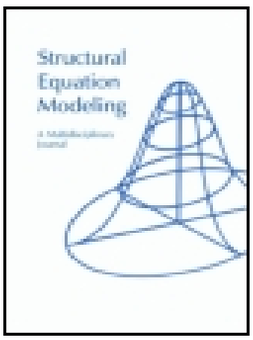

Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Two-Level Dynamic Structural Equation Models with Small Samples

Daniel McNeish

To cite this article: Daniel McNeish (2019): Two-Level Dynamic Structural Equation Models with Small Samples, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2019.1578657

To link to this article: https://doi.org/10.1080/10705511.2019.1578657

Structural Equation Modeling: A Multidisciplinary Journal, 00: 1–19, 2019

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2019.1578657

Two-Level Dynamic Structural Equation Models with Small Samples

Daniel McNeish
Arizona State University

Advances in data collection have made intensive longitudinal data easier to collect, unlocking potential for methodological innovations to model such data. Dynamic structural equation modeling (DSEM) is one such methodology but recent studies have suggested that its small N performance is poor. This is problematic because small N data are omnipresent in empirical applications due to logistical and financial concerns associated with gathering many measurements on many people. In this paper, we discuss how previous studies considering small samples have focused on Bayesian methods with diffuse priors. The small sample literature has shown that diffuse priors may cause problems because they become unintentionally informative. Instead, we outline how researchers can create weakly informative admissible-range-restricted priors, even in the absence of previous studies. A simulation study shows that metrics like relative bias and non-null detection rates with these admissible-range-restricted priors improve small N properties of DSEM compared to diffuse priors.

Keywords: Time-Series Analysis, DSEM, small sample, intensive longitudinal data, prior distribution

Dynamic Structural Equation Modeling (DSEM) is a recent methodological development that blends multilevel, structural equation, time-series, and time-varying effects modeling in one broad integrated framework (Asparouhov, Hamaker, & Muthén, 2018). DSEM has been shown to be particularly useful in the analysis of intensive longitudinal data (ILD), which has flourished in recent years as designs based on ecological momentary assessment (EMA), experience sampling methods (ESM), and ambulatory assessments increase in popularity (Bolger & Laurenceau, 2013; Hamaker & Wichers, 2017; Trull & Ebner-Priemer, 2014; Walls & Schafer, 2006). In these designs, the goal is generally to obtain a large number of measures, *T*, from the same individuals, *N*, over a relatively brief period of time (where *T* is usually greater than 10 but can exceed 100 in some studies). Such time-series designs have historically been popular in econometrics and engineering, but their potential in the behavioral sciences is just beginning to be fully realized as advances in data collection proliferate (Hamaker & Dolan, 2009). Models for ILD tend to differ from the traditional conceptualization of longitudinal data in behavioral sciences, which is closely associated with growth modeling. In *developmental process* data that are modeled with growth models, the common structure is a smaller number of measurement occasions taken over a relatively long period of time (e.g., several months or years). In *stable process* data such as ILD, the focus lies in modeling variability present in many measurement occasions taken in a short duration (e.g., days or weeks; Hamaker & Wichers, 2017). Though growth models for developmental processes commonly include parameters to capture within-person and between-person variance, the primary focus typically is on how scores change over time; the variability that is assessed is the variability in the growth curves (Asparouhov et al., 2018). In stable process data and time-series analysis, growth is secondary or a nuisance to account for, if it is even present at all (Kuljanin, Braun, & DeShon, 2011). That is, the outcome variable(s) of interest ebb and flow over time but the net change across the observation window may be null (Nesselroade, 1991).

Whether a developmental or stable process is of interest, in all longitudinal designs, small samples are prevalent

Correspondence should be addressed to Daniel McNeish, Department of Psychology, Arizona State University PO Box 871104, Tempe, AZ, 85287, USA. E-mail: dmcneish@asu.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hsem.

because of financial and logistical constraints present when following people over time (e.g., McNeish & Matta, 2018). Because of the intensive nature of the data collection with ILD, the risk of small samples similarly exists (Jongerling, Laurenceau, & Hamaker, 2015; Schultzberg & Muthén, 2018). For instance, an informal survey on ILD reported in Muthen (2017, Slide 27) found that 39% (23/59) of participants had ILD with $N < 100$ and 24% (14/59) with $N < 50$. Schultzberg and Muthén (2018) conducted a large-scale simulation to explore a small sample performance of nine different two-level DSEM models. From their results, they conclude that studies with large N and small T perform substantially better than models with small N and large T (p. 495). These results align with results from the multilevel modeling literature which suggests that the higher level sample size is much more influential than the lower level sample size (Maas & Hox, 2005; Scherbaum & Ferreter, 2009), though it does go against typical recommendations in single-level time-series analysis where large T (e.g., 50 or 100) are recommended (e.g., Chatfield, 2016). Specifically, the results of Schultzberg and Muthén (2018) find that random effect variance estimates are particularly biased for small N (less than 50 for the models used in their simulation) and the direction of the bias of these random effect variance estimates were consistently positive (i.e., the estimates are larger than the specified population values).

Models in these simulations were estimated with Bayesian Markov Chain Monte Carlo (MCMC) because DSEM tends to include several random effects to fully capture all aspects of within-person and between-person variance. With maximum likelihood estimation, models with many random effects can make estimation quite difficult or even impossible given the use of the joint distribution (Asparouhov et al., 2018; Muthén & Asparouhov, 2012). MCMC with Gibbs sampling, on the other hand, works with a series of conditional distributions rather than a single joint distribution, which can facilitate estimation of DSEM models, especially as model complexity increases.

A notable step of Bayesian estimation (MCMC or otherwise) is that researchers must specify prior distributions for each parameter (e.g., van de Schoot et al., 2014). Though these priors can be informative to incorporate findings from previous studies or judgments from experts, a more common approach is to use diffuse or non-informative priors so that the posteriors are primarily driven by the data (van De Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017). In user-friendly software with Bayesian modules like M*plus*, if researchers do not specify prior distributions, the software will provide diffuse priors by default (Asparouhov & Muthén, 2010). This approach is not inherently flawed and is quite reasonable considering the broad classes of models that can be handled by general structural equation modeling software. However, diffuse priors have been shown to have unintended consequences with smaller samples sizes (Depaoli & van de Schoot, 2017; McNeish, 2016a, 2017; Zondervan-Zwijnenburg, Peeters, Depaoli, & van de Schoot, 2017).

We delve into more detail shortly, but the basis of the issue is that Bayesian methods combine two sources of information to form posterior distributions: the data (via the likelihood) and prior beliefs (via prior distributions). Small sample data do not contribute much information to the likelihood (van de Schoot et al., 2014), so the likelihood is relatively less informative and carries relatively less weight within the posterior distribution compared to situations with larger samples. This means the prior distribution receives higher relative weight in the posterior distribution with smaller samples and therefore can become unintentionally informative (McNeish, 2017). The effect can be particularly pronounced at higher levels of the hierarchy in multilevel models (McNeish & Stapleton, 2016). As has been shown in previous studies, the effect of diffuse priors with small samples can lead to highly positively biased estimates of random effect variances (McNeish, 2016b; van De Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015; Zondervan-Zwijnenburg, Peeters et al., 2017), as was found in Schultzberg and Muthén (2018).

The main goal of the current paper is to show how researchers can improve the small sample size performance of DSEM models by proactively specifying the prior distributions themselves, especially for random effect variances. We begin with a brief overview of small sample issues concomitant with Bayesian estimation of latent variable models. We then discuss admissible-range-restricted priors as one possible strategy for forming prior distributions with smaller samples, even if one does not have prior literature or expert opinions on which to rely. We then perform a simulation study to show how even rudimentary methods of specifying priors and not relying on the software to provide priors with smaller samples can improve performance by reducing bias (particularly of the random effect variances), decreasing the variance in the posterior distributions (the Bayesian analog of efficiency) and increasing the ability of the model to detect non-null effects (the Bayesian analog of power).

OVERVIEW OF BAYESIAN METHODS WITH SMALL SAMPLES

In the traditional frequentist framework, probability is defined in terms of long-run frequencies (e.g., Spiegelhalter, Myles, Jones, & Abrams, 1999). Maximum likelihood estimation (a leading frequentist estimation method) bases its estimation on finding point estimates for each model parameter that maximize some likelihood function provided by the model and its underlying assumptions. In contrast, the Bayesian definition of probability represents a state of knowledge based on

updating prior beliefs with newly acquired knowledge (e.g., Gelman, Carlin, Stern, & Rubin, 2004). Bayesian estimation of statistical models is based on a posterior distribution (rather than a single point estimates) for each parameter, which updates previous beliefs about a parameter (captured by the prior distribution) with new information added from data (captured by the likelihood).

Before considering the contribution of the data, researchers specify the prior distribution for each parameter in the model of interest. These priors can be *informative* if researchers have a strong inclination regarding plausible values for each parameter. Or, priors can be *diffuse* such that the distribution covers a wide range of possible values if researchers do not have any prior beliefs (a prior distribution is placed on each parameter in the model so it is also possible for some priors to be informative while allowing others to be diffuse). When the data are subsequently modeled, the likelihood from the specified model is computed (this is the same likelihood used in maximum likelihood estimation and is similarly based on the model and any distributional assumptions). The likelihood contains information strictly from the data and is not affected by prior distributions. A weighted combination of the prior and the likelihood form the posterior distribution, which is conceptually similar to the sampling distribution in a frequentist framework.

Bayesian methods estimate distributions instead of point estimates, so assumptions concerning asymptotics with frequentist methods (e.g., central limit theorem, Fisher information) are absent with Bayesian MCMC. Instead, MCMC yields an empirical sampling distribution that is completely independent of assumptions about sample size being sufficiently large. This property leads to recommendations that Bayesian methods are preferable at smaller sample sizes (Hox, van de Schoot, & Matthijsse, 2012; Lee & Song, 2004; Muthén & Asparouhov, 2012). This freedom from asymptotic assumptions gives Bayesian methods the *potential* to be more appropriate and more trustworthy than frequentist methods with small samples (van de Schoot et la., 2014), but the benefit is not bestowed upon researchers simply from switching the estimator for their model (e.g., Depaoli & Clifton, 2015; McNeish, 2016a; van De Schoot et al., 2015).

