

A Primer on Two-Level Dynamic Structural Equation Models  
for Intensive Longitudinal Data in *Mplus*

Daniel McNeish  
Arizona State University, USA

Ellen L. Hamaker  
Utrecht University, the Netherlands

© 2019, American Psychological Association.

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/met0000250

DISCLOSURE OF INTEREST: Dr. Hamaker has been involved in the development of DSEM in *Mplus*, which is the topic of this paper.

CONTACT INFORMATION:

Daniel McNeish, Department of Psychology, PO Box 871104, Tempe, AZ, USA 85287. Email: [dmcneish@asu.edu](mailto:dmcneish@asu.edu).

## Abstract

Technological advances have led to an increase in intensive longitudinal data and the statistical literature on modeling such data is rapidly expanding, as are software capabilities. Common methods in this area are related to time-series analysis, a framework that historically has received little exposure in psychology. There is a scarcity of psychology-based resources introducing the basic ideas of time-series analysis, especially for datasets featuring multiple people. We begin with basics of  $N=1$  time-series analysis and build up to complex dynamic structural equation models available in the newest release of *Mplus* Version 8. The goal is to provide readers with a basic conceptual understanding of common models, template code, and result interpretation. We provide short descriptions of some advanced issues, but our main priority is to supply readers with a solid knowledge base so that the more advanced literature on the topic is more readily digestible to a larger group of researchers.

**KEYWORDS:** Dynamic Structural Equation Modeling; Time-Series Analysis; Intensive Longitudinal Data; Multilevel Modeling

## **A Primer on Two-Level Dynamic Structural Equation Models for Intensive Longitudinal Data in *Mplus***

As technological advances have continued to emerge at a rapid rate, data collection methods such as experience sampling, ambulatory assessment, daily diaries and ecological momentary assessment are allowing different types of longitudinal data to be collected, in far less intrusive manners, with fewer barriers (Conner & Barrett, 2012; Hamaker & Wichers, 2017; Mehl & Conner, 2012; Trull & Ebner-Priemer, 2014). Whereas traditional longitudinal data tend to consist of a small number of repeated measures with relatively large time intervals between them, the innovative data collection techniques result in much larger numbers of repeated measures which tend to be closely spaced in time. As the nature of longitudinal data collection changes, the methods for modeling such data have similarly changed, requiring psychologists and methodologists to adapt to evolving data structures (Jebb, Tay, Wang, & Huang, 2015). Though growth modeling has historically been the dominant method by which to model longitudinal data in psychology and related fields (e.g., Jebb & Tay, 2017), the focus in the newer forms of longitudinal research has shifted from modeling trends over time that reflect developmental processes, to investigating the dynamics of stable processes (Hamaker & Wichers, 2017; Molenaar, 2004; Molenaar & Campbell, 2009). This has led to a surge of recent advances in statistical models for *intensive longitudinal data* (ILD; e.g., Asparouhov, Hamaker, & Muthén, 2017; Gates & Molenaar, 2012; Jongerling, Laurenceau, & Hamaker, 2015; Oravecz, Tuerlinckx, & Vanderkerckhove, 2011, 2016; Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016), roughly defined as data with 20 or more measurement occasions (Collins, 2006; Walls & Schafer, 2006).

Software programs have similarly adapted to facilitate these types of analyses, such as the newly released *Mplus* Version 8, which now has a dedicated module on dynamic structural equation modeling (DSEM) for ILD (Asparouhov, Hamaker, & Muthén, 2018). DSEM integrates three different modeling approaches, that is: a) *time-series analysis*, which allows for modeling the lagged

relation in single subject data that consist of a large number of repeated measures; b) *multilevel modeling*, which can be used to simultaneously model longitudinal data of multiple individuals while allowing for individual differences in the parameters that describe the individual processes; and c) *structural equation modeling (SEM)*, which offers the opportunity to have multiple outcome variables, latent variables, and mediation effects. DSEM can be used for  $N = 1$  time series data, but also to model the data from multiple individuals simultaneously, where the within-person dynamics are modeled using a time-series model at the within-person level, while individual differences in individual parameters are modeled at the between-person level.

Multilevel modeling and SEM, as well as the combination of the two, are fairly well known among psychological researchers. In contrast however, time-series analysis is still relatively novel for most in this field (Jebb & Tay, 2017), although the literature and applications of these methods are rapidly expanding, and statistical research continues to push the boundaries (Boker et al., 2011; Driver, Oud, & Voelkle, 2017; Hamaker & Wichers, 2017; Hamaker, Asparouhov, Brose, Schmiedek, & Muthén, 2018; Ou, Hunter, & Chow, 2018). Nevertheless, conceptual understanding of these new techniques remains difficult for empirical and quantitative researchers alike as many of the concepts that are fundamental to time-series analysis—and therefore also to DSEM—are not common in psychology (Jebb et al., 2015). Furthermore, existing treatments in fields with a richer history of time-series models tend to focus on forecasting (Shmueli, 2010), an interest that is relatively uncommon in psychology when compared to explanation (though this may be changing, see, Yarkoni & Westfall, 2017). As a result, many researchers are left to start nearly from scratch with minimal resources at their disposal that are targeted specifically to an audience with little to no background on the subject. Alternatively, as noted by Hamaker, Ceulmans, Grasman, and Tuerlinckx (2015), many sources will provide an overview of why and how to collect ILD but will not include details on how to analyze such data to answer relevant research questions (p. 317).

The goal of this paper is to follow the call of Sharpe (2013) to bridge the gap from the high-end statistical work being done in this area to researchers who are just beginning to use these methods or who may not yet be aware of power of these techniques to answer research questions in their field. Essentially, our aim is to provide readers with a primer on the very basics of DSEM, how these models are fit in *Mplus*, and how the estimates are interpreted so that readers can more comfortably and confidently dive into the more comprehensive and nuanced discussions of these models (e.g., the recent work by Asparouhov et al., 2017, 2018 or Hamaker et al., 2018). That is, though some resources exist that are targeted to an intermediate audience where some familiarity is assumed, this paper is intended to readers at the novice level where we assume that readers possess knowledge of the basics of structural equation and multilevel modeling but no familiarity with ILD or time-series analysis.

To outline the structure of this paper, we first compare *Developmental Process* longitudinal data, which has historically been associated with growth modeling with the focus on a developmental trajectory and individual differences therein, to *Stable Process* longitudinal data, which is more typically associated with ILD and that is characterized by a focus on regulatory mechanisms and the dynamics of this. We then begin with some of the fundamental terminology for analysis of ILD and overview a single-level analysis where  $N = 1$ . We then extend the example to the more common case where multiple people are followed over time such that the data possess a multilevel structure with measurement occasions nested within people. We show how some straightforward DSEM models for ILD can be written in a very similar manner as growth models, though the focus and interpretation differ. We proceed with additional features that can be added to DSEM models, some of which are difficult to accommodate in other modeling frameworks and software. We then highlight some of the common difficulties with these models and how particular advanced features of DSEM can address them. Throughout the paper, example analyses and annotated software code are provided to facilitate application of these models to empirical datasets. Detailed interpretations and explanations of

idiosyncrasies of the DSEM framework that are not common in other frameworks are also provided as this basic but crucial information is often glanced over in more technical treatments of the topic.

### **Developmental Processes vs. Stable Processes**

Data collected over time on the same people fall under the umbrella of longitudinal data and the characteristics of such data can fall into one of two general classifications: developmental processes and stable processes (Jongerling et al., 2015). Table 1 summarizes the major differences between the two processes, details of which are expanded upon in this section. There are, of course, nuances and exceptions to what is presented here. We aim to provide a broad overview of the distinctions and differences in structure for different types of longitudinal data, so these characteristics describe more-or-less textbook cases, but should not be taken as rigid classifications.<sup>1</sup>

[Table 1 about here]

### **Developmental Processes**

Developmental process data (also known as panel data) typically feature a few measurement occasions that are widely spaced over the observations window (often months or years apart). The primary focus is questions about the *means* of the outcome variable over the course of the observation window (Nesselroade, 1991) – how much do the means change from the first to the last measurement occasion? What shape or trajectory does the change in the means follow over time? Common models for such data include mixed-effect models, latent growth models, or latent change score models (Grimm, Ram, & Estabrook, 2017). These models tend to take a *nomothetic* approach in describing the overall mean across people (McNeish, Stapleton, & Silverman, 2017). When expanded to person-specific growth curves, the focus is on within-person *change* not on within-person *variability* (Ram & Gertsoff, 2009). That is, the growth curve of each person is assessed and

---

<sup>1</sup> For example, the measurement burst design of Nesselroade (1991) would not squarely fit into either of the classifications we present here. Also, the data structure for a cross-lagged panel model may have several features of developmental processes but the interest may not necessarily focus on change.

considered a trait score but the variability around each person's growth curve is considered noise (Wang, Hamaker, & Bergeman, 2012). In models for developmental processes, it is typical for little consideration to be given to the covariance structures (Grimm & Widaman, 2010; Preacher & Hancock, 2015) and this information is frequently not reported in articles (McNeish et al., 2017). Moreover, when covariates are added to the model, the predominant focus is variables that affect the shape of the growth curve (Curran, Bauer, & Willoughby, 2014) – which variables predict higher or lower intercepts? How do particular variables affect the steepness of the growth curve? However, covariates that affect the variability *around* the growth curve are much less frequent, partially because data requirements to do so may not be present unless there are more measurement occasions (Hedeker, Mermelstein, Berbaum, & Campbell, 2009). Plainly stated, covariates for developmental process commonly explain between-person variability (factors that lead to differently shaped growth curves) but less commonly explain within-person variability (deviation around the growth curve).

With developmental processes, the important aspect of data collection is the length of the observations window – collecting data more frequently within a window that is too narrow provides little benefit with developmental processes given the focus on long-term change. Data must be collected long enough to allow mean changes to occur. The number of measurement occasions collected depends mostly on the expected functional form of the mean change over time (Curran, Obeidat, & Losardo, 2010; Duncan & Duncan, 2009). If the change is expected to be linear, it is efficient to collect only 3 or 4 measurement occasions that are spread out over a lengthy period of time – the cost of collecting additional time-points is unlikely to be worth the information additional measurement occasions provide. In most cases, the number of measurement occasions rarely enters double digits because even complex non-linearly trajectories can be sufficiently modeled with fewer data points.

### **Stable Processes**

Stable process data (also known as ILD) are characterized by collecting many measurement occasions over a short observations window (i.e., measurements taken several times per day over the course of a few days or weeks). This strategy is taken because the stable processes are mean-reverting (Stroe-Kunold, Gruber, Stadnytska, Werner, & Brosig, 2012), meaning that the outcome variable may be higher or lower than the mean at specific measurement occasions, but values do not systematically change from the first measurement occasion to the last. That is, the mean is *stable* with respect to time even though values fluctuate over measurement occasions. If there is a trend in the data such that the mean does change, this is problematic for many types of models and the data are transformed or the model is altered to correct-out the trend over time (Jebb et al., 2015). Therefore, unlike developmental processes, many measurements are taken in quick succession to capture the *variability* of the outcome across time. It is precisely this *within-person variability* that is of interest with stable processes (Ram & Gertsoff, 2009) – Why do some people have taller peaks and lower valleys on the outcome? Which covariates predict spikes in the outcome variable? To what extent do spikes and valleys persist over time? The within-person deviations from the mean are primary interest with stable processes, which is opposed to developmental processes where this information is treated as noise (Wang et al., 2012).

Because the focus is on within-person variability instead of long-term change, taking many measurements close together does not affect the ability to address research questions of interest as there is no change for which researchers must wait. The outcome is dynamic and constantly in flux, so every new measurement provides new information about within-person variability (provided that the measure is sensitive enough to detect possibly subtle changes), regardless of how long it has been since the previous measurement (Nesselroade & Ram, 2004). For this type of data, *time-series* models tend to be applied.

### **Comparing Developmental and Stable Processes**

To visually compare how data from different processes appear, consider the two plots in Figure 1. The left panel shows a hypothetical trajectory plot for a developmental process and the right panel shows a trace plot for stable process data. The trajectory plot on the left shows a growth curve which captures how the means at each measurement occasion change over time. The trace plot on the right has a constant mean over time (the dashed horizontal line) with time-specific values fluctuating around this mean (i.e., it is mean-reverting). This exemplifies the difference between models for the different types of processes. Growth models are interested in change and covariates affecting change; time-series models are interested in variability and covariates predicting when values will deviate from the mean.

[Figure 1 about here]

In the next section, we begin with an overview of time-series models for stable process data collected from a single person to define some of the essential terminology. We follow this overview by then extending the model to data containing multiple people.

### ***N* = 1 Model for ILD**

Before extending the model to a multilevel structure where measurement occasions are nested within people, first consider  $N=1$  ILD where a single participant is repeatedly measured 50 times regarding their Urge to Smoke. Data for this type of analysis in *Mplus* should be structured in the ‘long’ format such that each measurement occasion occupies one row, the repeated measures are all contained in a single column, and there is a column containing an index variable for the time to which the measurement occasions corresponds. This is opposed to the ‘wide’ format used in some growth modeling frameworks (e.g., latent growth models) where each person occupies one row and each measurement occasion occupies a unique column. The time-series plot for the data that we will use were featured in the right panel of Figure 1 ( $M = -.03$ ,  $SD = 2.90$ ). The data, all syntax files, and

all output files are available on the first author's Open Science Framework page,

<https://osf.io/wuprx/>.

### Stationarity

Before applying any statistical model, it is important to assess some characteristics of the data. The first step is typically to check whether the series is *stationary* (Hamilton, 1994). For a series of a normally distributed to be stationary, three conditions must be satisfied (Chatfield, 2003):

1.  $E(y_t) = \mu$  : the expected value of the outcome is constant for any arbitrary time  $t$
2.  $Var(y_t) = \sigma^2$  : the variance of the outcome is constant at any arbitrary time  $t$
3.  $Corr(y_t, y_{t-j}) = \rho_j$  : the autocorrelation between outcome values at time  $t$  and time  $t - j$  (for  $j$  an arbitrary integer  $j < t$ ) is only dependent upon how far apart the values are in time. This means that the autocorrelation is the same between any two observations that are  $j$  measurement occasions apart, regardless of where they are located in the series.

Essentially, these three properties require that the mean, variance, and autocorrelations of the outcome do not systematically change over time. If these conditions are not met, then the process does not necessarily fit under the “Stable” umbrella and most of the models we discuss in this paper cannot be applied because effects will depend on time-specific behavior of the series.

One common violation of stationarity is a *trend* such that values are consistently increasing or decreasing across measurement occasions. Linear trends can be checked with a simple linear regression model with Time as the only predictor. For the example data here, the slope of time is quite close to zero and not statistically significant ( $\beta = -.01, p = .63$ ) which comfortably suggests that there is no linear trend over time. Researchers should also consider other types of non-linear trends such as seasonality (e.g., relations between measures at constant intervals) or cyclical trends (e.g., sine or cosine that can also violate stationarity. If a trend is present, the data must be de-trended so that the first property ( $E(y_t) = \mu$ ) is re-established. Linear trends and potential remedies are

discussed at the end of this paper and the discussion that follows assumes stationarity is reasonably upheld.

### **Bayesian Estimation**

In *Mplus*, the relevant models that we cover are sufficiently complex that Bayesian Markov Chain Monte Carlo (MCMC) is used for estimation because traditional frequentist methods like maximum likelihood (ML) often encounter convergence issues or are intractable. We will not fully review Bayesian statistics since they are used in *Mplus* more so as a computational tool rather than for its philosophical principles (Asparouhov et al., 2018). However, we will cover some relevant differences in terminology for readers who are not well-versed in Bayesian statistics to facilitate proper interpretation the output. Readers interested in a deeper understanding of Bayesian statistics are referred to Kruschke, Aguinis, and Joo, (2012), van de Schoot et al. (2014), or Zyphur, Oswald, and Rupp (2015) for pedagogical introductions. Readers familiar with Bayesian methods may skip this section without any loss of generality.

A major difference between ML and Bayesian MCMC is that ML yields a single point estimate for each parameter whereas MCMC yields an entire distribution of possible values for each unknown parameter. This distribution is referred to as the *posterior distribution* or just “the posterior” for short. This posterior distribution is formed by updating researchers’ prior beliefs for each parameter (prior distributions) with information from the data (the likelihood, the same likelihood used in ML estimation). If researchers do not have specific prior beliefs, *Mplus* will assign parameters with *uninformative* priors such that all admissible values are about equally plausible. Instead of a point estimate for each parameter, in MCMC, the values of the posterior distribution are summarized with a measure of central tendency (i.e., mean, median, or mode) to provide a single representative estimate for the parameter. By providing a posterior distribution instead of point estimates, MCMC provides more intuitive analogs of standard errors or confidence intervals that do not rely on assumptions or asymptotic theory. Instead, the MCMC analog of the frequentist standard

error is the standard deviation of the posterior and the analog of the frequentist confidence interval is the Bayesian credible interval, for which the bounds of a 95% interval are the 2.5 and 97.5 percentiles of the posterior. With MCMC, there are no null hypotheses for any effects as with ML and, consequently, there are no  $p$ -values.<sup>2</sup> To determine whether an estimate is null in the population, a common test is whether 0 is within the 95% credible interval for the parameter of interest (this method is often ineffective for variances because they are typically assigned priors that preclude negative values, so 0 is always outside the 95% credible interval).

Model fit operates differently with MCMC estimation than with ML. Whereas ML has familiar indices like RMSEA, SRMR, or CFI and likelihood ratio tests for comparison of nested models, *Mplus* does not provide Bayesian analogs of these methods. Instead, model comparisons are based on the *deviance information criteria* (DIC). The DIC operates similar to the AIC and BIC in frequentist contexts such that lower values of the DIC indicate relatively better model fit, even for non-nested models so long as the models have the same latent variables (Asparouhov et al., 2018). Readers are cautioned not to rely too heavily on the DIC with time-series data until further advances are made, however, because it tends to be unstable and can lead to different conclusions for different seed values for the Monte Carlo chains.

