

# 测量时间与样本量对参数估计的影响： 基于密集追踪数据分析的方法比较

徐浩桢

西南大学心理学部，重庆

收稿日期：2024年1月21日；录用日期：2024年3月20日；发布日期：2024年3月31日

## 摘要

在心理学实验的进行过程中，被试的情绪、想法、行为和生理功能，往往不是保持不变，而是随时间变化一直波动的。传统的追踪或是横断研究能够对群体中存在的现象进行描述解释，但难以解释个体短时间内心理过程的动态变化。密集追踪是在短时间内对个体进行多次测量的方法，测得的数据更利于探究个体在实验过程中心理动态变化的过程作用机制。近年来，随着科学技术的发展，密集追踪测量的难度降低，目前已经成为心理学研究的一大热点。目前针对密集追踪研究的数据分析主要有传统的多层线性模型(Multilevel Modeling, MLM)的方法，以及新兴的动态结构方程模型(Dynamic Structural Equation Modeling, DSEM)的分析方法。二者均可以方便地对密集追踪数据中的自回归和交叉滞后效应进行建模。但目前尚未有研究探讨过时间点和样本量对于这两个模型参数估计的影响，因此本研究以模拟研究的形式，比较在不同时间点和样本量条件下两个模型参数估计的优劣，为研究者在实际研究中选择和使用模型提出建议。

## 关键词

密集追踪，时间序列，动态结构方程模型

# The Influence of Measurement Time and Sample Size on Parameter Estimation: Comparison of Methods Based on Intensive Tracking Data Analysis

Haozhen Xu

Faculty of Psychology, Southwest University, Chongqing

Received: Jan. 21<sup>st</sup>, 2024; accepted: Mar. 20<sup>th</sup>, 2024; published: Mar. 31<sup>st</sup>, 2024

## Abstract

In the process of conducting psychological experiments, the emotions, thoughts, behaviors, and physiological functions of the subjects often do not remain unchanged, but fluctuate continuously over time. Traditional tracking or cross-sectional research can describe and explain phenomena that exist within a group, but it is difficult to explain the dynamic changes in individual psychological processes over a short period of time. Intensive tracking is a method of measuring individuals multiple times in a short period of time, and the measured data is more conducive to exploring the process and mechanism of psychological dynamic changes in individuals during the experimental process. In recent years, with the development of science and technology, the difficulty of intensive tracking and measurement has decreased, and it has become a hot topic in psychological research. At present, the data analysis for intensive tracking research mainly includes traditional Multilevel Modeling (MLM) methods and emerging dynamic structural equation modeling (DSEM) analysis methods. Both can conveniently model autoregressive and cross lagged effects in intensive tracking data. However, there is currently no research exploring the impact of time points and sample size on the parameter estimation of these two models. Therefore, this study compares the advantages and disadvantages of parameter estimation of the two models under different time points and sample size conditions in the form of simulation research, and provides suggestions for researchers to choose and use models in practical research.

## Keywords

Intensive Tracking, Time Series, Dynamic Structural Equation Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在心理学实验的进行过程中, 被试的情绪、想法、行为和生理功能, 往往不是保持不变, 而是随时间变化一直波动的(Vallacher et al., 2002)。传统的追踪或是横断研究往往只能描述或解释被试群体中较为稳定的, 在短时间内恒定的变量间的关系, 难以揭示事件对于情绪、欲望等随时间动态变化的变量的持续性影响(Setodji et al., 2019)。这可能导致我们在研究过程中, 将变量由时间推移所导致的动态变化, 归因于实验事件对于变量的影响, 进而导致研究者对于干预效果及其潜在机制产生错误解读(张银普等, 2016; Setodji et al., 2019)。因此, 为了更深入地探讨变量变化的原因到底是因为时间推移, 还是受到实验事件的影响, 研究者往往需要采用密集追踪测量以得出个体状态变量随时间推移动态变化的细节(Trull & Ebner-Priemer, 2013)。

密集追踪是在短时间内对于个体进行多次测量的方法, 测得的数据更有利于我们探究个体在实验过程中心理动态变化的过程及其作用机制。近年来, 随着手机等通讯设备的发展, 数据采集的难度和成本大大降低, 研究者可以更加频繁地对被试生理和心理状态等方面的数据进行收集; 随着数据采集越来越便捷, 场景越来越多样, 这些都有利于研究者通过密集追踪的方式探索心理随时间变化的动态过程。如今在心理学的各个领域的研究中都有使用密集追踪数据(Chen & Zhang, 2020; Windt et al., 2018), 如, 成瘾行为(Lanza et al., 2014; Shiffman, 2009; Weinstock et al., 2017)、压力症状(Chun, 2016)、

心理资本(Malmberg & Martin, 2019; Sturgeon et al., 2014)饮食障碍问题(Munsch et al., 2009)、亲密关系(Howland & Rafaeli, 2010)、人格障碍(Trull et al., 2008)和语言使用(Lazarević et al., 2020; 安媛媛, 徐慰, 2017)等。

目前针对密集追踪研究的数据分析方法主要有传统的多层线性模型(Multilevel Modeling, MLM)的方法, 以及新兴的动态结构方程模型(Dynamic Structural Equation Modeling, DSEM)的分析方法。二者均可以方便地对密集追踪数据中的自回归和交叉滞后效应进行建模(刘源等, 2022)。但目前尚未有研究探讨过时间点和样本量对于这两个模型中参数估计的影响, 也未有研究比较在不同时间点和样本量条件下两个模型参数估计的优劣, 为研究者在实际研究中选择和使用模型提出建议。为了更直观地比较 MLM、DSEM 在不同时间点和样本量条件下的估计表现, 本研究利用这两种模型方法在不同时间点和样本量条件下进行模拟研究和比较, 并对这两种模型方法的优缺点和适用范围进行总结, 帮助研究者针对自身的实际研究情况灵活选择适合的模型方法。

