

Literature Summary

Benedikt Lugauer

Table of contents

Summary	1
Background and Objectives	2
Key Concepts and Terminology	2
Data Decomposition	2
Dynamic Multilevel Modeling Using DSEM	3
Model Extensions	3
1. Random Innovation Variances and Covariance	3
2. Indirect Effects Through Random Effects	4
Key Findings	4
Discussion of Challenges and Unresolved Issues	5
Conclusion	6
Summary	6
Introduction	6
Objective	7
Background	7
1. AR(1) Model	7
2. Multilevel AR(1) Model	7
Alternative Data-Generating Models (DGMs)	8

Summary

The article “At the Frontiers of Modeling Intensive Longitudinal Data: Dynamic Structural Equation Models for the Affective Measurements from the COGITO Study” by Hamaker et al. (2018) explores advanced modeling techniques for intensive longitudinal data using dynamic structural equation modeling (DSEM). The authors analyze affective data from the COGITO study to illustrate how DSEM can model individual differences in affect dynamics, including means, autoregressive effects, cross-lagged effects, and innovation variances and covariances.

Background and Objectives

- **COGITO Study:** An intensive longitudinal study involving two samples—101 younger adults (aged 20–31) and 103 older adults (aged 65–80)—each providing data over approximately 100 days.
- **Data Collected:** Daily self-reports of positive affect (PA) and negative affect (NA) using the Positive and Negative Affect Schedule (PANAS).
- **Aim:** To model within-person processes over time and to capture individual differences in affect dynamics using dynamic multilevel modeling with DSEM in Mplus software.

Key Concepts and Terminology

- **Within-Person Variability (Intraindividual Variability):** Variations in an individual's affect over time.
- **Between-Person Variability (Interindividual Variability):** Stable, trait-like differences between individuals.
- **Cross-Sectional Variability:** Variability across individuals at a single time point, combining both within-person and between-person variability.
- **Total Variability:** Variability across both individuals and time points.

Data Decomposition

The observed affect scores are decomposed into between-person means and within-person deviations:

$$PA_{it} = \mu_{PA,i} + PA_{it}^{(W)} \quad NA_{it} = \mu_{NA,i} + NA_{it}^{(W)}$$

- $\mu_{PA,i}$ and $\mu_{NA,i}$: Individual mean levels of PA and NA (between-person components).
- $PA_{it}^{(W)}$ and $NA_{it}^{(W)}$: Within-person deviations from the individual means at time (t) (time-varying components).

Dynamic Multilevel Modeling Using DSEM

- **Within-Person Level Model:** A first-order vector autoregressive (VAR(1)) model captures the dynamics of PA and NA over time for each individual.

$$\begin{pmatrix} PA_{it}^{(W)} \\ NA_{it}^{(W)} \end{pmatrix} = \begin{pmatrix} \phi_{PP,i} & \phi_{PN,i} \\ \phi_{NP,i} & \phi_{NN,i} \end{pmatrix} \begin{pmatrix} PA_{it-1}^{(W)} \\ NA_{it-1}^{(W)} \end{pmatrix} + \begin{pmatrix} \zeta_{PA,it} \\ \zeta_{NA,it} \end{pmatrix}$$

- $\phi_{PP,i}$ and $\phi_{NN,i}$: Individual autoregressive coefficients for PA and NA, respectively.
- $\phi_{PN,i}$ and $\phi_{NP,i}$: Individual cross-lagged effects between PA and NA.
- $\zeta_{PA,it}$ and $\zeta_{NA,it}$: Innovation terms (residuals) at time (t) for individual (i), assumed to be multivariate normally distributed.
- **Between-Person Level Model:** Captures individual differences in the means and dynamic parameters.

$$\begin{pmatrix} \mu_{PA,i} \\ \mu_{NA,i} \\ \phi_{PP,i} \\ \phi_{PN,i} \\ \phi_{NP,i} \\ \phi_{NN,i} \end{pmatrix} = \begin{pmatrix} \gamma_{PA} \\ \gamma_{NA} \\ \gamma_{PP} \\ \gamma_{PN} \\ \gamma_{NP} \\ \gamma_{NN} \end{pmatrix} + \begin{pmatrix} u_{PA,i} \\ u_{NA,i} \\ u_{PP,i} \\ u_{PN,i} \\ u_{NP,i} \\ u_{NN,i} \end{pmatrix}$$

- γ : Fixed (average) effects across individuals.
- u : Random effects capturing individual deviations from the average.