Regardless of sample size, researchers often specify diffuse prior distributions to avoid unduly intervening in the analysis or to let the data (via the likelihood) be the driving force in formation of the posterior –van De Schoot et al. (2017) report that 73% of papers utilized prior distributions that were diffuse. Put another way, researchers commonly employ Bayesian methods for their *computational* advantages rather than their *philosophical* advantages. With moderate or large samples, diffuse prior distributions are innocuous and minimally influence the posterior distributions (Muthén & Asparouhov, 2012). This is due to the relative dominance of the likelihood in such cases: with larger samples, the likelihood contains much information by virtue of contributions from many observations. Even though the posterior distribution is a combination of the prior and the likelihood, the contribution is not necessarily equal. When the likelihood contains a large amount of information as with large samples, if the prior is diffuse, the likelihood essentially drowns out the prior and the posterior is based almost entirely on the likelihood (Browne & Draper, 2006).

With small samples, the likelihood contains much less information because fewer observations contribute to it. In terms of the posterior distribution, the likelihood has much less relative weight and the prior distribution plays a much larger relative role in the posterior distribution with smaller samples. Given the dominant preference in empirical studies to avoid informative priors, poor statistical properties can result (van De Schoot et al., 2015). Due to the increased relative importance of the prior with smaller sample sizes, *any* prior distribution becomes informative with small samples and influences the posterior distribution (McNeish, 2016a). That is, the uncertainty contained with the diffuse prior is propagated into the posterior distribution because the likelihood is non-descript. From a Bayesian perspective, the newly acquired information is used to update previous beliefs. However, if that newly acquired information is lacking, the updated beliefs will look very similar to the previous beliefs. However, when Bayes is evoked on computational merits rather than philosophical merits, the priors do not usually contain previous beliefs but instead are specified to mimic frequentist principles while circumventing challenges associated with frequentist estimation. Therefore, specifying priors in this way increases posterior distribution sampling variability, entertains implausible parameter value in the MCMC chain, and affects the ability to detect non-null effects (McNeish, 2017). The differential effect of diffuse priors for large samples and small samples is shown conceptually in Figure 1.

NARROWING THE SUPPORT OF THE PRIOR

Basing prior distributions on review studies, previous research, or expert opinions is the optimal approach for specifying informative prior distributions (van de Schoot et al., 2018; Veen, Stoel, Zondervan-Zwijnenburg, & van de Schoot, 2017; Zondervan-Zwijnenburg, van de Schoot-Hubeek, Lek, Hoijtink, & van de Schoot, 2017). However, in many contexts – especially for recently developed methodologies like DSEM – this type of information is not always readily available to researchers. In the absence of ideal information, another alternative for specifying priors with small samples is to choose a prior that restricts the support of the prior based on boundaries of the possible or plausible parameter space (e.g., Berger, 2006).

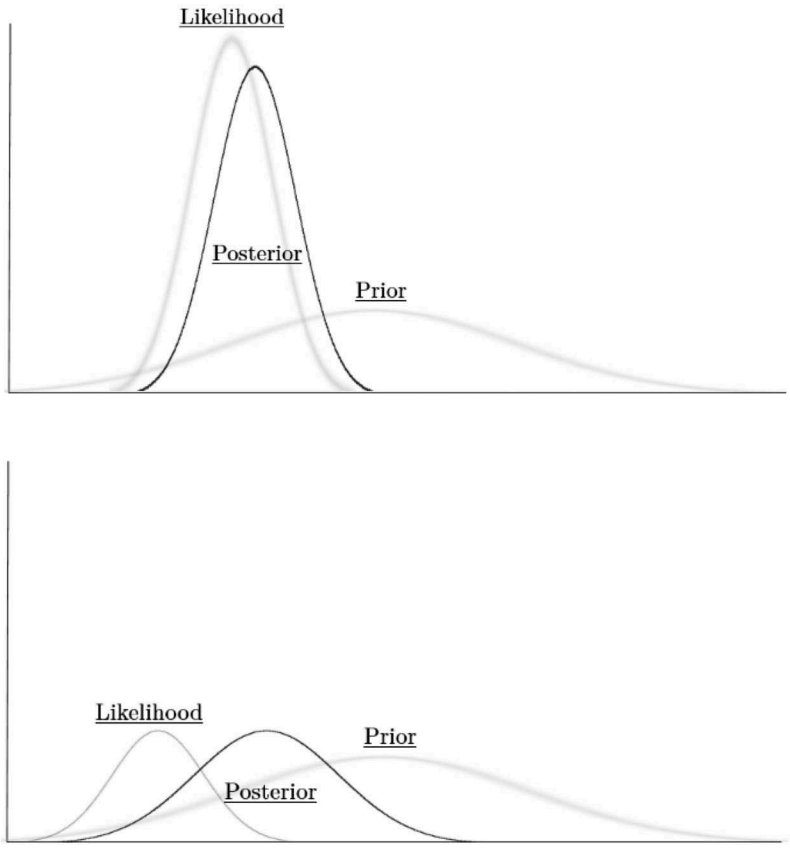

FIGURE 1 Conceptual comparison of the influence of diffuse priors for large samples (top panel) and small samples (bottom panel). With large samples, the likelihood dominates the prior so the posterior is strongly influenced by the likelihood. In small samples, the prior becomes unintentionally informative when the posterior is computed.

As an example in the context of DSEM and time-series analysis, assuming stationarity, the magnitude of the autoregressive inertia paths from the outcome at time $t - 1$ to the outcome at time t must fall between −1 and 1, but diffuse priors allow for random effect variances that could lead to subject-specific values that fall outside of this range (e.g., the M*plus* default prior for random effect variance has support on the interval $[0, \infty]$). Instead, the prior distribution of the random effect variance of this path can be manually restricted to a distribution where values fall between −1 and 1 with near certainty.

These so-called admissible-range-restricted priors can be applied without requiring any previous studies on the area of interest since they follow purely from the boundary of admissible or plausible parameter estimates. This same idea can be generalized to other parameters in the model that do not necessarily mandate that values must be within a particular range but that have known constraints from other sources. Variance components such as the intercept variance or the residual variance, for instance, cannot exceed the total variance of the outcome. Therefore, a similar principle could be applied such that the support of the prior for these parameters could be restricted based on the descriptive variance of the outcome. From inspecting descriptive statistics of other variables of interest, researchers can begin to see some intuitive boundaries (or at least an idea of what values would be vaguely realistic) for variance terms in the model (which are particularly affected by diffuse priors). In the simulation study in the next section, we will outline how these admissible-range-restricted priors can be created and how the small sample performance of models that use them can often be superior to small sample performance from models that use default diffuse priors.

SIMULATION DESIGN

Data generation model

The data generation model is based on the model featured in example 9.31 on page 358 of the M*plus* Version 8 User's Guide (but it is not identical). The model is a two-level time series with a univariate first-order autoregressive model for the continuous outcome. The model includes one time-varying

covariate that is also measured at each time-point and two continuous time-invariant covariates. All covariates are generated from a standard normal distribution. Four random effects are included in the model: random intercepts, random autoregressive inertia slopes, random time-varying covariate slopes, and random residual variances. All generated individuals have the same number of fixed time-points and with no missing data (i.e., the data are balanced and time-structured). Figure 2 shows the path diagram of the conceptual model. Data generation and simulation output files are available on the Open Science Framework. Data were generated and fit using M*plus* Version 8 (Muthén & Muthén, 1998–2018).

As a model equation, the complete data generation model at the Within-Level is

$$y_{ti} = \alpha_i + \varphi_i y^{(c)}_{t-1,i} + \beta_i x_{ti} + e_{ti} \quad (1)$$

where y_{ti} is the outcome variable at time t (for $t = 1, \ldots, T$) for person i (for $i = 1, \ldots N$), α_i is the intercept for the ith person, φ_i is the autoregressive inertia slope for the ith person, $y^{(c)}_{i,t-1}$ is outcome at time $t - 1$ which is latent-centered as $y^{(c)}_{i,t-1} = y_{i,t-1} - \alpha_i$, x_{ti} is a time-varying covariate, β_i is the effect of the time-varying covariate for person i, and $e_{ti} \sim N(0, \sigma_i^2)$. The Between-Level model equations with population values (which are taken from example 9.31 in the M*plus* Version 8 User's Guide) for the effects are

$$\begin{aligned} \alpha_i &= (.50 \times W_{1i}) + (.30 \times W_{2i}) + u_{0i} \\ \varphi_i &= .20 + (.10 \times W_{1i}) + (.05 \times W_{2i}) + u_{1i} \\ \beta_i &= .70 + (.30 \times W_{1i}) + (.20 \times W_{2i}) + u_{2i} \\ \ln(\sigma_i^2) &= (.30 \times W_{1i}) + (.10 \times W_{2i}) + u_{3i} \end{aligned} \quad (2)$$

where W_{1i} and W_{2i} are time-invariant covariates and

$$\mathbf{u}_i \sim N\left(\mathbf{0}, \begin{bmatrix} .30 & & & \\ 0 & .01 & & \\ 0 & 0 & .50 & \\ 0 & 0 & 0 & .10 \end{bmatrix}\right) \quad (3)$$

These population values lead to R^2 values of 40% for y_{ti}, 25% for α_i, 50% for φ_i, 10% for β_i, and 30% for $\ln(\sigma_i^2)$.

Simulation conditions

Our sample size conditions will generally follow those used by Schultzberg and Muthén (2018) and Jongerling et al. (2015) though will we focus more heavily on the Level-2 sample sizes (N) and less on the difference in the number of time-points (T). Our Level-2 sample size conditions for will be N= 10, 20, 30, 40, and 50. The largest sample size condition in our simulation was informed by results in Schultzberg and Muthén (2018) and the general multilevel modeling small sample literature where small sample bias in the variance components estimates begins to dissipate around N =50 (McNeish, 2016b). Our time-point conditions will be T = 25, 50, and 100 which represents the moderate number of time-point conditions from Schultzberg and Muthén (2018). These conditions are fully crossed to create 30 cells of the simulation design (5 sample sizes conditions × 3 time-point conditions × 2 prior distribution conditions) with 500 replications to be performed in each cell.

