A Single-Level Random Effects Cross-Lagged Panel Model for Longitudinal Mediation Analysis

Wei Wu Ian A. Carroll Po-Yi Chen

University of Kansas

Correspondence concerning this article should be addressed to Wei Wu, Psychology Department, Indiana University Purdue University Indianapolis, 402 N Blackford LD 120c, Indianapolis, IN 46202. E-mail: wwu1@iu.edu.

Keywords:

Random effects, cross-lagged panel model, mediation, longitudinal data

---

This is the author's manuscript of the article published in final edited form as:

Wu, W., Carroll, A. I., & Chen, P-Y. A Single-Level Random Coefficients Cross-Lagged Panel Models for Longitudinal Mediation Analysis. (2017). Behavioral Research Methods. https:/doi.org/10.3758/s13428-017-0979-2

Abstract

Cross-lagged panel models (CLPMs) are widely used to test mediation with longitudinal panel data. One major limitation of the CLPMs is that the model effects are assumed to be fixed across individuals. This assumption is likely to be violated (i.e., the model effects are random across individuals) in practice. When this happens, the CLPMs can potentially yield biased parameter estimates and misleading statistical inferences. This article proposes a model named a random effects cross-lagged panel model (RE-CLPM) to account for random effects in CLPMs. Simulation studies show that the RE-CLPM outperforms the CLPM in recovering the mean indirect and direct effects in a longitudinal mediation analysis when random effects exist in the population. The performance of the RE-CLPM is robust to a certain degree even when random effects are not normally distribution. In addition, the RE-CLPM does not produce harmful result when the model effects are in fact fixed in the population. Implications of the simulation studies and potential directions for future research are discussed.

Keywords: cross lagged panel model, longitudinal mediation, random effects, heteroscedasticity

A Single-Level Random Effects Cross-Lagged Panel Models for Longitudinal Mediation Analysis

Mediation analysis is often used in social and behavioral sciences to explain the mechanism for how and why an effect occurs (MacKinnon, 2008). Mediation is said to occur when the effect of an independent variable (X) on a dependent variable (Y) is transmitted by a third variable (M; Baron & Kenny, 1986). The third variable M is termed a mediator. When mediation occurs, the total effect of X on Y is partitioned into two components: the indirect effect of X on Y through M and the direct effect of X on Y that cannot be explained by the indirect effect (denoted by *c*'). The indirect effect is denoted by *ab* because it is often quantified by the product of two effects: the effect of X on M (*a* effect) and the effect of M on Y controlling for X (*b* effect; MacKinnon, 2008). Given a non-zero *ab,* if *c*' is zero in the population, M is said to completely mediate the X and Y relationship. Otherwise, M partially mediates the effect of X on Y.

In practice, it is strongly recommended to establish mediation with longitudinal data (Cole & Maxwell, 2003; Maxwell, & Cole, 2007) for at least two reasons: 1) the effects (e.g., *a*, *b*, *ab*, and *c*') involved in a mediation analysis are causal effects which need time to unfold. Temporally, X precedes M, and M precedes Y. 2) Prior levels of M or Y are confounding variables which should be controlled when testing the indirect effect and direct effect. Failure to do so can lead to substantial bias in parameter estimates in mediation analysis (Cole & Maxwell, 2003; Gollob & Reichardt, 1991).

Note that there are two types of longitudinal designs: longitudinal panel designs and intensive longitudinal designs. For the former designs, individuals are repeatedly measured for a limited number of time points each separated by several months or more, while for the latter

designs, individuals are measured for a large number of time points with relatively brief intervals between occasions (Collins, 2006; Wu, Selig, & Little, 2013). Models suitable for the data collected from the two designs may be different. The current study is focused on mediation models designed for the data resulted from longitudinal panel designs.

With longitudinal panel designs, the most popular model for mediation analysis is the cross-lagged panel model (CLPM; Preacher, 2015). As pointed out by Selig and Preacher (2009), "the CLPM allows time for causes to have their effects, supports stronger inference about the direction of causation in comparison to models using cross-sectional data, and reduces the probable parameter bias that arises when using cross-sectional data". However, there is a limitation of the CLPM which has been overlooked in previous research: the effects (e.g., direct and indirect effects) in the CLPM are assumed to be fixed (i.e., constant) across individuals. As a result, researchers can neither account for potential individual variability on the effects nor evaluate covarying relationships that may have arisen out of this variability. If any of the effects is random (i.e., variant) across individuals, the CLPM may yield biased parameter estimates and misleading statistical inferences.

There exists an alternative model for longitudinal mediation analysis that can accommodate individual-specific direct and indirect effects. This model is a multilevel model (MLM) which is proposed based on the fact that longitudinal data are clustered in nature: the repeated measures are nested within individuals (Kenny, Korchmaros, & Bolger, 2003; Bauer, Preacher, & Gil, 2006). However, the MLM has several limitations. First, the relationships among X, M, and Y are examined concurrently in the MLM proposed in the previous literature. This is only appropriate when the causal effects occur in a very short time frame. Second, the previous measures on M and Y are not controlled in the model. Third, although effects are

allowed to be random across individuals in the MLM, these effects are constrained to be equal over time. Thus, it cannot capture possible changes of these effects over time. These limitations make MLM a less desirable method for longitudinal mediation analysis than the CLPM.

In this article, we propose a new model to incorporate random effects in longitudinal mediation analysis. This model is a direct extension of the CLPM by allowing random effects in the model. Unlike the MLM, the proposed model is a single-level model, thus we refer to the model a single level random effects cross-lagged panel model (RE-CLPM). Through the model, we emphasize that it is possible to accommodate random effects even when the model is a single level model. The proposed model can be implemented in Mplus (Muthén, Muthén, & Asparouhov, 2015), thus is readily available to applied researchers.

The rest of the article is organized as follows. We first review the traditional CLPM for longitudinal mediation analysis followed by an introduction of the proposed method. We then present three simulation studies conducted to evaluate the performance of the proposed model under a variety of conditions, in comparison to the traditional CLPM. We conclude this article by discussing limitation and implications of the simulation studies as well as potential directions for future research.

Traditional Cross-Lagged Panel Model (CLPM)

Fitting a cross lagged panel model for longitudinal mediation typically requires at least three waves of data, although it is possible to run the analysis with two waves of data with additional assumptions imposed (Cole & Maxwell, 2003). Three waves of data are also most common in practice. Thus we illustrate the traditonal CLPM using a longitudinal study with three waves of data on the variables X, M, and Y.

To fit a CLPM, the longitudinal data must be organized in a way so that the repeated measures at each of the three time points are represented by separate variables such as $X_1 \ldots X_3$, $M_1 \ldots M_3$, and $Y_1 \ldots Y_3$, respectively. This way to organize the data is called wide format. Although the longitudinal data are multilevel in nature, with data organized in wide format, the CLPM is a single level model. This is analogous to the comparison between the MLM and latent curve model approaches for growth curve modeling where the latter is viewed a single level model. Figure 1 shows a typical CLPM that can be fit to the wide format data to test longitudinal mediation. With only manifest variables, this model is simply a path analysis model in structural equation modeling (SEM).

