



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

조 영 일 교수 지도

석사학위 청구논문

보조변수를 활용한

결측치 처리 과정에서

주성분 분석의 적용에 대한 연구

2020

성신여자대학교 대학원

심리학과

유 주 연

보조변수를 활용한  
결측치 처리 과정에서  
주성분 분석의 적용에 대한 연구

조영일 교수 지도

이 논문을 석사학위논문으로 제출함

2020년 6월

성신여자대학교 대학원

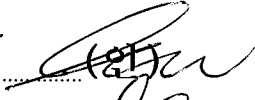


심리학과

유주연

# 인 준 서

유주연의 석사학위 논문으로 인준함

2020년 6월

심사위원장 박성희 (인)   
심사위원 서수연 (인)   
심사위원 조영일 (인) 

성신여자대학교 대학원

## 논문개요

본 논문은 완전정보최대가능성법을 활용한 결측치 추정에서 보조변수 간 상관에 따른 주성분 분석의 수행을 비교하는 시뮬레이션 연구다. 이를 위해 측정변수와 보조변수 간 상관의 크기, 보조변수 간 상관의 형태, 결측치의 비율, 표본의 크기, 모형의 분석 방법에 따라 조건을 설정했다. 이에 따라 통계프로그램 Mplus를 활용해 576( $4 \times 3 \times 4 \times 4 \times 3$ )개의 조건을 만족하는 자료를 생성했으며, 통계프로그램 R을 통해 500회씩 반복 시행 및 분석했다.

분석 결과, 주성분 분석은 연구에서 설정한 모든 조건에서 보조변수로 지정한 9개의 변수를 2개의 주성분으로 축소했다. 주성분 분석을 통해 구한 주성분을 보조변수로 활용해 결측치를 추정할 경우, 보조변수를 사용하지 않고 결측치를 추정할 때보다 추정이 더 정확했으며 오차와 편향이 작았다. 주성분을 보조변수로 활용한 결측치 추정은 보조변수를 그대로 투입한 추정과 유사한 수행 능력을 보였다. 또한 표본의 크기가 작고 결측치의 비율이 높은 극단적인 조건에서 보조변수를 그대로 투입할 경우 모형이 수렴하지 않는 문제가 발생하나, 주성분 분석을 활용할 경우 이를 해결할 수 있었다.

다만 보조변수 간 상관이 이질적인 조건에서 주성분 분석을 활용한 결측치 추정은 보조변수를 그대로 사용한 추정 방법에 비해 수행이 다소 떨어지는 것을 확인했다. 따라서 본 논문은 결측치 추정 과정에서 주성분 분석의 수행 능력을 확인하고, 어떤 상황에서 유용하게 활용될 수 있는지 지표를 제시했다. 마지막으로 본 연구의 한계와 후속 연구에 대한 제언을 논의했다.

**주요어** : 결측치, 보조변수, 주성분 분석, 완전정보최대가능성법, 시뮬레이션, 통계 프로그램 R, 통계 프로그램 Mplus

# 목 차

## 논문개요

I. 서론 .....	1
1. 연구의 필요성 및 목적 .....	1
II. 이론적 배경 .....	10
1. 결측치 메커니즘 .....	10
1) Missing completely at random (MCAR) .....	11
2) Missing at random (MAR) .....	13
3) Missing not at random (MNAR) .....	14
2. 결측치 처리 방법 .....	16
1) 다중대체법(Multiple Imputation: MI) .....	16
2) 완전정보최대가능성법(Full Information Maximum Likelihood: FIML) .....	19
3. 보조변수(Auxiliary Variable) .....	21
1) Inclusive Strategy .....	25
2) Restrictive Strategy .....	25
4. 주성분분석(Principal Component Analysis: PCA) .....	26
1) 주성분분석 .....	26
2) 주성분의 수 결정 .....	28

3) 보조변수 투입에서의 주성분분석 활용 .....	29
<b>III. 연구문제 및 가설 .....</b>	<b>32</b>
1. 연구문제 1 .....	32
2. 연구문제 2 .....	32
<b>IV. 연구방법 .....</b>	<b>33</b>
1. 자료 생성 .....	33
1) 몬테카를로 시뮬레이션 .....	33
2) 시뮬레이션 조건 .....	34
2. 자료 분석 .....	39
1) 모수 편향(Parameter bias) .....	39
2) RMSE(Root-mean square error) .....	39
3) 신뢰구간(Confidence Interval: CI) 범위 .....	40
4) 수렴(Convergence) 여부 .....	40
<b>V. 연구결과 .....</b>	<b>41</b>
1. 주성분의 수 .....	41
2. 수렴(Convergence) 여부 .....	41
3. 모수 편향(Parameter bias) .....	42
1) X와 Y의 상관 .....	42
2) Y의 평균 .....	46
3) Y의 분산 .....	48

4. RMSE .....	51
1)X와 Y의 상관 .....	51
2)Y의 평균 .....	53
3)Y의 분산 .....	56
3. 신뢰구간(Confidence Interval) .....	58
1)신뢰구간의 범위 .....	58
2)신뢰구간의 너비 .....	61
VI. 논의 .....	64

참고문헌

ABSTRACT

부록

## 표 목 차

- <표 1> 시뮬레이션 조건
- <표 2> Uniform 조건에서의 보조변수 간 상관
- <표 3> Moderate 조건에서의 보조변수 간 상관
- <표 4> Low 조건에서의 보조변수 간 상관
- <표 5> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 모수 편향
- <표 6> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 모수 편향
- <표 7> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 모수 편향
- <표 8> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 RMSE
- <표 9> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 RMSE
- <표 10> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 RMSE
- <표 11> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간

<표 12> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 범위

<표 13> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 모수 편향(1)

<표 14> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 모수 편향(2)

<표 15> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 모수 편향(3)

<표 16> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y$ 의 모수 편향(1)

<표 17> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y$ 의 모수 편향(2)

<표 18> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y$ 의 모수 편향(3)

<표 19> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(1)

<표 20> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(2)

<표 21> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(3)

<표 22> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 모수 편향(1)

<표 23> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 모수 편향(2)

<표 24> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 모수 편향(3)

<표 25> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y$ 의 모수 편향(1)

<표 26> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y$ 의 모수 편향(2)

<표 27> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y$ 의 모수 편향(3)

<표 28> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(1)

<표 29> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(2)

<표 30> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(3)

<표 31> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 모수 편향(1)

<표 32> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 모수 편향(2)

<표 33> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 모수 편향(3)

<표 34> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y$ 의 모수 편향(1)

<표 35> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y$ 의 모수 편향(2)

<표 36> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y$ 의 모수 편향(3)

<표 37> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(1)

<표 38> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(2)

<표 39> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(3)

<표 40> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 모수 편향(1)

<표 41> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 모수 편향(2)

<표 42> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 모수 편향(3)

<표 43> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 모수 편향(1)

<표 44> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 모수 편향(2)

<표 45> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 모수 편향(3)

<표 46> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(1)

<표 47> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(2)

<표 48> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 모수 편향(3)

<표 49> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 RMSE(1)

<표 50> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 RMSE(2)

<표 51> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 RMSE(3)

<표 52> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y$ 의 RMSE(1)

<표 53> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y$ 의 RMSE(2)

<표 54> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y$ 의 RMSE(3)

<표 55> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2$ 의 RMSE(1)

<표 56> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2$ 의 RMSE(2)

<표 57> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2$ 의 RMSE(3)

<표 58> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 RMSE(1)

<표 59> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 RMSE(2)

<표 60> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 RMSE(3)

<표 61> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y$ 의 RMSE(1)

<표 62> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y$ 의 RMSE(2)

<표 63> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y$ 의 RMSE(3)

<표 64> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2$ 의 RMSE(1)

<표 65> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2$ 의 RMSE(2)

<표 66> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2$ 의 RMSE(3)

<표 67> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 RMSE(1)

<표 68> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 RMSE(2)

<표 69> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 RMSE(3)

<표 70> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y$ 의 RMSE(1)

<표 71> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y$ 의 RMSE(2)

<표 72> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y$ 의 RMSE(3)

<표 73> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2$ 의 RMSE(1)

<표 74> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2$ 의 RMSE(2)

<표 75> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2$ 의 RMSE(3)

<표 76> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 RMSE(1)

<표 77> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 RMSE(2)

<표 78> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 RMSE(3)

<표 79> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 RMSE(1)

<표 80> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 RMSE(2)

<표 81> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y$ 의 RMSE(3)

<표 82> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 RMSE(1)

<표 83> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 RMSE(2)

<표 84> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2$ 의 RMSE(3)

<표 85> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(1)

<표 86> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(2)

<표 87> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(3)

<표 88> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(1)

<표 89> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(2)

<표 90> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(3)

<표 91> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(1)

<표 92> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(2)

<표 93> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(3)

<표 94> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(1)

<표 95> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(2)

<표 96> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
(3)

<표 97> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의  
너비(1)

<표 98> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의  
너비(2)

<표 99> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의  
너비(3)

<표 100> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
의 너비(1)

<표 101> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간  
의 너비(2)

<표 102> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간

의 너비(3)

<표 103> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의 너비(1)

<표 104> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의 너비(2)

<표 105> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의 너비(3)

<표 106> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의 너비(1)

<표 107> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의 너비(2)

<표 108> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}$ 의 90% 신뢰구간의 너비(3)

## 그림 목 차

<그림 1> 다중대체법의 대체값 생성 .....	4
<그림 2> 직사각형 형태로 표현한 결측치 패턴 .....	11
<그림 3> MCAR 결측치 .....	12
<그림 4> MAR 결측치 .....	14
<그림 5> MNAR 결측치 .....	15
<그림 6> 다중대체법 .....	16
<그림 7> Extra DV 모델(관찰 변수) .....	22
<그림 8> Extra DV 모델(잠재 변수) .....	23
<그림 9> Saturated correlates 모델(관찰 변수) .....	24
<그림 10> Saturated correlates 모델(잠재 변수) .....	24
<그림 11> 주성분 분석 .....	27

<그림 12> 연구 모형 ..... 38

# I. 서론

## 1. 연구의 필요성 및 목적

심리학은 인간의 마음과 행동에서 나타나는 다양성을 이해하고 그 원인과 과정을 설명하는 데에 목적을 두고 있다. 이를 위한 자료를 수집하는 과정에서 연구자는 문헌 조사, 참여 관찰, 면접법 등 다양한 방법을 이용하는데, 그 중에서도 설문지법은 시간과 비용적인 측면에서 효율적이고 참여 관찰, 면접법에 비해 연구자의 주관의 개입 가능성이 낮으며 표본의 크기가 큰 경우에도 적용이 가능하다는 이점을 가지고 있어 널리 사용되고 있다. 그러나 설문지법은 다른 조사 방법에 비해 무응답률이 높은 편인데, 이러한 자료의 결측은 문항에 대한 이해 부족, 응답 거부, 조사 불참 등 다양한 원인에 기인한다(Ibrahim, 2001). 특히 여러 회차에 걸쳐 조사가 진행되는 패널 자료의 경우 조사가 진행될수록 표본의 수가 감소하는 현상이 나타나는데, 한국아동청소년패널(KCYPS), 한국복지패널조사(KOWEPS), 한국노동패널조사(KLIPS) 등 국내에서 수집되고 있는 패널 자료에서도 1차 조사시기 이후 지속적으로 탈락이 증가하는 것을 확인할 수 있다(이경희, 민인식, 2016; 이화정, 강석복, 2014; 조태경, 손창균, 2015).

발생한 결측치를 무시하고 관찰된 자료에 근거하여 분석을 실시할 경우, 본래 연구자가 계획한 것보다 적은 수의 표본을 분석에 이용하게 되어 검증력이 감소하는 것은 물론 표본이 모집단을 제대로 대표할 수 없게 된 경우 추정에 편향(bias)이 발생할 수 있다(Little & Rubin, 2002). 또한 결측치는 발생 메커니즘에 따라 완전무작위결측(Missing completely at random, MCAR), 무작위결측(Missing at random, MAR), 작위결측(Missing not at

random, MNAR)으로 구분되는데(Rubin, 1976), 연구자가 이를 고려하지 않을 경우 결측치 처리 과정을 거처도 분석 결과에 편향이 발생한다. 따라서 연구자가 자신의 연구에서 발생한 결측치를 어떻게 처리했는지 보고하지 않을 경우, 해당 연구를 참고하거나 후속 연구를 진행하는 과정에서 결과를 정확하게 해석하는 데 어려움이 발생한다(Schlomer, Bauman & Card, 2010).

이러한 문제점을 극복하고자 결측치의 발생을 줄이기 위해 자료 수집 단계에서 대체 가능한 자료를 함께 수집하거나 분석 단계에서 단일 문항 대신 다수의 문항이 조합된 결과를 이용하는 등 여러 기법을 사용해 정보의 누락을 줄일 순 있으나(Hardy, Allore & Studenski 2009; Little et al 2012), 현실적으로 결측치를 완전히 예방하기는 어렵다. 이에 연구자들은 자료에 존재하는 결측치를 분석단계에서 처리하기 위해 여러 방법을 이용해왔는데, 전통적인 방법으로는 완전제거법(Listwise Deletion), 대응쌍제거법(Pairwise Deletion), 평균대체법(Mean Imputation), 회귀대체법(Regression Imputation) 등이 있다. 하지만 전통적인 방법들은 결측치를 처리하는 과정에서 다양한 문제들이 보고되었다.

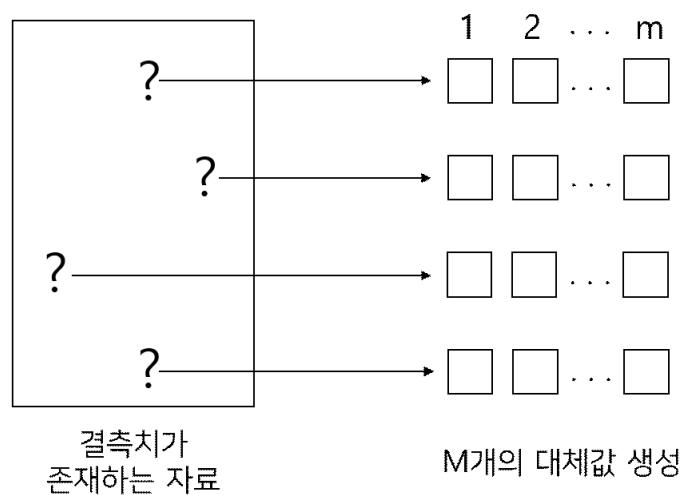
완전제거법은 전체 자료에서 결측치가 하나라도 존재할 경우, 해당 응답자를 분석에서 제외한다. 완전제거법은 결측치를 처리하는 가장 간편한 방법이나, 다수의 표본이 분석에서 제외되기 때문에 그만큼 정보의 누락이 발생하며 검증력이 떨어지고 분석 결과 또한 편향될 가능성이 높다(King, Joseph & Scheve. 1998). 몬테카를로 시뮬레이션을 활용한 연구에 의하면, 자료의 2%가 무작위로 발생한 결측치고 연구자가 이를 단순히 표본을 분석에서 제외하는 방식으로 처리할 경우, 최대 18.3%의 자료를 손실하게 된다(Kim & Curry, 1977). 그럼에도 불구하고 완전제거법은 통계 프로그램 패키지에서 기본 옵션으로 설정되어 있는 경우가 많다(Little & Rubin, 2002).

평균대체법과 회귀대체법은 단일대체법의 일종으로 해당 변수의 평균 혹은 회귀식을 통해 구해진 값을 투입하는 처리 방법이다. 결측치를 특정 값으로 대체하여 사용하기 때문에 완전한 형태의 데이터를 분석에 사용할 수 있지만, 결측치의 비율이 높을수록 분석에 투입되는 표본의 수 N은 증가하나 편차는 증가하지 않기 때문에 해당 변수의 분산이 감소하여 추정에 편향이 발생하거나 표준오차가 과소추정되는 문제가 발생한다(Baraldi & Enders, 2010; Roth, 1994; Scheffer, 2002).

완전제거법, 단일대체법 등 전통적인 방법들이 결측치를 올바르게 처리하기보다는 검증력을 떨어트리거나 편향을 발생시키는 등의 문제를 발생시킨다. 이러한 문제를 극복하기 위하여 최근에 많이 활용되는 방법은 다중대체법(Multiple Imputation; MI)과 완전정보최대가능성법(Full Information Maximum Likelihood Estimation; FIML)이다.

다중대체법은 결측치에 투입할 대체값을 다수 생성한 후, 각각의 대체값을 투입한 완전한 형태의 데이터를 기존의 방식대로 추정에 사용하는 방법으로 투입 단계, 추정 단계, 통합 단계를 거친다(Baraldi & Enders, 2010). 일반적으로 대체값은 10개 이상 생성하는 것을 권장하며 결측치의 비율이 높을수록 더 많은 대체값이 필요하다(Barnes, Lindborg, & Seaman, 2006). 다중대체법은 다양한 상황에서 적용이 용이하며 기존에 사용된 평균대체법, 회귀대체법과 달리 확정된 하나의 값을 투입하는 대신, <그림 1>과 같이 다수의 대체값을 투입하고 그 추정결과를 하나로 통합하기 때문에 편향과 표준오차의 과소추정 문제를 해결할 수 있다는 장점이 있다(Allison, 2003). 다양한 학문 분야에서 완전제거법, 단일대체법, 다중대체법의 결측치 처리 수행을 비교했을 때 다중대체법은 전통적인 방법에 비해 추정치의 편향이 적고 표준오차가 과소추정되거나 과대추정되는 경향이 적다는 일관적인 연구 결과가 보고되고 있다(Cheema, 2014; Newman, 2003).

이러한 긍정적인 효과에 반하여, 여전히 자료에 임의의 숫자를 투입한다는 점에서 다중대체법은 모수 추정에 있어 완전히 효과적인 방법이라고 보기는 어렵다(Allison, 2003). 대체값을 생성하는 과정에 분산이 감소하는 것을 방지하기 위한 임의의 변동을 포함하고 있기 때문에 확정적인 결과를 산출하지 않아 동일한 자료와 모형을 사용하더라도 통합단계에선 매번 추정치가 다르게 구해진다는 점도 문제로 꼽힐 수 있다. 또한 자료의 분포가 한쪽으로 치우친 형태일 때 수행 결과가 다소 편향되는 결과 또한 보고되고 있다(Yuan, Yang, & Bentler, 2012). 더불어 다중대체법을 구조방정식 장면에서 활용할 경우, 합치도 지수를 구하는 데에 어려움이 있으며 관련 연구도 미흡하다(Enders & Mansolf, 2018). 본 연구에서는 다중대체법을 직접적으로 활용하고 있진 않으나, 주요하게 다룰 보조 변수(Auxiliary Variable)가 다중대체법의 맥락에서 설명될 수 있기 때문에 이에 대해 이론적 배경에서 설명하고자 한다.



<그림 1> 다중대체법의 대체값 생성

완전정보최대가능성법은 Maximum Likelihood(ML)를 활용한 결측치 처리 방법으로 결측치를 처리하기 위해 가장 널리 쓰이는 방법이다. Likelihood는 일반적으로 우도로 번역하는 경우가 대부분이나, 본 논문에선 가능성으로 번역하고자 한다. 단일대체법, 다중대체법처럼 특정값을 결측치에 투입하는 대신, 완전정보최대가능성법은 표본 수준에서 관찰된 모든 변수를 활용해 해당 표본을 산출할 가능성이 가장 높은 모수를 추정한다(Enders, 2010). 완전정보최대가능성법은 표준 오차를 포함한 일반적인 통계 추정치, 모델에 대한 합치도 지수를 모두 구할 수 있다. 또한, 결측치에 특정한 값을 부여하거나 정보가 누락되지 않는다는 장점이 있다. 구조방정식 모형에서 결측치를 추정할 때에도 완전정보최대가능성법은 다른 처리법과 비교했을 때 가장 안정적인 수행을 보인다(Allison, 2003; Duncan, Duncan & Li, 1998; Enders & Bandalos 2001; Newman, 2003). 덧붙여서, 다중대체법과 비교했을 때에도 완전정보최대가능성법은 유사하거나 더 나은 결과를 보여주는 것으로 알려져 있다. 다만 표본의 크기가 작을 경우 추정치가 편향되는 경향이 있으며 (Yuan, Yang & Bentler, 2012), EM 알고리즘에 기반하고 있기 때문에 시작값에 따라 표준편차가 커지거나 잘못된 추정치로 수렴할 수 있다(Schafer & Graham, 2002). 또한 다중대체법과 완전정보최대가능성법은 모두 결측치가 MAR일 때 가장 효과적인 것이며 MNAR일 경우, 편향될 가능성이 있다.

다중대체법과 완전정보최대가능성법은 전통적인 방법에 비해 자료의 누락이 적고 편향이 적은 추정치를 얻을 수 있기 때문에 결측치를 처리하기 위해 연구자가 가장 많이 선택하는 방법이다. 이와 더불어 연구자는 조금 더 정확한 추정을 위해 보조변수를 활용할 수 있다. 보조변수는 데이터엔 포함되어 있으나 연구자의 이론적 모델엔 포함되지 않은 변수로 결측치를 다루는 과정에서 활용할 수 있는 변수다(Enders, 2008). 보조변수는 다중대체법

을 활용하는 과정에서 투입되기 시작했으며(Meng, 1994; Rubin, 1996), 완전 정보최대가능성법을 활용한 결측치 추정에서도 이용될 수 있음이 선행연구를 통해 밝혀졌다(Graham, 2003)

많은 선행 연구들이 보조변수를 사용할 경우, 결측치 추정에 도움이 될 수 있으며, 결과적으로 결측치로 인해 발생하는 편향을 줄일 수 있음을 확인했다. 이러한 이유로 결측치 추정을 위해 보조변수를 투입할 것을 권장하고 있다(Collins, Schafer & Kam 2001; Graham, 2003; Little et al, 2012; Savalei & Bentler, 2009).

결측치 추정에서 보조변수의 투입은 편향과 표준오차를 줄일 수 있지만, 표본의 수가 제한된 상황에서 너무 많은 보조변수는 오히려 편향을 증가시키거나 수렴이 되지 않아 추정 결과를 산출해내지 못하는 문제점을 보이기도 한다(Howard, Rherntulla & Little, 2015; Thoemmes & Rose, 2014). 연구자는 어떤 변수가 결측치 추정에 도움을 줄 수 있는지 확인할 수 없기 때문에 최대한 많은 보조변수를 투입하는 것이 보다 안전하다. 그러나 대다수의 심리학 연구는 표본의 수를 늘리기 어려워 보조변수의 투입이 제한되는 경우가 빈번하다. 이러한 맥락에서 주성분분석(Principle Component Analysis: PCA)이 보조변수의 수를 효과적으로 줄일 수 있는 방법으로 제시되고 있다. 주성분분석은 다수의 변수들의 공분산을 요약해주는 분석 방법이며, 결측치 처리의 맥락에서 주성분이라는 새로운 변수를 보조 변수로 대체하여 분석에 사용할 수 있다. 주성분분석의 목표는 원 변수들 간의 분산을 최대한 유지하면서 그 수는 축소하는 것으로 차원을 축소한다고 표현한다(Jolliffe, 2002). 주성분의 수는 각 주성분의 고유분산 크기와 주성분들이 원 변수의 분산을 얼마나 설명하는지를 기준으로 결정하는데, 일반적으로 고유분산의 크기가 1 이상이고 원 변수의 분산을 80%이상 설명할 경우, 분석 과정에서 본래 변수들과 동일한 수행 결과를 도출해낼 수 있다고 가정

한다(Jolliffe, 2002).

결측치 처리의 맥락에서 주성분분석은 보조변수들의 분산은 최대한 유지하면서 투입될 변수의 수를 줄여줄 수 있다는 점에서 표본이 작은 경우 발생하는 수렴과 편향 문제를 완화할 수 있다는 장점이 있다. 선행연구 결과, 주성분이 투입된 보조변수의 전체 분산 중 40%를 설명하면 이를 새로운 보조변수로 활용해 결측치를 추정해도 보조변수를 그대로 사용하는 것과 동일한 수준의 수행을 보일 뿐만 아니라 표본의 크기가 작은 경우에 발생하는 수렴 문제도 해결할 수 있는 것으로 나타났다(Howard, Rherntulla & Little, 2015).

주성분분석은 변수 간 상관이 높을 경우 적은 수의 주성분으로 충분한 크기의 분산을 얻을 수 있으나, 상관이 낮을 경우 요약될 수 있는 정보가 적기 때문에 같은 크기의 분산을 얻기 위해 더 많은 주성분이 필요해진다. 주성분 분석의 목표가 최소한의 주성분으로 원 자료의 정보를 최대한 끌어내는 것임을 감안할 때, 다수의 주성분을 사용한다는 것은 효율성의 문제를 제기할 수 있다. 다시 말해, 주성분분석은 변수 간의 상관이 높을수록 효과적인 방법이 될 수 있다.

한편 보조변수의 경우, 결측치를 설명할 수 있는 다수의 변수를 투입하는 과정에서 연구자는 각각의 보조변수가 서로 낮은 상관을 가지길 기대한다. 결측치를 추정하는 과정을 결측치를 준거 변수, 보조변수를 예측 변수로 두는 회귀분석이라고 생각한다면, 보조변수들은 결측치와 일정 수준의 상관을 갖고 있어 결측치를 설명하고 예측해 줄 수 있음과 동시에 서로 낮은 상관을 갖고 있는 것이 바람직하다. 상관이 높은 두 보조변수를 함께 사용한다는 것은 모형의 자유도는 낮추면서 결측치에 대해 새롭게 얻는 정보의 양은 적음을 의미하며 다중공선성의 문제가 발생할 수 있기 때문이다. 많은 선행 연구들이 다중공선성이 발생할 경우 추정에 부정적인 영향을 미친다는 결과

를 보고해왔다.(Farrar & Glauber, 1967; Mansfield & Helms, 1982; Marsh, Dowson & Pietsch, 2004).

일반적으로 중다회귀분석에서 다중공선성 문제가 발생할 경우 주성분분석을 활용할 수 있다고 알려져 있으나, 여전히 요약된 주성분들이 다중공선성을 제대로 해결하지 못하는 경우가 존재한다(Hadi & Ling, 1998). 반대로 보조변수 간 상관이 낮은 경우, 주성분분석을 사용하더라도 투입될 변수의 수를 효과적으로 줄이긴 어렵다. 또한 이 과정을 통해 구해진 주성분이 전체 분산 중 40%를 설명한다고 할 때, 해당 주성분들이 보조변수들이 가지고 있는 고유한 정보를 누락 없이 요약하고 있는지 확인이 필요하다. 그러나 보조변수 투입과정에서 주성분분석을 이용한 시뮬레이션 연구들은 보조변수와 결측치 간 상관을 달리하고 보조변수들 간 상관은 하나의 수치로 통일한 것으로 보고되고 있다(안혜영, 2016; Howard, Rherntulla & Little, 2015). 심리학 연구 장면에서 사용되는 다양한 변수들이 서로 유사한 크기의 상관을 가지고 있기는 어렵다는 것을 감안할 때, 보조변수 간 상관의 크기와 동질성에 대한 논의가 필요하다. 즉 보조변수 간의 상관이 높을 때와 낮을 때, 그리고 다수의 보조변수들이 서로 다른 크기의 상관을 가질 때 주성분분석의 활용 방안에 대한 추가적인 연구가 필요하다.

이러한 맥락에서 본 연구는 결측치 처리 과정에서 보조변수 간 상관의 크기가 이질적일 때 주성분분석이 효과적으로 보조변수를 요약하는지 살펴보고, 기존의 연구 결과를 근거로 전체 분산의 40%를 설명하는 주성분을 사용했을 때의 결측치 처리 결과를 보조변수를 그대로 사용했을 때와 비교하고자 한다. 이를 검증하기 위해 통계 프로그램 Mplus와 R을 활용하여 몬테카를로 시뮬레이션을 진행할 것이다. 몬테카를로 시뮬레이션은 연구자가 설정한 조건에 따라 무작위로 추출된 표본을 모형에 적용 후, 반복 분석한 결과를 보고해준다(이현숙, 김수진, 전수현, 2010). 몬테카를로 시뮬레이션은

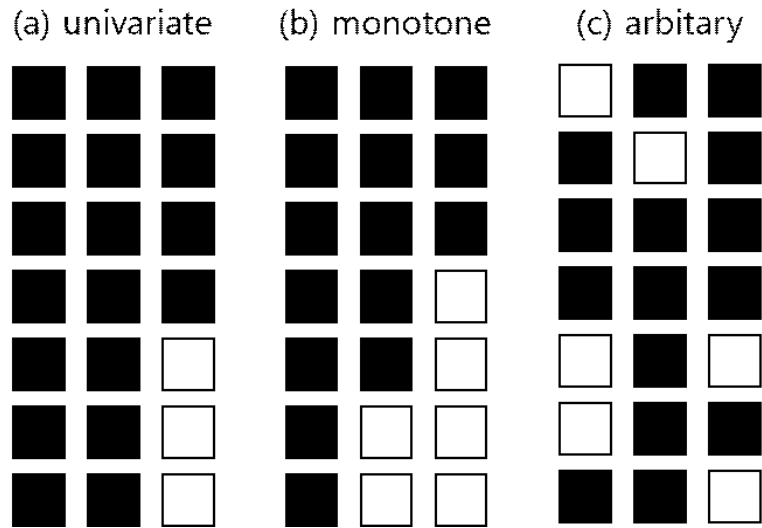
확인하고자 하는 분석 방법 외의 조건들을 통제할 수 있다는 이점이 있다. 본 연구에서는 몬테카를로 시뮬레이션 기능을 통해 생성된 표본을 연구모형에 적용하여 보조변수 간 상관에 따라 주성분 분석의 수행이 어떻게 변하는지 확인하고, 보조변수를 사용하지 않는 경우, 보조변수를 그대로 투입하는 경우의 결측치 처리 결과와 비교할 것이다.

각 조건별로 보조변수 간 상관의 크기와 동질성이 다르게 설정될 것이며, 다양한 상황에서의 결측치 처리 수행 능력을 비교하기 위해 결과변수와 보조변수 간의 상관, 결측치의 비율, 표본의 크기를 다르게 설정할 것이다. 결측치의 형태는 Linear MAR로 통일할 것이며 X가 Y를 설명하는 단순회귀모형을 사용할 것이다. 결측치 처리는 ML법을 활용해 이뤄질 것이며, Saturated model(Graham, 2003)을 활용하여 Y의 평균, 분산, X와 Y간 상관계수를 구해 모수가 편향된 정도, 평균편차제곱근, 신뢰구간을 확인해 수행능력을 비교할 것이다. 또한 표본의 크기가 작은 상황에서 다수의 보조변수를 투입했을 때 발생하는 수렴 문제를 주성분분석이 해결할 수 있는지 확인하고자 한다.

## II. 이론적 배경

### 1. 결측치 메커니즘

대다수의 자료는 표본을 의미하는  $n$ 개의 행과 변수를 의미하는  $p$ 개의 열로 이루어진 직사각형 모양의 행렬  $Y$ 로 표현이 가능하며  $Y_{ij}$ 는 행렬  $Y$ 에서  $i$ 번째 표본의  $j$ 번째 변수의 관찰값을 의미한다. 결측치의 패턴은 행렬  $Y$ 를 바탕으로 값이 관찰됐을 경우엔 검정색, 관찰되지 않았을 경우엔 흰색으로 표시한 <그림 2>를 통해 확인할 수 있다(Schafer & Graham, 2002). (a)는 전체 자료 중 하나의 변수에서만 결측치가 발생한 단변량 패턴이다. (b)는 단조 패턴으로  $j$ 번째 열에서 결측치가 발생하고 이후 마지막 열까지 모두 결측된 형태다. 단조 패턴의 대표적인 예로 중단 연구에서 발생하는 탈락(drop out)이 있다. 마지막으로 (c)는 임의 패턴으로 모든 행과 열에서 무작위로 결측치가 발생한 패턴이다. <그림 2>를 행렬의 형태로 나타낸  $M$ 은  $Y$ 와 동일한 크기로  $M_{ij}$ 는  $Y_{ij}$ 가 값을 가질 경우엔 0,  $Y_{ij}$ 가 결측치일 경우 1의 값을 가지며 결측치 메커니즘(원리)을 이해하고 구분하는 지표로 사용된다. 결측치가 발생하는 메커니즘을 고려하지 않을 경우, 처리 과정을 거치더라도 편향된 결과로 이어질 수 있기 때문에(Schafer & Graham, 2002), 결측치를 올바르게 다루기 위해서는 그 발생 메커니즘을 이해할 필요가 있다.



<그림 2 > 직사각형 형태로 표현한 결측치 패턴. 각 직사각형의 행은 표본을, 열은 변수를 의미한다.