All models will be fit with M*plus* version 8 using MCMC with two chains. We set the minimum number of iterations to 1000 and used the Potential Scale Reduction criteria (Gelman & Rubin, 1992) to terminate chains thereafter using the M*plus* default value of 1.05.[1] Posterior distributions are summarized using the median and chains were not thinned. Each generated dataset is modeled with two different sets of prior distributions. The first uses the M*plus* defaults that were used in Schultzberg and Muthén (2018). The default priors are $\Gamma^{-1}(-1, 0)$ for random effect variances and $N(0, \infty)$ for intercepts and fixed slopes (Asparouhov & Muthén, 2010).[2] The $\Gamma^{-1}(-1, 0)$ distribution is improper and equivalent to uniform on $[0, \infty)$(Asparouhov & Muthén, 2010, p. 23). For the second condition, we used admissible-range-restricted prior distributions for the random effect variances but retain the M*plus* defaults for intercepts and fixed slopes. The process of creating these admissible-range-restricted priors for the random effect variances is described in the detail in the next section.

Creating admissible-range-restricted priors

Using only observed descriptive information from the data and assuming no other prior information from previous studies, weakly informative priors can be created based to reduce the support of the prior distributions. For the remainder of this section, we discuss the process for arriving at these types of priors in general and the process for arriving at the prior distributions for the simulation model specifically. Visual representations of the ultimate prior distributions for the random effect variances (intercept, time-varying covariate, autoregressive parameter, log-residual) are shown in Figure 3.

[1] PSR can sometimes stop the algorithm prematurely, so we inspected trace plots for some of the replications to ensure that the number of replications was sufficient. Also note that 1000 was the minimum number of iterations, the number of iterations typically exceeded 1000 as necessary to satisfy the PSR criteria.

[2] There are several different parameterizations of the inverse gamma distribution. When used in this paper, to facilitate implementation in M*plus*, we follow the definition used in their documentation such that the probability density function is $f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{\left(-\beta/x\right)}$ where Γ is the gamma function, α is the shape parameter, β is the scale parameter, and x is the value at which the function is being evaluated for $x > 0$ (Asparouhov & Muthén, 2010, p. 52).

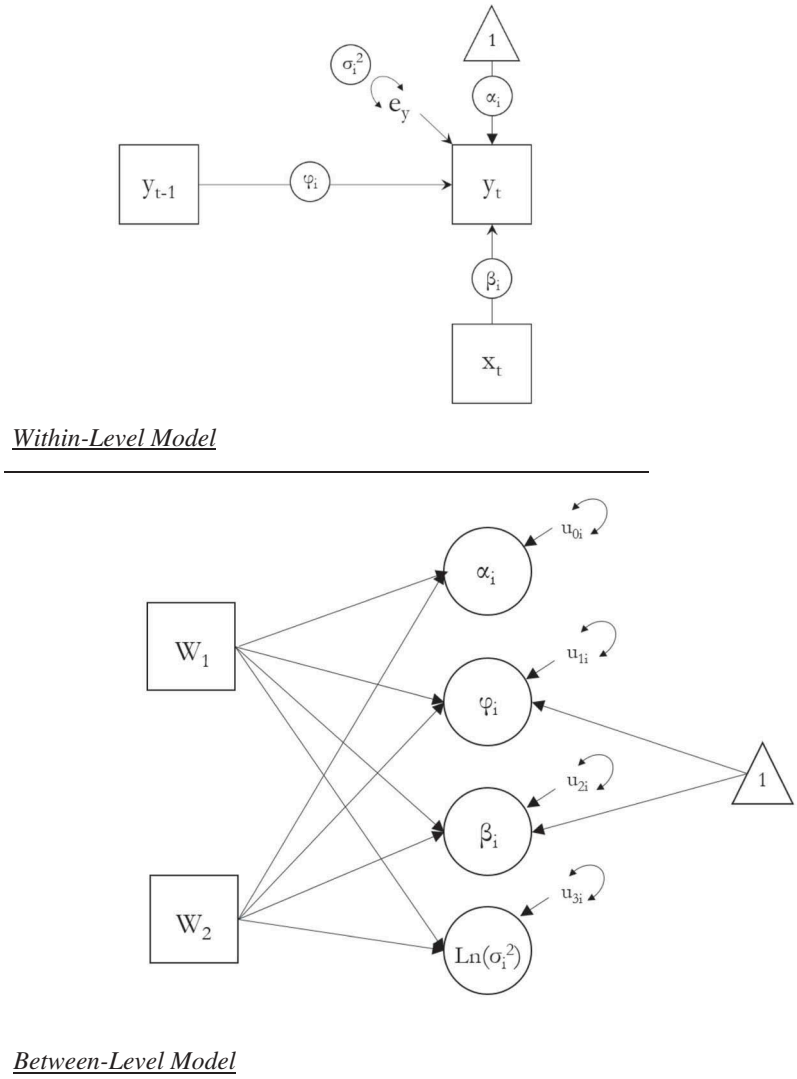

FIGURE 2 Conceptual path diagram for data generation model. Paths with black dots indicate the path has random effects that vary by person.

General strategy

For random effect variances, the inverse gamma distribution is most commonly used when separately modeling each variance (we consider the case of modeling all variances at once later on). This distribution is commonly used because it does not allow negative values, which corresponds to the definition of variance. The inverse gamma distribution takes two parameters, a shape parameter (labeled α) and a scale parameter (labeled β). These parameters as not as intuitive as the mean and variance parameters of a normal distribution. However, broadly speaking, the shape parameter controls where the density of the distribution lies. The larger the shape parameter, the more concentrated the density is towards 0. The scale parameter controls the dispersion of the distribution with larger values increasing the spread. With the inverse gamma distribution, the inherent skew of the distribution (as seen in Figure 3) typically makes the mode more informative than the mean when describing central tendency. The mode is calculated by $\beta/(\alpha+1)$ and is defined for any values of shape and scale (unlike the mean, which is only defined for shape values greater than 1).

The simple form of the mode and the fact that it contains both parameters make it a useful starting point for creating admissible-range-restricted priors.[3] Using the scale of the

[3] To clarify the terminology, there is a difference between "restricted" and "bounded" priors. Uniform priors bound the support of the prior by making the probability of particular values exactly 0; if the prior probability is 0, the posterior probability of that value will necessarily be 0 as well. This is opposed to restricting the support as we describe with inverse gamma. By "restrict", we mean to heavy concentrate the prior probability to a specific area, but note that this does not make the prior probability of values outside the restricted range to be exactly 0. In this way, researchers make assign arbitrary low probabilities for unlikely values without necessarily excluding them.

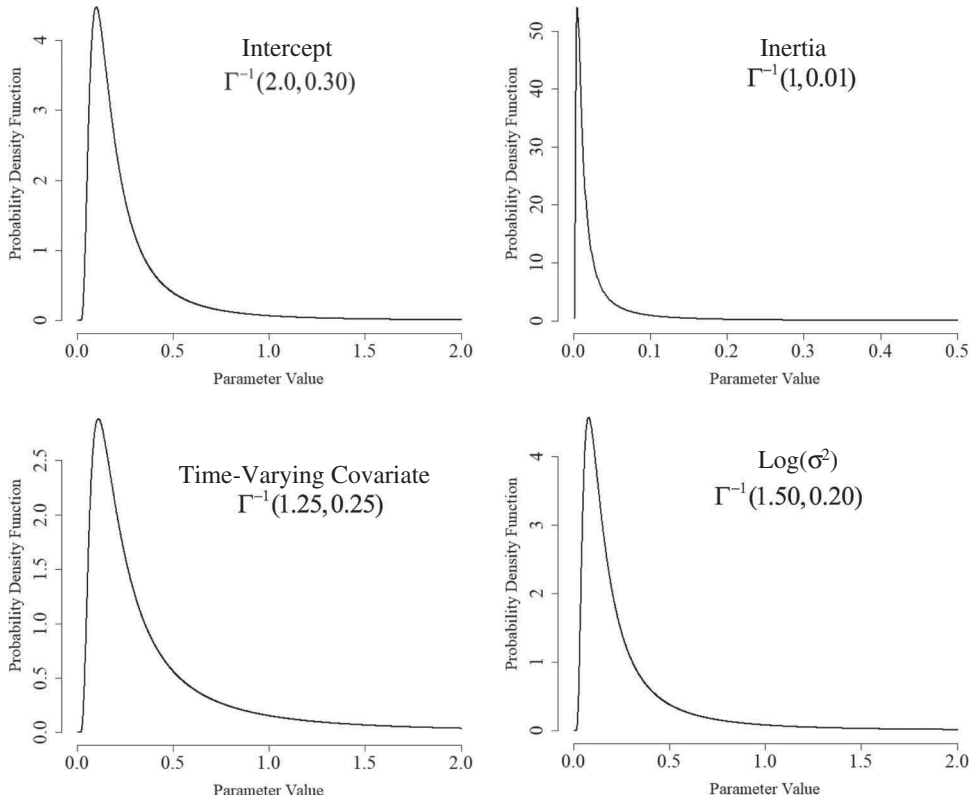

FIGURE 3 Visual representation of prior distributions for random effect variances.

variables or descriptive statistics, researchers first select the mode of the prior – the value they believe is reasonable for the parameter of interest. If no value immediately comes to mind, a value slightly above 0 may be a good starting point. There are multiple invers gamma distributions that produce that mode. Therefore, researchers can inspect the properties of several of these possible choices to determine which may be best for the model. For instance, shape parameters from .50 to 4.00 in .25 increments could be selected. To keep the mode constant, selecting a shape parameter means the scale parameter must be a particular value, specifically $\beta = Mode(\alpha + 1)$.

To help select an appropriate distribution, the 95% highest density interval (HDI) is helpful. The 95% HDI is similar to a 95% credible interval. The difference is that the 95% credible interval bounds are taken from the 2.5th percentile and 97.5th percentile of the posterior whereas the HDI uses bounds such that every value within the interval has a higher probability density than all the values outside the interval. The major difference between the HDI and credible interval is that the HDI does not use equal tails, which can be more informative for non-symmetric distributions like inverse gamma. The HDInterval R package (Meredith & Kruschke, 2017) is a quick way to assess the HDI of a particular inverse gamma distribution.[4] From the HDI, the goal is to choose a distribution whose most plausible values span the possible values of the parameter of interest (as informed by plausible values or descriptive statistics). This process is described using the simulation model to give readers a concrete example of how this process would be carried out with data.