## 2. 文献综述

### 2.1. 密集追踪数据的分析方法

#### 2.1.1. 多层线性模型

以两水平的多层线性模型为例, 在利用多层线性模型分析纵向数据时, 以不同时间点下的追踪结果为第一层数据, 以不随时间变化的稳定的个体特征或所接受的实验处理为第二层数据, 形成两层数据结构(Raudenbush & Bryk, 2002)。以存在随机自回归效应的两水平数据为例, 个体  $i$  在时间  $t$  的方程表达式如下:

$$Y_{it} = \beta_{oi} + \beta_{1i}Y_{it-1} + \varepsilon_{it} \quad (1)$$

在方程(1)中,  $\beta_{oi}$  代表截距,  $\beta_{1i}$  表示斜率,  $X_{it}$  表示第  $i$  个被试在第  $t$  次测量时自变量的水平,  $\varepsilon_{it}$  表示残差, 代表第  $i$  个被试在第  $t$  次测量中因变量不能被自变量所解释的部分。方程(1)与一般的回归方程很相似, 区别在于(1)中的截距和斜率都不是定值, 而是随机变量, 他们的取值是由第二水平的变量所决定的。常见的第二水平的变量有个体特征和实验处理, 以他们为自变量建立两个第二水平的回归方程, 方程表达式如下:

$$\beta_{oi} = \gamma_{oo} + \gamma_{o1}W_{1i} + \mu_{oi} \quad (2)$$

$$\beta_{1i} = \gamma_{1o} + \gamma_{11}W_{1i} + \mu_{1i} \quad (3)$$

其中,  $\gamma_{oo}$  是方程(2)的截距, 是个体水平的自变量  $W_1$  为 0 时因变量  $Y$  的平均值。 $\beta_{1i}$  是方程(2)中自变量  $W_1$  的回归系数, 表示因变量  $Y$  自回归效应的强弱。 $\mu_{oi}$  是方程(2)中的残差, 是个体水平上因变量  $Y$  不被自变量  $W_1$  解释的部分,  $\gamma_{1o}$  是方程(3)的截距, 是自变量  $W_1$  为 0 时因变量自回归效应的均值。 $\gamma_{11}$  是方程(3)中自变量  $W_1$  的回归系数, 可以被理解为自变量  $W_1$  对因变量自回归效应的影响大小。 $\mu_{1i}$  是方程(3)中的残差, 是因变量  $Y$  的自回归效应中未被  $W_1$  解释的部分。

在使用多层线性模型分析纵向数据时, 假设不同个体在水平 1 估计的残差方差是一致的  $\varepsilon_{it} \sim (0, \sigma^2)$ , 但对某些变量来说, 不同个体间在短期内的波动轨迹有较大差异(例如情绪)。此时多层线性模型的假设可能导致参数有偏估计(Jahng & Wood, 2017)。因此当研究问题关注的是个体心理动态变化过程的变异性和不稳定性时, 有可能会得到不正确的参数估计(Jahng, 2008), 影响研究者对变量干预效果和机制的探索(Jahng, 2008)。其次, MLM 通常假设随机效应服从一定的正态分布, 但在实际的研究中, 这一假设往往不能得到满足, 因此以这一假设为前提得到的随机效应的估计参数往往也是不准确的(Piccirillo & Rode-

baugh, 2019)。此外,在多层线性模型当中,我们往往需要对数据进行中心化,以此来分解组间效应和组内效应(Curran & Bauer, 2011)。传统的中心化方法有个体均值中心化、总均值中心化等。但在一些情况下,比如数据内部存在自回归效应(Nickell, 1981)或情境效应(Contextual Effects)时,使用个体均值中心化或是总均值中心化处理数据往往会产生较大的估计偏差(Asparouhov & Muthén, 2019)。

### 2.1.2. 动态结构方程模型

基于传统的多层线性模型在密集型追踪数据分析过程中存在的诸多问题,研究者们提出了新的统计分析方法,动态结构方程模型(Dynamic Structural Equation Modeling, DSEM)。DSEM 模型最初由于计算机技术的限制没有得到广泛应用,直到研究者使用 Mplus 软件进行 DSEM 模型的研究给予了指导。

DSEM 模型包含三个子模型,最一般的模型是交叉分类的 DSEM 模型,它包含了个体和特定时间的随机效应。第二个模型是两级 DSEM 模型,它只包含个体特异的随机效应。两级 DSEM 模型也是在现有研究中被最多使用,应用价值最大的一个模型,因为相比交叉分类的 DSEM 模型,他更容易估计、识别和解释,相比单级 DSEM 模型它可以从数据中获取更多的信息。第三个模型是针对来自单个个体的时间序列数据的单级 DSEM 模型(Zhang & Nesselroade, 2007)。在后一种模型中,没有随机效应。在这里,我们描述了最一般的交叉分类的 DSEM 模型。两级和单级 DSEM 模型是交叉分类 DSEM 模型的特殊情况。

完整的 DSEM 模型会将观测分数分解为 3 个部分,其表达式如下:

$$Y_{it} = Y_{1,it} + Y_{2,i} + Y_{3,t} \quad (4)$$

其中,  $Y_{it}$  表示个体  $i$  在  $t$  时刻的观测分数,  $Y_{2,i}$  表示的是在个体水平上个体  $i$  对于观测分数的特异贡献,  $Y_{3,t}$  表示的是在时间水平上时间  $t$  对于观测分数的特异贡献,  $Y_{1,it}$  是在观测分数去除了个体水平的特异贡献  $Y_{2,i}$  和时间水平的特异贡献  $Y_{3,t}$  后的剩余部分。两层的 DSEM 模型只包含前 2 个部分  $Y_{1,it}$  和  $Y_{2,i}$ , 不考虑时间  $t$  对于观测分数的特异贡献,而单层 DSEM 模型只包含  $Y_{1,it}$ , 个体  $i$  和时间  $t$  对于观测分数的特异贡献都不考虑。这 3 个部分都是满足正态分布的潜在随机向量,被用于形成三组结构方程模型。

根据刚才的分解,我们可以得到一个第一水平的变量  $Y_{1,it}$  和两个第二水平的变量,分别是个体  $i$  的特异贡献  $Y_{2,i}$  和时间  $t$  的特异贡献  $Y_{3,t}$ , 他们的方程表达式如下:

$$Y_{1,it} = v_1 + \sum_{l=0}^L \lambda_{1,l} \eta_{1,i,t-l} + \sum_{l=0}^L R_l Y_{1,i,t-l} + \sum_{l=0}^L K_{1,l} X_{1,i,t-l} + \varepsilon_{1,it} \quad (5)$$

$$\eta_{1,it} = \alpha_1 + \sum_{l=0}^L B_{1,l} \eta_{1,i,t-l} + \sum_{l=0}^L Q_l Y_{1,i,t-l} + \sum_{l=0}^L \Gamma_{1,l} X_{1,i,t-l} + \xi_{1,it} \quad (6)$$

$$Y_{2,i} = v_2 + \lambda_2 \eta_{2,i} + K_2 X_{2,i} + \varepsilon_{2,i} \quad (7)$$

