

## 구조방정식 모형을 활용한 매개효과 검정 방법의 비교: 백분위수 부트스트랩과 편향조정 부트스트랩\*

심 미 경<sup>1)</sup>      서 영 숙<sup>2)</sup>      김 수 영<sup>1)\*</sup>  
<sup>1)</sup>이화여자대학교 심리학과      <sup>2)</sup>한국교육&심리연구소

매개모형이 심리학, 교육학, 경영학 등 다양한 분야에서 활발하게 이용되고 있다. 매개효과를 검정하는 방법은 Baron과 Kenny의 단계적 검정, 다변량 델타(Sobel의 방법) 등 전통적인 방법들을 거쳐 최근 부트스트랩이 표준 절차로 자리잡았다. 백분위수 부트스트랩으로 추정된 신뢰구간에 존재하는 잠재적 편향을 보완하기 위하여 편향조정 부트스트랩이 제안된 이후, 조정된 방법이 더 정확한 신뢰구간과 높은 검정력을 제공한다라는 연구 결과들을 바탕으로 국내에서도 빈번히 사용되고 있다. 하지만 편향조정 부트스트랩이 검정력 뿐만 아니라 1종오류 또한 더 높아 적절한 매개효과 검정 방법이 아닐 수 있다는 사실이 최근 보고되고 있다. 본 연구는 매개효과의 검정에 관련된 최신 연구 흐름을 살펴보고, 잠재변수를 사용한 매개모형의 다양한 조건을 반영한 시뮬레이션을 실시하여 백분위수 부트스트랩과 편향조정 부트스트랩의 수행을 비교함으로써 매개효과 검정 방법에 대한 연구자들의 이해를 높이고자 하였다. 연구 1의 결과 편향조정 부트스트랩, 백분위수 부트스트랩, 다변량 델타 방법 순으로 검정력이 높았으며, 지표변수 개수가 많을수록, 효과크기가 클수록, 요인부하가 클수록, 표본크기가 클수록 검정력이 증가하는 경향을 보였다. 연구 2의 결과 1종오류 역시 편향조정 부트스트랩, 백분위수 부트스트랩, 다변량 델타 방법 순으로 높아 구조방정식 매개모형에서도 편향조정 부트스트랩이 가장 큰 1종오류를 발생시킨다는 것을 확인할 수 있었다. 마지막으로 본 연구의 제한점을 기술하고 연구 결과를 바탕으로 한 가이드라인을 제공한다.

주요어 : 매개모형, 매개효과, 부트스트랩, 검정력, 1종오류

\* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임  
(NRF-2019S1A5A2A03041362).

† 교신저자: 김수영, 이화여자대학교 심리학과, 서울시 서대문구 이화여대길 52  
Tel: 02-3277-3792, E-mail: [suyoung.kim@ewha.ac.kr](mailto:suyoung.kim@ewha.ac.kr)

매개모형이란 독립변수와 종속변수 사이에 존재하는 인과관계를 제3의 변수인 매개변수를 이용하여 설명하는 모형으로서(Baron & Kenny, 1986; MacKinnon, 2012; Preacher & Hayes, 2004) 심리학, 교육학, 경영학, 의학 등 다양한 분야에서 활발하게 이용되고 있다. 매개모형을 사용하는 연구자의 가장 큰 관심 중 하나는 연구자가 설정한 매개변수가 독립변수와 종속변수 사이에서 통계적으로 유의한 중개(intermediate) 역할을 하는가이다. 더 정확하고 안정적으로 매개효과를 추정하고 검정하기 위하여 다양한 방법들이 제안되고 발전되어 왔다(Baron & Kenny, 1986; Cohen & Cohen, 1983; Efron & Tibshirani, 1986; Judd & Kenny, 1981; MacKinnon et al., 2002; MacKinnon et al., 2007; Preacher & Hayes, 2004, 2008a; Preacher & Selig, 2012; Sobel, 1982; Tofighi & Kelly, 2020). Sobel(1982)의 방법이나 Baron과 Kenny(1986)의 단계적 검정 방법 등 전통적인 방법들을 거쳐 Efron(1979)이 개발한 부트스트랩(bootstrap)이 매개효과 검정의 표준 절차로 자리잡았다. 경험적으로 구한 신뢰구간(confidence interval, CI)을 이용하여 매개효과를 검정하는 부트스트랩은 전통적인 방법들이 갖는 비현실적인 가정(예를 들어, 매개효과 추정치의 정규성)을 필요로 하지 않는다(Fritz & MacKinnon, 2007; Shrout & Bolger, 2002). 일반적으로 부트스트랩이라 하면 백분위수 부트스트랩(percentile bootstrap)을 의미하는데 이 방법으로 구한 신뢰구간에는 편향이 존재하여 양쪽 꼬리부분이 모수를 포함하지 않을 확률이 동일하지 않다. 이를 보완하기 위해 Efron과 Tibshirani(1986)는 신뢰구간의 편향을 수정한 편향조정 부트스트랩(bias-corrected bootstrap)을 제안하였다. 편향을 조정하는 방법이 이론적으로 더 정확한 신뢰

구간을 제공하고 검정력을 높인다는 여러 연구 결과들이 존재한다(Cheung & Lau, 2008; MacKinnon et al., 2004). 하지만 최근 연구 흐름을 살펴보면, 편향조정 부트스트랩이 검정력은 더 높을지라도, 부풀려진(inflated) 1종오류를 발생시키기 때문에 최선의 선택은 아니라는 지적을 받고 있다(Biesanz et al., 2010; Falk & Biesanz, 2015; Hayes & Scharkow, 2013; Taylor et al., 2008; Tofighi & Kelly, 2020). 본 연구는 잠재변수가 있는 매개모형의 다양한 조건을 반영한 시뮬레이션을 통하여 백분위수 부트스트랩과 편향조정 부트스트랩의 수행도(performance)를 비교함으로써 매개효과 검정 방법 중 특히 부트스트랩에 대한 연구자들의 이해도를 높이고자 한다. 이를 위하여 정반합의 과정을 통해 발전하는 학문의 관점에서 매개효과의 검정 방법과 부트스트랩에 관련된 최신 연구 흐름을 살펴보고, 부트스트랩 사용과 관련한 실용적인 가이드라인을 제공하는 것이 목적이다.

매개모형은 독립변수  $X$ 가 종속변수  $Y$ 에 미치는 영향이 매개변수  $M$ 을 통해 전달되는 현상을 설명하는데 유용한 모형으로 수십 년간 방법론 연구(Fritz & MacKinnon, 2007; Lachowicz et al., 2018; Liu & Wang, 2019; Muthén & Asparouhov, 2015) 및 내용영역 연구(Duckworth et al., 2016; Lockman & Servaty-Seib, 2018; Malone et al., 2016)에서 활발하게 사용되고 있다. 그림 1은 하나의 매개변수를 포함하는 단순매개모형을 나타낸다. 독립변수가 매개변수에 미치는 영향을 경로  $a$ , 독립변수의 영향을 통제된 상태에서 매개변수가 종속변수에 미치는 영향을 경로  $b$ 라고 할 때, 매개효과(간접효과)는 두 경로의 곱인  $ab$ 로 정의된다. 이때 독립변수가 매개변수를 통하지 않고

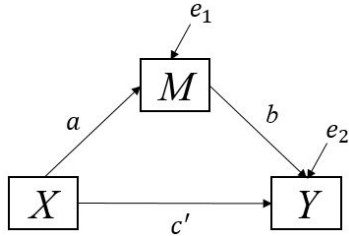


그림 1. 매개모형 경로도

종속변수에 미치는 영향을 직접효과  $c'$  이라고 한다. 단순매개모형에서 독립변수가 종속변수에 미치는 효과인 총효과  $c$ 는 간접효과와 직접효과의 합으로 표현할 수 있다( $c = ab + c'$ ).

매개효과  $ab$ 를 직접 검정하는 방법 중 하나로 Sobel의 방법 또는 다변량 델타 방법(multivariate delta method)이 있다(Cramer, 1946; Sobel, 1982).<sup>1)</sup> 이 방법은 매개효과 추정치가 정규분포를 따른다는 가정하에 매개효과의 표준오차를 이론적으로 계산하여 검정하는 방법이다. Sobel의 방법은 다양한 분야에서 오랜 기간 사용되었고, 여전히 많은 영역에서 이용되고 있지만 검정력이 낮다는 한계가 있으며(Falk & Biesanz, 2015; Fritz & MacKinnon, 2007; Hayes & Scharkow, 2013), 두 모수의 곱으로 이루어진 매개효과는 이론적으로도 경험적으로도 정규성 가정을 만족하지 않는다는 문제가 있다(Biesanz et al., 2010; Lomnicki, 1967; MacKinnon et al., 2004; Shrout & Bolger, 2002; Springer & Thompson, 1966; Stone & Sobel, 1990).

이러한 비현실적인 가정을 극복하기 위하

1) 델타 방법에 비해 Sobel의 방법은 경로  $a$ 와  $b$  추정치의 공분산이 0임을 가정한다는 차이가 있다(Sobel, 1982). 많은 연구자들이 델타 방법과 Sobel의 방법을 엄격하게 구분하지 않는 경향이 있다.

여 부트스트랩 방법이 제안되었다(Efron, 1979; MacKinnon et al., 2002). 부트스트랩은 연구자가 가진 표본을 가상의 모집단(pseudo population)으로 가정하고 매개효과 추정치의 경험적인(empirical) 표집분포를 도출한다. 95% 부트스트랩 신뢰구간을 구하기 위해 연구자는 복원추출을 통해 만든 부트스트랩 표본들로부터 매개효과 추정치들을 구하고, 이 중 2.5%와 97.5%에 해당하는 백분위수를 각각 하한과 상한으로 하는 신뢰구간을 만든다. 위와 같은 방법을 백분위수 부트스트랩(percentile bootstrap)이라고 한다. 이 방법은 표집분포에 대한 가정이 필요 없고, Sobel의 방법을 비롯한 전통적인 방법들보다 더 높은 검정력을 보인다는 장점이 있으나, 이렇게 구한 부트스트랩 신뢰구간은 비대칭적일 수 있다는 특성이 있다.

부트스트랩 신뢰구간 추정치의 정확성을 높이기 위한 시도로 편향조정 부트스트랩 방법이 제안되었다(Efron & Tibshirani, 1986). 원칙적으로 “편향을 조정한” 부트스트랩 신뢰구간보다 더 우수할 것이라고 기대할 수 있다. 실제로 2000년대까지 편향조정 부트스트랩이 더 정확한 신뢰구간과 더 높은 검정력을 제공한다는 이유로 권장되었다(Cheung & Lau, 2008; MacKinnon et al., 2004; Preacher & Hayes, 2008a). 하지만 이 방법은 실제로 존재하지 않는 효과가 통계적으로 유의하다고 잘못 판단할 확률인 1종오류율을 크게 증가시킨다는 치명적인 단점이 지적되었다(Taylor et al., 2008). 2010년대 이후 진행된 시뮬레이션 연구들을 살펴보면 편향조정 부트스트랩이 부풀려진 1종오류 또는 비일관적이거나 낮은 커버리지를 보인다는 이유로 여러 비판을 받았다

(Biesanz et al., 2010; Falk & Biesanz, 2015; Fritz et al., 2012; Hayes & Scharkow, 2013). 사회과학 분야의 많은 연구자가 사용하고 있는 통계 프로그램인 SPSS PROCESS Macro에서도 매개효과 검정시 편향조정 부트스트랩 신뢰구간을 디폴트로 사용하다가 버전 3부터는 백분위수 부트스트랩으로 변경된 바 있으며<sup>2)</sup>, *Mplus*도 이용자들을 위한 제언에서 2014년까지는 편향조정 부트스트랩을 권장했지만, 2016년 이후로는 다시 백분위수 부트스트랩을 권장하고 있다(B. Muthén, 2014, 2018; L. Muthén, 2008, 2014, 2016). 그럼에도 불구하고 국내에서는 편향조정 부트스트랩이 여전히 빈번하게 사용되고 있으며(Kim et al., 2015; Park & Bak, 2014; Park & Hur, 2021; Oh et al., 2018), 여러 구조방정식 서적에서 추천되고 있는 방법이기도 하다(Bae, 2016; Lee, 2018).