[Figure 1 is about here]

The relationships among the variables in the CLPM can be described using the following equations (see also Selig & Preacher, 2009).

$$X_{it} = d_{xt} + s_{x(t-1)} X_{i(t-1)} + e_{xit}$$

$$M_{it} = d_{mt} + a_{(t-1)} X_{i(t-1)} + s_{m(t-1)} M_{i(t-1)} + e_{mit} \qquad (1)$$

$$Y_{it} = d_{yt} + s_{y(t-1)} Y_{i(t-1)} + c'_{(t-2)} X_{i(t-2)} + b_{(t-1)} M_{i(t-1)} + e_{yit}$$

Where *i* represents individual and *t* represents time point; the *s* coefficients represent *autoregressive* effects which capture the stability of the constructs over time in terms of rank orders of the scores; the *a, b,* and *c'* coefficients represent *cross-lagged* effects which are the focus of a mediation analysis. The *d* coefficients are intercepts, and the *e*s are residuals. Let $\mathbf{e}_{it} = (e_{xit}, e_{mit,} e_{yit})'$ be the vector of residuals for individual *i*. The residuals are assumed to follow a multivariate normal distribution with

$$\boldsymbol{\mu}(\mathbf{e}_{it}) = (0,0,0)', \text{ and } \boldsymbol{\Sigma}(\mathbf{e}_{it}) = \begin{pmatrix} \sigma^2_{ext} & & \\ \sigma_{emt,ext} & \sigma^2_{emt} & \\ \sigma_{eyt,ext} & \sigma_{eyt,emt} & \sigma^2_{eyt} \end{pmatrix}.$$

Where $\boldsymbol{\mu}(\mathbf{e}_{it})$ is the mean vector and $\boldsymbol{\Sigma}(\mathbf{e}_{it})$ is the covariance matrix of the residuals. Note that following common practice, we also assume that the residuals are not correlated over time although this assumption may be relaxed.

The model shown in Figure 1 allows researchers to examine the indirect effect of X_1 on Y_3 through M_2 (i.e.,a_1b_2) and the leftover direct effect of X_1 on Y_3 (i.e., c`). As mentioned above, the indirect and direct effects, although they may vary across time, are assumed to be constant across individuals in this model. In this article, we show that this assumption can be relaxed using the random effects cross lagged panel models (RE-CLPM).

Random Effects Cross Lagged Panel Models

The RE-CLPM is an extension of the CLPM by allowing effects in the model to be random. Using the same example shown in Equation (1), if all of the effects are random, then Equation (1) should be modified as follows.

$$\begin{aligned} X_{it} &= d_{xt} + s_{xi(t-1)}X_{i(t-1)} + e_{xit} \\ M_{it} &= d_{mt} + a_{i(t-1)}X_{i(t-1)} + s_{mi(t-1)}M_{i(t-1)} + e_{mit} \qquad (2) \\ Y_{it} &= d_{yt} + s_{yi(t-1)}Y_{i(t-1)} + c'_{i(t-2)}X_{i(t-2)} + b_{i(t-1)}M_{i(t-1)} + e_{yit} \end{aligned}$$

Comparing Equation (2) to Equation (1), an *i* subscript is attached to each of the effects that are allowed to be random (i.e., *s*, *a*, *b*, and *c*' effects), indicating that they are individual specific. The RE-CLPM allow more parameters to be estimated than the CLPM. For each random effect, mean and variance will be estimated in addition to possible covariances with any

of the other random effects. For simplicity, let's assume that all of the random effects are constant across time, and $\mathbf{u_i} = (s_{xi,} s_{mi,} s_{yi,} a_i, b_i, c_i)'$ represents the vector of random effects for individual *i*. The mean vector of the random effects [$\mathrm{E}(\mathbf{u_i})$] can be then written as

$$E(\mathbf{u_i}) = (s_{x,} s_{m,} s_{y,} a, b, c)' \quad ,$$

where *a*, *b and c* represent the means of the *a*, *b, and c* effects across individuals. The covariance matrix of the random effects [$\mathbf{\Sigma(u_i)}$] can be written as

$$\mathbf{\Sigma(u_i)} = \begin{pmatrix} \sigma^2_{sx} & & & & & \\ \sigma^2_{sm,sx} & \sigma^2_{sm} & & & & \\ \sigma_{sy,sx} & \sigma_{sy,sm} & \sigma^2_{sy} & & & \\ \sigma_{a,sx} & \sigma_{a,sm} & \sigma_{a,sy} & \sigma^2_a & & \\ \sigma_{b,sx} & \sigma_{b,sm} & \sigma_{b,sy} & \sigma_{b,a} & \sigma^2_b & \\ \sigma_{c,sx} & \sigma_{c,sm} & \sigma_{c,sy} & \sigma_{c,a} & \sigma_{c,b} & \sigma^2_c \end{pmatrix},$$

where σ^2_a and σ^2_b are variances of *a* and *b* effects, respectively, and $\sigma_{b,a}$ is the covariance between *a* and *b* effects. Based on these parameter estimates, the mean and variance of the indirect effect (ab_i) can be computed using the following formula (Kendall and Stuart, 1969).

$$E(ab_i) = a \times b + \sigma_{b,a} \qquad (3)$$

$$\mathrm{var}(a_i b_i) = b^2\sigma^2_a + a^2\sigma^2_b + \sigma^2_a\sigma^2_b + 2ab\sigma_{b,a} + (\sigma_{b,a})^2 \qquad (4)$$

Note that for Equation 4 to be valid, the random effects are assumed to be normally distributed. However, this assumption does not need to be true for Equation 3 to be valid (Kendall & Stuart, 1969).

One limitation of the RE-CLPM is that with random effects, popular model fit indices that are available to the CLPM such as the chi-square test statistic, root mean square error of approximation (RMSEA), and comparative fit index (CFI) cannot be computed. However, researchers can use information criteria such as Akaike Information Criterion (AIC, Akaike, 1973) and Bayesian Information Criterion (BIC, Raftery, 1995) to compare the RE-CLPM to CLPM. Note that there is also a sample size adjusted version for either BIC or AIC. The adjusted BIC and AIC are referred to as ABIC and AICc, respectively. Past research has found that ABIC could outperform BIC with a small sample size or a large number of parameters (Yang, 2006), and AICc could outperform AIC with a small sample size (Hurvich & Tsai, 1989; Burnham & Anderson, 2004). The performance of these information criteria in selecting the right model between the traditional and RE-CLPM is examined in the simulation studies. Note that for any of the criteria, a smaller value indicates a better model.

$$AIC = -2LL + 2k$$

$$BIC = -2LL + k\ln(n)$$

$$ABIC = -2LL + k\ln(\frac{n+2}{24})$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

where *LL* stands for log likelihood, *n* is sample size[1], and *k* is the number of parameters.

[1] Note that in the multilevel modeling framework, there is a debate over what "n" should be used in computing the information criteria (the number of individuals, the number of repeated observations, or something in between). The same debate may apply to the RE-CLPM. Since sample size is used as the n to calculate the information criteria for the traditional CLPM, to facilitate the comparison, sample size is also used as the n to compute the information criteria for the RE-CLPM.

As mentioned above, the RE-CLPM is a single-level model which is different from the multilevel cross-lagged panel models for mediation analysis (Fall, 2011; Preacher, Zyphur, & Zhang, 2010; 2011) that are designed for cases where individuals are nested within higher level units (e.g., classrooms). So with only a single level, why should random effects be modeled and how the parameters related to the random effects are estimated?

Why should Random Effects be Modeled?

If effects are random but treated as fixed in the CLPM, there are two potential consequences. First, the covariance between the *a* and *b* effects that comprise an indirect effect cannot be estimated (i.e., it is treated as 0). Since the covariance should be taken into account when computing the mean indirect effect (see Equation 3), this will result in a biased estimate for the mean indirect effect. Second, random effects will cause heteroscedasticity in the residuals. Failure to take into account this heteroscedasticity can bias the standard error estimates, leading to misleading statistical inferences. The second point is further illustrated below using simple regression analysis and path analysis as examples.

Simple regression model. Let Y be the dependent variable and X be the independent variable in a simple regression analysis. The regression equation is written as follows when the X effect is fixed across individuals.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (5)$$

where β_0 and β_1 represent the fixed intercept and regression coefficient, respectively, and ε_i represents the residual from the fixed effect for individual *i*. Here ε_i is assumed to be independent of *X* and have a constant variance over X (homoscedasticity assumption).