### **AR(1) Model**

When the outcome is repeatedly measured for the same individuals in a stationarity series, the data are typically *autocorrelated* within people such that the variable at time  $t$  for a particular person is correlated with itself at previous times for the same person (but especially so at times with the closest proximity). In such case, it follows that the variable at a previous measurement occasion

---

<sup>2</sup> The *Mplus* output does provide a Bayesian analog to a one-side  $p$ -value, though its interpretation does not directly coincide with the interpretation provided by a traditional frequentist  $p$ -value. The Bayesian  $p$ -value in *Mplus* corresponds to the proportion of the posterior distribution on the opposite side of 0 than the posterior mean. For example, a value of .03 for a positive estimate would mean that 3% of the posterior distribution is below 0 (Muthén, 2010 p. 7)

may serve as a good predictor of the same variable at the current measurement occasions. For instance, if Urge to Smoke is high the first measurement occasion, Urge to Smoke is likely to also be high at the second measurement occasion. Predicting a variable from itself at earlier measurement occasions is referred to as *autoregression*. Autoregressive Lag-1 – denoted by AR(1) – means that the outcome at time  $t$  is predicted by the outcome at the immediate preceding time and no other previous measurement occasions. The number of previous measurement occasions used for prediction can be increased; for instance, an AR(2) model would predict time  $t$  from both times  $t - 1$  and  $t - 2$ .

For the example data where the outcome is Urge to Smoke, the AR(1) model can be written as

$$Urge_t = \alpha + \varphi Urge_{t-1} + e_t . \quad (1)$$

This model states that the Urge to Smoke at time  $t$  is equal to an intercept  $\alpha$ , plus an autoregressive slope  $\varphi$  times Urge to Smoke at time  $t - 1$ , plus an normally distributed time-specific residual  $e_t \sim N(0, \sigma^2)$  that has constant variance over time. The mean of the series can be expressed as a

function of the intercept and the autoregressive slope, that is,  $\mu = \frac{\alpha}{(1-\varphi)}$ . The model can be fit in the

DSEM framework in *Mplus* with the following code

```
VARIABLE:
LAGGED = urge(1); !Create Lag-1 for Urge to Smoke;
ANALYSIS:
ESTIMATOR=BAYES;!Lagged variables can only be estimated with Bayes in Mplus;
BITERATIONS=(1000);!Run at least 1000 iteration of the MCMC algorithm;
MODEL:
urge ON urge&1; !Autoregressive effect, Phi;
urge; ! residual variance, sigma^2;
[urge]; !intercept value, alpha;
```

The `LAGGED = urge(1)` statement tells *Mplus* to create a new variable that is equal to `urge` at the previous measurement occasion. Including a `LAGGED =` statement tells *Mplus* to use DSEM, which requires the use of Bayesian estimation, so `ESTIMATOR=BAYES` is included in the

ANALYSIS command. The BITERATIONS=(1000) statement is not required, but can be helpful to ensure that the estimation is more stable; when the trace plots lead to doubts about whether the model converged, the user can increase the number of iterations.

In the MODEL command, Lag-1 Urge is specified by including "&1" after urge. It is assumed that the autoregressive effect is constant, so that a regression coefficient can capture the relation for any number of measurement occasions, for any arbitrary time. Table 2 shows the estimates for the model.

[Table 2 about here]

From Table 2, the interpretation is that Urge to Smoke at time  $t$  is positively related to Urge to Smoke at time  $t - 1$  such that a one-unit increase in Urge to Smoke at time  $t - 1$  is expected to increase Urge to Smoke at time  $t$  by .31 points. This model assumes no trend over time, so the

expected value for Urge to Smoke is a horizontal line at  $\frac{\mu}{(1-\varphi)} = \frac{-.03}{(1-.31)} = -.04$ . Urge to Smoke is

on a scale where 0 correspond to average so this person's Urge to Smoke is about average in this population. The outcome variable at each measurement occasions fluctuates around the horizontal line at -.04 with a variance of 8.18, after accounting for Lag-1 of Urge to Smoke.

### Time-Varying Covariates

The model can be extended to include time-varying covariates that are also collected at each measurement occasion to explain why Urge to Smoke rises and falls across measurement occasions. For instance, if a Z-scored measure of Depression was collected at each measurement occasion,<sup>3</sup> the model could be extended such that

---

<sup>3</sup> We use the time-varying covariate measured at the same time as the outcome, but note that causality is usually attributed to temporal precedence. It may be preferable to use the time-varying covariates from time  $t - 1$  instead of time  $t$  if the variables are truly measured simultaneously. Using the time-varying covariate from time  $t$  becomes more sensible when different variables at the same measurement occasion refer to differ timespans. For instance, simultaneity would be more justified if the outcome at time  $t$  were phrased "what is your urge to smoke right now?" while the time-varying covariate were phrased "what has your level of depression been since the last response?"

$$Urge_t = \alpha + \phi Urge_{t-1} + \beta Depression_t + e_t \quad (2)$$

Equation 2 builds upon Equation 1 by adding  $\beta$ , which captures the effect of Depression at time  $t$  on Urge to Smoke at time  $t$ . In *Mplus*, the code would be expanded to:

```
VARIABLE:
LAGGED = urge(1); !Create Lag-1 for Urge to Smoke;
ANALYSIS:
ESTIMATOR=BAYES;!Lagged variables can only be estimated with Bayes in Mplus;
BITERATIONS=(1000);!Run at least 1000 iteration of the MCMC algorithm;
MODEL:
urge ON urge&1; !Autoregressive effect, Phi;
urge ON dep; !Time-varying effect, Beta;
urge; ! residual variance, sigma^2;
[urge]; !intercept value, alpha;
```

To help visualize the model, a path diagram with the *Mplus* code superimposed on the paths shown in Figure 2. Even though there are 50 measurement occasions, the model reduces to two measurement occasions with at arbitrary times  $t$  and  $t - 1$ . This is accomplished by making an assumption that the autoregressive effect  $\phi$  and the time-varying covariate effect  $\beta$  are constant at each time-point. That is, the effect of Depression at Time 1 is constrained to be the same as the Depression effect at Time 2, and so on; this also becomes clear from the fact that these parameters do not have a subscript for time in Equation 2. This assumption is reasonable in many contexts but can be relaxed if necessary as well (Asparouhov et al., 2018). Unconstraining effects across time is discussed at the end of this paper, but note that it can present computational challenges.

[Figure 2 about here]

Table 3 shows the estimates for the model with Depression as a time-varying covariate. The horizontal mean line around which Urge to Smoke varies is equal to

$$\frac{\alpha}{(1-\phi)} + \frac{(\beta \times \overline{Dep})}{(1-\phi)} = \frac{.07}{(1-.35)} + \frac{(2.43 \times (-.03))}{(1-.35)} = .00 \text{ for } -.03 \text{ the mean of Depression for this}$$

person (even though the scale is standardized, the mean is not exactly 0). This value is slightly different from the results from the model in Table 2 because it is now conditional of Depression. The

autoregressive path (i.e.,  $\varphi$ ) in Table 3 (.35) is about the same as in Table 2 (.31). The effect of Depression (i.e.,  $\beta$ ) is quite large: A one-unit increase in Depression is predicted to yield a 2.43 unit increase in Urge to Smoke, indicating that this person is much more likely to feel a strong urge to smoke when they feel more depressed. The residual variance in Table 3 dramatically decreased (1.36) compared to Table 2 (8.18), indicating that Depression explains much of the variability in Urge to Smoke across measurement occasions.

[Table 3 about here]

As a reminder, note that the model is not directly accounting for mean changes – the intercept of Urge to Smoke (i.e.,  $\alpha$ ) is included as a constant across all 50 measurement occasions. There could be a trend in Urge to Smoke, if there is a trend in Depression. When Depression is not characterized by a trend, the model can be used to focus on the way that momentary fluctuations in Urge to Smoke are related to momentary fluctuations in Depression.

This type of  $N=1$  model is invaluable in disciplines like economics, finance, and meteorology where ILD are collected over time for a single entity. In psychology, such  $N=1$  research could be of interest if we want to closely monitor a particular individual, for instance with the purpose to intervene in time when there are indications that the individual is at increased risk for relapse. However, in psychology and related behavioral sciences, it is more typical to collect ILD on multiple people, meaning that  $N > 1$  and that the data structure becomes multilevel in nature such that measurement occasions are nested within people. For the remainder of the paper, we focus on the multilevel extension of models for ILD.

### **Two-Level DSEM for $N > 1$**

For researchers who associate longitudinal data with developmental processes and growth modeling, two-level models are a familiar modeling framework (Bliese & Ployhart, 2002; Singer, 1998). Several different models fall under the umbrella of “two-level model” (e.g., multilevel

models, mixed effect models, latent growth models, random coefficient models, or hierarchical linear models), and while there are some nuanced differences (e.g., McNeish & Matta, 2018), the underlying idea behind each of the models is the same (Curran, 2003). The average growth trajectory across all participants is modeled with fixed effects, while person-specific random effects are included to capture each person's deviation from the average trajectory. This results in each person in the data being described by a unique growth trajectory that follows the same functional form as the average growth trajectory of all participants but with potentially different values for the coefficients.

The basic two-level growth model can be written as<sup>4</sup>

$$\begin{aligned} y_{it} &= \beta_{0i} + \beta_{1i} \text{Time}_{it} + e_{it} \\ \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned} \quad (3)$$

The first equation specifies that the outcome variable at time  $t$  for person  $i$  ( $y_{it}$ ) is equal to the person-specific intercept for person  $i$  ( $\beta_{0i}$ ) plus a person-specific slope for person  $i$  ( $\beta_{1i}$ ) multiplied by the Time variable plus a person-specific residual at time  $t$  ( $e_{it}$ ) that captures the difference between the predicted value and the observed value. The second equation then specifies a mode for determining the person-specific intercept for person  $i$ . In this basic model, the person-specific intercept ( $\beta_{0i}$ ) is equal to the overall average intercept across all people,  $\gamma_{00}$ , plus a person-specific random intercept  $u_{0i}$ . The third equation is similar to the second in that it models the person-specific linear slope for Time. Similar to the second equation, in the basic model, the person-specific slope ( $\beta_{1i}$ ) is equal to the average linear slope across all participants,  $\gamma_{10}$ , plus a person-specific random

---

<sup>4</sup> The model can similarly be specified as a single-level model using a multivariate data structure (i.e., as a latent curve model), but we focus on the multilevel approach to better connect the model to the multilevel format specification used in DSEM.

slope  $u_{1i}$ . A notable assumption of the model is that the vector of random effects,  $\mathbf{u}_i = \begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix}$ , is

assumed to follow multivariate normal distribution with mean vector 0 and covariance matrix  $\boldsymbol{\tau}$  [i.e.,

$\mathbf{u}_i \sim MVN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$ . There is also an assumption about the within-time residual though it

is common to model these residuals as having constant variance [i.e.,  $e_{it} \sim (0, \sigma^2)$ ].

When extending DSEM to  $N > 1$ , two-level models can similarly be utilized. However, if the process is stationary, the focus is not on the association with mean change and the Time variable (or some function thereof) as in a growth model. Instead, increased attention is paid to the variability, especially within-person variability that is often of little consequence in growth models. The next section outlines how to extend the AR(1) model to  $N > 1$  with a two-level model, how the model can incorporate both time-varying and time-invariant covariates, how random effects can be extended to the residual variance, and how the model extends to multiple outcomes.

### Multilevel AR(1) Model

Though the focus of growth models and ILD models differ, the multilevel extension of each can be modeled within the same general two-level framework. Take the AR(1) model in Equation 2 featuring Urge to Smoke predicted by Lag-1 Urge to Smoke and a time-varying value of Depression. Instead of having 50 measurement occasions on a single person as in the previous sections, now imagine that there are 50 measurement occasions on 100 different people. The model in Equation 2 can be extended to a multilevel framework such that the multilevel AR(1) model would be

$$\begin{aligned}
 Urge_{it} &= \alpha_i + \varphi_i Urge_{(t-1)i} + \beta_i Dep_{it} + e_{it} \\
 \alpha_i &= \gamma_{00} + u_{0i} \\
 \varphi_i &= \gamma_{10} + u_{1i} \\
 \beta_i &= \gamma_{20} + u_{2i}
 \end{aligned} \tag{4a}$$

This basic multilevel AR(1) model has many similarities to the growth models to which psychologists are accustomed. The first equation specifies the *Within-Level model* such that the outcome variable Urge to Smoke at time  $t$  for person  $i$  is equal a person-specific intercept ( $\alpha_i$ ) plus a person-specific autoregressive term ( $\varphi_i$ ) multiplied by Urge to Smoke for person  $i$  at time  $t - 1$  plus a person-specific coefficient ( $\beta_i$ ) multiplied by the Depression score for person  $i$  at time  $t$  plus a time-point specific residual ( $e_{it}$ ) for person  $i$  which is assumed to be normally distributed with constant variance,  $e_{it} \sim N(0, \sigma^2)$ .

The Lag-1 predictor and the time-varying covariate have an ‘ $n$ ’ to denote that these variables are the raw, *non-centered* version of the variable. When extending to datasets with multiple people – as with multilevel models in general – how covariates in the Within-Level model are centered can have major consequences for the results, and changes the substantive interpretation of the results (e.g., Curran & Bauer, 2011; Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995; Raudenbush and Bryk (2002). The model in Equation 4a is not quite how *Mplus* fits the model by default, which is important to note if comparing results to other software like SAS Proc Mixed. We will return to centering and how it is handled in *Mplus* shortly.

Every coefficient in the first equation (the Within-Level equation) with only an  $i$  subscript subsequently becomes an outcome of a separate equation in the *Between-Level* model equations so that between-person differences in the person-specific coefficients can be modeled. In this basic Between-Level model, each person-specific coefficient is equal to the fixed effect ( $\gamma$ ) capturing the average effect across all people, and a person-specific random effect ( $u_i$ ) capturing the deviation of the specific person’s coefficient from the average.

The random effect covariance structure can be manipulated by the researcher. For this example, we assume uncorrelated random effects and constant residual variance such that

$$\mathbf{u}_i \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ 0 & \tau_{11} & \\ 0 & 0 & \tau_{22} \end{bmatrix} \right). \text{ As a difference between growth models and models for ILD, note}$$

that Equation 4a does *not* include Time as a covariate in the equation for  $Urge_{it}$ . This reflects the difference in focus: While growth curve modeling is focused on capturing the growth trajectories of individuals over time, the current model is focused on capturing the momentary dynamics of a process. The latter is concerned with the way the current state of a person can be predicted by their preceding state, and how external influences perturb the process.

**Centering.** When the focus of the model is on within-person quantities as in the models we discuss in this paper, it is important to distinguish within-person effects from between-person effects. That is, we want to interpret the covariates in a relative fashion for each person. For instance, imagine hypothetically that Person 1 has low Depression in general but that their Depression spikes at Time 6 (relative to their average Depression). If Depression is in its raw metric, this spike may appear uninteresting when compared to the entire dataset – high depression for someone who is, on average, low on Depression, may appear like an average Depression score when compared to the grand mean on Depression. However, if we want to understand the within-person process, we are interested in whether an elevated Depression score relative to the person’s mean (rather than the grand mean) predicts an increase in this person’s Urge to Smoke. In that case it is important to capture that – although the Depression value at Time 6 may be average in general – the Depression value at Time 6 *for Person 1* is high, relative to their typical value.

In the general multilevel modeling framework, this interest can be accommodated by *centering* covariates in the Within-Level model (e.g., Curran & Bauer, 2011; Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995). In particular, *person-mean centering* can be used such that each person’s mean value on the covariate (across all measurement occasions) is subtracted from the raw value at each measurement occasion. This rescales the covariate such that 0 represents the

average for each person. That is, after person-mean centering, Person 6 and Person 12 may have 0 values for Depression at the same measurement occasion, which would represent that each person is at their mean for the covariate value, even though the raw values of Depression are likely to be different for each person. Essentially, the application of person-mean centering rescales the covariate values so that the values represent deviation from the person's mean. Note that centering is not the same as standardization because it does not transform the variance to one or any other specific value.

A common approach to centering in multilevel models is to use *observed* person-mean centering whereby the sample mean for each person is calculated and subtracted from the raw value. For the Lag-1 Urge to Smoke covariate in Equation 4a, this would be conducted by calculating the mean of the outcome for each person across all measurement occasions,  $\overline{Urge}_i$  and subtracting this value from the raw Lag-1 Urge to Smoke:  $(Urge_{(t-1)i} - \overline{Urge}_i)$ . Similarly, the Depression score could be centered with the individual's sample mean:  $(Depression_{it} - \overline{Depression}_i)$ . However, this method can lead to multiple forms of bias in the estimates of  $\phi$  and  $\beta$  that can be avoided by using *latent* person-mean centering, which is the default in DSEM in Mplus (cf. Asparouhov et al., 2018; Nickell, 1981).