$$\eta_{2,i} = \alpha_2 + B_2 \eta_{2,i} + K_2 X_{2,i} + \xi_{2,i} \quad (8)$$

$$Y_{3,t} = v_3 + \lambda_3 \eta_{3,t} + K_3 X_{3,t} + \varepsilon_{3,t} \quad (9)$$

$$\eta_{3,t} = \alpha_3 + B_3 \eta_{3,t} + K_3 X_{3,t} + \xi_{3,t} \quad (10)$$

其中,方程(4)和(5)是第一水平变量  $Y_{1,it}$  的表达式,  $\eta_{1,it}$  表示  $Y_{1,it}$  对应的潜变量,  $\Lambda$  为因子负载矩阵,  $R$ 、 $K$ 、 $B$ 、 $Q$ 、 $\Gamma$  表示回归系数矩阵。  $Y_{2,i}$  是个体特异的时不变协变量的向量,  $X_{3,t}$  是时间特异而个体不变的协变量的向量。同样的,  $\eta_{2,i}$  和  $\eta_{3,t}$  是个体和时间特异的潜变量。剩余的  $\varepsilon_{1,it}$ ,  $\xi_{1,it}$ ,  $\varepsilon_{2,i}$ ,  $\xi_{2,i}$ ,  $\varepsilon_{3,t}$ ,  $\xi_{3,t}$  代表均值为 0 的残差。

具体来说, DSEM 在第一水平当中的各个参数的随机效应(比如因子载荷和截距项等)均可以在第二水平和第三水平用个体特异和时间特异的变量进行解释, DSEM 模型可以同时考虑不同个体  $i$  和时间点  $t$



的特异贡献。与此同时，DSEM 模型还允许研究者对于残差方差的随机效应进行建模，残差方差代表的是因变量短时间内的波动大小。

相比传统的分析方法，DSEM 在密集追踪数据的分析上存在很多优势。首先，DSEM 是多种模型方法的结合，能更便捷、全面地分析模型中潜变量与显变量，潜变量与潜变量，显变量与显变量之间的关系(Hamaker et al., 2018; Zhou et al., 2021)。与此同时它不仅能够对两变量的模型进行建模分析，三个及以上的变量也可以用 DSEM 建模分析。在各个时间点测量的观测变量，均可以被分解为时间特异的部分，个体特异的部分，他们之间也都可以通过 DSEM 建立回归路径，大大提升了研究者对于纵向密集数据建模的灵活性。其次，使用 DSEM 建模可以很方便地将纵向密集数据中的特征纳入模型(比如心理过程变化的动态趋势等)，不仅可以考虑变量本身的自回归效应，还可以考虑变量间关系随时间的动态变化(Zhou et al., 2021)。另外，DSEM 对模型参数的估计采用贝叶斯模型的框架，当模型中的待估参数较多，模型较为复杂时，贝叶斯方法相比传统的频率学派更有优势(张沥今等, 2019; Zhou et al., 2021)。Song 和 Zhang 在 2014 年也曾提出多层动态因子模型(Multilevel Dynamic Factor Model, MDFM)，但后来发现当 MDFM 中存在大量的随机效应时，模型往往无法收敛。而 DSEM 解决了这一问题，研究者可以在该模型中设置任意的测量节点和随机效应，没有数量限制，即使研究中有较多的测量时间点和随机效应，对 DSEM 造成的计算量负担也相对较小。Schultzberg 和 Muthén 在 2018 年的研究中表明，DSEM 在各个测量时间点数量下，对于随机参数的均值均有一个较为良好的估计(Schultzberg & Muthén, 2018)。

然而，DSEM 也存在一些不足与局限。首先是在面对测量间隔不等距的数据时，我们选择的最小时间间隔单位  $\delta$  会对 DSEM 第一水平的参数估计产生一定的影响，不同的  $\delta$  可能会产生不同的参数估计结果(Asparouhov & Muthén, 2019)。其次，相比于传统的多层线性模型，DSEM 对于测量数据的时间点和样本量往往有更高的要求，一般至少需要 10 个以上的时间点模型才可以收敛。

## 2.2. 模型估计的影响因素

首先，样本量在大多数情况下都会影响到参数估计结果。关于积极感受和消极感受与个体抑郁之间关系的研究中，他们使用积极消极感受量表(PANA)对 101 名青年被试 104 名老年被试进行了 60 次测量。Joly-Burra et al. (2018)进行的关于健康老年人的前瞻记忆与认知抑制功能关系的研究中，使用 GO/NO GO 范式对 92 名法国老年人测量得到了共 368 个时间序列。在 2018 年进行的关于睡眠时长和儿童久坐行为的双向影响的研究中，对于 277 个家庭的孩子进行了测量，连续两周通过使孩子佩戴动力加速仪判断孩子每天久坐和睡眠的时间(Armstrong et al., 2019)。

除此之外，时间点对参数估计也会有一定影响。在 Öhrlund et al. (2019)进行的关于居民用电量与电价指定策略和关税收取策略影响的研究中，他们 212 名用户在 2 年时间内每小时的用电量，共计 7981 个时间点的用电数据。在 Mun 等(2019)进行的关于疼痛体验的个体内变化和个体间差异的研究中，研究采用了三种不同的测量模式，分别是时间偶然抽样、事件偶然抽样和混合抽样，对连续 10 天内被试的疼痛体验进行了测量。

## 3. 问题提出

在上文中我们介绍了纵向密集型数据的特点和优势，传统的密集数据分析方法 MLM，新型的分析方法 DSEM，以及这两种方法各自的优缺点。由上文可知，DSEM 模型使研究者可以更灵活地对密集纵向数据进行分析建模，但因为 DSEM 模型更为复杂，存在更多的待估参数，使用 DSEM 模型需要数据具备更多的时间节点和样本量。但目前尚未有研究探索过在不同的时间点样本量条件下，DSEM 和 MLM 参

数估计性能的优劣,对于 DSEM 模型达到相对准确的参数估计所需的时间点和样本量数量并未有研究进行过讨论。

在现有的 DSEM 实证研究中,被试样本量和时间节点数量的跨度很大。例如,在 Hamaker et al. (2018)的研究当中指出,通过每日日记的方式采集数据,连续对 100 名被试进行了 100 次施测(即  $N = 100$  和  $T = 100$ )。在 McAdams & Constantian (1983)的研究当中,对 50 个被试的亲密关系和隶属关系,进行了连续 7 天,每天 7 个随机时间点的测量(即  $N = 50$ ,  $T = 49$ )。在 Bolger & Schilling (1991)进行的研究中,对 339 名被试,连续 6 周每天测量他们的神经质和感知压力(即  $N = 339$ ,  $T = 42$ )。Shiffman & Waters (2004)进行的关于吸烟行为和复发的研究当中,通过生态瞬时评估对 215 名被试进行了 100 次数据采集。