소개한 바와 같이 최근 여러 연구에서 백분위수 부트스트랩으로 다시 돌아가야 한다는 흐름이 있고, 이 주장이 여러 학자에 의하여 받아들여지기 시작하고 있지만, 아직 충분한 근거가 제시되었다고 보기에는 미흡하다. 백분위수 부트스트랩 및 편향조정 부트스트랩 등을 비교하는 기존의 시뮬레이션 연구들을 살펴보면 대부분 경로모형 위주의 단순한 구조거나 상당히 제한된 조건에서만 실시가 되었고, 다양하고 충분한 근거를 고려하지 못했다는 문제가 있다. 예를 들어, Hayes와 Scharkow(2013)는 몬테카를로 시뮬레이션을 통해 일곱 가지 간접효과 검정 방법의 상대적 신뢰성(relative trustworthiness)을 비교하여 연구자들에게 실행가능한 조언을 제공하였는데, 사회과학에서 자주 사용되는 잠재변수를 반영하지 않은 모형으로만 시뮬레이션을 진행했다

는 점에서 한계가 있다. Falk와 Biesanz(2015)는 잠재변수가 있는 매개모형을 이용해 여러 검정 방법의 수행도를 살펴보는 시뮬레이션을 실시하였지만, 지표변수 개수와 잠재변수의 내적 일관성 조건이 단지 두 개였고, 표본크기는 500 이하만 적용하는 등 구조방정식 모형 자체의 조건이 상당히 제한되었다는 아쉬움이 있다. 또한 Cheung(2007)은 매개효과를 검정하는 네 가지 신뢰구간 생성 방법을 설명하고 이 방법들의 수행도를 비교하였는데, 각 검정 방법의 수행도를 비교하는 준거로서 오직 신뢰구간의 커버리지 확률(coverage probability) 하나만을 적용했다는 한계가 있다.

본 연구의 목적은 매개모형을 사용하는 연구자들의 매개효과 검정 방법에 대한 이해도를 높이고, 적절한 부트스트랩 검정 방법을 선택할 수 있도록 가이드라인을 제공하는 데에 있다. 이를 위하여 매개효과 검정 방법과 이들을 비교하는 최근 연구들을 소개한다. 이후 구조방정식 모형의 다양한 조건을 반영한 몬테카를로 시뮬레이션을 통해 매개효과 검정 방법의 수행도를 비교한다. 검정 방법으로는 백분위수 부트스트랩과 편향조정 부트스트랩을 기본으로 고려하고, 비교 목적으로 다변량 델타 방법을 포함한다. 그리고 구조방정식 매개모형의 자료생성 조건으로 표본크기, 효과 크기의 수준(경로계수의 크기), 지표변수의 개수, 요인부하의 크기를 사용한다. 각 검정 방법의 수행도를 검정력 및 1종오류 측면에서 비교하고 비수렴률을 확인한다(Fritz et al., 2012; Tofighi & Kelly, 2020; Valente et al., 2016). 마지막으로 시뮬레이션 결과를 바탕으로 연구상황에 적절한 부트스트랩 방법을 제안하고 본 연구의 한계를 기술한다.

2) <http://processmacro.org/faq.html>

## 매개효과 검정 방법과 연구 흐름

매개모형의 인기만큼이나 다양한 매개효과 검정 방법들이 최근까지도 개발되며 발전되고 있다. 모든 방법을 소개하기에는 지면의 한계가 있으므로 그 중 몇 가지 방법만 소개한다. 먼저 단일 표본을 사용하는 방법(single sample methods)으로 매개효과 내의 경로계수들을 개별적으로 검정하는 인과단계 접근법(causal step approach)과 매개효과 모수 자체(두 경로계수의 곱인  $ab$ )를 검정하는 계수곱 접근법(product of coefficients approach)이 있다. 인과단계 접근법에는 일련의 회귀분석을 통해 회귀계수들을 단계적으로 검정하는 방법(Baron & Kenny, 1986; Judd & Kenny, 1981)과 경로계수  $a$ 와  $b$ 만 유의하면 매개효과가 유의하다고 보는 결합 유의성 검정(joint significance test, Hayes, 2018; MacKinnon et al., 2002)이 있다. 계수곱 접근법에는 매개효과 추정치가 정규분포를 따른다고 가정하는 점근적 정규 방법(asymptotic normal methods, 예를 들어 다변량 델타 방법과 Sobel의 방법 등)과 분포에 대해 가정하지 않는 곱분포 방법(distribution of the product method, 예를 들어  $M$ 검정, 경험적  $M$ 검정 등)이 있다. 다음으로 재표집을 통해 수많은 표본으로부터 매개효과를 추정하는 방법(resampling methods)에는 원자료를 거짓 모집단으로 보고 이를 이용해 재표집하는 부트스트랩(bootstrap)과 원자료에서 얻은 모수 추정치들을 이용해 재표집하는 몬테카를로(monte carlo) 방법이 있다. 이 외에도 우도기반 신뢰구간을 이용하는 방법(likelihood-based confidence intervals, Cheung, 2007), 베이저안 가정을 이용하는 방법(bayesian mediation, Yuan & MacKinnon, 2009), 모형기반 제약 최적화 절차(the model-based constrained

optimization procedure, Tofghi & Kelly, 2020) 등이 있다.

## Baron과 Kenny의 단계적 검정

매개효과 검정의 발전에서 초창기 Baron과 Kenny(1986)가 제안한 단계적 방법은 다음과 같다. 첫 번째로 독립변수가 종속변수에 미치는 총효과( $c$ )가 통계적으로 유의함을 확인한 후, 두 번째로 독립변수가 매개변수에 미치는 영향( $a$ )이 유의한지 확인한다. 세 번째로 독립변수를 통제된 상태에서 매개변수의 종속변수에 대한 영향( $b$ )이 유의한지 확인한다. 총효과  $c$ 와 경로  $a$ ,  $b$ 가 각각 통계적으로 유의할 때 마지막으로 직접효과( $c'$ )의 크기가 총효과의 크기보다 작아졌는지 확인한다. 이와 같은 절차는 매개모형에 있는 각각의 경로를 검정하는 과정에서 경로들의 성질을 확인할 수 있고, 모형에 대한 이론적 탐색과 통찰의 기회를 제공한다. 장점이 있는 반면(Zhao et al., 2010), 매개효과의 검정력이 낮다는 한계가 있다(Fritz & MacKinnon, 2007; MacKinnon et al., 2002). 이 뿐만 아니라, 매개효과와 직접효과의 부호가 다른 비일관적 매개효과가 발생하여 총효과가 유의하지 않은 경우(억제효과), 매개효과가 실제로 존재함에도 이를 확인하지 못한 채 분석을 종료하는 오류를 범할 수 있다(Kim & Kim, 2020). 이런 이유로 많은 연구자들이  $ab$  검정을 매개효과 검정의 결정적 단계로 보고 있으며(Cerin & MacKinnon, 2009; Fairchild & McDaniel, 2017; Rucker et al., 2011), Hayes(2018)는 Baron과 Kenny의 방법에 대하여 통계적으로도 철학적으로도 이상적이지 않다는 인식이 커지고 있으며, 더 이상 이 방법을 권하지 않는다고 하였다.

### 결합유의성 검정

결합유의성 검정은 총효과에 대한 검정 없이 매개효과를 구성하는 경로  $a$ 와  $b$ 만 각각 통계적으로 유의하면( $a \neq 0$  and  $b \neq 0$ ) 매개효과가 존재한다고 판단한다(Kenny, Kashy, & Bolger, 1998; MacKinnon et al., 2002). 이 방법은 Baron과 Kenny의 방법보다도 적은 단계를 확인하여 검정이 간편하고 쉽다는 장점이 있으며 MacKinnon 등(2002)은 결합유의성 검정이 인과단계 접근법들 중에서는 높은 검정력과 정확한 1종오류율을 보인다고 하였다. 하지만 이 방법도 경로들을 개별적으로 검정하는 인과단계 접근법의 한계로부터 자유로울 수 없다. 인과관계 접근법은 연구자가 검정하려는 매개효과 자체의 크기에 대한 추정치,  $p$ 값( $p$ -value), 신뢰구간 등을 제공하지 않아 매개효과에 대한 직접적인 증거를 제공하지 못한다. 또한  $a$ 와  $b$ 가 모두 통계적으로 유의하지 않은 경우일지라도  $ab$ 가 통계적으로 유의할 수 있다는 사실을 간과하고 있으며(Hayes, 2009; Kim, 2016 재인용) 중다매개모형으로 확장이 어렵다는 한계도 있다(MacKinnon et al., 2002).

### 다변량 델타 방법

다변량 델타 방법은 표본크기가 커질수록 매개효과 추정치  $\hat{ab}$ 가 점근적으로 정규분포를 따른다고 가정한다. 이 방법은 매개효과 추정치를 추정치의 표준오차  $SE_{\hat{ab}}$ 으로 나눈 값을 검정통계량으로 하여  $z$ 검정을 실시한다. 대표적으로 Sobel의 방법(Sobel, 1982)이 많이 알려져 있으며 국내에서는 매개효과 추정치의 정규성을 가정하는 방법들을 통칭하여 Sobel의

방법이라 부르기도 한다. Sobel의 방법은 델타 방법에 비해 경로  $a$ 와  $b$ 가 서로 독립적이라는 가정이 추가되고(Sobel, 1982, 1987), 식 1과 같이 매개효과 추정치의 표준오차를 계산한다.

$$SE_{\hat{ab}} = \sqrt{\hat{b}^2 SE_a^2 + \hat{a}^2 SE_b^2} \quad (1)$$

Sobel의 방법과 비교하여 다변량 델타 방법(Cramer, 1946)은 경로  $a$ 와  $b$  사이에 상관성이 있을 수 있음을 가정한다. 구조방정식 매개모형에서는 경로  $a$ 와  $b$ 가 서로 관련이 있을 수 있으므로 매개효과를 정확히 검정하기 위해서는 두 경로 간 상관관계를 고려하여야 한다(MacKinnon, 2008; Valente et al., 2016). 다변량 델타 방법에 따른 매개효과 추정치의 표준오차는 식 2와 같다.

$$SE_{\hat{ab}} = \sqrt{\hat{b}^2 SE_a^2 + \hat{a}^2 SE_b^2 + SE_a^2 SE_b^2} \quad (2)$$

다변량 델타 방법은 이론적으로 간단하고 표준오차의 계산이 쉽다는 장점이 있는 반면, 충족될 수 없는 가정을 바탕으로 하기에 발생하는 불가피한 한계들이 있다. 첫째, 정규분포를 따르는 임의의 두 변수(예,  $a$ 와  $b$ )의 곱으로 만들어진 변수(예,  $ab$ )는 이론적으로도 경험적으로도 정규분포를 따르지 않는다(Biesanz et al., 2010; Lomnicki, 1967; MacKinnon et al., 2004; Shrout & Bolger, 2002; Springer & Thompson, 1966). Stone과 Sobel(1990)은 시뮬레이션을 통해 간접효과 모수가 0이 아닐 때 간접효과 추정치의 분포가 편포됨을 보였다. 둘째, 매개효과 추정치가 좌우대칭을 만족하지 않는 표집분포를 보임에도 불구하고 정규성

가정에 기반한 대칭적인 신뢰구간을 이용함으로써 비대칭적인 오류율 문제가 발생한다 (MacKinnon & Dwyer, 1993). 정규이론에 기반한 매개효과  $ab$ 의 신뢰구간은 매개효과 모수 분포의 왜도와 첨도에 영향을 받는다. 왜도의 절대값이 커질수록 신뢰구간의 커버리지는 낮아지고 불균형은 커지며, 첨도의 값이 커질때에도 신뢰구간 커버리지의 불균형이 커진다. 심지어 왜도와 첨도의 상호작용이 발생하기도 한다(Cheung, 2007; Fritz & MacKinnon, 2007; Hayes & Scharkow, 2013; MacKinnon et al., 2004). 셋째, 시뮬레이션 연구 결과(Biesanz et al., 2010; Fritz & MacKinnon, 2007), 부트스트랩 등의 다른 검정 방법들과 비교해 정규성을 가정한 방법들의 검정력이 매우 낮다. 이런 여러 가지 이유로 인해 방법론 연구자들은 매개효과의 검정 상황에서 더이상 정규성 가정에 기반한 방법들을 권하지 않는다(Biesanz et al., 2010; Hayes, 2018; Hayes & Scharkow, 2013).