If the regression coefficient β_1 is random, then Equation (5) should be modified to

$$Y_1 = \beta_0 + (\beta_1 + \delta_i)X_i + e_i = (\beta_0 + \beta_1 X_i) + (\delta_i X_i + e_i) \quad (6)$$

where the effect of X on Y for individual *i* contains two components: β_1 (the mean effect of X on Y across individuals) and δ_i (the deviation of the effect for individual *i* from the mean effect). The residual from the random effect is denoted by e_i which is assumed to be normally distributed and independent of X. Based on Equations (5) and (6), it is clear that if a fixed effect model is fit to the data generated from a random effect model, then the homoscedasticity assumption will be automatically violated (i.e., the residuals will be heteroscedastic). Specifically, ε_i will be dependent on X ($\varepsilon_i = \delta_i X_i + e_i$), and the variance of ε_i will be no longer constant over X but a quadratic function of X. The quadratic function is presented in Equation (7) and displayed in Figure 2. As shown in Equation (7), the quadratic effect of X on the residual variance is determined by the variance of the variance of the random coefficient [i.e., var(δ_i)]. As the latter increases, the severity of the heteroscedasticity also increases.

$$\text{var}(\delta_i X_i + e_i) = \sigma_\delta^2 X_i^2 + 2\sigma_{\delta,e} X_i + \sigma_e^2 . \quad (7)$$

[Figure 2 is about here]

Path analysis model. Since the CLPM is a path analysis model, we extend the illustration to a path analysis model where two or more regression equations are estimated simultaneously. Considering a simple path analysis model with two regression equations in which Y_1 is regressed on X_1 and Y_2 is regressed on X_2. The two equations with random effects can be written as follows.

$$Y_{1i} = (\beta_1 + \beta_2 X_{1i}) + (\delta_{1i} X_{1i} + e_{1i})$$

$$Y_{2i} = (\beta_3 + \beta_4 X_{2i}) + (\delta_{2i} X_{2i} + e_{2i})$$

With two regression equations, in addition to residual variances (see Equations 8 and 9), the residual covariance may also be heterogeneous over Xs (see Equation 10). Specifically, if there was a covariance between the random effects ($\sigma_{\delta_1,\delta_2}$). X_1 and X_2 will interact with each other to influence the residual covariance. The larger the covariance, the bigger the interaction effect is.

$$\text{var}(\delta_{1i} + e_{1i}) = \sigma^2_{\delta_1} X^2_{1i} + 2\sigma_{\delta_1,e_1} X_{1i} + \sigma^2_{e_1} \quad (8)$$

$$\text{var}(\delta_{2i} + e_{2i}) = \sigma^2_{\delta_2} X^2_{2i} + 2\sigma_{\delta_2,e_2} X_{2i} + \sigma^2_{e_2} \quad (9)$$

$$\text{cov}(\delta_{1i} X_{1i} + e_{1i}, \delta_{2i} X_{1i} + e_{2i}) = \sigma_{\delta_1,\delta_2} X_{2i} X_{1i} + \sigma_{\delta_1,e_2} X_{1i} + \sigma_{\delta_2,e_1} X_{2i} + \sigma_{e_1,e_2} \quad (10)$$

How to Estimate a Single-Level Random Coefficient Model?

Two approaches have been developed to estimate a single-level random coefficient model. The first approach is to model the heteroscedasticity directly based on the pattern of the heteroscedasticity. This approach was initially developed in economics to model random effects in a single-level linear regression analysis (Goldstein, 2003; Hildreth & Houck, 1968; Johnston, 1984; Swamy, 1971; Weisberg, 2014). Taking the example shown in Equation (8), this approach will explicitly specify the residual variance of Y on X to be a quadratic function of X, and uses the estimated quadratic effect as the estimate for the variance of the random effect of X on Y [i.e., $\text{var}(\delta_i)$]. Following the same logic, the covariance between the two random effects shown in Equation (13) can be estimated as the interaction effect between X_1 and X_2 on the residual covariance. As one can imagine, this approach can soon become cumbersome when the number of random effects increases.

The second approach is to model the random effects directly which indirectly accounts for the heteroscedasticity. This is the approach implemented in M*plus* (Muthén, Muthén, & Asparouhov, 2015). This approach treats random effects as unobserved or latent variables, which

is essentially no different from how random effects in a multilevel structural equation model are handled. Because it is computationally challenging to estimate models with random effects, an iterative algorithm called an EM algorithm is used to obtain the maximum likelihood estimates of model parameters. EM stands for expectation and maximization which are the two steps involved in each iteration of the EM algorithm. Simply speaking, the expectation step predicts the values of the random effects from the observed data. The maximization step then treats the predicted values for the random effects as if they were known and uses them to estimate the model parameters through the maximum likelihood estimation method. The EM algorithm iterates between the two steps until the parameter estimates stop changing meaningfully from one iteration to the next. In order to mitigate some of the computational intensity of the EM algorithm and hasten convergence, an accelerated version of the EM algorithm is used in Mplus. Readers can refer to Jamshidian and Jennrich (1997) or Lee and Poon (1998) for information on the ways to accelerate an EM algorithm.

Simulation Studies

In this section, three simulation studies are conducted to assess the performance of the RE-CLPM. The first study is designed to examine whether the RE-CLPM performs sub-optimally when the model effects are in fact fixed in the population. The second study aims to examine whether RE-CLPM performs as expected when the effects are random in the population, and to what extent using the traditional CLPM can lead to biased result. The third study is a preliminary investigation on whether the RE-CLPM is robust to the violation of the normality assumption of random effects. All of the simulation studies are conducted using SAS 9.4 and M*plus* 7.4 (Muthén & Muthén, 1998-2015).

Study 1

Recall that the purpose of study 1 is to examine how the random effects models will perform if the effects were fixed (i.e., no heteroscedasticity) in the population. Two models are compared in this study: RE-CLPM and CLPM.

Data Generation Model

The data generation model in this study is similar to the CLPM shown in Figure 1 with three waves of data. The variables at the first time point (i.e., X_1, M_1 and Y_1) were all in standardized metric and had a correlation of .3 with one another. For simplicity, all of the path coefficients were constrained to be equal over time, hereinafter, we use *a*, *b*, *c*', and *ab* to represent the constant coefficients over time. Following an example provided in Maxwell and Cole (2007, Figure 4), the autoregressive effects were set at .8, .3, and .7 for X, M, and Y, respectively, representing a wide range of stability coefficients. The effect size for *a* or *b* (note that the size of *a* and *b* are kept equal) was varied at three levels (explained below). The direct effect (i.e., *c*') was fixed at 0, representing a complete mediation scenario. The residual variances on X, M, and Y (at the 2nd, 3rd, and 4th time points) were constant over time. They were set up in a way so that the outcome variables were in a standardized metric at any of the time points (see the population values for the residual variances below). The data generation model with population values for the parameters is presented in Figure 3.

[Figure 3 is about here]

Design Factors

Two design factors are manipulated in this study: sample size and effect size of *a* and *b*. Sample size was varied at three levels (100, 200, or 500), reflecting small, moderate, and large sample sizes in social and behavior sciences. The effect size of *a* and *b* was varied at two

levels: .3 (*ab* = .09), or .6 (*ab* =.36), representing small and large effects, respectively. Correspondingly, the residual variances on X, M and Y were set at .36, .82, and .42, respectively, when *a* and *b* = .3, and at .36, .55. and .15, respectively, when *a* and *b* = .6, to make the total variances of the variables to be 1. In total, there are 3 × 2 = 6 conditions. One thousand replications were generated for each condition and the two models were fit to each replication. The CLPM is the same as the population generating model. The RE-CLPM was specified as the same as that used in simulation study 2. Specifically, all of the effects (both autoregressive and cross-lagged effects) in the RE-CLPM were freed to be random across individuals but equated across time. In addition, the correlation between a_i and b_i is freely estimated. Note that the RE-CLPM is an overspecification of the population model in this case. The M*plus* syntaxes for fitting the two models can be found in the Appendices.