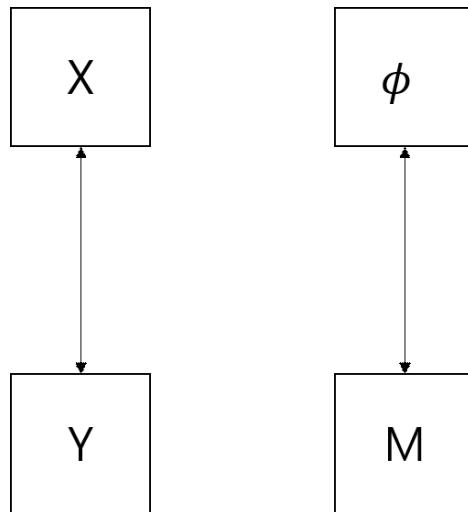
결측치는 발생하는 메커니즘과 자료 내 다른 변수들 간의 관계에 따라 완전무작위결측(Missing completely at random, MCAR), 무작위결측(Missing at random, MAR), 작위결측(Missing not at random, MNAR)으로 구분한다(Rubin, 1976). 본 연구에서는 각각 MCAR, MAR, MNAR로 표기한다.

### 1) Missing completely at random (MCAR)

이해의 편의를 위해, 성별과 수학점수, 두 개의 변수로만 이루어져 있는 학생 100명에 대한 데이터를 예시로 들고자 한다. 표본 중 30%가 수학점수에 누락이 있으며, 이를 자료 내에서 관찰된 변수로 설명할 수 없을 때, 해당 결측치를 MCAR이라고 한다. MCAR 가정 하에서 결측치는 관찰된 자료와 관찰되지 않은 자료를 모두 포함한 전체 데이터와 무관하게 발생한다. 이를 수식으로 나타낼 경우 식 (1)과 같다.

$$f(M|Y, \phi) = f(M|\phi) \text{ for all } Y, \phi \quad \text{식 (1)}$$

$f(M|Y, \phi)$ 는 결측치의 패턴을 나타내는 행렬  $M$ 에 대한 조건부분포며,  $\phi$ 는 자료 내에서 확인할 수 없는 알려지지 않은 행렬 벡터다. 식 (1)에서  $M$ 이 결측치를 의미하는 1을 가질 확률은 전체 자료를 의미하는  $Y$ 와 무관하게 결정된다. MCAR을 그림으로 나타낼 경우, <그림 3>과 같다.  $X$ 는 관찰된 변수,  $Y$ 는 결측치가 존재하는 변수이며  $M$ 은  $Y$ 의 결측치 패턴을 나타내는 행렬,  $\phi$ 는 자료 내에 존재하지 않는 변수, 행렬 벡터다. 결측치 패턴 행렬  $M$ 은  $X$ 나  $Y$ 와 상관이 없으며 자료 내에선 나타나지 않는 변수  $\phi$ 의 영향만 받는다.



<그림 3> MCAR 결측치 (Enders, 2010)

결측치가 MCAR일 경우, 분석 과정에서 결측치를 무시해도 편향되지 않은 결과를 얻을 수 있다(Little, 1988). 그러나 MCAR은 연구 설계 과정에서

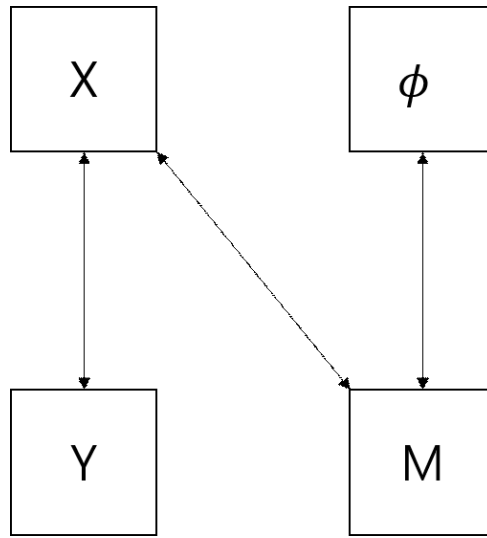
자연스럽게 발생한 결측치를 설명하기에 적절한 형태로, 다수의 변수가 자료에 포함된 실제 연구 장면에서 결측치가 MCAR일 가능성은 낮다.

## 2) Missing at random (MAR)

MAR은 MCAR보다 유연한 형태로 결측치의 발생이 자료 내의 다른 변수로 설명될 수 있음을 인정한다. 앞서 예시로 든 성별과 수학점수로 이루어진 자료에서 표본이 남성일 경우 수학 점수가 결측되는 경향이 있다면 수학 점수의 결측치는 MAR로, 발생한 결측치를 성별로 설명이 가능하다고 판단할 수 있다. MAR을 수식으로 나타낼 경우 식 (2)와 같다.

$$f(M|Y,\phi) = f(M|Y_{obs},\phi) \text{ for all } Y_{mis},\phi \quad \text{식 (2)}$$

Y를 관찰된 값인  $Y_{obs}$ 와 결측치인  $Y_{mis}$ 로 구분했을 때, 결측치 행렬 M은 다른 변수에서 관찰된 값인  $Y_{obs}$ 로 설명가능하다. MAR 결측치를 그림으로 나타낼 경우 <그림 4>와 같다. 각 표기는 <그림 3>과 동일하다. MAR 결측치의 경우 결측치 패턴 M이 관찰된 변수 X와 상관을 가지고 있다. 즉 Y에서 발생하는 결측치의 패턴을 X를 통해 설명이 가능함을 의미한다.



<그림 4> MAR 결측치 (Enders, 2010)

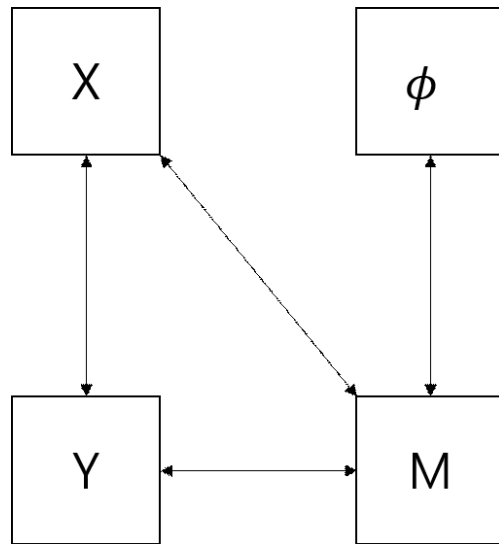
MAR은 다른 변수와의 상관을 인정하기 때문에 MCAR보다 현실적인 조건이라고 볼 수 있다. 또한 다중대체법과 완전정보최대가능성법 등 대부분의 결측치 처리 방법은 결측치의 형태가 MAR이라고 가정한다. 그러나 결측치의 원인에 대한 정보가 없을 경우, 연구자는 자료에 존재하는 결측치가 MAR라고 판단할 수 없다(Bollen & Curran, 2006).

### 3) Missing not at random (MNAR)

MNAR은 결측치 메커니즘 중 가장 유연한 조건으로 결측치가 발생한 원인이 해당 변수 자체에도 있는 경우다. 예시로 든 자료에서 수학점수가 낮은 학생들이 수학점수에 대한 응답을 거부했을 경우, 결측치의 원인이 누락된 수학점수 자체에 있기 때문에 해당 결측치는 MNAR이 된다. MNAR을 수식으로 나타낼 경우 식 (3)과 같다.

$$f(M|Y, \phi) \neq f(M|Y_{obs}, \phi) \text{ for all } Y_{mis}, \phi \quad \text{식 (3)}$$

전체 데이터와 결측치 행렬의 조건부분포는 관찰된 데이터와 결측치 행렬의 조건부분포와 다르다. 다시 말해, 관찰되지 않은 데이터  $Y_{mis}$ 에 결측치에 대한 추가적인 정보가 있음을 의미한다. MNAR 결측치를 그림으로 나타낼 경우 <그림 5>와 같으며 표기는 위의 MCAR, MAR과 동일하다. 결측치의 패턴을 나타내는 행렬  $M$ 은 관찰된 변수인  $X$  뿐만 아니라 결측치를 가지고 있는  $Y$  변수와도 상관을 가진다. 즉 MNAR 결측치를 처리하기 위해서는 자료 내에 나타난  $X$  변수뿐만 아니라 자료에 나타나지 않은  $Y$ 에 대한 정보도 필요함을 의미한다.



<그림 5> MNAR 결측치 (Enders, 2010)

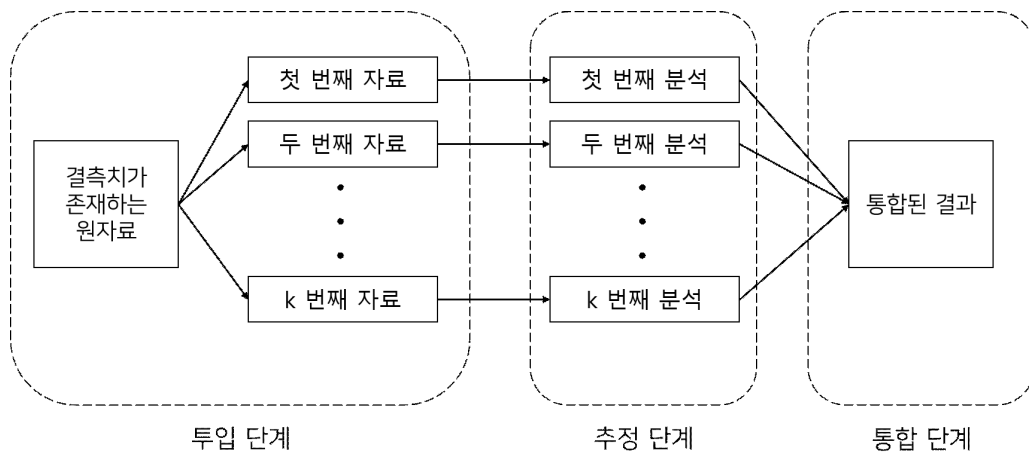
결측치가 MNAR인 경우 가지고 있는 자료에 연구자의 관심사에 대한 정보가 누락됐음을 의미하기 때문에, 해당 자료 그대로 분석에 사용될 경우

결과 또한 편향되어 나타날 가능성이 매우 높다.

## 2. 결측치 처리 방법

### 1) 다중대체법(Multiple Imputation: MI)

다중대체법은 자료가 다변량 정규분포를 따르고 결측치가 MAR라는 가정을 두고 있으며, 투입 단계, 추정 단계, 통합 단계를 거친다.



<그림 6> 다중대체법 (강민아, 김경아, 2006)

결측치를 대체할 값들을 구하는 투입 단계에선 자료가 정규분포를 따른다는 전제하에 I(imputation step)단계와 P(posterior step)단계를 거쳐 서로 독립된 대체값을 생성한다. I단계에선 특정 함수를 활용해 하나의 대체값을 생성한다. 대체값을 생성하기 위해 다양한 함수를 사용할 수 있으나 본 연구에서는 식 (4)와 같은 회귀식의 형태를 사용하고 있다.

$$X = b_0 + b_1 Y + S_{xy} E \quad \text{식 (4)}$$

결측치  $X$ 에 대한 대체값은 식 (4)를 따라 생성된다.  $b_1$ 은 결측치  $X$ 에 대한  $Y$ 의 회귀계수,  $b_0$ 는 잔차항이다. 다중대체법에서 대체값은 결측치가 존재하는 변수에 대한 회귀식인  $b_0 + b_1 Y$ 에  $E$ 라는 정규분포를 따르는 오차항을 구해 더하는데, 이때  $E$ 는  $X = b_0 + b_1 Y$  회귀식에서 구한 잔차인  $b_0$ 의 표준편차  $S_{xy}$ 를 반영한다. 회귀대체법을 통해 생성한 대체값을  $X$ 축과  $Y$ 축으로 이루어진 2차 평면 위에 표현했을 때, 그 위치가 회귀선과 정확히 일치하는 것과 달리 다중대체법을 통해 생성된 대체값은 오차를 반영한  $S_{xy}E$ 를 포함하고 있어 전체 자료의 분산이 감소하는 현상을 완화할 수 있다. 다중대체법은 서로 독립된 다수의 대체값을 요구하는데, 이는 곧 새로운 대체값을 만들 때마다 I단계의 식 (4)에서 활용하는 회귀계수  $b_1$ 이 달라져야 함을 의미한다. 따라서 새로운 회귀식을 생성하기 위해 P단계를 거친다. P단계는 베이저안 추정(Bayesian Estimation)을 기반으로 한다. P단계에서는 앞서 I단계에서 추정된 대체값을 채워 넣은 자료를 활용해 새로운 평균벡터와 공분산 행렬을 구한다. 이후 각각의 평균벡터와 공분산 행렬의 추정치에 임의의 잔차를 투입해 새로운 모수 값을 생성하는데, 모수 값에 임의로 오차를 줌으로써 새로운 표본분포 혹은 사후분포를 추정하는 것과 동일한 기능을 한다. P단계에서 생성된 새로운 모수 값은 다시 I단계로 돌아가 새로운 대체값을 생성하며 해당 대체값은 다시 P단계에 사용되는데, 투입 단계에서는 이러한 과정을 필요한 대체값의 수만큼 반복한다.

$$X_t^* \sim p(X_{mis} | X_{obs}, \theta_{t-1}^*) \quad \text{식 (5.1)}$$

$$\theta_t^* \sim p(\theta | X_{obs}, X_t^*) \quad \text{식 (5.2)}$$

투입 단계를 식으로 요약하자면 식 (5.1), 식 (5.2)와 같다. 식 (5.1)은 I단

계, 식 (5.2)는 P단계를 의미한다. 식 (5.1)의  $X_t^*$ 는 t번째 I단계에서 생성된 투입값이며,  $X_{mis}$ 는 결측치,  $X_{obs}$ 는 관찰된 값을 의미한다.  $\theta_{t-1}^*$ 는 이전 t-1 번째 P단계에서 생성된 평균 벡터와 공분산 행렬을 의미한다. 즉 I단계에서 생성되는 임의의 값은 관찰된 자료와 이전 단계에서 추정된 모수의 분포를 따른다. 식 (5.2)의  $\theta_t^*$ 는 t번째 P단계에서 생성되는 모수값이며, 평균 벡터와 공분산 행렬을 의미한다.  $X_{obs}$ 는 관찰된 자료,  $X_t^*$ 는 이전 I 단계에서 생성된 대체값이다. 즉, t번째 P단계에서 생성되는 모수값은 관찰된 자료와 이전 단계에서 생성된 대체값의 분포로부터 임의로 추출된 값이다.

추정 단계에서는 각각의 대체값이 투입된 데이터들을 일반적인 분석 방법과 동일한 절차를 거쳐 연구자가 구하고자 하는 추정치와 표준오차를 구할 수 있다. 마지막으로 통합 단계에서는 앞에서 구한 다수의 추정치와 표준오차를 하나로 통합하는데, 추정치의 경우 평균을 사용한다. 표준오차의 경우 추정된 표준오차들의 평균과 추정치들의 분산을 사용한다(Enders, 2010).  $M$  개의 대체값을 이용했을 때 표준오차를 통합하는 식은 식 (6)과 같다(Rubin, 1987).

$$SE = \sqrt{\frac{1}{M} \sum_{k=1}^M s_k^2 + \left(1 + \frac{1}{M}\right) \left(\frac{1}{M-1}\right) \sum_{k=1}^M (x_k - \bar{x})^2} \quad \text{식 (6)}$$

$x_k$ 는  $M$ 개의 대체값 중  $k$ 번째 대체값을 통해 구한 추정치이며  $s_k$ 는  $x_k$ 의 표준오차다. 선행 연구에 따르면, 대체값의 개수가 증가할수록 다중대체법의 수행력은 증가하며 20개 이상 사용할 경우 완전정보최대가능성법과 유사한 수행을 보이는 것으로 나타났다(Graham, Olchowski, & Gilreath, 2007).

2) 완전정보최대가능성법(Full Information Maximum Likelihood: FIML)

최대가능성추정법(Maximum Likelihood Estimation: MLE)은 최대가능성 함수를 활용해 관찰된 자료와 연구자가 구하고자 하는 모수간의 거리를 표준화한다. 그리고 해당 거리를 최소화하는 모수를 찾는 것이 목표다. 최대가능성함수는 식 (7.1)과 같다.

$$\log L = \sum_{i=1}^N \log \left[ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5 \left( \frac{y_i - \mu}{\sigma} \right)^2} \right] \quad \text{식 (7.1)}$$

대괄호 안은 정규분포 형태로 나타나는 확률밀도함수로, 특정한 평균과 표준편차를 가진 정규분포 내에서 단일 표본이 가지고 있는 정보와 동일한 값을 가질 확률을 나타낸다.  $\left( \frac{y_i - \mu}{\sigma} \right)^2$ 은 표본과 구하고자 하는 모수와의 거리를 나타내는데, 모수와 표본의 차이가 작을수록 표준화된 거리는 작아지고 해당 확률함수에 로그를 취한  $\log L$ 의 값은 커진다. 최대가능성추정법은 각각의 표본에 대한  $\log L$ 값을 모두 더해 그 값이 최대가 되는 지점의 모수를 구할 때까지 반복하는 과정을 거친다. 식 (7.1)은 단변량 추정인 경우를 표현한 식이며, 다변량 추정의 경우 모수와 표본 사이의 거리를 나타내는 스칼라 값이 (7.2)과 같은 행렬의 형태로 대체된다.

$$(Y_i - \mu)^T \Sigma^{-1} (Y_i - \mu) \quad \text{식 (7.2)}$$

$\mu$ 는 전집의 평균 벡터며  $\Sigma$ 는 전집의 공분산 행렬이다.  $Y_i$ 는 단일 표본의 벡터다.  $T$ 는 전치행렬을 의미하며  $^{-1}$ 은 역행렬 표기다. 완전정보최대가능성

법은 동일한 최대가능성함수를 사용하나, 관찰된 변수들만을 활용해 모수를 추정한다. 다변량정규분포를 따르는 자료에서 개인 수준의 최대가능성 함수는 식 (7.3.1)이며 동등한 함수인 식 (7.3.2)로도 표현이 가능하다.

$$\log L_i = \log \left\{ \frac{1}{(2\pi)^{k/2} |\Sigma|^{k/2}} e^{-.5(Y_i - \mu)' \Sigma^{-1} (Y_i - \mu)} \right\} \quad \text{식 (7.3.1)}$$

$$\log L_i = K_i - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} [x_i - \mu_i]' \Sigma_i^{-1} [x_i - \mu_i] \quad \text{식 (7.3.2)}$$

$K_i$ 는  $i$ 번째 표본에서 관찰된 변수의 수이며,  $x_i$ 는 관찰된 개별 점수,  $\mu$ 와  $\Sigma$ 는 전체 자료에서의 평균 벡터와 공분산 행렬이다. 식 (7.3.1)을 통해 모든 표본에 대한 가능성함수를 구한 후 이를 모두 합해 전체 표본에 대한 가능성함수를 구할 수 있다.

$$\log L(\mu, \Sigma) = \sum_{i=1}^N \log L_i \quad \text{식 (7.4)}$$

전체 표본에 대한 가능성함수를 구한 후에는 ML과 동일한 과정을 통해 추정치를 구할 수 있다. 완전정보최대가능성법은 MAR 결측치를 가정하고 있으나 MCAR일 때에도 다른 결측치 처리 방법과 비교했을 때 가장 높은 검증력을 보이는 것으로 알려져 있다. 결측치가 MNAR인 경우 수행력이 떨어지긴 하나, 이는 다른 결측치 처리 방법도 마찬가지이며 MNAR 결측치를 바람직하게 추정하는 방법은 아직 연구가 미흡하기 때문에 완전정보최대가능성법만의 문제는 아니다(Enders, 2010).

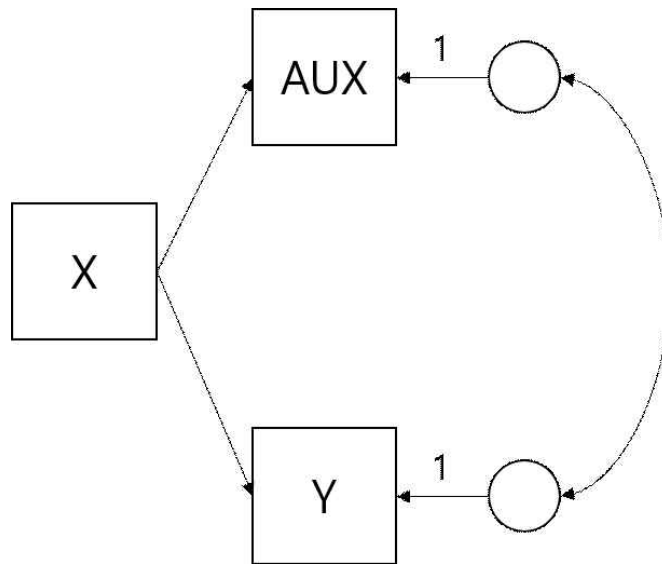
### 3. 보조변수(Auxiliary Variable)

보조변수는 데이터에는 포함되어 있으나 연구자의 이론적 모델에는 포함되지 않은 변수로, 결측치를 다루기 위해 ML 혹은 MI를 사용할 때 투입할 수 있는 변수다(Enders, 2008). 변수는 두 가지 기준에 근거하여 보조변수로 투입될 수 있는데, 하나는 해당 변수가 결측과 상관이 있거나 결측의 원인 그 자체가 될 수 있다고 판단될 때이다(Collins, Schafer & Kam, 2001; Schafer, 1997). 앞서 사용한 성별과 수학점수 예시에서 남성 표본이 수학점수를 보고하지 않는 경향이 더 많은 것으로 나타났을 때, 성별을 수학점수의 결측치에 대한 보조변수로 활용하여 성별에 따라 응답률이 다름으로 인해 나타나는 편향을 줄일 수 있다. 두 번째로 결측치 자체와 상관이 있다고 판단되는 변수를 보조변수로 사용할 수 있다. 예시로 사용 중인 성별과 수학점수로 이루어진 자료에 과학점수라는 변수가 하나 더 존재하고 수학점수와 과학점수 사이에 유의미한 정적상관이 발견됐다면, 과학점수는 수학점수의 결측치를 메우는 과정에서 정보를 제공할 수 있다 판단되어 보조변수로 활용이 가능하다. 선행 연구를 통해 보조변수는 결측치가 MAR일 때 추정에 긍정적 효과를 미치는 것으로 나타났으며 MNAR일 때 발생하는 편향도 완벽한 제거는 어렵지만 완화하는 것으로 알려졌다(Collins, Schafer & Kam, 2001). 즉 결측치 처리에 있어서 보조변수의 포함은 대부분의 상황에서 추정에 도움을 준다고 말할 수 있다. 보조변수는 대부분의 상황에서 추정에 도움을 주고 검증력을 높여주지만 결측치가 존재하는 변수와의 상관이 .4 이상인 것을 권장한다(Enders, 2010). 다중대체법을 사용하는 맥락에서 보조변수는 가능한 모든 변수를 투입하는 것을 권장하며 그 방법 또한 비교적 간단하다(Rubin, 1996; Dempster, Laird & Rubin, 1977). 그러나 완전정보최대가능성법을 사용하는 경우 보조변수를 모형에 투입하는 과정이 다중대체법에 비해 복잡하고 모델 식별의 문제가 발생하기 때문에 그 수에 제한

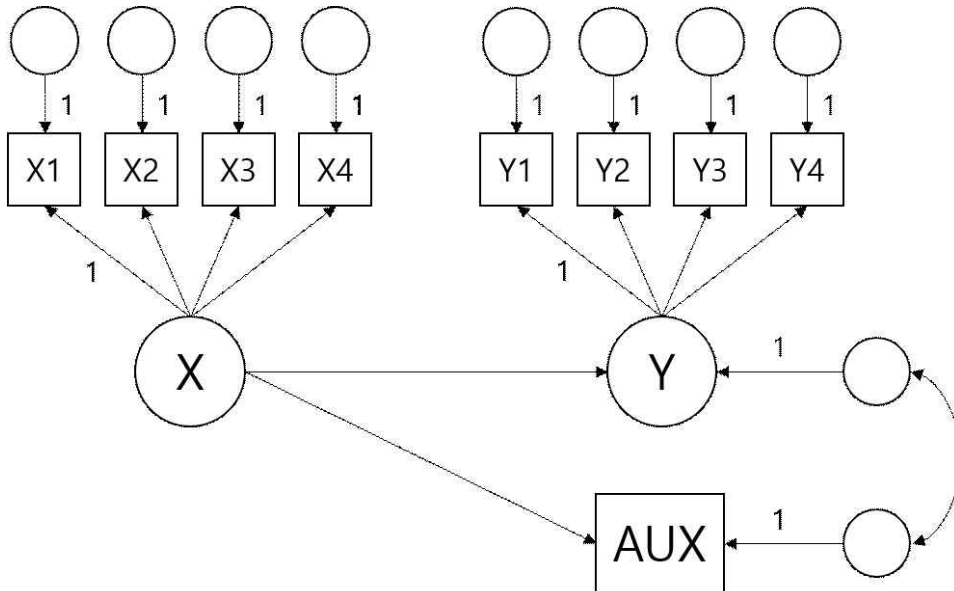
이 생길 수 있다(Enders, 2010).

완전정보최대가능성법과 구조방정식 모형을 사용하는 장면에서 보조변수를 모형에 어떤 형태로 투입할 것인가에 대하여 새로운 종속변수의 형태로 넣는 ‘Extra DV’모형과 모형의 일부분으로 통합된 ‘Saturated correlates’모형이 제시됐다(Graham, 2003). 두 모델 모두 구조방정식 모형 자체엔 영향을 주지 않으면서 보조변수를 포함할 수 있는 형태다.

Extra DV모형은 보조변수를 또 다른 종속변수로서 투입하는 모델이다. 보조변수는 X로 설명이 가능하며 보조변수의 잔차는 결측치가 존재하는 변수 Y의 잔차와 상관이 존재한다. 이때 투입되는 보조변수의 조건은 다음과 같다. (a) 모든 보조변수는 모형 내 모든 예측변수로 설명이 가능하다. (b) 모든 보조변수의 잔차는 잠재된 외생변수의 잔차와 상관을 가지고 있다. (c) 모든 보조변수의 잔차는 서로 상관을 갖는다. Extra DV 모델을 그림으로 나타낼 경우 <그림 7>, <그림 8>과 같다.

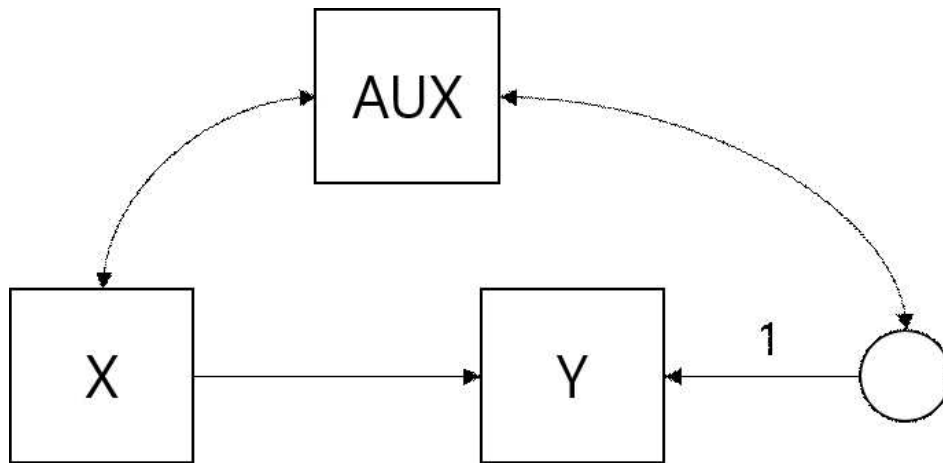


<그림 7> Extra DV 모델 (관찰변수) (Graham, 2003)

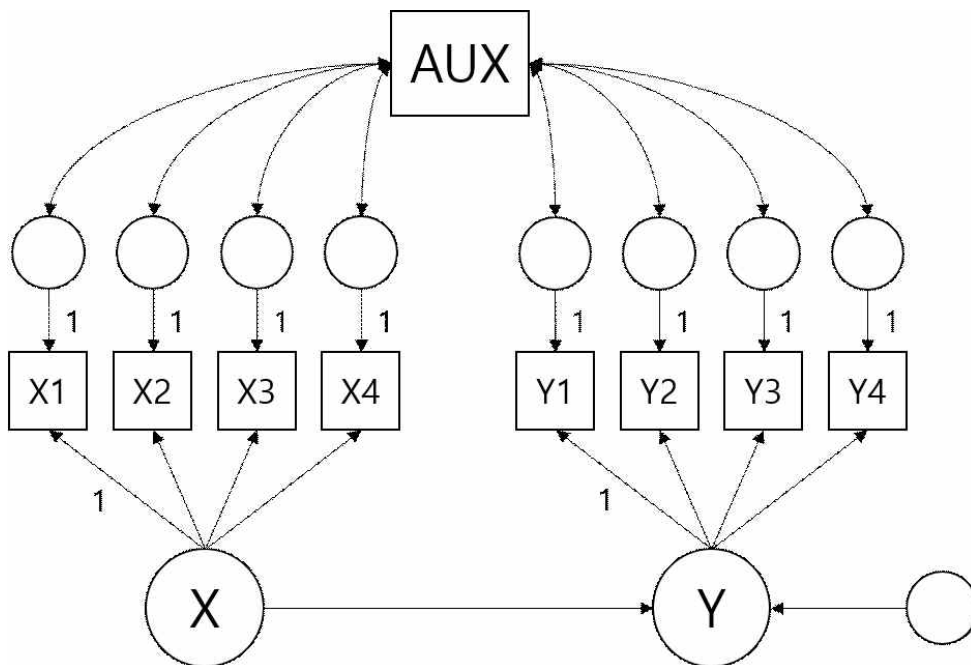


<그림 8> Extra DV 모델 (잠재변수) (Graham, 2003)

Saturated correlates 모델은 보조변수를 구조방정식 모형의 일부로 통합하는 방법이다. Saturated correlates 모델에 투입되는 보조변수들은 다음과 같은 조건을 따른다. (a) 모든 관찰된 외생변수는 보조변수와 상관을 가진다. (b) 모든 내생 변수는 보조변수와 상관을 가진다. (c) 모든 보조변수들은 서로 상관을 가진다. Saturated correlates 모델을 그림으로 나타낼 경우 <그림 9>, <그림 10>과 같다.



<그림 9> Saturated correlates 모델 (관찰 변수) (Graham, 2003)



<그림 10> Saturated Correlates 모델 (잠재 변수) (Graham, 2003)

두 모델의 수행을 비교해 본 결과, 두 모델 모두 결측치 추정엔 유사한 수행을 보였으나 전체 모형의 합치도 지수의 경우 Saturated correlates 모

델이 더 좋은 합치도를 가지는 것으로 나타났다. Extra DV 모델은 자유도,  $X^2$  검증과 RMSEA에서 나쁜 합치도를 보였다(Graham, 2003). 따라서 완전 정보최대가능성법과 구조방정식을 활용한 연구에선 Saturated correlates 모델의 형태로 보조변수를 투입한다(Collins, Schafer, & Kam, 2003; Savalei & Bentler, 2009).

### 1) Inclusive Strategy

보조변수를 선택할 때 연구자는 포괄적 혹은 제한적 입장을 취할 수 있다. 포괄적인(Inclusive) 입장을 취할 경우, 연구자는 분석 과정에 가능한 많은 보조변수를 투입한다는 것을 의미한다. 다수의 변수를 보조변수로 활용할 경우, 추정이 좀 더 정확해지고 표준오차가 감소하는 것은 물론이고 결측치를 설명할 수 있는 중요한 변수를 빠트릴 가능성이 적어진다(Collins, Schafer & Kam, 2001). 실제 연구 장면에서 연구자들은 어떤 변수가 결측치를 가장 잘 예측할 것인지 확인할 수 없기 때문에 이는 매우 중요한 장점이다. 그러나 한정된 수의 표본을 가지고 있는 모형에 보조변수를 투입한다는 것은 연구자의 직접적인 관심사와 관련이 없는 변수를 추가한다는 것을 의미하기 때문에 일종의 ‘잡음’이 되어 수행을 저하시킬 위험이 있다(Savalei & Bentler, 2009).