Application to the simulation model

This section demonstrates the general process described in the above section to the four random effect variances in the simulation model. In DSEM generally, the intercept variance and the inertia variance are the most straightforward parameters for which to determine the admissible values whereas admissible values for other parameters involves more subjectivity.

Intercept variance

Admissible values for the intercept variance are straightforward to determine with descriptive statistics from the data. Using one replication of generated data from our simulation, the total variance in the outcome was 3.25 with an unconditional intraclass correlation of 0.20. The intercept variance in a multilevel model like DSEM cannot exceed the total variance (3.25) and is likely around 0.65 (unconditional intraclass correlation × total variance = 0.20

[4] R code for obtaining the HDI is included in the appendix.

× 3.25 = 0.65) or below because the model has time-invariant covariates to explain between-person variance (i.e., the random intercept is capturing between-person variance that is not explained by the time-invariant covariates). Given this information, the prior distribution for the intercept variance can be specified so that its support extends only to values that would be plausible, given the characteristics of the data (e.g., the intercept variance is probably somewhere between 0 and a little less than 1).

To be conservative with the mode of the intended prior, we chose a value of .10. A value near 0.65 might be more likely; however, the covariates might explain some of the this variability, so we want to ensure that a value of 0 remains realistic if there is little difference in the intercepts, after accounting for covariates. We will inspect the inverse gamma distributions with a mode of .10 and shape parameters from .50 to 4.00 in .25 increments. We then solve for the scale parameter that is necessary to retain the desired mode and calculate the HDI in R. Table 1 shows the results of this analysis.

From here, selection of the appropriate distribution is going to involve some subjectivity, depending on how wide or narrow researchers want their prior to be. From Table 1, shape parameters between 1.50 and 3.00 appear to give reasonable HDI intervals whose upper bound is near the value we are expecting (about 0.65) based on the descriptive statistics from the outcome. We will proceed with the $\Gamma^{-1}(2, 0.30)$distribution because its HDI of [.03, .85] appears to narrow plausible values without being overly narrow or generous.

TABLE 1
95% HDIs for Several Hypothetical Inverse Gamma Distributions with a Mode of .10

Mode	*Shape*	*Scale*	*HDI Lower*	*HDI Upper*
.10	.25	.13	<.01	30816.00
.10	.50	.15	.01	76.30
.10	.75	.18	.02	10.83
.10	1.00	.20	.02	3.90
.10	1.25	.23	.02	2.18
.10	1.50	.25	.02	1.42
.10	1.75	.28	.03	1.08
.10	2.00	.30	.03	.85
.10	2.25	.33	.03	.72
.10	2.50	.35	.03	.61
.10	2.75	.38	.03	.55
.10	3.00	.40	.04	.49
.10	3.25	.43	.04	.46
.10	3.50	.45	.04	.42
.10	3.75	.48	.04	.40
.10	4.00	.50	.04	.37

Note: The mode is calculated by $\beta/(\alpha + 1)$, this table keeps the mode constant at .10 and selects Shape parameters in .25 increments, which means that the Scale parameter must take a specific value. The 95% HDIs are calculated with the HDInterval R package.

Inertia slope variance

The inertia slope of the outcome (y at time t) autoregressed on the outcome at the immediately preceding time (y at time $t - 1$) cannot exceed 1, provided that the assumption of stationarity has been met. Therefore, the goal with setting the prior distribution for this parameter is to select values that are narrow enough to exclude out-of-bounds values without being overly restrictive. This involves a little reverse engineering. The goal is to select values such that the entire person-specific distribution of random effects could not exceed a magnitude of 1 in either direction. Based on a normal distribution of random effects and a null intercept, this would mean that a variance of between .20 and .25 would satisfy this requirement. That is, assuming normality, a variance of .20 would mean the 95% for the random effects would be $[-2\sqrt{.20}, 2\sqrt{.20}] = [-0.90, 0.90]$. Though the prior could be specified to span the whole [−1,1] spectrum, choosing limits that are slightly below accommodates a fixed slope (around which the random effects are centered) that will unlikely be exactly equal to zero. Otherwise, to the extent that the fixed effect is non-null, the prior would allow for out-of-bounds values.

Though we will not go through the process as thoroughly as we did with the intercept variance, we selected a mode of .005 (due to the very small scale of this parameter). When testing different possible inverse gamma distribution that yields this mode, a $\Gamma^{-1}(1, .01)$ provided a 95% HDI of [.001, .195], which closely aligned with the intention to choose the prior that allowed the variance to be between 0 and .20.

Time-varying covariate slope variance

For the time-varying covariate slope variance, the descriptive statistics alone are not sufficient to restrict the range to admissible values. Still, knowing the variances of the covariate and the outcome can help to restrict the range of the prior to some degree based on plausible values (the time-varying covariate was generated from standard normal distributions, so the variance is equal to 1). Given the variances of the outcome (3.25) and covariate (1), it would seem unlikely that the magnitude of a person-specific slope would exceed 3. If we similarly reverse engineer which variance we would need so that most normally distributed person-specific slopes would not have a magnitude greater than 3 if the null effect were null, this would equate to a variance of about 2.25. That is, 95% of the person-specific slopes (assuming a null fixed effect) would be $[-2\sqrt{2.25}, 2\sqrt{2.25}] = [-3.00, 3.00]$

Similar to the intercept variance, we will presume a mode of .10 for the prior because we do not have strong prior information for what the value should be. Using the same prior mode also allows us to reuse Table 1 to aid the didactic intent of this paper. Looking at Table 1, a shape parameter of

1.25 appears to give a 95% HDI that most closely aligns with our goal for this parameter. With a mode of .10, the 95% HDI is a little short of the value for which we are aiming (2.18 vs 2.25). If we increase the scale parameter a little to 0.25 instead of 0.23 so that the prior is $\Gamma^{-1}(1.25, .25)$, the mode increases slightly to 0.11 but also extends to the 95% HDI to [.02, 2.37], which covers of definition of plausible values for this parameter. If the variance were 2.37, assuming a normal distribution of random effects and a null fixed slope, 95% of the person-specifics slopes would be $[-2\sqrt{2.37}, 2\sqrt{2.37}] = [-3.08, 3.08]$. Of course, the fixed slope for the time-varying covariate is unlikely to be exactly null, so the range would shift depending on the fixed slope estimate. Thus, we continue with $\Gamma^{-1}(1.25, .25)$ as the prior for the time-varying covariate slope.

Log-residual random effect variance

The prior for the random effect variance of the log-residual is also difficult to definitively specify given only descriptive statistics, so we will similarly try to restrict the prior based upon plausible value restrictions (as with the time-varying covariate slope variance) instead of descriptive statistics (as with the intercept and inertia variances). In the unconditional model, the intraclass correlation suggests that the average residual variance would be 2.60 (0.80 × 3.25) or lower after considering possible within-person covariates and inertia slopes. To restrict the range of the prior, we place an upper bound on the plausible value of a person-specific residual variance of 20. The log-residual random effect variance possesses an additional challenge in that the values are the log scale, rather than the raw metric of the outcome. On the log scale, the unconditional residual variance of 2.60 would correspond to about 0.96 and 20 corresponds to a value of about 3.0. The goal would be to specify a prior such that the standard deviation of the normally distributed random effects is about 1.0. That is, if we are expecting the fixed effect for the log-residual variance to be .96 (from the descriptive statistics), a standard deviation of 1.0 would make the upper bound of the 95% interval close to 3.0, which when converted back to the original metric would equal 20 as we specified as a plausible upper bound. Put another way, we want to select an inverse gamma distribution whose HDI upper bound is near 1 because larger values will lead to implausibly large person-specific values.

Again, seeing as we do not have strong prior information for where the prior should be centered, we will use .10 as with other parameters, which also enables us to re-use Table 1. In Table 1, it appears that shape parameters between 1.50 and 2.00 yield 95% HDI upper bounds near 1. We will choose the shape parameter of 1.50 in order to keep more prior density away from 0. To more reasonably narrow the support of the prior, we will change the scale parameter to .20 (because smaller scale parameters contract the distribution). Using this $\Gamma^{-1}(1.50, .20)$ shrinks the mode slightly to .08 and yields a 95% HDI of [.02, 1.14]. Though our preference was to use the smallest shape parameter to avoid concentrating the density near 0, the ultimate choice is admittedly subjective and there are alternative reasonable choices.

Note on covariates and explained variance

Note that this method is not narrowing the support of the prior to its absolute minimum. For instance, in the model we are fitting here, there are time-invariant covariates that will likely explain some of the variance in the between-level model. That is, the intercept variance is a residual variance that encapsulates the variance not attributable to the covariates. To the extent that the time-invariant covariates explain variability, the prior could be further reduced. Irrespective of how much variance is ultimately explained by the covariates, the values we specify here remain plausible and remain more informative than those of default diffuse priors. Regardless of the presence or absence of covariates, the same general guidelines can be followed. Do note that if covariates are expected to explain a large amount of variance, priors can be chosen more aggressively. As we discuss later, a more aggressive approach may be required for more complex models and/or smaller sample sizes.

Simulation outcomes

Four metrics will be followed and reported for the simulation study. The first is the relative bias of estimated parameters. Relative bias is calculated by $R^{-1}\sum_{r=1}^{R}\left(\frac{\hat{\theta}_r}{\theta}\right)$ for R the number of replications (500), r an index for replication number, $\hat{\theta}_r$ the estimated parameter value in replication r, and θ the population value for the parameter as specified in the simulation. This metric is the mean relative bias across all 500 replications in a particular condition. The metric will equal 1 if bias is absent from the estimates (the estimated value is equal to the population value). Values between .90 and 1.10 will be considered negligibly biased and acceptable based on suggestions provided in Flora and Curan (2004). Values below .90 are considered non-negligibly biased downward and values above 1.10 are considered non-negligibly biased upward.

