



불균등한 시간 간격을 고려하는 역동적 구조방정식 모형 비교 검증

Comparison of DSEM Approaches Considering Unequally Spaced Time Intervals

류호성, 전남대학교 심리학과 석사과정

a24801624@gmail.com

1. 서론

최근 기술에 발달로 인해 집중 종단 자료(Intensive Longitudinal Data; 이하 ILD)의 수집이 용이해졌다. 이에 따라 사회과학 분야에서는 ILD를 분석하는 방법론에 대한 관심과 수요가 증가하였으며, 그 중 역동적 구조방정식(Dynamic Strcutrual Equation Modeling; 이하 DSEM)이 큰 주목을 받고 있다.

일반적인 DSEM 모형들은 관측값의 시간 간격이 일정하게 측정되는 것을 가정한다. 그러나, 사회과학 분야에서는 생태순간평가(EMA)와 경험 표집법(ESM)등을 통해 데이터를 수집하는 경우처럼 관측값의 시간 간격이 동일하지 않는 경우가 많다. 이를 무시하고 분석할 경우 모수 추정이 불안정해질 수 있기에 불균등한 시간 간격을 고려하는 다양한 방법들이 존재한다.

대표적으로 Asparouhov, Hamaker & Muthén (2018)이 시행한 바와 같이 관측값들의 연속적인 시간 간격을 이산적인 시간 간격으로 재조정하는 메커니즘이 있으며, 이는 Mplus에서 TINTERVAL 옵션을 통해 이루어진다. 하지만 적절한 재조정 시간 간격을 설정해야 한다는 점과 데이터가 측정된 실질적인 시간을 반영하지 못한다는 한계가 있다.

또 다른 방법으로는 데이터가 측정된 시간 자체를 분석에 반영하는 연속 시간 모델링(Continuous Time Modeling)이 있다. 그 중에서도 CT-RDSEM의 경우, within-level에서 시간에 따른 동적 구조를 잔차 수준에서 모델링하여 잔차 간 자기회귀관계를 드리프트 계수(drift parameter; 이하 d)를 통해 설명한다. d 는 연속적인 시간 간격에 따라 잔차 간 자기상관이 얼마나 빠르게 감소하는 지를 나타내며, 이산적 시간 간격에서 자기회귀계수 r 과 다음과 같은 식을 통해 서로 변환이 가능하다. $r(\Delta t)=EXP(d \cdot \Delta t)$. CT-RDSEM을 사용하여 연속 시간 간격 데이터를 분석할 경우 관측값 사이 시간을 재조정하지 않고 실제 간격을 반영하며, 불필요한 결측값이 발생하지 않게 되어 추정 과정에서 이점이 생긴다.

CT-RDSEM을 사용하여 불균등한 시간 간격을 고려한 시뮬레이션 연구는 Asparouhov & Muthén (2024)이 시행한 바 있으나, 해당 연구는 일부 제한된 조건에서 모형의 성능을 평가하였다는 한계가 있다.

이에 본 연구는 DSEM에서 불균등한 시간 간격을 고려하는 방법에 관하여 다양한 조건을 설정하여 시뮬레이션 연구를 수행하였다. 모형은 two-level AR(1) regression 모형에 초점을 맞추어 연구를 수행하였다. 연구는 불균등한 시간 간격을 가진 데이터를 생성한 후 세 가지 모형(CT-RDSEM, RDSEM(TIN=1), RDSEM(TIN=0.5))의 추정 성능을 비교하였다.

본 연구에서 설정한 연구 문제는 두 가지 이다.

- 불균등한 시간 간격을 고려한 DSEM 모형들의 모수 추정치의 편향 정도는 어떠한가?
- 다양한 시뮬레이션 조건에서 모델링할 때 세 모형의 추정 성능은 어떻게 달라지는가?