Specifically, *Mplus* will use latent person-mean centering for all variables that are not specified as either a within-level or between-level variable (i.e., that do not appear in a WITHIN= or BETWEEN= statement). Latent centering has the advantage that it: (a) provides more substantively interpretable within-person effects when compared to not centering or grand mean centering; (b) eliminates Nickell's bias and Lüdtke's bias<sup>5</sup> for the autoregressive effect and the effect(s) of other

---

<sup>5</sup> Nickell's Bias refers to the phenomenon where negative bias is present in autoregressive estimates when a lagged covariate is observed mean centered. Lüdtke's Bias is the phenomenon where contextual effect estimates are bias with observed mean centering due to possible unreliability in the observed mean.

time-varying covariates respectively; and (c) leads to an intercept that can be interpreted as the person's mean, which can be further modeled at the between level.<sup>6</sup>

Using Equation 4a as an example, the Within-Level lagged Urge variable and Depression covariate would be partitioned in a Within-Level component and a Between-Level component:

$$\begin{aligned} Urge_{(t-1)i}^n &= Urge_{(t-1)i}^c + Urge_i^b \\ Dep_{ii}^n &= Dep_{ii}^c + Dep_i^b \end{aligned} \quad (4b)$$

where  $Urge_{(t-1)i}^n$  and  $Dep_{ii}^n$  are the raw non-centered values of the lagged Urge variable and the Depression variable in the data for the  $i$ th person at the  $(t-1)$ th and  $t$ th measurement occasion respectively,  $Urge_{(t-1)i}^c$  and  $Dep_{ii}^c$  are the Within-Level components that have a mean of 0 by definition, and  $Urge_i^b$  and  $Dep_i^b$  are the Between-Level components that are latent variables with an estimated mean and variance. The latter parts are conceptually similar to the descriptive sample mean of Urge to Smoke and Depression, but creating latent variables for these components allows measurement error to be accounted for so that the mean estimate is as accurate as possible (Asparouhov & Muthén, 2018).

Thus, instead of Equation 4a, *Mplus* uses

$$\begin{aligned} Urge_{ii} &= \alpha_i + \varphi_i Urge_{(t-1)i}^c + \beta_i Dep_{ii}^c + e_{ii} \\ \alpha_i &= \gamma_{00} + u_{0i} \\ \varphi_i &= \gamma_{10} + u_{1i} \\ \beta_i &= \gamma_{20} + u_{2i} \\ Dep_i^b &= \gamma_{30} + u_{3i} \end{aligned} \quad (4c)$$

---

<sup>6</sup> An important note here is that the *Mplus* User's Guide at the time of this writing is based on Version 8.0 and provides examples using observed person-mean centering. Observed person-mean centering is conducted by including a variable in a WITHIN statement and then adding DEFINE: dep (GROUPMEAN) to the VARIABLE command. The examples we present utilize the latent person-mean centering approach given its aforementioned advantages (Asparouhov & Muthen, 2018), but the discrepancy between the current User's Guide and what is described in this paper is important to keep in mind.

The ‘c’ superscript is used to indicate that these covariates are person-mean centered. For each of the four Between-Level variables, the fixed effect represents the overall mean across persons (i.e., the  $\gamma$ 's), and a random effect that represents the variance of person-specific deviations around this mean (i.e., the  $u$ 's). Hence, the covariance matrix of the random effects will now be expanded to be  $4 \times 4$  instead of  $3 \times 3$ . This specification will yield estimates that have substantively meaningful within-person effects. For instance, after centering in this way, the intercept  $\alpha_i$  is equal to the mean of Urge to Smoke for individual  $i$ :  $\alpha_i = Urge_i^b$ . Otherwise, the effects in Equation 4a will contain a mix of non-partitioned within-person and between-person effects that cannot be disentangled (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002 p. 139).

**Path diagram & code.** The path diagram representation of the multilevel AR(1) model is presented in Figure 3. Because the data have a multilevel structure, the path diagram is split into two levels, each mirroring different lines of Equation 4c: the Within-Level features relations in the equation with  $Urge_{it}$  as the outcome and the Between-Level models the equations with  $\alpha_i$ ,  $\varphi_i$ ,  $\beta_i$ , and  $Dep_i^b$  as the outcomes. Following conventions from Curran and Bauer (2007) for visually representing models with multiple levels, the intercept, the Lag-1 slope, and the time-varying covariate slope all have superimposed circles.<sup>7</sup> This indicates that this path is treated as a latent variable with a mean and variance in the Between-Level model. Each circle in the Within-Level model becomes a latent variable in the Between-Level model. The intercept of the Between-Level latent variables are equivalent to fixed effects  $\gamma$  parameters in a mixed effect model equation.

[Figure 3 about here]

---

<sup>7</sup> In this notation and framework, any Within-Level predictor would have a circle drawn over the path because it will be a latent variable in the Between-Level model. If no random effect is desired for certain predictor in the Within-Level, the latent variable variance in the Between-Level model would be constrained to zero (but the latent variable would not be removed).

*Mplus* code for fitting the model in Equation 4 is shown below (on the top of the next page).

The Within-Level model corresponds to the  $Urge_{it}$  equation in Equation 4c. The `urge on dep` and `urge on urge&1` code informs *Mplus* that Urge to Smoke is predicted by latent person-mean centered Depression and latent person-mean centered Lag-1 Urge to Smoke. Prior to each of these lines, a vertical pipe is included to signify that these paths are person-specific, and they will be latent variables in the Between-Level model. The label that precedes the pipe may be designated with whichever label the user wishes because the latent variable is not observed; here we use Greek letters for clarity. In the terms of Equation 4c, the “`phi |`” and “`beta |`” labels signify that `phi` and `beta` will appear as latent variables in the Between-Level equation (i.e., that  $\varphi$  and  $\beta$  have  $i$  subscripts). The outcome variable by itself in the Within-Level model (i.e., “`urge ;`”) corresponds to the residual Within-Level variance,  $\sigma^2$ . There is no “`alpha |`” code in the Within-Level model because *Mplus* uses the outcome variable label to specify the intercept, the Between-Level intercept variance, and the residual variance.

```
VARIABLE:
LAGGED = urge(1) ;!Created a Lag-1 variable for urge to smoke;
CLUSTER = person;!The Between-Level ID variable is Person;
ANALYSIS:
TYPE = TWOLEVEL RANDOM; !Two-Level Model with paths that are latent variables;
ESTIMATOR = BAYES;!Lagged variables can only be estimated with Bayes in Mplus;
BITERATIONS = (1000);!Run at least 1000 iteration of the MCMC algorithm;
MODEL:
%WITHIN%
phi | urge ON urge&1;! urge is regressed on Lag-1 urge, the slope is latent;
beta | urge ON dep;! urge is regressed on Depression, the slope is latent;
urge; !within-level variance, sigma^2;
%BETWEEN%
[urge]; !mean intercept, gamma_00;
[phi]; !mean of autoregressive slope,gamma_10;
[beta]; !mean of TVC slope, gamma_20;
[dep]; !the overall mean of Depression, gamma_30;
urge; !intercept variance, tau_00;
phi; !autoregressive slope variance, tau_11;
beta; !TVC slope variance, tau_22;
dep; !Between-person variance of Depression, tau_33;
```

In the Between-Level model, each variable is modeled using an intercept and a variance. The intercepts of the latent variables are defined by placing their labels in square brackets: These correspond to the fixed effects  $\gamma_{00}, \gamma_{10}, \gamma_{20}$ , and  $\gamma_{30}$  in Equation 4c for  $\alpha_i, \varphi_i, \beta_i$ , and  $Dep_i^b$

respectively. The variance of the latent variables (the  $\tau$  parameters) are included by specifying the latent variable label by itself outside of brackets: So `urge`, `beta`, `phi`, and `dep` correspond to  $\tau_{00}$ ,  $\tau_{11}$ ,  $\tau_{22}$ , and  $\tau_{33}$  associated with Equation 4c. In the current model, all the random effects are uncorrelated, which may be considered a rather restrictive assumption; this can be relaxed by adding `WITH` statements to allow the random effects to covary with each other. To concretize the function of each line of code in this model, Figure 4 shows the path diagram for this model with *Mplus* code placed over the relevant paths.

[Figure 4 about here]

**Results & interpretation.** The results are presented in Table 4. The average horizontal mean line across all people around which Urge to Smoke varies is  $-.01$  and there is non-null variability around this line ( $.60$ ) meaning that some people have consistently higher or lower horizontal mean lines than the average. That is, there are between-person differences for average Urge to Smoke – we did not include a trend, and the line remains horizontal constant across time for all people. However, each person is characterized by their own constant deviation from this line. Following from the normality assumption of the latent variables, 95% of people in this data have person-specific intercepts between  $[-1.53, .151]$  (calculated from;  $-.01 \pm 1.96\sqrt{.60}$ ; the intercept plus and minus 1.96 deviations).

The average autoregressive coefficient across all people is  $.21$  and there is a small amount of variability ( $.02$ ). Between-person variance for the autoregressive path will tend to be small because the definition of stationarity requires an autoregressive parameter to be bound between  $-1$  and  $1$ , so small variances do not necessarily mean the between-person variance is unimportant. In these data, an average value of  $.21$  with a variance of  $.02$  would mean that 95% of people in this data have person-specific autoregressive coefficients between  $[-.07, .49]$ . Even though the variability estimate of  $.02$  seems small in absolute terms, the substantive result seems rather meaningful considering that

the autoregressive coefficient can only range from -1 to 1 and the [-.07, .49] range covers a large portion of admissible autoregressive coefficient values. Regarding the effect of Depression, the average effect across all people is .80 with a between-person variance of .79. This means that the 95% of the person-specific effects are between [-.94, 2.54]. The intercept and variance of  $Dep_i^b$  are also included in the output but are typically more important for proper person-mean centering of the Within-Level covariates than for their substantive effect on the outcome of interest. Covariances of  $Dep_i^b$  with other Between-Level latent variables could be included in the model, if they are of interest (e.g., to assess whether people who are generally higher on depression systematically have higher autoregressive coefficients).

[Table 4 about here]

Whether effects should be considered large or small, and whether the individual differences are large or small is difficult to ascertain. This distinction will depend on the variances of Depression and Urge to Smoke, which themselves may vary across individuals. As a result, even if two individuals have the same value for the effect of Depression on Urge to Smoke, this may mean something different when their variances differ. To make the interpretation easier, *Mplus* can also produce the standardized parameters and variance explained measures by including `OUTPUT : STANDARDIZED` in the code. Standardized within-person estimates are based on first standardizing the regression coefficients for each person separately, taking into account their particular variances on the covariate and the outcome, and then taking the average of this across individuals (it is also possible to output each individual's standardized parameter estimate). This approach was first proposed by Schuurman et al. (2016), and is referred to as *within-person standardization*. Standardizing between-person estimates is done using the between-person variances on the predictor and outcome variables. An  $R^2$  variance explained measure is also provided for each outcome variable. For outcomes in the Within-Level model, variance explained is calculated

separately for each person and then averaged across all people. Because Bayesian estimation is used, 95% credible intervals are provided for all standardized estimates and variance explained measures.

### **Time-Invariant Covariates**

The model in the previous section included two time-varying covariates (Lag-1 Urge to Smoke, Depression) to explain the variation in Urge to Smoke. Time-invariant covariates that are constant across all measurement occasions within a person but that vary between people can also be added to the model to predict intercepts, autoregressive slopes, or time-varying covariate slopes. For instance, imagine that measures of Job Stress and Home Stress were also collected in the Urge to Smoke data, but that these variables were collected only once at the beginning of the study instead of at each of the 50 measurement occasions. Each person will have a different value for Job Stress and Home Stress but the values will be constant for all measurement occasions within a person (e.g., the Job Stress value for Person 16 will be the same at the first measurement occasion as at the last measurement occasion). In this context, the variables would be time-invariant based on how they were collected and would be included into the model as such. Note that nearly any variable could be time-varying or time-invariant and the distinction is based upon how the data are collected rather than the nature of the variable itself.

Centering of time-invariant covariates functions differently than centering of time-invariant covariates. First, time-invariant covariates can only be centered around the grand-mean and not the person-mean. Because the value of the time-invariant covariates are constant within a person, centering around the person-mean would not make sense because it would yield a 0 value for every person in the data (i.e., the person-specific mean *is* the value of the variable for each person). In multilevel modeling in general, centering of time-invariant covariates can be done to improve interpretation of the intercept of the latent variable that is regressed on these predictors: When all predictors have a mean of zero (due to grand mean centering), the intercept of the outcome variable

is identical to its mean, as the intercept is the expected score when the predictors have a score of zero.

The model in Equation 4c can be expanded to accommodate time-invariant covariates in the Between-Level portion of the model:

$$\begin{aligned}
 Urge_{it} &= \alpha_i + \phi_i Urge_{(t-1)i} + \beta_i Dep_{it}^c + e_{it} \\
 \alpha_i &= \gamma_{00} + \gamma_{01} JobStress_i^c + \gamma_{02} HomeStress_i^c + \gamma_{03} Dep_i^b + u_{0i} \\
 \phi_i &= \gamma_{10} + \gamma_{11} JobStress_i^c + \gamma_{12} HomeStress_i^c + u_{1i} \\
 \beta_i &= \gamma_{20} + \gamma_{21} JobStress_i^c + \gamma_{22} HomeStress_i^c + u_{2i} \\
 Dep_i^b &= \gamma_{30} + u_{3i}
 \end{aligned} \tag{5}$$

Where the “c” superscripts on  $JobStress_i^c$  and  $HomeStress_i^c$  indicate that they have been grand-mean centered such that  $JobStress_i^c = JobStress_i^n - \overline{JobStress}$  and  $HomeStress_i^c = HomeStress_i^n - \overline{HomeStress}$ . Because time-invariant covariates do not vary within a person (i.e., they have “i” subscripts, not “it” subscripts), they can only appear in the Between-Level model whose equations feature an outcome variable that only have “i” subscripts as well. In Equation 5, Job Stress and Home Stress are included as predictors of the person-specific intercept, the person-specific autoregressive coefficient, and the person-specific time-varying covariate coefficient.

Also note that  $Dep_i^b$  has now been included as a covariate of the intercept in the  $\alpha_i$  equation in Equation 5. Including the latent person-mean centered  $Dep_{it}^c$  in the Within-Level model and the latent person mean  $Dep_i^b$  in the Between-Level model allows the Depression effect to be fully partitioned. This allows researchers to inspect whether a one-unit change in a covariate at a specific measurement occasion affects the outcome (the within-person effect) and how a one-unit change in the covariate mean across all measurement occasions affects the average of the outcome variable (the between-person effect). By including both effects in the model simultaneously, it is possible to investigate whether the within-person and between-person effects differ. For example, that might

mean that changes in Depression across measurement occasions could affect Urge to Smoke, reflecting a momentary within-person process, but having a generally high value of Depression (across all measurement occasions) may have no effect on people's average Urge to Smoke, meaning that if we interpret these person-specific means as traits, these traits are unrelated.

Substantively, the  $\alpha_i$  Between-Level equation models the person-specific intercept (the location of the horizontal mean line around which measurement occasions vary) as a function of Job Stress, Home Stress, and the person-specific mean on Depression. Perhaps people with more stress in their lives are more likely to have higher mean Urge to Smoke so their horizontal mean line is elevated. Note that these time-invariant covariates do not state that the Urge to Smoke linearly increases over time. Rather, it simply specifies a vertical shift in the location of the horizontal mean line as a function of Job Stress, Home Stress, and person-specific average on Depression. The  $\varphi_i$  equation models the autoregressive coefficient as a function of Job Stress and Home Stress to test whether people with more stress have larger or smaller carryover effects in their Urge to Smoke. The  $\beta_i$  equation similarly models whether the effect of Depression differs for people with different stress levels.

**Path diagram & code.** Figure 5 shows the path diagram for the model in Equation 5. In Figure 5, the time-invariant predictors are restricted to the Between-Level model because they do not vary within people and only predict differences between people.

[Figure 5 about here]

The *Mplus* code for the model including time-invariant covariates would be

```
VARIABLE:
Lagged = urge(1); !create variable for Lag-1 of Urge to Smoke;
BETWEEN = js hs;!Job Stress and Home Stress are time-invariant, used only in !the Between-Model;
CLUSTER = person; !The Between-Level ID variable is Person;
Define: CENTER js hs (GRANDMEAN);!Person-mean center to time-varying covariate to get a !within-
person effect;
ANALYSIS:
TYPE = TWOLEVEL RANDOM;!Two-Level Model with paths that are latent variables;
ESTIMATOR = BAYES;!Lagged variables can only be estimated with Bayes in Mplus;
BITERATIONS = (1000);!Run at least 1000 iteration of the MCMC algorithm;
MODEL:
```

```

%WITHIN%
phi | urge ON urge&1;! urge is regressed on Lag-1 urge, the slope is latent;
beta | urge ON dep;! urge is regressed on Depression, the slope is latent;
urge; !within-level variance, sigma^2;
%BETWEEN%
[urge]; !mean intercept, gamma_00;
[phi]; !mean of autoregressive slope,gamma_10;
[beta]; !mean of TVC slope, gamma_20;
[dep]; !the overall mean of Depression, gamma_30;
urge; !intercept variance, tau_00;
phi; !autoregressive slope variance, tau_11;
beta; !TVC slope variance, tau_22;
dep; !Between-person variance of Depression, tau_33;
urge ON js hs dep; !Job Stress and Home Stress predict the intercept,gamma_01 and gamma_02
!The latent person mean of Depression gives the !between-person effect of the TVC, gamma_03;
phi ON js hs;!Job Stress and Home Stress predict the autoregressive slope,gamma_11 and gamma_12;
beta ON js hs;!Job Stress and Home Stress predict the TVC slope,gamma_01 and gamma_02;

```

The difference between this code and the code in the previous section is that the time-invariant covariates must be declared in a BETWEEN= statement in the VARIABLE command. The time-invariant covariates have no within-person variance, so including these variables in the Within-Level model will yield a warning message. Specifying the time-invariant covariates in a BETWEEN= statement properly acknowledges that these variables only differ across people and will only appear in the portion of the model corresponding to between-person differences. In the Between-Level model, Job Stress (JS) and Home Stress (HS) are included as predictors of the Between-Level latent variables in a standard regression equation. Also in the VARIABLE command, a DEFINE statement is included to properly grand mean center the time-invariant covariates.