### 곱분포 방법

MacKinnon 등(2002; 2004)은 두 회귀계수  $a$ 와  $b$ 의 곱인  $ab$ 의 분포가 실제로 정규분포를 따르지도 않고 종종 높은 첨도를 갖는 비대칭 분포를 따름에도 불구하고 다변량 델타 방법이 정규분포를 가정하기 때문에 검정력이 낮고 신뢰구간이 불균형을 이룬다고 보고 이러한 가정을 필요로 하지 않는 대안적 방법으로 곱분포 방법과 재표집 방법(부트스트랩)을 사용할 것을 제안하였다. 먼저 곱분포 방법은 표본으로부터 두 정규확률변수  $a$ 와  $b$ 의 추정치(평균과 표준편차)를 구하여 각각 표준점으로 변환하고( $z_a = \hat{a}/\hat{\sigma}_a$ 와  $z_b = \hat{b}/\hat{\sigma}_b$ ) 이들

을 곱한 값( $z_a z_b$ )을 이용해 Meeker 등(1981; 1994)이 제시한 두 확률변수의 곱분포에서 임계값(Meeker 상한, Meeker 하한)을 구한다. 이렇게 구한 임계값들을 이용하여  $M$  신뢰한계를 구하고 이 신뢰한계가 0을 포함하지 않으면 매개효과가 유의하다고 판단한다. 시뮬레이션 연구들에 따르면 곱분포 방법은 정규성을 가정한 검정 방법보다 더 정확한 신뢰한계를 제공하므로(Falk & Biesanz, 2015; Hayes & Scharkow, 2013) 연구자가 원자료를 가지고 있지 않아 재표집 방법(예, 부트스트랩)을 사용할 수 없는 경우에 권장된다(MacKinnon et al., 2004; Tofighi & MacKinnon, 2011). Tofighi와 MacKinnon(2011)은 두 정규확률변수의 평균과 표준오차, 두 변수 간 상관관계, 유의수준을 지정하면 정규 방법, 곱분포 방법, 몬테카를로 방법의 신뢰한계, 백분위수, 분위수 등을 계산해주는 R 패키지 RMediation을 개발하였다.

### 백분위수 부트스트랩

정규성을 가정하는 매개효과 검정 방법들이 가지는 한계를 극복하기 위한 또 다른 대안으로 부트스트랩이 제안되었다(Bollen & Stine, 1990; Efron, 1979; MacKinnon et al., 2002; MacKinnon et al., 2004; Preacher & Hayes, 2004; Shrout & Bolger, 2002). 부트스트랩은 매개효과 추정치의 표집분포에 대한 가정이 필요 없는 비모수적이고 경험적인 재표집 방법이다. 이 방법은 연구자가 가진 표본(표본크기  $n$ )을 거짓 모집단으로 가정하고 이로부터 무선적으로 복원추출하여 새로운 표본을 표집한다. 이렇게 만들어진 하나의 부트스트랩 표본(마찬가지로 표본크기  $n$ )으로부터 첫 번째 매개효

과 추정치( $\hat{ab}_1$ )를 구한다. 이 과정을  $B$ 번(예, 1,000번) 반복하여  $B$ 개의 부트스트랩 표본을 생성하고 각 부트스트랩 표본으로부터 총  $B$ 개의 매개효과 추정치(예,  $\hat{ab}_1, \hat{ab}_2, \dots, \hat{ab}_{1,000}$ )를 구한다. 이렇게 얻은  $B$ 개의 매개효과 추정치를 작은 값부터 오름차순으로 나열한다. 유의수준 5%를 가정하는 백분위수 부트스트랩 방법에서 95% 신뢰구간을 구하기 위해, 크기순으로 나열된 매개효과 추정치 중 아래로부터 2.5%, 위로부터 2.5%에 해당하는 값을 각각 하한(lower bound)과 상한(upper bound)으로 사용한다. 이렇게 구한 신뢰구간이 0을 포함하지 않으면 영가설( $H_0: ab = 0$ )을 기각하고 매개효과가 통계적으로 유의하다고 판단한다.

연구자가 가지고 있는 자료를 바탕으로 하는 부트스트랩 신뢰구간 방법은 매개효과 추정치의 분포에 대한 사전적인 가정을 필요로 하지 않고, 경로  $a$ 와  $b$ 의 독립성에 대한 가정도 필요로 하지 않는다. 또한 백분위수 부트스트랩은 계산이 명료하고 전통적인 방법들보다 검정력이 높다는 장점이 있다. 이론적으로 매개분석을 위한 모든 가정이 충족된다면 간접효과 추정치는 간접효과 모수의 편향되지 않은 추정치이다. 하지만 특히 작은 표본크기에서 매개효과 추정치의 부트스트랩 표집분포는 치우치거나(skewed) 비대칭인 경향이 있어(Bollen & Stine, 1990; MacKinnon & Dwyer, 1993; MacKinnon et al., 2007; Shrout & Bolger, 2002), 이 방법으로 구한 모수 추정치는 좋은 모수 추정치가 되기 위한 조건 몇 가지를 만족하기 어렵다. 예를 들어, 부트스트랩으로 구한 표본 통계치들이 대응하는 모수의 불편향 추정치가 아니기 쉬우며( $\hat{\theta} \neq \theta$ ), 추정치로부터

신뢰구간의 하한, 상한까지의 거리가 동일하지 않을 수 있다(DiCiccio & Efron, 1996; Efron, 1987; Kirby & Gerlanc, 2013). 비대칭적인 분포를 가지고 대칭적인(위, 아래 각각 2.5%) 신뢰구간을 계산하면 신뢰구간 양 끝의 오차율이 동일하지 않은 편향된 신뢰구간과 낮은 검정력이 발생한다(MacKinnon & Dwyer, 1993; Shrout & Bolger, 2002; Stone & Sobel, 1990). 이러한 점을 보완하고 추정의 정확성을 높이기 위하여 편향조정 부트스트랩과 편향조정가속 부트스트랩이 제안되었다.

#### 편향조정 부트스트랩

Efron과 Tibshirani(1986)는 백분위수 부트스트랩 신뢰구간에 존재하는 잠재적 편향을 조정하는 편향조정 부트스트랩(bias-corrected bootstrap, BC)을 제안하였다. 원자료로 그린 표본의 분포와 부트스트랩 추정치들로 그린 표집분포는 그 모양이 일치하지 않는다. 부트스트랩 표집분포는 일반적으로 치우치고 비대칭이다. 이때  $B$ 개의 부트스트랩 추정치들( $\hat{ab}_i$ ) 중 원자료로부터 구한 매개효과 추정치( $\hat{ab}$ )보다 크기가 작은 부트스트랩 추정치들의 비율을 이용해 편향 상수(bias constant)를 계산하고 이를 이용해 신뢰구간 양 끝의 오차율이 같아지도록 부트스트랩 신뢰구간을 수정한다. 만약 부트스트랩 표집분포에 편향이 없어  $\hat{ab}$ 보다 작은  $\hat{ab}_i$  개수의 비율이 50%라면 편향상수는 0이 되고 백분위수 부트스트랩 신뢰구간과 편향조정 부트스트랩 신뢰구간은 일치하게 된다. 참고로 편향조정 가속 부트스트랩(bias-corrected and accelerated bootstrap, BCa)은 부트스트랩 표집분포의 지나친 치우침을 보정



하기 위하여 가속 상수(acceleration constant)를 이용해 한 번 더 수정한 방법이다.

### 몬테카를로 방법

몬테카를로 신뢰구간(MacKinnon et al., 2004; Preacher & Selig, 2012)은 대부분의 통계 소프트웨어 패키지들이 매개효과와 관련된 모수 추정치의 점근공분산행렬(asymptotic covariance matrix)을 제공한다는 사실을 이용한다. 먼저 표본을 이용해 매개효과를 구성하는 요소들의 추정치( $\hat{a}$ ,  $\hat{b}$ ,  $SE_{\hat{a}}$ ,  $SE_{\hat{b}}$ )를 구하고, 이들을 모수값으로 갖는 무작위 표본을  $M$ 개(예, 1,000개) 생성한다. 이렇게 만들어진 표본들로부터  $ab$ 의 표집분포를 만들고 분포의 백분위수를 이용하여 매개효과 추정치의 몬테카를로 신뢰구간을 계산한다. 몬테카를로 방법도 부트스트랩과 마찬가지로 매개효과 추정치의 분포에 대한 가정이 필요없다. 이 방법은 요약정보만 가지고 있을 때에도 매개효과를 검정할 수 있어 계산이 빠르며, 부트스트랩 사용이 어려운 경우(예, 다층모형)에도 이용할 수 있다는 장점이 있다. 이 방법은 앞서 언급한 RMediation 패키지(Tofighi & MacKinnon, 2011)를 이용해 추정할 수 있다.

몬테카를로 신뢰구간 방법이 제안된 이후에도 새로운 방법들이 개발되고 시뮬레이션을 통해 각 방법들의 수행이 검증되고 있다(Biesanz et al., 2010; Cheung, 2007; Tofighi & Kelly, 2020; Yuan & MacKinnon, 2009). 본 연구는 수많은 매개효과 검정 방법들 중에서 현실점을 기준으로 충분히 검증되고, 또 SPSS, Mplus 등 통계 프로그램에서도 사용이 권장되고 있는 방법인 부트스트랩 방법을 중심으로 연구 흐름을 살펴본다.

### 부트스트랩 연구의 흐름

부트스트랩 신뢰구간의 정확성을 높이기 위한 방법으로 제안된 편향조정 부트스트랩은 검정력이 높다는 장점이 있다. Efron과 Tibshirani(1986)에 의해 개발되었던 이 방법은 MacKinnon 등(2004)에 의해 재조명된 이후 심리학을 비롯한 여러 사회과학 분야에서 가장 정확한 매개효과 검정 방법으로 인식되어 왔다. 특히 표본크기가 작아 백분위수 부트스트랩의 검정력이 낮은 경우 편향조정 부트스트랩은 더 높은 검정력을 근거로 우월한 방법으로 추천되었다. MacKinnon 등(2004)은 재표집 방법들 중 편향조정 부트스트랩으로 구한 신뢰구간이 가장 정확하므로 이 방법이 최선(the best method)이라고 하였다. Cheung과 Lau(2008)는 잠재변수를 사용한 매개모형에서 여덟 가지 매개효과 검정 방법<sup>3)</sup> 중 편향조정 부트스트랩이 매개효과와 억제효과 검정에 가장 정확한 신뢰한계와 높은 검정력을 제공한다고 하였다. Preacher와 Hayes (2008)는 다중매개모형에서 개별 간접효과에 대한 신뢰한계를 구하는 가장 검정력이 높고 합리적인 방법은 부트스트랩이며 그중에서도 특히 편향조정 부트스트랩을 사용할 것을 권하였다.