### 2) Restrictive Strategy

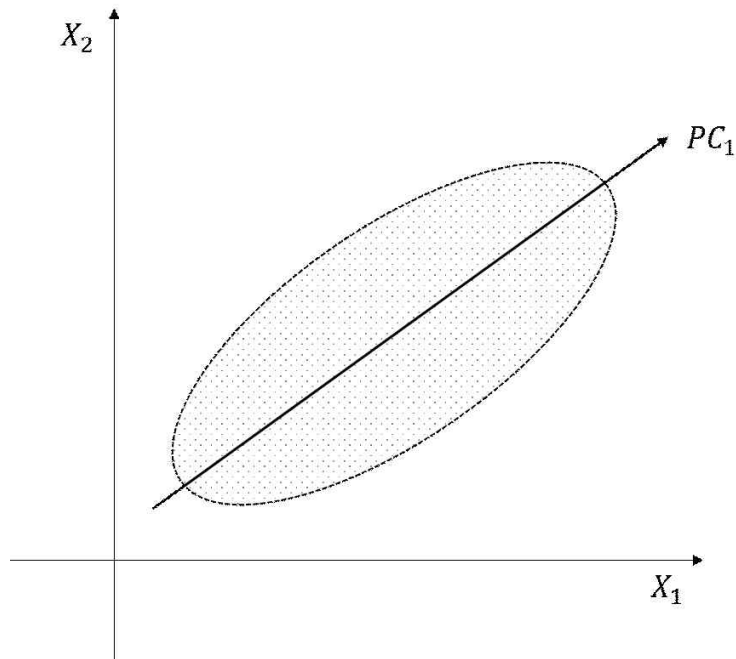
제한적 입장은 이러한 맥락에서 제시된 것으로, 연구자는 최소한의 보조변수만을 투입해야한다는 입장이다. 포괄적인 입장을 취해 가능한 많은 보조변수를 모형에 투입하는 것이 추정 결과에 큰 도움을 주지는 않는다는 연구 결과(Mustillo, 2012)와 더불어, 변수 간의 관계에 따라서는 오히려 악영향을 미쳐 편향이 발생할 위험이 있다(Thoemmes & Rose, 2014)는 것이 선

행 연구를 통해 밝혀졌다. 또한 표본이 작은 모형에 너무 많은 보조변수를 투입할 경우 추정의 정확성이 감소하거나(Hardt, Allore & Studenski 2012) 추정치가 수렴되지 않는 문제가 발생한다(Howard, Rherntulla & Little, 2015). 때문에 제한적 입장에서선 보조변수가 결측치를 포함한 변수와 .4 이상의 상관과 이론적인 뒷받침을 가지고 있어야하며, 보조변수를 투입했을 때 모형 구조에 미치게 될 영향까지 고려할 것을 권장한다(Thoemmes & Rose, 2014). 그러나 포괄적 입장에서 언급된 것과 같이, 실제 연구 장면에서는 연구자가 변수간의 관계를 정확하게 파악하기는 어려우며, 어떤 변수가 결측치를 가장 잘 설명할 것인지 모르는 경우가 일반적이다. 즉 변수 간 상관을 확인하고 결측치가 존재하는 변수와 .40 가량의 상관을 가진 변수들을 추림으로써 대략적인 후보군은 추릴 수 있으나 연구자가 자료를 보고 보조변수로 활용할 변수들을 임의로 선택해 투입하기는 어려운 것이다. 때문에 제한적으로 보조변수를 선택하는 과정에서 중요 변수가 누락될 가능성이 존재한다는 위험이 있다. 선행 연구 결과, 처리하고자 하는 결측치와 상관이 높은 변수가 분석에서 누락될 경우, 추정 결과가 부정확하고 표준오차가 커지는 것이 밝혀졌다(Collins, Schafer & Kam, 2001; Howard, Rherntulla & Little, 2015).

#### **4. 주성분분석(Principal Component Analysis: PCA)**

##### **1) 주성분분석**

주성분분석은 다수의 변수가 서로 상관을 갖고 있는 자료에서 본래 자료가 가지고 있는 변산은 최대한 유지하며 원변수 대신 주성분이라는 새로운 변수를 사용해 자료의 차원을 축소하는 통계기법이다(Jolliffe, 2002).



<그림 11> 주성분 분석. 두 개의 변수  $X_1$ 과  $X_2$ 를 가지고 있는 자료를 주성분 1개로 요약할 수 있다.

이는 원변수의 변량의 합이 주성분들의 변량의 합과 동일하다는 전제를 갖는다.  $p$ 개의 변수로 이루어진 자료가 있고  $x$ 는 각 변수들의 벡터라고 할 때, 주성분분석은 최대한 많은 변산을 가지는  $x$ 의 선형 결합  $\alpha_1^T x$ 를 구하고자 한다. 이 때  $\alpha_1$ 은  $p$ 개의 상수,  $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1p}$ 로 이루어진 벡터이며  $T$ 는 전치행렬을 의미한다.  $\alpha_1^T x$ 는 식 (8)로 표현할 수 있다.

$$\alpha_1^T x = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1p}x_p = \sum_{j=1}^p \alpha_{1j}x_j \quad \text{식 (8)}$$

다음으로  $\alpha_1^T x$ 와 상관을 가지지 않으면서 최대로 큰 변산을 갖는  $\alpha_2^T x$ 를

구한다. 이러한 과정을 반복해  $k$ 번째로 구해진 선형 함수  $\alpha_k^T x$ 는  $k$ 번째 주성분을 의미한다.  $p$ 개의 변수를 가진 자료는 최대  $p$ 개의 주성분을 구할 수 있으며 주성분분석의 목표는  $p$ 개보다 적은 수의 주성분으로 원자료의 변산을 최대한 설명하는 것이다.

선형 함수  $\alpha_1^T$ , 즉 주성분을 구하기 위해선 모집단에 대한 공분산 행렬  $\Sigma$ , 혹은 자료에 대한 공분산 행렬  $S$ 를 이용해야 한다. 각 행렬의 성분은  $i$ 번째 변수와  $j$ 번째 변수의 분산 혹은 공분산을 의미한다.  $k$ 번째 주성분인 선형 함수  $\alpha_k^T x$ 의  $\alpha_k$ 는 공분산 행렬  $\Sigma$ 에서  $k$ 번째로 큰 고윳값(eigen value)에 대응하는 고유벡터(eigen vector)다. 공분산행렬  $\Sigma$ 는 대칭행렬이며, 대칭행렬의 서로 다른 고윳값의 고유벡터는 서로 직교하기 때문에 주성분들의 상관은 모두 0이다.

## 2) 주성분의 수 결정

주성분분석의 목표는 원 자료의 변수보다 적은 수의 주성분으로 자료의 변량을 설명하는 데에 있기 때문에 주성분의 수를 결정하는 과정 또한 변량을 참고한다. 가장 일반적으로 사용되는 방법은 주성분을 통해 가져갈 변량의 비율을 결정하고 해당 비율을 충족시키는 최소한의 주성분을 고르는 방법이다.  $m$ 개의 주성분을 사용했을 때의 누적 변량  $t_m$ 은 식 (9)으로 구한다.

$$t_m = 100 \frac{\sum_{k=1}^m l_k}{\sum_{j=1}^p s_{jj}} = 100 \frac{\sum_{k=1}^m l_k}{\sum_{k=1}^p l_k} = \frac{100}{p} \sum_{k=1}^m l_k \quad \text{식 (9)}$$

$s_{jj}$ 는  $p$ 개의 변수를 가진 자료에서  $j$ 번째 변수의 분산이며  $l_k$ 는  $k$ 번째 주성분의 변량이다. 앞서 언급한대로 주성분분석은 원자료의 변량의 합이 주

성분들의 변량의 합과 동일하다는 전제를 갖기 때문에  $\sum_{j=1}^p s_{jj}$ 는  $\sum_{k=1}^p l_k$ 로 대체 가능하다. 컷오프로 사용할 누적 변량  $t^*$ 에 대한 명확한 기준은 없으며 주성분분석을 활용한 논문들은 대부분 누적 변량이 80%이상을 기준으로 사용하고 있다(Petroni & Braglia, 2000).

주성분의 수를 결정하는 기준으로 누적된 변량의 양과 함께 카이저 규칙(Kaiser's rule)을 참고할 수 있다. 이는 벡터  $x$ 의 성분들이 모두 독립일 경우, 상관계수행렬에 기초해서 추려진 주성분 또한 원변수와 동일하게 각각 고유한 분산을 가진다는 전제를 둔다. 이러한 전제 하에 분산이 1보다 작은 주성분들은 원변수에 비해 정보의 양이 적으며 따라서 사용할 가치가 없다고 판단하며 주성분 중 변량  $l_k$ 가 1을 초과하는 주성분들만 선택한다. 그러나 카이저규칙이 주성분분석이 아닌 요인 분석에 조금 더 적절하게 이용될 수 있다는 점에서(Kaiser, 1960)  $l_k = 1$ 이라는 컷 오프 기준이 엄격한 기준일 수도 있다(Jolliffe, 2002).

이 외에도 스크리 그래프(Scree Graph), 바틀렛 검정(Bartlett's test), 편상관계수 등 여러 방법을 이용해 주성분의 수를 결정할 수 있으나, 대부분 명확한 가이드라인이 없기에 대부분 누적된 변량의 비율과 주성분의 분산의 크기를 참고한다.

### 3) 보조변수 투입에서의 주성분분석 활용

결측치 처리를 위한 보조변수의 투입에 대해 취할 수 있는 두 가지 입장은 다음과 같다. Inclusive Strategy의 경우 연구자가 결측치 추정에 도움이 될 수 있는 변수들을 확인할 방법이 없기 때문에 보조변수는 많이 투입할수록 좋다는 입장이다. 선행연구를 통해 결측치 변수, 모형과 전혀 상관없는 변수를 투입한다 해도 도움이 안 될 뿐 추정에 악영향을 미치지 않는다는 것이 밝혀져(Collins, Schafer & Kam, 2001), 대부분의 경우 보조변수는 가

능한 많이 투입하는 것을 권장한다. 그러나 Restrictive Strategy에 관련된 선행 연구들은 변수들의 상관만 보고 보조변수를 선택하는 경우 오히려 보조변수가 추정에 편향을 미칠 수 있음을 제기한다(Thoemmes & Rose, 2014). 또한 제한된 적은 표본을 사용하는 연구의 경우, 너무 많은 보조변수를 투입하면 추정결과가 수렴하지 않는 문제가 발생한다. 이러한 맥락에서 주성분분석을 활용할 경우, 보조변수들이 갖고 있는 정보, 변량은 최대한 유지하면서 모형에 투입되는 보조변수의 수를 효과적으로 줄일 수 있다(Howard, Rherntulla & Little 2015). 선행 연구에 따르면 주성분 분석을 활용해 보조변수의 수를 축소하고자 할 때, 원자료인 보조변수들의 변량을 40% 가량 설명하는 주성분을 새로운 보조변수로 활용할 경우, 보조변수를 그대로 사용한 추정과 유사한 수행을 보였으며 표본의 수가 50이하인 소표본 조건에서 발생하는 수렴 문제 또한 해결한다(Howard, Rherntulla & Little 2015). 즉 주성분분석은 특히 다중대체법에 비해 투입할 수 있는 보조변수의 수가 제한적인 완전정보최대가능성법을 활용할 때 유용하게 활용될 수 있다. 그러나 이러한 선행 연구들은 모두 보조변수 간 상관이 동일하다는 전제하에 이루어진 시뮬레이션 연구이며 보조변수 간 상관의 동질성(Homogeneity of Correlation)이 주성분 분석을 활용한 결측치 추정에 어떤 영향을 주는지에 대해선 연구가 미흡하다.

이에 따라 본 연구는 Howard와 동료들(2015)의 선행 연구를 바탕으로, 보조변수 간 상관이 동질적이지 않은 경우에도 주성분 분석은 효과적으로 보조변수의 수를 축소하는지, 또 이때 사용하는 원 자료의 변량 중 40% 라는 컷오프가 유의한지 확인하고자 한다. 이를 위해 몬테카를로 시뮬레이션을 진행할 것이며, 생성된 자료를 활용하여 보조변수 간 상관에 따른 주성분 분석의 수행 결과를 비교할 것이다. 또한 구해진 주성분을 보조변수로 사용해서 결측치를 추정했을 때, 원 보조변수를 그대로 사용한 추정결과와 비교

함으로써, 연구자가 결측치 처리 과정에서 보조변수의 수를 줄이기 위해 주 성분 분석을 활용하고자 할 때 참고할 수 있는 가이드라인을 제시하고자 한다.

### Ⅲ. 연구문제 및 가설

**연구문제 1: 보조변수 간 상관성이 이질적일 때에도 주성분 분석은 효과적으로 보조변수의 수를 줄일 수 있는가?**

[가설 1-1] 보조변수 간 상관성이 이질적일 경우, 전체 변량의 40% 컷오프를 만족하기 위해선 보조변수 간 상관성이 동질적일 때보다 더 많은 수의 주성분이 필요할 것이다.

**연구문제 2: 보조변수 간 상관성이 이질적일 때, 전체 변량의 40%를 설명하는 주성분은 효과적으로 결측치를 처리할 것인가?**

[가설 2-1] 보조변수 간 상관성이 높고 동질적일 때, 주성분 분석을 통해 구해진 40%의 변량을 포함한 주성분은 보조변수를 그대로 활용할 때와 유사한 수준의 수행을 보일 것이다.

[가설 2-2] 보조변수 간 상관성이 높고 이질적일 때, 주성분 분석을 통해 구해진 40%의 변량을 포함한 주성분은 보조변수를 그대로 활용할 때보다 편향이 클 것이다.

[가설 2-3] 보조변수 간 상관성이 낮고 이질적일 때, 주성분 분석을 통해 구해진 40%의 변량을 포함한 주성분은 보조변수를 그대로 활용할 때보다 편향이 클 것이다.

## IV. 연구 방법

### 1. 자료 생성

#### 1) 몬테카를로 시뮬레이션

통계 프로그램 R과 Mplus 8.1를 통해 가상자료를 생성하고 이를 활용하여 가설을 검증하는 몬테카를로 시뮬레이션(Monte Carlo Simulation) 방법을 사용하였다. R은 다양한 통계모형에 대한 분석에 대하여 결과 및 그래프를 제공하는 무료 소프트웨어로, 이미 여러 분야에서 폭 넓게 활용되고 있다(조민호, 2016). Mplus 또한 통계 자료를 분석하는데 널리 이용되고 있는 프로그램으로 종단 데이터, 다층 데이터, 결측치가 포함된 데이터 등 다양한 형태의 자료를 유연하게 분석할 수 있으며 몬테카를로 시뮬레이션을 지원하고 있다(Muthén & Muthén, 2016). 무엇보다도 R은 다수의 사용자들이 등록한 다양한 통계 분석 패키지를 다운로드받아 사용할 수 있다는 것이 큰 장점이다. 이러한 이유로 본 논문에서는 두 통계 프로그램을 활용하여 몬테카를로 시뮬레이션을 진행하였다.

몬테카를로 시뮬레이션은 연구자가 통제된 조건에 부합하는 자료를 무선적으로 생성하고 이를 분석에 활용하는 방법으로, 일종의 실험이라고 말할 수 있다. 몬테카를로 시뮬레이션은 연구자가 정의내린 특정 조건을 만족하는 자료를 무선적으로 다수 생성할 수 있으며, 현실에선 수집하기 어렵거나 불가능한 자료도 임의로 생성해내어 통계적 분석이 가능하다는 이점이 있다(Stephenson & Holbert, 2003).

## 2) 시뮬레이션 조건

선행 연구를 통해 보조변수의 총 변량 중 40%를 설명하는 주성분을 새로운 보조변수로 취할 경우, 보조변수를 그대로 사용한 것과 유사하게 결측치를 추정한다는 것을 확인했다. 그러나 기존 연구들의 경우 보조변수 간 상관의 형태를 모두 동일한 크기로 고정했던 한계가 있다. 실제 연구 장면에서 연구자의 자료에 포함된 변수들이 유사한 크기의 상관을 갖고 있길 기대하긴 어렵기 때문이다. 따라서 본 논문에서는 상관이 동질적이지 않은 다수의 변수로부터 주성분을 구하고 이를 결측치 추정 과정에서 보조변수로 사용하고자 할 때, 선행연구와 동일하게 총 변량의 40%를 설명하는 주성분을 취하는 것이 효율적인지 알아보하고자 하였다. 이를 위해 몬테카를로 시뮬레이션을 활용하여 다양한 조건의 자료를 생성했으며, 일부 조건은 모든 모형에서 동일하게 적용됐다.

분석을 위해  $X$ 가  $Y$ 를 설명하는 간단한 회귀모형을 설정했으며  $Y$ 는 결측치를 포함하고 있는 변수다. 다수의 문항이  $X$ 와  $Y$ 를 구성하는 구조방정식 형태도 가능하나, 본 연구의 목적은 보조변수의 역할을 검증하는 것에 있기 때문에  $X$ 와  $Y$ 는 이론변수가 아닌 단일 변수인 측정변수로 사용했다.  $Y$ 에 존재하는 결측치는 모두 Linear MAR를 따르며 9개의 보조변수 중 첫 번째 보조변수인 A1의 영향을 받는다. 보조변수는 총 9개며, 주성분 분석을 통해 전체 변량 중 40%를 설명하는 지점을 절단점(cut-off score)로 설정하고 이를 최소한의 수로 만족하는 주성분을 새로운 보조변수로 활용했다. 분석 방법은 ML을 활용했으며, 보조변수는 Graham이 제안한 Saturated Correlates Model을 따라 투입됐다.

<표 1> 시뮬레이션 조건

Y와 보조변수 간 상관	결측치 비율	표본 크기	보조변수 간 상관의 동질성
.10	15%	50	Uniform
.20	25%	200	Moderate
.30	50%	400	
.40	75%	600	Low

주. Y와 보조변수 간 상관(4)×결측치 비율(4)×표본 크기(4)×보조변수 간 상관의 동질성(3)=192개의 조건

각 시뮬레이션 조건은 <표 1>와 같다. 결측치가 존재하는 변수 Y와 보조변수 9개간의 상관은 .10에서 .40까지 .10씩 증분한다. 이는 낮은 상관부터 높은 상관까지 다양한 조건에서 결측치 추정 수행 능력을 보기 위함이다. 9개 보조변수 간 상관은 동질성을 기준으로 세 가지 조건으로 나뉘었으며 각각 Uniform, Moderate Homogeneity(이하 Moderate로 표기), Low Homogeneity(이하 Low로 표기)로 명명한다. Uniform 조건은 선행연구를 반영하여 9개의 보조변수 간 상관을 모두 .40으로 고정한 조건이다(Collins, Schafer & Kam, 2001). Moderate 조건은 9개의 보조변수를 3개의 묶음으로 나누어 각 묶음 내에서는 중간 정도의 상관(.40)을, 묶음 간에는 낮은 상관(.10)을 갖도록 고정한 조건이다. 마지막으로 Low 조건에선 9개의 보조변수가 변수 간 거리에 따라 다른 상관을 가진다. 조건 별 보조변수 간 상관의 크기는 <표 2>, <표 3>, <표 4>에서 확인할 수 있다.

<표 2> Uniform 조건에서의 보조변수 간 상관

	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	1.00								
A2	.40	1.00							
A3	.40	.40	1.00						
A4	.40	.40	.40	1.00					
A5	.40	.40	.40	.40	1.00				
A6	.40	.40	.40	.40	.40	1.00			
A7	.40	.40	.40	.40	.40	.40	1.00		
A8	.40	.40	.40	.40	.40	.40	.40	1.00	
A9	.40	.40	.40	.40	.40	.40	.40	.40	1.00

<표 3> Moderate 조건에서의 보조변수 간 상관

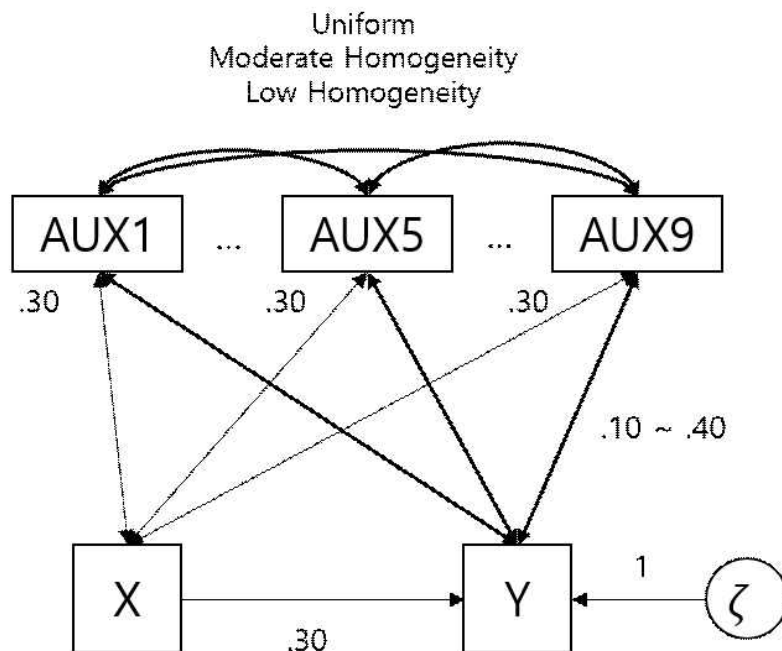
	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	1.00								
A2	.40	1.00							
A3	.40	.40	1.00						
A4	.10	.10	.10	1.00					
A5	.10	.10	.10	.40	1.00				
A6	.10	.10	.10	.40	.40	1.00			
A7	.10	.10	.10	.10	.10	.10	1.00		
A8	.10	.10	.10	.10	.10	.10	.40	1.00	
A9	.10	.10	.10	.10	.10	.10	.40	.40	1.00

<표 4> Low 조건에서의 보조변수 간 상관

	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1	1.00								
A2	.50	1.00							
A3	.45	.50	1.00						
A4	.40	.45	.50	1.00					
A5	.35	.40	.45	.50	1.00				
A6	.30	.35	.40	.45	.50	1.00			
A7	.25	.30	.35	.40	.45	.50	1.00		
A8	.20	.25	.30	.35	.40	.45	.50	1.00	
A9	.15	.20	.25	.30	.35	.40	.45	.50	1.00

또한 결측치의 비율이 높아질수록 추정에 오차가 발생한다는 선행연구 (Collins, Schafer & Kam 2001;) 결과를 고려하여 Y에 존재하는 결측치의 비율을 15%, 25%, 50%, 75% 네 가지 조건으로 나뉘었으며 전체 표본의 크기는 50, 200, 400, 600 네 가지 조건에서 살펴봤다(Thoemmes & Rose, 2014). 결측치 비율 75%는 연구 장면에서 굉장히 높은 비율인데, 이를 포함한 이유는 주성분분석을 결측치 처리에 활용한 기존의 선행 연구가 결측치의 비율을 10%에서 80%까지 확인했기 때문에 본 연구에서도 극단적인 조건에서 주성분 분석의 수행능력을 확인하기 위함이다(Howard, Rherntulla & Little, 2015). 또한 실제 데이터를 활용해 보조변수의 결측치 처리 수행능력을 확인한 경험적 연구에서 사용한 종단 자료의 결측치의 비율이 72%였기 때문에 결측치 비율 75%를 포함하기로 했다(Thoemmes & Rose, 2014). 이 뿐만 아니라 표본의 크기가 작고 결측치의 비율이 높은 극단적인 조건에서 다수의 보조변수를 투입했을 때 수렴에 실패하는 문제가 발생하는데, 이러한 수렴문제를 주성분분석이 해결할 수 있는지, 또 구해진 추정치는 편향되지

않았는지 확인하고자 했다. 따라서 이를 위해 극단적인 조건을 부여하고자 결측치 비율 75%도 연구에 포함했다. 표본의 크기 중 50은 작은 표본을 대표하며 600은 충분히 큰 표본을 의미한다. 이는 표본이 작을 때 여러 개의 보조변수를 한꺼번에 투입하면 추정결과가 수렴되지 않는 경우가 발생한다는 선행 연구 결과를 반영한 것이다(Howard, Rherntulla & Little, 2015). 언급된 모든 조건을 적용한 연구 모형은 <그림 12>와 같다.



<그림 12> 연구 모형

결측치는 세 가지 조건에서 추정되는데, 보조변수를 전혀 사용하지 않고 추정하는 조건(이하 No-AUX), 보조변수 9개를 모두 활용해서 추정하는 조건(이하 Inclusive), 마지막으로 주성분 분석을 활용한 조건(이하 PCA)로 구분한다.

## 2. 자료 분석

각 조건별로 결측치 추정 수행 결과를 비교하기 위해  $Y$ 에 결측치가 존재하는 자료를 생성한 후, 이를 세 가지 분석방법을 활용해 추정하여  $Y$ 의 평균  $\mu_Y$ ,  $Y$ 의 분산  $\sigma_Y^2$ ,  $X$ 와  $Y$ 의 상관인  $\rho_{XY}$ 를 구하고 이를 평가했다. 사용된 분석방법은 보조변수의 투입 없이 완전정보최대가능성법을 사용한 FIML 분석, 보조변수를 그대로 투입한 Inclusive 분석, 주성분을 새로운 보조변수로 활용한 PCA 분석이다. 평가 기준은 선행 연구 결과를 바탕으로 모수 편향(Parameter bias), 평균 제곱근 편차(Root-mean square error, 이하 RMSE), 신뢰구간(Confidence Interval: CI), 추정 결과의 수렴(Convergence) 여부다(Collins, Schafer & Kam, 2001; Graham, 2003; Howard, Rherntulla & Little, 2015).

### 1) 모수 편향(Parameter bias)

가장 먼저 추정을 통해 구해진 모수가 전집의 모수와 비교했을 때 얼마나 유사한지 확인했다.  $\mu_Y, \sigma_Y^2, \rho_{XY}$ 의 평균값을 구하고 전집의 값과 비교했으며, 식 (10)과 같다(Howard, Rherntulla & Little, 2015).  $\theta$ 는 전집의 모수이며  $\bar{\theta}$ 는 추정된 모수의 평균이다.

$$Bias = \bar{\theta} - \theta \quad \text{식 (10)}$$

### 2) RMSE(Root-mean square error)

RMSE는 추정을 통해 구해진 모수와 전집의 편차 평균을 제공한 값으로, 추정이 얼마나 정밀했는지 평가할 수 있는 지표다(Collins, Schafer & Kam,

2001). RMSE는 식 (11)을 사용해 구한다.  $\theta$ 는 전집의 모수,  $\hat{\theta}_i$ 는 각 추정마다 구해진 모수다. 수식에서 사용된  $N$ 은 전체 조건 중 추정 결과가 수렴된 조건의 수다.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta)^2}{N}} \quad \text{식 (11)}$$

### 3) 신뢰구간(Confidence Interval: CI) 범위

추정된 모수가 포함된 정규분포에 모집단의 모수가 포함될 확률을 의미하며 추정의 정확도와 편향 여부를 가늠할 수 있다. 90% 신뢰구간은  $\hat{\theta}_i \pm (1.645)\sigma_{\theta_i}$ 로 구하며  $\hat{\theta}_i$ 는 각 조건에서 구해진 추정치,  $\sigma_{\theta_i}$ 는 추정치들의 편차를 의미한다.

### 4) 수렴(Convergence) 여부

표본의 수가 작은 조건에서 다수의 보조변수를 전부 모형에 투입할 경우, 추정결과가 수렴하지 않을 수 있음이 선행 연구에서 밝혀졌기 때문에 수렴 여부도 함께 고려하고자 한다(Howard, Rherntulla & Little, 2015). 또한 보조변수를 전부 투입하면 수렴에 실패하는 모형이 주성분 분석을 활용했을 때에는 수렴에 성공하는지 여부를 확인함으로써 결측치 처리 과정에서 주성분 분석 효용을 살펴봤다.

## V. 연구결과

### 1. 주성분의 수

연구에서 설정한 모든 조건에서 주성분 분석은 보조변수 간 상관의 형태 (Uniform, Moderate, Low)에 관계없이 9개의 보조변수를 2개의 주성분으로 축소했다. 이는 전체 변량 40%를 주성분의 수 결정 기준으로 했을 때, 보조변수 간 상관의 형태는 주성분 분석의 수행에 유의미한 영향을 미치지 못함을 의미한다. 주성분 분석의 수행 능력을 확인하기 위해 연구 조건 외에 추가적으로 보조변수 간 상관의 크기를 .20으로 동일하게 고정하고 주성분 분석을 500회 반복 시행했다. 그 결과 보조변수 간 상관의 크기가 .20일 때, 전체 변량의 40%를 설명하기 위해선 최소 2개, 최대 3개의 주성분이 필요한 것으로 확인됐으며, 500회의 시행 평균은 2.7개였다.

### 2. 수렴(Convergence) 여부

선행 연구에서 다수 언급된 작은 표본에서의 수렴 실패 문제를 본 연구에서도 확인할 수 있었다 구조방정식과 같은 연구 모형을 수립하고 이를 가지고 있는 자료에 적용하고자 할 때, 간명성을 가지기 위해선 모형의 식별과 평가는 필수적이다(MacCallum, 1995). 수렴 실패는 이러한 모형 식별 단계에서 발생하는 문제로 일반적으로 표본의 크기가 작거나 변수들의 분포가 불균형적일 때 발생한다(Allison, 2008). 모형이 수렴에 실패했다는 것은 연구자가 자료에 비해 너무 복잡한 모형을 설계했다는 의미이며 해당 모형으로는 분석 단계를 진행할 수 없음을 의미한다. 분석 결과, 표본의 크기가 50이고 결측치의 비율이 75%인 매우 극단적인 조건에서 보조변수 9개를 모두

투입할 경우, 500회의 시행 중 평균 440회의 시행이 모형 수렴에 실패했다. 또한 선행 연구와 마찬가지로 동일한 조건에서 9개의 보조변수 대신 전체 분산의 40%를 설명하는 주성분을 추려 보조변수로 활용할 경우, 모형 수렴에 성공하여 추정치를 구할 수 있었다.

표본의 크기가 50인 작은 표본 조건에서도 결측치가 15%, 25%, 50%일 때 보조변수간 상관의 형태, Y와 보조변수 간 상관의 크기와 관계없이 9개의 보조변수를 모두 포함해도 500회의 반복시행 모두 수렴에 성공했다. 또한 결측치의 비율이 75%여도 표본의 크기가 200, 400, 600인 나머지 조건에선, 9개의 보조변수를 모두 사용한 Inclusive 추정도 500회의 시행 모두 수렴에 성공했다.

결측치가 존재하는 변수 Y와 보조변수들 간의 상관 크기에 따른 수렴 여부의 차이는 보이지 않았다. 보조변수들 간 상관 형태의 경우, 9개의 보조변수가 모두 동일한 크기의 상관(.40)을 가지고 있는 Uniform 조건에선 보조변수를 모두 사용한 Inclusive 추정만 수렴에 실패했으나 보조변수 간 상관의 크기가 다른 Moderate와 Low 조건에선 500회의 시행 중 2회 꼴로 PCA를 활용해도 수렴에 실패하는 것으로 나타났다.

### 3. 모수 편향(Parameter bias)

#### 1) X와 Y의 상관

보조변수 간 상관이 동질적일 때, 혹은 이질적일 때에 따라 각 결측치 처리 방법의 수행을 비교하기 위해 X와 Y의 상관인  $\rho_{XY}$ 의 모수 편향을 확인했다. 모수 편향은 구해진 모수가 전집과 비교했을 때 얼마나 유사한지 확인하는 지표이며 값이 작을수록 전집과 유사하게 추정했음을 의미한다. Y와 보조변수 간 상관을 .10부터 .40까지 .10씩 증분하며 추정 결과를 비교했으나, 네 조건 모두 결과가 유사했기 때문에 본문에선 Y와 보조변수 간 상관

이 .40인 조건을 기준으로 표와 해석을 정리했다. 또한 결측치 비율이 25%, 50%인 조건에서의 추정 결과가 대체로 결측치의 비율이 15%인 조건과 비슷한 양상을 보였기 때문에 본문의 표엔 결측치의 비율이 15%, 75%일 때의 결과만 추려 정리했다. 본문에 첨부하지 않은 나머지 모든 조건에서의 결과는 부록에서 확인이 가능하다.

모수 편향의 경우 표본의 크기가 클수록 감소하는 경향이 일관되게 관찰되었다. 사례수가 50이고 결측치의 비율이 75%인 조건에선 세 가지 추정 방법 모두 다른 조건에 비해 모수 편향 값이 크게 나왔는데(Y와 보조변수 간 상관인 .40인 조건에서, No-AUX=-0.032, Inclusive=0.360, PCA=0.245), 이는 설정한 조건이 사례수가 매우 적고 결측치의 비율은 매우 높은 극단적인 조건이기 때문이다. 해당 조건에서 보조변수를 모두 투입한 Inclusive 추정은 다른 추정 방법에 비해 모수편향이 특히 크게 나타났다. 또한 Inclusive 추정의 경우 표본의 크기가 50인 조건에서 모수편향이 크게 나타나는 것을 확인할 수 있었다. 표본의 크기가 200, 400, 600인 조건에선 보조변수를 활용하지 않은 No-AUX보다 작고 PCA 추정과 유사한 크기의 모수편향 값을 갖는 것으로 봤을 때, 이는 Inclusive 추정이 표본의 크기에 민감하며 특히 연구 내에서 50으로 대표되는 작은 표본에선 추정이 부정확해질 수 있음을 시사한다.

추정치와 모수 편향은 보조변수 간 상관의 영향도 받는 것을 확인할 수 있었다. 특히 PCA를 활용한 경우, 보조변수 간 상관이 이질적일 때 편향이 심각했다. 표본의 크기가 50이고 결측치의 비율이 75%일 때, 보조변수 간 상관이 낮은 Low조건에서 No-AUX 조건의 모수편향은 .246, Inclusive 조건의 모수편향은 .328이었으나, PCA 조건은 .504로 매우 큰 값을 가졌다. 이는 동일한 PCA 조건 내에서도 보조변수간 상관이 모두 동일했던 Uniform 조건(0.245)이나 중간 정도의 동질성을 갖고 있는 Moderate 조건(0.124)와

비교해봤을 때에도 유의미하게 큰 편차다. 표본의 크기가 50이고 결측치의 비율이 75%인 조건의 경우 Inclusive 추정이 500회의 시행 중 수렴에 실패한 시행이 다수 있다는 것을 고려해 보조변수 간 상관의 이질적인 조건 내에서 사례수와 결측치의 비율이 다른 조건에서의 수행 결과도 확인을 했다. 그 결과, 보조변수 간 상관의 형태가 이질적인 Low 조건에서 표본의 크기가 400, 600일 때 PCA 추정의 모수편향이 Inclusive 추정의 모수편향보다 커지는 것을 확인했다. 표본의 크기가 400일 때 결측치의 비율이 15%일 경우 Inclusive 추정의 모수편향은 0.029였으며 PCA 추정의 모수편향은 0.036이었다. 결측치의 비율이 25%인 조건에서의 모수편향은 각각 0.066과 0.072였다. 이러한 패턴은 보조변수 간 상관의 형태가 가장 이질적인 Low 조건에서만 나타났으며 Moderate 조건에선 나타나지 않았다. 그러나 보조변수 간 상관의 크기와 형태가 동일한 Uniform 조건에서 PCA 추정이 Inclusive 추정과 비슷하거나 더 작은 모수편향 값을 가졌던 것과 비교했을 때, 보조변수 간 상관의 동질성은 PCA를 활용한 결측치 추정에 영향을 미친다고 해석할 수 있다.