2. 연구방법

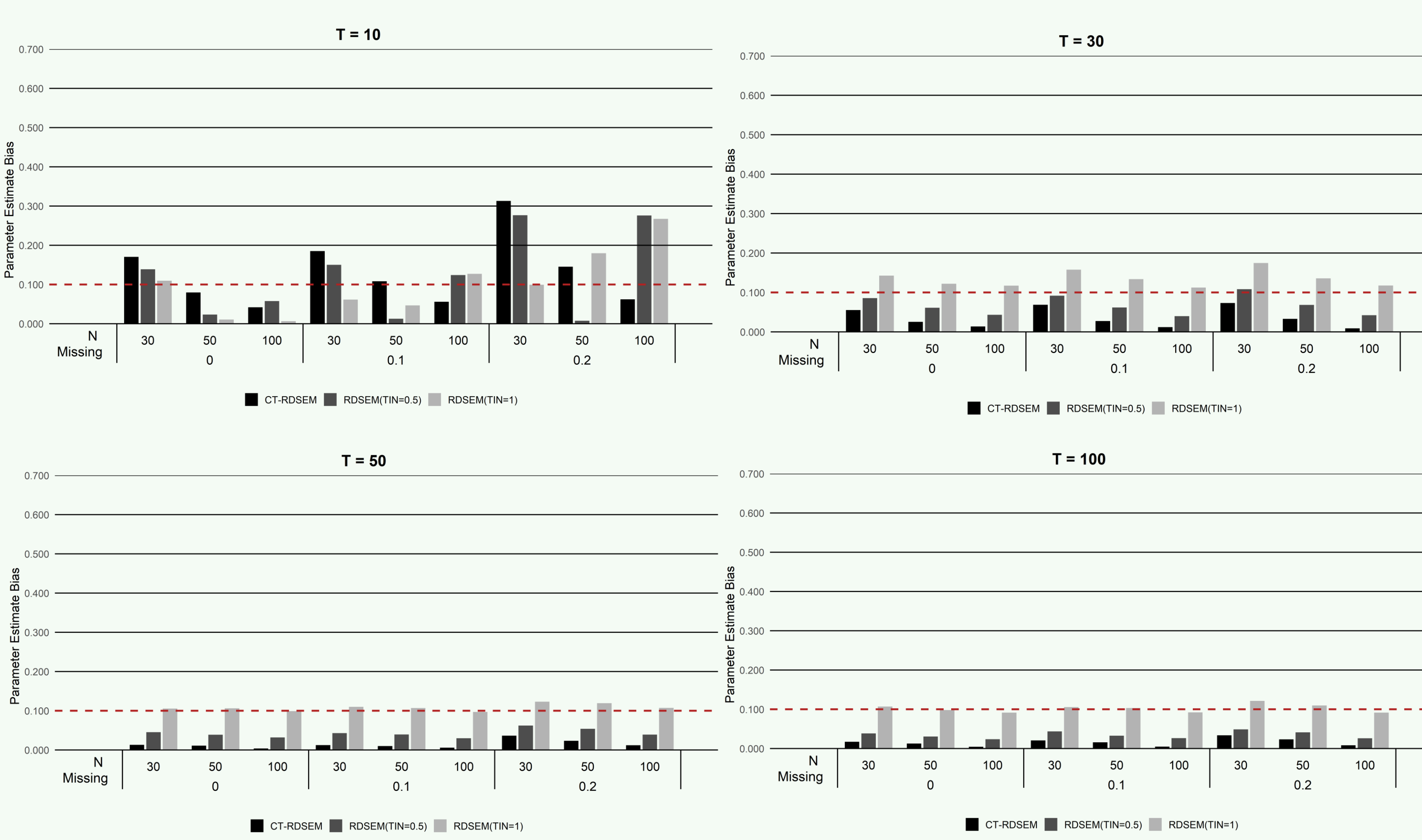
시뮬레이션 조건은 표본 수(3가지), 측정 시점 수(4가지), 결측율(3가지)으로 설정하 였다. 따라서 총 조건의 수는 $3 \times 4 \times 3 = 36$ 개이며, 각 조건 당 100개의 자료를 생성하였다. 각 모형의 추정 성능은 Bias, 95% Coverage, MSE, DIC로 판단하였다. 데이터는 CT-RDSEM 모형을 기반으로 연속 시간 간격을 가진 자료를 생성하였으며, 데이터의 생성과 분석은 Mplus 8.11과 R의 Mplusautomation 패키지를 사용하였다.

| Model Settings | |
|------------------------------|--|
| Drift parameter of X, Y | -0.7 |
| Regression of Y on X | (Within-level) 1 (Between-level) -1 |
| Mean of intercepts (X, Y) | 3 |
| Variance of intercepts (X,Y) | 1 |

| Simultion Conditions | |
|----------------------|-----------------|
| Number of N | 30, 50, 100 |
| Number of T | 10, 30, 50, 100 |
| Missing rate | 0, 0.1, 0.2 |