하지만 연구자들 사이에서 편향조정 부트스트랩의 사용에 대한 우려의 목소리가 나타나면서, 백분위수 부트스트랩과 편향조정 부트스트랩에 대한 새로운 비교 연구결과가 나오기 시작하였다. Taylor 등(2008)은 3개의 매개효과가 있는 다중매개모형에서 신뢰구간을 이용해 매개효과를 검정하는 방법으로 백분위수

3) 정규성을 가정하는 네 가지 방법(Sobel, Aroian, Goodman, jackknife)과 부트스트랩 네 가지 방법(percentile, bootstrap- $t$ , BC, BCa)

부트스트랩과 편향조정 부트스트랩이 선호되며, 후자의 방법이 더 높은 검정력을 보이긴 하지만 일부 조건에서 1종오류도 높다고 지적하였다. Biesanz 등(2010)은 편향조정 가속 부트스트랩은 부풀려진 1종오류율과 비일관적인 커버리지를 보이므로 사용을 추천하지 않으며, 1종오류율을 유지하면서도 합리적인 검정력과 안정적인 커버리지를 보이는 백분위수 부트스트랩이 전반적으로 가장 좋은 수행을 보인다고 결론지었다. Fritz 등(2012)은 편향조정 부트스트랩에서 1종오류를 부풀리는 조건을 탐색하기 위하여 시뮬레이션 연구를 수행하였다. 이들은 0이 아닌 경로(non-zero path)<sup>4)</sup>의 효과크기가 작을 때(small, 경로계수는 0.14)에는<sup>5)</sup> 표본크기가 500, 1,000인 조건에서 1종오류가 높았고, 한 경로의 효과크기가 보통이거나 클 때(medium 또는 large, 경로계수는 0.36 또는 0.51)에는 표본크기가 100 이하로 작은 조건에서 가장 높은 1종오류를 보인다고 보고하였다. 이전의 연구 결과들을 바탕으로 낮은 검정력을 보이는 작은 표본크기 조건에서 편향조정 부트스트랩이 종종 사용되어왔기 때문에 작은 표본크기에서 부풀려진 1종오류를 보이는 것은 문제로 인식된다(Chen & Fritz, 2021). Hayes와 Scharkow(2013)는 검정력이 가장 중요한 경우라면 편향조정 부트스트랩이 좋은 선택이겠지만, 검정력과 1종오류, 신뢰구간의 커버리지

모두를 균형있게 고려한다면 백분위수 부트스트랩 신뢰구간이 좋은 검정 방법이 될 것이라고 하였다. Tofighi와 Kelly(2020) 역시 편향조정 부트스트랩 방법이 다른 방법들보다 조금 더 높은 검정력을 보이긴 하지만 이는 1종오류율을 희생하고 얻은 것이므로 사용을 권하지 않는다고 하였다. 지금까지 연구된 결과를 종합하면, 백분위수 부트스트랩과 편향조정 부트스트랩이 모두 좋은 방법이지만, 검정력의 관점에서는 편향조정 방법이 약간 더 우세하고, 제1종오류의 관점에서는 백분위수 방법이 더 낫다고 볼 수 있다. 다만 소개된 연구 대부분이 경로모형만을 이용하여 검정 방법들을 비교하였고, 잠재변수가 포함된 구조방정식 모형을 이용한 일부의 연구들도 매우 제한된 자료생성 조건에서만 비교가 진행되었다는 점에서 한계가 있다.

## 방 법

### 몬테카를로 시뮬레이션

다양한 조건에서 생성된 구조방정식 모형의 매개효과를 검정하는 방법들을 비교하기 위하여 몬테카를로 시뮬레이션을 실시하였다. 매개효과 검정 방법에 따른 수행 결과를 비교하는 준거에 따라 두 가지 연구를 수행하였는데, 하나는 검정력을 비교하는 연구이고, 다른 하나는 1종오류를 비교하는 연구였다. 기존의 연구 중 신뢰구간의 양 끝 부분이 모수를 포함하지 않을 확률(오차율)을 보고한 연구도 있었지만(Cheung & Lau, 2008), 본 연구에서는 검정력과 1종오류를 중심으로 시뮬레이션 결과를 비교하였다. 사용한 모형은 독립변수, 매개

- 4) 1종오류는 매개효과가 없다는 영가설이 진실인 상황에서 영가설을 기각하는 오류이므로 매개효과  $ab = 0$ 인 상황, 즉  $a$  또는  $b$ 가 0이거나  $a$ 와  $b$ 가 모두 0인 경우를 조건으로 한다. 예를 들어,  $a = 0$ 일 때에는 경로  $b$ 가 non-zero path가 될 수 있다.
- 5) 효과크기의 수준에 따라 경로계수와 설명오차 분산의 크기가 계산된다.

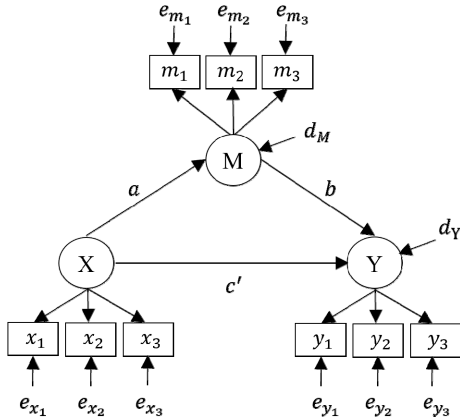


그림 2. 잠재변수마다 지표변수를 3개씩 가지는 단순매개모형 경로도

변수, 종속변수가 각각 하나씩 있는 단순매개 모형이었으며 잠재변수가 있는 구조방정식 모형을 적용하였다(그림 2 참조). Sim 등(2022)에 따르면 2016년부터 2018년까지 국내외 7가지 주요 학술지에서 발표된 매개모형 연구 중 57% 이상이 잠재변수를 사용한 매개모형이었다. 지표변수와 잠재변수는 모두 연속형 변수이며, 평균 0, 분산 1을 갖는 표준화 변수를 가정하였다. 표준화 변수를 사용함으로써 표준화 효과크기와 표준화 요인부하를 사용하는 장점을 누릴 수 있다. 표준화 효과크기는 변수의 단위에 의존하지 않고(Preacher & Hayes, 2008b), 변수의 선형 변환에도 영향을 받지 않는다(Lachowicz et al., 2018). 또한 내용영역 연구에서도 주로 보고되는 결과이므로 다른 연구 결과들과 비교하기 쉽다(Liu & Wang, 2019).

잠재변수가 있는 단순매개모형에서 표준화 변수를 사용하면, 측정모형 단계에서는 지표 변수와 잠재변수의 분산을 1로 설정함에 따라 표준화 요인부하  $\lambda$ 의 제곱은 곧 지표변수의

신뢰도가 된다(L. Muthén & B. Muthén, 2002). 또한 표준화 요인부하의 값을 설정함에 따라 측정오차의 분산이  $1 - \lambda^2$ 으로 계산된다. 예를 들어, 표준화 요인부하가 .5라면 지표변수의 신뢰도는 .5<sup>2</sup>인 .25가 되고, 측정오차의 분산은  $1 - .5^2$ 인 .75가 된다. 참고로 이때 합성 신뢰도(composite reliability)는 .5로 계산된다. 구조모형 단계에서는 경로계수 값이  $R^2$ 과 같아지고 설명오차의 분산이  $1 - R^2$ 이 되어, 계산과 해석이 쉬워진다는 장점이 있다.

매 시뮬레이션 조건마다 1,000개의 자료세트를 사용하였고, 각 자료세트마다 부트스트랩 표본의 개수( $B$ )도 1,000으로 설정하였다(Kim, 2012; Shrout & Bolger, 2002; Tofighi & Kelley, 2020). 자료의 생성과 추정은 *Mplus* 8.3(L. Muthén & B. Muthén, 1998-2021)과 Python(Van Rossum & Drake, 1995)을 이용하였으며, 모형의 추정은 최대우도 추정(maximum likelihood estimation, ML) 방법을 이용하였다.

#### 자료분석의 준거

연구 1에서는 각 시뮬레이션 조건에서 매개 효과 검정 방법의 검정력을, 연구 2에서는 1종오류를 비교하고, 두 연구 모두에서 비수렴률(nonconvergence rate)을 확인하였다. 검정력이란 어떤 효과가 존재하는 것이 진실인 상황에서 영가설( $H_0 : \theta = 0$ )을 기각함으로써 효과가 존재한다고 올바른 의사결정을 내릴 확률을 의미하며 대략 80% 또는 90% 이상을 확보하는 것이 추천된다. 1종오류란 어떤 효과가 존재하지 않는 것이 진실인 상황에서 영가설( $H_0 : \theta = 0$ )을 기각함으로써 효과가 존재한다고 잘못된 의사결정을 내릴 확률로서 일반적

으로 5% 이내로 통제되기를 기대한다. *Mplus*에서 시뮬레이션을 실시한 후 나오는 결과 부분의 가장 마지막 열(% Sig Coeff)은 모든 자료세트를 이용해 모형을 추정했을 때 각 모수가 통계적으로 유의한 비율을 나타낸다. 따라서 이 값은 모수가 0이 아니라면 검정력을, 모수가 0이라면 1종오류를 의미하게 된다. 비수렴률은 시뮬레이션에서 생성된 각 자료세트에 모형을 추정할 때 수렴에 실패한 자료세트의 비율을 의미하며, 값이 작을수록 추정이 안정적으로 이루어졌음을 확인한다.

## 연구 1

### 시뮬레이션 설계 및 자료 생성

연구 1에서는 잠재변수를 사용한 매개모형의 다양한 조건에서 세 가지 검정 방법의 검정력을 비교한다. 시뮬레이션에서는 검정 방법을 포함한 총 다섯 가지 조건을 고려하였다. 첫 번째 조건인 검정 방법은 고전적 접근법 중 여전히 가장 많이 사용되는 방법 중 하나인 다변량 델타 방법을 사용하고, 부트스트랩 방법으로는 백분위수 부트스트랩과 편향을 조정한 부트스트랩을 사용한다. 편향을 조정한 방법 중에서는 그동안 국내외 연구자들에게 더 많이 사용되어 온 편향조정 부트스트랩을 사용한다(Kim et al., 2015; MacKinnon et al., 2004; Park & Bak, 2014; Preacher & Hayes, 2008). 두 번째 조건은 잠재변수 하나당 지표 변수 개수로서 세 가지 수준으로 설정하였다. Bollen(1989)이 제안했던 2지표변수 법칙(2-indicator rule)에 따라 최소 2개부터 한 개씩

증가시켜 각 잠재변수마다 2, 3, 4개씩의 지표 변수를 설정하였다. 모형의 단순화를 위하여 하나의 조건 안에서 잠재변수마다 갖는 지표 변수의 개수는 동일하게 하였다. 잠재변수마다 3개의 지표변수를 갖는 매개모형의 경로도가 그림 2에 제공된다. 세 번째 조건은 매개효과의 효과크기(effect size)이다. 매개효과의 크기를 측정하기 위하여 총효과에 대한 간접효과의 비율( $P_M = ab/c$ ; Alwin & Hauser, 1975), 직접효과에 대한 간접효과의 비율( $R_M = ab/c'$ ; Sobel, 1982)을 비롯해 다양한 효과크기 측정치가 소개되었다(Lachowicz et al., 2018; MacKinnon, 2008; Miočević et al., 2018; Preacher & Kelly, 2011). 본 연구에서는 다른 연구들과의 비교가능성을 위하여 매개모형을 다룬 시뮬레이션 연구에서 가장 일반적으로 사용되고 있는 측정치인 매개효과로 설명되는 종속변수 분산의 비율, 즉 종속변수에 대한 설명력( $R^2$ )의 크기를 사용하였다(Cohen, 1988; Fairchild et al., 2009). 효과크기도 세 가지 수준 (small 2%, medium 13%, large 26%)을 설정하였고, 이에 따라 경로  $a$ 와  $b$ 의 계수가 0.14, 0.36, 0.51로 계산되어 설정되었다. 경로계수 역시 모형의 단순화를 위하여 두 경로의 크기는 같다고 가정하였다. 직접효과  $c'$ 의 크기는 small에 해당하는 0.14로 가정하였다. Fritz와 MacKinnon(2007)에 따르면 직접효과의 크기는 매개효과 검정에 큰 영향을 미치지 않기 때문에 모든 조건에서 하나의 경우(small)만 고려하였다. 예를 들어, 효과크기가 작을(small) 때 독립변수  $X$ 와 매개변수  $M$ 의 관계에서 종속변수에 해당하는 매개변수의 분산 중 2%를 독립변수가 설명한다고 가정하고 경로  $a$ 와 매개변수의 설명오차분산  $Var(d_M)$ 을 계산한다.