**Results & interpretation.** The model estimates are provided in Table 5. Each of the time-invariant effects were non-null. The covariates for the intercept show that the value of the person's mean (the horizontal line) around which people vary tends to be higher in people who experience more stress at their job or at home. A one-unit increase in Job Stress predicts an increase in the person's horizontal mean line of .50 points; a one-unit increase in Home Stress is associated with an increase in the person's horizontal mean line of .33 points. A similar pattern emerges for the autoregressive slope – both time-invariant coefficients are positive meaning that increases in Job Stress and Home Stress are expected to strengthen the carryover effect of Urge to Smoke (by .12 and .06 points, respectively). Increases in Job Stress and Home Stress are similarly predictive of a

stronger effect of Depression on Urge to Smoke such that people with higher stress level tend to respond with larger changes in their Urges to Smoke when their Depression changes (by .29 and .35 points, respectively). Put another way, the time-invariant covariates can be seen as moderators of the time-varying effects in the model, these are sometimes referred to as cross-level interactions, as they involve the interaction of a time-varying Within-Level covariate and a time-invariant Between-Level covariate. The person mean of Depression does not appear to have an effect on the individual intercept of Urge to Smoke ( $\gamma_{03}$ ), as 0 is within the 95% credible interval for the effect of regressing Alpha on Depression.

[Table 5 about here]

Note that the Between-Level variances decreased in this model compared to the estimates in Table 4. This is because part of the random variation from person to person captured by the Between-Level variance in Table 4 has been explained by the time-invariant covariates in Table 5. The intercept variance dropped from .60 to .34, meaning that Job Stress and Home Stress account for  $\frac{(.60 - .34)}{.60} = 43\%$  of the variability in people's means. Applying the same formula to the autoregressive slope and time-varying covariate slope yields explained variances of 50% and 19%, respectively (using estimates rounded to two decimal places). In the next section, we expand the model further to allow the residual variance to be person-specific rather than constant across all people by using location-scale models.

### **Multilevel Location-Scale Model**

The DSEM models presented in the previous section mirror two-level growth models and assume that the residual variance is the same for each person in the data. That is, the within-person variance  $\sigma^2$  had no “*i*” subscript meaning that it is the same for all people. Substantively, a constant residual variance across all people means that each person's behavior is equally predictable

(Jongerling et al., 2015). Abstractly, this would mean that – conditional on the covariates in the model – the peaks and valleys of the series have about the same height for each person. However, this may often be violated in psychological research because some individuals are less predictable and prone to larger swings across time compared to other individuals, possibly due to individual differences in sensitivity external factors (e.g., Wichers et al., 2009). Variability is directly of interest in models for ILD, so it is commonly of interest to know who is experiencing larger swings and if it is possible to predict why these large swings occur.

To demonstrate, Figure 6 shows overlaid trace plots for two people in the example Urge to Smoke data (Person 5 and Person 96). For both of these people, the average Urge to Smoke is somewhat similar (.32 for Person 5, -.08 for Person 96), both are stationary (the slope for Time is -.01 for both people, which is negligible), but there is a noticeable difference in the behavior of the trace plot over time. Person 5 (in grey) has much higher peaks, much lower valleys, and more variability across measurement occasions than Person 96. Because of the increased emphasis on variability in models for ILD, it is often important to allow the within-person residual variance to be person-specific and explicitly modeled rather than constant (as is typical in growth modeling). This type of model is referred to as a *multilevel location-scale model* (Hedeker, Mermelstein, & Demirtas, 2008, 2012; Jongerling et al., 2015; Wang et al., 2012). The model is called “location-scale” because it includes covariates both for the mean structure (the location) and also for the residual variance (the scale). It is possible to incorporate modeling features for these types of questions in multilevel software; however, prepackaged software like Proc Mixed are limited in their functionality (e.g., no random effects) or require users to write a custom likelihood in a procedure like SAS Proc Nlmixed (Hedeker et al., 2012). The DSEM framework can fully accommodate these types of modeling extensions with minimal additional code.

[Figure 6 about here]

The multilevel location-scale model looks similar to the model in the previous section except that the residual variance is now also an outcome in the Between-Level equation. Because the residual variance is a Within-Level quantity, it is permitted to be latent by taking an “*i*” subscript –  $e_i \sim (0, \sigma_i^2)$  – and become an outcome in the Between-Level model. This allows the residual variance to be person-specific and allows time-invariant covariates (or person means of time-varying covariates) to be added to model differences in residual variability for different people. For instance, Job Stress and Home Stress can be included as time-invariant covariates of the residual variance such that the additional Between-Level equation would be

$$\sigma_i^2 = \exp(\omega_0 + \omega_1 \text{JobStress}_i + \omega_2 \text{HomeStress}_i + u_{4i}) \quad (6)$$

The Between-Level component of Depression ( $Dep_i^b$ ) could also be included as a predictor of  $\sigma_i^2$ , but we omit it here (a) to keep the model and associated figures succinct and (b) for parsimony because the effect of  $Dep_i^b$  is null were it included in the model.

An important point in Equation 6 is that the variance is modeled as the exponential of the linear regression on the right; alternatively we could take the log on both sides, such that the log of the variance is modeled as a linear regression. This is necessary because variances cannot be negative, and using a log-linear model precludes negative values. It is important to remember this transformation when interpreting coefficients. In Equation 6  $\exp(\omega_0)$  is the expected residual variance for a person with 0 values on Job Stress and Home Stress (the mean residual variance if the covariates are centered),  $\exp(\omega_1)$  is the expected *multiplicative* change in the residual variance for a one-unit change in Job Stress, and  $\exp(\omega_2)$  is the expected *multiplicative* change in the residual variance for a one-unit change in Home Stress. Equation 6 also features a random effect –  $u_{4i}$  – which allows the within-person residual variance to vary randomly for reasons not explained by the included covariates. We emphasize “multiplicative” because using a log-linear model has implications

for interpretation, similar to other log-linear models like those for count data (e.g., Poisson or negative binomial). We will cover the difference in interpretation in more detail after fitting this model.

**Path diagram & code.** The path diagram for the multilevel location-scale model is shown in Figure 7. It is similar to the path diagram in Figure 5 but the within-person residual variance specified as latent variable,<sup>8</sup> allowing it to vary for each person.

[Figure 7 about here]

The DSEM code in *Mplus* would be:

```
VARIABLE:
LAGGED= urge(1);!Create Lag-1 variable for outcome;
BETWEEN = js hs;!Job Stress and Home Stress only appear in the Between-Level model;
CLUSTER = person;!The ID variable at the person-level is PERSON;
Define: CENTER js hs (GRANDMEAN); !Person-mean center to time-varying covariate to get a
!within-person effect;
ANALYSIS:
TYPE = TWOLEVEL RANDOM; !Two-Level Model with paths that are latent variables;
ESTIMATOR = BAYES;!Lagged variables can only be estimated with Bayes in Mplus;
ITERATIONS = (1000);!Run at least 1000 iteration of the MCMC algorithm;

MODEL:
%WITHIN%
phi | urge ON urge&1;!urge is regressed of Lag-1 urge, the slope is latent, phi_i;
beta | urge ON dep;!urge is regressed on Depression, the slope is latent, beta_i;
logv| urge;
!the Within-Level residual variance is latent, sigma^2_i;

%BETWEEN%
[urge]; !mean intercept, gamma_00;
[phi]; !mean of autoregressive slope,gamma_10;
[beta]; !mean of TVC slope, gamma_20;
[dep]; !the overall mean of Depression, gamma_30;
[logv]; !log of the mean of the residual variance, omega_0;
urge; !intercept variance, tau_00;
phi; !autoregressive slope variance not explained by JS and !HS, tau_11;
beta; !TVC slope variance not explained by JS and HS, tau_22;
dep; !Between-person variance of Depression, tau_33;
logv; ! between-person variance of sigma^2_i not explained by !JS and HS, tau_44;
urge ON js hs dep; !The person-specific intercept is predicted by Job Stress and !Home Stress
and the latent person mean of Depression;
phi ON js hs;
!The person-specific autoregressive slope is predicted by Job Stress and Home Stress;
beta ON js hs;!The person-specific TVC slope is predicted by Job Stress and !Home Stress;
logv ON js hs;
!The person-specific residual variance is predicted by Job !Stress and Home Stress;
```

The model looks similar to the model in the previous section featuring the time-invariant covariates.

The difference for this model is that the residual variance in the Within-Level model is now specified

---

<sup>8</sup> To clarify, the Within-Level residual (the difference between observed and predicted values) is a latent variable. The difference here is that the residual *variance* is also a latent variable in addition to the residual itself.

to be a latent variable and allowed to vary by person via the “logv|” prefix. The label does not need to begin with “log” nor does it need to contain “v”; whichever label this parameter is given, *Mplus* will note that this latent variable is associated with a variance and will use a log-linear model in the Between-Level model. In the Between-Level model, the intercept of the log residual variance is specified by placing the logv label in square brackets. Including logv by itself denotes that there is unexplained variability in the person-specific residual variances. The last line of the code specifies that the log residual variance is predicted by Job Stress and Home Stress (both grand-mean centered).

**Results & interpretation.** Table 6 shows the estimates of the model. The parameters new to the location-scale model are the  $\omega$  parameters involving the residual variance as an outcome in the Between-Level model and the person-specific variance,  $\tau_{44}$ . Recall that these parameters are on the log scale, so they must be exponentiated to place them back on the raw scale. The estimate of  $\omega_0$  is the expected log residual variance for a person who has mean values on Job Stress and Home Stress (as these predictors were grand mean centered); on the raw scale, the expected residual variance is  $\exp(\hat{\omega}_0) = \exp(.05) = 1.05$ . The between-person variance is .09, meaning that assuming normality, 95% of people’s residual variance are between  $\exp\left(.05 - 1.96\sqrt{.09}\right) = .58$  and  $\exp\left(.05 + 1.96\sqrt{.09}\right) = 1.89$ . In a log-linear model, the addition or subtraction must occur prior to apply the exponential function – it is not appropriate to exponentiate first and then add [e.g.,  $\exp(.05) + 1.96 * \exp(\sqrt{.09}) = 1.05 + 2.64 = 3.69$  is not the correct upper limit].

[Table 6 about here]

The raw residual variance is expected to change multiplicatively by  $\exp(\hat{\omega}_1) = \exp(.33) = 1.39$  for a one-unit change in Job Stress. Note, this means that a one-unit change in Job Stress predicts a residual variance of  $1.05 \times 1.39 = 1.46$  *not*  $1.05 + 1.39 = 2.44$ . The differential interpretation is the result of converting for a log-linear model (similar to Poisson

models). To emphasize this point, a two-unit change in Job Stress would predict a residual variance on the raw scale of  $1.05 \times 1.39^2 = 1.05 \times 1.39 \times 1.39 = 2.03$  - each additional one-unit increase in the covariate raises the exponent of the effect (the exponent does not need to be an integer). Similarly, a one-unit change in Home Stress is predicted to multiplicatively increase the residual variance by  $\exp(\hat{\omega}_2) = \exp(.05) = 1.05$ . The 95% credible interval for Home Stress contained 0, so the effect would be interpreted as being null. Substantively, this would mean that people with higher Job Stress are expected to have more unexplained within-person variability but higher Home Stress does not appear to affect within-person variability. In other words, Urge to Smoke is less predictable for people with higher values of Job Stress than people with lower reported Job Stress.

### Multilevel VAR (1) Model

Each of the models in the previous section have been univariate in nature such that there was a single outcome variable in the Within-Level model. In terms of the path diagram, this means that single-headed arrows only point to a single variable. Models for ILD may be extended to multiple outcomes simultaneously. The multilevel VAR(1) model is an example of a multivariate model that can easily be accommodated in the DSEM framework but that would be difficult to specify within standard multilevel modeling software like SAS Proc Mixed or `lme4` in R (the `m1VAR` package in R can accommodate this by fitting each outcome variable separately and then combining information together, Epskamp, Desrno, & Bringmann, 2017). One example would be the so-called Lag-1 *multilevel vector autoregressive model* [denoted as multilevel VAR(1)]. In the multilevel VAR(1) model, there are two outcome variables collected at each measurement occasion and the interest is whether the Lag-1 of either outcome predicts either outcome at the current time.<sup>9</sup> Using our running

---

<sup>9</sup> Note that the relation between Depression and Urge to Smoke has changed from a Lag-0 relation (when both variables are collected at the same measurement occasion) to a Lag-1 relation (Depression at the previous measurement occasion predicts Urge to Smoke at the current measurement occasion). This section intends to demonstrate the VAR(1) model using the same running example, but theoretical relations between variables should be considered when deciding which lag relation exists between variables.

example, the two outcomes would be Urge to Smoke and Depression. Urge to Smoke would be regressed on Lag-1 Urge to Smoke *and* Lag-1 Depression. Similarly, Depression would be regressed on Lag-1 Depression and Lag-1 Urge to Smoke. As in previous models, any of the intercepts or slopes can be specified to vary from person to person, as can the residual variances like in the location-scale model. A Within-Level portion of the path diagram for this model is shown in Figure 8.

[Figure 8 about here]

If we wanted to fit the multilevel VAR(1) model in Figure 8 with latent variables for the residual variances, this is a straightforward extension of DSEM in *Mplus*. The model equations would be

$$\begin{aligned}
 Urge_{it} &= \alpha_{1i} + \varphi_{1i} Urge_{(t-1)i}^c + \varphi_{4i} Dep_{(t-1)i}^c + e_{1it} \\
 Dep_{it} &= \alpha_{2i} + \varphi_{2i} Dep_{(t-1)i}^c + \varphi_{3i} Urge_{(t-1)i}^c + e_{2it} \\
 \alpha_{1i} &= \gamma_{00} + u_{0i} \\
 \alpha_{2i} &= \gamma_{10} + u_{1i} \\
 \varphi_{1i} &= \gamma_{20} + u_{2i} \\
 \varphi_{2i} &= \gamma_{30} + u_{3i} \\
 \varphi_{3i} &= \gamma_{40} + u_{4i} \\
 \varphi_{4i} &= \gamma_{50} + u_{5i}
 \end{aligned} \tag{7}$$

where  $e_{1it} \sim N(0, \sigma_{1i}^2)$  and  $e_{2it} \sim N(0, \sigma_{2i}^2)$ . Because both Urge to Smoke and Depression are outcome, both are latent person-mean centered by default such that  $Urge_{(t-1)i}^c = (Urge_{(t-1)i}^n - \alpha_{1i})$  and  $Dep_{(t-1)i}^c = (Dep_{(t-1)i}^n - \alpha_{2i})$ . The residual variances have  $i$  subscripts as in the location-scale model and become outcomes in the Between-Level model,

$$\begin{aligned}
 \sigma_{1i}^2 &= \exp(\omega_0 + u_{6i}) \\
 \sigma_{2i}^2 &= \exp(\omega_1 + u_{7i})
 \end{aligned} \tag{8}$$

This model is bivariate, so there are two residual variances – one for each outcome. Because the VAR(1) has multiple residual variances, covariances between the within-level residual random effects ( $u_{6i}$  and  $u_{7i}$  in Equation 8) can also be included if desired. We omit this term for simplicity, but more detail on how to expand the model to include it can be found in Section 4.2 of Hamaker et al. (2018). Similar to the location-scale model, the model for the person-specific residual variances is log-linear so that values cannot be negative. Covariates could be included in these equations as well. The Between-Level latent variable covariance matrix used for this model is  $\mathbf{u}_i \sim MVN(\mathbf{0}_8, \text{diag}[\tau_{00} \dots \tau_{77}])$  meaning that there is no correlation between the latent variables, though the latent variables are permitted to covary if desired.

The *Mplus* code for this model would be

```
VARIABLE:
LAGGED = urge(1) dep(1); !Created a Lag-1 variable for urge and depression;
CLUSTER=person; !The Between-Level ID variable is Person;
ANALYSIS:
TYPE = TWOLEVEL RANDOM; !Two-Level Model with paths that are latent variables;
ESTIMATOR = BAYES; !Lagged variables can only be estimated with Bayes in Mplus;
BITERATIONS=(1000); !Run at least 1000 iteration of the MCMC algorithm;
MODEL:
%WITHIN%
phi1 | urge ON urge&1; !urge to smoke is regressed on Lag-1 urge to smoke, phi_1i
!the slope is latent
phi2 | dep ON dep&1; !depression is regressed on Lag-1 depression, phi_2i the slope is latent;
phi3 | dep ON urge&1; !Depression is regressed on Lag-1 urge to smoke, phi_3i the slope is
latent;
phi4 |urge ON dep&1; !urge to smoke is regress on Lag-1 depression, phi_4ithe slope is latent;
logv1|urge; !the Within-Level residual variance of urge to smoke is latent, !sigma^2_1i;
logv2|dep; !the Within-Level residual variance of depression is latent, !sigma^2_2i;
%BETWEEN%
[urge]; !mean intercept of urge to smoke, gamma_00;
[dep]; !mean intercept of depression, gamma_10;
[phi1]; !mean of urge autoregressive slope, gamma_20;
[phi2]; !mean of depression autoregressive slope, gamma_30;
[phi3]; !mean of dep on urge&1 slope, gamma_40;
[phi4]; !mean of urge on dep&1 slope, gamma_50;
[logv1]; !log of the mean of the urge residual variance, omega_0;
[logv2]; !log of the mean of the urge residual variance, omega_1;
urge; !urge to smoke intercept variance, tau_00;
dep; !depression intercept variance, tau_11;
phi1; !urge to smoke autoregressive slope variance, tau_22;
phi2; !depression autoregressive slope variance, tau_33;
phi3; !dep on urge&1 slope variance, tau_44;
phi4; !urge on dep&1 slope, tau_55;
logv1; !between-person variance of urge to smoke residual variance, tau_66;
logv2; !between-person variance of depression residual variance, tau_77;
```

In the code for this model, both Urge and Depression are included in the LAGGED statement because each are predicted by their Lag-1 component. The Within-Level model includes more relations now

because there are two autoregressive slopes (urge ON urge&1 and dep ON dep&1) and two cross-lagged paths (urge ON dep&1 and dep ON urge&1). Each of these relations are specified to be person-specific by placing a latent variable label followed by a vertical pipe before the particular path. The remaining Between-Level model code is similar to previous DSEM models we have discussed: each latent variable in the Within-Level model receives an intercept by placing the label in square brackets and a between-person variance by placing the label by itself outside of square brackets.