Evaluation Criteria

The following criteria were used to evaluate the examined models. We focus on the extent to which the models recover the mean effects *a, b, c',* and *ab* given that they are often of most interest to applied researchers.

Convergence rate. Convergence rate is calculated as the proportion of replications that successfully converged. With more parameters to be estimated, we expect that it is more challenging for RE-CLPM to converge than the CLPM, especially under small sample sizes.

Percentage relative bias. The percentage relative bias for a specific parameter θ is computed as follows if the population value for the parameter (θ_0) is not zero or close to zero.

$$\%bias = 100 \times (\bar{\theta} - \theta_0) / \theta_0$$

where $\bar{\theta}$ is the average effect over replications. If θ_0 is zero or close to zero (< .01), the % relative bias is computed by $100 \times (\bar{\theta} - \theta_0)$. Following Muthén, Kaplan and Hollis (1987), a relative bias less than 10% is considered acceptable.

Confidence interval coverage. The 95% confidence interval (CI) coverage rates are reported for the target effects. For each parameter, the 95% CI coverage is calculated as the proportion of the 95% CIs that cover the population value across replications. Upper and lower limits of the CIs were computed based on the corresponding standard error estimates. Note that the standard error for *ab* is computed using the delta method which is a popular method for obtaining standard errors of nonlinear combinations of estimated coefficients (Sobel, 1982). The CI for *ab* computed in this way is not optimal because it cannot reflect the nonnormality of *ab*. Given this practical concession as it relates to simulation design, it suffices for the purpose of the study to differentiate the performances of the examined methods. Based on the acceptable range for the type I error rate, a coverage rate between 92.5% to 97.5% is considered acceptable.

Power or type I error rate. Power for any non-zero effect or type I error rate for any zero effect was calculated as the proportion of replications that yield significant effect under each condition. Following Bradley (1978), a type I error rate within the range of 0.025 to 0.075 is considered as acceptable for a 0.05 nominal rate.

Model Selection

In addition to parameter estimates, we also examined the extent to which the true model (i.e., the CLPM in this study) can be correctly selected using AIC, BIC, ABIC, and AICc. Specifically, the success rate of model selection (i.e., the proportion of replications for which an information criterion led to the correct model selection) was used to quantify the accuracy of model selection.

Result

The result from the first study is summarized in Table 1. Since the same result pattern was observed for different effect sizes and sample sizes, we only reported the result for effect size = 0.3 and n = 200 in Table 1. In addition, the result for *a* effect was consistent with that for the *b* effect, thus, we did not reported the result for the *b* effect in Table 1. As shown in Table 1, the RE-CLPM and CLPM were comparable regarding parameter estimates, CI coverage, and type I error rates, indicating that allowing random effects when the effects were in fact fixed in the population was not particularly harmful. However, the CLPM was more efficient since it yielded higher power to detect the target effects.

[Table 1 is about here]

Convergence rate. The convergence rate for the CLPM was 100% under all conditions. The estimation for the RE-CLPM researched saddle points for some of the replications where the negative of the matrix of the second derivatives with respect to the model parameters is not positive definite, especially when the sample size was small. For the RE-CLPM, saddle points occurred in 53.0%, 15.6%, and 0.4% of the replications for n = 100, 200, and 500, respectively. When saddle points occur, Mplus uses an ad-hoc estimator to compute the standard error estimates. A detail explanation for saddle points and possible causes for saddle points are provided in Asparouhov and Muthén (2012). Since we are not sure about the accuracy of the ad-hoc estimator, the replications with saddle points were not included in the final result.

Model selection. For all of the conditions examined in the study, all information criteria led to above 98% success rate of selecting the right model, indicating that the information criteria all performed very well in differentiating the traditional and random effects models when the misspecified model is an overspecification of the true model.

Study 2

As mentioned above, the purpose of study 2 is to compare the performance of the RE-CLPM and CLPM when random effects exist in the population.

Data Generation Model

Using the CLPM shown in Figure 3 as the baseline model, we freed the autoregressive (i.e., *s*) and cross-lagged (i.e., *a*, *b*, and *c*') effects to be random and used the modified model to generate data in this study. For simplicity, the random effects were normally distributed and equated across time, and only the *a* and *b* effects were allowed to correlate. The variance of *c*' was fixed at .04. The variances of the *s* effects were all fixed at .01. The variances of *a* and *b* were varied at two levels as shown below.

Design Factors

Besides the two factors (sample size and effect size of *a* and *b*) varied at study 1, we manipulated two additional factors in this study. The population values for the design factors are selected based on the study conducted by Bauer, Preacher, and Gil (2006).

1. Variance of *a* or *b* (σ_a^2 or σ_b^2) is varied at two levels: .04, .16. These values are chosen to represent low and high levels of heterogeneity in the data. Since the RE-CLPM is not currently used by researchers, there exist no conventions or individual cases to aid in determining "low" and "high" heterogeneity in this context. To this end, the values for the variances are simply motivated to induce differences in results by method, such that one can infer general themes in the performance of CLPM vs. RE-CLPM for data of this type.

2. The correlation between the *a* and *b* effects ($r_{a,b}$) is varied at five levels: -6, -.3, 0, .3, .6, representing large negative, small negative, zero, small positive, and large positive correlations between the two effects.

In sum, there are 3 (sample size) × 2 (effect size) × 2 (variance) × 5 (correlation) = 60 conditions. One thousand replications were generated for each condition, and the same two models in study 1 were fit to each replication. The same set of information criteria was also used to compare the models. Similar to study 1, the performance of AIC, BIC, ABIC, and AICc in recovering the true model (RE-CLPM in this case) was examined.

Result

The result from the second simulation study is summarized in Table 2. Similar to Study 1, because the result patterns are the same across effect sizes and sample sizes, only the result for effect size = .3 and n = 200 was reported. As shown in Table 2, it is evident that the RE-CLPM performed well under all examined conditions. In contrast, the CLPM, because it fails to account for random effects, did not yield satisfactory result in most of the conditions. We describe the result in detail below.

[Table 2 is about here]

Convergence rate. Again, the CLPM converged 100% of the time under all conditions. The estimation for the RE-CLPM researched saddle points in about 19%, 1.9%, and 0% of the replications on average for *n* = 100, 200, and 500, respectively. Thus considering saddle points, n ≥ 200 seemed preferred for the RE-CLPM.

Bias. The relative biases from RE-CLPM were all within 5% (see Table 2). In comparison, although the estimates of *a*, *b*, or *c*' from the CLPM were accurate, the estimates for *ab* were biased when the covariance between *a* and *b* was nonzero. This is not surprising because

the CLPM could not account for the covariance between *a* and *b* in the computation of *ab* (see Equation 3). Consequently, it underestimated *ab* when the covariance was positive (relative bias ranged from -2.84 % to -55.55%), and overestimated *ab* when the covariance was negative (relative bias ranged from 8.41% to 1401.67%). The bias was larger as the variance increased and size of the correlation increased.

CI coverage. The CI coverage rates from the RE-CLPM all fell in the acceptable range (see Table 3). In contrast, the CI coverage rates from the CLPM were below the acceptable range under most of the conditions. The CI coverage rate from the CLPM tended to decrease as the variance of *a* or *b* increased for the focal effects. In the most extreme case, the CI coverage rate for *ab* was 25% when the variance of *a* or *b* was large and the covariance between a_i and b_i was positive. This low CI coverage was attributable to both the bias in the parameter estimate and underestimated sampling variability due to the failure to account for random effects. Even under the conditions where the parameter estimates from the CLPM were accurate, the CI coverage from the CLPM could drop way below 90%. For example, the CI coverage rates for c` could be as low as 65% when the variance of *a* or *b* was large, although the relative biases for c` were very small (see Table 2).