The second metric is the coverage of the 95% credible interval. For each replication for each parameter, a 95% credible interval is estimated. The coverage metric records the proportion of these intervals across all 500 replications in which the population value appears within the interval. This is a measure of how well the variability of the estimates is estimated. The ideal value for coverage is 95%,

TABLE 2
Relative Bias for T = 50 across Sample Size and Prior Distribution Conditions

		N= 10		*N= 20*		*N= 30*		*N= 40*		*N= 50*	
Effect	*Population Value*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*
φ_i on W_{1i}	.10	.99	.99	.98	.98	.99	.98	.99	.99	1.00	.99
φ_i on W_{2i}	.05	1.07	1.07	.98	.98	.98	.99	.97	1.03	.98	.98
β_i on W_{1i}	.30	1.03	1.03	1.01	1.01	1.00	1.00	1.01	.99	1.00	1.00
β_i onW_{2i}	.40	1.02	1.02	.98	.98	.98	.98	.99	1.02	.99	.99
$\ln(\sigma_i^2)$ on W_{1i}	.30	1.04	1.04	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02
$\ln(\sigma_i^2)$ on W_{2i}	.10	1.03	1.02	1.03	1.04	1.03	1.03	1.03	1.02	1.05	1.04
α_i on W_{1i}	.50	1.02	1.02	1.00	1.01	1.00	1.00	1.00	.99	1.00	1.00
α_i on W_{2i}	.30	1.00	1.00	1.01	1.01	1.00	1.00	.99	1.00	.99	.99
Int(β_i)	.70	.98	.98	.98	.98	.99	.99	.99	.99	.99	.99
Int(φ_i)	.20	1.00	1.00	1.01	1.00	1.01	1.00	1.01	.99	1.00	1.00
Var (α_i)	.30	**.82**	**1.67**	.91	**1.22**	.93	**1.12**	.95	1.10	.96	1.07
Var (β_i)	.50	.92	**1.70**	.99	**1.25**	.97	**1.13**	.98	1.09	.98	1.07
Var (φ_i)	.01	**1.11**	**2.47**	1.04	**1.45**	1.01	**1.26**	.99	**1.14**	.99	**1.13**
Var [$\ln(\sigma_i^2)$]	.10	**1.23**	**1.87**	1.10	**1.24**	1.07	**1.14**	1.05	**1.14**	1.04	1.08

Note: ARR = Admissible-Range-Restricted, DD = Default Diffuse. Bold entries indicate values outside the [.90,1.10] negligible range.

and Bradley (1978) suggests that values between 92.5% and 97.5% are acceptable.

The third metric is the average posterior distribution standard deviation. The posterior distribution standard deviation measures how much uncertainty is contained within the posterior distribution and is the Bayesian analog of the standard error in the frequentist framework. There is no cut-off where values that are considered acceptable or unacceptable; however, if the coverage rates of the 95% interval and relative bias are acceptable, then small posterior distribution standard deviations are more desirable because they will decrease uncertainty and increase the probability of finding non-null effects when they are present.

The fourth metric is the non-null detection rate, which is the Bayesian analog to power in the frequentist framework. The non-null detection rate is computed by the proportion of replications in which an effect with a non-null population value is detected as being non-null as adjudicated by 0 falling outside the 95% credible interval. There is no set cut-off for acceptable non-null detection rates, but similar to power, higher values are better.

SIMULATION RESULTS

Relative bias

Table 2 shows the relative bias of each model parameter across sample size for each prior distribution condition when $T = 50$. The pattern was similar across conditions of T, and the number of time-points was not a primary interest, as this has been covered in Schultzberg and Muthén (2018), so we only show the $T = 50$ condition to keep the results and the interpretation thereof to be as concise as possible. Output files from other conditions can be found in the supplemental materials should they be of interest.

As anticipated, the relative bias of the fixed coefficients was negligibly biased across conditions. We, therefore, focus on the relative bias of the random effect variance estimates in this section, which tends to be the most problematic with smaller samples sizes and diffuse prior distributions. Figure 4 shows relative bias plots for each of the four random effect variance estimates across sample size conditions in the T= 50 condition. The Flora and Curran (2004) 0.90 and 1.10 thresholds are superimposed as dashed line lines on each plot.

As was found in Schultzberg and Muthén (2018), random effect variances with diffuse priors and small samples exhibit large positive relative bias, especially with $N < 30$. At $N = 40$, the inertia slope variance and variance of the residual still maintained relative biases exceeding 1.10. On the other hand, the admissible-range-restricted priors led to much smaller biases for all four parameters, across conditions of N. Each of the random effect variances is negligibly biased at $N = 20$ or larger with admissible-range-restricted priors. Some non-negligible bias does remain at $N = 10$, though the magnitude is much reduced compared to results from diffuse priors (e.g., the time-varying covariate slope relative bias is 2.47 with diffuse priors vs. 1.11 with admissible-range-restricted priors).

Coverage and posterior standard deviations

Table 3 shows the 95% credible interval coverage with values outside Bradley's range appearing in bold. Coverage in the default diffuse condition is too large in the N = 10 condition, but coverage is within Bradley's range for all other sample sizes in this condition. The

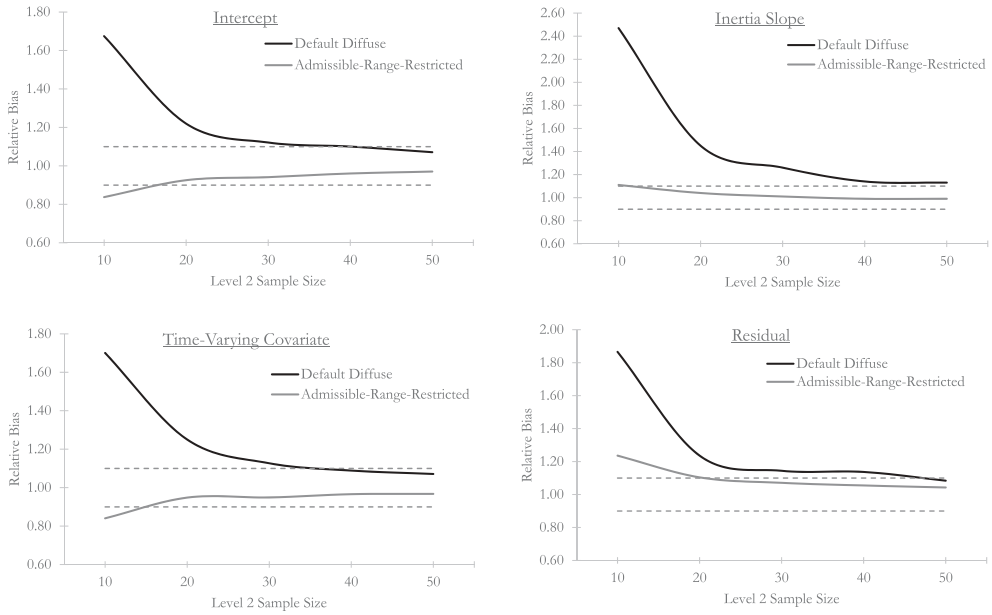

FIGURE 4 Relative bias plots for random effect variances of the intercept (upper left), inertia path (upper right), time-varying covariate (lower left), and residual (lower right).

TABLE 3
Coverage Rates for T = 50 across Sample Size and Prior Distribution Conditions

	N= 10		*N= 20*		*N= 30*		*N= 40*		*N= 50*	
Effect	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*
φ_i on W_{1i}	95.8	**98.4**	94.8	96.6	95.0	96.2	93.6	97.0	94.6	94.6
φ_i on W_{2i}	95.6	**99.2**	96.0	96.2	95.8	96.6	94.2	95.8	95.0	95.0
β_i on W_{1i}	94.0	**98.4**	93.6	96.2	94.8	96.0	96.6	95.4	95.2	95.8
β_i on W_{2i}	**92.0**	**97.6**	93.2	95.0	92.8	94.2	93.8	96.4	93.6	95.2
$\ln(\sigma_i^2)$ on W_{1i}	95.8	**98.6**	95.4	95.8	95.4	96.0	95.2	96.2	96.0	96.0
$\ln(\sigma_i^2)$ on W_{2i}	96.8	**97.6**	**97.6**	97.4	95.8	96.0	95.0	95.2	95.4	95.8
α_i on W_{1i}	93.0	**98.2**	95.0	97.0	94.6	96.2	93.6	97.0	94.4	95.4
α_i on W_{2i}	93.4	**97.8**	94.0	96.6	95.0	96.8	95.6	93.2	95.0	95.4
Int(β_i)	93.2	97.2	95.2	**98.0**	94.2	96.0	96.0	95.4	94.0	95.6
Int(φ_i)	96.4	**99.2**	95.2	**98.0**	94.6	97.0	93.6	95.6	92.6	94.0

Note: ARR = Admissible-Range-Restricted, DD = Default Diffuse. Cell values represent percentages, bold entries indicate values outside of Bradley's [92.5, 97.5] range.

admissible-range-restricted condition had two instances that deviated from Bradley's range, but otherwise fell within the acceptable range across conditions. Overall, it appears that the credible interval coverage is performing well across conditions.

Given that coverage is reasonable, the next question is about the width of the credible intervals. This is related to efficiency in the frequentist framework – the desired credible intervals are those with proper coverage that are also as small as possible, which aids in minimizing uncertainty and maximizing the ability to detect non-null effects. Table 4 shows the average posterior standard deviations across all 500 replications for each parameter, for each condition. Notably, the average posterior standard deviations of the random effect variances are much larger in the default diffuse condition than in the admissible-range-restricted condition. This uncertainty (along with large positive relative bias) permeates to the between-model effects where the default diffuse average posterior standard deviations are also inflated for $N \leq 30$ even though these effects had identical prior distributions across conditions (i.e., only the priors of the random effect variances were manipulated across conditions; priors on fixed paths were not manipulated). As discussed in the next subsection, these artificially uncertain posterior distributions will affect the ability of the model to detect non-null effects.