<표 5> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.249	0.117	0.295	0.151	0.298	0.113
		200	0.256	0.140	0.305	0.156	0.305	0.141
		400	0.253	-0.025	0.301	-0.023	0.301	-0.024
		600	0.255	-0.039	0.302	-0.049	0.302	-0.045
	75	50	0.303	-0.032	0.211	0.360	0.336	0.245
		200	0.300	0.179	0.298	0.089	0.298	0.146
		400	0.292	0.072	0.294	0.052	0.294	0.061
		600	0.292	0.014	0.294	0.003	0.293	-0.007
Mod	15	50	0.246	0.125	0.289	0.131	0.289	0.127
		200	0.261	-0.086	0.308	-0.082	0.307	-0.083
		400	0.255	-0.044	0.302	-0.029	0.302	-0.028
		600	0.254	-0.046	0.300	-0.064	0.300	-0.060
	75	50	0.313	-0.188	0.169	0.107	0.303	0.124
		200	0.299	0.269	0.300	0.135	0.300	0.130
		400	0.290	-0.060	0.297	-0.064	0.297	-0.065
		600	0.291	0.016	0.298	0.042	0.297	0.045
Low	15	50	0.251	0.102	0.295	0.133	0.251	0.101
		200	0.254	0.137	0.304	0.150	0.294	0.140
		400	0.253	-0.024	0.301	-0.029	0.290	-0.036
		600	0.255	-0.035	0.302	-0.038	0.292	-0.038
	75	50	0.322	0.246	0.142	0.328	0.271	0.504
		200	0.300	0.198	0.298	0.104	0.297	0.125
		400	0.292	0.052	0.293	0.027	0.293	0.051
		600	0.291	0.014	0.294	0.010	0.292	-0.013

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

## 2) Y의 평균

Y의 평균,  $\mu_Y$  또한  $\rho_{XY}$ 와 유사한 결과를 보였다. 표본 크기 50, 결측치 75% 조건의 경우 보조변수를 모두 투입한 Inclusive 추정은 수렴에 실패해 추정치를 구할 수 없는 시행이 다수 발생했으며, 구해진 추정치 또한 부정확하고 편향된 것을 확인할 수 있었다. 그러나 PCA를 활용할 경우 수렴문제를 해결할 수 있었다. Y와 보조변수 간 상관의 크기가 .40이고 보조변수 간 상관이 동질적인 Uniform 조건에서 보조변수를 활용하지 않은 No-AUX 추정의 Y 평균 모수편향은 -0.188, 보조변수를 모두 투입한 Inclusive 추정은 -1.299, 주성분을 보조변수로 활용한 PCA 추정은 -0.618이었다. 다만 수렴에 성공해 추정치를 얻은 PCA 추정도 모수 편향이 큰 값을 가져 추정 결과가 부정확할 수 있음을 고려해야 한다(No-AUX=0.837(-0.188), Inclusive=1.036(-1.299), PCA=0.986(-0.618)).

PCA 추정은 보조변수 간 상관이 동질적인 Uniform 조건에서, Inclusive 추정과 유사하거나 더 나은 수행을 보였다. 그러나 보조변수 간 상관이 이질적일수록 PCA를 활용한 추정은 보조변수를 모두 활용한 Inclusive 추정에 비해 다소 편향된 것을 확인했다. 보조변수 간 상관이 이질적이고(Low) 표본의 크기가 50, 결측치의 비율이 15%조건에서 보조변수를 모두 사용할 경우 추정치는 1.018, 모수 편향은 0.053이었으나, PCA를 활용할 경우, 추정치는 1.214, 모수 편향은 -0.057이었다. 앞과 동일한 조건에서 표본의 크기가 200인 조건에서도 Inclusive 추정의 Y 평균 추정치는 0.997, 모수편향은 0.008이었으며, PCA 추정의 Y 평균 추정치는 1.037, 모수편향은 0.041이었다. 다만 이러한 패턴이 항상 일정하게 나타나는 것은 아니었다.

표본의 크기가 50일 때 Inclusive 추정의 모수편향이 다른 두 추정방법에 비해 크게 나타나는 현상은 Y의 평균에서도 발견됐다.

<표 6> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 모수 편향

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	1.212	-0.075	1.019	0.025	1.014	-0.093
		200	1.174	-0.023	0.996	-0.008	0.998	0.022
		400	1.175	-0.039	1.001	-0.063	1.001	-0.054
		600	1.164	0.044	0.999	0.066	0.999	0.050
	75	50	0.837	-0.188	1.036	-1.299	0.986	-0.618
		200	0.896	-0.136	0.995	0.035	0.998	-0.047
		400	0.920	-0.133	1.004	-0.075	1.003	-0.060
		600	0.928	-0.005	1.007	-0.061	1.006	-0.035
Mod	15	50	1.219	-0.061	1.020	-0.144	1.032	-0.141
		200	1.166	0.109	0.995	0.068	0.998	0.069
		400	1.170	0.046	1.002	0.020	1.003	0.019
		600	1.167	0.060	1.003	0.083	1.004	0.072
	75	50	0.825	0.000	1.074	0.021	0.996	-0.184
		200	0.897	-0.199	1.002	0.055	1.000	0.053
		400	0.921	-0.020	1.004	0.042	1.003	0.036
		600	0.930	-0.003	1.004	-0.051	1.003	-0.055
Low	15	50	1.215	-0.058	1.018	0.053	1.214	-0.057
		200	1.179	-0.013	0.997	0.008	1.037	0.041
		400	1.176	-0.030	1.001	-0.044	1.040	-0.031
		600	1.164	0.040	0.999	0.052	1.035	0.045
	75	50	0.655	0.014	0.879	0.048	0.922	0.045
		200	0.896	-0.154	0.996	0.024	0.983	0.001
		400	0.920	-0.118	1.004	-0.056	0.990	-0.066
		600	0.929	-0.004	1.007	-0.061	0.994	-0.018

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

### 3) Y의 분산

$\sigma_y^2$  또한 앞선 두 결과와 유사했다. 표본의 크기가 작고 결측치의 비율이 높아 수렴의 문제가 발생하지 않는 이상, 보조변수를 모두 투입한 Inclusive 추정과 주성분을 보조변수로 활용한 PCA 추정은 보조변수 없이 추정한 No-AUX 추정과 비교했을 때 유사하거나 더 나은 수행을 보였으며, Inclusive 추정에서 수렴문제가 발생할 경우, PCA를 활용해 이를 해결하고 추정치를 구할 수 있었다. Y와 보조변수 간 상관의 크기가 .40이고 보조변수 간 상관이 모두 동일한 Uniform 조건에서 표본의 크기가 200이고 결측치의 비율이 15%일 때, No-AUX 추정의 모수편향은 -0.012였다. 동일한 조건에서 Inclusive 추정의 모수편향은 -0.006이었고 PCA 추정의 경우 -0.019였다. 표본의 크기가 50이고 결측치의 비율이 75%일 경우, 보조변수 9개를 모두 투입한 Inclusive 추정은 500회의 시행 중 60회 가량만 수렴에 성공했으며 모수편향은 1.718이었다. 한편 PCA 추정의 경우 9개의 보조변수를 2개로 축소해 500회의 시행 모두 수렴에 성공했으며 모수편향 또한 0.458로 Inclusive 추정에서 발생한 수렴 문제를 해결할 수 있었다. 다만 이 경우, 두 추정방법 모두 보조변수를 투입하지 않은 No-AUX 추정의 모수편향인 0.156보다 큰 값을 가지기 때문에 보조변수 투입에 대한 재고가 필요해 보인다.

PCA를 활용한 추정은 보조변수 간 상관이 이질적인 조건 외엔 보조변수를 모두 투입한 추정과 유사하거나 더 나은 추정치를 보였다. 보조변수 간 상관이 이질적인 Low 조건일 때, 결측치의 비율이 15%, 표본의 크기가 600인 시행에서 Inclusive 추정의 모수 편향은 0.015였으나 PCA 추정의 모수편향은 0.020으로 수행이 다소 떨어지는 것을 확인할 수 있었다. 그러나 보조변수 간 상관이 이질적인 조건에서 PCA 추정의 모수편향이 Inclusive 추정의 모수편향보다 큰 현상은 일관적이지 않았다.

모수 편향치는 표본 크기의 영향을 가장 크게 받는 것으로 보였으며 표본 크기가 클수록 모수 편향치는 전반적으로 낮아지는 것을 확인할 수 있었다. 특히 Inclusive 추정의 경우 표본의 크기가 50인 조건에서 다른 두 방법에 비해 모수 편향이 큰 것을 확인할 수 있었다. 결측치의 비율이 15%이고 표본의 크기가 50일 때, 보조변수 간 상관의 형태가 동질적인 Uniform 조건에서 세 추정방법의 모수편향은 No-AUX 추정이 -0.230, Inclusive 추정이 -0.307, PCA 추정이 -0.268이었다. 보조변수 간 상관의 형태가 중간 정도의 동질성을 갖고 있는 Moderate 조건에서의 모수편향은 보조변수를 투입하지 않은 No-AUX 추정이 -0.228, 보조변수를 모두 활용한 Inclusive 추정이 -0.253, PCA 추정이 -0.222였다. 마지막으로 보조변수 간 상관의 형태가 이질적인 Low 조건에선 No-AUX 추정이 -0.218, Inclusive 추정이 -0.296, PCA 추정이 -0.218이었다. 다만 이러한 현상은 표본의 크기가 50인 조건에서 두드러지며 표본의 크기가 200이상일 경우 다른 두 추정방법과 유사하거나 더 나은 모수 편향을 갖는 것으로 확인됐다.

<표 7> Y와 보조변수 간 상관인 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.875	-0.230	0.967	-0.307	0.943	-0.268
		200	0.917	0.012	0.983	-0.006	0.981	-0.019
		400	0.926	0.054	0.989	0.097	0.989	0.082
		600	0.932	0.029	0.994	0.015	0.993	0.031
	75	50	0.777	0.156	1.979	1.718	1.028	0.458
		200	0.945	-0.029	0.998	-0.142	0.966	-0.100
		400	0.971	0.095	0.988	0.047	0.982	0.089
		600	0.983	-0.120	0.990	-0.102	0.987	-0.113
Mod	15	50	0.876	-0.228	0.957	-0.253	0.949	-0.222
		200	0.933	0.134	0.995	0.132	0.995	0.127
		400	0.936	0.120	0.998	0.156	0.999	0.158
		600	0.936	0.014	0.997	0.015	0.997	0.021
	75	50	0.775	0.228	0.920	-0.011	0.965	0.100
		200	0.944	-0.178	0.987	-0.134	0.983	-0.126
		400	0.973	0.180	0.991	0.158	0.991	0.163
		600	0.984	-0.109	0.993	0.015	0.993	0.025
Low	15	50	0.874	-0.218	0.965	-0.296	0.874	-0.218
		200	0.917	-0.007	0.983	-0.019	0.958	-0.034
		400	0.926	0.043	0.989	0.081	0.965	0.066
		600	0.932	0.028	0.994	0.015	0.970	0.020
	75	50	0.669	0.784	1.027	1.265	1.332	0.308
		200	0.945	-0.054	0.998	-0.175	0.963	-0.158
		400	0.972	0.110	0.990	0.097	0.982	0.103
		600	0.984	-0.119	0.990	-0.078	0.988	-0.088

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

## 4. RMSE

### 1) X와 Y의 상관

RMSE(Root-mean square error)는 추정의 정밀성을 평가하는 지표로 크기가 작을수록 추정이 정밀했음을 의미한다. 표본 크기 50, 결측치 75%인 조건에서 보조변수를 모두 투입한 Inclusive 추정은 500회의 시행 대부분 수렴에 실패했는데, RMSE는 그 중 수렴에 성공한 시행만 계산한다.

RMSE 또한 모수편향과 마찬가지로 표본의 크기가 증가함에 따라 값이 감소하는 것을 확인할 수 있었다. 가장 극단적인 조건인 표본 크기 50, 결측치 비율 75%인 시행을 제외한 모든 조건에서 Inclusive 추정과 PCA 추정은 보조변수를 투입하지 않았을 때에 비해 더 작은 RMSE 값을 가졌다. Y와 보조변수 간 상관이 .40이고 표본의 크기가 50, 결측치의 비율이 15%인 조건에서 보조변수를 투입하지 않은 No-AUX 추정의 RMSE는 0.183, 보조변수를 모두 투입한 Inclusive 추정의 RMSE는 0.172, PCA를 활용한 추정의 RMSE는 0.167이었다. 이는 가능한 보조변수를 모형에 투입하는 것이 결측치 추정에 도움이 된다는 선행연구를 지지하는 결과다.

한편 세 가지 추정방법 모두 표본의 크기가 50이고 결측치의 비율이 75%인 가장 극단적인 조건에선 다른 조건에 비해 RMSE가 크게 추정된 것을 확인할 수 있었다. 보조변수 간 상관이 모두 동일한 크기를 가지는 Uniform 조건에서 No-AUX 추정의 RMSE는 .470, Inclusive 추정의 RMSE는 .523, PCA 추정의 RMSE는 .434였다.

보조변수 간 상관의 형태는 RMSE에 큰 영향을 주지 않는 것으로 밝혀졌다. 세 가지 추정방법 모두 RMSE는 사례수가 증가함에 따라 감소하는 패턴을 보였으며 보조변수를 활용한 Inclusive 추정과 PCA 추정은 보조변수를 투입하지 않은 No-AUX 조건에 비해 더 작은 RMSE 값을 가졌다.

<표 8> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.249	0.183	0.295	0.172	0.298	0.167
		200	0.256	0.087	0.305	0.077	0.305	0.077
		400	0.253	0.074	0.301	0.056	0.301	0.056
		600	0.255	0.064	0.302	0.046	0.302	0.046
	75	50	0.303	0.470	0.211	0.523	0.336	0.434
		200	0.300	0.157	0.298	0.141	0.298	0.131
		400	0.292	0.106	0.294	0.094	0.294	0.090
		600	0.292	0.089	0.294	0.074	0.293	0.072
Mod	15	50	0.246	0.179	0.289	0.149	0.289	0.149
		200	0.261	0.087	0.308	0.072	0.307	0.073
		400	0.255	0.073	0.302	0.055	0.302	0.055
		600	0.254	0.067	0.300	0.046	0.300	0.046
	75	50	0.313	0.465	0.169	0.287	0.303	0.286
		200	0.299	0.162	0.300	0.084	0.300	0.085
		400	0.290	0.102	0.297	0.060	0.297	0.060
		600	0.291	0.084	0.298	0.047	0.297	0.048
Low	15	50	0.251	0.183	0.295	0.171	0.251	0.183
		200	0.254	0.088	0.304	0.077	0.294	0.077
		400	0.253	0.074	0.301	0.056	0.290	0.056
		600	0.255	0.064	0.302	0.046	0.292	0.046
	75	50	0.322	0.449	0.142	0.324	0.271	0.514
		200	0.300	0.157	0.298	0.138	0.297	0.131
		400	0.292	0.105	0.293	0.092	0.293	0.089
		600	0.291	0.089	0.294	0.073	0.292	0.072

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

## 2) Y의 평균

Y의 평균 또한  $\rho_{XY}$ 와 마찬가지로 표본의 크기가 커질수록 RMSE가 감소하는 것을 일관되게 확인할 수 있었다. 특히 표본 크기가 50, 200인 조건과 400, 600인 조건의 차이가 컸다. Uniform 조건에서 결측치가 15%일 때, Inclusive 추정의 경우 표본의 크기가 50일 경우 RMSE는 0.262였으며 200일 경우 0.111이었다. 이에 비해 표본 크기가 400과 600인 조건에선 각각 0.084, 0.068로 크게 감소했으며 추정방법과 관계없이 모든 조건에서 유사한 변화 패턴을 보였다.

$\mu_Y$ 의 RMSE는 Y와 보조변수 간 상관의 영향도 받는 것으로 나타났다. Y와 보조변수 간 상관이 .10, ~.30일 경우, 표본의 크기가 50, 200인 조건에서 세 가지 추정방법의 RMSE는 유사했다. Y와 보조변수 간 상관이 .30이고 보조변수 간 상관이 Uniform인 조건에서 결측치가 15%일 경우, No-AUX 조건의 RMSE는 0.298였으며 보조변수를 모두 활용한 Inclusive 추정과 주성분을 활용한 PCA 추정은 각각 0.278와 0.252였다. 동일한 조건에서 표본 크기가 400, 600으로 증가할 경우 Inclusive 추정과 PCA 추정은 보조변수를 사용하지 않은 No-AUX 추정보다 더 정교한 추정을 했다. 사례수가 600일 때 No-AUX 추정의 RMSE가 0.138인 반면, Inclusive와 PCA 추정의 RMSE는 각각 0.071, 0.069였다. Y와 보조변수 간 상관이 .40일 경우 두 추정방법은 표본 크기가 작은 조건에서도 보조변수를 활용하지 않고 추정할 때보다 더 정교했다. 보조변수 간 상관의 크기가 동일한 Uniform 조건에서 표본크기가 50, 결측치 비율이 15%일 때. No-AUX 추정의 RMSE는 0.325였으며 Inclusive 추정과 PCA 추정의 RMSE는 각각 0.262, 0.239였다.

앞선 결과와 마찬가지로 표본의 크기가 작고 결측치의 비율이 매우 높은 극단적인 조건에선 Inclusive 추정과 PCA 추정 또한 RMSE가 크게 증가했

으며 보조변수를 활용하지 않고 추정할 때보다 더 불안정한 추정을 하는 것으로 나타났다. 보조변수 간 상관성이 동질적일 Uniform 조건에서 PCA 추정은 Inclusive 추정만큼 불안정하진 않았으나(No-AUX=0.545, Inclusive=0.675, PCA=0.548), 보조변수 간 상관성이 이질적인 Low 조건에선 Inclusive 추정보다 더 큰 RMSE를 가졌다(No-AUX=0.683, Inclusive=0.586, PCA=0.674). 보조변수 간 상관성이 이질적인 Low 조건에서 PCA 추정은 Inclusive 추정에 비해 큰 RMSE 값을 갖는 것을 확인할 수 있었는데 이러한 패턴은 다소 불규칙하게 나타났다. 결측치의 비율이 15%고 표본의 크기가 50인 조건에서 Inclusive 추정의 RMSE는 0.261이었으며 PCA 추정의 RMSE는 그보다 큰 0.329였다. 표본의 크기가 증가하거나 결측치의 비율이 증가해도 이러한 패턴을 확인할 수 있었으나 표본의 크기가 50인 경우, 결측치의 비율이 25%, 50%인 조건에선 Inclusive 추정은 PCA 추정보다 큰 RMSE 값을 가졌다. 각 결측치 비율에 따른 Inclusive 추정의 RMSE는 0.276, 0.347이었으며 PCA 추정의 RMSE는 0.253, 0.306였다. 이는 앞서 확인한 모수편향과 마찬가지로 Inclusive 추정이 다른 두 추정방법에 비해 표본 크기의 영향을 많이 받으며 특히 표본의 크기가 50으로 작을 경우, 그 차이가 두드러지게 나타난다고 해석할 수 있다.

<표 9> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 RMSE

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	1.212	0.325	1.019	0.262	1.014	0.239
		200	1.174	0.204	0.305	0.111	0.998	0.109
		400	1.175	0.193	1.001	0.084	1.001	0.081
		600	1.164	0.176	0.999	0.068	0.999	0.067
	75	50	0.837	0.545	1.036	0.675	0.986	0.548
		200	0.896	0.241	0.995	0.201	0.998	0.183
		400	0.920	0.168	1.004	0.124	1.003	0.120
		600	0.928	0.140	1.007	0.104	1.006	0.101
Mod	15	50	1.219	0.330	1.020	0.213	1.032	0.221
		200	1.166	0.199	0.995	0.103	0.998	0.107
		400	1.170	0.188	1.002	0.076	1.003	0.078
		600	1.167	0.179	1.003	0.060	1.004	0.062
	75	50	0.825	0.549	1.074	0.400	0.996	0.363
		200	0.897	0.240	1.002	0.121	1.000	0.122
		400	0.921	0.164	1.004	0.082	1.003	0.083
		600	0.930	0.136	1.004	0.068	1.003	0.069
Low	15	50	1.215	0.329	1.018	0.261	1.214	0.329
		200	1.179	0.209	0.997	0.111	1.037	0.115
		400	1.176	0.194	1.001	0.083	1.040	0.090
		600	1.164	0.177	0.999	0.068	1.035	0.074
	75	50	0.655	0.683	0.879	0.586	0.922	0.674
		200	0.896	0.241	0.996	0.196	0.983	0.184
		400	0.920	0.167	1.004	0.121	0.990	0.119
		600	0.929	0.139	1.007	0.101	0.994	0.102

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

### 3) Y의 분산

추정된 분산의 RMSE도  $\mu_Y$ 와 유사한 결과를 보였다. 세 추정방법 모두 표본 크기가 증가할수록 RMSE가 감소하는 것을 확인했으며 표본의 크기가 50인 조건을 제외하면 보조변수를 활용한 Inclusive 추정과 PCA 추정은 보조변수를 투입하지 않은 No-AUX 추정보다 작은 RMSE를 가졌다. 표본 크기가 50, 200, 400, 600으로 증가할 때, 보조변수를 투입하지 않은 No-AUX 추정의 RMSE는 0.242, 0.131, 0.103, 0.088이었으며 보조변수를 모두 투입한 Inclusive 추정의 경우 0.250, 0.115, 0.079, 0.064였다. PCA 추정 또한 앞선 두 추정과 동일하게 사례수가 증가할수록 RMSE는 작아졌다. PCA 추정의 RMSE는 표본크기가 50일 때 0.242, 200일 때 0.113, 400일 때 0.078이었으며 표본크기가 600일 땐 0.063이었다.

표본 크기가 50이고 결측치의 비율이 75%며 보조변수 간 상관성이 동질적일 때, 보조변수를 모두 활용한 Inclusive 추정의 RMSE가 나머지 두 추정에 비해 매우 크게 나타났다. No-AUX 추정의 RMSE가 0.485, PCA 추정의 RMSE가 0.896이었던 것에 비해 Inclusive 추정의 RMSE는 2.458이었다. 동일한 조건에서 PCA를 활용할 경우 Inclusive 추정에 비해 추정치가 안정적이거나, 보조변수가 상관성이 이질적인 Low 조건에서 PCA 추정은 다른 두 방법에 비해 큰 RMSE를 가졌다. 표본 크기가 50, 결측치의 비율이 75%이며 보조변수간 상관성이 이질적인 Low 조건에서 PCA 추정의 RMSE는 1.200이었다. 보조변수를 사용하지 않은 No-AUX 추정의 RMSE는 0.535였으며 보조변수를 모두 투입한 Inclusive 추정의 RMSE는 0.869였다. 그러나 이러한 패턴은 일정하지 않았다.

표본의 크기가 50일 때, 다른 두 추정에 비해 Inclusive 추정의 RMSE가 다소 큰 패턴 또한 일관적으로 확인할 수 있었다.

<표 10> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_y^2(1)$ 의 RMSE

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.875	0.242	0.967	0.250	0.943	0.242
		200	0.917	0.131	0.983	0.115	0.981	0.113
		400	0.926	0.103	0.989	0.079	0.989	0.078
		600	0.932	0.088	0.994	0.064	0.993	0.063
	75	50	0.777	0.485	1.979	2.458	1.028	0.896
		200	0.945	0.219	0.998	0.226	0.966	0.210
		400	0.971	0.151	0.988	0.146	0.982	0.143
		600	0.983	0.124	0.990	0.115	0.987	0.114
Mod	15	50	0.876	0.243	0.957	0.215	0.949	0.214
		200	0.933	0.117	0.995	0.096	0.995	0.098
		400	0.936	0.095	0.998	0.068	0.999	0.069
		600	0.936	0.088	0.997	0.059	0.997	0.059
	75	50	0.775	0.496	0.920	0.502	0.965	0.418
		200	0.944	0.220	0.987	0.139	0.983	0.138
		400	0.973	0.151	0.991	0.094	0.991	0.094
		600	0.984	0.125	0.993	0.075	0.993	0.077
Low	15	50	0.874	0.243	0.965	0.247	0.874	0.243
		200	0.917	0.132	0.983	0.114	0.958	0.116
		400	0.926	0.103	0.989	0.079	0.965	0.083
		600	0.932	0.088	0.994	0.064	0.970	0.068
	75	50	0.669	0.535	1.027	0.869	1.332	1.200
		200	0.945	0.219	0.998	0.226	0.963	0.209
		400	0.972	0.151	0.990	0.144	0.982	0.144
		600	0.984	0.124	0.990	0.113	0.988	0.114

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

## 5. 신뢰구간(Confidence Interval)

### 1) 신뢰구간의 범위

추정의 정확도와 편향 여부를 확인하기 위해 90% 신뢰구간과 그 길이를 확인했다. 일반적으로 여러 추정방법을 비교할 때, 신뢰구간 내에 모수가 포함될 확률이 높으면서 그 구간의 너비가 더 짧은 방법이 수행이 더 좋다고 평가한다(Collins, Schafer & Kam, 2001). 95% 신뢰구간이 가장 일반적으로 보고되나, 계량심리학에선 90% 신뢰구간도 빈번하게 사용된다. 또한 사례수가 작은 경우 90% 신뢰구간을 사용하기도 하며 선행연구 또한 90% 신뢰구간을 보고하고 있기에 이에 따라 90% 신뢰구간을 보고한다(Collins, Schafer & Kam, 2001; Howard, Rherntulla & Little, 2015).

신뢰구간이 0을 포함하는지 여부에 따라 비교하고자하는 두 집단 사이에 유의미한 차이가 있는지 확인할 수 있다. 본 연구에서는 세 가지 결측치 추정 방법이 자료 생성 단계에서 설정한 모수와 근접하게 값을 추정하는지 알아보는 데에 목적이 있기 때문에 신뢰구간이 0을 포함하지 않을 때, 즉 두 집단 사이에 유의미한 차이가 없다고 할 때 추정이 정확했다고 해석한다.

Y와 보조변수 간 상관의 크기가 .40인 조건에서 결측치 비율과 표본의 크기, 보조변수 간 상관의 형태에 따른 세 가지 추정방법의 평균 신뢰구간 범위는 <표 11>과 같다. 보조변수를 투입하지 않을 경우, 표본의 크기가 가장 작은 50인 조건에서 신뢰구간은 항상 0을 포함했다. 보조변수간 상관이 동질적인 조건에서 결측치의 비율이 15%일 때, 보조변수를 투입하지 않은 No-AUX 추정의 신뢰구간의 하한은 -0.013, 상한은 0.510이었다. 동일한 조건에서 결측치의 비율이 25%일 때 신뢰구간의 하한은 -0.018, 상한은 0.530이었으며 결측치의 비율이 50%일 때 신뢰구간의 하한은 -0.053, 상한은 0.604였다. 결측치의 비율이 75%인 조건에서 신뢰구간의 하한과 상한은

각각  $-0.268$ 과  $0.881$ 이었다. 이러한 패턴은 보조변수와 Y간 상관이  $.30$ 일 때도 나타나며  $.20$ 인 조건에서도 표본의 크기가  $50$ 인  $16$ 개의 조건 중  $7$ 개의 조건에서 No-AUX 추정의 신뢰구간은  $0$ 을 포함했다. Y와 보조변수간 상관이  $.10$ 일 경우 표본의 크기가  $50$ 인  $16$ 개의 조건 중  $6$ 개의 조건이 신뢰구간에  $0$ 이 포함됐다. 결측치가 존재하는 변수와  $.40$ 이상의 상관이 있을 경우 보조변수로서 투입할 것을 권장하는 선행 연구결과로 고려했을 때(Collins, Schafer & Kam 2001), 결측치가 존재하는 변수 Y와  $.30 \sim .40$  이상의 상관을 갖는 변수를 추정과정에서 보조변수로 사용하지 않을 경우, 보조변수를 사용한 다른 추정방법보다 정확도가 떨어진다고 해석할 수 있다.

보조변수를 활용한 Inclusive와 PCA 추정방법 또한 결측치의 비율이  $50\%$ ,  $75\%$ 인 경우, 표본의 크기가  $50$ 인 조건에서 신뢰구간이  $0$ 을 포함했다. Inclusive 추정의 경우 표본의 크기가  $50$ 이고 결측치의 비율이  $75\%$ 인 극단적인 조건에서 신뢰구간이  $0$ 을 포함하지 않는 경우도 있었으나,  $500$ 회의 시행 중 수렴에 성공한 시행이  $60$ 회 가량이기 때문에 다른 두 방법에 비해 결측치 추정 수행 능력이 더 좋다고 평가하기는 어렵다. 이러한 패턴은 보조변수와 Y간 상관의 크기가  $.10$ ,  $.20$ ,  $.30$ 일 때에도 일관적으로 나타났다. 즉 표본의 크기가  $50$ 인 조건에서 결측치의 비율이  $50\%$ 를 초과할 경우, 보조변수를 활용하는 Inclusive 추정과 PCA 추정의 결과 또한 부정확할 가능성을 염두에 두어야함을 의미한다.

<표 11> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA		
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L	
Uniform	15	50	-0.013	0.510	0.035	0.556	0.042	0.554	
		200	0.130	0.382	0.180	0.430	0.180	0.429	
		400	0.164	0.341	0.213	0.390	0.213	0.389	
		600	0.183	0.327	0.230	0.374	0.231	0.374	
	75	50	-0.268	0.881	-0.064	0.779	-0.139	0.761	
		200	0.051	0.549	0.093	0.504	0.093	0.503	
		400	0.118	0.467	0.150	0.437	0.150	0.437	
		600	0.150	0.434	0.177	0.410	0.177	0.410	
	Mod	15	50	-0.015	0.508	0.053	0.525	0.052	0.527
			200	0.135	0.387	0.190	0.426	0.189	0.426
			400	0.167	0.344	0.219	0.386	0.219	0.386
			600	0.182	0.326	0.232	0.368	0.232	0.368
75		50	-0.250	0.875	0.030	0.586	-0.021	0.598	
		200	0.052	0.547	0.165	0.434	0.162	0.438	
		400	0.116	0.464	0.202	0.392	0.201	0.394	
		600	0.149	0.432	0.220	0.375	0.219	0.375	
Low		15	50	-0.012	0.514	0.036	0.555	0.175	0.327
			200	0.128	0.381	0.179	0.429	0.170	0.417
			400	0.164	0.342	0.213	0.389	0.204	0.377
			600	0.183	0.327	0.230	0.373	0.221	0.363
	75	50	-0.303	0.964	0.361	0.916	-0.196	0.724	
		200	0.051	0.549	0.097	0.499	0.094	0.500	
		400	0.118	0.466	0.153	0.433	0.150	0.435	
		600	0.149	0.433	0.180	0.407	0.177	0.408	

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

## 2) 신뢰구간의 너비

신뢰구간의 너비는 표본의 크기와 결측치의 비율의 영향을 크게 받는 것으로 확인됐다. Y와 보조변수 간 상관의 크기가 .40인 조건을 기준으로, No-AUX, Inclusive, PCA 세 가지 추정방법 모두 표본이 크기가 커질수록 신뢰구간의 길이는 짧아졌다. 결측치의 비율이 15%이고 표본의 크기가 50, 200, 400, 600으로 증가할 때 보조변수를 투입하지 않은 No-AUX 추정의 신뢰구간 길이는 0.523, 0.251, 0.177, 0.144로 감소했으며 보조변수 9개를 투입한 Inclusive 추정의 경우 0.521, 0.250, 0.177, 0.144로 감소했다. 주성분을 보조변수로 활용한 PCA 추정도 각각 0.521, 0.249, 0.176, 0.143으로 감소하는 경향을 보였다.

동일한 표본의 크기를 가질 때, 결측치의 비율이 증가할수록 신뢰구간의 범위는 넓어지는 것을 확인했다. 표본의 크기가 200으로 동일한 조건에서 결측치의 비율이 15%, 25%, 50%, 75%로 증가할수록 세 추정방법의 신뢰구간 폭은 점점 넓어진다. 보조변수를 투입하지 않은 No-AUX 추정의 경우 0.251, 0.270, 0.333, 0.498로 신뢰구간의 폭이 넓어졌다. 보조변수를 사용하는 Inclusive와 PCA 추정도 신뢰구간의 폭이 넓어지는 것을 확인할 수 있었다. Inclusive 추정의 경우 결측치의 비율이 증가함에 따라 신뢰구간의 너비 또한 0.250, 0.262, 0.298, 0.411로 증가했으며, PCA 추정도 마찬가지로 0.249, 0.261, 0.298, 0.410으로 증가했다.