Power/type I error rate. On average, the power to detect *ab* was higher for CLPM (see Table 2), however, this gain in power was not really meaningful considering the poor CI coverage rates from the CLPM. For the direct effect (*c*'), the CLPM resulted in inflated type I error rates, particularly for the conditions where the correlation between a_i and b_i was nonzero. When the variance of *a* or *b* was large, the type I error rate for *c*' could be as high as .35. In comparison, the type I error rates for *c*' from RE-CLPM were all close to 5%.

Model selection. As mentioned above, we examined the extent to which information criteria can lead to a selection of the correct model (i.e., RE-CLPM in this study). Table 3 shows the success rates of model selection with the four indices. The correlation between *a* and *b* did not have a strong influence on the success rates. Thus the results were collapsed over the different levels of correlations in Table 3. As expected, the accuracy of model selection by all four indices improved as sample size increased, given that there was less sampling error with a larger sample size (Preacher & Merkle, 2012). It also improved as the variance of *a* or *b* increased, because as the variance of *a* or *b* increased, the level of heteroscedasticity also increased which led to a greater degree of discrepancy between the RE-CLPM and CLPM. In addition, the accuracy of model selection by all four indices increased as the effect size increased because a larger effect size was associated with smaller measurement errors, due to how the data were generated in the study. In general, with a small sample size (n = 100), ABIC had highest success rates, making it the best index to use among the four indices examined. With a moderate or large sample size (n = 200 or 500), AIC performed much better than BIC, and slightly better than ABIC and AICc, making it the most desirable index to use.

[Table 3 is about here]

Study 3

The purpose of study 3 is to examine whether the RE-CLPM is robust to the violation of the normality assumption of random effects. We used the same data generation model as in study 2, except that *a* and *b* effects were generated from non-normal distributions. Similar to study 2, the sample size was varied at three levels (100, 200, 500), and the variance of *a* or *b* was varied at two levels (.04, .16). Since the effect size of *a* or *b* did not influence the result in the first two studies, we fixed the effect size of *a* or *b* at .3. As a preliminary investigation, we considered

only two levels of correlation between *a* and *b:* 0 and .3. The same set of criteria were used to evaluate the performance of the RE-CLPM.

The non-normal distributions for *a* and *b* were generated in a way so that the skewness and excess kurtosis of the distributions were 2.25 and 7, respectively. This degree of non-normality was considered moderate in Finch, West, and MacKinnon (1997) or West, Finch, and Curran (1995). Specifically, the non-normal distributions were generated using Fleishman's (1978) power transformation approach. To ensure that correlation between *a* and *b* were comparable between the non-normal distribution conditions in this study and normal distribution conditions examined in study 2, we computed intermediate correlation effects following Vale and Maurelli (1983) and imposed Fleishman's (1978) power transformation on the data generated from the intermediate correlation effects.

Result

The result from study 3 is presented in Table 4. For comparison purposes, we also included in Table 4 the result for the corresponding normal distribution conditions obtained in study 2. Given that sample size did not affect the result, we only reported the result for $n = 200$.

As shown in Table 4, the RE-CLPM was robust to moderate deviation from normality for the *a* and *b* effects. The parameter estimates were unbiased. The CI coverage rates were slightly decreased, the power rates for the indirect effect were slightly overestimated, and the type I error rates for the direct effect were slightly inflated by non-normality, however, they were all very close to those obtained under the normality assumption.

[Table 4 is about here]

Empirical Example

In this section, an empirical example is provided to demonstrate the proposed method. The original dataset was from Little (2013, p.303) which was used to test the hypothesis that substance use (M) mediated the effect of family conflict (X) on victimization (Y). Specifically, family conflict is expected to increase substance use, which in turn increases victimization. In this illustration, we used three waves of data collected 6 months apart from 1132 middle school students. Note that family conflict, substance use, and victimization were latent constructs in the original analysis. Given that the RE-CLPM cannot be applied to latent variables yet (discussed later), we created scale scores for the constructs. The scale scores are the weighted sum of the manifest indicators using the factor loadings reported in the book chapter as the weights.

We first fit the traditional CLPM which is similar to that shown in Figure 1 to the data. We then freed the autoregressive and cross-lagged effects to be random. Note that the autoregressive effects associated with the mediator were fixed in the model because freeing them to be random caused convergence problems. In addition, we specified a less restricted model than the RE-CLPM used in the simulation studies by allowing all of the random effects to vary across time. We also explored the covariance structure of the random effects by allowing any significant covariances to be freely estimated.

The results from the RE-CLPM and CLPM are summarized in Table 5. Consistent with the simulation studies, we focus on reporting the result for the mean indirect and direct effects. The complete outputs for the fitted models are available upon request. As shown in Table 5, the estimated effects from the CLPM and RE-CLPM were consistent in terms of their directions. However, there were two noteworthy differences between results from CLPM and RE-CLPM. The mean effect $M_2 \rightarrow Y_3$ was significant in the CLPM but not in the RE-CLPM. Consequently, the indirect effect $X_1 \rightarrow M_2 \rightarrow Y_3$ was significant in the CLPM but not in the RE-CLPM. Given that

all information criteria were smaller for the RE-CLPM (see Table 5), and the variances of several random effects [e.g., a_1 ($X_1 \rightarrow M_2$) and the autoregressive effects associated with the X and Y variables] were significant, we believe that the result from the RE-CLPM is more trustable.

[Table 5 is about here]

Conclusions and Discussions

The current article tackles an overlooked limitation of the CLPM for longitudinal mediation analysis: the CLPM does not accommodate random or individually varying direct and indirect effects. It showed that random effects would give rise to heteroscedasticity in the data. Ignoring this heteroscedasticity can potentially result in biased estimates for the indirect effects, unacceptably low CI coverage rates, and highly inflated type I error rates for the target effects.

We propose the RE-CLPM to solve the problem. Our simulation studies show that the RE-CLPM exhibited desirable properties. It yielded accurate parameter estimates and reliable statistical inferences for both mean indirect and direct effects when random effects exist in the population. When random effects do not exist, the RE-CLPM did not produce harmful result. It also showed robustness to moderate deviation from normality assumption of random effects. However, these desirable properties do not necessarily mean that the RE-CLPM is categorically superior to the CLPM. With only fixed effects, there is a gain in statistical efficiency and power by using the CLPM. In addition, researchers can assess the data-model fit via model fit indices for the CLPM. Furthermore, even with random effects, if the degree of heteroscedasticity introduced by the random effects was low, there may not be meaningful differences in the results between the CLPM and RE-CLPM.

So when the RE-CLPM should be adopted? In practice, it is usually difficult to know a priori whether there are true random effects and the degree of heteroscedasticity introduced by

the random effects. One may examine the residual plot for each outcome variable at each time point. However, these residual plots cannot reflect heteroscedasticity in the residual covariance structure for multiple outcome variables jointly. A practical strategy is to fit both the RE-CLPM and the CLPM to the data, and then compare the results from the two models. One should start with the CLPM to ensure that the CLPM fits the data first, and then free the effects in the model to be random. Two things should be examined when selecting between the models. First are information criteria. The current study suggests that if the sample size is small, ABIC is the best choice. Otherwise, AIC should be used. Researchers should also examine whether there are meaningful differences between the two sets of results. For example, whether there is a significant variance or covariance for any of the random effects specified in the model, and whether the two models lead to diverse conclusions regarding the key effects. If there is no meaningful difference, it is reasonable to report the CLPM result.

There are a few limitations of the simulation studies that worth mentioning. First, the current study only examined RE-CLPMs with 3 waves of data in the simulation studies. We cannot informedly extrapolate whether the result will apply to datasets with more waves of data. For example, the simulation studies showed that it is desirable to have $n \geq 200$ for the RE-CLPM with 3 waves of data. The sample size requirement would be presumably larger with more than 3 waves of data.