**Results & interpretation.** Table 7 shows the estimates from this model. From Table 7, the results show that the Lag-1 Urge to Smoke affects Urge to Smoke at time  $t$  (0 is outside the credible interval for the mean of Phi1) but Lag-1 Depression does not affect Depression at time  $t$  (0 is inside the credible interval for the mean of Phi2). Lag-1 Urge to Smoke does affect Depression at time  $t$  (0 is outside the credible interval for the mean of Phi3) such that higher Urge to Smoke at time  $t-1$  predicts lower Depression at time  $t$ . Lag-1 Depression does not appear to affect Urge to Smoke at time  $t$  (0 is inside the credible interval for the intercept of Phi4). As in the Multilevel Location-Scale Model, the residual variance of both Urge to Smoke and Depression vary from person to person. The average residual variance of Urge to Smoke is  $\exp(.66) = 1.94$  but there is between-person variability as indicated by the .51 random effect variance estimate ( $\tau_{66}$ ). Similarly, the average residual variance of Depression is  $\exp(-.01) = .99$  but the between-person variability is much smaller with a random effect variance estimate of .01 ( $\tau_{77}$ ).

[Table 7 about here]

The default prior distributions in *Mplus* do not allow for variances in the Between-Level model to drop below 0, so the credible intervals of variances will never contain zero. As a result, these credible intervals are not always informative for determining whether there is a non-null between-person variance in particular paths. The Between-Level variances of  $\alpha_{2i}$ ,  $\varphi_{2i}$ ,  $\varphi_{3i}$ , and  $\varphi_{4i}$

each have credible intervals of [0.00, 0.01], suggesting that these paths hardly vary across people and the credible interval is only positive because *Mplus* is preventing inadmissible negative values.<sup>10</sup> The Lag-1 effect of Urge to Smoke on Urge to Smoke at time  $t$  ( $\varphi_{1t}$ ) shows between-person variability, meaning that some people have less carry-over than others do.

When considering the variances of the intercepts (i.e., the individual differences in person-specific mean over time), we see a large variance for the intercept of Urge to Smoke, which means that the horizontal mean lines around which values fluctuate within a person are considerably different across people. In contrast, the small variance of the intercept of Depression indicates that the horizontal mean line around which Depression scores vary is about the same across people (e.g., people do not exhibit much difference on average Depression). The large between-person variance of  $\sigma_{1i}^2$  indicates that there are marked differences across people in the degree to which their Urge to Smoke cannot be predicted (based on their Urge to Smoke and Depression at the preceding time point). The small between-person variance of  $\sigma_{2i}^2$  indicates that there are very small between-person differences in the unpredictability of Depression.

### **Issue to Consider: Trend in the Series**

At the onset of this paper, we mentioned that most models for ILD require the series to be stationary, one condition of which is that the expected value across the series is constant ( $E(y_t) = \mu$ ). Though the Urge to Smoke examples we have covered met this assumption, it is not uncommon for ILD to have a trend such that the expected value is a function of time (Stroe-Kunold et al., 2012). For instance, imagine that the context of the Urge to Smoke example data were changed

---

<sup>10</sup> This model did not feature any time-invariant covariates. If time-invariant covariates are included in the model, these Between-Level variances represent the unexplained variance present after accounting for effects of the covariates. Therefore, if the time-invariant covariates have non-null coefficients and the Between-Level variance are quite close to 0, it is still possible that there are between-person differences but that these differences are near-perfectly predicted by the time-invariant covariates present in the model.

such that participants quit cold turkey immediately preceding the observation window.

Hypothetically, the Urge to Smoke might decrease linearly over time and the time-series plot for a single person may hypothetically look like the left panel of Figure 9. If there is a constant linear trend over time, the series may be made stationary by *de-trending* the series. In  $N = 1$  context, a simple way to do this is to use a regression model with time as the covariate, save the residuals, and model the trend-free residuals instead of the raw outcome.

[Figure 9 about here]

This process is not always as simple to implement in  $N > 1$  settings because there are multiple residuals that could be used if de-trending with a two-level growth model. Rather than performing the analysis in two steps, the model can be estimated in a single step by including Time as a time-varying covariate in the model (e.g., Hamaker et al., 2018; Wang & Maxwell, 2015). To demonstrate with a multilevel AR(1) without time-invariant covariates, a trend could be accounted for with the following model,

$$\begin{aligned}
 Urge_{ii} &= \alpha_i + \varphi_i Urge_{(t-1)i}^c + \beta_{1i} Dep_{ii}^c + \beta_{2i} Time_{ii} + e_{ii} \\
 \alpha_i &= \gamma_{00} + u_{0i} \\
 \varphi_i &= \gamma_{10} + u_{1i} \\
 \beta_{1i} &= \gamma_{20} + u_{2i} \\
 \beta_{2i} &= \gamma_{30} + u_{3i} \\
 Dep_i^b &= \gamma_{40} + u_{4i}
 \end{aligned} \tag{9}$$

A simulation by Asparouhov (2018) shows that this method recovers the coefficients accurately but does not estimate the variance terms correctly. A complete explanation is given in Appendix D of Asparouhov et al. (2018), but the general idea is that it is best to model the trend separately from the autoregression to ensure stationarity. This can be done using *residual* DSEM (RDSEM) which allows for modeling the autoregression on the Within-Level *residuals* rather than on the variable itself. In equation form, the model would be

$$\begin{aligned}
 Urge_{it} &= \alpha_i + \beta_{1i} Dep_{it}^c + \beta_{2i} Time_{it} + e_{it} \\
 e_{it} &= \varphi_i e_{(t-1)i} + \delta_{it} \\
 \alpha_i &= \gamma_{00} + u_{0i} \\
 \varphi_i &= \gamma_{10} + u_{1i} \\
 \beta_{1i} &= \gamma_{20} + u_{2i} \\
 \beta_{2i} &= \gamma_{30} + u_{3i} \\
 Dep_i^b &= \gamma_{40} + u_{4i}
 \end{aligned} \tag{10}$$

Note that  $Urge_{(t-1)i}$  no longer appears as a covariate in the first equation; instead, the residual of the first equation appears as the outcome in the second equation and it is regressed on the residual at the previous time-point with some residual  $\delta_{it} \sim (0, \sigma^2)$ . This specification separates the trend ( $\beta_{2i}$ ) from the autoregression ( $\varphi_i$ ) into different equations. Similar to DSEM, the individual differences in the autoregression path can be further modeled as a latent variable in the Between-Level model. To make this process as clear as possible, Figure 10 shows the Within-Level path diagram for this model. *Mplus* code for this is listed below – note that RDSEM is available beginning in Version 8.1 or higher.

[Figure 10 about here]

```

VARIABLE:
LAGGED = urge(1); !Created a Lag-1 variable for urge to smoke;
WITHIN = time; ! Time is only used in the Within-Level model;
CLUSTER = person; !The Between-Level ID variable is Person;
ANALYSIS:
TYPE = TWOLEVEL RANDOM; !Two-Level Model with paths that are latent variables;
ESTIMATOR = BAYES; !Lagged variables can only be estimated with Bayes in Mplus;
BITERATIONS = (1000); !Run at least 1000 iteration of the MCMC algorithm;
Define: CENTER dep (GROUPMEAN); !Person-mean center to time-varying covariate to get a !within-
person effect;
MODEL:
%WITHIN%
beta | urge ON dep; ! urge is regressed on Depression, the slope is latent;
phi | urge^ ON urge^1; !The ` ^ ` symbol is shorthand for residual;
! The residual of Urge to Smoke is regressed on Lag-1 residual !of Urge to Smoke, the slope is
latent;
trend | urge ON time; ! urge is regressed on time to capture the trend over time, the !slope is
latent;
Urge; !within-level variance, sigma^2;

%BETWEEN%
[urge]; !mean intercept, gamma_00;
[phi]; !mean of autoregressive slope, gamma_10;
[beta]; !mean of TVC slope, gamma_20;

```

```
[trend]; !mean of the linear trend slope, gamma_30;
[dep]; !the overall mean of Depression, gamma_40;
urge; !intercept variance, tau_00;
phi; !autoregressive slope variance, tau_11;
beta; !TVC slope variance, tau_22;
trend; !linear trend slope variance, tau_33;
dep; !Between-person variance of Depression, tau_44;
```

### Issue to Consider: Unequally Spaced Measurement Occasions

The Urge to Smoke example data was ideally structured in that each of the 100 people had 50 measurements at the same occasions. However, this is not always the case, especially when the measurement occasions are very close together and not all participants respond at the same measurement occasions. For instance, some techniques like ecological momentary assessment intentionally use random measurement occasions so that participants cannot anticipate the timing of the next prompt. When the data are unequally spaced in time, or when participants are not measured at the same occasions, the model setup and interpretation provided in the Urge to Smoke example may be inaccurate. Specifically, the issue is that the two-level DSEM assumes that the Lag-1 effect is constant across all measurement occasions.

To explain, imagine the first few measurement occasions for Person 1 occurred at 11am, 3pm, and 4pm. If these times were coded at Time 1, Time 2, and Time 3, the Lag-1 effect for 11am to 3pm would be assumed to be the same as the effect of 3pm to 4pm in a two-level DSEM, even though the elapsed time is four times as long between Time 1 and Time 2 as Time 2 and Time 3. The two-level DSEM setup discussed earlier would not account for the fact that the distance between two adjacent measurement occasions is potentially different. This is highlighted in Figure 11. The standard DSEM operates as in the top panel such that measurement occasions are equally spaced. However, the proper interpretation would be closer to that shown in the bottom panel where there are unobserved potential measurement occasions between unequally spaced observed data. In the bottom panel, the carryover from 11am to 3pm would be equal to  $\varphi^4 = \varphi \times \varphi \times \varphi \times \varphi$  rather than  $\varphi$  as shown in the top panel.

[Figure 11 about here]

To accommodate this situation, users can use the `TINTERVAL` statement in the `VARIABLE` command of *Mplus*. In this statement, the time variable immediately follows `TINTERVAL =` with the desired time metric in parentheses. If the data were coded in one-hour increments such that 11am was coded as 11, 3pm as 15, and 4pm as 16,<sup>11</sup> including `TINTERVAL = time(1)` would be added to the code to rescale time into one hour increments. *Mplus* would do this by appending the dataset and inserting all increments of the parenthetical value into the data. In the hypothetical example where the “1” is specified in parentheses, *Mplus* would insert rows for `Time=12, 13, and 14` even though there are no observed data. The outcome is simply coded as missing. However, by appending the data in this way, it allows the autoregressive term to maintain a constant interpretation as shown in the bottom panel of Figure 11 –  $\varphi$  would be the carryover from one-hour prior rather than being some complex weighted average of the diverse intervals (De Haan-Rietdijk, Voelkle, Keijsers & Hamaker, 2017) as would be the case in the top panel of Figure 11.

The value included in parentheses of the `TINTERVAL` statement depends on how the data are coded. For instance, imagine the same data discussed in the preceding paragraph were coded daily instead of hourly. In this case, a one-hour increase would correspond to  $\frac{1}{24} \approx .04$  increase in time instead of 1, so .04 would be included in the parenthetical argument to get the one-hour lag effect instead.

Asparouhov et al. (2018) suggest that the transformed time interval be selected such that the interval be (a) interpretable from a substantive or theoretical point of view, (b) be large enough to not produce large amounts of missing data, and (c) be small enough to preserve the original time metric. The definition of “large” amounts of missing data is suggested to be 90% or larger in simulations by

---

<sup>11</sup> If coding this way, 0 would be the start of the study so 10am on day 2 would be coded as 34.

Asparouhov et al. (2018). For instance, if people are only measured twice per day, using an hourly interval may not be helpful because there will be 22 missing values, per person, per day. But choosing an interval of an entire day (24 hours) would be too large because it would not preserve the fact that multiple observations exist per day. Depending on when the measurement occasions were taken, perhaps it would be best to use 6 hour or 12 hour intervals to transform time in an interpretable manner.

Essentially, when time is treated discretely, the results will depend on the time interval that is used in the study because this will dictate the distance between measurement occasions and how lags are interpreted (Kuiper & Ryan, 2018), which has been referred to as the *lag problem* (Reichardt, 2011). The lag problem can be circumvented by adopting *continuous time models* rather than the operating in the discrete time framework we have used throughout this paper (e.g., Voelkle, Oud, Davidov, & Schmidt, 2012; Voelkle & Oud, 2013). By treating time as continuous, the benefit is that no interval needs to be specified, though the disadvantage is that estimation tends to be much more intensive and the model tends to be more difficult to specify in software. We will not go into detail about continuous time modeling here given the depth required to adequately discuss its nuances, but interested readers are referred to Ryan, Kuiper, and Hamaker (2018) for additional detail.

### **Discussion**

DSEM combines time-series techniques for long sequences of observations from a single person with multilevel structural equation modeling, thereby allowing for individual differences in parameters that define an individual's time series process. This makes DSEM particularly suitable for many forms of ILD in psychological research, as these data are formed from large numbers of repeated measures from multiple individuals. The structural equation modeling component implies that DSEM can handle multiple outcome variables and allows for the inclusion of latent variables in the model, in both the Within-Level and Between-Level models. Thus, while DSEM overlaps with existing and well-known approaches, it combines unique strengths of these different modeling

approaches, resulting in a flexible and sophisticated modeling framework that is specifically designed to tackle the challenges posed by ILD. Below, we compare DSEM to existing methods in more detail, noting where DSEM holds advantages over other traditional methods that researchers may alternatively consider. We also discuss potential drawbacks that researchers may want to note before embarking on a DSEM analysis.

### **DSEM Compared to Time-Series Analysis**

Like time-series analysis, DSEM was specifically designed to model the autocorrelation structure in ILD through the use of lagged regressions between observed variables and/or latent variables. However, time-series analysis is in essence an  $N = 1$  technique, which implies that when ILD are obtained from multiple individuals, one must perform separate  $N = 1$  analyses for each person. In contrast, DSEM combines the  $N = 1$  time-series techniques with multilevel modeling, allowing for the simultaneous modeling of ILD of multiple people. Specifically, DSEM features a time-series model in the Within-Level model while modeling individual difference in the time-series parameters (such as means, autoregression and cross-lagged regressions) as latent variables in the Between-Level model. As such, DSEM allows researchers to study differences in the dynamics of individuals in a straightforward way, all within a single model.

### **DSEM Compared to SEM**

Longitudinal data in psychology are often analyzed using SEM approaches such as latent growth curve modeling or cross-lagged panel modeling, which allow for the inclusion of multiple outcome variables as well as latent variables (McNeish & Matta, 2018). Such SEM approaches are based on having the data in wide format, where each row represents a separate person (or dyad) and the different waves are represented as different columns (and thus treated as different variables). Although SEM with data in wide format is suitable for traditional panel data that consist of a relatively small number of repeated measures (say, between two and ten), it becomes less practical as the number of time points increases, although this approach has been used for  $N=1$  time series data

(cf. Hamaker, Dolan & Molenaar, 2005). However, when the number of time points increases, especially when dealing with multivariate data, at some point, computational problems will emerge (Asparouhov & Muthén, 2015 p. 182). Alternatively, researchers have used a Toeplitz approach (e.g., Brown & Nesselroade, 2005; Gonzales & Ferrer, 2014, Molenaar, 1985), which is known to lead to moment estimates instead of maximum likelihood estimates (Hamaker, Dolan & Molenaar, 2003).

In contrast, DSEM is based on data in long format, where repeated measures are included as separate rows; this makes an increase in the number of time points arbitrary from a computational point of view. In addition, while SEM allows for individual differences in means or intercepts and for individual differences in the slope over time, the (lagged) regression coefficients are invariant across people (i.e., they cannot take random effects and must be modeled solely with fixed effects). In contrast, due to the multilevel features of DSEM, we can easily incorporate individual differences in (lagged) regression coefficients in our longitudinal analysis, and investigate how these differences are related to other individual differences.

### **DSEM Compared to Multilevel Modeling**

Multilevel modeling is a regression technique for analyzing data where observations are clustered within higher level units (Hox, Moerbeek, & van de Schoot, 2017). Longitudinal data are a special case where repeated measures are clustered in individuals (or dyads, etc.), and multilevel models are often used to model such data (Bolger & Laurenceau, 2013; Walls & Schafer, 2006). However, as indicated above, the key focus of time-series analysis and DSEM is the autocorrelation structure of the data. In DSEM, this is modeled through decomposing the data into within-person and between-person variance, and subsequently modeling the remaining autocorrelation in the within-person component through the use of lagged relations, either between a variable and itself at an earlier time point (autoregression), or between two different variables at different time points (cross-

lagged regression). To some extent, this approach can be mimicked in standard multilevel regression software (e.g., Bringmann et al., 2013; Kuppens, Allen & Sheeber, 2010; Rovine & Walls, 2005; Suls, Green & Hillis, 1998). Below, we briefly describe how such an analysis can be performed using standard multilevel regression software, and then we discuss the limitations of this approach.

**Including Lagged Regressions in Standard Multilevel Software.** To be able to study lagged relations and individual differences therein with ILD in standard multilevel software, researchers first need to create lagged variables that can be used as predictors. For instance, when considering the VAR(1) model in Equation 7, the researcher needs to create two lagged predictors:  $Urge_{(t-1)i}^n$  and  $Dep_{(t-1)i}^n$  (recall that an  $n$  superscript refers to the raw, non-centered version of the variable). Subsequently, the researcher should person-mean center these lagged predictors using the observed means ( $\overline{Urge}_i$  and  $\overline{Dep}_i$ , as this is the only mean that is available with multilevel modeling), to yield  $Urge_{(t-1)i}^{omc} = Urge_{(t-1)i}^n - \overline{Urge}_i$  and  $Dep_{(t-1)i}^{omc} = Dep_{(t-1)i}^n - \overline{Dep}_i$ , where the superscript “omc” refers to “observed mean centered”. Then, these centered predictors can be included as Within-Level predictors in a multilevel model. Though person mean centering in multilevel software is the only option, this leaves estimates susceptible to both Nickell’s bias and Ludtke’s bias. These disadvantages are overcome by DSEM in *Mplus*, as it uses latent mean centering by default to similarly partition within-person and between-person effects, while also guarding against these biases (Asparouhov et al., 2018).