세 가지 추정방법 중 보조변수를 활용한 두 추정방법(Inclusive와 PCA)은 보조변수를 활용하지 않을 때보다 좁은 신뢰구간 범위를 가지는 것을 확인할 수 있었다. 이는 결측치의 비율이 15%인 조건에선 뚜렷하게 드러나지 않으나 결측치의 비율이 높아질수록 그 차이폭이 커졌다. 표본의 크기가 400인 조건에서 결측치의 비율이 15%일 때 보조변수를 활용하지 않은 No-AUX 추정의 신뢰구간 너비는 0.177, 보조변수를 활용한 Inclusive 추정

과 PCA 추정의 신뢰구간 너비는 각각 0.177, 0.176으로 소수 세 자리 기준 큰 차이가 없었으나, 결측치의 비율이 25%로 증가할 경우, 세 추정 방법의 신뢰구간 너비는 각각 0.192, 0.186, 0.185로 나타났다. 결측치의 비율이 50%인 조건에선 각각 0.238, 0.212, 0.212였다. 유사한 추정치와 편향, RMSE를 가진다고 할 때, 신뢰구간의 너비가 다른 방법보다 좁은 추정방법이 수행이 더 좋다고 평가할 수 있다. 즉 결측치를 추정할 때 보조변수를 활용하는 것이 보조변수를 활용하지 않고 추정하는 것보다 더 정확하고 좋은 수행 결과를 기대할 수 있음을 의미한다.

보조변수 간 상관의 동질성은 신뢰구간의 너비에 유의미한 영향을 주진 않는 것으로 확인됐다. 보조변수 9개를 모두 활용한 Inclusive 추정과 주성분을 새로운 보조변수로 활용한 PCA 추정은 결측치의 비율이 극단적으로 높고 표본의 크기가 작아 Inclusive 추정이 수렴에 실패한 경우가 아닌 이상 대부분의 조건에서 신뢰구간의 너비는 유사했다.

<표 12> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA	
Uniform	15	50	0.523	0.521	0.512	
		200	0.251	0.25	0.249	
		400	0.177	0.177	0.176	
		600	0.144	0.144	0.143	
	75	50	0.284	0.233	0.233	
		200	0.498	0.411	0.41	
		400	0.349	0.287	0.287	
		600	0.284	0.233	0.233	
	Mod	15	50	0.523	0.472	0.475
			200	0.252	0.236	0.237
			400	0.178	0.167	0.167
			600	0.144	0.136	0.136
75		50	1.112	0.034	0.546	
		200	0.495	0.269	0.276	
		400	0.348	0.189	0.192	
		600	0.283	0.154	0.156	
Low		15	50	0.526	0.518	0.152
			200	0.252	0.250	0.247
			400	0.178	0.176	0.174
			600	0.144	0.143	0.141
	75	50	0.170	0.001	0.081	
		200	0.498	0.402	0.406	
		400	0.349	0.281	0.285	
		600	0.284	0.228	0.231	

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

## VI. 논의

본 연구는 보조변수를 활용한 결측치 추정방법에서 주성분 분석의 활용방안에 대한 지표를 제시하고자했다. 자료에 존재하는 결측치를 무시하고 분석을 할 경우, 구해진 추정 결과가 편향될 가능성이 높기 때문에 분석 단계에서 결측치를 올바르게 처리하는 것이 중요하다(Little & Rubin, 2002). 기존에 사용되고 있거나 통계 프로그램의 기본으로 설정되어 있는 제거법은 정보의 누락이 크기 때문에 효율성이 떨어지는 단점이 있어 최근엔 다중대체법 혹은 완전정보최대가능성법을 활용하는 것을 권장하고 있다. 이중 완전정보최대가능성법은 구조방정식에 적용하고 모형의 평가가 가능하다는 점에서 이점을 갖고 있다. 그러나 완전정보최대가능성법은 보조변수를 투입하는데 제한을 갖고 있다는 단점이 있다(Enders, 2010). 다중대체법과 달리 완전정보최대가능성법은 너무 많은 보조변수를 투입하면 모형이 수렴하지 않기 때문이다. 이를 해결하기 위해선 투입되는 보조변수의 수를 줄이는 것밖에 없지만, 실제 연구 장면에서 어떤 변수가 결측치에 영향을 주는지 연구자가 확인할 수 없기 때문에 선택적으로 변수를 골라 보조변수로 투입하기는 어려운 점이 있다. 이러한 문제점을 해결하기 위해 선행 연구에선 주성분 분석(Principle Component Analysis)를 제시했다(Howard, Rherntulla & Little, 2015). 선행 연구를 통해 보조변수로 사용 가능한 변수들을 대상으로 주성분 분석을 실시해 전체 변량의 40%를 설명하는 주성분들을 추려내면 기존 보조변수와 동일한 수행을 보이는 것을 확인할 수 있었다. 그러나 결측치 추정에 주성분 분석을 활용한 연구는 아직 소수에 불과하며 이들 모두 보조변수 간 상관은 동일하게 고정했다. 그러나 실제 연구 장면에서 모든 보조변수가 서로 동일한 크기의 상관을 가질 거라 기대하긴 어렵다.

따라서 본 연구는 결측치 처리 과정에서 주성분 분석이 어떻게 활용될 수 있는지, 어떤 조건에서 사용하면 효과적이고 어떤 조건에선 수행이 떨어지는지 확인하기 위해 기존 선행 연구에서 확인한 조건뿐만 아니라 보조변수 간 상관의 크기와 동질성에도 차이를 두고 다양한 조건에서 주성분 분석을 활용한 결측치 처리의 수행 능력을 확인했다. 시뮬레이션 결과에 따른 연구 가설의 기각 및 지지 여부는 다음과 같다.

연구문제 1에선 보조변수 간 상관이 이질적일 때에도 주성분 분석이 보조변수의 수를 효과적으로 줄일 수 있는지 확인하고자 했다. 보조변수와 결측치가 존재하는 변수 Y와의 상관, 보조변수 간 상관의 형태, 결측치의 비율, 표본의 크기에 따라 조건을 나누고 각각 500회씩 반복 시행한 결과, 주성분 분석은 전체 변량 중 40%를 포함하는 것을 기준으로 잡았을 때, 보조변수로 생성한 9개의 변수를 항상 2개의 주성분으로 축소했다. 따라서 ‘가설 1-1. 보조변수 간 상관이 이질적일 경우, 전체 변량의 40%를 설명하기 위해선 보조변수 간 상관이 동질적일 때보다 더 많은 수의 주성분이 필요할 것이다.’는 지지되지 않았다.

연구문제 2에선 보조변수 간 상관의 형태에 따라, 전체 변량의 40%를 설명하는 주성분이 효과적으로 결측치를 처리할 것인지 확인하고자 했다. 시뮬레이션 결과, 보조변수 간 상관의 크기가 동일한 Uniform 조건일 때 주성분 분석은 보조변수와 Y간 상관의 크기, 표본의 크기, 결측치의 비율에 관계없이 대체로 보조변수 9개를 그대로 사용한 추정방법과 유사하거나 더 나은 수행을 보였다. 특히 표본의 크기가 작고 결측치의 비율이 높은 극단적인 조건에서 보조변수 9개를 모두 투입할 경우, 500회의 시행 중 400회 이상의 시행이 모형 수렴에 실패해 추정치가 구해지지 않거나 신뢰구간을 구할 수 없는 문제가 발생했다. 그러나 동일한 조건에서 주성분 분석을 활용할 경우 모형에 투입되는 변수의 수가 줄어들기 때문에 모형 수렴 문제를

해결할 수 있었으며 보조변수를 사용하지 않은 추정방법보다 정확함은 물론이고 추정치 또한 연구에서 설정한 값과 유사하게 구해졌다. 주성분 분석을 활용한 추정방법은 보조변수를 그대로 투입한 추정방법과 유사한 크기의 모수편향과 RMSE 값을 가졌으며 신뢰구간의 너비 또한 보조변수를 활용하지 않은 추정방법과 비교했을 때 더 좁았다. 따라서 ‘가설 2-1. 보조변수 간 상관성이 높고 동질적일 때, 주성분 분석을 통해 구해진 40%의 변량을 포함한 주성분은 보조변수를 그대로 활용할 때와 유사한 수준의 수행을 보일 것이다.’는 지지되었다.

보조변수 간 상관의 크기와 동질성에 따라 주성분분석을 활용한 결측치 추정의 수행능력을 확인하기 위해 Moderate 조건과 Low 조건을 새로 추가했다. Low 조건은 보조변수 간 상관성이 높고 이질적인 조건이며, Moderate 조건은 보조변수 간 상관성이 낮고 이질적인 조건으로 설정했다. 이를 위해 자료 생성 단계에서 Moderate조건은 9개의 보조변수를 3개씩 묶어 묶음 내 상관은 .40, 묶음 간 상관은 .10으로 고정했다. Low 조건은 9개의 보조변수를 .15부터 .50까지 .05씩 증분하여 상관을 고정했다. 시뮬레이션 결과, 보조변수 간 상관성이 높고 이질적인 Low 조건에서 PCA를 활용한 결측치 추정의 수행이 다소 떨어지는 것을 확인했다. PCA추정을 통해 구해진 Y와 X간 상관, Y의 평균 그리고 Y의 분산의 모수편향이 Inclusive 추정의 모수편향보다 크게 나타났으며 RMSE 또한 Inclusive 추정에 비해 다소 큰 것을 확인할 수 있었다. 다만 이러한 현상은 일정한 패턴을 갖기보단 다소 불규칙하게 나타났기 때문에 추가적인 연구가 필요할 것으로 보인다. 또한 신뢰구간의 범위나 너비에 대해선 유의미한 차이가 없었다. 따라서 ‘가설 2-2. 보조변수 간 상관성이 높고 이질적일 때, 주성분 분석을 통해 구해진 40%의 변량을 포함한 주성분은 보조변수를 그대로 활용할 때보다 편향이 클 것이다.’는 부분적으로 지지가 됐으나, 추가적인 연구가 필요하다고 볼 수 있다.

한편 보조변수 간 상관성이 낮고 이질적일 때 PCA를 활용한 결측치 추정 수행을 확인하기 위해 설정한 Moderate 조건에선 유의미한 차이를 볼 수 없었다. 대부분의 조건에서 보조변수 간 상관성이 동질적일 때와 유사한 패턴을 보였으며 보조변수를 그대로 투입한 Inclusive 추정과 비슷하거나 더 나은 수행을 보였다. 따라서 ‘가설 2-3. 보조변수 간 상관성이 낮고 이질적일 때, 주성분 분석을 통해 구해진 40%의 변량을 포함한 주성분은 보조변수를 그대로 활용할 때보다 편향이 클 것이다.’는 기각됐다.

본 시뮬레이션 연구의 주요 조건이었던 보조변수 간 상관성의 동질성을 포함하여 Y와 보조변수 간 상관, 결측치의 비율, 표본의 크기에 따른 연구 결과를 정리하자면 다음과 같다. Y와 보조변수 간 상관성이 .30~.40이상일 때, 결측치 추정에 보조변수를 투입하지 않을 경우 추정이 부정확하고 편향되는 것을 확인할 수 있었다. 이는 기존의 선행연구에서 결측치가 존재하는 변수와 .40이상의 상관성을 가진 변수라면 가능한 추정에 반영하는 것이 편향을 줄일 수 있다는 제안을 지지하는 결과다(Collins, Schafer & Kam, 2001).

모수편향과 RMSE의 경우 추정 방법에 상관없이 표본의 크기가 증가할수록 감소하는 것을 확인했다. 특히 보조변수 9개를 모두 사용하는 Inclusive 추정의 경우 표본의 크기가 50인 조건에서 다른 두 추정방법에 비해 모수편향과 RMSE의 값이 큰 것을 확인할 수 있었다. 표본의 크기가 200, 400, 600인 조건에선 보조변수를 투입하지 않은 No-AUX 추정보다 작은 값을 가지고 PCA를 활용한 추정과 유사하거나 더 나은 결과를 보이는 것으로 미루어보아 보조변수를 모두 투입하는 Inclusive 추정은 표본 크기에 민감하며 특히 50이하의 작은 표본에선 편향이 심해진다고 해석할 수 있다.

신뢰구간의 경우 보조변수를 투입하지 않을 경우 추정치의 신뢰구간에 0이 포함되는 것을 확인할 수 있었다. 이는 곧 구해진 추정치가 자료의 모수와 유의미하게 이질적임을 의미하며 결측치를 올바르게 처리하는 것이 목적

인 해당 연구에선 결측치가 제대로 추정되지 않았음을 의미한다고 해석할 수 있다. 보조변수를 활용하는 Inclusive 추정과 PCA 추정에서도 표본의 크기가 50일 때 결측치의 비율이 50%, 75%일 경우 신뢰구간에 0이 포함되는 것을 확인할 수 있었다. 이는 표본의 크기가 작을 때 결측치의 비율이 50%를 초과할 경우, 보조변수를 활용해도 결과에 편향이 존재할 수 있음을 의미한다. 신뢰구간 너비의 경우 보조변수 간 상관의 형태나 결측치가 존재하는 변수 Y와 보조변수 간 상관의 크기와 관계없이 표본의 크기가 클수록, 결측치의 비율이 낮을수록 좁아지는 것을 확인할 수 있었다.

결측치 처리에서 주성분 분석의 활용을 중심으로 연구 결과를 정리하자면 다음과 같다. 완전정보최대가능성법을 활용하여 결측치를 추정할 경우, 모형이 제대로 수렴하기 위해선 투입할 수 있는 보조변수의 수에 제한이 생긴다. 실제로 시뮬레이션 중 가장 극단적인 조건이었던 표본의 크기 50, 결측치의 비율이 75%인 조건에서 보조변수 9개를 그대로 사용할 경우 모형이 수렴하지 않는 것을 확인할 수 있었다. 그러나 9개의 보조변수에서 주성분 분석을 통해 전체 변량의 40%를 설명하는 주성분을 추려 새로운 보조변수로 활용할 경우, 모형이 수렴에 실패하는 문제를 해결할 수 있었다. 이는 주성분분석과 모형 수렴여부의 관계를 확인한 선행연구에서 확인한 바와 일치하는 결과다(Howard, Rherntulla & Little, 2015). 연구 내에서 설정한 모든 조건에서 주성분 분석은 9개의 보조변수를 2개의 주성분으로 축소했으며 요약된 주성분을 보조변수로 활용한 결측치 추정은 모수편향과 RMSE, 신뢰구간을 비교했을 때 보조변수를 그대로 사용한 추정과 유사하거나 더 나은 수행을 보였다. 그러나 보조변수 간 상관의 크기가 이질적인 조건에선 부분적으로 수행이 다소 떨어지는 것을 확인할 수 있었다. 또한 표본의 크기가 50인 조건에서 결측치의 비율이 50%를 초과할 경우, 주성분 분석을 활용한 추정방법 또한 보조변수를 투입하지 않은 추정과 보조변수 9개를 모두 투입

한 추정과 마찬가지로 신뢰구간에 0이 포함되는 것을 확인할 수 있었다. 따라서 본 연구는 결측치 처리 과정에서의 주성분 분석 활용 방안을 다음과 같이 제시한다. 첫째, 주성분 분석은 모형에 투입되는 보조변수의 수를 효과적으로 줄일 수 있다. 따라서 연구자의 자료가 표본의 크기가 작으며 결측치의 비율이 높을 경우, 주성분 분석을 활용해 모형이 수렴에 실패하는 것을 방지할 수 있다. 둘째, 투입된 보조변수들의 전체 변량 40%를 설명하는 주성분을 새로운 보조변수로 활용할 경우 기존의 보조변수와 유사한 수행을 보인다. 셋째, 보조변수 간 상관의 형태가 이질적일 경우, 주성분을 활용한 결측치 추정은 보조변수를 그대로 활용한 결측치 추정에 비해 편향이 다소 존재할 수 있다. 이를 방지하기 위해 자료 내에 존재하는 변수들의 상관의 크기와 형태를 먼저 확인해볼 것을 권장한다. 마지막으로 보조변수를 투입해도 표본의 크기가 50미만이며 결측치의 비율이 50%를 초과할 경우, 추정 결과가 부정확할 수 있다. 따라서 해당 조건에서의 추정치를 해석할 땐 주의가 필요하다.

본 연구의 의의는 다음과 같다. 첫째, 기존 선행연구에서 반영되지 않은 보조변수 간 상관의 크기와 동질성에 따른 주성분 분석의 수행 능력을 확인하기 위해 세 가지 조건을 두고 연구를 진행했으며 이에 따라 실제 연구 장면에서 결측치 추정을 위해 주성분 분석을 활용하고자 할 때 지표를 제공하고자 했다. 본 연구는 기존 선행연구(Howard, Rherntulla & Little, 2015)의 결과를 재확인함과 동시에 보조변수 간 상관의 크기와 동질성을 세 가지 조건으로 구분함으로써 보다 현실적인 조건에서 주성분분석이 보조변수를 어떻게 축소할 것인지, 또 축소된 주성분은 보조변수로서 적절하게 기능할 것인지 확인할 수 있었다. 보조변수 간 상관의 크기뿐만 아니라 표본의 크기, 결측치의 비율, 결측치가 존재하는 변수 Y와의 상관 크기와 같은 다양한 조건으로 세분화함으로써, 보다 다양한 상황에서 결측치를 처리하는 데에 있

어 어떤 수행 방법을 사용하는 것이 좋을지, 또 주성분 분석은 어떤 상황에서 활용하는 것이 좋은지 그 지표를 만들어내고자 했다.

둘째, 국내에서는 아직 다소 생소할 수 있는 주성분 분석을 활용한 결측치 처리 방법을 소개하고자 했다. 기존에도 다양한 연구에서 사용되고 있는 설문지법과 더불어 다양한 분야에서 종단조사 패널 자료를 수집하고 제공하고 있는 만큼, 이를 정보의 누락 없이 연구에 활용하고 정확한 결과를 얻기 위해선 자료에 이미 존재하는 결측치를 적절하게 처리하는 것이 중요하다. 본 논문은 이러한 상황에서 결측치를 처리하기 위해 연구자가 사용할 수 있는 다양한 방법을 제시하고자 했다.

본 연구의 한계는 다음과 같으며 이를 반영해 후속연구를 제안하고자 한다. 첫째, 본 연구에선 보조변수 간 상관의 크기와 동질성에 차이를 주기 위해 연구자가 임의로 상관의 크기를 정해 자료를 생성했다. 실제 연구 장면에선 보조변수 간 상관의 형태는 더욱 복잡하고 다양할 것이며 하나 혹은 그 이상의 분포를 따르고 있을 가능성이 높다. 따라서 이를 반영해 조금 더 현실적인 조건에서의 연구가 필요하다.

둘째, 본 연구에선 결측치가 존재하는 변수 Y와 보조변수 간 상관의 크기는 하나로 고정했다. 보조변수로 생성한 9개의 변수는 모두 Y에 대해 .10부터 .40까지 .10씩 증분하며 동일한 크기의 상관을 갖도록 자료를 생성했으나, 실제 연구에선 Y와 보조변수 간 상관의 크기 또한 다양할 것이다. 보조변수로 활용하기 위해선 Y와 보조변수 간 상관의 크기가 .40 이상일 것을 권장하는 선행 연구 결과를 고려했을 때, 다수의 보조변수들이 Y와 각기 다른 크기의 상관을 갖도록 자료를 생성하고 이에 따른 결측치 추정 방법들의 수행 능력을 확인해 볼 필요성이 있다.

셋째, 본 연구에선 결측치 처리에서 주성분 분석의 수행 능력을 확인하는데 중점을 뒀기에 단일 변수 X가 단일 변수 Y를 설명하는 단순한 조건을

설정했으며 결측치는 Y에만 설정했다. 하나의 개념을 다루기 위한 지표로 단일 문항을 사용하기보단 여러 문항을 종합하여 합계지표를 사용하는 것이 일반적인만큼 본 연구에서 사용한 모형은 실제 연구 장면에서 연구자가 살펴보고자 하는 모형에 비해 매우 단순한 형태다. X와 Y가 하나의 측정변수가 아닌 여러 개의 변수로 구성된 이론 변수일 경우 보조변수까지 더해졌을 때, 더욱 복잡한 형태의 모형이 될 것이며 이에 따른 상관의 크기와 동질성 또한 매우 다양해질 것이다. 따라서 X와 Y를 다수의 측정변수를 포함한 이론 변수로 설정해 결측치 처리 수행 능력을 확인해보는 것이 필요하다. 또한 실제 연구 장면에서 활용될 자료엔 연구자가 원하는 변수 Y 뿐만 아니라 Y를 설명하는 변수인 X, 혹은 보조변수에도 결측치가 존재할 가능성이 높기 때문에 이를 반영하는 후속 연구가 필요할 것이다.

마지막으로 본 연구에선 Y가 단일 변수 A1에 따라 결측치를 갖는 Linear MAR 메커니즘을 따라 결측치를 생성했다. 실제 연구 장면에서 발생하는 결측치는 단일 변수의 영향을 받기보단 다양한 변수들의 복합적인 영향을 받을 가능성이 높기 때문에 이를 반영하는 연구가 필요할 것이다. 결측치의 발생이 선형적인 형태를 따르지 않는 Non-Linear MAR 결측치 혹은 결측치가 존재하는 변수 자체에 결측치의 원인이 존재하는 MNAR 결측치를 처리하는 수행 능력 또한 평가되어야 할 것이다.

## 참 고 문 헌

- 송주원. (2014). 다중응답 문항에서 발생하는 무응답에 대한 대체 방법 비교. **Journal of The Korean Data Analysis Society**, 16(2), 691-701.
- 안혜영 (2016). Deciding the number of principal components as auxiliary variables in analyzing missing data, Graduate School, Korea University.
- 이경희, 민인식 (2016). 패널조사 응답지속성에 관한 연구: 한국노동패널조사를 중심으로, **조사연구**, 17(3), 11-24.
- 이현숙, 김수진, 전수현, & 공역. (2010). *구조방정식모형 원리와 적용*. 서울: 학지사.
- 이화정, 강석복 (2014). 한국아동 · 청소년패널조사자료에서 단위무응답의 실태 및 가중치 적용, **한국데이터정보과학회지**, 25(6), 1397-1405.
- 조민호 (2016). *빅데이터 분석을 위한 R 프로그래밍*. 서울: 정보문화사
- 조태경, & 손창균. (2015). 한국복지패널조사 자료를 이용한 패널자료의 대표성평가. **Journal of The Korean Data Analysis Society**, 17(5), 2461-2471.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology*, 545-557.
- Allison, P. D. (2008). Convergence failures in logistic regression. In *SAS Global Forum*, 1-11.
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of school psychology*, 5-37.
- Barnes, S. A., Lindborg, S. R., & Seaman Jr, J. W. (2006). Multiple

- imputation techniques in small sample clinical trials. *Statistics in medicine*, 233-245.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons. Hoboken, NJ.
- Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research*, 487-508.
- Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 330-351.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1-22.
- Duncan, T. E., Duncan, S. C., & Li, F. (1998). A comparison of model and multiple imputation based approaches to longitudinal analyses with partial missingness. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-21.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling*, 430-457.
- Enders, C. K. (2008). A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 434-448.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.
- Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural

- equation models with multiply imputed data. *Psychological methods*, 76-93.
- Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107.
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, 80-100.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention science*, 206-213.
- Hadi, A. S., & Ling, R. F. (1998). Some cautionary notes on the use of principal components regression. *The American Statistician*, 15-19.
- Hardy, S. E., Allore, H., & Studenski, S. A. (2009). Missing data: a special challenge in aging research. *Journal of the American Geriatrics Society*, 722-729.
- Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research*, 285-299.
- Ibrahim, J. G., Chen, M. H., & Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, 551-564.
- Jolliffe, I. T. (2002). *Principal component analysis*. Springer. New York.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 141-151.

- Kim, J. O., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Research*, 215-240.
- King, G., Honaker, J., Joseph, A., & Scheve, K. (1998, July). List-wise deletion is evil what to do about missing data in political science. *In Annual Meeting of the American Political Science Association, Boston.*
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, 1198-1202.
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley. New York.
- Little, R. J., D'agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., ... & Neaton, J. D. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 1355-1360.
- MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. *Structural equation modeling: Concepts, issues, and applications*, 16-36.
- Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, 158-160.
- Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: a reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, 518.
- Meng, X. L. (1994). Multiple-imputation inferences with uncongenial

- sources of input. *Statistical Science*, 538–558.
- Mustillo, S. (2012). The effects of auxiliary variables on coefficient bias and efficiency in multiple imputation. *Sociological Methods & Research*, 335–361.
- Muthén, L. K., & Muthén, B. O. (2016). *Mplus user's guide*. Mutéñ & Muthén. Los Angeles.
- Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, 328–362.
- Petroni, A., & Braglia, M. (2000). Vendor selection using principal component analysis. *Journal of supply chain management*, 63–69.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel psychology*, 537–560.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 581–592.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, 473–489.
- Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 477–497.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 147.
- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the*

*Information and Mathematical Sciences*, 153-160.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling psychology*, 1-10.

Stephenson, M. T. & Holbert, R. L. (2003). A Monte Carlo simulation of observable versus latent variable structural equation modeling techniques. *Communication Research*, 332-354.

Thoemmes, F., & Rose, N. (2014). A cautious note on auxiliary variables that can increase bias in missing data problems. *Multivariate Behavioral Research*, 4443-459.

Yuan, K. H., Yang-Wallentin, F., & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods & Research*, 598-629.

# ABSTRACT

## Application of Principal Components as Auxiliary Variable in Missing data Handling

JooYeon Yoo

Department of Psychology

The Graduate School of

Sungshin University

This paper was planned to do a simulation study in which the effectiveness of the principal component analysis were compared according to the correlation coefficients of auxiliary variables with complete information, missing value estimation methods utilizing the maximum likelihood method. Conditions formed as follows for this; Size of correlation between measurement variable and auxiliary variable, Shape of correlation between auxiliary variables, Ratio of missing data, Size of a sample, Analysis method of a model. According to these conditions, the data that satisfies 576(4×3×4×4×3) conditions are generated by using statistics program Mplus, then repeatedly implemented and analyzed for 500 times by statistics program R.

As a result of the analysis, PCA downscaled 9 auxiliary variables to 2

principal components under every condition of this study. In the case of estimates missing data using principal component gotten from PCA as an auxiliary variable, the estimation was more accurate than estimate missing data without using auxiliary variable and less errorful or biased. The estimation of missing data using the principal component as an auxiliary variable showed similar performance ability with the estimation that putting an auxiliary variable as it is. Also, the model was not converged if the auxiliary variable put as it is under the extreme condition that is a small size sample and a high ratio of missing data, but this can be solved by using PCA.

However, estimation of missing data by using PCA under the condition that the correlation between auxiliary variables is different is slightly less effective than estimation using an auxiliary variable as it is.

So this study verified the performance ability of PCA in a process of estimates missing data and suggested the index about what kind of circumstance PCA can be used usefully. Lastly, the limitation of this study and proposal about follow up study is discussed.

**Key Words:** Missing Data, Auxiliary Variable, Principal Component Analysis, Full Information Maximum Likelihood, Monte Carlo Simulation, Program R, Program Mplus

## 부 록

<부록 1> 시뮬레이션 결과 표

<부록 2> 시뮬레이션을 위한 R 코드

## <부록 1> 시뮬레이션 결과 표

<표 13> Y와 보조변수 간 상관인 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.285	0.106	0.301	0.147	0.298	0.114
		200	0.292	0.114	0.305	0.107	0.305	0.116
		400	0.289	-0.050	0.301	-0.045	0.301	-0.038
		600	0.289	-0.025	0.301	-0.032	0.301	-0.030
	25	50	0.288	0.073	0.300	0.099	0.298	0.087
		200	0.293	0.111	0.306	0.107	0.305	0.113
		400	0.288	-0.047	0.301	-0.038	0.300	-0.034
		600	0.290	-0.008	0.302	-0.016	0.301	-0.014
	50	50	0.284	0.114	0.292	0.083	0.291	0.097
		200	0.301	0.116	0.305	0.094	0.304	0.107
		400	0.294	-0.075	0.299	-0.092	0.298	-0.073
		600	0.295	0.036	0.301	0.038	0.300	0.033
	75	50	0.293	0.091	0.212	0.633	0.298	0.534
		200	0.302	0.124	0.305	0.065	0.302	0.124
		400	0.297	0.043	0.298	0.015	0.297	0.027
		600	0.294	0.030	0.295	0.058	0.294	0.040

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 14> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.283	0.137	0.298	0.188	0.294	0.150
		200	0.294	-0.098	0.306	-0.112	0.306	-0.105
		400	0.290	-0.018	0.302	-0.015	0.301	-0.009
		600	0.287	-0.057	0.299	-0.050	0.298	-0.050
	25	50	0.302	0.021	0.311	-0.039	0.312	0.027
		200	0.294	-0.029	0.306	-0.043	0.305	-0.040
		400	0.289	-0.051	0.301	-0.053	0.301	-0.046
		600	0.287	-0.059	0.300	-0.053	0.299	-0.057
	50	50	0.297	-0.059	0.302	-0.063	0.304	-0.128
		200	0.302	-0.109	0.307	-0.116	0.306	-0.119
		400	0.294	-0.079	0.299	-0.100	0.298	-0.079
		600	0.295	0.042	0.300	0.045	0.300	0.039
	75	50	0.294	0.070	0.190	1.208	0.298	0.127
		200	0.303	0.159	0.304	0.119	0.300	0.111
		400	0.296	0.017	0.297	-0.009	0.296	0.019
		600	0.295	0.013	0.294	0.046	0.293	0.024

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 15> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.285	0.102	0.301	0.141	0.296	0.112
		200	0.292	0.117	0.305	0.111	0.302	0.114
		400	0.289	-0.054	0.301	-0.048	0.299	-0.048
		600	0.289	-0.020	0.301	-0.027	0.299	-0.026
	25	50	0.288	0.070	0.299	0.094	0.297	0.076
		200	0.293	0.113	0.306	0.108	0.303	0.108
		400	0.289	-0.051	0.301	-0.043	0.298	-0.049
		600	0.287	-0.055	0.300	-0.049	0.296	-0.056
	50	50	0.285	0.109	0.293	0.079	0.287	0.067
		200	0.300	0.127	0.305	0.107	0.304	0.118
		400	0.294	-0.081	0.299	-0.099	0.298	-0.085
		600	0.295	0.038	0.300	0.042	0.299	0.033
	75	50	0.329	0.233	0.187	-0.044	0.240	0.318
		200	0.302	0.145	0.304	0.092	0.299	0.094
		400	0.296	0.030	0.298	0.002	0.296	0.029
		600	0.294	0.019	0.294	0.051	0.293	0.025

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 16> Y와 보조변수 간 상관인 .10일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	1.057	-0.093	0.998	-0.065	1.006	-0.118
		200	1.040	0.062	0.995	0.107	0.997	0.074
		400	1.043	0.007	1.000	0.004	1.000	-0.015
		600	1.041	0.018	1.000	0.042	1.000	0.025
	25	50	1.069	-0.019	1.010	0.073	1.018	-0.042
		200	1.049	0.037	0.991	0.069	0.996	0.040
		400	1.053	-0.008	1.000	-0.028	1.001	-0.035
		600	1.051	-0.037	0.998	-0.019	0.999	-0.033
	50	50	1.069	-0.062	1.012	0.083	1.024	-0.081
		200	1.041	-0.064	0.994	0.002	1.000	-0.037
		400	1.048	-0.006	1.001	0.026	1.002	-0.008
		600	1.045	-0.139	1.000	-0.148	1.001	-0.158
	75	50	0.977	-0.312	1.019	-1.310	1.013	-1.054
		200	0.968	-0.113	0.986	-0.075	0.987	-0.136
		400	0.980	-0.105	0.998	-0.088	0.996	-0.082
		600	0.972	-0.047	1.003	-0.108	1.004	-0.076

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 17> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	1.059	-0.129	1.002	-0.122	1.013	-0.165
		200	1.042	0.101	1.000	0.148	1.000	0.134
		400	1.047	0.032	1.003	0.014	1.004	0.000
		600	1.047	0.067	1.004	0.048	1.005	0.049
	25	50	1.043	0.015	1.001	0.236	1.000	-0.047
		200	1.052	-0.007	0.996	0.044	0.999	0.045
		400	1.057	0.066	1.003	0.035	1.005	0.030
		600	1.057	0.086	1.002	0.048	1.005	0.058
	50	50	1.047	0.374	1.002	0.345	0.997	0.381
		200	1.047	0.154	0.997	0.165	0.999	0.214
		400	1.048	-0.001	1.001	0.042	1.004	0.019
		600	1.046	-0.145	1.000	-0.158	1.001	-0.165
	75	50	0.976	-0.265	1.048	-2.116	1.007	-0.537
		200	0.966	-0.134	0.986	-0.087	0.988	-0.057
		400	0.981	-0.088	0.998	-0.063	0.997	-0.080
		600	0.983	-0.038	1.005	-0.100	1.004	-0.067