Second, the CI for the indirect effect *ab* was computed using the delta method. There are better ways to establish the CI. For example, the CI may be established using a Monte Carlo simulation approach (Preacher & Selig, 2012) or bootstrapping approach (MacKinnon, Lockwood, & Williams, 2004). Kenny, Korchmaros, and Bolger (2003) did not find meaningful differences between the Monte Carlo approach and the delta method for the MLM. It is

interesting to investigate whether the Monte Carlo simulation and bootstrapping approaches would provide better CIs for the RE-CLPM, especially in the case where *a* or *b* does not follow a normal distribution.

The proposed RE-CLPM is also limited in several ways. First, it can be only applied to manifest variables so far. As a result, presence of measurement errors can potentially bias the results. To correctly account for measurement errors, methods to accommodate latent variables in the RE-CLPM need to be developed. A possible direction towards solving the problem is to use Bayesian structural equation modeling given the computational advantages of Bayesian estimation method in handling complex models (Zhang, Wang, & Bergeman, 2016). Second, the RE-CLPM also has limited use to categorical variables. Muthén, Muthén, and Asparouhov (2015) noted that random coefficient models may not be identified with categorical data.

Third, like the traditional CLPM, the RE-CLPM assumes that the lags between observations are constant across individuals, and the effects detected using the RE-CLPM are limited to the lags. Deboeck and Preacher (2016) have termed the traditional CLPM *discrete time model*, and proposed a so-called *continuous time model* (CTM) to overcome the limitations of the CLPM. The CTM allows random lags (i.e., individually varying lags), and produces autoregressive and cross-lagged effects estimates that are independent of lags. Although the autoregressive and cross-lagged effects obtained from the CTM have different interpretations than those obtained from the CLPM, they are mathematically related. Based on the effects estimated in the CTM model, one may compute the corresponding effects in the discrete time model of any given lag through nonlinear transformations. Deboeck and Preacher (2016) only considered the CTM with fixed effects, however, they pointed out that it is possible to accommodate random effects in the CTM. It would be interesting to compare the continuous

time and discrete time models when random effects are included in both models. The only downside of the CTM is that it is less flexible than the CLPM in modeling time varying effects.

A few other extensions of the RE-CLPM deserve future investigation. First, the RE-CLPM can be extended to incorporate potential covariates for any of the random effects (e.g., indirect effects). If a covariate effect is significant, the random coefficient will vary as a function of the covariate, and the covariate is said to moderate the effect (e.g., moderated indirect effects). This is analogous to a moderated mediation analysis with the traditional CLPM (Hayes, 2013). However, the RE-CLPM may be able to produce more accurate standard error estimates for covariate effects than the CLPM. In addition, with the RE-CLPM, researchers can examine the proportion of variance in the random effect explained by the covariate, which may serve as an effect size measure for the moderation effect. Future research needs to be conducted to examine to what extent the RE-CLPM can correctly recover moderation effects.

Second, given that missing data are ubiquitous in longitudinal studies, it is worth noting that the RE-CLPM may be fitted with the presence of missing data by using the full information maximum likelihood estimation method (Yuan, & Bentler, 2000; Yuan, Yang-Wallentin, & Bentler, 2012; Asparouhov, & Muthén, 2003). Future research is warranted to examine whether the RE-CLPM will perform as well with missing data. Another principle missing data method is multiple imputation (MI). However, researchers should be cautious about using MI in this case because the commonly used imputation models (e.g., regression models) do not take random effects or heteroscedasticity into account, which can potentially introduce bias into the analysis result.

Third, the article only addresses heteroscedastic residuals induced by random effects, thus heteroscedasticity will disappear after accounting for the random effects. It is possible that

different causes of heteroscedastic residuals may coexist in one model (e.g., non-normality of outcome variables), with random effects being only one of the causes. In this case, we suspect that modeling random effects may mitigate but not completely solve the heteroscedasticity problem. To account for heteroscedastic residuals that are not due to random effects while modeling random effects at the same time in a single level model would be an interesting yet challenging venue for future research. It is also likely that heteroscedastic residuals, if they exist, are not induced by random effects at all. In this scenario, it would be interesting to examine whether the presence of heteroscedastic residuals will lead to wrong detection of random effects.

Finally, the current study only incorporated random effects in the cross-lagged panel model. A recent study conducted by Hamaker, Kuiper, and Grasman (2015) pointed out that it is also necessary to account for random intercepts in CLPMs if constructs under study contain trait like or time invariant components. Failure to do so may result in spurious effects, misleading conclusions regarding the size and direction of a cross-lagged effect. They suggested that random intercepts can be included as latent factors underlying corresponding repeated measures with loadings fixed at 1s. The random intercepts extract the time-invariant components for each individual out of the repeated measures. The autoregressive and cross-lagged effects can be then tested on the disturbances on the repeated measures after accounting for the random intercepts. The cross-lagged effects detected in this way would reflect the mediation process controlling for the time invariant components. Given that the disturbances are modeled as latent variables in the random intercept model proposed by Hamaker et al. (2015), there is no good way yet to allow random autoregressive and cross lagged effects in the random intercept model. Future research is warranted to develop ways to accommodate both random intercepts and random effects in the longitudinal mediation model.

References

Akaike H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov B. N., Csaki F. (Eds), *Second international symposium on information theory* (pp. 267-281), Akademiai Kiado: Budapest.

Asparouhov, T., & Muthén, B. (2003). *Full-information maximum-likelihood estimation of general two-level latent variable models with missing data*. Technical Report. Muthén & Muthén, Los Angeles.

Asparouhov, T., & Muthén, B. (2012). *Saddle Points*. Technical Report. Muthén & Muthén, Los Angeles.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological methods, 11*(2), 142 - 163. doi: 10.1037/1082-989X.11.2.142

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Burnham, K. P., Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in Model Selection, *Sociological Methods & Research, 33: 261–304,*doi*:*10.1177/0049124104268644

Cole, D., & Maxwell, S. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112(4)*, 558-577. doi: 10.1037/0021-843X.112.4.558

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.

Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 61-75.

Fall, E. C. (2011). Modeling Heterogeneity in Indirect Effects: Multilevel Structural Equation *Modeling Strategies* (Doctoral dissertation). Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/9787/Fall_ku_0099D_11774_DATA_1.pdf?sequence=1.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling, 4*(2), 87–107.

Fleishman, A. I. (1978). A method for simulating nonnormal distributions. *Psychometrika, 43*(4), 521–532.

Goldstein, H. (2003). *Multilevel Statistical Models* (3rd ed.). London: Edward Arnold.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances unanswered questions, future directions* (pp. 243–259). Washington, DC: American Psychological Association.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological methods, 20*(1), 102 -116.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*: Guilford Press.

Hildreth, C. and Houck, J.P. (1968). Some estimators for a linear model with random coefficientsrandom effects. *JASA, 63,* 584–595.

Hurvich, C. M. & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, 76,* 297-307.

Jamshidian, M., & Jennrich, R. I. (1997). Acceleration of the EM Algorithm by using Quasi-Newton Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59*(3), 569-587.

Johnston, J. (1984), *Econometric Methods*. New York: McGraw-Hill.

Kendall, M.G. & Stuart, A. (1969). *The advanced theory of statistics.* Charles Griffin, London.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. Psychological methods, 8(2), 115 – 128, doi: 10.1037/1082-989X.8.2.115.

Lee, S.-Y., & Poon, W.-Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica 8*, 749-766.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York, NY: Guilford.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39(1),* 99–128.