표 1. 연구 1 시뮬레이션에서 사용한 모수

모수	효과크기		
	Small	Medium	Large
$a$	0.14	0.36	0.51
$b$	0.14	0.36	0.51
$c'$	0.14	0.14	0.14
$Var(d_M)$	0.98	0.87	0.74
$Var(d_Y)$	0.95	0.81	0.65

마찬가지로 독립변수, 매개변수 그리고 종속 변수의 관계에서는 종속변수  $Y$ 의 분산 중 독립변수가 설명하는 분산이 2%, 매개변수가 설명하는 분산이 2%라고 가정하고 경로  $b$ 와  $c'$ , 종속변수의 설명오차분산  $Var(d_Y)$ 를 계산한다. 이렇게 계산된 효과크기에 따른 경로계수와 설명오차분산의 크기는 표 1과 같으며 자세한 계산과정은 Sim 등(2022)의 연구 부록에서 확인할 수 있다. 네 번째 조건인 표준화 요인부하의 크기는 0.5와 0.7 두 가지 수준으로 설정하였다. Hair 등(2019)에 따르면 수렴타당도를 확보하기 위하여 표준화 요인부하 추정치가 0.5 이상이어야 하고, 이상적으로는 0.7 이상이어야 한다. Kline(2016) 역시 각 표준화 요인부하 추정치가 0.7 이상이어야 수렴타당도를 확보할 수 있다고 하였다. 마지막으로 표본크기는 100, 200, 500, 1,000 네 가지 수준으로 설정하였다(Cheung, 2007; MacKinnon et al., 2002). 결과적으로 총 216 ( $3 \times 3 \times 3 \times 2 \times 4$ )가지 시뮬레이션 조건이 연구 1에서 생성되었다.

결과

표 2는 다섯 가지 시뮬레이션 조건하에서의

검정력을 나타낸다. 가독성을 높이기 위해 검정력이 .8 미만인 결과는 음영처리하였다. 검정 방법에 따른 검정력을 살펴보면, 모든 조건에서 편향조정 부트스트랩이 가장 높았고, 백분위수 부트스트랩과 다변량 델타 방법이 그 뒤를 이었다. 지표변수 개수가 많을수록, 효과크기가 커질수록, 요인부하가 커질수록, 표본크기가 커질수록 검정력이 증가하는 경향을 보였다. 지표변수 개수도 적고(요인당 2개) 효과크기도 작을(small) 때 또는 효과크기가 작으면서(small) 요인부하도 작을(0.5) 때에는 표본크기가 1,000으로 커져도 모든 검정 방법에서 검정력 80%를 달성하지 못하였다. 효과크기가 작거나(small), 요인부하가 작거나(0.5), 표본크기가 작은 경우(100)에는 지표변수 개수가 늘어남에도 불구하고 검정력이 감소하는 경우가 나타나기도 하였다.

그림 3은 지표변수 개수, 효과크기, 요인부하 크기, 표본크기의 조건에 따른 세 가지 검정 방법의 검정력 평균을 나타낸다. 모든 조건에서 편향조정 부트스트랩, 백분위수 부트스트랩, 다변량 델타 방법 순으로 높은 검정력을 보이는 것을 확인할 수 있다. 지표변수 개수, 요인부하 크기, 표본크기는 상대적으로 열악한 조건(요인당 지표변수 2개, 요인부하 0.5, 표본크기 100)에서 검정 방법에 따른 검정력 차이가 크고, 지표변수 개수가 많아질수록, 요인부하가 커질수록, 표본크기가 커질수록, 검정력이 낮았던 다변량 델타 방법에서는 가파르게 증가하였고, 검정력이 가장 높았던 편향조정 부트스트랩에서는 완만하게 증가하여 그 차이가 점점 감소하였다. 예를 들어, 편향조정 부트스트랩과 백분위수 부트스트랩의 평균 검정력 차이는 표본크기가 100일 때에는 0.105, 200일 때에는 0.084, 500일 때에는

표 2. 연구 1의 시뮬레이션 조건에 따른 검정력

효과 크기	요인 부하	표본 크기	2 지표변수			3 지표변수			4 지표변수		
			Delta	PBS	BCBS	Delta	PBS	BCBS	Delta	PBS	BCBS
Small	0.5	100	0.000	0.045	0.113	0.000	0.004	0.037	0.000	0.003	0.018
		200	0.000	0.025	0.053	0.000	0.003	0.028	0.002	0.016	0.033
		500	0.000	0.079	0.117	0.017	0.041	0.110	0.040	0.121	0.205
		1,000	0.016	0.130	0.249	0.135	0.277	0.410	0.288	0.461	0.576
	0.7	100	0.000	0.054	0.056	0.003	0.007	0.026	0.005	0.025	0.046
		200	0.001	0.072	0.117	0.017	0.060	0.116	0.031	0.075	0.147
		500	0.074	0.180	0.347	0.204	0.359	0.487	0.279	0.445	0.569
		1,000	0.422	0.504	0.722	0.695	0.828	0.869	0.760	0.861	0.918
Medium	0.5	100	0.000	0.070	0.223	0.000	0.032	0.194	0.003	0.058	0.156
		200	0.000	0.128	0.274	0.039	0.200	0.383	0.252	0.365	0.494
		500	0.151	0.530	0.735	0.843	0.870	0.926	0.967	0.973	0.982
		1,000	0.923	0.934	0.968	1.000	0.998	1.000	1.000	1.000	1.000
	0.7	100	0.031	0.261	0.376	0.254	0.368	0.491	0.382	0.520	0.614
		200	0.485	0.648	0.810	0.835	0.875	0.905	0.901	0.933	0.956
		500	0.992	0.991	0.994	1.000	0.999	1.000	1.000	1.000	1.000
		1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Large	0.5	100	0.000	0.100	0.369	0.000	0.118	0.409	0.010	0.189	0.390
		200	0.000	0.227	0.531	0.135	0.494	0.673	0.618	0.733	0.797
		500	0.424	0.755	0.862	0.977	0.973	0.984	1.000	0.997	1.000
		1,000	0.984	0.975	0.985	1.000	1.000	1.000	1.000	1.000	1.000
	0.7	100	0.157	0.525	0.671	0.694	0.762	0.825	0.857	0.885	0.911
		200	0.913	0.908	0.948	0.988	0.982	0.987	1.000	0.999	1.000
		500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

주. 검정력이 0.8 미만인 결과에 음영 처리. Delta = 다변량 델타 방법; PBS = 백분위수 부트스트랩; BCBS = 편향조정 부트스트랩.

0.056, 1,000일 때에는 0.040으로 감소한다. 효과크기는 점점 커짐에 따라 세 검정 방법의 검정력(절대값) 차이가 유지되거나 근소하게 증가하여 효과크기가 작을(small) 때에는 편향

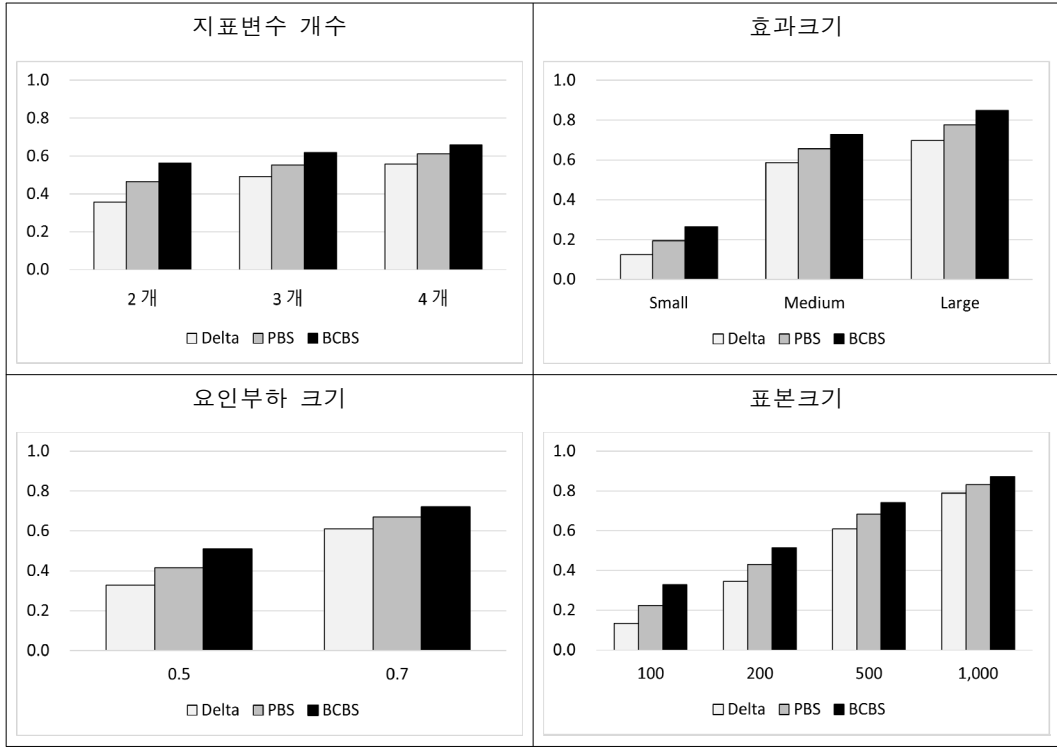


그림 3. 지표변수 개수, 효과크기, 요인부하 크기, 표본크기의 변화에 따른 매개효과 검정 방법간 검정력 비교  
 주. Delta = 다변량 델타 방법; PBS = 백분위수 부트스트랩; BCBS = 편향조정 부트스트랩.

조정 부트스트랩과 다변량 델타 방법의 검정력 차이가 0.140, 중간(medium)일 때에는 0.142, 클(large) 때에는 0.150이다. 하지만 상대적인 비율을 감안한다면 효과크기 역시 커질수록 세 가지 검정 방법의 검정력의 상대적 차이는 점점 감소하였다.

시뮬레이션 조건에 따른 비수렴률은 부록의 표 A에 제공된다. 비수렴률은 다변량 델타 방법이 부트스트랩 방법보다는 높았으며, 백분위수 부트스트랩과 편향조정 부트스트랩은 거의 비슷하였다. 지표변수 개수가 늘어날수록, 효과크기가 커질수록, 요인부하가 커질수록, 표본크기가 커질수록 비수렴률은 감소하는 경향을 보였다. 특히 요인당 지표변수가 2개일

때 높은 비수렴률을 보였는데, 효과크기가 작으면서(Small), 요인부하 0.5, 표본크기 100일 때에는 90%에 가까운 비수렴률을 보이기도 했고, 효과크기(Large)와 요인부하(0.7)가 함께 커지더라도 표본크기가 100으로 작을 때에는 10%가 넘는 비수렴률을 보였다. 요인당 지표변수 개수가 4개일 때에는 요인부하가 0.5이면서 동시에 표본크기가 100인 경우들을 제외하고는 0%의 비수렴률을 보이며 안정적인 추정 결과를 나타냈다.

