 방법 간 수행도가 달라지는 것으로 확인되었다(Savalei, 2010). 안정적인 추정을 위해 최대우도 방법과 부트스트랩은 일정 이상의 표본크기가 요구된다는 점(Nevitt & Hancock, 2001)을 고려할 때, 표본크기가 작은 경우라면 베이지안(Lee & Song, 2004)이 보다 적합한 선택이 될 수 있다.

다음으로, 정규성 가정이 약간 위배된 조건에서는 부트스트랩과 수정된 최대우도 방법의 사용이 권장된다. 구체적으로, 모형 적합도의 수행 측면에서 지표변수의 개수가 적을 때는 부트스트랩이 추천되고, 지표변수의 개수가 많을 때 일정 이상의 표본크기만 주어진다면 수정된 최대우도 방법의 사용도 적절할 것으로 기대된다. 반면, 베이지안은 일관적으로 낮은 수행을 보였다는 점에서 사용이 지양된다. 표준오차의 편향 측면에서는 상대적으로 모든

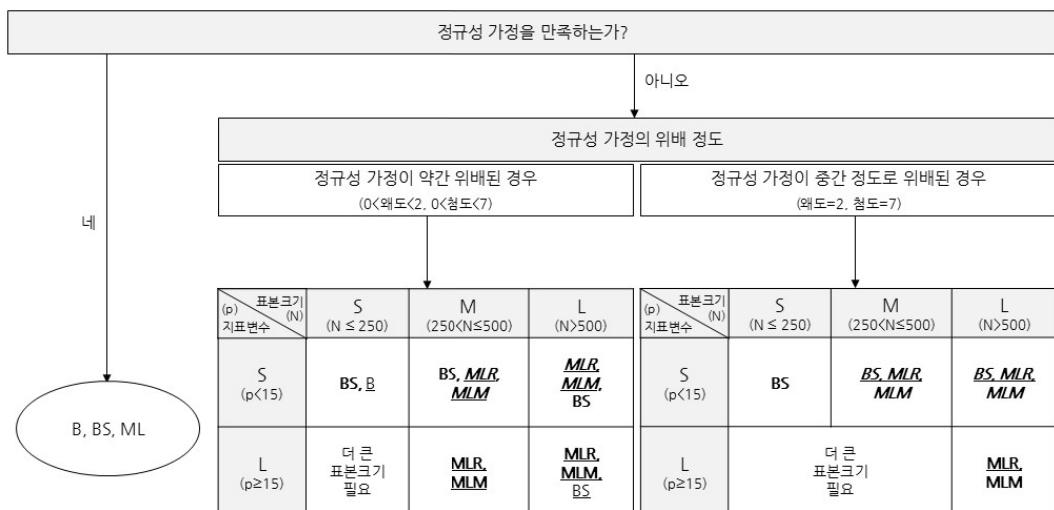


그림 2. 정규성 가정 위배 정도, 표본크기, 지표변수 개수에 따른 선택적 가이드라인. 1종오류가 통제되는 방법은 굵은 글씨체로, 검정력이 확보된 방법은 기울임체로, 표준오차 편향 측면에서 추천되는 방법은 밑줄로 표시하여 구분. BS = 부트스트랩; B = 베이지안.

방법의 수행이 양호하게 나타났으나, 전반적인 조건에서 양호한 수행을 보인 수정된 최대우도 방법의 사용이 권장된다. 즉, 두 기준을 모두 고려했을 때, 가장 추천될 수 있는 추정 방법은 모형이 작을 때는 부트스트랩, 모형이 클 때는 수정된 최대우도 방법이다.