Maxwell, S., & Cole, D. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological methods, 12*(1), 23-44. doi: 10.1037/1082-989X.12.1.23

Muthén, B. & Muthén, L., & Asparouhov, T. (2015). *Random coefficient regression.* Retrieved from https://www.statmodel.com/download/Random_coefficient_regression.pdf

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*(3), 431-462.

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling, 4*, 599–620.

Muthén, L.K. and Muthén, B.O. (1998-2015). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Preacher, K. J. (2015). Advances in mediation analysis: a survey and synthesis of new developments. *Annual Review of Psychology, 66*, 825-852.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological methods*, 17(1), 1.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*(2), 77-98.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*(2), 161-182.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods, 15*(3), 209.

Raftery, A. (1995). Bayesian model selection in ocial research. *Sociological methodology, 25*, 111–196.

Selig, J. P., Preacher, K. J., & Little, T. D. (2012). Modeling time-dependent association in longitudinal data: A lag as moderator approach. *Multivariate Behavioral Research, 47*(5), 697-716.

Selig, J., & Preacher, K. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development, 6(2),* 144-164.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 290–312). San Francisco, CA: Jossey-Bass.

Swamy, P.A.V.B. (1971), *Statistical Inference in Random Coefficient Regression Models*. New York: Springer.

Vale, C. D. and Maurelli, V.A. (1983) Simulating multivariate nonormal distributions, *Psychometrika, 48,* 465-471.

Weisberg, S. (2014). *Applied Linear Regression*, fourth edition, Hoboken NJ: John Wiley.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equations with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues and applications* (pp. 56–75). Newbury Park, CA: Sage.

Wu, W., Selig, J. P., & Little, D. (2013). Longitudinal data analysis. *The Oxford Handbook of Quantitative Methods, Vol. 2: Statistical Analysis*, 387 - 410.

Yang, C. C., (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50,* 1090-1104.

Yuan, K. H., & Bentler, P. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology, 30(1),* 165-200.

Yuan, K. H., Yang-Wallentin, F., & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. Sociological Methods & Research, 41(4), 598-629.

Zhang, Q., Wang, L. J. & Bergeman, C. (2016). Multilevel autoregressive mediation models: Specification, estimation, and applications. Paper presented at the Modern Modeling Methods (M3) Conference, Storrs, CT.

Table 1
The Result from Simulation Study 1 (n = 200, effect size of a/b=0.3)

Effect	% Relative Bias		CI coverage		Power/Type I error[a]	
	CLPM	**RE-CLPM**	**CLPM**	**RE-CLPM**	**CLPM**	**RE-CLPM**
ab	0.36	0.80	0.95	0.98	1.00	0.98
a	0.21	0.20	0.93	0.94	1.00	1.00
c	-0.13	-0.13	0.95	0.96	0.05	0.04

Note that the relative biases bigger than 10%, the CI coverage rates lower than .90, and the type I error rates larger than .1 were highlighted.
[a] for a and ab, the values in this column are power rates. For c', the values are type I error rates.
RE-CLPM = random effects cross-lagged panel model; CLPM = cross-lagged panel model;

Table 2
The Result from Simulation Study 2 (effect size of a/b = 0.3, n = 200)

Effect	σ_a^2	$r_{a,b}$	Relative bias		CI Coverage		Power/Type I error rate[a]	
			CLPM	RE-CLPM	CLPM	RE-CLPM	CLPM	RE-CLPM
ab	0.16	0.6	**-53.55**	-1.45	**0.05**	0.95	1.00	0.98
		0.3	**-38.02**	0.20	**0.30**	0.94	1.00	0.88
		0	-8.78	0.56	**0.83**	0.95	1.00	0.53
		-0.3	**90.24**	-5.34	**0.52**	0.95	1.00	0.14
		-0.6	**1401.67**	0.05	**<.01**	0.95	1.00	0.05
	0.04	0.6	**-21.40**	-3.39	**0.65**	0.97	1.00	0.92
		0.3	**-13.62**	-3.07	**0.82**	0.98	1.00	0.87
		0	-2.84	1.49	0.90	0.97	1.00	0.80
		-0.3	**11.34**	3.30	0.92	0.97	1.00	0.70
		-0.6	**31.04**	9.00	**0.85**	0.97	1.00	0.57
a	0.16	0.6	-7.69	-1.47	**0.86**	0.94	1.00	0.99
		0.3	-5.21	1.21	**0.86**	0.95	1.00	1.00
		0	-6.85	-0.56	**0.86**	0.95	1.00	1.00
		-0.3	-6.41	-0.22	**0.86**	0.93	1.00	1.00
		-0.6	-6.56	-0.88	**0.85**	0.94	1.00	0.99
	0.04	0.6	-1.30	0.99	0.92	0.94	1.00	0.99
		0.3	-1.80	0.12	0.93	0.95	1.00	1.00
		0	-2.03	-0.19	0.93	0.96	1.00	1.00
		-0.3	-1.74	0.14	0.94	0.96	1.00	1.00
		-0.6	-1.53	0.46	0.94	0.95	1.00	0.99
c'	0.16	0.6	8.43	0.07	**0.69**	0.95	**0.31**	0.05
		0.3	4.11	-0.71	**0.87**	0.96	**0.13**	0.04
		0	-0.55	0.09	0.93	0.95	0.07	0.05
		-0.3	-4.71	0.04	**0.85**	0.95	**0.15**	0.05
		-0.6	-9.55	1.01	**0.65**	0.95	**0.35**	0.05
	0.04	0.6	2.29	0.65	0.92	0.96	**0.08**	0.04
		0.3	1.08	-0.03	0.91	0.95	**0.09**	0.05
		0	-0.33	0.24	0.93	0.95	**0.08**	0.05
		-0.3	-0.93	0.11	0.90	0.95	**0.10**	0.05
		-0.6	-2.32	0.32	0.90	0.96	**0.10**	0.04

Note that relative biases bigger than 10% and CI coverage rates lower than .90 were highlighted. Type I error rates larger than .07 were highlighted. [a] for *a* and *ab*, the values in this column are power rates. For *c*', the values are type I error rates. RE-CLPM = random effects cross-lagged panel model; CLPM = cross-lagged panel model; σ_a^2 represents the population variance of *a* effect**,** and $r_{a,b}$ represents the population correlation between *a* and *b*.

Table 3
The success rates of model selection in study 2

n	σ_a^2	***Effect size of a/b***	**AIC**	**BIC**	**ABIC**	**AICc**
100	0.04	0.3	0.19	0.00	**0.35**	0.02
		0.6	0.69	0.12	**0.84**	0.28
	0.16	0.3	0.88	0.33	**0.95**	0.55
		0.6	1.00	0.96	**1.00**	0.99
200	0.04	0.3	**0.36**	0.01	0.32	0.20
		0.6	**0.96**	0.42	0.96	0.91
	0.16	0.3	**1.00**	0.80	0.99	0.99
		0.6	**1.00**	1.00	1.00	1.00
500	0.04	0.3	**0.87**	0.08	0.66	0.83
		0.6	**1.00**	0.98	1.00	1.00
	0.16	0.3	**1.00**	1.00	1.00	1.00
		0.6	**1.00**	1.00	1.00	1.00

Note that the result was collapsed over different levels of correlation between *a* and *b*. The highest success rate for each condition was highlighted. σ_a^2 represents the population variance of *a* effect.