Model Extensions

1. Random Innovation Variances and Covariance

- **Rationale:** Individuals may differ in their reactivity to unobserved influences or in the variability of these influences.
- **Model Adjustment:** The variances and covariance of the innovation terms ζ are allowed to vary across individuals.
- **Implementation:**
 - Introduce a latent factor η_{it} to model the shared component of the innovations.
 - Decompose the innovations:

$$\begin{pmatrix} \zeta_{PA,it} \\ \zeta_{NA,it} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \eta_{it} + \begin{pmatrix} \delta_{PA,it} \\ \delta_{NA,it} \end{pmatrix}$$

- The unique residuals $\delta_{PA,it}$ and $\delta_{NA,it}$ are uncorrelated.
- Individual variances and covariance are functions of the variances of (η_{it}) and (δ) :

$$\begin{aligned} Var(\zeta_{PA,it}) &= \psi_i + \pi_{PA,i}, \\ Var(\zeta_{NA,it}) &= \psi_i + \pi_{NA,i}, \\ Cov(\zeta_{PA,it}, \zeta_{NA,it}) &= \psi_i. \end{aligned}$$

- ψ_i : Variance of the common factor η_{it} for individual (i).
- $\pi_{PA,i}$ and $\pi_{NA,i}$: Variances of the unique residuals for PA and NA.
- **Between-Person Level Extension:**
- The log of the variances $\log \pi_{PA,i}$, $\log \pi_{NA,i}$, $\log \psi_i$ are modeled as random effects to ensure positivity.

2. Indirect Effects Through Random Effects

- **Objective:** To assess whether prior depression (pre-CESD scores) predicts later depression (post-CESD scores) through the random effects of affect dynamics.
- **Modeling Approach:**
- Include pre-CESD as a predictor of the random effects at the between-person level.
- Include post-CESD as an outcome variable predicted by the random effects and pre-CESD.
- Compute indirect effects as the product of the effects from pre-CESD to the random effect and from the random effect to post-CESD.

Key Findings

- **Model 1 (Basic VAR(1) Model):**
- Significant individual differences in means, autoregressive parameters, and cross-lagged effects.
- Negative within-person correlation between PA and NA, indicating that on days with higher PA, individuals tend to report lower NA, and vice versa.

- Between-person correlations differed from within-person correlations, emphasizing the importance of separating these sources of variance.
- **Model 2 (Including Random Innovation Variances and Covariance):**
- Individual differences in innovation variances and covariance were significant.
- The random effects (means, lagged parameters, variances) were interrelated at the between-person level.
- Ignoring random innovation variances can bias estimates of other model parameters.
- **Model 3 (Indirect Effects Through Random Effects):**
- In the younger sample, higher pre-CESD scores predicted higher post-CESD scores through:
- Increased autoregressive parameter of NA $\phi_{NN,i}$: Suggesting that individuals with higher depression symptoms have more persistent NA over time.
- Stronger negative covariance between innovations ψ_i : Indicating that higher depression is associated with more intertwined fluctuations of PA and NA.
- No significant indirect effects through the means of PA and NA.

Discussion of Challenges and Unresolved Issues

- **Data Requirements:**
- **Sample Size:** Balancing the number of individuals and the number of time points is crucial for reliable estimates, especially when modeling random effects.
- **Non-Normal Data:** Handling skewed or non-normal distributions (common with affect data) requires careful modeling choices, such as treating variables as categorical or using transformations.
- **Model Evaluation:**
- **Model Comparison:** Need for appropriate criteria (e.g., Deviance Information Criterion) to compare nested and non-nested models, though challenges exist due to the complexity of random effects.
- **Model Fit:** Assessing model fit is complicated in time series models, and traditional fit indices may not be directly applicable.
- **Theoretical Concerns:**
- **Causality:** Caution is advised when interpreting lagged effects as causal due to potential omitted variables and the observational nature of the data.