**Unequal Intervals and Missing Observations.** Two common—and related—characteristics of ILD are unequal intervals between the observations and missing observations. Unequal intervals are a key feature of certain measurement designs, such as experience sampling, which is based on using random measurement times to ensure the participants are not anticipating the next measurement occasion. But even when the intervals between the measurement occasions are equal in the design, there will be some missing observations, which also result in unequal intervals between

the observations present in the data. When the interest is in lagged relations, this is a concern as it is well known that the strength of lagged relations (whether autoregressive or cross-lagged) depend on the time interval between occasions (Gollob & Reichardt, 1987; Ryan et al., 2018; Voelkle et al., 2012). That is, the more temporally distant two observations are, the less strongly they will tend to be related.

To account for different intervals between observations, researchers may decide to add missing values to the data file in order to make the distance between the observations (approximately) equal (cf. de Haan et al., 2017). However, this implies that when a lagged version of this variable is used as a predictor, there will also be a missing value on the predictor for the subsequent record. This means that the missing data problem is actually much larger than it would be in ordinary multilevel modeling without lagged regressions, and could lead to listwise deletion in programs that use conditional versions of maximum likelihood estimation (Bauer, 2003).

DSEM circumvents this problem by using a Kalman filter approach in which missing data are easily incorporated. Heuristically, the approach can be understood as follows (for details, see Hamaker & Grasman, 2012; Kim & Nelson, 1999). The Kalman filter makes a prediction of the next observation, based on the lagged predictors. This prediction is compared to the observation for that occasion and updated in light of it. If there is no observation, the Kalman filter simply continues with the prediction it had; if there is an observation, the Kalman filter continues with the updated prediction. In both cases, the filter moves forward to the next occasion and now makes a prediction based on the observations and predictions of the previous occasion.

Due to this procedure, none of the observations are lost, even if the majority of occasions are characterized by either a missing outcome or predictor (see Asparouhov et al., 2018). Furthermore, there is no need for the user to add records with missing data to make the records equally spaced in time; instead, the user can allow *Mplus* to handle the unequal intervals between observations by

adding a variable that represents time since the first observation, and use this in the `TINTEVAL` command described earlier.

**Multivariate Outcomes.** In comparison to standard multilevel modeling, DSEM extends easily to more complex types of models. For instance, the multilevel VAR(1) model has two (or more) outcome variables, a feature that is often not straightforward to accommodate within a single model with standard multilevel software and could require fitting a separate model to each outcome. Fitting separate models for each outcome can be particularly undesirable it does not allow for a direct estimation of the correlations between the innovations, nor of the correlations between the random coefficients associated with different outcome variables (for instance, the means and/or autoregressive parameters of different variables). Other multivariate models involving ILD would similarly require multiple univariate models to be combined, while DSEM allows for fitting a single, multivariate model.

**Latent Variables & Measurement Error.** The outcome variable and predictors in standard multilevel software are observed and each is featured within a column in the data. In contrast, this is not necessary with DSEM as the predictors and/or outcome(s) at each time could be any combination of observed and latent variables. This implies that autoregression can be performed by regressing the latent variable at time  $t$  on the same latent variable at time  $t - 1$ , while time-varying observed covariates could then predict latent variable at time  $t$ . If latent variables were treated with a multistage approach where latent variables are scored in the first stage and treated as known in the second stage, this could lead to biased coefficients (Devlieger, Mayer, & Rosseel, 2016) and a Croon correction is recommended (Croon, 2002).

While latent variables are typically measured using multiple indicators (and if so, researchers would need to assess measurement invariance to ensure that the latent variable has the same meaning over time), this need not be the case. Schuurman, Houtveen, and Hamaker (2015) note that

measurement error in observed variables may be a pervasive problem in ILLD, which can be partitioned out via latent variables, using only one observed indicator for the latent variable. This can also be done in the multilevel case with DSEM (Schuurman & Hamaker, 2019). Additionally, moving average components—which were not discussed in this paper, but are often considered in the time-series literature—can also be considered as latent variables: Moving average models include delayed effects of the residual at time  $t - 1$  on the outcome at time  $t$ . From the time-series literature it is known that an AR(1) model with measurement error can be rewritten as a model with a moving average component, although the reverse is not necessarily true (see also Asparouhov et al., 2018; Schuurman et al., 2015).

**Structural Relations in the Between-Level Model.** Standard multilevel software allows for observed time-invariant variables to be included at the between-level model, meaning that the latent variables can only be treated as outcomes (Sadikaj, Wright, Dunkley, Zuroff, & Moskowitz, 2019). In DSEM, one can specify any structural relations of interest in the Between-Level model, meaning that one can specify any relations desired using either latent variables or time-invariant observed variables, without restriction. That is, either latent or observed variables can be as predictors, mediators, or outcomes. Multivariate relations are permissible at the Between-Level, including factor models or a path models (such as mediation; Preacher, Zyphur, & Zhang, 2010). Essentially, DSEM removes constraints on what can be modeled in the Between-Level model compared to the relatively strict constraints of a standard multilevel model.

### **Advantages of DSEM in Comparison to Multilevel SEM**

*Mplus* includes the possibility of multilevel SEM, which combines aspects of multilevel models with aspects of SEM to allow an SEM at the Within-Level with latent variables that can be further modeled in the Between-Level model. Many of the advantages of DSEM in comparison to SEM or multilevel modeling that were discussed above are also possible with multilevel SEM (e.g., Sadikaj et al., 2019). The main advantage of DSEM over multilevel SEM is that DSEM is

specifically designed to model lagged relations and to account for the time-order of the observations. Typical DSEM tools that do not exist in multilevel SEM are the commands that creation lagged observed and latent variables, the `TINTERVAL` command that can be used to account for unequal intervals (Asparouhov et al., 2018), the Kalman filter foundation that handles missing observations in an elegant and appropriate manner (Hamaker & Grasman, 2012; Harvey, 1989), and many plotting options that can be used to investigate the observed and latent time sequences.

### **Potential Concerns with DSEM**

**Constant Autoregressive Coefficients.** Two-level DSEM models that we covered assume that the effects are constant between any arbitrary pair of measurement occasions. For instance, the autoregressive slope (averaged across individuals) for Urge to Smoke on Lag-1 Urge to Smoke in Table 7 was .19. This means that the autoregressive slope between Time 2 and Time 3 is .19, Time 3 and Time 4 is .19, and so forth. This similarly applies to any time-varying covariates such as the effect of Depression in our examples, which would be assumed to have a constant effect across time. Asparouhov et al. (2018) note that this is reasonable in a large number of application areas but may be violated in others.

If effects in the Within-Level model are thought to change across people *and* across time, this can be accommodated with a *cross-classified model* within *Mplus*. A cross-classified DSEM is similar to the time-varying effect model proposed in Shiyko, Lanza, Tan, Li and Shiffman (2012; see also Bringmann, Ferrer, Hamaker, Borsboom & Tuerlinckx, 2018), except that the time-specific effects are captured with random effects in DSEM, meaning that researchers do not need to specify the functional form of the effects over time. Such a cross-classified DSEM model would estimate the average autoregressive coefficient over all measurement occasions but would allow each measurement occasion to have a potentially different effect so that the values are not constrained (Asparouhov et al., 2018).

**Heterogeneity in Series.** DSEM follows the random effect tradition, which means that it has implicit assumptions about time-series processes across people (Nesselroade & Molenaar, 1999). Specifically, DSEM, like most multilevel models, posits that random effects (e.g., means, autoregressive parameters, cross-lagged regression parameters) that follow a multivariate normal distribution (Molenaar, 2004). Researchers may be concerned about this assumption, and wonder about possible ramifications when the random effects are actually not normally distributed, but perhaps are formed for instance a mixture of two normal distributions.

If one has a variable that may be considered as predictive of this mixture (e.g., gender or age group), this could be included as a predictor in the Between-Level model (Voelkle, Brose, Schmiedek, & Lindenberger, 2014). Alternatively, a two-step approach may be employed, which similarly follows suggested practice for clustered data prior to the proliferation of multilevel models. Here, this process involves modeling each person separately (i.e., running  $N$  single-subject time-series models, either with DSEM or with standard time-series analysis software), saving the estimates for each person, and then performing subsequent analyses on these estimates. This approach may however be rather time-consuming, and it is difficult to combine the results in a single analysis afterwards, as one should account for the uncertainty of the individual parameter estimates before combining them. Also, this may not be necessary – even if the random effects are not normally distributed in the population: When the individual time-series are of sufficient length (say 30 observation or more), the normality assumption is overruled by the data, and one can find non-normal distributed random effects when plotting the person-specific estimates, which is possible with DSEM in *Mplus*. However, this does require researchers to obtain these estimates and inspect plots created from them, which arguably should become default practice when employing DSEM.

### Concluding Remarks

This paper intended to serve as a basic introduction to two-level DSEM, how these models are fit in *Mplus*, and how the estimates are interpreted. Given our intended purpose and audience, we

admittedly did not cover some of the more advanced features of DSEM framework and constrained the discussion to two-level models with observed outcomes. The DSEM framework extends further than what is presented here and we strongly encourage readers interested in more advanced models to consult Asparouhov et al. (2018) for a complete and nuanced treatment of these capabilities within *Mplus* once they feel they have a grasp on the basic elements of DSEM.

The DSEM framework is broad and encompasses many models. All nuances of DSEM cannot be covered within a single paper and we only scratched the surface with the models that we covered. However, we hope that this primer provides readers with a foundational understanding of DSEM needed to dive deeper in this expanding literature.

## References

- Asparouhov, T. (2018). *Dynamic structural equation modeling of intensive longitudinal data using Mplus version 8*. Retrieved from <http://www.statmodel.com/download/Part%205%20Asparouhov.pdf>
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling, 25*, 359-388.
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic latent class analysis. *Structural Equation Modeling, 24*, 257-269.
- Asparouhov, T. & Muthén, B.O. (2015). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J.R. Harring, L.M. Stapleton & S.N. Beretvas (Eds.), *Advances in Multilevel Modeling for Educational Research* (pp. 163-192). Information Age Publishing.
- Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28*, 135-167.
- Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment, 23*, 447-458.
- Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods, 5*, 362-387.
- Boker, S. M., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., ...J., F. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*, 306-317.
- Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford Press.
- Bringmann, L. F., Ferrer, E., Hamaker, E. L., Borsboom, D., & Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivariate Behavioral Research, 53*, 293-314.
- Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS One, 8*: e60188.
- Browne, M.W., & Nesselroade, J.R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In A. Maydeu-Olivares & J.J. McArdle (Eds.), *Advances in psychometrics: A Festschrift for Roderick P. McDonald* (pp. 415-452). Mahwah, NJ: Erlbaum.
- Chatfield, C. (2003). *The analysis of time series: an introduction*. CRC press.

- Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505-528.
- Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*, 327-337.
- Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195- 223). Mahwah, NJ: Lawrence Erlbaum
- 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.
- Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods*, *9*, 220-237.
- Curran, P. J., & Bauer, D. J. (2007). Building path diagrams for multilevel models. *Psychological Methods*, *12*, 283-297.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along?. *Multivariate Behavioral Research*, *38*, 529-569.
- Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121-136.
- De Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., and Hamaker, E. L. (2017). Discrete- vs. Continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849, 19p.
- Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*, 741-770.
- Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*, 1-35.
- Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: an introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass*, *3*, 979-991.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*, 121-138.
- Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2017). mlVAR: multi-level vector autoregression (R package version 0.4). Retrieved from <https://cran.r-project.org/package=mlVAR>

- Gates, K.M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *65*, 310–319.
- Gonzales, J. E., & Ferrer, E. (2014). Individual pooling for group-based modeling under the assumption of ergodicity. *Multivariate Behavioral Research*, *49*, 245-260.
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. New York: Guilford Press
- Grimm, K. J., & Widaman, K. F. (2010). Residual structures in latent growth curve modeling. *Structural Equation Modeling*, *17*, 424-442.
- Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*, 80–92.
- 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*, advance online publication.
- Hamaker, E. L., Ceulemans, E., Grasman, R. P. P. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, *7*, 316-322.
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points T exceeds the number of cases N: Raw data maximum likelihood. *Structural Equation Modeling*, *10*, 352-379.
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*, 207-233.
- Hamaker, E. L., & Grasman, R. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492.
- Hamaker, E. L. & Grasman, R. P. P. P. (2012). Regime switching state-space model applied to psychological processes: Handling missing data and making inferences. *Psychometrika*, *77*, 400-422.
- Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*, 10-15.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University

Press.

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009). Modeling mood variation associated with smoking: an application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction, 104*, 297-307.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine, 31*, 3328-3336.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics, 64*, 627-634.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis techniques and applications*. Mahwah, NJ: Routledge.

Jebb, A. T., & Tay, L. (2017). Introduction to time series analysis for organizational research: Methods for longitudinal analyses. *Organizational Research Methods, 20*, 61-94.

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. *Frontiers in Psychology, 6*, 727.

Jongerling, J., Laurenceau, J. P., & Hamaker, E. L. (2015). A multilevel AR (1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research, 50*, 334-349.

Kim, C.J., & Nelson, C.R. (1999). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. Cambridge, MA: MIT Press.

Kreft, I. G., De Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*, 1-21.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods, 15*, 722-752.

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-Considering the role of the time-interval. *Structural Equation Modeling, 25*, 809-823.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science, 21*, 984-991.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*, 203-229.

McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods*, *50*, 1398-1414.

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114-140.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201-218.

Molenaar, P. C. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181-202.

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*, 112-117.

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, *17*, 313-335.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92-105). Washington, DC: American Psychological Association.

Nesselroade, J.R., & Molenaar, P.C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R.H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223–250). Thousand Oaks, CA: Sage.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*, 9-29.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 1417-1426.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2016). Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model. *Multivariate Behavioral Research*, *51*, 106–119.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic difference equation model for affective dynamics. *Psychological Methods*, *16*, 468–490.

Ou, L., Hunter, M., & Chow, S.-M. (2018). dynr: Dynamic modeling in R. (R-package version 0.1.12-5). Retrieved from: <https://cran.r-project.org/web/packages/dynr/>

Preacher, K. J., & Hancock, G. R. (2015). Meaningful aspects of change as novel random coefficients: A general method for reparameterizing longitudinal models. *Psychological Methods, 20*, 84-101.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods, 15*, 209-233.

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging, 24*, 778.

Reichardt, C. S. (2011). Commentary: Are three waves of data sufficient for assessing mediation? *Multivariate Behavioral Research, 46*, 842-851.

Rovine, M. J., and Walls, T. A. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. In T. A. Walls and J. L. Schafer (Eds.). *Models for Intensive Longitudinal Data* (p. 124-147). New York: Oxford University Press.

Ryan, O., Kuiper, R.M., & Hamaker, E.L. (2018). A continuous time approach to intensive longitudinal data: What, Why and How? In K.V. Montfort, J.H.L. Oud, & M.C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 27-54). New York: Springer.

Sadikaj, G., Wright, A. G., Dunkley, D., Zuroff, D., & Moskowitz, D. S. (2019). Multilevel structural equation modeling for intensive longitudinal data: A practical guide for personality researchers. *Psyarxiv preprint*, <https://doi.org/10.31234/osf.io/hwj9r>.

Shmueli, G. (2010). To explain or to predict?. *Statistical Science, 25*, 289-310.

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n= 1 psychological autoregressive modeling. *Frontiers in Psychology, 6*, 1038.

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods, 21*, 206-221.

Schuurman, N. K. & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive models. *Psychological Methods, 24*, 70-91

Sharpe, D. (2013). Why the resistance to statistical innovations? Bridging the communication gap. *Psychological Methods, 18*, 572-582.

Shiyko, M. P., Lanza, S. T., Tan, X., Li, R., & Shiffman, S. (2012). Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: Differences between successful quitters and relapsers. *Prevention Science, 13*, 288-299.

- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323-355.
- Stroe-Kunold, E., Gruber, A., Stadnytska, T., Werner, J., & Brosig, B. (2012). Cointegration methodology for psychological researchers: An introduction to the analysis of dynamic process systems. *British Journal of Mathematical and Statistical Psychology, 65*, 511-539.
- Suls, J., Green, P., and Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin, 24*, 127-136.
- Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science, 23*, 466-470.
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*, 842-860.
- Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research, 49*, 193-213.
- Voelkle, M. C., & Oud, J. H. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology, 66*, 103-126.
- Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. *Psychological Methods, 17*, 176-192.
- Walls, T. A., & Schafer, J. L. (Eds.). (2006). *Models for intensive longitudinal data*. New York, NY: Oxford University Press.
- Wang, L. P., Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods, 17*, 567-581.
- Wang, L.P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Multivariate Behavioral Research, 20*, 63-83
- Wichers, M., Geschwind, N., Jacobs, N., Kenis, G., Peeters, F., Derom, C., ... & van Os, J. (2009). Transition from stress sensitivity to a depressive state: longitudinal twin study. *The British Journal of Psychiatry, 195*, 498-503.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*, 1100-1122.

Zyphur, M. J., Oswald, F. L., & Rupp, D. E. (2015). Rendezvous overdue: Bayes analysis meets organizational research.