因此本文旨在通过模拟研究的形式探索 DSEM 和 MLM 在处理纵向密集型数据时,其中的各个参数达到稳定且准确估计所需的时间节点和样本量数量。本研究将基于嵌套模型比较,同时考察包含随机残差方差的潜变量模型(DSEM)和不包含随机残差方差的模型(MLM)。

## 4. 研究设计

为了解决以上问题,本文希望通过比较在不同的时间点和样本量条件下 MLM 和 DSEM 模型各个参数估计的 MSE, SIG 和 95%覆盖率,以判断 2 个模型对于各个参数估计的稳定性和精确性,找到两个模型准确估计所需的最小时间点和样本量数量,为研究者进行密集纵向数据分析时的模型使用提供建议。

### 4.1. 模型定义

模拟研究中考虑的模型是基于 Hamaker 等人(2018)的模型。图 1 显示了在蒙特卡罗模拟中考虑的 4 种模型变化的模型图。这些模型符合路径分析的规范,与 Mplus 用户指南中的符号相一致。矩形是观测变量,圆是潜变量。从 A 到 B 的单向箭头意味着 A 影响 B,并对应一个回归斜率。一个从观察到的或潜在变量开始的单向箭头是一个具有相应方差的残差,一个填充的小圆是一个随机系数。线中间的填充圆代表的是随机斜率,单向箭头末端的填充圆代表的是随机截距。从填充圆开始的单向箭头代表的是一个随机残差方差。

在模型 1 到模型 2 中,允许三个研究者感兴趣的参数是随机的。模型 1 中包含了随机均值和随机自回归系数,模型 2 增加了一个随机残差方差。在模型 3 到 4,同样的模式重复,但添加了会影响随机均值、自回归系数的协变量  $W$ 。所有的模型表达式都嵌套在下面的 2 个第一水平和第二水平的方程表达式当中:

$$Y_{it} = \beta_{oi} + \beta_{li}Y_{it-1} + \varepsilon_{it} \quad (11)$$

$$\beta_{oi} = \gamma_{oo} + \gamma_{ol}W_{li} + \mu_{oi} \quad (12)$$

$$\beta_{li} = \gamma_{lo} + \gamma_{ll}W_{li} + \mu_{li} \quad (13)$$

其中,公式(12)代表的是第一水平,第  $i$  个被试在第  $t - 1$  个时间节点的观测变量  $Y_{it-1}$  与第  $i$  个被试在第  $t$  个时间节点的观测变量间的自回归关系,  $\beta_{oi}$  代表截距,  $\beta_{li}$  表示斜率,  $X_{it}$  表示第  $i$  个被试在第  $t$  次测量时自变量的水平,  $\varepsilon_{it}$  表示残差,代表第  $i$  个被试在第  $t$  次测量中因变量不能被自变量所解释的部分。公式(12)(13)代表的是第二水平,  $\gamma_{oo}$  是方程(2)的截距,是个体水平的自变量  $W_1$  为 0 时因变量  $Y$  的平均值。  $\beta_{li}$  是方程(2)中自变量  $W_1$  的回归系数,表示因变量  $Y$  自回归效应的强弱。  $\mu_{oi}$  是方程(2)中的残差,是个体水平上因变量  $Y$  不被自变量  $W_1$  解释的部分,  $\gamma_{lo}$  是方程(3)的截距,是自变量  $W_1$  为 0 时因变量自回归效应的

均值。 $\gamma_{11}$  是方程(3)中自变量  $W_1$  的回归系数，可以被理解为自变量  $W_1$  对因变量自回归效应的影响大小。 $\mu_{i_t}$  是方程(3)中的残差，是因变量  $Y$  的自回归效应中未被  $W_1$  解释的部分。