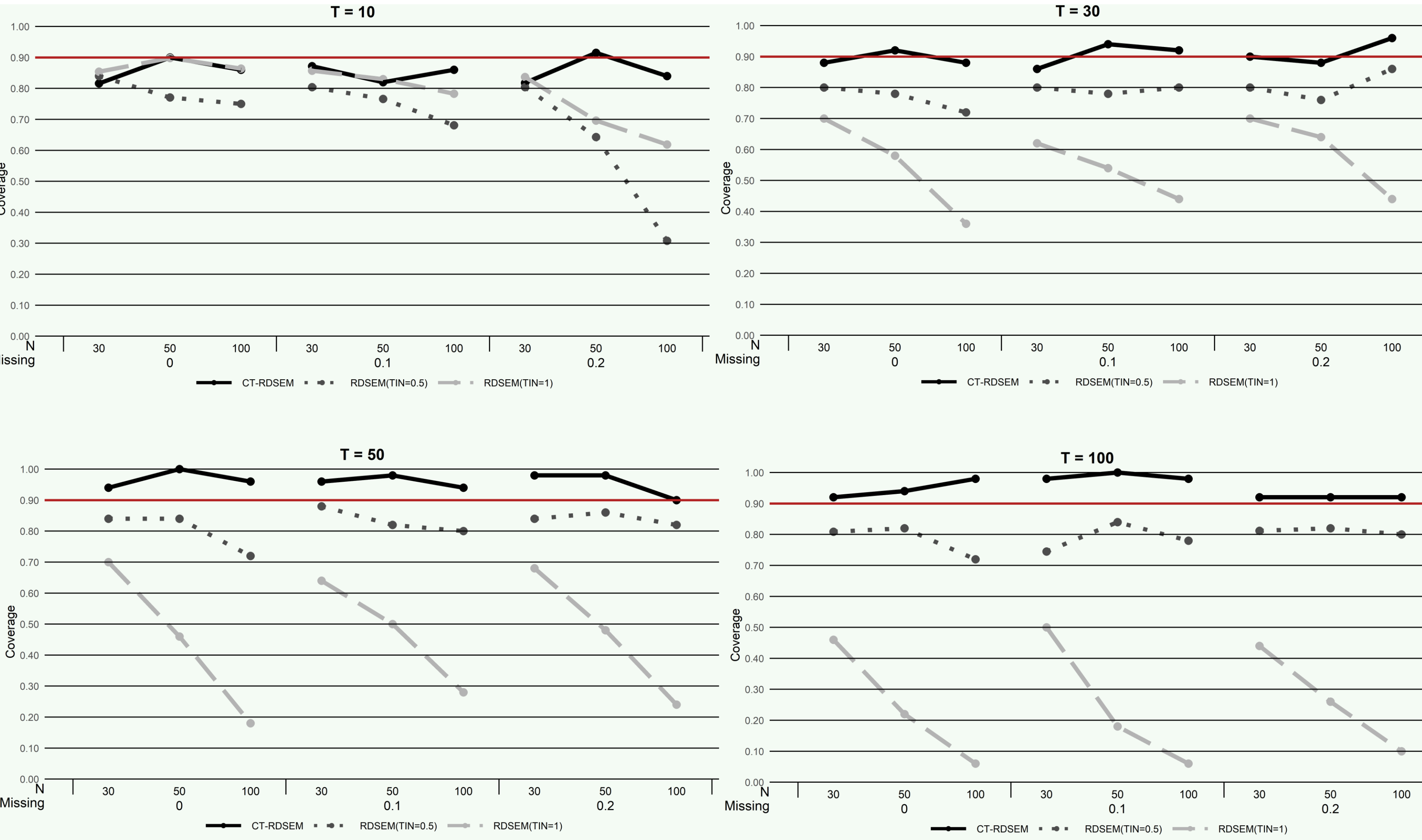
3. 연구결과

● Parameter Estimate Bias



종속변수 Y의 Drift paramete의 bias로, 추정이 정확할수록 bias는 0에 가까우며, 0.1이내에 있을 경우 수용가능한 범위에 있는 것을 의미한다 (Muthén & Muthén, 2002)

● Coverage



종속변수 Y의 Drift parameter의 95% corverage로, 0.90~0.98이내에 있을 경우 수용가능한 범위에 있는 것을 의미한다 (Muthén & Muthén, 2002)

4. 결론 및 실용적 제언

본 연구는 ILD를 분석할 때 연속 시간 간격을 고려하는 이산 시간 간격 재조정 방식과 연속 시간을 반영하는 CT-RDSEM의 모수 추정 정확도를 다양한 조건 속에서 비교하였다는 점에서 의의가 있다. 본 연구 결과, TINTERVAL을 사용하는 RDSEM보다 CT-RDSEM이 전반적으로 모수 추정의 편향이 적었다. 또한 결측율이 높고, N이 많을수록 CT-RDSEM이 비교적 정확하게 추정하였다. 본 연구 결과를 바탕으로 한 제언은 다음과 같다. 첫째, 본 연구에서 활용한 모형의 경우 정확한 추정을 위해서는 $N=30$, $T=50$ 은 확보되어야 한다. 둘째, 수집된 데이터의 시간 간격이 일정하지 않고 불균등한 분포를 이루고 있다면 연속 시간 모델링을 활용한 CT-RDSEM이 권장된다. 단, 연구자는 CT-RDSEM과 RDSEM은 변환가능한 관계에 있다는 점에서 각 모형의 특징 (추정에 걸리는 속도, CT-RDSEM의 제한점 등)을 고려하여 선택하여야 한다.

References

1)Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. Structural equation modeling: a multidisciplinary journal, 25(3), 359–388.
2)Asparouhov, T., & Muthén, B. (2024). Continuous time dynamic structural equation models. Technical report, <https://www.statmodel.com/download/CTRDSEM.pdf>.
3)Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural equation modeling, 9(4), 599–620.