Table 1  
*Comparison of characteristics of developmental and stable process*

	Developmental Processes	Stable Processes
Number of Measurement Occasions	Few (3-8)	Many (10+)
Spacing of Measurements	Further apart (months or years)	Closer together (hours or days)
Longitudinal Focus	Change in outcome across entire observation window	Instantaneous changes from time-point to time-point
Mean Change	Primary focus; goal is to capture changes across the observation window	Non-interest because the process is mean-reverting and stable across the observation window. A nuisance to accommodate if present
Variability	Secondary focus to mean change and typically regarded as noise; commonly not reported	Variability around a stable mean is the main focus; goal is capture why there are peaks and valleys across time
Covariate Focus	Variables that change the growth curve (explaining between-person variability)	Variables predicting fluctuations from time to time (explaining within-person variability)

Table 2

*Estimates and 95% credible intervals for N = 1 AR(1) model*

Effect	Notation	Posterior Median	95% Credible Interval
Urge on URGE&1	$\varphi$	0.31	[.03,.60]
Intercept (URGE)	$\alpha$	-0.03	[-.81,.81]
Res. Var. (URGE)	$\sigma^2$	8.18	[5.59,12.26]

Table 3

*Estimates and 95% credible intervals for  $N = 1$  AR(1) model with a time-varying covariate*

Effect	Notation	Posterior Median	95% Credible Interval
URGE on URGE&1	$\varphi$	.35	[.25, .46]
Intercept (URGE)	$\alpha$	.07	[-.24, .39]
URGE on DEP	$\beta$	2.43	[2.12, 2.75]
Res. Var. (URGE)	$\sigma^2$	1.36	[.92, 2.20]

Table 4

*Estimates and 95% credible intervals for multilevel AR(1) model*

Effect	Notation	Posterior Median	95% Credible Interval
Intercept (Alpha)	$\gamma_{00}$	-.01	[-.18, .16]
Intercept (Phi)	$\gamma_{10}$	.21	[.17, .24]
Intercept(Beta)	$\gamma_{20}$	.80	[.61, .95]
Intercept(Dep)	$\gamma_{30}$	.01	[-.02, .04]
Var. (Alpha)	$\tau_{00}$	.60	[.44, .83]
Var. (Phi)	$\tau_{11}$	.02	[.01, .03]
Var. (Beta)	$\tau_{22}$	.79	[.61, .95]
Var. (Dep)	$\tau_{33}$	.01	[.00, .01]
Res. Var (Urge)	$\sigma^2$	1.14	[1.09, 1.19]

Table 5

*Estimates and 95% credible intervals for multilevel AR(1) model with time-invariant covariates*

Effect	Notation	Posterior Median	95% Credible Interval
Intercept (Alpha)	$\gamma_{00}$	.06	[-.14,.30]
Intercept (Phi)	$\gamma_{10}$	.19	[.16,.22]
Intercept(Beta)	$\gamma_{20}$	.79	[.62,.95]
Intercept(Dep)	$\gamma_{30}$	.02	[-.01, .05]
Alpha on Job Stress	$\gamma_{01}$	.50	[.35,.65]
Alpha on Home Stress	$\gamma_{02}$	.33	[.19,.46]
Alpha on Dep	$\gamma_{03}$	-2.35	[-11.22, 4.35]
Phi on Job Stress	$\gamma_{11}$	.12	[.09,.15]
Phi on Home Stress	$\gamma_{12}$	.06	[.03,.08]
Beta on Job Stress	$\gamma_{21}$	.29	[.11,.48]
Beta on Home Stress	$\gamma_{22}$	.35	[.17,.51]
Var. (Alpha)	$\tau_{00}$	.34	[.16,.49]
Var. (Phi)	$\tau_{11}$	.01	[.00,.01]
Var. (Beta)	$\tau_{22}$	.64	[.47,.88]
Var. (Dep)	$\tau_{33}$	.01	[.00, .01]
Res. Var. (Urge)	$\sigma^2$	1.14	[1.09,1.18]

Table 6

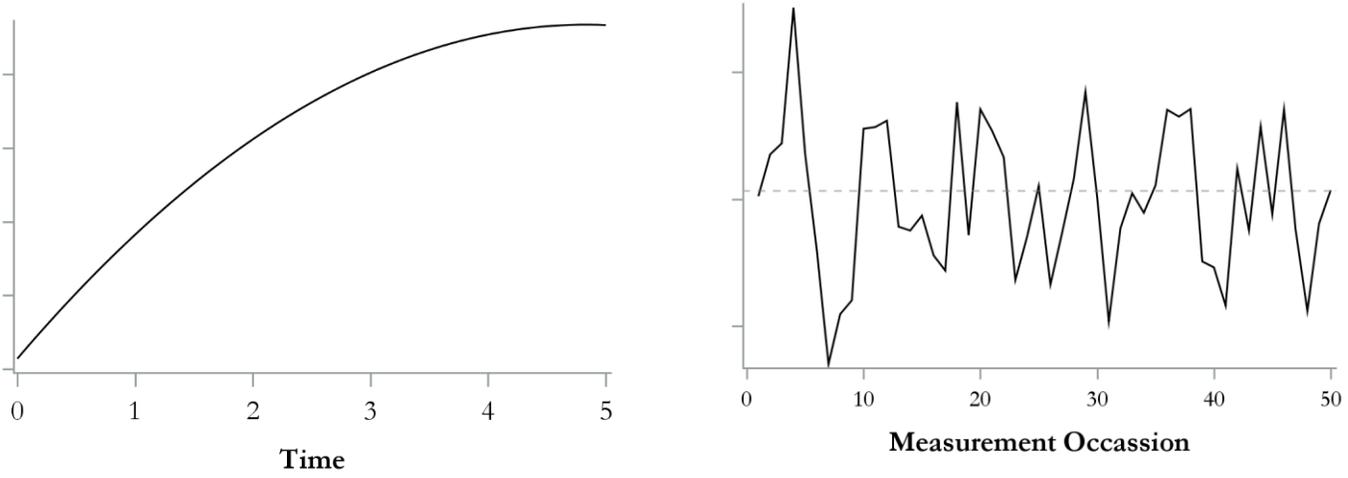
*Estimates and 95% credible intervals for multilevel location-scale model*

Effect	Notation	Posterior Median	95% Credible Interval
Intercept (Alpha)	$\gamma_{00}$	.03	[-.18, .26]
Intercept (Phi)	$\gamma_{10}$	.19	[.17, .22]
Intercept(Beta)	$\gamma_{20}$	.79	[.64, .95]
Intercept(Dep)	$\gamma_{30}$	.02	[-.01, .05]
Intercept(LogV)	$\omega_0$	.05	[-.04, .12]
Alpha on Job Stress	$\gamma_{01}$	.50	[.34, .65]
Alpha on Home Stress	$\gamma_{02}$	.31	[.18, .44]
Alpha on Dep	$\gamma_{03}$	-.77	[-8.62, 8.08]
Phi on Job Stress	$\gamma_{11}$	.12	[.09, .15]
Phi on Home Stress	$\gamma_{12}$	.06	[.03, .09]
Beta on Job Stress	$\gamma_{21}$	.28	[.10, .49]
Beta on Home Stress	$\gamma_{22}$	.35	[.19, .51]
LogV on Job Stress	$\omega_1$	.33	[.23, .40]
LogV on Home Stress	$\omega_2$	.05	[-.03, .12]
Var. (Alpha)	$\tau_{00}$	.34	[.21, .47]
Var. (Phi)	$\tau_{11}$	.01	[.00, .01]
Var. (Beta)	$\tau_{22}$	.63	[.49, .86]
Var. (Dep)	$\tau_{33}$	.01	[.00, .01]
Var. (LogV)	$\tau_{44}$	.09	[.06, .15]

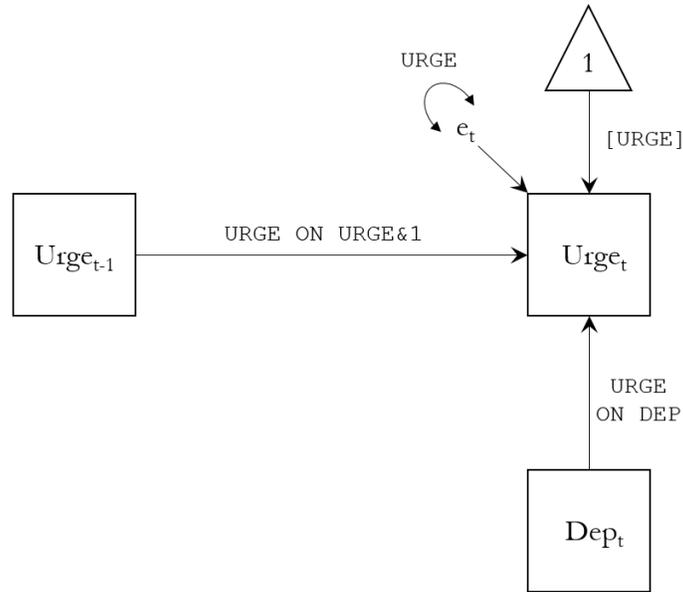
Table 7

*Estimates and 95% credible intervals for multilevel VAR(1) model*

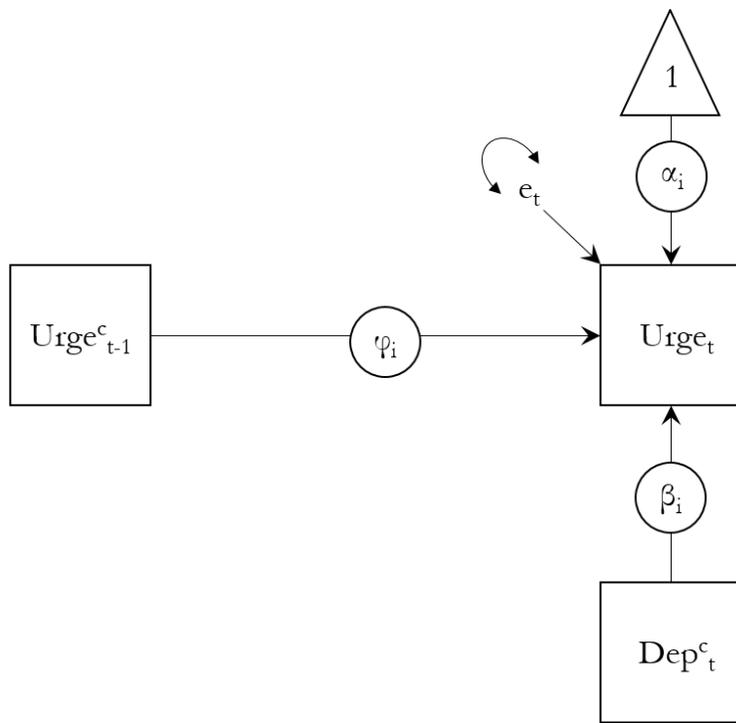
Effect	Notation	Posterior Median	95% Credible Interval
Intercept (Alpha Urge)	$\gamma_{00}$	-.01	[-.19, .15]
Intercept (Alpha Dep)	$\gamma_{10}$	.01	[-.02, .05]
Intercept(Phi1)	$\gamma_{20}$	.18	[.13, .22]
Intercept(Phi2)	$\gamma_{30}$	.01	[-.03, .05]
Intercept(Phi3)	$\gamma_{40}$	-.03	[-.05, -.01]
Intercept(Phi4)	$\gamma_{50}$	.02	[-.02, .06]
Intercept(LogV Urge)	$\omega_0$	.66	[.52, .81]
Intercept(LogV Dep)	$\omega_1$	-.01	[-.05, .03]
Var. (Alpha Urge)	$\tau_{00}$	.62	[.46, .88]
Var. (Alpha Dep)	$\tau_{11}$	.00	[.00,.01]
Var. (Phi1)	$\tau_{22}$	.02	[.01, .03]
Var. (Phi2)	$\tau_{33}$	.00	[.00,.01]
Var. (Phi3)	$\tau_{44}$	.00	[.00,.01]
Var. (Phi4)	$\tau_{55}$	.00	[.00,.01]
Var. (LogV Urge)	$\tau_{66}$	.51	[.38,.70]
Var. (LogV Dep)	$\tau_{77}$	.01	[.00,.02]



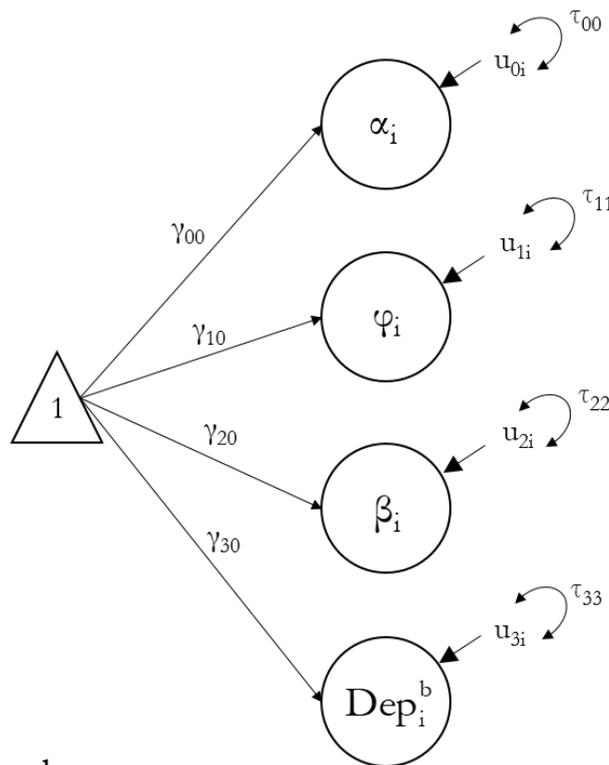
*Figure 1.* Comparison of developmental process data for 6 measurements occasions (left) and stable process data for 50 measurement occasions (right)



*Figure 2.* Path diagram of  $N=1$  AR(1) model with time-varying covariate based on Equation 2 with labels corresponding to *Mplus* code. The means and variances of the exogenous variables (Lage-1 Urge to Smoke and Depression) are not shown to focus on parameters of interest in the model.

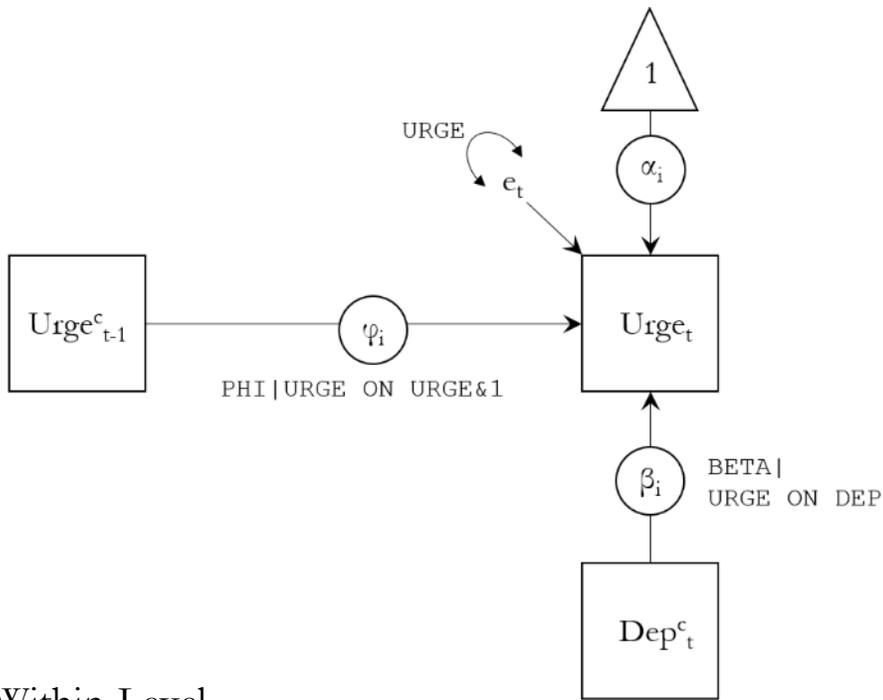


Within-Level

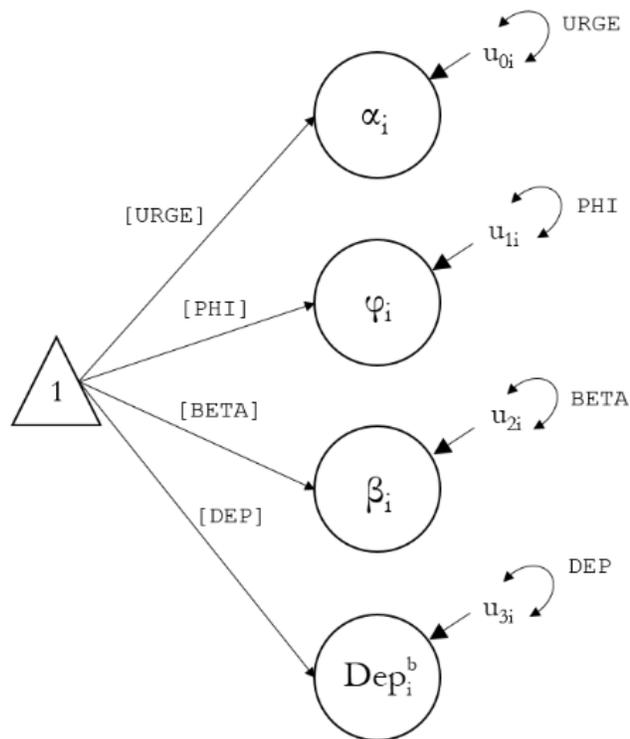


Between-Level

Figure 3. Path diagram of multilevel AR(1) model from Equation 4c. The means and variances of the exogenous variables (Lage-1 Urge to Smoke and Depression) are not shown to focus on parameters of interest in the model.