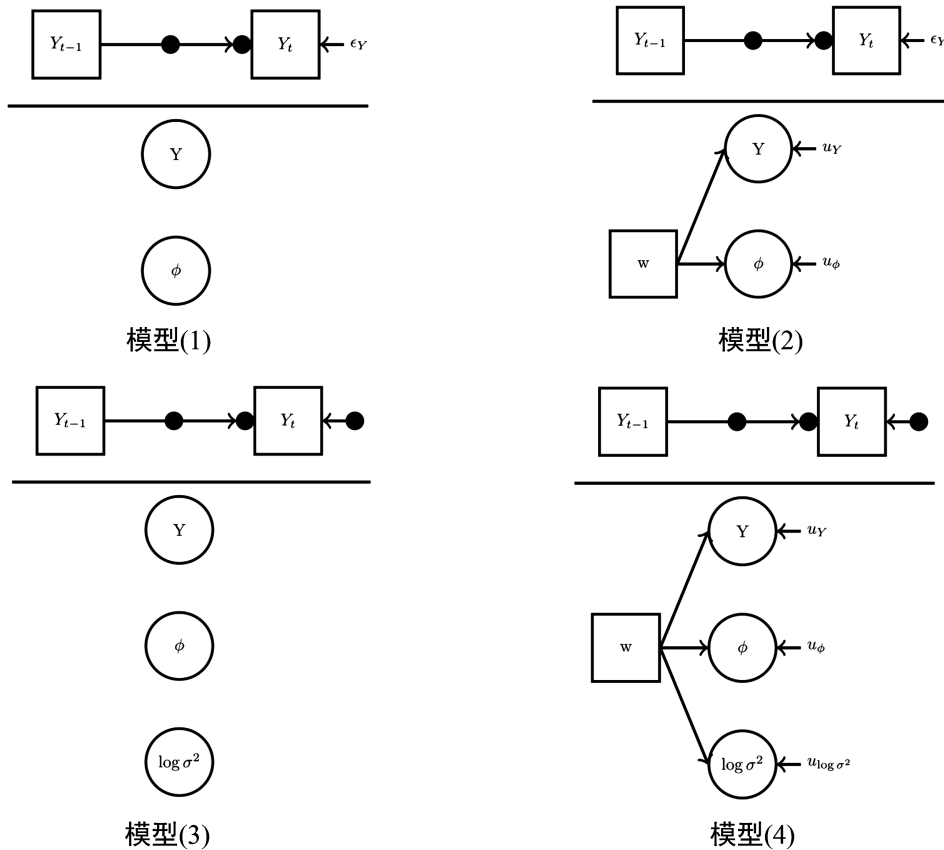


Figure 1. Diagram of DSEM and MLM models  
图 1. DSEM 和 MLM 模型示意图

## 4.2. 参数设定

在上述四个模型的基础上，本模拟研究的参数设定主要结合目前已有的纵向密集型数据的研究以及 2017 年 Schultzberg 和 Muthén 进行的两水平 DSEM 模型的时间和样本数量的研究以及 Mplus Examples (Monte Carlo Counterparts)。本研究假设存在两个变量  $Y$  和  $W$ ， $Y$  变量存在自回归效应， $W$  会对  $Y$  变量的均值，自回归斜率的大小产生影响。在此将  $Y$  在时间序列上的自回归参数的均值设定为 0.2/0.5，方差设定为 0.04，而变量  $Y$  的均值设为 0.5，方差设置为 0.07，协变量  $W$  的均值设为 0，方差设为 0.119，协变量  $W$  对随机自回归斜率的回归系数为 0.31，对因变量  $Y$  的回归系数为 0.41，残差方差的  $\log$  值的均值设为 -1.18，方差设为 0.02。

## 4.3. 考察指标

本研究的考察指标为 4 个参数估计指标。DSEM 模型的拟合优度指标一般采用 DIC 指标，但由于 DIC 的值会受到模型自由度和待估参数数量的影响，因此目前没有办法通过 DIC 对不同方法间模型拟合进行比较，DSEM 相关的模型研究此前都没有探讨比较过模型的拟合指标，因此在本研究中我们也未对其拟合指标进行比较和探讨 (Schultzberg & Muthén, 2018)。主要依据参数估计的误差均方根、95%覆盖率

和统计检验力这几个指标对不同模型进行比较。

### 1) 误差均方

误差均方(Mean square error, MSE)为描述参数估计精度的指标。MSE 越小则表明参数估计越精确, 相对更加稳健可以用于比较不同条件下模型参数估计的稳定性, 其计算公式如下,

$$\text{MSE} = \frac{\sum_{r=1}^R (\hat{\theta}_r - \theta)^2}{\text{NR}} \quad (14)$$

其中, NR 表示模拟研究中条件的重复次数,  $\hat{\theta}_r$  表示所考察参数在第  $r$  次上的估计值。

### 2) 统计检验力

显著性系数(Significant Coefficients)提供了对统计检验力和一类错误率的讨论, 这一指标也可通过 Mplus8.3 软件进行计算生成。统计检验力(Statistical Power)是 100 次重复当中参数估计的 95%置信区间不包含 0 的比例。本研究主要考察自回归系数的均值、随机效应, 以及协变量对于自回归系数和因变量的回归系数的统计检验力。

### 3) 95%覆盖率

95%覆盖率用于表示 100 次重复次数中参数估计的 95%置信区间包含真值的比例, 这一指标可以直接通过 Mplus8.3 软件生成。

$$\text{coverage}(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R CI_{95}(\hat{\theta}_r) \quad (15)$$

其中, 若  $\theta - 1.96 \times SE < \hat{\theta}_r < \theta + 1.96 \times SE$ , 则  $CI_{95}(\hat{\theta}_r) = 1$ ; 否则,  $CI_{95}(\hat{\theta}_r) = 0$ 。