TABLE 4
Posterior Standard Deviations for T = 50 across Sample Size and Prior Distribution Conditions

	N= 10		*N= 20*		*N= 30*		*N= 40*		*N= 50*	
	ARR	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*
Between-Model Paths										
φ_i on W_{1i}	.066	.091	.038	.042	.029	.031	.025	.025	.022	.022
φ_i on W_{2i}	.063	.087	.037	.041	.029	.031	.024	.025	.021	.022
β_i on W_{1i}	.296	.438	.184	.208	.143	.153	.121	.126	.107	.111
β_i on W_{2i}	.284	.418	.180	.203	.141	.152	.120	.127	.106	.110
$\ln(\sigma_i^2)$ on W_{1i}	.174	.222	.100	.105	.077	.079	.065	.067	.057	.058
$\ln(\sigma_i^2)$ on W_{2i}	.168	.215	.098	.103	.077	.079	.064	.066	.056	.057
α_i on W_{1i}	.232	.351	.015	.168	.115	.125	.098	.104	.087	.091
α_i on W_{2i}	.220	.334	.141	.162	.112	.122	.096	.103	.085	.089
Int(β_i)	.272	.272	.175	.175	.138	.138	.117	.117	.103	.103
Int(φ_i)	.055	.055	.036	.036	.029	.029	.024	.024	.022	.022
Var (α_i)	.168	.096	.113	.192	.090	.123	.078	.098	.069	.083
Var (β_i)	.338	1.502	.200	.304	.150	.192	.126	.150	.111	.127
Var (φ_i)	.013	.063	.007	.012	.006	.008	.005	.006	.004	.005
Var [$\ln(\sigma_i^2)$]	.097	.392	.053	.078	.041	.051	.035	.041	.030	.035

Note: ARR = Admissible-Range-Restricted, DD = Default Diffuse.

Non-null detection rate

Related to the issue of positively biased random effect variances with small samples and diffuse priors, inflated variances result in overestimation of the uncertainty in fixed coefficients, as seen in the posterior standard deviation reported in Table 4. Overestimating the uncertainty in the parameters has a direct effect on the non-null detection rate. Table 5 shows the standardized effect for each non-null effect as well as the non-null detection rate at each sample size for each prior distribution condition. Similar to relative bias, the different time-point conditions maintained similar patterns, so we report the $T = 50$ condition only in the interest of succinctness. In Table 5, it can be observed that the admissible-range-restricted priors yield higher non-null detection rates for $N \leq 30$. For $N > 30$, the rates are mostly within two percentage points (though there were four cases where the admissible-range-restricted condition maintained an advantage with respect to non-null detection).

Because the model contains many effects, we plot three representative effects from different portions of the model to help visualize the difference in non-null detection rates between prior distribution conditions across sample sizes:

- The effect of W_{2i} on $\ln(\sigma_i^2)$ (standardized effect = 0.10)
- The effect of W_{2i} on β_i (standardized effect = 0.20)
- The effect of W_{2i} on α_i (standardized effect = 0.30)

TABLE 5
Non-Null Detection Rates for Between-Person Parameters, for T = 50 across Sample Size and Prior Distribution Conditions

		N= 10		*N= 20*		*N= 30*		*N= 40*		*N= 50*	
Effect	*Standardized Coefficient*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*	*ARR*	*DD*
φ_i on W_{1i}	.60	35	17	74	67	92	89	96	97	99	99
φ_i on W_{2i}	.30	14	5	28	21	41	36	53	55	63	61
β_i on W_{1i}	.20	22	5	40	31	57	52	70	65	81	78
β_i on W_{2i}	.30	35	15	60	51	77	73	88	89	95	95
$\ln(\sigma_i^2)$ on W_{1i}	.55	45	28	86	83	96	95	99	99	100	100
$\ln(\sigma_i^2)$ on W_{2i}	.10	8	4	16	14	29	27	36	33	47	46
α_i on W_{1i}	.40	62	32	90	85	98	97	100	99	100	100
α_i on W_{2i}	.30	34	12	59	47	75	69	87	80	92	90

Note: ARR = Admissible-Range-Restricted, DD = Default Diffuse. Cell values represent percentages. The variance terms are not included because their definition is not consistent with the aim of the non-null detection rate. Variances are bounded below by 0, so their credible intervals necessarily do not include 0, and the non-null rate would uniformly be 100.

Figure 5 shows the non-null detection rate for these three effects across sample size and prior distribution conditions for the $T = 50$ condition. The plots show that the admissible-range-restricted prior detection rates are about twice the default diffuse rates at $N = 10$ and continue to be higher across all samples sizes, though the advantage diminishes over time and the detection rates ultimately converge. These results echo previous small sample studies showing that the positively biased random effect variance estimates adversely affect the model's ability to detect non-null effects. Specifying prior distributions that restrict the range of the prior is effective for augmenting non-null detection rates while also maintaining credible intervals with good coverage.

COVARYING RANDOM EFFECTS

The simulation in the previous section used data that were generated to have no covariance between the various random effects in the model. This simplifies the specification of the prior distributions because the priors can be chosen *univariately* such that each random effect variance receives a unique prior distribution. However, in many DSEM applications, it may be of interest to fit models that have random effects that covary. For instance, it may be of interest to test whether people with higher intercepts also tend to have stronger inertia slopes as well. This section overviews two methods to specify priors for random effects that covary and conducts a small simulation to show how results change with covarying random effects.

Inverse wishart priors

With covarying random effects, the default prior in M*plus* is no longer a univariate inverse gamma for each random effect variance but rather a multivariate inverse Wishart prior on the entire random effect covariance matrix. Inverse Wishart is a multivariate distribution whose diagonal elements are inverse gamma distributions, but the distribution also includes potentially non-null off-diagonal elements to account for relations between different random effects. An inverse Wishart distribution is defined as $\mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$ where $\boldsymbol{\Psi}$ is a $p \times p$ scale matrix for p the dimension of the matrix (i.e., the number of random effects) and ν the degrees of freedom (larger degrees of freedom make the prior more informative). In M*plus*, the default Inverse Wishart distribution for a random effect covariance matrix is $\mathcal{W}^{-1}(\mathbf{0}, -p-1)$. This results in covariances that can be anywhere in the $(-\infty, \infty)$ interval provided that the overall matrix is positive-definite. The kth diagonal element of the matrix can be represented as an inverse gamma distribution,

$$\Gamma^{-1}\left(\frac{\nu - p + 1}{2}, \frac{\psi_{kk}}{2}\right) \tag{4}$$

for ψ_{kk} the kth diagonal element of the $\boldsymbol{\Psi}$ matrix (Asparouhov & Muthén, 2010, p. 57). It is relevant to note that using M*plus* defaults or similarly diffuse specifications for the inverse Wishart distribution has been found to lead to poor performance (McNeish, 2016a; Schuurman, Grasman, & Hamaker, 2016), so researchers are similarly

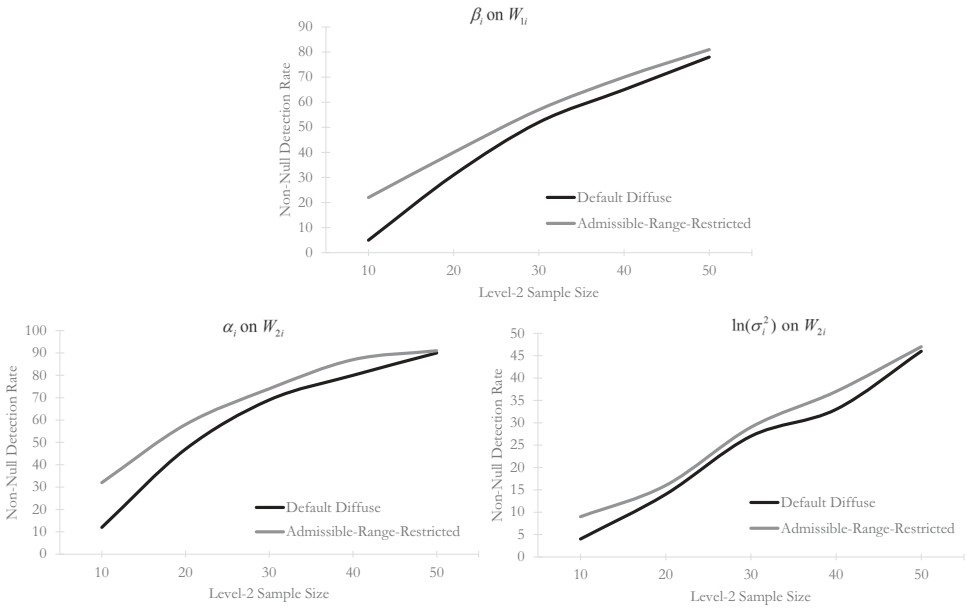

FIGURE 5 Three representative non-null detection rate plots across simulation conditions depicting $\beta_i$on W_{1i} (top), $\alpha_i$on W_{2i} (lower left), and $\ln(\sigma_i^2)$on W_{2i} (lower right). Note that the scale of the vertical axes is different to facilitate locating differences between conditions.

encouraged to restrict the support of the prior for inverse Wishart priors in the context of small samples should random effects be considered to have non-null relations with each other.

In M*plus*, an advantage of the Inverse Wishart distribution is that it guarantees that the random effect covariance matrix will be positive-definite because all the random effect terms are treated in one large block (e.g., Gelman et al., 2013). As disadvantages, the degrees of freedom are constant for all entries of the random effect covariance matrix, meaning that all the inverse gamma distributions upon the diagonal terms must share the same scale parameter (i.e., p is part of the scale parameter in Equation 4). This property limits the ability to feature prior distributions with different shape parameters in the inverse Gamma parameterization. In DSEM, this could present problems because the random effects can be on vastly different scales (i.e., the inertia slope variance is necessarily constrained to smaller values but the intercept variance could be quite large depending on the scale of the outcome). Additionally, the meaning of the inverse Wishart scale and degrees of freedom are not straightforward to interpret, so setting informative priors to aid small sample analyses can be quite difficult (e.g., Zhang, 2013).

Separation priors

As an alternative to the inverse Wishart prior, a separation prior strategy has been suggested (Barnard, McCulloch, & Meng, 2000). Separation priors decompose the covariance matrix into two separate components: variances (or standard deviations) and covariances (or correlations). Barnard et al. (2000) outlined a popular method to decompose the covariance matrix into a vector of correlations; log-normal priors are then applied to the standard deviations and a Uniform [−1,1] prior is applied to the correlations. A simulation by Liu, Zhang, and Grimm (2016) found separation priors to be preferable to inverse Wishart priors (i.e., smaller bias, better credible interval coverage) for linear and non-linear growth models in most cases.