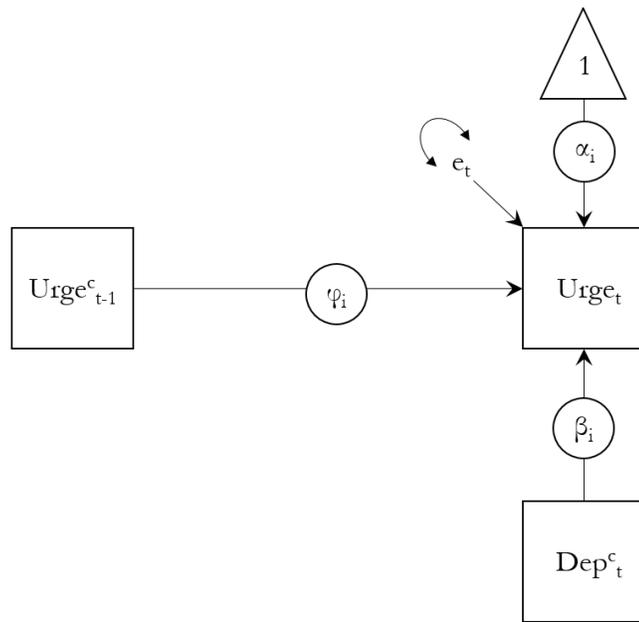


Within-Level

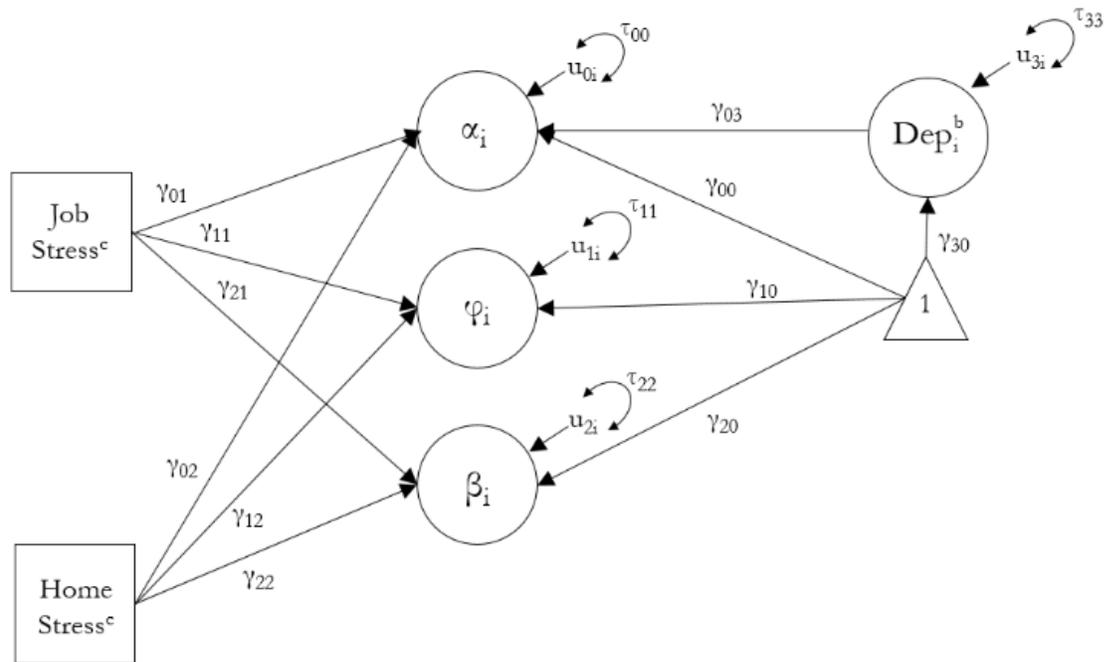


Between-Level

Figure 4. Multilevel AR(1) path diagram for the model in Equation 4c with labels corresponding to *Mplus* code. The means and variances of the exogenous variables (Lage-1 Urge to Smoke and Depression) are not shown to focus on parameters of interest in the model.

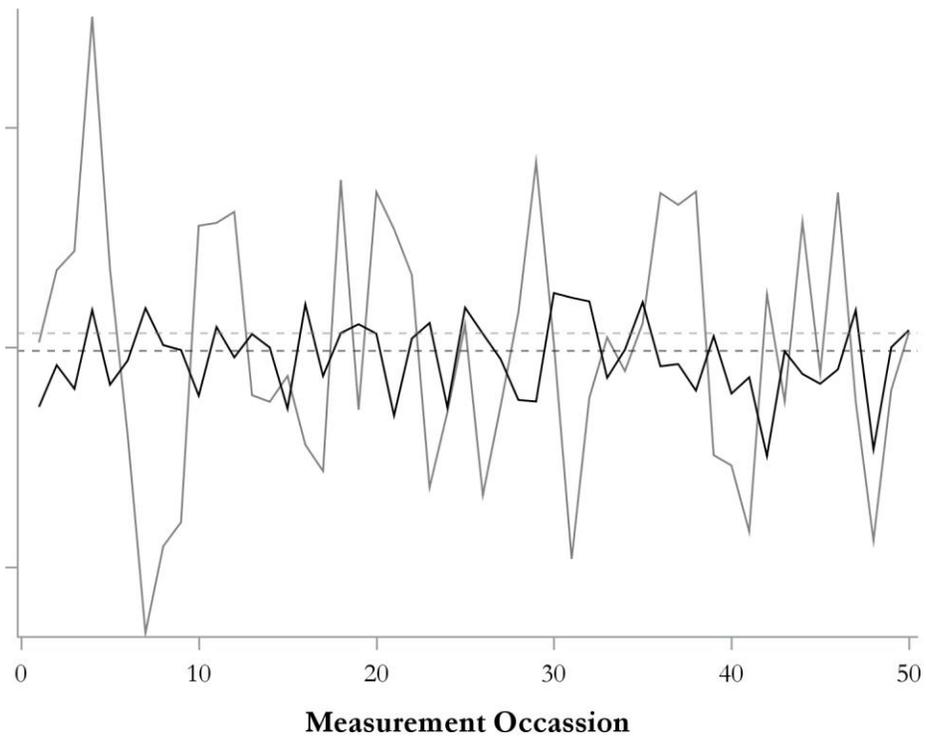


Within-Level



Between-Level

Figure 5. Path diagram for multilevel AR(1) model with time-invariant covariates based on Equation 5. The means and variances of the exogenous variables (Lage-1 Urge to Smoke, Depression, Home Stress, and Job Stress) are not shown to focus on parameters of interest in the model.



*Figure 6.* Comparison of trace plots for Person 5 (grey) and Person 96 (black) to highlight differences in variability across people when  $N > 1$ . Means for each person are shown as dashed lines

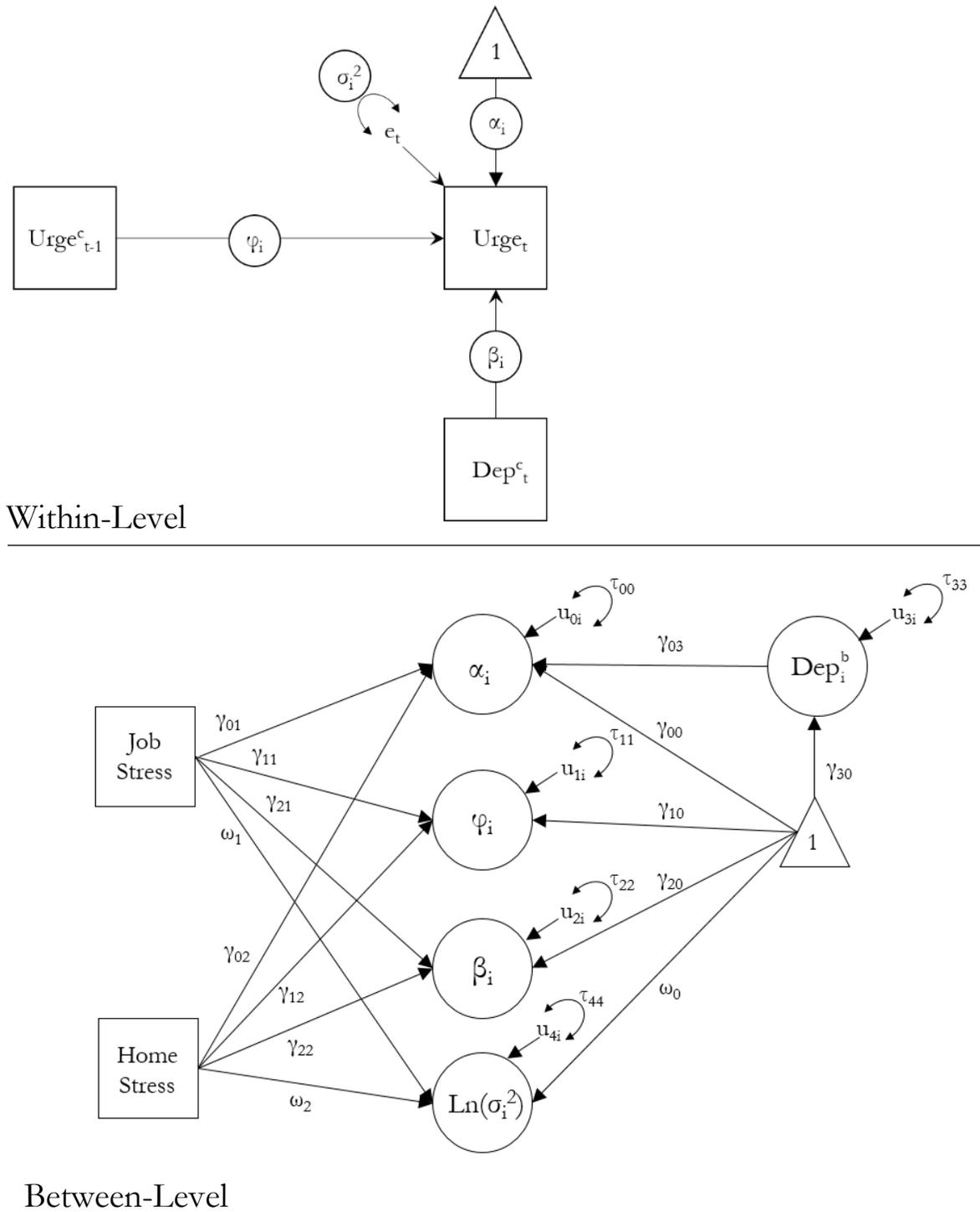
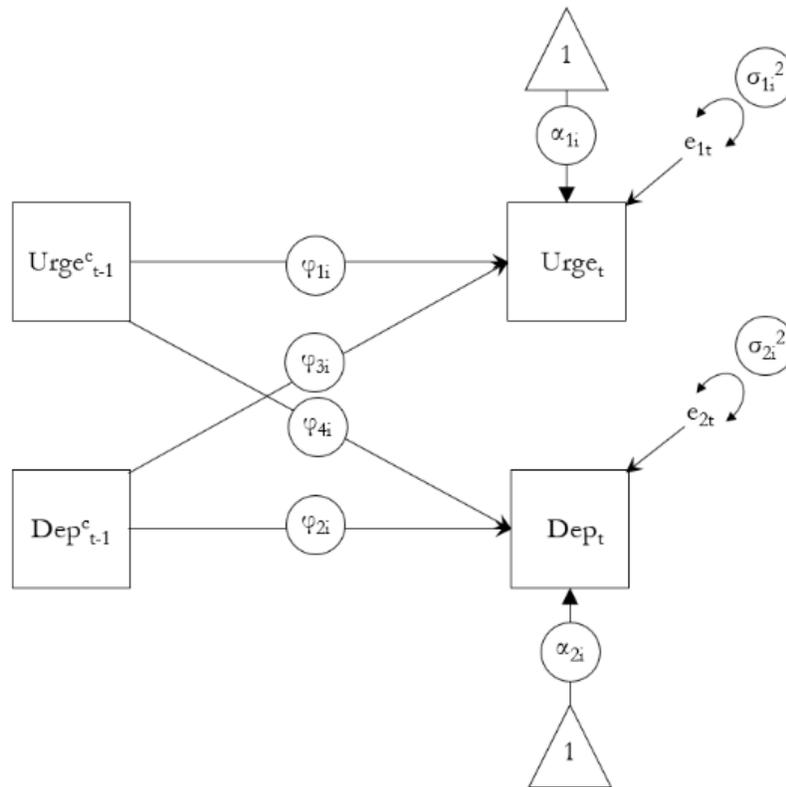
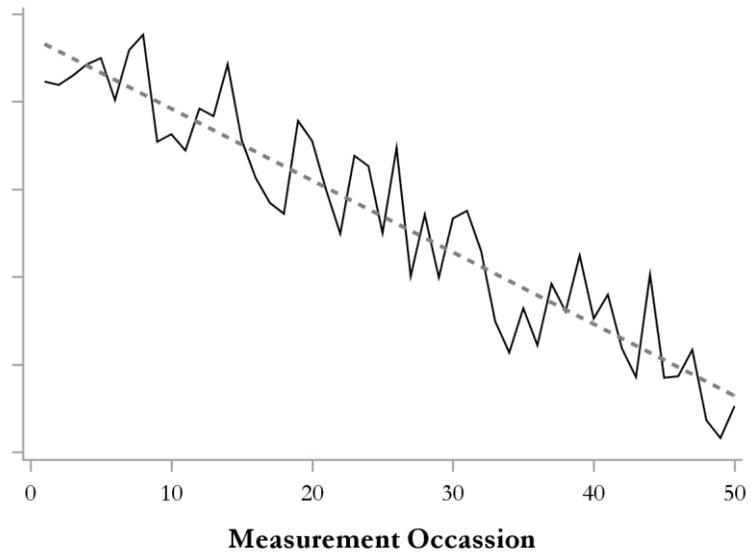


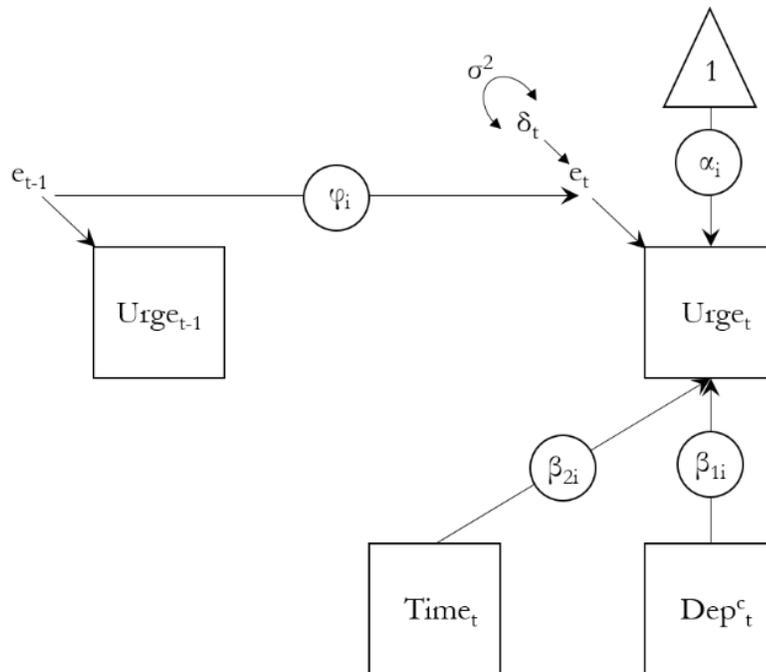
Figure 7. Path diagram for multilevel location-scale model based on Equations 5 and 6. The means and variances of the exogenous variables (Lage-1 Urge to Smoke, Depression, Home Stress, and Job Stress) are not shown to focus on parameters of interest in the model.



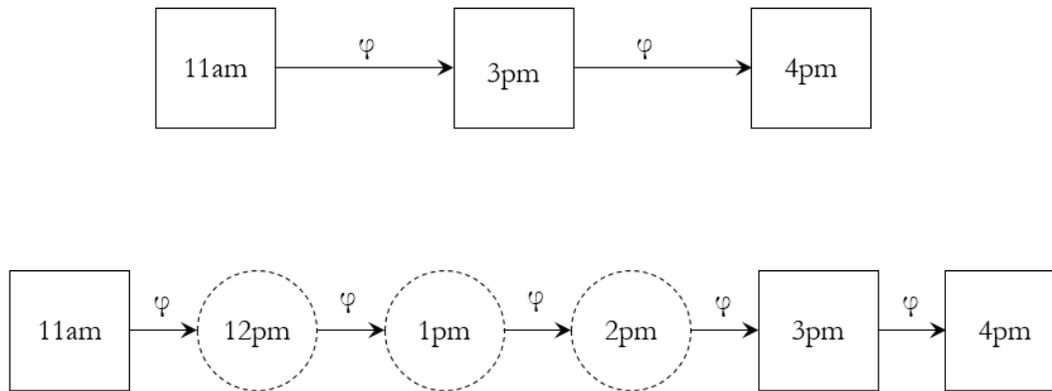
*Figure 8.* Within-Level path diagram for the multilevel VAR(1) model based on Equation 7. The means and variances of the exogenous variables (Lage-1 Urge to Smoke, Lag-1 Depression) are not shown to focus on parameters of interest in the model. Major changes occur in the Within-Level model, so we do not include the Between-Level model in order to focus on the relevant pedagogical information. The Between-Model also has no covariates and is comprised only of means and variances.



*Figure 9.* Trace plot for a single hypothetical individual with a negative linear trend superimposed with a dashed grey line.



*Figure 10.* Within-Level path diagram for RDSEM model with a linear trend for the model shown in Equation 10. The means and variances of the exogenous variables (Lage-1 Urge to Smoke, Depression, and Time) are not shown to focus on parameters of interest in the model.



*Figure 11.* Model specification if treating unequally spaced measurement occasions as if they are equally spaced (top) vs. using the TINTERVAL statement in *Mplus* to accommodate unequally spaced measurement occasions. In the bottom panel, the dashed circles are not observed in the data but are included as missing observations to facilitate the interpretation of the autoregressive term.