마지막으로, 정규성 가정이 중간 정도로 위배된 조건에서도 부트스트랩과 수정된 최대우도 방법이 추천된다. 모형 적합도의 측면에서, 지표변수의 개수가 적을 때는 부트스트랩이, 많은 경우에는 충분한 표본크기의 확보를 전제로 수정된 최대우도 방법의 사용이 적절할 것으로 기대된다. 반면, 베이지안은 여전히 일관되지 않으면서도 기준을 벗어나는 결과들이 대부분이었다는 점에서 추천되지 않는다. 표준오차의 편향 측면에서 전반적으로 우수한 수행을 보인 MLR이 가장 적절하였으며, 모든 조건에서 베이지안이 가장 양호하지 않은 결과를 보였다. 이에 따라 추천될 수 있는 방법은 부트스트랩과 MLR이다. 정규성 가정이 심각하게 위배된 조건에서도 여러 방법의 수행도를 탐색하는 연구들이 진행되어 왔으나(Curran et al., 1996; Nevitt & Hancock, 2001, 2004; Yu, 2002), 이를 실제 자료에 적용했을 때는 심각한 추정 문제의 발생 가능성이 보고되기도 하였다(김재철, 라종민, 2018). 이러한 점을 고려할 때, 해당 조건에서는 비정규성으로 인한 왜곡을 보정하는 접근을 넘어선 새로운 모형의 사용이 검토된다. 이에 비정규성 자체를 모형화하는 skew-SEM(Asparouhov & Muthén, 2016)이 고려될 수 있으며, 이는 평균과 공분산으로 자료를 모형화하는 기존 구조방정식 모형을 확장한 접근법이다. 비대칭 t 분포 등을 통해 왜도와 첨도까지 모수로 설정함으로써 분포 형태 자체를 모형화한다는 특

징을 가진다(Yuan & Gomer, 2021).

결론적으로 비정규성이 가정된 시뮬레이션 연구 결과를 종합했을 때, 부트스트랩이 가장 적합한 방법으로 간주되며, 베이지안 방법은 권장되지 않는다. 한편, 특정 조건에서는 수정된 최대우도 방법의 사용이 추천되는데, 특히 표본크기와 지표변수의 개수가 일정 수준 이상인 경우가 이에 해당된다. 이러한 조건에서는 MLR이 다른 방법들보다 우수한 수행을 보였으므로, MLR이 더 적합한 선택이 될 수 있다. 다만, 이는 대략적인 기준에 기반한 분류의 결과로, 모든 연구 상황에 적용될 수 있는 것은 아니라는 점에 유의해야 한다. 또한, 본 연구의 논의는 다수의 선행 연구에서 공통적으로 설정한 조건들을 중심으로 전개되었으며, 본 논의의 범위에 포함되지 않는 조건들에서의 수행도는 다른 연구들을 통해 확인할 수 있다. 결측치 조건의 경우, Savalei(2010)는 완전한 데이터에서는 방법들 간 수행도에 점근적으로 유의한 차이가 없으나, 결측치가 있는 경우 MLR의 수행도가 우수함을 보였다. 또한, 모형의 부적합 정도를 살펴본 연구(Lai, 2019)는 부적합 정도와 비정규성 정도가 모두 심각한 조건에서 부트스트랩이 MLM과 MLR보다 나은 성능을 나타낸다고 보고하였다. 이러한 점들을 고려하여 연구자는 앞선 이론적인 고찰을 바탕으로 추정 방법 간의 결과를 비교함으로써 각 방법의 장점과 한계점을 종합적으로 검토하여 상황에 맞는 적절한 방법을 결정할 수 있을 것이다.

본 연구는 대안적인 추정 방법을 이용한 여러 결과들을 통합하여 비교하는 과정에서 몇 가지 제한점을 가지고 있다. 먼저, 비교에 이용된 추정 방법의 종류가 한정적이다. 본 연구에서는 세 개의 접근법을 대표하기 위한 추

정 방법으로 MLM, MLR과 비모수 부트스트랩 및 Bollen-Stine 부트스트랩, 그리고 디폴트 사전분포 설정을 적용한 베이지안을 이용하였다. 그러나 MLMV나 편향 조정 부트스트랩 등 더 다양한 강건한 추정 방법이 존재하며, 이들 간의 수행 차이가 존재한다는 것이 선행 연구에서 보고된 바 있다(Falk, 2018; Ferraz et al., 2022). 또한, 베이지안은 사전분포로 정보 사전분포를 지정할 경우 프로그램의 디폴트 분포를 이용하는 것보다 더욱 정확하고 안정적인 추정 결과를 제공하는 것으로 알려져 있다. 김재철, 라종민(2018)은 정규성 가정이 심각하게 위배된 경우, MLMV 등과 같은 빈도주의 접근보다 적절한 사전분포를 지정한 베이지안 접근이 추정 문제를 해결하는 데 더 효과적일 수 있음을 보고하였다. 이러한 결과에 비추어 볼 때, 베이지안의 결과가 가장 좋지 않았던 본 연구에서의 결론과는 다른 결과가 확인될 가능성이 있다. 다음으로, 본 연구가 제안한 방향성이 보편적인 원리를 반영하고 있는 것은 아니라는 점에 있다. 본 연구는 다양한 연구 상황에서의 결과를 분류하기 위해 다수의 선행 연구에서 공통적으로 이용한 요인들을 바탕으로 기준을 설정하여, 이를 비교에 이용하였다. 그러나 경험적인 기준을 적용하는 과정에서, 일부 연구의 결과는 분류 기준에 완전히 부합하지 않아 유사한 범주로 분류되었으며, 특정 연구에서 개별적으로 설정한 결측치 비율 등과 같은 조건들을 충분히 반영하지 않았다는 점이 한계로 작용할 수 있다. 하지만 모든 연구에서 발생할 수 있는 조건을 고려해 하나의 기준으로 통합하여 비교하는 것은 현실적으로 가능하지 않으며, 본 연구의 주된 목적은 여러 시뮬레이션 결과들이 공통적인 조건에서 보이는 전반적인 수행 경향을