## 연구 2

### 시뮬레이션 설계 및 자료 생성

연구 2에서는 구조방정식 매개모형의 다양한 조건에서 세 가지 검정 방법의 1종오류 크기를 비교하였다. 시뮬레이션에서는 세 가지 조건이 고려되었다. 첫 번째 조건인 검정 방법은 연구 1과 마찬가지로 다변량 델타 방법, 백분위수 부트스트랩, 편향조정 부트스트랩 세 가지를 고려하였다. 두 번째 조건인 효과 크기는 매개효과  $ab$ 를 0으로 만드는 경로  $a$ 와  $b$ 의 경로계수 조합 일곱 가지를 사용하였다. 종속변수에 대한 설명력( $R^2$ )의 크기에 따라 네 가지 수준의 효과크기(Zero 0%, Small 2%, Medium 13%, Large 26%)를 설정하고, 이에 따라  $a$ 와  $b$ 의 경로계수가 0, 0.14, 0.36, 0.51로 계산되었다. 두 경로의 조합에 따라 일곱 가지 경우(ZZ, ZS, ZM, ZL, SZ, MZ, LZ; Z = Zero; S = Small; M = Medium; L = Large)가 생성되었다. 직접효과  $c'$ 의 크기는 연구 1에서와 마찬가지로 0.14로 설정하였다. 표본크기는 100, 200, 500, 1,000, 2,000, 3,000, 5,000 일곱 가지 수준으로 설정하였다(Yuan et al., 2017). 기존 문헌들을 살펴보면 검정력은 표본크기가 커짐에 따라 증가하는 뚜렷한 경향을 보이는 것과 달리, 1종오류는 표본크기 조건이 200까지인 연구에서는 표본크기가 커짐에 따라 1종오류도 증가하였고(Tofighi & MacKinnon, 2016), 표본크기 조건이 2,500까지인 연구에서는 변화의 방향이 일관적이지 않았다(Fritz et al., 2012). 이런 경향은 특히 부풀려진 1종오류가 발생할 때 그러했다. 따라서 연구 2에서는 표본크기 증가에 따른 1종오류의 변화와 부풀려진 1종오류가 발생하는 조건을 더 다양한 표본크기 조건에서 확인하기 위하여 연구 1보다 더 큰 표본크기 조건을 추가

하였다. 이외에 지표변수의 개수는 3개, 표준화 요인부하의 크기는 0.7을 공통적으로 적용하였다. 결과적으로 연구 2에서는 총 147 ( $3 \times 7 \times 7$ )가지 조건에서 시뮬레이션을 실시하였다.

### 결과

표 3은 시뮬레이션 조건에 따른 1종오류 크기를 나타낸다. 가독성을 위하여 1종오류가 5%를 초과하는 결과는 기울임체로 표기하였고, Serlin(2000)의 기준(3.5~6.5%)을 벗어나는 값은 굵은 글씨체로 구분하였다. 두 기준에 모두 해당하는 경우 회색 바탕으로 음영처리하였다. 참고로 1종오류의 크기는 무조건 낮다고 하여 좋은 개념이 아니며, 연구자들이 통계적 검정에서 원하는 상황은 수용가능한 1종오류율의 수준을 넘지 않는 선에서 이를 충분히 이용하면서 검정력을 높이는 것이다. 검정 방법에 따른 1종오류를 비교하면, 대부분의 조건에서 다변량 델타 방법이 가장 낮았고, 편향조정 부트스트랩이 가장 높았다. 세 가지 검정방법 모두에서  $a = b = 0$ 일 때(ZZ) 모든 표본크기에서 현저히 낮은 1종오류율을 보였고, 하한 기준보다 낮은 1종오류는 다변량 델타 방법에서 가장 많이 나타났으며(49가지 경우 중 25번), 상한 기준보다 높은 1종오류는 편향조정 부트스트랩에서만 발생하였다. 이는 편향조정 부트스트랩의 부풀려진 1종오류를 경고했던 기존의 시뮬레이션 연구들과 동일한 결과이다(Biesanz et al., 2010; Fritz et al., 2012; Hayes & Scharkow, 2013; Tofighi & Kelly, 2020). 다변량 델타 방법과 백분위수 부트스트랩에서는 표본크기가 커짐에 따라 1종오류율이 높아지다가, 표본크기 1,000 또는 2,000 이상인



표 3. 연구 2의 시뮬레이션 조건에 따른 1종오류

검정 방법	경로계수			표본크기							
	a 경로	b 경로	조합	100	200	500	1,000	2,000	3,000	5,000	
Delta	0.00	0.00	ZZ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.14	ZS	0.000	0.000	0.005	0.015	0.039	0.036	0.049	
	0.00	0.36	ZM	0.010	0.018	0.037	0.044	0.052	0.046	0.054	
	0.00	0.51	ZL	0.015	0.033	0.042	0.050	0.052	0.048	0.054	
	0.14	0.00	SZ	0.000	0.002	0.003	0.018	0.031	0.034	0.038	
	0.36	0.00	MZ	0.003	0.018	0.035	0.043	0.048	0.049	0.046	
	0.51	0.00	LZ	0.009	0.028	0.035	0.046	0.051	0.052	0.047	
PBS	0.00	0.00	ZZ	0.002	0.001	0.002	0.000	0.000	0.000	0.001	
	0.00	0.14	ZS	0.004	0.007	0.023	0.043	0.054	0.053	0.059	
	0.00	0.36	ZM	0.026	0.040	0.046	0.050	0.055	0.052	0.059	
	0.00	0.51	ZL	0.039	0.049	0.046	0.051	0.054	0.051	0.059	
	0.14	0.00	SZ	0.002	0.007	0.027	0.039	0.051	0.049	0.046	
	0.36	0.00	MZ	0.014	0.038	0.046	0.050	0.051	0.052	0.050	
	0.51	0.00	LZ	0.028	0.040	0.042	0.049	0.049	0.054	0.050	
BCBS	0.00	0.00	ZZ	0.002	0.004	0.004	0.001	0.004	0.004	0.006	
	0.00	0.14	ZS	0.010	0.024	0.041	0.071	0.077	0.069	0.070	
	0.00	0.36	ZM	0.050	0.067	0.054	0.056	0.056	0.051	0.061	
	0.00	0.51	ZL	0.058	0.053	0.048	0.050	0.051	0.055	0.060	
	0.14	0.00	SZ	0.004	0.020	0.057	0.068	0.071	0.064	0.055	
	0.36	0.00	MZ	0.041	0.054	0.054	0.054	0.056	0.055	0.047	
	0.51	0.00	LZ	0.045	0.049	0.047	0.050	0.057	0.057	0.051	

주. 1종오류가 .05를 초과하는 결과는 기울임체로, Serlin의 기준(3.5 ~ 6.5%)을 벗어나는 결과는 굵은 글씨체로 구분하고, 두 기준에 모두 해당하는 경우 음영처리. Delta = 다변량 델타 방법; PBS = 백분위수 부트스트랩; BCBS = 편향조정 부트스트랩. Z = Zero; S = Small; M = Medium; L = Large.

조건에서 5%를 초과하는 결과가 발생했다. 편향조정 부트스트랩에서는 0이 아닌 경로계수 크기가 작고(ZS, SZ) 표본크기가 1,000 이상일 때에 높은 1종오류를 보였으며, 표본크기가 2,000일 때에 가장 부풀려진 1종오류를

나타낸 후 다시 감소하는 모습을 보였다. 0이 아닌 경로계수의 크기가 보통 이상(ZM, ZL, MZ, LZ)이면 표본크기에 관계없이(표본크기가 가장 작은 100에서도) .05에 가깝거나 .05를 초과하는 1종오류를 보였다. 이는 그림 4에서

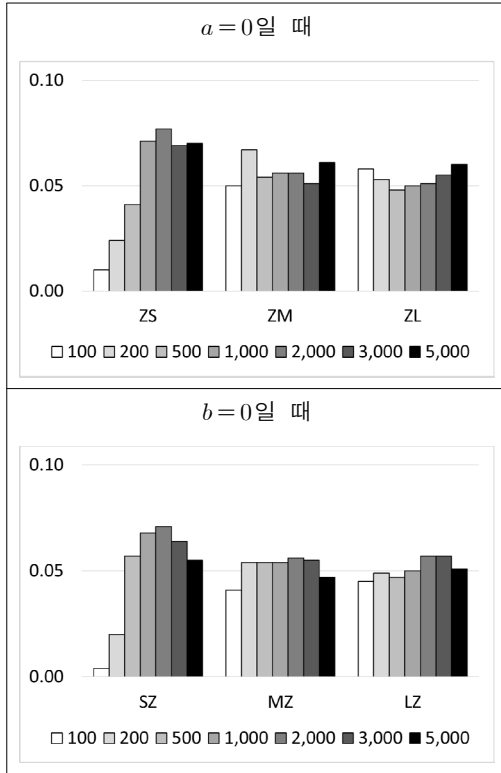


그림 4 편향조정 부트스트랩에서 경로계수 크기와 표본크기에 따른 1종오류

주. Z = Zero; S = Small; M = Medium; L = Large.

도 확인할 수 있다. 직접효과를 0으로 고정된 경로모형을 이용했던 Fritz 등(2012)의 연구 결과와 비교하면, 그들의 연구에서는 0이 아닌 경로계수가 보통 이상이고(medium, large) 표본 크기가 작을 때(50, 100)에 가장 부풀려진 1종오류를 보였고, 0이 아닌 경로계수가 작을 때 (small)에는 표본크기 500, 1,000에서 높은 1종오류를 보인 후 .05로 수렴해가는 모습을 보였다.

그림 5는  $a$  또는  $b$ 가 0일 때 나머지 경로의 계수 크기에 따른 세 가지 검정 방법의 1종오류를 표본크기에 대하여 평균낸 것이다.

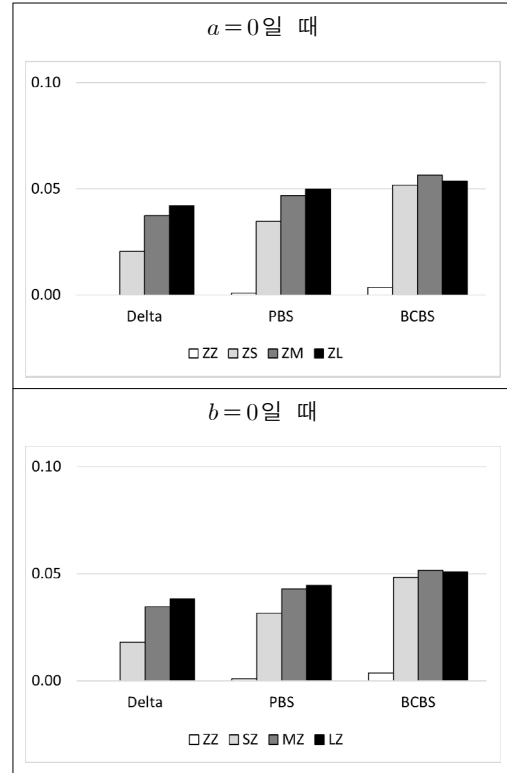


그림 5. 0이 아닌 경로(non-zero path)의 크기에 따른 매개효과 검정 방법간 1종오류 비교

주. Delta = 다변량 델타 방법; PBS = 백분위수 부트스트랩; BCBS = 편향조정 부트스트랩. Z = Zero; S = Small; M = Medium; L = Large.

다변량 델타 방법과 백분위수 부트스트랩에서는 0이 아닌 경로의 계수 크기가 커짐에 따라 1종오류가 높아지는 경향을 보였다. 편향조정 부트스트랩에서는 0이 아닌 경로의 계수가 작을(small) 때부터 이미 5%에 가까워지거나 5%를 초과하는 1종오류를 보였다. 이는 0이 아닌 경로계수의 크기가 작을(ZS, SZ) 때 특히 부풀려진 1종오류를 보였기 때문이다(그림 4). 한 가지 흥미로운 점은 0이 아닌 경로의 계수 크기가 같아도,  $a$ 가 0일 때의 1종오류가  $b$ 가 0일 때의 1종오류보다 높은 경향을

보였다. 예를 들어, 표본크기가 200이고 편향 조정 부트스트랩을 사용했을 때,  $ZM(a = 0, b = 0.36)$ 인 경우의 1종오류는 0.067이지만,  $MZ(a = 0.36, b = 0)$ 인 경우의 1종오류는 0.054이다. 경로계수의 체계적 변화에 따른 1종오류율의 변화를 더 구체적으로 확인하고 싶다면 Caron(2019)의 연구를 참고하길 바란다.