Though conceptually appealing, such a strategy is difficult to implement within M*plus* (Schuurman, 2016, p. 28). That is, M*plus* provides a user-friendly approach to running complex models with Bayesian estimation, but the trade-off of this user-friendliness is that it is not always possible to implement custom or more general methods. However, a similar (but not exact) approach in the spirit of separation priors can be carried out in M*plus*. With Bayesian estimation within M*plus*, users have the option of specifying a prior distribution for each individual parameter in the model. Rather than modeling the random effect covariance matrix multivariately with one inverse Wishart prior, each of the variances can receive their own admissible-range-restricted inverse gamma prior just as outlined above in the prior simulation. When individual pieces of the random effects covariance matrix are given univariate priors (like inverse gamma), M*plus* will assign a normal prior with mean 0 and infinite variance to all covariances.

Of course, allowing the covariance to be any value will increase the chance that the entire covariance matrix becomes non-positive definite because the covariances are not dependent on each other. To keep covariances within the appropriate range, a MODEL CONSTRAINTS statement can be used to limit the covariance. A correlation of +1.0 would correspond to the product of the square roots of the two variances involved. For instance, the previous simulation specified that the variance of the intercepts was .30 and the variance of the time-varying covariate was .50. The product of the square roots would be $\sqrt{.30} \times \sqrt{.50} = .547 \times .707 = 0.387$, so the covariance between random effects of the intercept and time-varying covariate would need to fall between [−.387, .387]. These values are not known ahead of time because they are estimated, so specific scalar values cannot be used in the MODEL CONSTRAINTS statement. Instead, the variance parameters can be given a parenthetical label and the MODEL CONSTRAINTS statement can be built with the parameter labels. Specific code for implementing this method can be found in the supplemental materials. A small simulation is provided next to show the benefit of such an approach with small samples and covarying random effects and to show how sample size requirements change as a function of model complexity.

COVARYING RANDOM EFFECT SIMULATION

To demonstrate the viability of the aforementioned separation prior strategy in M*plus* and to show some weaknesses of the inverse Wishart diffuse default, we will conduct a small simulation with covarying random effects. The conditions are similar to the previous simulation in that we will include the five sample size conditions which are slightly larger because the model is more complex (N = 30, 40, 50, 75, 100) and two conditions for the prior distributions (default diffuse and admissible-range-restricted). The default diffuse condition is $\mathcal{W}^{-1}(\mathbf{0}, -5)$; the admissible-range-restricted condition uses the same univariate inverse gamma priors for the variances as the first simulation and normal distributions for the covariances. Because models with correlated random effects are more difficult and time-intensive to estimate, we only included the 50 time-point condition and simplified the model slightly. The data generation model for this simulation will be

$$y_{ti} = \alpha_i + \varphi_i\, y^{(c)}_{t-1,\,i} + \beta_i\, x_{ti} + e_{ti} \tag{5}$$

$$\begin{aligned}\alpha_i &= 0 + u_{0i}\\ \varphi_i &= .20 + u_{1i}\\ \beta_i &= .70 + u_{2i}\\ \ln(\sigma_i^2) &= 0 + u_{3i}\end{aligned} \tag{6}$$

$$\mathbf{u}_i \sim N\left(\mathbf{0}, \begin{bmatrix} .300 & & & \\ .014 & .010 & & \\ .097 & .018 & .500 & \\ .043 & .008 & .056 & .100 \end{bmatrix}\right) \tag{7}$$

The model no longer features any time-invariant covariates, and all off-diagonal elements of the random effect covariance matrix in Equation 7 are non-null with values selected to produce a correlation of 0.25 for each entry. 500 replications will be performed for each cell of the design, all models are estimated in M*plus* Version 8, all models are fit without misspecifications, and a minimum of 2500 iterations are completed for each replication with convergence judged by a PSR value less than 1.05 thereafter. More iterations were used in this simulation because the model was more complex and inspections of replications with 1000 iterations indicated some evidence that chains may be converging too early. Given the more detailed inspection of the first simulation, we will focus on results related to bias of the random effect covariance parameters and the coverage intervals of the fixed effects.

Simulation results

Similar to the first data generation model, the generation model for this simulation had four random effect variance terms: the intercept variance, the time-varying covariate variance, the inertia slope variance, and the log-residual variance. Figure 6 shows relative bias plots for each of these four random effect variance estimates across sample size conditions. The Flora and Curran (2004) 0.90 and 1.10 thresholds are superimposed as dashed line lines on each plot.

As a common theme in multilevel models in general, the more complex the model becomes, the larger the small sample bias grows and the sample size at which it remains also increases. This is directly observed here, as there is very noticeable small sample bias even as N approaches 100, which is quite different from the simpler model without random effect covariances in Figure 4. That being said, a similar pattern emerges with covarying random effects as with independent random effects. The admissible-range-restricted priors were able to notably reduce the bias in the random effect variances compared to the M*plus* default inverse Wishart. Unlike the simpler model, the admissible-range-restricted priors that we implemented could not completely remove all the bias from these estimates until a sample size of around 50 or 75.

Regarding the bias of the random effect covariances, the admissible-range-restricted priors provided better estimates of the covariances than the default diffuse priors based on the data generation values. However, the covariances are influenced by the variance estimates, so we also converted the covariances to correlations and we computed the standard root mean square residual (SRMR) that is commonly used in structural equation modeling to compare the estimated correlation matrix to the observed correlation matrix (the population correlations in this case). The SRMR values were reasonable for all sample sizes in the

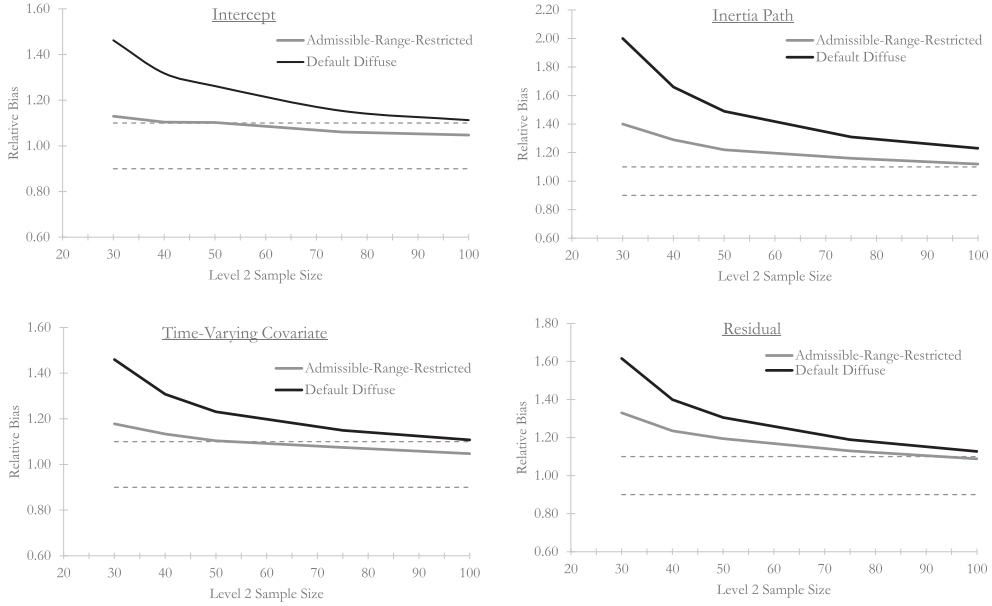

FIGURE 6 Relative bias plots for random effect variances of the intercept (upper left), inertia path (upper right), time-varying covariate (lower left), and residual (lower right) for the simulation with covarying random effects.

simulation (the maximum SRMR was .048 for the admissible-range-restricted condition with N= 30), though the default diffuse inverse Wishart did have a slightly smaller SRMR at smaller samples (where smaller indicates more accurate estimates). Overall, it seems that the random effect correlations were not greatly affected with smaller samples, even if the variances were biased.

Two conclusions can be drawn from the second simulation. First, sample size requirements needed to use the M*plus* diffuse default priors increases as a function of model complexity, especially if random effect covariances are desired. Second, the admissible-range-restricted method we outline can reduce the bias in the variance estimates, but it will not automatically eliminate the bias. The priors we used in the simulations were rather liberal with respect to what we considered "plausible". As model complexity increases with small samples, researchers likely have to be more aggressive with which values they consider plausible when constructing their priors in order to eliminate as much potential bias as possible.

DISCUSSION

The main take-home point from our simulation results is that the statistical properties of DSEM estimates are not necessarily poor with small N. Poor properties *can* occur if researchers allow default diffuse priors distributions to be used when small N is present. For the first model used in this simulation, small sample issues occurred roughly at $N < 50$, though small sample issues can remain at larger sample sizes for more complex models that can easily be fit within the DSEM framework. This was seen in the second simulation where small sample issues were present for $N = 100$. With small N relative to model complexity, the likelihood provided by the data is not sufficiently strong to outweigh the prior, so the prior becomes unintentionally informative, leading to increased bias of the estimates and decreased the ability to detect non-null effects due to unnecessarily large credible intervals. These issues can be circumvented if researchers override default priors. If available, the ideal way to do so is through literature review, meta-analysis, or expert opinion. Seeing as these are often difficult to come by, we demonstrated how even specifying a prior as simple as restricting support only to admissible or plausible values can notably improve small sample performance. As such, DSEM models seem to follow patterns seen in other multivariate random effects models – namely that Bayesian methods can be adopted to reduce sample size requirements but this strategy is only effective when researchers specify informative or weakly informative priors rather than relying on diffuse priors.

Given the novelty of DSEM modeling, there are several nuances and limitations of the study and results that we present here. First, the models we incorporated are restricted to two-level DSEM models with no missing data and each individual having the same number of time-points that were collected at the same intervals. DSEM modeling in M*plus* is also able to accommodate data where individuals have a different number of time-points or where each person's data come from a different set of time-points by using the TINTERVAL option. Other models within the broader DSEM framework also exist such as cross-classified time-series models and residual DSEM models, both of which are unexplored in the current study. These additional models may feature unique features or idiosyncrasies that make their small sample performance deviate from the results found with two-level DSEM simulation. Potential issues related to other DSEM models are worthy of additional study to determine how sample size affects the statistical properties of their estimates (e.g., Hamaker, Asparouhov, Brose, Schmiedek, & Muthén, 2018).