제시하는 데 있었다. 이에 추정 방법의 적절성에 대한 보다 다양하고 현실적인 조건에서의 평가는 추후 후속 연구와 경험적 증거를 통해 이루어질 것으로 기대한다.

이러한 제한점에도 불구하고 본 연구는 국내에서 아직까지 심도 있게 다뤄지지 않은 정규성 가정 위배의 문제와 이에 대안적으로 사용될 수 있는 방법들을 소개하고, 이들 방법에 대한 종합적인 논의를 시도하였다는 점에서 의의가 있다. 또한, 선행 연구에서 실시된 시뮬레이션 결과들을 종합적으로 정리함으로써 각 방법들의 수행 결과를 체계적으로 비교하였다. 단순히 비교하는 것에 그치지 않고, 해당 방법들의 적절성을 논의하여 간략한 가이드라인을 제공하였다는 장점이 있다. 이에 따라 본 연구 내용을 토대로 연구자들이 정규성 가정의 위배 상황에서 대안적인 추정 방법들에 대한 이해와 각 방법의 장점과 한계점을 종합적으로 고려하여 적합한 방법을 결정할 수 있을 것으로 기대한다.

참고문헌

- 김수영 (2016). 구조방정식 모형의 기본과 확장: MPLUS 예제와 함께. 서울: 학지사.
- 김재철, 라종민 (2018). 청소년 연구에서 베이지안 SEM과 전통적 SEM의 모수추정 안정성과 적합도 비교. 청소년학연구, 25(2), 419-445.
<https://doi.org/10.21509/KJYS.2018.02.25.2.419>
- 이지윤, 김수영 (2021). 베이지안 분석에서 사전분포의 이해와 적용. 한국심리학회지: 일반, 40(4), 567-596.
<https://doi.org/10.22257/kjp.2021.12.40.4.567>

- Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference*. Washington, DC. Retrieved from https://nces.ed.gov/FCSM/pdf/2005FCSM_Asparouhov_Muthen_IIA.pdf
- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. Retrieved from <https://www.statmodel.com/download/Bayes3.pdf>
- Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 1-19. <https://doi.org/10.1080/10705511.2014.947375>
- Bandalos, D. L., & Leite, W. (2013). The role of simulation in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 625-666). Charlotte, NC: IAP.
- Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika*, 48(4), 493-517. <https://doi.org/10.1007/BF02293875>
- Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, 13, 95-115. <https://doi.org/10.1214/aos/1176346579>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, 21, 205-229. <https://doi.org/10.1177/0049124192021002004>
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152. <https://doi.org/10.1111/j.2044-8317.1978.tb00581.x>
- Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83. <https://doi.org/10.1111/j.2044-8317.1984.tb00789.x>
- Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, 41(2), 193-208. <https://doi.org/10.1111/j.2044-8317.1988.tb00896.x>
- Cain, M. K., & Zhang, Z. (2018). Fit for a Bayesian: An Evaluation of PPP and DIC for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(1), 39-50. <https://doi.org/10.1080/10705511.2018.1490648>
- Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44(2), 347-357. <https://doi.org/10.1111/j.2044-8317.1991.tb0096>