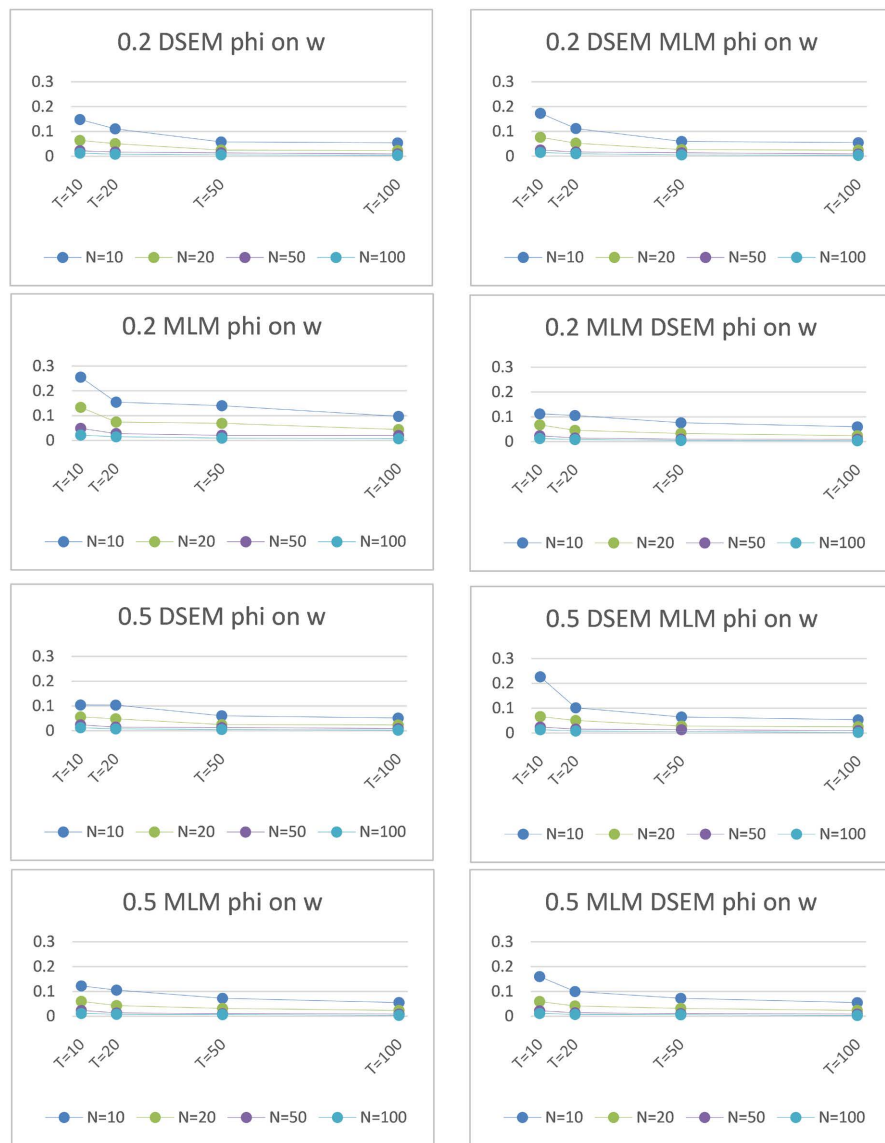
## 5. 研究结果

### 5.1. 误差均方

在这里, 模拟研究误差均方的结果见图 2。当因变量  $Y$  的自回归效应较弱时( $\gamma_{10} = 0.2$ ), 协变量对于随机自回归斜率的影响( $\gamma_{01} = 0.31$ )的 MSE 根, 在存在随机残差方差但残差方差又未被估计时大于 0.2, 其余情况这一参数的估计都是较为稳定的; 协变量对于因变量  $Y$  的均值的 MSE, 当错误假设是否存在随机残差方差时, 在各个时间点样本量条件下该参数估计的稳定性普遍较差(MSE > 0.2), 只有在被试数在 50 以上时, 这一参数估计的 MSE 才接近于 0.2, 而当正确假设是否存在随机残差方差时, MLM 在  $N \geq 50$ ,  $T \geq 20$  的条件下可以或者稳定的参数估计结果, 而 DSEM 在  $N \geq 20$ ,  $T \geq 50$  的条件下就可以获得稳定的参数估计结果; 随机自回归斜率均值的 MSE, 无论是 DSEM 还是 MLM, 在各个时间点样本量条件下都可以获得比较稳的参数估计结果。

当因变量  $Y$  的自回归效应较强时( $\gamma_{10} = 0.5$ ), 协变量对于随机自回归斜率的影响( $\gamma_{01} = 0.31$ )的 MSE 在正确假设是否存在随机残差方差时, 在  $N \geq 50$ ,  $T \geq 20$  的条件下该参数的估计结果均很稳定, 当错误假设时, 在  $N \geq 20$ ,  $T \geq 50$  或是  $N \geq 50$ ,  $T \geq 20$  的条件下, 该参数的估计结果是稳定的; 协变量对于因变量  $Y$  的均值的 MSE, 当正确假设是否存在随机残差方差时,  $N * T \geq 200$  时, 该参数的估计结果是稳定的, 当错误假设是否存在随机残差方差时, 在  $N \geq 50$ ,  $T \geq 50$  或是在  $N \geq 100$ ,  $T \geq 20$  时, 该参数的估计结果是稳定的; 随机自回归斜率均值的 MSE, 在正确假设是否存在随机残差方差时, MLM 和 DSEM 在各个时间点样本量条件下, 参数估计的结果都很稳定, 而在错误假设是否存在随机残差方差时, 参数估计的结果都比较不稳定(MSE > 0.2)。





**Figure 2.** MSE of the effect of covariates on random autoregressive coefficients ( $\phi$  on  $w$ )  
**图 2.** 协变量对于随机自回归系数的影响( $\phi$  on  $w$ )的 MSE

## 5.2. 统计检验力

模拟研究统计检验力的结果见图 3，当因变量  $Y$  的自回归效应较弱时 ( $\gamma_{10} = 0.2$ )，协变量对于随机自回归斜率的影响 ( $\gamma_{01} = 0.31$ ) 的显著性水平，在各个时间点样本量条件下，使用 DSEM 模型与使用 MLM 模型相比差异不大，显著性水平更多的是受到时间点和样本量数量的影响。相比时间点数量，样本量对于参数显著性水平的影响更大，当样本量在 50 及以上且时间点数量在 20 及以上时，协变量对于因变量  $Y$  的影响的显著性水平大于 0.8。

随机自回归斜率的均值 ( $\phi$ ) 的显著性水平在各个模拟条件下均较高，使用正确的或错误的模型进行估计对于该参数的显著性水平的影响较小，该参数的显著性水平只受时间点和样本量的影响，样本量的影响要大于时间点数量。当随机自回归效应较强时， $N \geq 50$ ， $T \geq 20$  时，该参数的显著性水平大于 0.8，当随机自回归效应较弱时，当  $N \geq 50$ ， $T \geq 10$  或者  $N \geq 50$ ， $T \geq 50$  时，该参数的显著性水平大于 0.8。

协变量对于因变量的斜率(y on w)的显著性水平会受到估计模型的影响,当数据中存在随机残差方差却使用 MLM 模型进行估计时,参数的显著性水平会降低,但数据中不存在随机残差方差时,即便使用 DSEM 模型进行估计,参数的显著性水平不受影响,各个条件下 DSEM 模型估计的显著性水平均由于 MLM 模型。