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 18> Y와 보조변수 간 상관인 .10일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	1.057	-0.084	0.999	-0.049	1.014	-0.110
		200	1.040	0.059	0.995	0.104	1.005	0.091
		400	1.043	0.012	1.000	0.005	1.009	0.003
		600	1.041	0.013	1.000	0.036	1.009	0.029
	25	50	1.069	-0.012	1.010	0.086	1.026	-0.015
		200	1.049	0.035	0.991	0.069	1.006	0.066
		400	1.053	-0.002	0.999	-0.023	1.012	0.000
		600	1.057	0.077	1.002	0.039	1.016	0.062
	50	50	1.069	-0.049	1.012	0.100	1.033	-0.005
		200	1.041	-0.076	0.993	-0.009	1.006	-0.007
		400	1.048	0.002	1.001	0.035	1.011	0.025
		600	1.046	-0.141	1.000	-0.152	1.009	-0.161
	75	50	1.059	-0.239	1.052	1.436	0.989	-0.522
		200	0.967	-0.132	0.986	-0.095	0.986	-0.076
		400	0.980	-0.097	0.998	-0.079	0.995	-0.090
		600	0.984	-0.044	1.005	-0.102	1.002	-0.062

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 19> Y와 보조변수 간 상관인 .10일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.939	0.225	0.988	-0.173	0.952	-0.179
		200	0.977	0.109	0.985	-0.032	0.982	-0.028
		400	0.987	0.075	0.992	0.020	0.991	0.023
		600	0.990	0.060	0.994	-0.006	0.994	-0.001
	25	50	0.935	0.240	1.005	-0.250	0.948	-0.195
		200	0.977	0.120	0.986	-0.002	0.982	0.005
		400	0.986	0.081	0.992	0.043	0.991	0.043
		600	0.989	0.065	0.994	-0.045	0.993	-0.041
	50	50	0.912	0.287	1.106	-0.248	0.926	-0.074
		200	0.973	0.145	0.984	-0.019	0.976	-0.024
		400	0.984	0.095	0.988	-0.059	0.986	-0.055
		600	0.990	0.079	0.993	-0.073	0.992	-0.070
	75	50	0.793	0.463	3.302	3.192	1.024	0.282
		200	0.947	0.219	1.009	-0.053	0.953	-0.046
		400	0.976	0.149	0.988	0.058	0.977	0.062
		600	0.981	0.130	0.988	-0.086	0.983	-0.091

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 20> Y와 보조변수 간 상관인 .10일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.940	0.224	0.987	-0.201	0.952	-0.200
		200	0.991	0.103	0.999	0.158	0.996	0.162
		400	0.994	0.073	1.000	0.127	0.998	0.131
		600	0.992	0.062	0.997	-0.003	0.996	-0.002
	25	50	0.959	0.221	1.008	0.368	0.968	0.245
		200	0.990	0.112	1.000	0.108	0.995	0.113
		400	0.993	0.079	0.999	0.133	0.997	0.134
		600	0.992	0.067	0.998	0.034	0.996	0.030
	50	50	0.941	0.261	1.114	-0.083	0.955	0.164
		200	0.990	0.137	1.000	-0.049	0.994	-0.044
		400	0.984	0.096	0.988	-0.056	0.986	-0.055
		600	0.990	0.080	0.993	-0.074	0.992	-0.070
	75	50	0.789	0.468	3.278	11.788	1.036	0.127
		200	0.946	0.219	1.007	-0.101	0.952	-0.083
		400	0.976	0.149	0.989	0.074	0.978	0.084
		600	0.985	0.124	0.989	-0.095	0.985	-0.097

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 21> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.939	0.225	0.988	-0.171	0.951	-0.178
		200	0.977	0.109	0.985	-0.035	0.981	-0.032
		400	0.987	0.075	0.992	0.023	0.990	0.023
		600	0.990	0.060	0.995	-0.008	0.992	-0.004
	25	50	0.935	0.240	1.004	-0.249	0.949	-0.199
		200	0.977	0.120	0.986	-0.005	0.981	0.000
		400	0.986	0.081	0.992	0.045	0.989	0.042
		600	0.992	0.067	0.998	0.027	0.995	0.021
	50	50	0.912	0.287	1.107	-0.251	0.927	-0.080
		200	0.973	0.145	0.985	-0.027	0.975	-0.035
		400	0.984	0.095	0.988	-0.057	0.986	-0.053
		600	0.990	0.080	0.993	-0.072	0.991	-0.070
	75	50	0.910	0.542	1.246	2.579	1.703	-0.304
		200	0.947	0.220	1.008	-0.075	0.952	-0.057
		400	0.976	0.149	0.989	0.066	0.977	0.073
		600	0.985	0.123	0.990	-0.097	0.985	-0.106

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 22> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.273	0.121	0.298	0.173	0.298	0.133
		200	0.281	0.123	0.307	0.130	0.306	0.128
		400	0.277	-0.042	0.302	-0.035	0.302	-0.028
		600	0.279	-0.034	0.302	-0.049	0.302	-0.044
	25	50	0.276	0.088	0.302	0.116	0.298	0.100
		200	0.281	0.124	0.305	0.118	0.304	0.122
		400	0.277	-0.038	0.301	-0.028	0.300	-0.024
		600	0.278	-0.007	0.301	-0.020	0.301	-0.018
	50	50	0.280	0.111	0.291	0.087	0.291	0.076
		200	0.298	0.076	0.306	0.030	0.305	0.048
		400	0.289	-0.062	0.298	-0.075	0.298	-0.057
		600	0.290	0.002	0.300	-0.002	0.299	-0.006
	75	50	0.293	0.101	0.203	1.191	0.302	0.565
		200	0.301	0.131	0.303	0.072	0.300	0.132
		400	0.295	0.035	0.295	0.020	0.294	0.029
		600	0.295	0.017	0.295	0.043	0.294	0.022

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 23> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.271	0.136	0.294	0.203	0.294	0.155
		200	0.282	-0.117	0.306	-0.115	0.305	-0.111
		400	0.277	-0.025	0.301	-0.023	0.301	-0.011
		600	0.275	-0.060	0.299	-0.054	0.298	-0.049
	25	50	0.292	0.042	0.313	-0.020	0.314	0.051
		200	0.284	-0.052	0.308	-0.070	0.307	-0.067
		400	0.278	-0.022	0.302	-0.029	0.301	-0.012
		600	0.276	-0.079	0.300	-0.074	0.299	-0.076
	50	50	0.289	-0.018	0.307	-0.081	0.304	-0.086
		200	0.297	-0.096	0.307	-0.109	0.306	-0.116
		400	0.288	-0.076	0.298	-0.096	0.298	-0.077
		600	0.290	0.009	0.299	0.010	0.299	0.003
	75	50	0.293	0.076	0.196	2.467	0.301	0.176
		200	0.302	0.171	0.301	0.123	0.298	0.113
		400	0.295	0.007	0.294	-0.012	0.293	0.019
		600	0.295	0.011	0.294	0.042	0.293	0.024

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 24> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.273	0.117	0.300	0.163	0.294	0.132
		200	0.281	0.127	0.306	0.134	0.301	0.128
		400	0.277	-0.047	0.301	-0.040	0.296	-0.043
		600	0.278	-0.031	0.302	-0.043	0.297	-0.040
	25	50	0.276	0.085	0.301	0.109	0.294	0.092
		200	0.280	0.112	0.304	0.118	0.299	0.115
		400	0.277	-0.043	0.301	-0.036	0.296	-0.045
		600	0.275	-0.069	0.300	-0.064	0.294	-0.070
	50	50	0.280	0.107	0.291	0.083	0.287	0.051
		200	0.298	0.086	0.305	0.043	0.304	0.064
		400	0.288	-0.076	0.299	-0.095	0.296	-0.084
		600	0.290	0.002	0.300	0.003	0.297	-0.006
	75	50	0.322	0.246	0.196	1.662	0.299	0.446
		200	0.301	0.151	0.302	0.093	0.298	0.101
		400	0.295	0.022	0.295	0.007	0.293	0.031
		600	0.294	0.017	0.295	0.048	0.293	0.019

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 25> Y와 보조변수 간 상관인 .20일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	1.107	-0.145	1.009	-0.163	1.010	-0.184
		200	1.082	0.040	0.992	0.066	0.994	0.057
		400	1.087	0.003	0.999	-0.010	1.000	-0.026
		600	1.081	0.042	0.999	0.081	0.999	0.057
	25	50	1.123	-0.075	1.009	0.003	1.017	-0.084
		200	1.105	0.038	0.995	0.078	0.999	0.060
		400	1.106	-0.001	0.999	-0.023	1.000	-0.025
		600	1.105	-0.029	0.998	-0.005	0.999	-0.026
	50	50	1.113	-0.064	1.015	0.005	1.023	-0.081
		200	1.083	-0.022	0.995	0.114	0.999	0.065
		400	1.092	-0.021	1.003	0.017	1.004	-0.014
		600	1.091	-0.087	1.001	-0.076	1.002	-0.095
	75	50	0.948	-0.297	1.030	-2.191	1.005	-1.047
		200	0.950	-0.102	0.988	-0.045	0.990	-0.109
		400	0.966	-0.095	1.002	-0.070	1.000	-0.061
		600	0.964	-0.023	1.004	-0.087	1.003	-0.053

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 26> Y와 보조변수 간 상관인 .20일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	1.113	-0.145	1.014	-0.195	1.022	-0.202
		200	1.084	0.128	0.999	0.117	1.001	0.116
		400	1.091	0.030	1.004	0.010	1.005	-0.023
		600	1.088	0.072	1.005	0.065	1.006	0.051
	25	50	1.092	-0.013	0.999	0.187	0.999	-0.100
		200	1.101	0.035	0.993	0.080	0.998	0.076
		400	1.107	0.030	1.002	0.016	1.004	-0.024
		600	1.108	0.108	1.002	0.076	1.005	0.073
	50	50	1.101	0.215	0.999	0.271	1.005	0.202
		200	1.094	0.114	0.996	0.137	1.000	0.183
		400	1.094	0.005	1.003	0.054	1.005	0.032
		600	1.092	-0.094	1.002	-0.091	1.003	-0.105
	75	50	0.949	-0.241	1.057	-2.858	0.999	-0.362
		200	0.949	-0.127	0.991	-0.025	0.993	-0.007
		400	0.966	-0.076	1.003	-0.031	1.002	-0.055
		600	0.964	-0.017	1.006	-0.088	1.004	-0.059

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 27> Y와 보조변수 간 상관인 .20일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	1.104	-0.133	1.007	-0.140	1.029	-0.180
		200	1.082	0.030	0.993	0.057	1.013	0.067
		400	1.086	0.003	0.999	-0.010	1.019	-0.007
		600	1.081	0.037	0.999	0.072	1.017	0.057
	25	50	1.123	-0.068	1.010	0.021	1.038	-0.065
		200	1.104	0.044	0.996	0.069	1.021	0.071
		400	1.106	0.005	0.999	-0.015	1.023	0.014
		600	1.107	0.092	1.001	0.060	1.026	0.071
	50	50	1.113	-0.051	1.016	0.026	1.040	-0.017
		200	1.083	-0.033	0.995	0.104	1.014	0.090
		400	1.094	0.005	1.002	0.044	1.021	0.034
		600	1.092	-0.087	1.002	-0.083	1.019	-0.098
	75	50	0.818	-0.150	0.952	-1.712	0.943	-0.224
		200	0.950	-0.120	0.989	-0.058	0.985	-0.051
		400	0.966	-0.087	1.002	-0.062	0.997	-0.072
		600	0.964	-0.023	1.004	-0.091	0.998	-0.044

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 28> Y와 보조변수 간 상관인 .20일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.923	-0.159	0.978	-0.167	0.946	-0.160
		200	0.966	-0.019	0.985	-0.035	0.982	-0.028
		400	0.974	0.026	0.991	0.031	0.990	0.033
		600	0.978	0.017	0.994	0.005	0.993	0.015
	25	50	0.922	-0.181	0.996	-0.260	0.944	-0.177
		200	0.965	0.006	0.986	-0.010	0.982	0.001
		400	0.976	0.033	0.993	0.042	0.991	0.036
		600	0.979	-0.017	0.996	-0.016	0.994	-0.006
	50	50	0.912	-0.060	1.101	-0.247	0.932	-0.046
		200	0.967	-0.009	0.983	-0.024	0.974	-0.023
		400	0.979	-0.061	0.988	-0.061	0.986	-0.057
		600	0.985	-0.059	0.992	-0.060	0.991	-0.054
	75	50	0.796	0.095	3.209	6.693	1.006	0.287
		200	0.948	-0.048	1.004	-0.072	0.955	-0.062
		400	0.975	0.064	0.988	0.045	0.978	0.059
		600	0.983	-0.103	0.987	-0.089	0.984	-0.101

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 29> Y와 보조변수 간 상관인 .20일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.922	-0.175	0.975	-0.203	0.945	-0.176
		200	0.979	0.161	0.999	0.160	0.996	0.164
		400	0.981	0.145	0.999	0.154	0.998	0.167
		600	0.980	0.019	0.997	0.018	0.996	0.025
	25	50	0.950	0.260	1.002	0.245	0.968	0.232
		200	0.977	0.172	0.999	0.162	0.994	0.166
		400	0.979	0.153	0.998	0.165	0.996	0.174
		600	0.978	0.012	0.996	0.015	0.995	0.015
	50	50	0.944	0.132	1.047	-0.068	0.967	0.095
		200	0.985	-0.053	1.000	-0.072	0.994	-0.068
		400	0.979	-0.066	0.988	-0.055	0.986	-0.063
		600	0.985	-0.061	0.992	-0.056	0.992	-0.051
	75	50	0.793	0.134	2.760	14.381	1.032	0.053
		200	0.947	-0.088	1.002	-0.145	0.957	-0.122
		400	0.975	0.086	0.989	0.063	0.979	0.079
		600	0.983	-0.100	0.987	-0.072	0.984	-0.075

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 30> Y와 보조변수 간 상관인 .20일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.923	-0.155	0.977	-0.167	0.941	-0.159
		200	0.966	-0.022	0.985	-0.038	0.977	-0.034
		400	0.974	0.026	0.991	0.033	0.984	0.029
		600	0.978	0.017	0.994	0.004	0.987	0.012
	25	50	0.921	-0.180	0.995	-0.261	0.941	-0.184
		200	0.965	0.015	0.985	0.002	0.976	0.004
		400	0.976	0.034	0.993	0.047	0.986	0.033
		600	0.978	0.014	0.997	0.022	0.989	0.017
	50	50	0.912	-0.061	1.101	-0.250	0.932	-0.063
		200	0.967	-0.016	0.983	-0.030	0.973	-0.032
		400	0.979	-0.066	0.988	-0.065	0.984	-0.065
		600	0.985	-0.058	0.992	-0.056	0.989	-0.054
	75	50	0.687	0.766	1.298	3.391	1.814	-0.130
		200	0.948	-0.067	1.003	-0.095	0.954	-0.078
		400	0.975	0.073	0.988	0.053	0.977	0.069
		600	0.983	-0.103	0.987	-0.082	0.984	-0.101

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 31> Y와 보조변수 간 상관인 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.262	0.122	0.297	0.180	0.298	0.126
		200	0.268	0.133	0.306	0.153	0.306	0.142
		400	0.265	-0.033	0.302	-0.024	0.302	-0.021
		600	0.268	-0.032	0.303	-0.046	0.303	-0.041
	25	50	0.273	0.071	0.301	0.100	0.300	0.093
		200	0.269	0.128	0.304	0.114	0.304	0.110
		400	0.266	-0.045	0.301	-0.038	0.301	-0.035
		600	0.266	-0.017	0.301	-0.030	0.301	-0.028
	50	50	0.276	0.121	0.293	0.083	0.291	0.068
		200	0.292	0.061	0.306	0.011	0.305	0.029
		400	0.285	-0.070	0.298	-0.074	0.298	-0.059
		600	0.287	0.017	0.299	0.007	0.299	0.006
	75	50	0.298	-0.030	0.205	0.422	0.313	0.396
		200	0.300	0.140	0.300	0.083	0.298	0.139
		400	0.294	0.055	0.294	0.038	0.294	0.052
		600	0.293	0.015	0.294	0.030	0.294	0.013

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 32> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.259	0.151	0.293	0.203	0.293	0.167
		200	0.271	-0.100	0.307	-0.083	0.306	-0.087
		400	0.266	-0.020	0.302	-0.025	0.301	-0.012
		600	0.265	-0.052	0.300	-0.054	0.299	-0.047
	25	50	0.277	0.093	0.311	0.024	0.310	0.059
		200	0.271	-0.050	0.308	-0.084	0.307	-0.081
		400	0.267	0.002	0.302	-0.034	0.302	-0.011
		600	0.264	-0.054	0.300	-0.067	0.300	-0.063
	50	50	0.284	-0.167	0.303	-0.163	0.304	-0.157
		200	0.291	-0.096	0.307	-0.112	0.305	-0.120
		400	0.284	-0.082	0.298	-0.071	0.297	-0.065
		600	0.286	0.030	0.299	0.018	0.299	0.015
	75	50	0.295	-0.083	0.191	0.711	0.310	0.104
		200	0.300	0.192	0.298	0.132	0.297	0.121
		400	0.292	0.024	0.294	-0.005	0.294	0.012
		600	0.293	0.015	0.294	0.037	0.294	0.027

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 33> Y와 보조변수 간 상관인 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.262	0.110	0.296	0.164	0.290	0.123
		200	0.268	0.135	0.306	0.152	0.298	0.140
		400	0.265	-0.033	0.302	-0.025	0.294	-0.031
		600	0.278	-0.031	0.302	-0.043	0.297	-0.040
	25	50	0.273	0.063	0.297	0.064	0.291	0.011
		200	0.269	0.128	0.304	0.113	0.297	0.109
		400	0.266	-0.048	0.301	-0.048	0.293	-0.054
		600	0.264	-0.054	0.300	-0.064	0.291	-0.069
	50	50	0.277	0.116	0.294	0.077	0.286	0.047
		200	0.292	0.071	0.305	0.025	0.303	0.044
		400	0.285	-0.078	0.298	-0.080	0.295	-0.076
		600	0.287	0.018	0.299	0.015	0.296	0.007
	75	50	0.288	0.280	0.174	1.201	0.236	0.558
		200	0.301	0.159	0.299	0.097	0.297	0.110
		400	0.293	0.053	0.294	0.043	0.293	0.062
		600	0.293	0.015	0.294	0.035	0.293	0.006

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 34> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	1.162	-0.078	1.017	0.016	1.014	-0.104
		200	1.130	0.008	0.994	0.010	0.996	0.024
		400	1.132	-0.020	1.000	-0.043	1.000	-0.046
		600	1.121	0.029	0.997	0.061	0.998	0.042
	25	50	1.171	-0.042	1.007	0.042	1.014	-0.056
		200	1.158	0.026	0.998	0.080	1.000	0.080
		400	1.160	-0.013	1.000	-0.035	1.001	-0.032
		600	1.159	-0.001	0.999	0.027	1.000	0.003
	50	50	1.157	-0.112	1.015	-0.062	1.024	-0.121
		200	1.130	0.028	0.997	0.166	1.000	0.132
		400	1.136	-0.029	1.004	0.013	1.004	-0.014
		600	1.134	-0.122	1.003	-0.104	1.003	-0.120
	75	50	0.895	-0.246	1.040	-1.577	1.001	-0.975
		200	0.921	-0.081	0.992	-0.008	0.996	-0.072
		400	0.941	-0.090	1.002	-0.047	1.001	-0.045
		600	0.942	-0.019	1.005	-0.082	1.005	-0.046

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 35> Y와 보조변수 간 상관인 .30일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	1.164	-0.107	1.019	-0.079	1.027	-0.188
		200	1.124	0.115	0.996	0.051	0.999	0.070
		400	1.132	0.007	1.003	0.003	1.004	-0.027
		600	1.127	0.061	1.003	0.073	1.005	0.049
	25	50	1.158	-0.030	0.995	0.109	1.006	-0.026
		200	1.156	0.009	0.995	0.071	0.999	0.074
		400	1.159	-0.005	1.002	0.014	1.004	-0.026
		600	1.159	0.088	1.003	0.099	1.005	0.072
	50	50	1.148	0.503	1.010	0.404	1.011	0.369
		200	1.142	0.113	0.997	0.136	1.001	0.160
		400	1.136	-0.013	1.004	0.039	1.006	0.046
		600	1.135	-0.135	1.003	-0.093	1.004	-0.096
	75	50	0.901	-0.175	1.062	-1.090	1.001	-0.174
		200	0.921	-0.115	0.999	0.033	1.001	0.039
		400	0.941	-0.092	1.004	-0.034	1.003	-0.041
		600	0.945	-0.008	1.007	-0.071	1.005	-0.055

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 36> Y와 보조변수 간 상관인 .30일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	1.161	-0.061	1.018	0.044	1.044	-0.112
		200	1.130	0.006	0.995	0.015	1.024	0.042
		400	1.132	-0.021	1.000	-0.045	1.029	-0.034
		600	1.081	0.037	0.999	0.072	1.017	0.057
	25	50	1.169	-0.006	1.013	0.034	1.046	0.010
		200	1.158	0.026	0.999	0.085	1.033	0.095
		400	1.160	-0.009	1.000	-0.023	1.036	0.003
		600	1.159	0.088	1.002	0.099	1.039	0.103
	50	50	1.156	-0.099	1.015	-0.035	1.051	-0.066
		200	1.130	0.018	0.998	0.156	1.024	0.165
		400	1.136	-0.019	1.004	0.025	1.029	0.023
		600	1.135	-0.123	1.003	-0.111	1.029	-0.121
	75	50	0.730	-0.062	0.902	-0.760	0.994	-0.075
		200	0.923	-0.100	0.992	-0.015	0.985	-0.020
		400	0.940	-0.112	1.003	-0.087	0.992	-0.087
		600	0.942	-0.018	1.006	-0.084	0.996	-0.031

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 37> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.899	-0.241	0.970	-0.230	0.941	-0.253
		200	0.945	-0.004	0.984	-0.016	0.981	-0.017
		400	0.953	0.040	0.989	0.061	0.988	0.057
		600	0.959	0.021	0.994	0.007	0.993	0.020
	25	50	0.905	-0.178	0.997	-0.298	0.941	-0.181
		200	0.945	0.019	0.983	-0.002	0.980	0.005
		400	0.956	0.032	0.991	0.058	0.990	0.044
		600	0.960	0.001	0.995	-0.002	0.994	0.015
	50	50	0.901	-0.053	1.069	-0.224	0.930	-0.033
		200	0.959	0.072	0.981	0.019	0.975	0.034
		400	0.972	-0.043	0.987	-0.029	0.985	-0.029
		600	0.976	-0.049	0.991	-0.048	0.990	-0.044
	75	50	0.781	0.152	2.474	1.634	1.007	0.408
		200	0.946	-0.052	1.003	-0.112	0.960	-0.087
		400	0.972	0.082	0.988	0.045	0.979	0.068
		600	0.985	-0.093	0.990	-0.073	0.987	-0.095

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 38> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.901	-0.258	0.964	-0.337	0.943	-0.262
		200	0.959	0.141	0.996	0.175	0.994	0.159
		400	0.961	0.158	0.998	0.164	0.997	0.176
		600	0.961	0.023	0.996	0.018	0.996	0.033
	25	50	0.931	0.132	0.991	0.001	0.968	0.063
		200	0.956	0.193	0.996	0.177	0.994	0.177
		400	0.961	0.142	0.998	0.177	0.997	0.174
		600	0.960	-0.001	0.996	-0.006	0.995	0.007
	50	50	0.930	0.146	1.025	-0.157	0.963	-0.082
		200	0.972	0.017	0.996	-0.017	0.992	0.001
		400	0.972	-0.037	0.988	0.006	0.986	-0.008
		600	0.976	-0.054	0.991	-0.039	0.991	-0.041
	75	50	0.780	0.191	1.882	1.575	1.027	0.986
		200	0.945	-0.109	0.999	-0.200	0.968	-0.167
		400	0.974	0.118	0.993	0.123	0.985	0.135
		600	0.984	-0.105	0.991	-0.045	0.989	-0.042

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 39> Y와 보조변수 간 상관인 .30일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.899	-0.231	0.969	-0.216	0.930	-0.240
		200	0.945	-0.006	0.984	-0.021	0.968	-0.026
		400	0.953	0.043	0.989	0.068	0.975	0.055
		600	0.978	0.017	0.994	0.004	0.987	0.012
	25	50	0.904	-0.141	0.996	-0.235	0.933	-0.119
		200	0.945	0.019	0.982	-0.003	0.968	0.000
		400	0.956	0.033	0.991	0.064	0.977	0.040
		600	0.960	-0.001	0.997	-0.009	0.982	-0.009
	50	50	0.901	-0.054	1.065	-0.224	0.927	-0.063
		200	0.959	0.065	0.981	0.014	0.970	0.028
		400	0.972	-0.039	0.987	-0.017	0.981	-0.023
		600	0.976	-0.048	0.991	-0.042	0.985	-0.046
	75	50	0.698	0.754	1.072	2.112	1.591	0.185
		200	0.946	-0.072	1.001	-0.136	0.959	-0.115
		400	0.974	0.092	0.990	0.055	0.980	0.078
		600	0.985	-0.092	0.991	-0.060	0.987	-0.084

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 40> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.249	0.117	0.295	0.151	0.298	0.113
		200	0.256	0.140	0.305	0.156	0.305	0.141
		400	0.253	-0.025	0.301	-0.023	0.301	-0.024
		600	0.255	-0.039	0.302	-0.049	0.302	-0.045
	25	50	0.256	0.088	0.299	0.128	0.300	0.099
		200	0.259	0.116	0.305	0.107	0.305	0.099
		400	0.255	-0.057 28	0.300	-0.060 76	0.300	-0.061 77
		600	0.255	-0.027	0.301	-0.037	0.301	-0.037
	50	50	0.276	0.128	0.291	0.087	0.293	0.063
		200	0.288	0.070	0.304	0.024	0.304	0.037
		400	0.280	-0.051	0.298	-0.057	0.298	-0.044
		600	0.282	0.018	0.299	-0.007	0.299	-0.007
	75	50	0.303	-0.032	0.211	0.360	0.336	0.245
		200	0.300	0.179	0.298	0.089	0.298	0.146
		400	0.292	0.072	0.294	0.052	0.294	0.061
		600	0.292	0.014	0.294	0.003	0.293	-0.007

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 41> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.246	0.125	0.289	0.131	0.289	0.127
		200	0.261	-0.086	0.308	-0.082	0.307	-0.083
		400	0.255	-0.044	0.302	-0.029	0.302	-0.028
		600	0.254	-0.046	0.300	-0.064	0.300	-0.060
	25	50	0.302	0.021	0.311	-0.039	0.312	0.027
		200	0.294	-0.029	0.306	-0.043	0.305	-0.040
		400	0.256	-0.006	0.302	-0.038	0.303	-0.033
		600	0.252	-0.048	0.300	-0.068	0.300	-0.066
	50	50	0.278	-0.161	0.313	-0.006	0.310	0.010
		200	0.285	-0.029	0.306	-0.092	0.307	-0.095
		400	0.279	-0.096	0.298	-0.063	0.298	-0.065
		600	0.281	0.053	0.299	0.022	0.299	0.024
	75	50	0.313	-0.188	0.169	0.107	0.303	0.124
		200	0.299	0.269	0.300	0.135	0.300	0.130
		400	0.290	-0.060	0.297	-0.064	0.297	-0.065
		600	0.291	0.016	0.298	0.042	0.297	0.045

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 42> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.251	0.102	0.295	0.133	0.251	0.101
		200	0.254	0.137	0.304	0.150	0.294	0.140
		400	0.253	-0.024	0.301	-0.029	0.290	-0.036
		600	0.255	-0.035	0.302	-0.038	0.292	-0.038
	25	50	0.256	0.081	0.298	0.116	0.289	0.092
		200	0.258	0.115	0.304	0.106	0.295	0.100
		400	0.254	-0.054	0.300	-0.066	0.290	-0.072
		600	0.252	-0.051	0.299	-0.068	0.288	-0.072
	50	50	0.277	0.123	0.291	0.081	0.288	0.048
		200	0.287	0.080	0.303	0.043	0.301	0.058
		400	0.280	-0.062	0.297	-0.065	0.295	-0.066
		600	0.282	0.020	0.299	0.001	0.295	-0.004
	75	50	0.322	0.246	0.142	0.328	0.271	0.504
		200	0.300	0.198	0.298	0.104	0.297	0.125
		400	0.292	0.052	0.293	0.027	0.293	0.051
		600	0.291	0.014	0.294	0.010	0.292	-0.013

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 43> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	1.212	-0.075	1.019	0.025	1.014	-0.093
		200	1.174	-0.023	0.305	-0.008	0.998	0.022
		400	1.175	-0.039	1.001	-0.063	1.001	-0.054
		600	1.164	0.044	0.999	0.066	0.999	0.050
	25	50	1.223	-0.035	1.014	0.146	0.996	-0.048
		200	1.212	0.018	0.998	0.079	0.999	0.096
		400	1.216	-0.008	1.002	-0.004	1.002	0.016
		600	1.213	0.015	1.000	0.042	1.001	0.027
	50	50	1.201	-0.160	1.026	-0.136	1.025	-0.149
		200	1.175	-0.012	0.999	0.156	1.000	0.138
		400	1.179	-0.052	1.004	0.002	1.004	-0.013
		600	1.179	-0.146	1.003	-0.099	1.003	-0.103
	75	50	0.837	-0.188	1.036	-1.299	0.986	-0.618
		200	0.896	-0.136	0.995	0.035	0.998	-0.047
		400	0.920	-0.133	1.004	-0.075	1.003	-0.060
		600	0.928	-0.005	1.007	-0.061	1.006	-0.035

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 44> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	1.219	-0.061	1.020	-0.144	1.032	-0.141
		200	1.166	0.109	0.995	0.068	0.998	0.069
		400	1.170	0.046	1.002	0.020	1.003	0.019
		600	1.167	0.060	1.003	0.083	1.004	0.072
	25	50	1.043	0.015	1.001	0.236	1.000	-0.047
		200	1.052	-0.007	0.996	0.044	0.999	0.045
		400	1.212	0.038	1.002	0.044	1.003	0.034
		600	1.212	0.069	1.004	0.090	1.004	0.077
	50	50	1.192	0.460	0.990	0.116	1.005	0.064
		200	1.191	0.081	0.997	0.098	1.000	0.097
		400	1.179	0.012	1.003	0.034	1.005	0.058
		600	1.181	-0.180	1.003	-0.044	1.004	-0.023
	75	50	0.825	0.000	1.074	0.021	0.996	-0.184
		200	0.897	-0.199	1.002	0.055	1.000	0.053
		400	0.921	-0.020	1.004	0.042	1.003	0.036
		600	0.930	-0.003	1.004	-0.051	1.003	-0.055

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 45> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	1.215	-0.058	1.018	0.053	1.214	-0.057
		200	1.179	-0.013	0.997	0.008	1.037	0.041
		400	1.176	-0.030	1.001	-0.044	1.040	-0.031
		600	1.164	0.040	0.999	0.052	1.035	0.045
	25	50	1.223	-0.022	1.015	0.177	1.057	-0.038
		200	1.212	0.019	0.999	0.087	1.043	0.107
		400	1.215	-0.005	1.002	0.002	1.049	0.029
		600	1.212	0.068	1.004	0.084	1.052	0.093
	50	50	1.200	-0.146	1.026	-0.102	1.062	-0.101
		200	1.175	-0.023	1.000	0.144	1.033	0.162
		400	1.179	-0.038	1.004	0.015	1.037	0.021
		600	1.180	-0.147	1.003	-0.103	1.037	-0.106
	75	50	0.655	0.014	0.879	0.048	0.922	0.045
		200	0.896	-0.154	0.996	0.024	0.983	0.001
		400	0.920	-0.118	1.004	-0.056	0.990	-0.066
		600	0.929	-0.004	1.007	-0.061	0.994	-0.018

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 46> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Uniform	15	50	0.875	-0.230	0.967	-0.307	0.943	-0.268
		200	0.917	0.012	0.983	-0.006	0.981	-0.019
		400	0.926	0.054	0.989	0.097	0.989	0.082
		600	0.932	0.029	0.994	0.015	0.993	0.031
	25	50	0.874	-0.249	0.972	-0.294	0.940	-0.260
		200	0.919	0.018	0.983	-0.017	0.981	-0.020
		400	0.930	-0.002	0.990	0.029	0.989	0.003
		600	0.932	0.002	0.994	-0.006	0.993	0.008
	50	50	0.889	-0.046	1.043	-0.187	0.930	-0.049
		200	0.949	0.083	0.982	0.006	0.977	0.019
		400	0.964	-0.029	0.988	0.001	0.987	-0.004
		600	0.966	-0.054	0.990	-0.047	0.990	-0.049
	75	50	0.777	0.156	1.979	1.718	1.028	0.458
		200	0.945	-0.029	0.998	-0.142	0.966	-0.100
		400	0.971	0.095	0.988	0.047	0.982	0.089
		600	0.983	-0.120	0.990	-0.102	0.987	-0.113