시뮬레이션 조건에 따른 비수렴률은 부록의 표 B에 제공된다. 표본크기가 100인 경우들을 제외하고는 모두 비수렴률 0%로 안정적인 결과를 보였다. 이는 연구 2의 자료생성 조건이 지표변수 3개와 요인부하 0.7로 한정되어 안정적인 추정이 이루어졌기 때문인 것으로 보인다.

### 결론 및 제언

본 연구는 매개효과 검정 방법과 그 중에서도 표준 절차로 자리잡은 부트스트랩에 대하여 연구자들의 이해도를 높이고 적절한 검정 방법을 선택하는 데에 도움이 되는 가이드라인을 제공하고자 하였다. 이를 위하여 몇 가지 매개효과 검정 방법을 소개하고 부트스트랩과 관련된 최신 연구 흐름을 살펴보았다. 또한 잠재변수를 사용한 매개모형에서 다양한 조건을 변화시키며 시뮬레이션을 실시하여 다변량 델타방법, 백분위수 부트스트랩과 편향조정 부트스트랩의 수행도를 비교하였다. 연구 1에서는 지표변수 개수, 효과크기, 요인부하, 표본크기 변화에 따른 세 가지 검정 방법의 검정력을 비교하고, 연구 2에서는 잠재변수마다 지표변수를 3개씩 갖는 매개모형에서 매개효과 모수를 0으로 만드는 경로  $a$ 와  $b$ 의 계수 조합(각 경로의 효과크기)과 표본크기 변

화에 따른 1종오류를 비교하였다.

시뮬레이션 결과를 요약하면 다음과 같다. 첫째, 다변량 델타 방법의 검정력은 현저히 낮았고, 편향조정 부트스트랩이 백분위수 부트스트랩보다 검정력은 높았지만(연구 1) 대부분의 시뮬레이션 조건에서 1종오류도 높은 경향을 보였다(연구 2). 이를 통해 잠재변수를 사용한 매개모형에서도 앞서 소개했던 2010년 이후 발표된 연구 결과들과 흐름을 같이 한다는 것을 확인할 수 있었다. 둘째, 다양한 요인이 매개분석의 검정력에 영향을 미치는 방식을 확인하였다. 지표변수 개수가 많을수록, 효과크기가 클수록, 요인부하가 클수록, 표본크기가 클수록 검정력이 증가는 경향을 보이며, 검정력이 가장 높은 방법인 편향조정 부트스트랩에서는 비교적 완만하게 검정력이 증가하였다. 셋째, 구조방정식 매개모형에서도 편향조정 부트스트랩이 부풀려진 1종오류를 발생시키는 것을 확인할 수 있었다. 다변량 델타 방법과 백분위수 부트스트랩에서는 표본크기가 작은 경우에는 1종오류가 작고 표본크기가 커질수록 1종오류가 증가하는 경향을 보였다. 반면 편향조정 부트스트랩은 0이 아닌 경로계수의 크기가 보통 이상이면 작은 표본크기(100)에서조차 1종오류가 이미 초과한 조건들이 있었고, 0이 아닌 경로계수가 작을 때(ZS, SZ) 특히 부풀려진 1종오류를 보였다.

본 연구가 매개효과 검정 방법에 대한 최근 연구흐름을 소개하고, 구조방정식 모형을 이용한 다양한 조건하에서의 시뮬레이션을 통해 백분위수 부트스트랩과 편향조정 부트스트랩의 수행을 확인함으로써 연구자들의 이해도를 높였지만, 시뮬레이션 연구 특성상 다음과 같은 한계가 있다. 첫째, 내용영역 연구에서 발생하는 모든 조건과 경우를 확인하기는 어렵

다는 현실적인 어려움이 있어 연구자의 판단 하에 구조방정식 매개모형의 대표적인 조건을 반영하여 시뮬레이션을 실시하였다. 둘째, 다양한 종류의 매개모형을 모두 포함시킬 수는 없었으므로 독립변수와 매개변수, 종속변수가 각각 한 개씩 있는 단순매개모형만을 사용하였다. 셋째, 현시점에서 추정의 정확성과 안정성을 동시에 만족시키면서도 SPSS, *Mplus* 같은 상용 프로그램을 통해 연구자들이 쉽게 이용할 수 있는 부트스트랩을 중심으로 비교함으로써 베이지안 매개분석 등의 최신 검정 방법은 포함시키지 않았다(Miočević et al., 2017). 이러한 제한점들을 바탕으로 향후 진행될 연구에서는 중다매개모형, 종단매개모형, 다층매개모형 등 모형과 관련된 범위를 확장하거나 더욱 다양한 검정 방법을 반영한 시뮬레이션을 통해 수행을 비교하는 연구로 확장할 수 있다.

본 연구 결과를 바탕으로 매개모형을 사용할 연구자들에게 제공할 수 있는 가이드라인은 다음과 같다. 첫째, 다변량 델타 방법의 사용을 피한다. 다변량 델타 방법이 가정하는 매개효과와 정규성은 이론적으로도 경험적으로도 만족될 수 없으며, 다변량 델타 방법의 낮은 검정력은 본 연구의 시뮬레이션을 통해서도 확인할 수 있었다. 둘째, 매개효과 검정 시 다변량 델타 방법, 백분위수 부트스트랩, 편향조정 부트스트랩 세 가지 방법 중에서는 백분위수 부트스트랩을 사용한다. 부풀려진 1종오류를 발생시키는 위험을 감수하고도 검정력을 높이는 방법을 선택하는 것은 적절하지 않기 때문이다. 국내에서 출간된 구조방정식 서적들이 편향조정 부트스트랩을 권장하고 있고 다수의 연구자들이 이를 받아들여 사용하고 있지만, 추정의 정확성과 안정성, 특정 매개분석 방법을 실시할 수 있는 프로그램에 대

한 접근성, 기존의 연구들을 통해 검증된 정도 등을 고려한다면 현시점에서 연구자가 선택할 수 있는 최선의 매개효과 검정 방법은 백분위수 부트스트랩일 것으로 판단된다. 만약 원자료가 없거나, 부트스트랩을 적용할 수 없는 경우라면 본 연구에서 소개했던 다른 검정 방법들을 참고할 수 있다.

본 연구는 잠재변수를 사용한 매개효과와 검정 방법, 특히 부트스트랩에 대한 연구자들의 이해를 높이기 위한 목적으로 최근 연구 흐름을 소개하고 시뮬레이션을 통해 그 수행도를 비교한 연구이다. 본 연구의 결과를 바탕으로 연구자들이 다변량 델타 방법, 백분위수 부트스트랩과 편향조정 부트스트랩의 특성을 이해하고 다양한 연구 장면에서 적절한 검정 방법을 선택할 수 있을 것으로 기대한다.

## 참고문헌

- Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, 40, 37-47.  
<https://doi.org/10.2307/2094445>
- Bae, B. (2016). *Structural equation modeling with Mplus 7.0*. Chungnam.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.  
<https://doi.org/10.1037//0022-3514.51.6.1173>
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and

- interval estimation for indirect effects. *Multivariate Behavioral Research*, 45, 661-701.  
<https://doi.org/10.1080/00273171.2010.498292>
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.  
<https://doi.org/10.1002/9781118619179>
- Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 20, 115-140.  
<https://doi.org/10.2307/271084>
- Caron, P.-O. (2019). A comparison of the type I error rates of three assessment methods for indirect effects. *Journal of Statistical Computation and Simulation*, 89(8), 1343-1356.  
<https://doi.org/10.1080/00949655.2019.1577858>
- Chen, D., & Fritz, M. S. (2021). Comparing alternative corrections for bias in the bias-corrected bootstrap test of mediation. *Evaluation & the Health Professions*, 44(4), 416-427.  
<https://doi.org/10.1177/01632787211024356>
- Cerin, E., & MacKinnon, D. P. (2009). A commentary on current practice in mediating variable analyses in behavioural nutrition and physical activity. *Public Health Nutrition*, 12(8), 1182-1188.  
<https://doi.org/10.1017/S1368980008003649>
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11(2), 296-325.  
<https://doi.org/10.1177/1094428107300343>
- Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(2), 227-246.  
<https://doi.org/10.1080/10705510709336745>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.  
<https://doi.org/10.4324/9780203771587>
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cramer, H. (1946). *Mathematical methods of statistics*. Princeton University Press.  
<https://doi.org/10.1515/9781400883868>
- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3), 189-228.  
<https://doi.org/10.1214/ss/1032280214>
- Duckworth, A. L., White, R. E., Matteucci, A. J., Shearer, A., & Gross, J. J. (2016). A stitch in time: Strategic self-control in high school and college students. *Journal of Educational Psychology*, 108(3), 329-341.  
<https://doi.org/10.1037/edu0000062>
- Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.  
<http://www.jstor.org/stable/2958830>
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171-185.  
<https://doi.org/10.1080/01621459.1987.10478410>
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical

- accuracy. *Statistical Science*, 1(1), 54-75.  
<http://www.jstor.org/stable/2245500>
- Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009).  $R^2$  effect-size measures for mediation analysis. *Behavior Research Methods*, 41, 486-498.  
<https://doi.org/10.3758/BRM.41.2.486>
- Fairchild, A. J., & McDaniel, H. L. (2017). Best (but oft-forgotten) practices: mediation analysis. *The American Journal of Clinical Nutrition*, 105(6), 1259-1271.  
<https://doi.org/10.3945/ajcn.117.152546>
- Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 24-38.  
<https://doi.org/10.1080/10705511.2014.935266>
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18(3), 233-239.  
<https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 47(1), 61-87.  
<https://doi.org/10.1080/00273171.2012.640596>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th edition). Cengage Learning, EMEA.
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. Guilford Press.
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24(10), 1918-1927.  
<https://doi.org/10.1177/0956797613480187>
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation review*, 5(5), 602-619.  
<https://doi.org/10.1177/0193841X8100500502>
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey. (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233-265). McGraw-Hill.
- Kim, H., & Kim, S.-Y. (2020). Exploring the possible direction of interpreting inconsistent mediational effect. *Korean Journal of Psychology: General*, 39(1), 91-115.  
<https://doi.org/10.22257/kjp.2020.3.39.1.91>
- Kim, H.-W., Lee, H.-B., & Shin, Y.-H. (2015). A mediation analysis of absorption capacity by bootstrapping technique in multiple mediator model. *Journal Of The Korea Society For Simulation*, 24(4), 89-96.  
<https://doi.org/10.9709/JKSS.2015.24.4.089>
- Kim, S.-Y. (2012). Sample size requirements in single- and multiphase growth mixture models: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3), 457-476.  
<https://doi.org/10.1080/10705511.2012.687672>
- Kim, S.-Y. (2016). *Fundamentals and extensions of structural equation modeling: With Mplus examples*. Hakjisa.
- Kirby, K. N., & Gerlanc, D. (2013). BootES: An