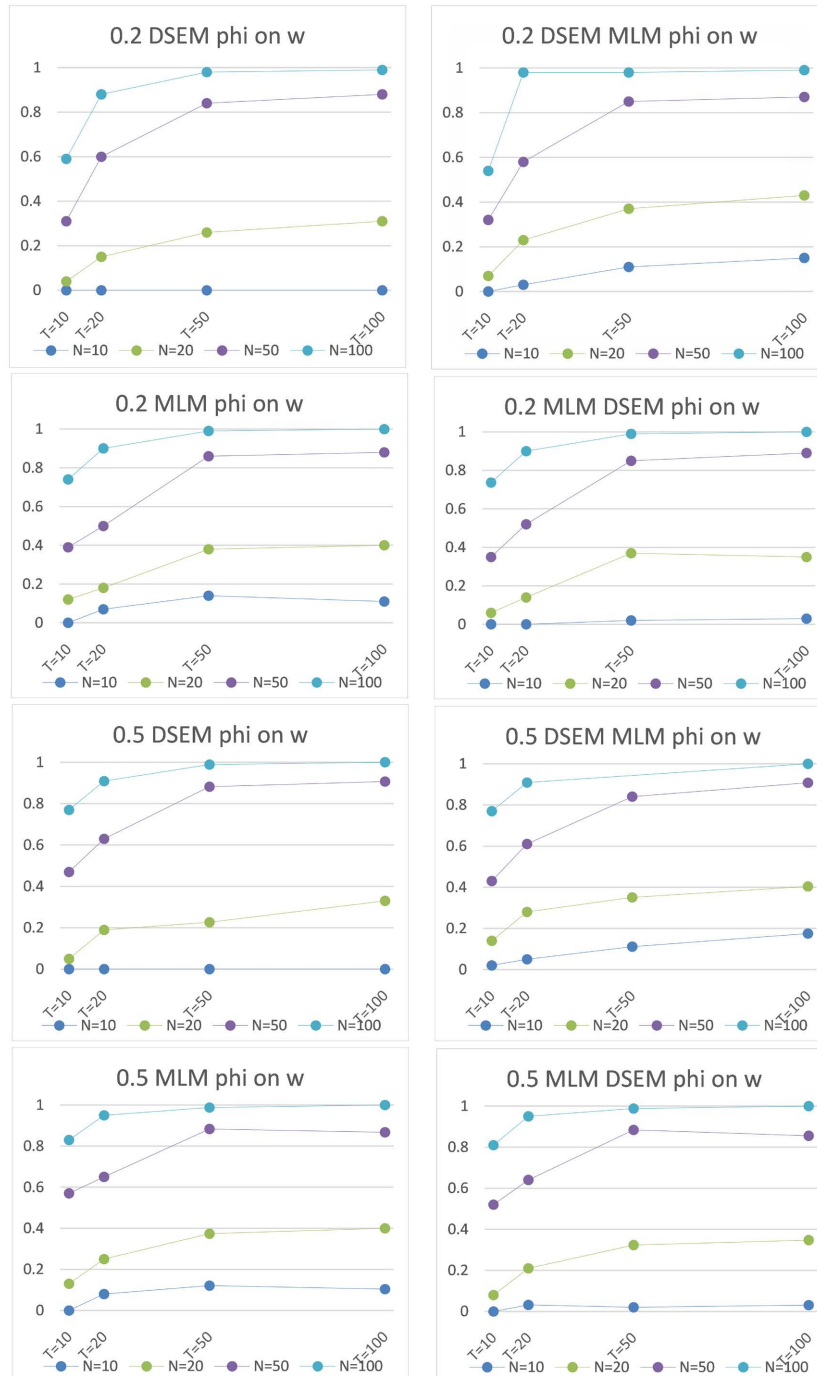


Figure 3. Statistical test power of the influence of covariates on the random autoregressive coefficients (phi on w)

图 3. 协变量对于随机自回归系数的影响(phi on w)的统计检验力

### 5.3. 95%覆盖率

时间点、样本量、因子载荷、对各个拟合模型参数估计的 95%覆盖率影响均不大, 只有在样本量为 100, 时间点数量为 10 的情况下 95%覆盖率才在 0.9 以下。在其余各个情况下, 所有参数的 95%覆盖率均大于 0.95, 绝大部分情况下 95%覆盖率的值在 1。

## 6. 讨论

模拟研究中我们发现无论是 DSEM 还是 MLM 模型, 对于模型中随机均值的估计都是比较准确的。随机均值在各个模型各个时间点样本量条件下都得到了良好的估计, 各个样本量时间点条件下, MSE 都接近于零, 显著性水平接近于 1。除了可以良好估计各个参数随机均值, 模型对于协变量对于随机自回归斜率的影响这一参数也有非常好的估计结果。对于  $N \geq 50$  时, 该参数在各个条件下均有良好的估计精度和稳定性。

此外, 由于 DSEM 研究条件的设置往往不是单纯的时间点和样本量的问题, 而是二者的组合。与此同时, 在现实的研究过程中, 往往会出现我的测量次数是固定的, 能够调整的只有被试数量, 或是被试数量是固定的, 能够调整的只有测量次数的情况, 因此很多研究者都很关注对于 DSEM 或是 MLM 模型而言,  $N$  和  $T$  之间是否存在互相补偿的关系。根据本模拟研究的结果而言, 大  $N$  小  $T$  的估计结果普遍要优于小  $N$  大  $T$ , 实际研究中研究者可以通过增大样本量来适当地减小测量次数, 这对于参数估计的准确度和稳定性不会有很大的影响。

因此当研究可以明确残差方差是否随机时, 建议在有残差方差时使用 DSEM 模型, 没有残差方差时使用 MLM 模型。而如果不明确是否有残差方差, 且时间点和样本量都大于等于 10 时, 一般使用 DSEM 模型估计会更好。

## 参考文献

- 安媛媛, 徐慰(2017). 雾霾和知觉压力加重疲劳感:来自生态瞬时评估的证据. *中国临床心理学杂志*, 25(6), 1130-1133. <https://doi.org/10.16128/j.cnki.1005-3611.2017.06.029>
- 刘源, 都弘彦, 方杰, 温忠麟(2022). 国内追踪数据分析方法研究与模型发展. *心理科学进展*, 30(8), 1734-1746. <https://doi.org/10.3724/sp.j.1042.2022.01734>
- 张沥今, 陆嘉琦, 魏夏琰, 潘俊豪(2019). 贝叶斯结构方程模型及其研究现状. *心理科学进展*, 27(11), 1812-1825.
- 张银普, 骆南峰, 石伟(2016). 经验取样法——一种收集“真实”数据的新方法. *心理科学进展*, 24(2), 305-316.
- Armstrong, B., Covington, L. B., Unick, G. J., & Black, M. M. (2019). Featured Article: Bidirectional Effects of Sleep and Sedentary Behavior among Toddlers: A Dynamic Multilevel Modeling Approach. *Journal of Pediatric Psychology*, 44, 275-285. <https://doi.org/10.1093/jpepsy/jsy089>
- Asparouhov, T., & Muthén, B. (2019). Latent Variable Centering of Predictors and Mediators in Multilevel and Time-Series Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 119-142. <https://doi.org/10.1080/10705511.2018.1511375>
- Bolger, N., & Schilling, E. A. (1991). Personality and the Problems of Everyday Life: The Role of Neuroticism in Exposure and Reactivity to Daily Stressors. *Journal of Personality*, 59, 355-386. <https://doi.org/10.1111/j.1467-6494.1991.tb00253.x>
- Chen, Y., & Zhang, S. (2020). A Latent Gaussian Process Model for Analysing Intensive Longitudinal Data. *British Journal of Mathematical and Statistical Psychology*, 73, 237-260. <https://doi.org/10.1111/bmsp.12180>
- Chun, C. A. (2016). The Expression of Posttraumatic Stress Symptoms in Daily Life: A Review of Experience Sampling Methodology and Daily Diary Studies. *Journal of Psychopathology and Behavioral Assessment*, 38, 406-420. <https://doi.org/10.1007/s10862-016-9540-3>
- Curran, P. J., & Bauer, D. J. (2011). The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology*, 62, 583-619. <https://doi.org/10.1146/annurev.psych.093008.100356>
- Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the Frontiers of Modeling Intensive