- 6.x
- Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, 57, 357-369. <https://doi.org/10.1007/BF02295424>
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16-29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Depaoli, S. (2021). *Bayesian structural equation modeling*. New York, NY: The Guilford Press.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26. <https://doi.org/10.1214/aos/1176344552>
- Falk, C. F. (2018). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation Modeling*, 25(2), 244-266. <https://doi.org/10.1080/10705511.2017.1367254>
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), 521-532. <https://doi.org/10.1007/BF02293811>
- Foldnes, N., & Olsson, U. H. (2016). A simple simulation technique for nonnormal data with prespecified skewness, kurtosis, and covariance matrix. *Multivariate Behavioral Research*, 51(2-3), 207-219. <https://doi.org/10.1080/00273171.2015.1133274>
- Ferraz, R. C., Maydeu-Olivares, A., & Shi, D. (2022). Asymptotic is Better than Bollen-Stine Bootstrapping to Assess Model Fit: The Effect of Model Size on the Chi-Square Statistic. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(5), 731-743. <https://doi.org/10.1080/10705511.2022.2053128>
- Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling*, 4(2), 87-107. <https://doi.org/10.1080/10705519709540063>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 439-492). Charlotte, NC: IAP.
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, 14(3), 275-299. <https://doi.org/10.1037/a0015825>
- Gao, C., Shi, D., & Maydeu-Olivares, A. (2020). Estimating the maximum likelihood root mean square error of approximation (RMSEA) with non-normal data: A Monte-Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2), 192-201. <https://doi.org/10.1080/10705511.2019.1637741>
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741. <https://doi.org/10.1109/TPAMI.1984.4767596>

- Gelman, A., Carlin, J. B., & Stern, H. S. (2013). *Bayesian Data Analysis* (3rd ed.). New York, NY: Chapman & Hall/CRC.
- Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733-807.
<https://www.jstor.org/stable/24306036>
- Grønneberg, S., & Foldnes, N. (2019). Testing Model Fit by Bootstrap Selection. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(2), 182-190.
<https://doi.org/10.1080/10705511.2018.1503543>
- Hancock, G. R., & Liu, M. (2012). Bootstrapping standard errors and data-model fit statistics in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 296-306). New York, NY: The Guilford Press.
- Hjort, N. L., Dahl, F. A., & Steinbakk, G. H. (2006). Post-processing posterior predictive p values. *Journal of the American Statistical Association*, 101(475), 1157-1174.
<https://doi.org/10.1198/016214505000001393>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research*, 26(3), 329-367.
<https://doi.org/10.1177/0049124198026003003>
- Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465-497). Academic Press.
<https://doi.org/10.1016/B978-012691360-6/50017-3>
- Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological bulletin*, 112(2), 351-362.
<https://doi.org/10.1037/0033-2909.112.2.351>
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(1), 221-233.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2, Pt.1), 183-202.
<https://doi.org/10.1007/BF02289343>
- Kaplan, D. (2023). *Bayesian statistics for the Social Sciences* (2nd ed.). New York, NY: The Guilford Press.
- Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650-673). New York, NY: The Guilford Press.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: The Guilford Press.
- Lai, K. (2018). Estimating standardized SEM parameters given nonnormal data and incorrect model: Methods and comparison. *Structural Equation Modeling*, 25(4), 600-620.
<https://doi.org/10.1080/10705511.2017.1392248>
- Lai, K. (2019). More Robust Standard Error and Confidence Interval for SEM Parameters Given Incorrect Model and Nonnormal Data. *Structural Equation Modeling: A Multidisciplinary*