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 47> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Mod	15	50	0.876	-0.228	0.957	-0.253	0.949	-0.222
		200	0.933	0.134	0.995	0.132	0.995	0.127
		400	0.936	0.120	0.998	0.156	0.999	0.158
		600	0.936	0.014	0.997	0.015	0.997	0.021
	25	50	0.959	0.250	1.008	0.368	0.968	0.245
		200	0.990	0.118	1.000	0.108	0.995	0.113
		400	0.933	0.081	0.998	0.145	0.998	0.141
		600	0.931	0.005	0.996	0.004	0.996	0.010
	50	50	0.916	0.159	0.985	-0.101	0.968	-0.043
		200	0.961	-0.025	0.995	0.048	0.993	0.063
		400	0.965	-0.012	0.991	0.034	0.992	0.036
		600	0.967	-0.064	0.994	-0.007	0.994	-0.004
	75	50	0.775	0.228	0.920	-0.011	0.965	0.100
		200	0.944	-0.178	0.987	-0.134	0.983	-0.126
		400	0.973	0.180	0.991	0.158	0.991	0.163
		600	0.984	-0.109	0.993	0.015	0.993	0.025

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 48> Y와 보조변수 간 상관인 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 모수 편향(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	P.Bias	Est	P.Bias	Est	P.Bias
Low	15	50	0.874	-0.218	0.965	-0.296	0.874	-0.218
		200	0.917	-0.007	0.983	-0.019	0.958	-0.034
		400	0.926	0.043	0.989	0.081	0.965	0.066
		600	0.932	0.028	0.994	0.015	0.970	0.020
	25	50	0.874	-0.245	0.969	-0.283	0.923	-0.277
		200	0.919	0.017	0.983	-0.019	0.960	-0.020
		400	0.929	-0.002	0.990	0.032	0.966	0.004
		600	0.931	0.003	0.996	-0.006	0.970	-0.009
	50	50	0.889	-0.048	1.030	-0.181	0.925	-0.094
		200	0.949	0.076	0.982	-0.002	0.969	0.020
		400	0.964	-0.024	0.988	0.026	0.980	0.003
		600	0.966	-0.053	0.991	-0.036	0.982	-0.053
	75	50	0.669	0.784	1.027	1.265	1.332	0.308
		200	0.945	-0.054	0.998	-0.175	0.963	-0.158
		400	0.972	0.110	0.990	0.097	0.982	0.103
		600	0.984	-0.119	0.990	-0.078	0.988	-0.088

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치, P.Bias=모수 편향(Parameter bias)

<표 49> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.285	0.173	0.301	0.185	0.298	0.174
		200	0.292	0.079	0.305	0.081	0.305	0.080
		400	0.289	0.060	0.301	0.059	0.301	0.059
		600	0.289	0.048	0.301	0.048	0.301	0.048
	25	50	0.288	0.189	0.300	0.202	0.298	0.189
		200	0.293	0.087	0.306	0.089	0.305	0.089
		400	0.288	0.066	0.301	0.066	0.300	0.065
		600	0.290	0.053	0.302	0.052	0.301	0.052
	50	50	0.284	0.232	0.292	0.263	0.291	0.230
		200	0.301	0.103	0.305	0.107	0.304	0.104
		400	0.294	0.079	0.299	0.080	0.298	0.079
		600	0.295	0.064	0.301	0.064	0.300	0.064
	75	50	0.293	0.391	0.212	0.682	0.298	0.458
		200	0.302	0.152	0.305	0.164	0.302	0.153
		400	0.297	0.105	0.298	0.110	0.297	0.105
		600	0.294	0.092	0.295	0.094	0.294	0.091

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 50> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.283	0.173	0.298	0.183	0.294	0.173
		200	0.294	0.080	0.306	0.080	0.306	0.080
		400	0.290	0.061	0.302	0.060	0.301	0.059
		600	0.287	0.050	0.299	0.049	0.298	0.049
	25	50	0.302	0.182	0.311	0.192	0.312	0.184
		200	0.294	0.083	0.306	0.083	0.305	0.082
		400	0.289	0.063	0.301	0.062	0.301	0.061
		600	0.287	0.053	0.300	0.052	0.299	0.051
	50	50	0.297	0.234	0.302	0.263	0.304	0.236
		200	0.302	0.099	0.307	0.099	0.306	0.097
		400	0.294	0.079	0.299	0.079	0.298	0.078
		600	0.295	0.064	0.300	0.063	0.300	0.062
	75	50	0.294	0.394	0.190	0.686	0.298	0.445
		200	0.303	0.153	0.304	0.164	0.300	0.153
		400	0.296	0.103	0.297	0.107	0.296	0.102
		600	0.295	0.088	0.294	0.089	0.293	0.086

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 51> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.285	0.173	0.301	0.184	0.296	0.174
		200	0.292	0.079	0.305	0.080	0.302	0.079
		400	0.289	0.060	0.301	0.059	0.299	0.059
		600	0.289	0.049	0.301	0.048	0.299	0.048
	25	50	0.288	0.188	0.299	0.201	0.297	0.188
		200	0.293	0.087	0.306	0.089	0.303	0.088
		400	0.289	0.066	0.301	0.065	0.298	0.065
		600	0.287	0.053	0.300	0.052	0.296	0.052
	50	50	0.285	0.231	0.293	0.262	0.287	0.232
		200	0.300	0.103	0.305	0.106	0.304	0.103
		400	0.294	0.079	0.299	0.079	0.298	0.078
		600	0.295	0.064	0.300	0.064	0.299	0.064
	75	50	0.329	0.488	0.187	0.420	0.240	0.697
		200	0.302	0.152	0.304	0.165	0.299	0.153
		400	0.296	0.105	0.298	0.109	0.296	0.104
		600	0.294	0.090	0.294	0.091	0.293	0.088

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 52> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	1.057	0.248	0.998	0.282	1.006	0.253
		200	1.040	0.119	0.995	0.119	0.997	0.116
		400	1.043	0.093	1.000	0.086	1.000	0.084
		600	1.041	0.078	1.000	0.070	1.000	0.068
	25	50	1.069	0.279	1.010	0.319	1.018	0.281
		200	1.049	0.133	0.991	0.136	0.996	0.131
		400	1.053	0.106	1.000	0.096	1.001	0.093
		600	1.051	0.090	0.998	0.079	0.999	0.075
	50	50	1.069	0.344	1.012	0.431	1.024	0.344
		200	1.041	0.165	0.994	0.170	1.000	0.163
		400	1.048	0.127	1.001	0.122	1.002	0.119
		600	1.045	0.105	1.000	0.099	1.001	0.095
	75	50	0.977	0.490	1.019	0.875	1.013	0.592
		200	0.968	0.217	0.986	0.236	0.987	0.217
		400	0.980	0.149	0.998	0.150	0.996	0.147
		600	0.972	0.126	1.003	0.128	1.004	0.126

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 53> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	1.059	0.249	1.002	0.279	1.013	0.251
		200	1.042	0.123	1.000	0.121	1.000	0.119
		400	1.047	0.095	1.003	0.086	1.004	0.084
		600	1.047	0.081	1.004	0.069	1.005	0.067
	25	50	1.043	0.255	1.001	0.296	1.000	0.259
		200	1.052	0.136	0.996	0.138	0.999	0.130
		400	1.057	0.105	1.003	0.096	1.005	0.090
		600	1.057	0.093	1.002	0.079	1.005	0.075
	50	50	1.047	0.324	1.002	0.404	0.997	0.334
		200	1.047	0.155	0.997	0.156	0.999	0.151
		400	1.048	0.127	1.001	0.121	1.004	0.116
		600	1.046	0.105	1.000	0.097	1.001	0.092
	75	50	0.976	0.493	1.048	0.855	1.007	0.562
		200	0.966	0.216	0.986	0.233	0.988	0.210
		400	0.981	0.148	0.998	0.146	0.997	0.142
		600	0.983	0.122	1.005	0.122	1.004	0.120

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 54> Y와 보조변수 간 상관이 .10일 때, 조건별  $\mu_Y(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	1.057	0.248	0.999	0.281	1.014	0.254
		200	1.040	0.119	0.995	0.119	1.005	0.116
		400	1.043	0.093	1.000	0.086	1.009	0.085
		600	1.041	0.079	1.000	0.070	1.009	0.069
	25	50	1.069	0.278	1.010	0.318	1.026	0.283
		200	1.049	0.133	0.991	0.136	1.006	0.129
		400	1.053	0.106	0.999	0.096	1.012	0.094
		600	1.057	0.092	1.002	0.080	1.016	0.078
	50	50	1.069	0.344	1.012	0.430	1.033	0.350
		200	1.041	0.164	0.993	0.169	1.006	0.162
		400	1.048	0.127	1.001	0.122	1.011	0.119
		600	1.046	0.105	1.000	0.099	1.009	0.096
	75	50	1.059	0.542	1.052	0.608	0.989	1.122
		200	0.967	0.216	0.986	0.235	0.986	0.216
		400	0.980	0.149	0.998	0.149	0.995	0.146
		600	0.984	0.123	1.005	0.124	1.002	0.122

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 55> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.939	0.225	0.988	0.236	0.952	0.226
		200	0.977	0.109	0.985	0.110	0.982	0.109
		400	0.987	0.075	0.992	0.075	0.991	0.075
		600	0.990	0.060	0.994	0.060	0.994	0.060
	25	50	0.935	0.240	1.005	0.265	0.948	0.243
		200	0.977	0.120	0.986	0.120	0.982	0.119
		400	0.986	0.081	0.992	0.081	0.991	0.081
		600	0.989	0.065	0.994	0.065	0.993	0.065
	50	50	0.912	0.287	1.106	0.455	0.926	0.291
		200	0.973	0.145	0.984	0.147	0.976	0.145
		400	0.984	0.095	0.988	0.096	0.986	0.096
		600	0.990	0.079	0.993	0.080	0.992	0.080
	75	50	0.793	0.463	3.302	4.944	1.024	0.956
		200	0.947	0.219	1.009	0.240	0.953	0.221
		400	0.976	0.149	0.988	0.153	0.977	0.151
		600	0.981	0.130	0.988	0.132	0.983	0.131

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 56> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.940	0.224	0.987	0.232	0.952	0.224
		200	0.991	0.167	0.999	0.104	0.996	0.104
		400	0.994	0.073	1.000	0.073	0.998	0.073
		600	0.992	0.062	0.997	0.062	0.996	0.062
	25	50	0.959	0.221	1.008	0.241	0.968	0.223
		200	0.990	0.112	1.000	0.114	0.995	0.112
		400	0.993	0.079	0.999	0.079	0.997	0.079
		600	0.992	0.067	0.998	0.067	0.996	0.067
	50	50	0.941	0.261	1.114	0.384	0.955	0.263
		200	0.990	0.137	1.000	0.138	0.994	0.137
		400	0.984	0.096	0.988	0.096	0.986	0.096
		600	0.990	0.080	0.993	0.081	0.992	0.081
	75	50	0.789	0.468	3.278	5.061	1.036	1.206
		200	0.946	0.219	1.007	0.242	0.952	0.222
		400	0.976	0.149	0.989	0.154	0.978	0.151
		600	0.985	0.124	0.989	0.126	0.985	0.124

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 57> Y와 보조변수 간 상관이 .10일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.939	0.225	0.988	0.236	0.951	0.226
		200	0.977	0.109	0.985	0.110	0.981	0.109
		400	0.987	0.075	0.992	0.075	0.990	0.075
		600	0.990	0.060	0.995	0.060	0.992	0.060
	25	50	0.935	0.240	1.004	0.265	0.949	0.243
		200	0.977	0.120	0.986	0.120	0.981	0.119
		400	0.986	0.081	0.992	0.081	0.989	0.081
		600	0.992	0.067	0.998	0.067	0.995	0.067
	50	50	0.912	0.287	1.107	0.458	0.927	0.292
		200	0.973	0.145	0.985	0.147	0.975	0.145
		400	0.984	0.095	0.988	0.096	0.986	0.096
		600	0.990	0.080	0.993	0.080	0.991	0.080
	75	50	0.910	0.542	1.246	1.227	1.703	3.002
		200	0.947	0.220	1.008	0.240	0.952	0.221
		400	0.976	0.149	0.989	0.153	0.977	0.150
		600	0.985	0.123	0.990	0.125	0.985	0.124

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 58> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.273	0.177	0.298	0.185	0.298	0.175
		200	0.281	0.080	0.307	0.080	0.306	0.080
		400	0.277	0.063	0.302	0.060	0.302	0.059
		600	0.279	0.052	0.302	0.048	0.302	0.048
	25	50	0.276	0.190	0.302	0.199	0.298	0.187
		200	0.281	0.087	0.305	0.088	0.304	0.088
		400	0.277	0.069	0.301	0.065	0.300	0.064
		600	0.278	0.056	0.301	0.052	0.301	0.051
	50	50	0.280	0.236	0.291	0.255	0.291	0.229
		200	0.298	0.104	0.306	0.106	0.305	0.103
		400	0.289	0.079	0.298	0.079	0.298	0.077
		600	0.290	0.064	0.300	0.062	0.299	0.061
	75	50	0.293	0.392	0.203	0.701	0.302	0.438
		200	0.301	0.151	0.303	0.158	0.300	0.148
		400	0.295	0.104	0.295	0.106	0.294	0.101
		600	0.295	0.089	0.295	0.086	0.294	0.085

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 59> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.271	0.179	0.294	0.180	0.294	0.172
		200	0.282	0.081	0.306	0.077	0.305	0.077
		400	0.277	0.064	0.301	0.058	0.301	0.058
		600	0.275	0.055	0.299	0.049	0.298	0.049
	25	50	0.292	0.182	0.313	0.185	0.314	0.180
		200	0.284	0.085	0.308	0.080	0.307	0.080
		400	0.278	0.065	0.302	0.059	0.301	0.059
		600	0.276	0.056	0.300	0.050	0.299	0.050
	50	50	0.289	0.205	0.307	0.218	0.304	0.199
		200	0.297	0.098	0.307	0.091	0.306	0.090
		400	0.288	0.079	0.298	0.075	0.298	0.074
		600	0.290	0.064	0.299	0.059	0.299	0.059
	75	50	0.293	0.399	0.196	0.620	0.301	0.423
		200	0.302	0.152	0.301	0.153	0.298	0.144
		400	0.295	0.101	0.294	0.099	0.293	0.094
		600	0.295	0.087	0.294	0.081	0.293	0.079

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 60> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.273	0.175	0.300	0.182	0.294	0.171
		200	0.281	0.080	0.306	0.080	0.301	0.079
		400	0.277	0.063	0.301	0.059	0.296	0.059
		600	0.278	0.052	0.302	0.049	0.297	0.048
	25	50	0.276	0.190	0.301	0.198	0.294	0.186
		200	0.280	0.087	0.304	0.087	0.299	0.087
		400	0.277	0.069	0.301	0.065	0.296	0.065
		600	0.275	0.056	0.300	0.051	0.294	0.051
	50	50	0.280	0.234	0.291	0.254	0.287	0.229
		200	0.298	0.104	0.305	0.105	0.304	0.102
		400	0.288	0.079	0.299	0.077	0.296	0.077
		600	0.290	0.064	0.300	0.062	0.297	0.062
	75	50	0.322	0.541	0.196	0.408	0.299	0.738
		200	0.301	0.151	0.302	0.158	0.298	0.147
		400	0.295	0.103	0.295	0.105	0.293	0.100
		600	0.294	0.089	0.295	0.086	0.293	0.084

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 61> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	1.107	0.269	1.009	0.278	1.010	0.254
		200	1.082	0.137	0.992	0.117	0.994	0.114
		400	1.087	0.120	0.999	0.087	1.000	0.085
		600	1.081	0.106	0.999	0.071	0.999	0.069
	25	50	1.123	0.295	1.009	0.310	1.017	0.274
		200	1.105	0.160	0.995	0.132	0.999	0.128
		400	1.106	0.141	0.999	0.097	1.000	0.094
		600	1.105	0.129	0.998	0.079	0.999	0.075
	50	50	1.113	0.359	1.015	0.422	1.023	0.339
		200	1.083	0.180	0.995	0.165	0.999	0.160
		400	1.092	0.150	1.003	0.120	1.004	0.117
		600	1.091	0.131	1.001	0.097	1.002	0.093
	75	50	0.948	0.499	1.030	0.849	1.005	0.571
		200	0.950	0.219	0.988	0.225	0.990	0.208
		400	0.966	0.151	1.002	0.145	1.000	0.141
		600	0.964	0.125	1.004	0.118	1.003	0.117

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 62> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	1.113	0.276	1.014	0.275	1.022	0.252
		200	1.084	0.141	0.999	0.117	1.001	0.115
		400	1.091	0.122	1.004	0.084	1.005	0.082
		600	1.088	0.110	1.005	0.067	1.006	0.066
	25	50	1.092	0.268	0.999	0.278	0.999	0.253
		200	1.101	0.159	0.993	0.129	0.998	0.123
		400	1.107	0.137	1.002	0.091	1.004	0.087
		600	1.108	0.129	1.002	0.075	1.005	0.072
	50	50	1.101	0.302	0.999	0.341	1.005	0.294
		200	1.094	0.174	0.996	0.149	1.000	0.145
		400	1.094	0.150	1.003	0.112	1.005	0.108
		600	1.092	0.132	1.002	0.092	1.003	0.088
	75	50	0.949	0.504	1.057	0.799	0.999	0.547
		200	0.949	0.218	0.991	0.215	0.993	0.194
		400	0.966	0.150	1.003	0.135	1.002	0.130
		600	0.964	0.124	1.006	0.112	1.004	0.110

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 63> Y와 보조변수 간 상관이 .20일 때, 조건별  $\mu_Y(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	1.104	0.264	1.007	0.272	1.029	0.249
		200	1.082	0.137	0.993	0.117	1.013	0.113
		400	1.086	0.119	0.999	0.086	1.019	0.087
		600	1.081	0.106	0.999	0.071	1.017	0.071
	25	50	1.123	0.295	1.010	0.309	1.038	0.274
		200	1.104	0.160	0.996	0.131	1.021	0.128
		400	1.106	0.140	0.999	0.096	1.023	0.096
		600	1.107	0.128	1.001	0.076	1.026	0.078
	50	50	1.113	0.357	1.016	0.420	1.040	0.342
		200	1.083	0.180	0.995	0.164	1.014	0.160
		400	1.094	0.150	1.002	0.118	1.021	0.118
		600	1.092	0.132	1.002	0.096	1.019	0.095
	75	50	0.818	0.578	0.952	0.671	0.943	1.055
		200	0.950	0.219	0.989	0.224	0.985	0.207
		400	0.966	0.151	1.002	0.143	0.997	0.140
		600	0.964	0.125	1.004	0.118	0.998	0.117

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 64> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.923	0.230	0.978	0.239	0.946	0.254
		200	0.966	0.112	0.985	0.112	0.982	0.114
		400	0.974	0.079	0.991	0.077	0.990	0.085
		600	0.978	0.063	0.994	0.062	0.993	0.069
	25	50	0.922	0.242	0.996	0.269	0.944	0.274
		200	0.965	0.121	0.986	0.121	0.982	0.128
		400	0.976	0.084	0.993	0.084	0.991	0.094
		600	0.979	0.067	0.996	0.067	0.994	0.075
	50	50	0.912	0.286	1.101	0.441	0.932	0.339
		200	0.967	0.146	0.983	0.148	0.974	0.160
		400	0.979	0.097	0.988	0.098	0.986	0.117
		600	0.985	0.080	0.992	0.081	0.991	0.093
	75	50	0.796	0.464	3.209	4.746	1.006	0.571
		200	0.948	0.216	1.004	0.237	0.955	0.208
		400	0.975	0.148	0.988	0.153	0.978	0.141
		600	0.983	0.122	0.987	0.124	0.984	0.117

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 65> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.922	0.232	0.975	0.240	0.945	0.252
		200	0.979	0.104	0.999	0.105	0.996	0.115
		400	0.981	0.075	0.999	0.075	0.998	0.082
		600	0.980	0.064	0.997	0.063	0.996	0.066
	25	50	0.950	0.217	1.002	0.234	0.968	0.253
		200	0.977	0.113	0.999	0.114	0.994	0.123
		400	0.979	0.080	0.998	0.079	0.996	0.087
		600	0.978	0.070	0.996	0.068	0.995	0.072
	50	50	0.944	0.245	1.047	0.293	0.967	0.294
		200	0.985	0.137	1.000	0.137	0.994	0.145
		400	0.979	0.096	0.988	0.096	0.986	0.108
		600	0.985	0.081	0.992	0.081	0.992	0.088
	75	50	0.793	0.469	2.760	3.868	1.032	0.547
		200	0.947	0.216	1.002	0.240	0.957	0.194
		400	0.975	0.148	0.989	0.151	0.979	0.130
		600	0.983	0.124	0.987	0.122	0.984	0.110

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 66> Y와 보조변수 간 상관이 .20일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.923	0.228	0.977	0.237	0.941	0.249
		200	0.966	0.112	0.985	0.111	0.977	0.113
		400	0.974	0.078	0.991	0.077	0.984	0.087
		600	0.978	0.063	0.994	0.062	0.987	0.071
	25	50	0.921	0.242	0.995	0.267	0.941	0.274
		200	0.965	0.120	0.985	0.121	0.976	0.128
		400	0.976	0.084	0.993	0.083	0.986	0.096
		600	0.978	0.069	0.997	0.068	0.989	0.078
	50	50	0.912	0.286	1.101	0.442	0.932	0.342
		200	0.967	0.146	0.983	0.148	0.973	0.160
		400	0.979	0.096	0.988	0.097	0.984	0.118
		600	0.985	0.080	0.992	0.081	0.989	0.095
	75	50	0.687	0.570	1.298	1.486	1.814	1.055
		200	0.948	0.217	1.003	0.238	0.954	0.207
		400	0.975	0.148	0.988	0.153	0.977	0.140
		600	0.983	0.123	0.987	0.124	0.984	0.117

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 67> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.262	0.180	0.297	0.179	0.298	0.172
		200	0.268	0.083	0.306	0.080	0.306	0.079
		400	0.265	0.068	0.302	0.059	0.302	0.058
		600	0.268	0.056	0.303	0.048	0.303	0.047
	25	50	0.273	0.196	0.301	0.203	0.300	0.189
		200	0.269	0.089	0.304	0.086	0.304	0.086
		400	0.266	0.073	0.301	0.063	0.301	0.062
		600	0.266	0.061	0.301	0.050	0.301	0.050
	50	50	0.276	0.236	0.293	0.239	0.291	0.220
		200	0.292	0.104	0.306	0.101	0.305	0.098
		400	0.285	0.080	0.298	0.075	0.298	0.075
		600	0.287	0.066	0.299	0.060	0.299	0.060
	75	50	0.298	0.443	0.205	0.615	0.313	0.451
		200	0.300	0.157	0.300	0.157	0.298	0.145
		400	0.294	0.105	0.294	0.102	0.294	0.097
		600	0.293	0.090	0.294	0.081	0.294	0.080

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 68> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.259	0.180	0.293	0.169	0.293	0.164
		200	0.271	0.084	0.307	0.075	0.306	0.074
		400	0.266	0.067	0.302	0.056	0.301	0.056
		600	0.265	0.060	0.300	0.047	0.299	0.048
	25	50	0.277	0.186	0.311	0.177	0.310	0.175
		200	0.271	0.087	0.308	0.077	0.307	0.076
		400	0.267	0.069	0.302	0.057	0.302	0.057
		600	0.264	0.062	0.300	0.049	0.300	0.049
	50	50	0.284	0.220	0.303	0.206	0.304	0.191
		200	0.291	0.099	0.307	0.084	0.305	0.083
		400	0.284	0.080	0.298	0.066	0.297	0.066
		600	0.286	0.066	0.299	0.054	0.299	0.054
	75	50	0.295	0.440	0.191	0.499	0.310	0.425
		200	0.300	0.158	0.298	0.136	0.297	0.129
		400	0.292	0.103	0.294	0.089	0.294	0.084
		600	0.293	0.087	0.294	0.069	0.294	0.068

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 69> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.262	0.180	0.296	0.179	0.290	0.170
		200	0.268	0.083	0.306	0.079	0.298	0.078
		400	0.265	0.068	0.302	0.058	0.294	0.058
		600	0.278	0.052	0.302	0.049	0.297	0.048
	25	50	0.273	0.198	0.297	0.203	0.291	0.190
		200	0.269	0.089	0.304	0.085	0.297	0.085
		400	0.266	0.072	0.301	0.063	0.293	0.063
		600	0.264	0.062	0.300	0.051	0.291	0.051
	50	50	0.277	0.234	0.294	0.236	0.286	0.219
		200	0.292	0.103	0.305	0.100	0.303	0.098
		400	0.285	0.080	0.298	0.074	0.295	0.074
		600	0.287	0.066	0.299	0.060	0.296	0.060
	75	50	0.288	0.434	0.174	0.383	0.236	0.603
		200	0.301	0.154	0.299	0.154	0.297	0.144
		400	0.293	0.105	0.294	0.101	0.293	0.097
		600	0.293	0.089	0.294	0.081	0.293	0.080

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 70> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	1.162	0.298	1.017	0.278	1.014	0.252
		200	1.130	0.170	0.994	0.116	0.996	0.114
		400	1.132	0.156	1.000	0.086	1.000	0.084
		600	1.121	0.138	0.997	0.071	0.998	0.069
	25	50	1.171	0.328	1.007	0.315	1.014	0.274
		200	1.158	0.200	0.998	0.129	1.000	0.127
		400	1.160	0.184	1.000	0.093	1.001	0.090
		600	1.159	0.175	0.999	0.076	1.000	0.073
	50	50	1.157	0.374	1.015	0.393	1.024	0.324
		200	1.130	0.205	0.997	0.157	1.000	0.153
		400	1.136	0.180	1.004	0.112	1.004	0.110
		600	1.134	0.165	1.003	0.092	1.003	0.089
	75	50	0.895	0.526	1.040	0.764	1.001	0.575
		200	0.921	0.230	0.992	0.220	0.996	0.202
		400	0.941	0.159	1.002	0.138	1.001	0.133
		600	0.942	0.134	1.005	0.113	1.005	0.111

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 71> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	1.164	0.301	1.019	0.257	1.027	0.243
		200	1.124	0.167	0.996	0.110	0.999	0.110
		400	1.132	0.154	1.003	0.081	1.004	0.080
		600	1.127	0.143	1.003	0.064	1.005	0.064
	25	50	1.158	0.302	0.995	0.267	1.006	0.249
		200	1.156	0.197	0.995	0.120	0.999	0.117
		400	1.159	0.180	1.002	0.086	1.004	0.083
		600	1.159	0.175	1.003	0.069	1.005	0.068
	50	50	1.148	0.338	1.010	0.314	1.011	0.275
		200	1.142	0.205	0.997	0.136	1.001	0.134
		400	1.136	0.179	1.004	0.097	1.006	0.096
		600	1.135	0.165	1.003	0.080	1.004	0.079
	75	50	0.901	0.534	1.062	0.622	1.001	0.522
		200	0.921	0.229	0.999	0.191	1.001	0.175
		400	0.941	0.158	1.004	0.116	1.003	0.112
		600	0.945	0.131	1.007	0.097	1.005	0.096

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 72> Y와 보조변수 간 상관이 .30일 때, 조건별  $\mu_Y(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	1.161	0.298	1.018	0.277	1.044	0.252
		200	1.130	0.170	0.995	0.116	1.024	0.115
		400	1.132	0.156	1.000	0.085	1.029	0.089
		600	1.081	0.106	0.999	0.071	1.017	0.071
	25	50	1.169	0.332	1.013	0.322	1.046	0.286
		200	1.158	0.200	0.999	0.128	1.033	0.130
		400	1.160	0.184	1.000	0.093	1.036	0.098
		600	1.159	0.175	1.002	0.075	1.039	0.082
	50	50	1.156	0.372	1.015	0.386	1.051	0.327
		200	1.130	0.205	0.998	0.155	1.024	0.153
		400	1.136	0.180	1.004	0.111	1.029	0.113
		600	1.135	0.165	1.003	0.091	1.029	0.093
	75	50	0.730	0.641	0.902	0.629	0.994	0.983
		200	0.923	0.227	0.992	0.216	0.985	0.201
		400	0.940	0.160	1.003	0.136	0.992	0.133
		600	0.942	0.134	1.006	0.112	0.996	0.112

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 73> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.899	0.237	0.970	0.246	0.941	0.240
		200	0.945	0.118	0.984	0.114	0.981	0.113
		400	0.953	0.087	0.989	0.079	0.988	0.078
		600	0.959	0.071	0.994	0.063	0.993	0.062
	25	50	0.905	0.253	0.997	0.288	0.941	0.258
		200	0.945	0.125	0.983	0.121	0.980	0.120
		400	0.956	0.089	0.991	0.082	0.990	0.082
		600	0.960	0.075	0.995	0.069	0.994	0.068
	50	50	0.901	0.282	1.069	0.392	0.930	0.293
		200	0.959	0.146	0.981	0.148	0.975	0.146
		400	0.972	0.099	0.987	0.099	0.985	0.099
		600	0.976	0.082	0.991	0.082	0.990	0.081
	75	50	0.781	0.476	2.474	3.258	1.007	0.951
		200	0.946	0.221	1.003	0.240	0.960	0.221
		400	0.972	0.150	0.988	0.153	0.979	0.149
		600	0.985	0.125	0.990	0.124	0.987	0.122

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 74> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.901	0.237	0.964	0.239	0.943	0.235
		200	0.959	0.109	0.996	0.104	0.994	0.104
		400	0.961	0.080	0.998	0.073	0.997	0.073
		600	0.961	0.073	0.996	0.064	0.996	0.063
	25	50	0.931	0.225	0.991	0.231	0.968	0.225
		200	0.956	0.118	0.996	0.112	0.994	0.111
		400	0.961	0.087	0.998	0.079	0.997	0.078
		600	0.960	0.076	0.996	0.066	0.995	0.065
	50	50	0.930	0.256	1.025	0.295	0.963	0.254
		200	0.972	0.138	0.996	0.129	0.992	0.129
		400	0.972	0.100	0.988	0.095	0.986	0.094
		600	0.976	0.083	0.991	0.079	0.991	0.078
	75	50	0.780	0.477	1.882	2.137	1.027	0.784
		200	0.945	0.221	0.999	0.229	0.968	0.207
		400	0.974	0.151	0.993	0.142	0.985	0.138
		600	0.984	0.125	0.991	0.112	0.989	0.112

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 75> Y와 보조변수 간 상관이 .30일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.899	0.237	0.969	0.245	0.930	0.236
		200	0.945	0.118	0.984	0.114	0.968	0.113
		400	0.953	0.087	0.989	0.078	0.975	0.080
		600	0.978	0.063	0.994	0.062	0.987	0.062
	25	50	0.904	0.255	0.996	0.291	0.933	0.255
		200	0.945	0.125	0.982	0.121	0.968	0.121
		400	0.956	0.089	0.991	0.082	0.977	0.083
		600	0.960	0.076	0.997	0.068	0.982	0.069
	50	50	0.901	0.282	1.065	0.386	0.927	0.291
		200	0.959	0.146	0.981	0.148	0.970	0.145
		400	0.972	0.099	0.987	0.099	0.981	0.099
		600	0.976	0.082	0.991	0.082	0.985	0.081
	75	50	0.698	0.541	1.072	0.933	1.591	1.711
		200	0.946	0.220	1.001	0.241	0.959	0.221
		400	0.974	0.151	0.990	0.153	0.980	0.151
		600	0.985	0.125	0.991	0.123	0.987	0.122

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 76> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.249	0.183	0.295	0.172	0.298	0.167
		200	0.256	0.087	0.305	0.077	0.305	0.077
		400	0.253	0.074	0.301	0.056	0.301	0.056
		600	0.255	0.064	0.302	0.046	0.302	0.046
	25	50	0.256	0.189	0.299	0.181	0.300	0.173
		200	0.259	0.091	0.305	0.082	0.305	0.082
		400	0.255	0.077	0.300	0.060	0.300	0.060
		600	0.255	0.067	0.301	0.048	0.301	0.048
	50	50	0.276	0.233	0.291	0.226	0.293	0.205
		200	0.288	0.103	0.304	0.094	0.304	0.093
		400	0.280	0.080	0.298	0.069	0.298	0.069
		600	0.282	0.065	0.299	0.055	0.299	0.055
	75	50	0.303	0.470	0.211	0.523	0.336	0.434
		200	0.300	0.157	0.298	0.141	0.298	0.131
		400	0.292	0.106	0.294	0.094	0.294	0.090
		600	0.292	0.089	0.294	0.074	0.293	0.072

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 77> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.246	0.179	0.289	0.149	0.289	0.149
		200	0.261	0.087	0.308	0.072	0.307	0.073
		400	0.255	0.073	0.302	0.055	0.302	0.055
		600	0.254	0.067	0.300	0.046	0.300	0.046
	25	50	0.302	0.182	0.311	0.192	0.312	0.184
		200	0.294	0.083	0.306	0.083	0.305	0.082
		400	0.256	0.074	0.302	0.055	0.303	0.055
		600	0.252	0.069	0.300	0.046	0.300	0.046
	50	50	0.278	0.218	0.313	0.159	0.310	0.157
		200	0.285	0.097	0.306	0.072	0.307	0.072
		400	0.279	0.080	0.298	0.054	0.298	0.055
		600	0.281	0.064	0.299	0.044	0.299	0.044
	75	50	0.313	0.465	0.169	0.287	0.303	0.286
		200	0.299	0.162	0.300	0.084	0.300	0.085
		400	0.290	0.102	0.297	0.060	0.297	0.060
		600	0.291	0.084	0.298	0.047	0.297	0.048