- Longitudinal Data: Dynamic Structural Equation Models for the Affective Measurements from the COGITO Study. *Multivariate Behavioral Research*, 53, 820-841. <https://doi.org/10.1080/00273171.2018.1446819>
- Howland, M., & Rafaeli, E. (2010). Bringing Everyday Mind Reading into Everyday Life: Assessing Empathic Accuracy with Daily Diary Data. *Journal of Personality*, 78, 1437-1468. <https://doi.org/10.1111/j.1467-6494.2010.00657.x>
- Jahng, S. (2008). *Multilevel Models for Intensive Longitudinal Data with Heterogeneous Error Structure: Covariance Transformation and Variance Function Models*. University of Missouri-Columbia.
- Jahng, S., & Wood, P. K. (2017). Multilevel Models for Intensive Longitudinal Data with Heterogeneous Autoregressive Errors: the Effect of Misspecification and Correction with Cholesky Transformation. *Frontiers in Psychology*, 8, Article 262. <https://doi.org/10.3389/fpsyg.2017.00262>
- Joly-Burra, E., Van Der Linden, M., & Ghisletta, P. (2018). Intraindividual Variability in Inhibition and Prospective Memory in Healthy Older Adults: Insights from Response Regularity and Rapidity. *Journal of Intelligence*, 6, Article 13. <https://doi.org/10.3390/jintelligence6010013>
- Lanza, S. T., Vasilenko, S. A., Liu, X., Li, R., & Piper, M. E. (2014). Advancing the Understanding of Craving during Smoking Cessation Attempts: A Demonstration of the Time-Varying Effect Model. *Nicotine & Tobacco Research*, 16, S127-S134. <https://doi.org/10.1093/ntr/ntt128>
- Lazarević, L. B., Bjekić, J., Živanović, M., & Knežević, G. (2020). Ambulatory Assessment of Language Use: Evidence on the Temporal Stability of Electronically Activated Recorder and Stream of Consciousness Data. *Behavior Research Methods*, 52, 1817-1835. <https://doi.org/10.3758/s13428-020-01361-z>
- Malmberg, L. E., & Martin, A. J. (2019). Processes of Students' Effort Exertion, Competence Beliefs and Motivation: Cyclic and Dynamic Effects of Learning Experiences within School Days and School Subjects. *Contemporary Educational Psychology*, 58, 299-309. <https://doi.org/10.1016/j.cedpsych.2019.03.013>
- McAdams, D. P., & Constantian, C. A. (1983). Intimacy and Affiliation Motives in Daily Living: An Experience Sampling Analysis. *Journal of Personality and Social Psychology*, 45, 851-861. <https://doi.org/10.1037/0022-3514.45.4.851>
- Mun, C. J., Suk, H. W., Davis, M. C., Karoly, P., Finan, P., Tennen, H., & Jensen, M. P. (2019). Investigating Intraindividual Pain Variability: Methods, Applications, Issues, and Directions. *Pain*, 160, 2415-2429. <https://doi.org/10.1097/j.pain.0000000000001626>
- Munsch, S., Meyer, A. H., Milenkovic, N., Schlup, B., Margraf, J., & Wilhelm, F. H. (2009). Ecological Momentary Assessment to Evaluate Cognitive-Behavioral Treatment for Binge Eating Disorder. *International Journal of Eating Disorders*, 42, 648-657. <https://doi.org/10.1002/eat.20657>
- Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica: Journal of the Econometric Society*, 49, 1417-1426. <https://doi.org/10.2307/1911408>
- Öhrlund, I., Schultzberg, M., & Bartusch, C. (2019). Identifying and Estimating the Effects of a Mandatory Billing Demand Charge. *Applied Energy*, 237, 885-895. <https://doi.org/10.1016/j.apenergy.2019.01.028>
- Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of Idiographic Methods in Psychology and Applications for Psychotherapy. *Clinical Psychology Review*, 71, 90-100. <https://doi.org/10.1016/j.cpr.2019.01.002>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (Vol. 1). Sage.
- Schultzberg, M., & Muthén, B. (2018). Number of Subjects and Time Points Needed for Multilevel Time-Series Analysis: A Simulation Study of Dynamic Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 495-515. <https://doi.org/10.1080/10705511.2017.1392862>
- Setodji, C. M., Martino, S. C., Dunbar, M. S., & Shadel, W. G. (2019). An Exponential Effect Persistence Model for Intensive Longitudinal Data. *Psychological Methods*, 24, 622-636. <https://doi.org/10.1037/met0000211>
- Shiffman, S. (2009). Ecological Momentary Assessment (EMA) in Studies of Substance Use. *Psychological Assessment*, 21, 486-497. <https://doi.org/10.1037/a0017074>
- Shiffman, S., & Waters, A. J. (2004). Negative Affect and Smoking Lapses: A Prospective Analysis. *Journal of Consulting and Clinical Psychology*, 72, 192-201. <https://doi.org/10.1037/0022-006X.72.2.192>
- Sturgeon, J. A., Zautra, A. J., & Arewasikporn, A. (2014). A Multilevel Structural Equation Modeling Analysis of Vulnerabilities and Resilience Resources Influencing Affective Adaptation to Chronic Pain. *Pain*, 155, 292-298. <https://doi.org/10.1016/j.pain.2013.10.007>
- Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory Assessment. *Annual Review of Clinical Psychology*, 9, 151-176. <https://doi.org/10.1146/annurev-clinpsy-050212-185510>
- Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective Instability: Measuring a Core Feature of Borderline Personality Disorder with Ecological Momentary Assessment. *Journal of Abnormal Psychology*, 117, 647-661. <https://doi.org/10.1037/a0012532>

- 
- Vallacher, R. R., Read, S. J., & Nowak, A. (2002). The Dynamical Perspective in Personality and Social Psychology. *Personality and Social Psychology Review*, 6, 264-273. [https://doi.org/10.1207/S15327957PSPR0604\\_01](https://doi.org/10.1207/S15327957PSPR0604_01)
- Weinstock, J., Farney, M. R., Elrod, N. M., Henderson, C. E., & Weiss, E. P. (2017). Exercise as an Adjunctive Treatment for Substance Use Disorders: Rationale and Intervention Description. *Journal of Substance Abuse Treatment*, 72, 40-47. <https://doi.org/10.1016/j.jsat.2016.09.002>
- Windt, J., Arden, C. L., Gabbett, T. J., Khan, K. M., Cook, C. E., Sporer, B. C., & Zumbo, B. D. (2018). Getting the Most Out of Intensive Longitudinal Data: A Methodological Review of Workload-Injury Studies. *BMJ Open*, 8, e022626. <https://doi.org/10.1136/bmjopen-2018-022626>
- Zhang, Z., & Nesselroade, J. R. (2007). Bayesian Estimation of Categorical Dynamic Factor Models. *Multivariate Behavioral Research*, 42, 729-756. <https://doi.org/10.1080/00273170701715998>
- Zhou, L., Wang, M., & Zhang, Z. (2021). Intensive Longitudinal Data Analyses with Dynamic Structural Equation Modeling. *Organizational Research Methods*, 24, 219-250. <https://doi.org/10.1177/1094428119833164>