- Journal*, 26(2), 260-279.
<https://doi.org/10.1080/10705511.2018.1505522>
- Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and Maximum Likelihood Approaches in Analyzing Structural Equation Models with Small Sample Sizes. *Multivariate Behavioral Research*, 39(4), 653-686.
https://doi.org/10.1207/s15327906mbr3904_4
- Lei, P.-W., & Wu, Q. (2012). Estimation in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 164-180). New York, NY: The Guilford Press.
- Levy, R., & Mislevy, R. J. (2016). *Bayesian Psychometric Modeling* (1st ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Liang, X., & Yang, Y. (2016). Confirmatory Factor Analysis under Violations of Structural and Distributional Assumptions: A Comparison of Robust Maximum Likelihood and Bayesian Estimation Methods. *Journal of Psychological Science*, 39(5), 1256-1267.
<https://doi.org/10.16719/j.cnki.1671-6981.20160536>
- Lüdtke, O., Ulitzsch, E., & Robitzsch, A. (2021). A Comparison of Penalized Maximum Likelihood Estimation and Markov Chain Monte Carlo Techniques for Estimating Confirmatory Factor Analysis Models With Small Sample Sizes. *Frontiers in psychology*, 12, Article 615162.
<https://doi.org/10.3389/fpsyg.2021.615162>
- MacCallum R. C. (2003). 2001 Presidential Address: Working with Imperfect Models. *Multivariate Behavioral Research*, 38(1), 113-139.
- https://doi.org/10.1207/S15327906MBR3801_5
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, 24(3), 383-394.
<https://doi.org/10.1080/10705511.2016.1269606>
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166.
<https://doi.org/10.1037/0033-2909.105.1.156>
- Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling*, 19(1), 86-98.<https://doi.org/10.1080/10705511.2012.634724>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313-335.
<https://doi.org/10.1037/a0026802>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.
- Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling*, 8(3), 353-377.
https://doi.org/10.1207/S15328007SEM0803_2
- Nevitt, J., & Hancock, G. R. (2004). Evaluating Small Sample Approaches for Model Test Statistics in Structural Equation Modeling. *Multivariate Behavioral Research*, 39(3), 439-478.

- https://doi.org/10.1207/S15327906MBR3903_3
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications to developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, 15(4), 352-367.
<https://doi.org/10.1037/a0020143>
- Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64(1), 37-52.
<https://doi.org/10.1007/BF02294318>
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465-471.
<https://doi.org/10.1007/bf022936>
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemse, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1), 1-26.
<https://doi.org/10.1038/s43586-020-00001-2>
- Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). Hoboken, NJ: Wiley.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.
<https://doi.org/https://doi.org/10.2307/1912934>
- Xu, M. (2019). *A Comparison of Frequentist and Bayesian Approaches for Confirmatory Factor Analysis* (Publication No. 27534819) [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations and Theses Global.
- Yang, M., Jiang, G., & Yuan, K. H. (2017). The Performance of Ten Modified Rescaled Statistics as the Number of Variables Increases. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 414-438.
<https://doi.org/10.1080/10705511.2017.1389612>
- Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles, CA.
- Yuan, K. H., & Bentler, P. M. (2000). Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data. *Sociological Methodology*, 30, 165-200.
<https://doi.org/10.1111/0081-1750.00078>
- Yuan, K. H., & Gomer, B. (2021). An overview of applied robust methods. *The British journal of mathematical and statistical psychology*, 74 Suppl 1, 199-246.
<https://doi.org/10.1111/bmsp.12230>
- Yuan, K. H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, 59(2), 397-417.
<https://doi.org/10.1348/000711005X85896>
- Yung, F.-Y., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean

and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.) *Advanced structural equation modeling techniques* (pp. 195-226). New Jersey: Erlbaum.

1차원고접수 : 2025. 09. 01

2차원고접수 : 2025. 11. 12

최종게재결정 : 2025. 12. 02

한국심리학회지: 일반
Korean Journal of Psychology: General
2025, Vol. 44, No. 4, 453-481
<http://dx.doi.org/10.2257/kjp.2025.12.44.4.453>

Exploring alternatives to ML when normality assumptions are violated in structural equation modeling

Eunkyung Shin Su-Young Kim

Department of Psychology, Ewha Womans University

Maximum likelihood (ML), which is commonly used to estimate structural equation models, is based on the assumption of normality in the data. However, violations of the normality assumption are frequently reported in psychology and the social sciences, which can lead to biased estimation results and undermine the validity of statistical inferences. Although alternative methods that can provide reliable results under non-normal conditions have been explored, the performance of these methods has shown inconsistent patterns across studies, making it difficult to establish clear criteria for selecting appropriate methods. This study aims to address the problems posed by violations of the normality assumption and to explore alternative methods for dealing effectively with such violations. By integrating studies from the last 30 years of research, the study attempts to provide practical guidelines for researchers confronted with non-normality in their data. It first discusses the importance of the normality assumption in ML and examines the impact of its violation on estimation results. It then presents several alternative methods that are applicable under non-normal conditions and analyses the principles by which these methods deal with non-normality. Furthermore, previously published studies are systematically reviewed and categorized according to specific conditions, with the results visualized through tables and figures to compare the performance of different methods. Finally, the study integrates these discussions to propose guidelines for researchers and highlight their implications and limitations.

Key words : Normality, MLM, MLR, Bootstrap, Bayesian