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 78> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.251	0.183	0.295	0.171	0.251	0.183
		200	0.254	0.088	0.304	0.077	0.294	0.077
		400	0.253	0.074	0.301	0.056	0.290	0.056
		600	0.255	0.064	0.302	0.046	0.292	0.046
	25	50	0.256	0.189	0.298	0.178	0.289	0.172
		200	0.258	0.091	0.304	0.080	0.295	0.081
		400	0.254	0.077	0.300	0.059	0.290	0.060
		600	0.252	0.069	0.299	0.050	0.288	0.051
	50	50	0.277	0.232	0.291	0.217	0.288	0.205
		200	0.287	0.103	0.303	0.092	0.301	0.092
		400	0.280	0.080	0.297	0.068	0.295	0.068
		600	0.282	0.066	0.299	0.055	0.295	0.055
	75	50	0.322	0.449	0.142	0.324	0.271	0.514
		200	0.300	0.157	0.298	0.138	0.297	0.131
		400	0.292	0.105	0.293	0.092	0.293	0.089
		600	0.291	0.089	0.294	0.073	0.292	0.072

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 79> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	1.212	0.325	1.019	0.262	1.014	0.239
		200	1.174	0.204	0.305	0.111	0.998	0.109
		400	1.175	0.193	1.001	0.084	1.001	0.081
		600	1.164	0.176	0.999	0.068	0.999	0.067
	25	50	1.223	0.346	1.014	0.280	0.996	0.252
		200	1.212	0.243	0.998	0.123	0.999	0.120
		400	1.216	0.234	1.002	0.090	1.002	0.087
		600	1.213	0.225	1.000	0.073	1.001	0.070
	50	50	1.201	0.394	1.026	0.365	1.025	0.300
		200	1.175	0.235	0.999	0.144	1.000	0.141
		400	1.179	0.214	1.004	0.103	1.004	0.101
		600	1.179	0.202	1.003	0.084	1.003	0.083
	75	50	0.837	0.545	1.036	0.675	0.986	0.548
		200	0.896	0.241	0.995	0.201	0.998	0.183
		400	0.920	0.168	1.004	0.124	1.003	0.120
		600	0.928	0.140	1.007	0.104	1.006	0.101

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 80> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	1.219	0.330	1.020	0.213	1.032	0.221
		200	1.166	0.199	0.995	0.103	0.998	0.107
		400	1.170	0.188	1.002	0.076	1.003	0.078
		600	1.167	0.179	1.003	0.060	1.004	0.062
	25	50	1.043	0.255	1.001	0.296	1.000	0.259
		200	1.052	0.136	0.996	0.138	0.999	0.130
		400	1.212	0.228	1.002	0.076	1.003	0.079
		600	1.212	0.223	1.004	0.061	1.004	0.063
	50	50	1.192	0.360	0.990	0.228	1.005	0.224
		200	1.191	0.241	0.997	0.108	1.000	0.110
		400	1.179	0.213	1.003	0.076	1.005	0.078
		600	1.181	0.204	1.003	0.063	1.004	0.065
	75	50	0.825	0.549	1.074	0.400	0.996	0.363
		200	0.897	0.240	1.002	0.121	1.000	0.122
		400	0.921	0.164	1.004	0.082	1.003	0.083
		600	0.930	0.136	1.004	0.068	1.003	0.069

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 81> Y와 보조변수 간 상관이 .40일 때, 조건별  $\mu_Y(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	1.215	0.329	1.018	0.261	1.214	0.329
		200	1.179	0.209	0.997	0.111	1.037	0.115
		400	1.176	0.194	1.001	0.083	1.040	0.090
		600	1.164	0.177	0.999	0.068	1.035	0.074
	25	50	1.223	0.345	1.015	0.276	1.057	0.253
		200	1.212	0.243	0.999	0.121	1.043	0.127
		400	1.215	0.233	1.002	0.088	1.049	0.100
		600	1.212	0.223	1.004	0.072	1.052	0.087
	50	50	1.200	0.392	1.026	0.347	1.062	0.306
		200	1.175	0.235	1.000	0.140	1.033	0.143
		400	1.179	0.214	1.004	0.100	1.037	0.107
		600	1.180	0.203	1.003	0.083	1.037	0.090
	75	50	0.655	0.683	0.879	0.586	0.922	0.674
		200	0.896	0.241	0.996	0.196	0.983	0.184
		400	0.920	0.167	1.004	0.121	0.990	0.119
		600	0.929	0.139	1.007	0.101	0.994	0.102

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 82> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Uniform	15	50	0.875	0.242	0.967	0.250	0.943	0.242
		200	0.917	0.131	0.983	0.115	0.981	0.113
		400	0.926	0.103	0.989	0.079	0.989	0.078
		600	0.932	0.088	0.994	0.064	0.993	0.063
	25	50	0.874	0.250	0.972	0.260	0.940	0.246
		200	0.919	0.137	0.983	0.124	0.981	0.122
		400	0.930	0.104	0.990	0.083	0.989	0.083
		600	0.932	0.092	0.994	0.069	0.993	0.068
	50	50	0.889	0.282	1.043	0.360	0.930	0.288
		200	0.949	0.146	0.982	0.143	0.977	0.142
		400	0.964	0.102	0.988	0.097	0.987	0.097
		600	0.966	0.085	0.990	0.081	0.990	0.080
	75	50	0.777	0.485	1.979	2.458	1.028	0.896
		200	0.945	0.219	0.998	0.226	0.966	0.210
		400	0.971	0.151	0.988	0.146	0.982	0.143
		600	0.983	0.124	0.990	0.115	0.987	0.114

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 83> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Mod	15	50	0.876	0.243	0.957	0.215	0.949	0.214
		200	0.933	0.117	0.995	0.096	0.995	0.098
		400	0.936	0.095	0.998	0.068	0.999	0.069
		600	0.936	0.088	0.997	0.059	0.997	0.059
	25	50	0.959	0.221	1.008	0.241	0.968	0.223
		200	0.990	0.112	1.000	0.114	0.995	0.112
		400	0.933	0.102	0.998	0.069	0.998	0.070
		600	0.931	0.095	0.996	0.060	0.996	0.060
	50	50	0.916	0.260	0.985	0.213	0.968	0.223
		200	0.961	0.139	0.995	0.104	0.993	0.105
		400	0.965	0.102	0.991	0.075	0.992	0.075
		600	0.967	0.087	0.994	0.062	0.994	0.062
	75	50	0.775	0.496	0.920	0.502	0.965	0.418
		200	0.944	0.220	0.987	0.139	0.983	0.138
		400	0.973	0.151	0.991	0.094	0.991	0.094
		600	0.984	0.125	0.993	0.075	0.993	0.077

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 84> Y와 보조변수 간 상관이 .40일 때, 조건별  $\sigma_Y^2(1)$ 의 RMSE(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			Est	RMSE	Est	RMSE	Est	RMSE
Low	15	50	0.874	0.243	0.965	0.247	0.874	0.243
		200	0.917	0.132	0.983	0.114	0.958	0.116
		400	0.926	0.103	0.989	0.079	0.965	0.083
		600	0.932	0.088	0.994	0.064	0.970	0.068
	25	50	0.874	0.250	0.969	0.256	0.923	0.242
		200	0.919	0.137	0.983	0.123	0.960	0.124
		400	0.929	0.104	0.990	0.082	0.966	0.086
		600	0.931	0.095	0.996	0.070	0.970	0.074
	50	50	0.889	0.282	1.030	0.345	0.925	0.285
		200	0.949	0.147	0.982	0.143	0.969	0.142
		400	0.964	0.102	0.988	0.096	0.980	0.097
		600	0.966	0.085	0.991	0.081	0.982	0.080
	75	50	0.669	0.535	1.027	0.869	1.332	1.200
		200	0.945	0.219	0.998	0.226	0.963	0.209
		400	0.972	0.151	0.990	0.144	0.982	0.144
		600	0.984	0.124	0.990	0.113	0.988	0.114

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, Est=추정치

<표 85> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Uniform	15	50	0.020	0.549	0.028	0.574	0.031	0.565
		200	0.163	0.421	0.175	0.436	0.175	0.434
		400	0.198	0.380	0.209	0.394	0.209	0.393
		600	0.215	0.364	0.226	0.376	0.227	0.376
	25	50	0.005	0.570	0.009	0.591	0.014	0.582
		200	0.156	0.431	0.167	0.445	0.167	0.444
		400	0.191	0.386	0.202	0.399	0.202	0.398
		600	0.210	0.369	0.221	0.382	0.221	0.381
	50	50	-0.048	0.617	-0.054	0.638	-0.040	0.623
		200	0.130	0.471	0.136	0.475	0.135	0.473
		400	0.176	0.411	0.182	0.417	0.181	0.416
		600	0.199	0.392	0.204	0.397	0.204	0.396
	75	50	-0.244	0.823	-0.232	0.797	-0.219	0.816
		200	0.057	0.546	0.061	0.548	0.060	0.544
		400	0.125	0.468	0.128	0.468	0.127	0.467
		600	0.145	0.443	0.147	0.443	0.146	0.441

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 86> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Mod	15	50	0.018	0.547	0.026	0.570	0.028	0.560
		200	0.164	0.424	0.175	0.437	0.176	0.436
		400	0.198	0.382	0.209	0.394	0.210	0.393
		600	0.213	0.361	0.224	0.374	0.224	0.373
	25	50	0.025	0.580	0.026	0.597	0.034	0.590
		200	0.156	0.431	0.167	0.445	0.168	0.443
		400	0.192	0.385	0.204	0.398	0.204	0.397
		600	0.208	0.366	0.220	0.379	0.220	0.377
	50	50	-0.043	0.637	-0.046	0.653	-0.032	0.639
		200	0.139	0.466	0.145	0.469	0.145	0.468
		400	0.176	0.411	0.183	0.416	0.182	0.415
		600	0.199	0.392	0.205	0.396	0.204	0.395
	75	50	-0.245	0.832	-0.235	0.774	-0.213	0.818
		200	0.059	0.547	0.065	0.544	0.061	0.539
		400	0.125	0.467	0.129	0.465	0.128	0.463
		600	0.153	0.436	0.155	0.432	0.154	0.431

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 87> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Low	15	50	0.020	0.549	0.028	0.574	0.029	0.563
		200	0.163	0.421	0.174	0.435	0.172	0.432
		400	0.198	0.380	0.209	0.393	0.207	0.390
		600	0.215	0.363	0.226	0.376	0.224	0.373
	25	50	0.006	0.569	0.009	0.590	0.013	0.581
		200	0.156	0.431	0.167	0.445	0.165	0.441
		400	0.191	0.386	0.202	0.399	0.200	0.396
		600	0.208	0.365	0.220	0.379	0.218	0.375
	50	50	-0.048	0.617	-0.052	0.638	-0.045	0.619
		200	0.130	0.470	0.136	0.475	0.134	0.473
		400	0.176	0.411	0.182	0.417	0.181	0.415
		600	0.199	0.392	0.204	0.396	0.203	0.395
	75	50	-0.277	0.881	NA	NA	-0.274	0.820
		200	0.058	0.546	0.061	0.547	0.057	0.541
		400	0.125	0.468	0.128	0.467	0.126	0.465
		600	0.152	0.436	0.153	0.434	0.153	0.433

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 88> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Uniform	15	50	0.008	0.537	0.028	0.569	0.032	0.563
		200	0.153	0.410	0.177	0.436	0.177	0.435
		400	0.186	0.368	0.210	0.393	0.210	0.393
		600	0.205	0.352	0.228	0.376	0.228	0.376
	25	50	-0.005	0.556	0.015	0.588	0.019	0.578
		200	0.143	0.419	0.166	0.443	0.167	0.442
		400	0.179	0.374	0.203	0.398	0.203	0.397
		600	0.198	0.357	0.222	0.381	0.222	0.380
	50	50	-0.054	0.613	-0.049	0.630	-0.036	0.617
		200	0.128	0.467	0.140	0.471	0.139	0.470
		400	0.170	0.408	0.182	0.414	0.182	0.413
		600	0.194	0.387	0.206	0.394	0.205	0.393
	75	50	-0.245	0.832	-0.263	0.733	-0.203	0.800
		200	0.057	0.546	0.067	0.539	0.065	0.535
		400	0.124	0.466	0.130	0.460	0.130	0.459
		600	0.153	0.436	0.159	0.431	0.158	0.430

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 89> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Mod	15	50	0.005	0.537	0.027	0.562	0.031	0.557
		200	0.153	0.411	0.177	0.434	0.178	0.433
		400	0.186	0.369	0.210	0.392	0.210	0.391
		600	0.201	0.349	0.225	0.372	0.225	0.371
	25	50	0.015	0.568	0.036	0.591	0.043	0.586
		200	0.147	0.421	0.173	0.443	0.173	0.441
		400	0.182	0.374	0.207	0.396	0.207	0.395
		600	0.197	0.354	0.223	0.377	0.223	0.376
	50	50	-0.023	0.601	-0.002	0.615	0.004	0.604
		200	0.134	0.460	0.153	0.462	0.152	0.460
		400	0.170	0.406	0.187	0.410	0.187	0.409
		600	0.194	0.386	0.208	0.390	0.208	0.390
	75	50	-0.251	0.835	-0.187	0.743	-0.192	0.785
		200	0.058	0.546	0.078	0.525	0.076	0.521
		400	0.123	0.466	0.138	0.450	0.137	0.449
		600	0.153	0.436	0.166	0.423	0.165	0.422

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 90> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Low	15	50	0.011	0.535	0.031	0.568	0.031	0.556
		200	0.153	0.409	0.177	0.435	0.172	0.429
		400	0.186	0.368	0.210	0.393	0.205	0.387
		600	0.205	0.352	0.227	0.376	0.223	0.370
	25	50	-0.004	0.556	0.015	0.586	0.015	0.574
		200	0.143	0.418	0.166	0.442	0.162	0.436
		400	0.180	0.374	0.204	0.398	0.199	0.392
		600	0.197	0.353	0.222	0.378	0.216	0.372
	50	50	-0.053	0.613	-0.047	0.630	-0.040	0.613
		200	0.128	0.467	0.140	0.471	0.138	0.469
		400	0.171	0.406	0.184	0.413	0.182	0.411
		600	0.194	0.386	0.206	0.393	0.203	0.391
	75	50	-0.289	0.879	0.384	0.939	-0.337	0.859
		200	0.057	0.546	0.067	0.537	0.064	0.533
		400	0.123	0.466	0.130	0.459	0.129	0.457
		600	0.153	0.436	0.159	0.430	0.158	0.429

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 91> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Uniform	15	50	-0.002	0.525	0.029	0.564	0.036	0.560
		200	0.141	0.396	0.178	0.434	0.179	0.433
		400	0.175	0.355	0.212	0.392	0.212	0.392
		600	0.195	0.341	0.230	0.376	0.230	0.376
	25	50	-0.017	0.563	0.012	0.590	0.018	0.582
		200	0.133	0.406	0.169	0.440	0.170	0.439
		400	0.169	0.362	0.205	0.397	0.206	0.396
		600	0.187	0.345	0.223	0.379	0.223	0.379
	50	50	-0.055	0.607	-0.033	0.619	-0.023	0.604
		200	0.125	0.460	0.147	0.464	0.147	0.462
		400	0.166	0.404	0.186	0.410	0.186	0.409
		600	0.190	0.384	0.208	0.391	0.208	0.390
	75	50	-0.246	0.847	-0.151	0.778	-0.176	0.799
		200	0.051	0.549	0.074	0.527	0.073	0.524
		400	0.120	0.467	0.137	0.452	0.137	0.451
		600	0.150	0.436	0.165	0.424	0.165	0.423

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 92> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Mod	15	50	-0.004	0.522	0.037	0.548	0.040	0.545
		200	0.143	0.399	0.183	0.431	0.183	0.430
		400	0.176	0.357	0.214	0.389	0.214	0.389
		600	0.192	0.338	0.228	0.371	0.228	0.370
	25	50	-0.008	0.562	0.043	0.580	0.044	0.576
		200	0.135	0.406	0.179	0.436	0.179	0.435
		400	0.171	0.362	0.212	0.393	0.212	0.392
		600	0.186	0.342	0.226	0.374	0.226	0.373
	50	50	-0.041	0.608	0.011	0.595	0.015	0.592
		200	0.130	0.453	0.165	0.448	0.164	0.447
		400	0.166	0.403	0.195	0.400	0.195	0.400
		600	0.189	0.383	0.215	0.382	0.215	0.382
	75	50	-0.266	0.853	-0.073	0.730	-0.108	0.734
		200	0.052	0.548	0.104	0.492	0.102	0.491
		400	0.118	0.467	0.158	0.430	0.157	0.430
		600	0.151	0.435	0.184	0.405	0.183	0.405

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 93> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Low	15	50	-0.002	0.525	0.030	0.562	0.030	0.551
		200	0.141	0.396	0.178	0.434	0.171	0.424
		400	0.175	0.355	0.212	0.392	0.204	0.383
		600	0.205	0.352	0.227	0.376	0.223	0.370
	25	50	-0.022	0.568	0.006	0.588	0.007	0.576
		200	0.133	0.406	0.169	0.439	0.163	0.431
		400	0.169	0.363	0.206	0.396	0.199	0.388
		600	0.186	0.342	0.224	0.377	0.215	0.368
	50	50	-0.054	0.607	-0.029	0.617	-0.026	0.599
		200	0.125	0.460	0.148	0.462	0.145	0.460
		400	0.166	0.403	0.186	0.409	0.184	0.407
		600	0.190	0.384	0.208	0.390	0.205	0.388
	75	50	-0.330	0.919	0.387	0.984	-0.337	0.814
		200	0.053	0.548	0.076	0.523	0.073	0.521
		400	0.119	0.468	0.137	0.450	0.135	0.451
		600	0.150	0.436	0.166	0.422	0.164	0.422

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 94> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(1)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Uniform	15	50	-0.013	0.510	0.035	0.556	0.042	0.554
		200	0.130	0.382	0.180	0.430	0.180	0.429
		400	0.164	0.341	0.213	0.390	0.213	0.389
		600	0.183	0.327	0.230	0.374	0.231	0.374
	25	50	-0.018	0.530	0.032	0.567	0.037	0.563
		200	0.123	0.394	0.174	0.436	0.175	0.436
		400	0.158	0.351	0.207	0.393	0.208	0.393
		600	0.177	0.333	0.225	0.377	0.226	0.376
	50	50	-0.053	0.604	-0.015	0.597	-0.004	0.589
		200	0.121	0.454	0.155	0.453	0.155	0.453
		400	0.161	0.399	0.192	0.404	0.192	0.403
		600	0.186	0.379	0.213	0.385	0.213	0.385
	75	50	-0.268	0.881	-0.064	0.779	-0.139	0.761
		200	0.051	0.549	0.093	0.504	0.093	0.503
		400	0.118	0.467	0.150	0.437	0.150	0.437
		600	0.150	0.434	0.177	0.410	0.177	0.410

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 95> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(2)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Mod	15	50	-0.015	0.508	0.053	0.525	0.052	0.527
		200	0.135	0.387	0.190	0.426	0.189	0.426
		400	0.167	0.344	0.219	0.386	0.219	0.386
		600	0.182	0.326	0.232	0.368	0.232	0.368
	25	50	0.025	0.580	0.026	0.597	0.034	0.590
		200	0.156	0.431	0.167	0.445	0.168	0.443
		400	0.161	0.350	0.218	0.387	0.218	0.387
		600	0.176	0.328	0.231	0.368	0.231	0.368
	50	50	-0.048	0.603	0.064	0.562	0.058	0.563
		200	0.124	0.446	0.184	0.429	0.184	0.429
		400	0.160	0.398	0.212	0.385	0.211	0.385
		600	0.184	0.377	0.228	0.369	0.228	0.370
	75	50	-0.250	0.875	0.030	0.586	-0.021	0.598
		200	0.052	0.547	0.165	0.434	0.162	0.438
		400	0.116	0.464	0.202	0.392	0.201	0.394
		600	0.149	0.432	0.220	0.375	0.219	0.375

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 96> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간(3)

$\rho_{AA}$	Mis(%)	N	No-AUX		Inclusive		PCA	
			CI.U	CI.L	CI.U	CI.L	CI.U	CI.L
Low	15	50	-0.012	0.514	0.036	0.555	0.175	0.327
		200	0.128	0.381	0.179	0.429	0.170	0.417
		400	0.164	0.342	0.213	0.389	0.204	0.377
		600	0.183	0.327	0.230	0.373	0.221	0.363
	25	50	-0.017	0.529	0.032	0.563	0.029	0.550
		200	0.123	0.393	0.174	0.434	0.166	0.424
		400	0.158	0.350	0.208	0.392	0.199	0.381
		600	0.175	0.328	0.225	0.374	0.215	0.362
	50	50	-0.051	0.605	-0.010	0.593	-0.006	0.582
		200	0.121	0.454	0.156	0.451	0.154	0.449
		400	0.161	0.399	0.193	0.402	0.189	0.400
		600	0.185	0.378	0.214	0.384	0.210	0.381
	75	50	-0.303	0.964	0.361	0.916	-0.196	0.724
		200	0.051	0.549	0.097	0.499	0.094	0.500
		400	0.118	0.466	0.153	0.433	0.150	0.435
		600	0.149	0.433	0.180	0.407	0.177	0.408

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기, CI.U=신뢰구간 상한, CI.L=신뢰구간 하한

<표 97> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(1)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Uniform	15	50	0.530	0.547	0.534
		200	0.258	0.261	0.259
		400	0.183	0.185	0.183
		600	0.148	0.150	0.149
	25	50	0.565	0.582	0.567
		200	0.276	0.279	0.277
		400	0.195	0.197	0.196
		600	0.159	0.161	0.160
	50	50	0.665	0.692	0.664
		200	0.340	0.339	0.338
		400	0.236	0.235	0.234
		600	0.193	0.192	0.192
	75	50	0.193	0.192	0.192
		200	0.489	0.487	0.484
		400	0.343	0.340	0.339
		600	0.298	0.295	0.295

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 98> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(2)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Mod	15	50	0.529	0.544	0.532
		200	0.260	0.262	0.260
		400	0.184	0.185	0.184
		600	0.149	0.150	0.149
	25	50	0.556	0.571	0.556
		200	0.276	0.277	0.275
		400	0.193	0.194	0.193
		600	0.158	0.158	0.157
	50	50	0.680	0.697	0.672
		200	0.327	0.323	0.322
		400	0.236	0.233	0.233
		600	0.193	0.191	0.191
	75	50	1.070	0.149	0.986
		200	0.488	0.480	0.477
		400	0.342	0.336	0.335
		600	0.283	0.277	0.277

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 99> Y와 보조변수 간 상관이 .10일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(3)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Low	15	50	0.529	0.546	0.535
		200	0.258	0.261	0.259
		400	0.183	0.185	0.183
		600	0.148	0.150	0.149
	25	50	0.564	0.581	0.568
		200	0.276	0.278	0.277
		400	0.195	0.197	0.196
		600	0.157	0.159	0.158
	50	50	0.665	0.690	0.665
		200	0.340	0.339	0.338
		400	0.236	0.235	0.234
		600	0.193	0.192	0.192
	75	50	0.155	NA	0.085
		200	0.488	0.486	0.484
		400	0.343	0.339	0.339
		600	0.284	0.281	0.281

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 100> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(1)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Uniform	15	50	0.529	0.541	0.530
		200	0.257	0.259	0.257
		400	0.182	0.183	0.182
		600	0.148	0.149	0.148
	25	50	0.561	0.574	0.560
		200	0.276	0.277	0.275
		400	0.195	0.195	0.194
		600	0.159	0.159	0.159
	50	50	0.667	0.679	0.653
		200	0.339	0.332	0.331
		400	0.238	0.232	0.232
		600	0.193	0.188	0.188
	75	50	1.074	0.140	0.967
		200	0.489	0.472	0.470
		400	0.343	0.329	0.329
		600	0.283	0.272	0.272

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 101> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(2)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Mod	15	50	0.532	0.535	0.526
		200	0.258	0.257	0.255
		400	0.183	0.182	0.180
		600	0.148	0.147	0.146
	25	50	0.553	0.554	0.543
		200	0.274	0.270	0.268
		400	0.192	0.189	0.188
		600	0.157	0.154	0.153
	50	50	0.625	0.617	0.600
		200	0.325	0.309	0.308
		400	0.235	0.222	0.222
		600	0.193	0.182	0.182
	75	50	1.081	0.136	0.942
		200	0.488	0.446	0.445
		400	0.343	0.312	0.312
		600	0.283	0.257	0.257

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 102> Y와 보조변수 간 상관이 .20일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(3)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Low	15	50	0.524	0.537	0.526
		200	0.257	0.259	0.257
		400	0.182	0.183	0.181
		600	0.148	0.148	0.147
	25	50	0.560	0.572	0.559
		200	0.275	0.276	0.274
		400	0.194	0.194	0.193
		600	0.156	0.156	0.155
	50	50	0.667	0.676	0.653
		200	0.339	0.331	0.331
		400	0.235	0.229	0.229
		600	0.193	0.188	0.188
	75	50	0.157	0.001	0.096
		200	0.489	0.470	0.469
		400	0.343	0.328	0.329
		600	0.283	0.271	0.271

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 103> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(1)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Uniform	15	50	0.527	0.535	0.524
		200	0.255	0.256	0.255
		400	0.180	0.181	0.180
		600	0.146	0.146	0.146
	25	50	0.580	0.578	0.564
		200	0.273	0.271	0.269
		400	0.194	0.191	0.190
		600	0.158	0.156	0.155
	50	50	0.662	0.651	0.627
		200	0.335	0.316	0.316
		400	0.238	0.224	0.224
		600	0.194	0.183	0.183
	75	50	1.089	0.106	0.917
		200	0.498	0.453	0.451
		400	0.347	0.315	0.315
		600	0.286	0.259	0.258

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 104> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(2)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Mod	15	50	0.526	0.510	0.505
		200	0.256	0.248	0.248
		400	0.181	0.176	0.175
		600	0.146	0.142	0.142
	25	50	0.569	0.537	0.532
		200	0.271	0.257	0.256
		400	0.191	0.181	0.181
		600	0.155	0.147	0.147
	50	50	0.650	0.585	0.576
		200	0.324	0.284	0.284
		400	0.238	0.205	0.205
		600	0.194	0.167	0.167
	75	50	1.117	0.085	0.778
		200	0.496	0.388	0.390
		400	0.349	0.272	0.273
		600	0.284	0.221	0.222

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 105> Y와 보조변수 간 상관이 .30일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(3)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Low	15	50	0.527	0.532	0.521
		200	0.255	0.256	0.253
		400	0.180	0.180	0.179
		600	0.148	0.148	0.147
	25	50	0.589	0.582	0.570
		200	0.273	0.270	0.268
		400	0.194	0.191	0.189
		600	0.155	0.154	0.152
	50	50	0.661	0.646	0.625
		200	0.335	0.314	0.315
		400	0.238	0.223	0.223
		600	0.194	0.182	0.182
	75	50	0.167	0.001	0.097
		200	0.496	0.447	0.448
		400	0.349	0.313	0.315
		600	0.286	0.256	0.258

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 106> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(1)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Uniform	15	50	0.523	0.521	0.512
		200	0.251	0.25	0.249
		400	0.177	0.177	0.176
		600	0.144	0.144	0.143
	25	50	0.547	0.536	0.526
		200	0.27	0.262	0.261
		400	0.192	0.186	0.185
		600	0.156	0.151	0.151
	50	50	0.657	0.612	0.593
		200	0.333	0.298	0.298
		400	0.238	0.212	0.212
		600	0.193	0.172	0.172
	75	50	0.284	0.233	0.233
		200	0.498	0.411	0.41
		400	0.349	0.287	0.287
		600	0.284	0.233	0.233

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 107> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(2)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Mod	15	50	0.523	0.472	0.475
		200	0.252	0.236	0.237
		400	0.178	0.167	0.167
		600	0.144	0.136	0.136
	25	50	0.556	0.571	0.556
		200	0.276	0.277	0.275
		400	0.189	0.168	0.168
		600	0.153	0.137	0.137
	50	50	0.651	0.498	0.505
		200	0.322	0.244	0.246
		400	0.238	0.173	0.174
		600	0.193	0.141	0.142
	75	50	1.112	0.034	0.546
		200	0.495	0.269	0.276
		400	0.348	0.189	0.192
		600	0.283	0.154	0.156

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

<표 108> Y와 보조변수 간 상관이 .40일 때, 조건별  $\rho_{XY}(.30)$ 의 90% 신뢰구간의 너비(3)

$\rho_{AA}$	Mis(%)	N	No-AUX	Inclusive	PCA
Low	15	50	0.526	0.518	0.152
		200	0.252	0.250	0.247
		400	0.178	0.176	0.174
		600	0.144	0.143	0.141
	25	50	0.547	0.531	0.521
		200	0.270	0.260	0.258
		400	0.192	0.184	0.183
		600	0.153	0.148	0.147
	50	50	0.657	0.603	0.588
		200	0.333	0.294	0.296
		400	0.238	0.209	0.210
		600	0.193	0.170	0.171
	75	50	0.170	0.001	0.081
		200	0.498	0.402	0.406
		400	0.349	0.281	0.285
		600	0.284	0.228	0.231

주.  $\rho_{AA}$ =보조변수 간 상관의 형태, Mis(%)=결측치의 비율, N=표본의 크기

## <부록 2> 시뮬레이션을 위한 R 코드

```
#package
library(MASS)
library(lavaan)
library(MplusAutomation)
library(semTools)
library(factoextra)
library(ggplot2)
library(caret)
library(dplyr)

#result data frame
result <- matrix(nrow=500, ncol=15)
result <- as.data.frame(result)
names(result) <- c("reg.Y.fiml", "Y.fiml", "Y.var.fiml", "CI_L.fiml",
"CI_U.fiml", "reg.Y.aux", "Y.aux", "Y.var.aux", "CI_L.aux", "CI_U.aux",
"reg.Y.pca", "Y.pca", "Y.var.pca", "CI_L.pca", "CI_U.pca")

#loop
for(i in 1:500)
{
  assign( paste("data name", i, sep=""),
```

```

data <- read.table(paste("data name",i, "dat", sep=".") )
names(data) <- c("Y", "A1", "A2", "A3", "A4", "A5", "A6", "A7",
"A8", "A9", "X")

#missing data code change
data[data==999]<-NA

#PCA
pca <- data[, 2:10]
PCAmodel <- princomp(pca, cor = FALSE, scores = TRUE)
summary(PCAmodel)

#export PC: factoextra package
res.pca <- get_pca_ind(PCAmodel)
PC <- res.pca$coord
PC <- as.data.frame(PC)

#Cumulative Proportion
CP <- preProcess(pca, method="pca", thresh=0.4)
summary(CP)
NumOfComp <- CP[["numComp"]]
NumOfPC <- as.data.frame(NumOfComp)

#PCA Data Generation
PCAUX <- PC[, c(1:NumOfComp)]

```

```
XY<- data[, c(1,11)]
PCAregr <- cbind(XY, PCAUX)

#auxiliary analysis
model<-"Y~X"
aux.vars <-c("A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8", "A9")
pcaux.vars <- colnames(PCAUX)

#fiml regression
fimlfit <- sem(model = model, data = data, missing="fiml")
summaryfiml <- summary(fimlfit)
PE.fiml <- parameterestimates(fimlfit, level=.90)

#aux_regression
auxfit <- sem.auxiliary(model = model, aux = aux.vars, data = data,
missing="fiml")
summaryAUX <- summary(auxfit)
PE.aux <- parameterestimates(auxfit, level=.90)

#pca_regression
pcafit <- sem.auxiliary(model = model, aux = pcaux.vars, data =
PCAregr, missing="fiml")
summaryPCAUX <- summary(pcafit)
PE.pca <- parameterestimates(pcafit, level=.90)
```

```
#Y ~ X
reg.Y.fiml <- summaryfiml$PE$est[1]
reg.Y.aux <- summaryAUX$PE$est[1]
reg.Y.pca <- summaryPCAUX$PE$est[1]

#estimated Y
Y.fiml <- summaryfiml$PE$est[4]
Y.aux <- summaryAUX$PE$est[4]
Y.pca <- summaryPCAUX$PE$est[4]

#Var Y
Y.var.fiml <- summaryfiml$PE$est[2]
Y.var.aux <- summaryAUX$PE$est[2]
Y.var.pca <- summaryPCAUX$PE$est[2]

#Confidence Interval
CL_L.fiml <- PE.fiml$ci.lower[1]
CL_U.fiml <- PE.fiml$ci.upper[1]
CL_L.aux <- PE.aux$ci.lower[1]
CL_U.aux <- PE.aux$ci.upper[1]
CL_L.PCA <- PE.pca$ci.lower[1]
CL_U.PCA <- PE.pca$ci.upper[1]

#result save
result[i,] <-c(reg.Y.fiml, Y.fiml, Y.var.fiml, CL_L.fiml, CL_U.fiml,
```

reg.Y.aux, Y.aux, Y.var.aux, CI\_L.aux, CI\_U.aux,  
reg.Y.pca, Y.pca, Y.var.pca, CI\_L.PCA, CI\_U.PCA)

}