- **Trends and Cycles:** Deciding whether to model trends and cycles as part of the dynamics or as separate components is important for accurate modeling.
- **Between-Person Differences:** Understanding and modeling individual differences at the between-person level can provide insights but also adds complexity.
- **Discrete vs. Continuous Time:** Recognizing that lagged effects are specific to the time interval used and considering continuous-time modeling as an alternative.

Conclusion

The study demonstrates the utility of DSEM for modeling intensive longitudinal data, capturing complex affect dynamics and individual differences. The approach allows researchers to disentangle within-person processes from between-person differences and to explore how these dynamics relate to broader psychological constructs like depression. However, the authors also highlight several methodological challenges and areas needing further research, including data requirements, model evaluation, and theoretical considerations in modeling dynamic processes.

Summary

Title: Skewness and Staging: Does the Floor Effect Induce Bias in Multilevel AR(1) Models?

Authors: M. M. Haqiqatkhah, O. Ryan, and E. L. Hamaker

Introduction

Multilevel autoregressive (AR(1)) models are widely used in psychology to analyze intensive longitudinal data. Researchers have observed a positive correlation between individuals' autoregressive parameters (inertia) and measures of psychopathology severity—a phenomenon known as the **staging effect**. This suggests that higher inertia in affective time series is associated with higher levels of psychological distress or disorder severity.

However, concerns have been raised that this staging effect might be a statistical artifact arising from the **floor effect**—where data distributions are skewed due to a clustering of low scores, leading to high skewness, low mean, and low variability. This issue is particularly relevant for variables like negative affect in healthy individuals, where many report minimal symptoms, resulting in skewed distributions.

Objective

The study aims to investigate whether the floor effect induces bias in multilevel AR(1) models, specifically whether it leads to a spurious positive correlation between the mean level of an affective measure and its autoregressive parameter (the staging effect). The authors explore this through a simulation study using alternative data-generating models that produce skewed data with floor effects.

Background

1. AR(1) Model

The univariate first-order autoregressive (AR(1)) model for a time series (X_t) is:

$$X_t = c + \phi X_{t-1} + \varepsilon_t$$

where:

- c is the intercept.
- ϕ is the autoregressive parameter (inertia).
- ε_t is the innovation term, assumed to be normally distributed: $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

Stationarity and Marginal Properties:

For stationarity, ($|\phi| < 1$). The stationary (long-run) mean and variance are:

$$E(X_t) = \mu = \frac{c}{1 - \phi}, \quad \text{Var}(X_t) = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

The marginal distribution of (X_t) is normal due to the normality of (ε_t).

2. Multilevel AR(1) Model

In multilevel modeling, data from multiple individuals are analyzed simultaneously. The observed score for individual (i) at time (t), ($X_{i,t}$), is decomposed into:

$$X_{i,t} = \mu_i + \tilde{X}_{i,t},$$

where:

- (μ_i) is the individual's mean (between-person effect).

- $(\tilde{X}_{i,t})$ is the within-person deviation, modeled as:

$$\tilde{X}_{i,t} = \phi_i \tilde{X}_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, \sigma_{\varepsilon,i}^2).$$

Assumptions:

- **Level 1 (Within-Person) Assumption:** The residuals $(\varepsilon_{i,t})$ are normally distributed.
- **Level 2 (Between-Person) Assumption:** The random effects μ_i , ϕ_i , and $\sigma_{\varepsilon,i}^2$ are normally distributed across individuals.

Violation of Assumptions:

Empirical data, especially for variables like distress or negative affect, often violate these assumptions due to the floor effect. Individuals with low mean scores exhibit distributions with high skewness and low variability, deviating from normality.

Empirical Illustration:

Using data from the COGITO study, the authors demonstrate that many individuals' distress scores are highly skewed with strong floor effects, violating normality assumptions at both levels.

Alternative Data-Generating Models (DGMs)

To investigate whether the floor effect leads to spurious correlations, the authors introduce three alternative DGMs that can produce skewed data with floor effects.