- R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, 45, 905-927.  
<https://doi.org/10.3758/s13428-013-0330-5>
- Kline, R. B. (2016). *Principles and practices of structural equation modeling* (4th ed.). The Guilford Press.
- Lachowicz, M. J., Preacher, K. J., & Kelley, K. (2018). A novel measure of effect size for mediation analysis. *Psychological Methods*, 23(2), 244-261.  
<https://doi.org/10.1037/met0000165>
- Lee, H. (2018). *Analysis of mediating effect, moderating effect, and moderated mediating effect using Mplus*. Shinyoungsa.
- Liu, X., & Wang, L. (2019). Sample size planning for detecting mediation effects: A power analysis procedure considering uncertainty in effect size estimates. *Multivariate Behavioral Research*, 54(6), 822-839.  
<https://doi.org/10.1080/00273171.2019.1593814>
- Lockman, J. D., & Servaty-Seib, H. L. (2018). Testing the predictions of the existential constructivist theory of suicide in a college student sample. *Journal of Counseling Psychology*, 65(3), 294-307.  
<https://doi.org/10.1037/cou0000278>
- Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 29(3), 513-524.  
<https://doi.org/10.1111/j.2517-6161.1967.tb00713.x>
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Routledge.  
<https://doi.org/10.4324/9780203809556>
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2), 144-158.  
<https://doi.org/10.1177/0193841X9301700202>
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39, 384-389.  
<https://doi.org/10.3758/BF03193007>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83-104.  
<https://doi.org/10.1037/1082-989X.7.1.83>
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99-128.  
[https://doi.org/10.1207/s15327906mbr3901\\_4](https://doi.org/10.1207/s15327906mbr3901_4)
- Malone, J. C., Liu, S. R., Vaillant, G. E., Rentz, D. M., & Waldinger, R. J. (2016). Midlife Eriksonian psychosocial development: Setting the stage for late-life cognitive and emotional health. *Developmental Psychology*, 52(3), 496-508.  
<https://doi.org/10.1037/a0039875>
- Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). *Selected tables in mathematical statistics: The product of two normally distributed random variables*. American Mathematical Society.
- Meeker, W. Q., & Escobar, L. A. (1994). An algorithm to compute the CDF of the product

- of two normal random variables. *Communications in Statistics-Simulation and Computation*, 23(1), 271-280.  
<https://doi.org/10.1080/03610919408813168>
- Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(5), 666-683.  
<https://doi.org/10.1080/10705511.2017.1312407>
- Miočević, M., O'Rourke, H. P., MacKinnon, D. P., & Brown, H. C. (2018). Statistical properties of four effect-size measures for mediation models. *Behavior Research Methods*, 50, 285-301.  
<https://doi.org/10.3758/s13428-017-0870-1>
- Muthén, B. O. (2014). *Bias-corrected bootstrap*. Mplus Discussion.  
<http://www.statmodel.com/discussion/messages/11/628.html?1557166741>
- Muthén, B. O. & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 12-23.  
<https://doi.org/10.1080/10705511.2014.935843>
- Muthén, L. K. (2014). *Which kind of bootstrapping?* Mplus Discussion.  
<http://www.statmodel.com/discussion/messages/9/19169.html?1407875009>
- Muthén, L. K., & Muthén, B. O. (1998-2021). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599-620.  
[https://doi.org/10.1207/S15328007SEM0904\\_8](https://doi.org/10.1207/S15328007SEM0904_8)
- Muthén, L. K., & Muthén, B. O. (2008-2018). *Test indirect effect using bootstrap*. Mplus Discussion.  
<http://www.statmodel.com/discussion/messages/11/3625.html?1550544811>
- Oh, M.-H., Kim, J., & Lee, M. (2018). Effects of family stress on family adaptation to a family member with mental illness: Mediating effects of family strengths and perceived stigma. *Journal of Korean Academy of Psychiatric and Mental Health Nursing*, 27(3), 274-283.  
<https://doi.org/10.12934/jkpmhn.2018.27.3.274>
- Park, S.-M., & Bak, B.-G. (2014). A trend analysis of the mediation effect researches and a suggestion for the new direction of them. *The Korean Journal of Educational Psychology*, 28(1), 41-70.  
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06763977>
- Park, S., & Hur, H. (2021). Relationship between clinical nurses' job stress and medication safety performance: Mediating effect of fatigue. *Korean Journal of Adult Nursing*, 33(3), 283-293.  
<https://doi.org/10.7475/kjan.2021.33.3.283>
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717-731.  
<https://doi.org/10.3758/BF03206553>



- Preacher, K. J., Hayes, A. F. (2008a). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879-891.  
<https://doi.org/10.3758/BRM.40.3.879>
- Preacher, K. J., & Hayes, A. F. (2008b). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13-54). Sage.
- Preacher, K. J., & Kelly, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, 16(2), 93-115.  
<https://doi.org/10.1037/a0022658>
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6(2), 77-98.  
<https://doi.org/10.1080/19312458.2012.679848>
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5(6), 359-371.  
<https://doi.org/10.1111/j.1751-9004.2011.00355.x>
- Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods*, 5(2), 230-240.  
<https://doi.org/10.1037/1082-989X.5.2.230>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422-445.  
<https://doi.org/10.1037/1082-989X.7.4.422>
- Sim, M., Kim, S.-Y., & Suh, Y. (2022). Sample size requirements for simple and complex mediation models. *Educational and Psychological Measurement*, 82(1).  
<https://doi.org/10.1177/00131644211003261>
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.  
<https://doi.org/10.2307/270723>
- Sobel, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociological Methods & Research*, 16(1), 155-176.  
<https://doi.org/10.1177/0049124187016001006>
- Springer, M. D., & Thompson, W. E. (1966). The distribution of independent random variables. *SIAM Journal on Applied Mathematics*, 14(3), 511-526.  
<https://doi.org/10.1137/0114046>
- Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum. *Psychometrika*, 55(2), 337-352.  
<https://doi.org/10.1007/BF02295291>
- Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11(2), 241-269.  
<https://doi.org/10.1177/1094428107300344>
- Tofighi, D., & Kelley, K. (2020). Indirect effects in sequential mediation models: Evaluating methods for hypothesis testing and confidence

- interval formation. *Multivariate Behavioral Research*, 55(2), 188-210.  
<https://doi.org/10.1080/00273171.2019.1618545>
- Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, 43(3), 692-700.  
<https://doi.org/10.3758/s13428-011-0076-x>
- Valente, M. J., Gonzalez, O., Miočević, M., & MacKinnon, D. P. (2016). A note on testing mediated effects in structural equation models: Reconciling past and current research on the performance of the test of joint significance. *Educational and Psychological Measurement*, 76(6), 889-911.  
<https://doi.org/10.1177/0013164415618992>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.
- Yuan, K.-H., Zhang, Z., & Zhao, Y. (2017). Reliable and more powerful methods for power analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 315-330.  
<https://doi.org/10.1080/10705511.2016.1276836>
- Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological methods*, 14(4), 301-322.  
<https://doi.org/10.1037/a0016972>
- Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197-206.  
<https://doi.org/10.1086/651257>
- 1차원고접수 : 2021. 11. 07.  
2차원고접수 : 2022. 02. 01.  
3차원고접수 : 2022. 04. 09.  
최종게재결정 : 2022. 04. 28.

## Comparison of methods for testing mediation effects under structural equation modeling framework: Percentile bootstrap and bias-corrected bootstrap

Mikyung Sim<sup>1)</sup>

Youngsuk Suh<sup>2)</sup>

Su-Young Kim<sup>1)</sup>

<sup>1)</sup>Department of Psychology, Ewha Womans University

<sup>2)</sup>Korean Education & Psychology Institute

Mediation models are widely used in various fields such as psychology, education, and business administration. Some traditional methods like Baron and Kenny's approach and multivariate delta method were used before, but now bootstrap methods have become a standard procedure. Since the bias-corrected bootstrap (BC bootstrap) was proposed to correct for the potential bias in the confidence intervals (CIs) obtained by the percentile bootstrap, the BC method has been frequently used domestically based on some old research that the BC bootstrap provides more accurate CIs and higher power. However, recent studies have reported that the BC method may not be appropriate for testing mediating effects because it produces inflated Type I error rates. The goal of the present study is to enhance researchers' understanding about testing methods for the mediation effect by examining the recent research trend of testing methods for mediating effects and by comparing the performance of the percentile bootstrap and the BC bootstrap through a series of simulations under various conditions. In Study 1, the results show that the power was high in the order of BC bootstrap, percentile bootstrap, and multivariate delta method. As the number of indicators, the effect size, the factor loading, and the sample size increased, the power also increased. Study 2 demonstrates that Type 1 error rates were also high in the order of BC bootstrap, percentile bootstrap, and multivariate delta method, confirming that the BC bootstrap also generates the highest Type 1 error in structural equation mediation models. Finally, the limitations of this study and guidelines based on the simulation results are discussed.

*Keywords* : mediation model, mediation effect, bootstrap, power, Type I error

부 록

표 A. 연구 1의 시뮬레이션 조건에 따른 비수렴률

효과 크기	요인 부하	표본 크기	2 지표변수			3 지표변수			4 지표변수		
			Delta	PBS	BCBS	Delta	PBS	BCBS	Delta	PBS	BCBS
Small	0.5	100	0.898	0.874	0.874	0.389	0.350	0.350	0.063	0.059	0.059
		200	0.844	0.815	0.815	0.048	0.047	0.047	0.000	0.000	0.000
		500	0.677	0.634	0.634	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.378	0.349	0.349	0.000	0.000	0.000	0.000	0.000	0.000
	0.7	100	0.869	0.865	0.865	0.011	0.011	0.011	0.000	0.000	0.000
		200	0.746	0.736	0.736	0.000	0.000	0.000	0.000	0.000	0.000
		500	0.401	0.400	0.400	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.108	0.107	0.107	0.000	0.000	0.000	0.000	0.000	0.000
Medium	0.5	100	0.732	0.682	0.682	0.205	0.179	0.179	0.032	0.029	0.029
		200	0.426	0.405	0.405	0.008	0.008	0.008	0.000	0.000	0.000
		500	0.065	0.063	0.063	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.002	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
	0.7	100	0.442	0.437	0.437	0.002	0.002	0.002	0.000	0.000	0.000
		200	0.119	0.119	0.119	0.000	0.000	0.000	0.000	0.000	0.000
		500	0.005	0.005	0.005	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Large	0.5	100	0.552	0.481	0.481	0.134	0.108	0.108	0.041	0.036	0.036
		200	0.172	0.148	0.148	0.004	0.003	0.003	0.000	0.000	0.000
		500	0.005	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.7	100	0.129	0.128	0.128	0.000	0.000	0.000	0.000	0.000	0.000
		200	0.009	0.009	0.009	0.000	0.000	0.000	0.000	0.000	0.000
		500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		1,000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

주. 비수렴률이 5%를 초과하는 결과에 음영 처리. Delta = 다변량 델타 방법; PBS = 백분위수 부트스트랩; BCBS = 편향조정 부트스트랩.

표 B. 연구 2의 시뮬레이션 조건에 따른 비수렴률

검정 방법	경로계수			표본크기						
	a 경로	b 경로	조합	100	200	500	1,000	2,000	3,000	5,000
Delta	0.00	0.00	ZZ	0.013	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.14	ZS	0.011	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.36	ZM	0.007	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.51	ZL	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	0.14	0.00	SZ	0.012	0.000	0.000	0.000	0.000	0.000	0.000
	0.36	0.00	MZ	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	0.51	0.00	LZ	0.004	0.000	0.000	0.000	0.000	0.000	0.000
PBS	0.00	0.00	ZZ	0.013	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.14	ZS	0.011	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.36	ZM	0.007	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.51	ZL	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	0.14	0.00	SZ	0.012	0.000	0.000	0.000	0.000	0.000	0.000
	0.36	0.00	MZ	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	0.51	0.00	LZ	0.004	0.000	0.000	0.000	0.000	0.000	0.000
BCBS	0.00	0.00	ZZ	0.013	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.14	ZS	0.011	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.36	ZM	0.007	0.000	0.000	0.000	0.000	0.000	0.000
	0.00	0.51	ZL	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	0.14	0.00	SZ	0.012	0.000	0.000	0.000	0.000	0.000	0.000
	0.36	0.00	MZ	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	0.51	0.00	LZ	0.004	0.000	0.000	0.000	0.000	0.000	0.000

주. 비수렴률이 5%를 초과하는 결과는 발생하지 않음. Delta = 다변량 델타 방법; PBS = 백분위수 부트스트랩; BCBS = 편향조정 부트스트랩. Z = Zero; S = Small; M = Medium; L = Large.