Table 4
The Result from Simulation Study 3 (effect size of a/b = 0.3, n = 200)

Effect	Conditions		Relative bias		CI coverage		Power/Type I error[a]	
	σ_a^2	$r_{a,b}$	normal	Non- normal	normal	Non-normal	normal	Non-normal
ab	0.16	0.3	0.20	-0.05	0.94	0.92	0.88	0.91
		0	0.56	-1.58	0.95	0.94	0.53	0.57
	0.04	0.3	-3.07	-1.10	0.98	0.98	0.87	0.91
		0	1.49	-1.13	0.97	0.98	0.80	0.80
a	0.16	0.3	1.21	-1.63	0.95	0.93	1.00	1.00
		0	-0.56	-1.15	0.95	0.95	1.00	1.00
	0.04	0.3	0.12	-0.34	0.95	0.95	1.00	1.00
		0	-0.19	0.29	0.96	0.95	1.00	1.00
c`	0.16	0.3	-0.71	-0.58	0.96	0.94	0.04	0.06
		0	0.09	0.25	0.95	0.96	0.05	0.04
	0.04	0.3	-0.03	-0.09	0.95	0.96	0.05	0.04
		0	0.24	0.10	0.95	0.95	0.05	0.05

[a] for *a* and *ab*, the values in this column are power rates. For *c*', the values are type I error rates.

Table 5
The Result for the empirical example

Effect	CLPM	RE-CLPM
a_1 ($X_1 \rightarrow M_2$)	.125* (.017)	.081* (.018)
a_2 ($X_2 \rightarrow M_3$)	-.031 (.013)	.002 (.024)
b_1 ($M_1 \rightarrow Y_2$)	.099* (.032)	.101* (.034)
b_2 ($M_2 \rightarrow Y_3$)	**.078* (.036)**	**-.016 (.062)**
c'($X_1 \rightarrow Y_3$)	.093* (.023)	.080* (.025)
a_1b_2 ($X_1 \rightarrow M_2 \rightarrow Y_3$)	**.010* (.005)**	**<.001 (.004)**
AIC	37633.97	36265.94
BIC	37840.27	36537.65
ABIC	37710.04	36366.14
AICc	37636.83	36270.28

Note. The numbers in the parentheses are standard residuals. * Significant at $p < .05$.
RE-CLPM = random effects cross-lagged panel model; CLPM = cross-lagged panel model;
X = family conflict; M = substance use; Y = victimization

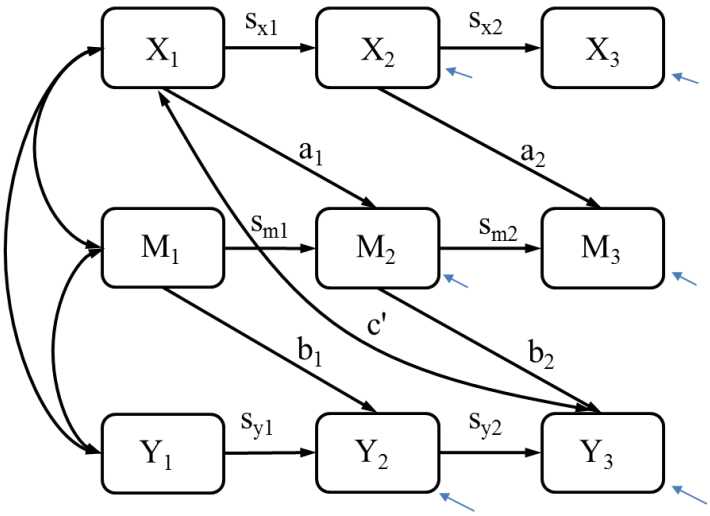

Figure 1. A fixed effect cross-lagged panel model with three waves of measurement.

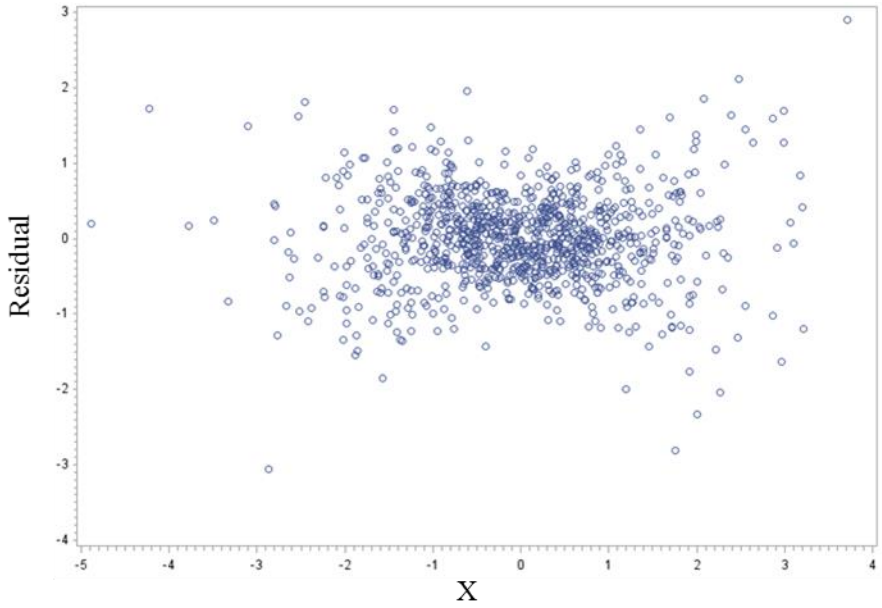

Figure 2. A plot of residuals on Y against X when a fixed effect regression model is fit to the data from a random effect regression model.

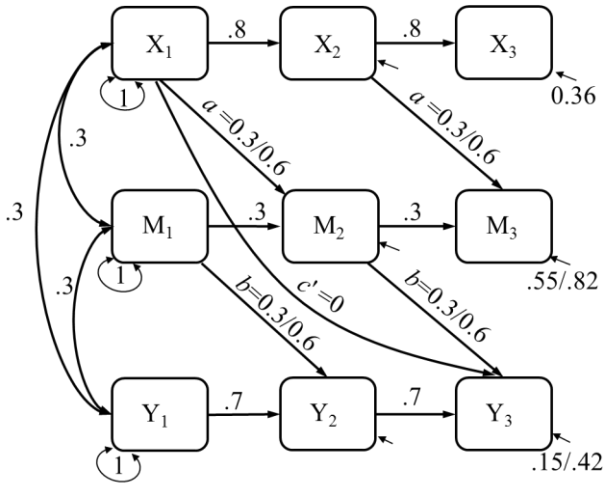

Figure 3. The data generation model in simulation study 1.

Appendix A

Example Mplus syntax for fitting the cross-lagged panel model in simulation study 1

```
DATA:
file is 'C:\mydata.dat';
VARIABLE:
names are x1 x2 x3 m1 m2 m3 y1 y2 y3 ;
usevariables are x1 x2 x3 m1 m2 m3 y1 y2 y3;
MODEL:
x2 on x1 (stx);
x3 on x2 (stx);
m2 on m1 (stm);
m3 on m2(stm);
m2 on x1 (am);
m3 on x2 (am);
y2 on y1(sty);
y3 on y2(sty);
y2 on m1 (bm);
y3 on m2 (bm);
y3 on x1 (c);
MODEL CONSTRAINT:
new(mab);
mab = am*bm;
```

Appendix B

Example Mplus syntax for fitting the random effects cross-lagged panel models

```
DATA:
file is 'C:\mydata.dat';
VARIABLE:
names are x1 x2 x3  m1 m2 m3  y1 y2 y3 ;
usevariables are x1 x2 x3  m1 m2 m3  y1 y2 y3;
ANALYSIS:
type = random;
Model:
sx1|x2 on x1;
sx1|x3 on x2;
sm1|m2 on m1;
sm1|m3 on m2;
sy1|y2 on y1;
sy1|y3 on y2;
b1| y2 on m1;
b1| y3 on m2;
a1| m2 on x1;
a1| m3 on x2;
[a1] (am);
[b1] (bm);
 a1 (av);
 b1 (bv);
a1 with b1 (covab);
c1|y3 on x1;
[c1] (cm);
c1 (cv);
MODEL CONSTRAINT:
new(mab varab);
mab = am*bm + covab;
varab =bm*bm*av+am*am*bv+bv*av+2*am*bm*covab + covab*covab;
```