Second, we chose a rather simple method by which to narrow the support of the prior distribution. For instance, the admissible-range-restricted prior led to some non-negligible bias in the random effect variances with $N = 10$ in the first simulation and non-negligible bias persisted for some parameters with $N < 50$ in the simulation with covarying random effects. This suggests that improved performance might be achieved by selecting a prior whose support is even more restricted or using more aggressive definitions of values that might be considered to be plausible. We used liberal definitions to demonstrate that the priors that are chosen do not necessarily have to be very accurate for performance to be improved for simpler models, even though more refinement is likely required for more complex models. We also tried to avoid choosing priors that were centered over population values since this would artificially inflate the evidence of the point we wished to make. With that being said, the priors we used in the admissible-range-restricted conditions are certainly not the only choices and many options exists that could effectively carry out the same intent. For instance, uniform priors may be useful for parameters whose range is expected to be quite small such as the inertia slope variance. Our main point is that researchers should try to take some approach besides relying on default diffuse priors when N is small, especially $N < 100$.

As described, our method for reducing the range of priors would classify as *data-dependent* because descriptive statistics from the data are used to inform the prior distributions (i.e., the priors could not be created before data collection and the values could change from sample to sample). It would be possible to reduce the range of the priors without relying on the data if researchers had intuition about which values of the parameters are permissible. Alternatively, other more intensive data-dependent prior

methods may be useful for creating priors at smaller samples (McNeish, 2016b; Schuurman et al., 2016). Whereas the "weakly" data-dependent method we describe and use in our simulation makes minimal use of the data by using descriptive statistics, more intensive data-dependent prior methods involve fitting an entire model with the specific parameters of interest and using the naïve model estimates from maximum likelihood or least squares estimates to create informative priors by centering the prior around the naïve estimate. More intensive data-dependent methods tend to improve performance, though stronger reuse of the data can overestimate precision of the estimates because the uncertainty of the naïve model is not always accurately incorporated (Kass & Steffy, 1989). Kass and Natarajan (2006) also suggest a default conjugate approach, though Schuurman et al. (2016) found that its performance in the context of time-series analysis could be improved upon.

CONCLUDING REMARKS

With any model intended for longitudinal data, small sample sizes are an ever-present threat due to the logistic difficulties of obtaining many repeated measures on the same individuals. Recent studies have suggested that DSEM might be susceptible to small sample issues; however, it appears that the issue is attributable to how Bayesian analyses are conducted. With smaller samples, we encourage researchers not to rely on diffuse priors or default priors provided within software programs. Though harmless at larger sample sizes, diffuse priors are unintentionally informative at smaller samples and have adverse effects on results. Using previous studies, reviews or meta-analyses, or expert opinions is ideal, but in the absence of this type of information, our results show that even constructing priors based on intuitive restrictions for admissible values can greatly improve accuracy of parameter estimates and the ability to detect non-null effects.

REFERENCES

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling*, *25*, 359–388. doi:10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus. Technical Report. Los Angeles, CA: Muthén & Muthén.

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*, 1281–1311.

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*, 385–402. doi:10.1214/06-BA115

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods*. New York, NY: Guilford Press.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144–152. doi:10.1111/bmsp.1978.31.issue-2

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*, 473–514. doi:10.1214/06-BA117

Chatfield, C. (2016). *The analysis of time series: An introduction*. London, UK: CRC press.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. doi:10.1080/10705511.2014.937849

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*, 240–261. doi:10.1037/met0000065

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. doi:10.1037/1082-989X.9.4.466

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472. doi:10.1214/ss/1177011136

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. London, UK: Chapman & Hall.

Hamaker, E. L., & Dolan, C. V. (2009). Idiographic data analysis: Quantitative methods from simple to advanced. In J. Valsiner, P. C. M. Molenaar, M. Lyra, & N. Chaudhary (Eds.), *Dynamic process methodology in the social and developmental sciences* (pp. 191–216). New York, NY: Springer-Verlag.

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 doi:10.1080/00273171.2018.1446819

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. doi:10.1177/0963721416666518

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods.*, *6*, 87–93.

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. doi:10.1080/00273171.2014.1003772

Kass, R. E., & Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models. *Bayesian Analysis*, *1*, 535–542. doi:10.1214/06-BA117B

Kass, R. E., & Steffy, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, *84*, 717–726. doi:10.1080/01621459.1989.10478825

Kuljanin, G., Braun, M. T., & DeShon, R. P. (2011). A cautionary note on modeling growth trends in longitudinal data. *Psychological Methods*, *16*, 249–264. doi:10.1037/a0023348

Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*, 653–686. doi:10.1207/s15327906mbr3904_4

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse Wishart and separation-strategy priors for Bayesian estimation of covariance

parameter matrix in growth curve analysis. *Structural Equation Modeling*, *23*, 354–367. doi:10.1080/10705511.2015.1057285

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92. doi:10.1027/1614-2241.1.3.86

McNeish, D. (2016a). On using Bayesian methods to address small sample problems. *Structural Equation Modeling*, *23*, 750–773. doi:10.1080/10705511.2016.1186549

McNeish, D. (2016b). Using data-dependent priors to mitigate small sample size bias in latent growth models: A discussion and illustration using M*plus*. *Journal of Educational and Behavioral Statistics*, *41*, 27–56. doi:10.3102/1076998615621299

McNeish, D. (2017). Challenging conventional wisdom for multivariate statistical models with small samples. *Review of Educational Research*, *87*, 1117–1151. doi:10.3102/0034654317727727

McNeish, D., & Matta, T. (2018). Differentiating between mixed effects and latent curve approaches to growth modeling. *Behavior Research Methods*, *50*, 1398–1414. doi:10.3758/s13428-017-0976-5

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, *51*, 495–518. doi:10.1080/00273171.2016.1167008

Meredith, M., & Kruschke, J. (2017). HDInterval: Highest (posterior) density intervals. Retrieved from https://cran.r-project.org/package=HDInterval

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. doi:10.1037/a0026802

Muthen, B. O. (2017). Dynamic structural equation modeling of intensive longitudinal data using Mplus Version 8 part 8. http://www.statmodel.com/download/Part%208%20Muthen.pdf

Muthén, L. K., & Muthén, B. O. (1998-2018). *Mplus [Computer program]*. Los Angeles, CA: Muthén & Muthén.

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

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, *12*, 347–367. doi:10.1177/1094428107308906

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling*, *25*, 495–515. doi:10.1080/10705511.2017.1392862

Schuurman, N. K. (2016). So you want to specify an inverse-Wishart prior distribution. Paper presented at the 2016 Mplus User Meeting, Utrecht, the Netherlands.

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, *51*, 185–206. doi:10.1080/00273171.2015.1065398

Spiegelhalter, D. J., Myles, J. P., Jones, D. R., & Abrams, K. R. (1999). Methods in health service research: An introduction to bayesian methods in health technology assessment. *British Medical Journal*, *319*, 508.

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, *23*, 466–470. doi:10.1177/0963721414550706

van De Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, *6*, 25216. doi:10.3402/ejpt.v6.25216

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. doi:10.1111/cdev.12169

van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & van Loey, N. E. (2018). Bayesian PTSD-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, *53*, 267–291. doi:10.1080/00273171.2017.1412293

van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*, 217–261. doi:10.1037/met0000100

Veen, D., Stoel, D., Zondervan-Zwijnenburg, M., & van de Schoot, R. (2017). Proposal for a five-step method to elicit expert judgment. *Frontiers in Psychology*, *8*, 2110. doi:10.3389/fpsyg.2017.02110

Walls, T. A., & Schafer, J. L. (Eds). (2006). *Models for intensive longitudinal data*. Oxford: Oxford University Press.

Zhang, Z. (2013). Bayesian growth curve models with the generalized error distribution. *Journal of Applied Statistics*, *40*, 1779–1795. doi:10.1080/02664763.2013.796348

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*, 305–320. doi:10.1080/15427609.2017.1370966

Zondervan-Zwijnenburg, M., van de Schoot-Hubeek, W., Lek, K., Hoijtink, H., & van de Schoot, R. (2017). Application and evaluation of an expert judgment elicitation procedure for correlations. *Frontiers in Psychology*, *8*, 90. doi:10.3389/fpsyg.2017.00090

APPENDIX

R Code to calculate 95% High Density Credible Intervals for Priors

*Note that R uses a different parameterization of the inverse gamma distribution. “rate” refers to scale in M*plus*. *

```
install.packages("HDInterval")
install.packages("invgamma")
library(invgamma)
library(HDInterval)

#Intercept variance#
int <- function(x)
qinvgamma(x,shape=2,rate=.30)
hdi(int)

#Inertia variance#
phi <- function(x)
qinvgamma(x,shape=1,rate=.01)
hdi(phi)

#TVC variance#
tvc <- function(x)
qinvgamma(x,shape=1.25, rate=.25)
hdi(tvc)
```

```
#log residual variance#
tvc <- function(x)
qinvgamma(x,shape=1.5,rate=.20)
hdi(tvc)
```

R Code to visualize a Prior

*Note that R uses a different parameterization of the inverse gamma distribution. "rate" refers to scale in M*plus*. *

```
install.packages("invgamma")
library(invgamma)
windowsFonts(A = windowsFont("Times New Roman"))
## Intercept Variance ##
s<-seq(0,2,.001)
plot(s,dinvgamma(s,shape=2,rate=.30),family="A",
type='l',xlab="Parameter Value",
ylab="Probability Density Function", lwd=2, lty=1,
cex.lab=1.5, cex.axis=1.75, bty="n")

## Inertia ##
s<-seq(0,2,.001)
plot(s,dinvgamma(s,shape=1,rate=.01),family="A",
type='l',xlab="Parameter Value",
ylab="Probability Density Function", lwd=2, lty=1,
cex.lab=1.5, cex.axis=1.75, bty="n")
## TVC ##
s<-seq(0,2,.001)
plot(s,dinvgamma(s,shape=1.25,rate=.25),
family="A", type='l',xlab="Parameter Value",
ylab="Probability Density Function", lwd=2, lty=1,
cex.lab=1.5, cex.axis=1.75, bty="n")

## Log Sigma ##
s<-seq(0,2,.001)
plot(s,dinvgamma(s,shape=1.5,rate=.2),family="A",
type='l',xlab="Parameter Value",
ylab="Probability Density Function", lwd=2, lty=1,
cex.lab=1.5, cex.axis=1.75, bty="n")
```