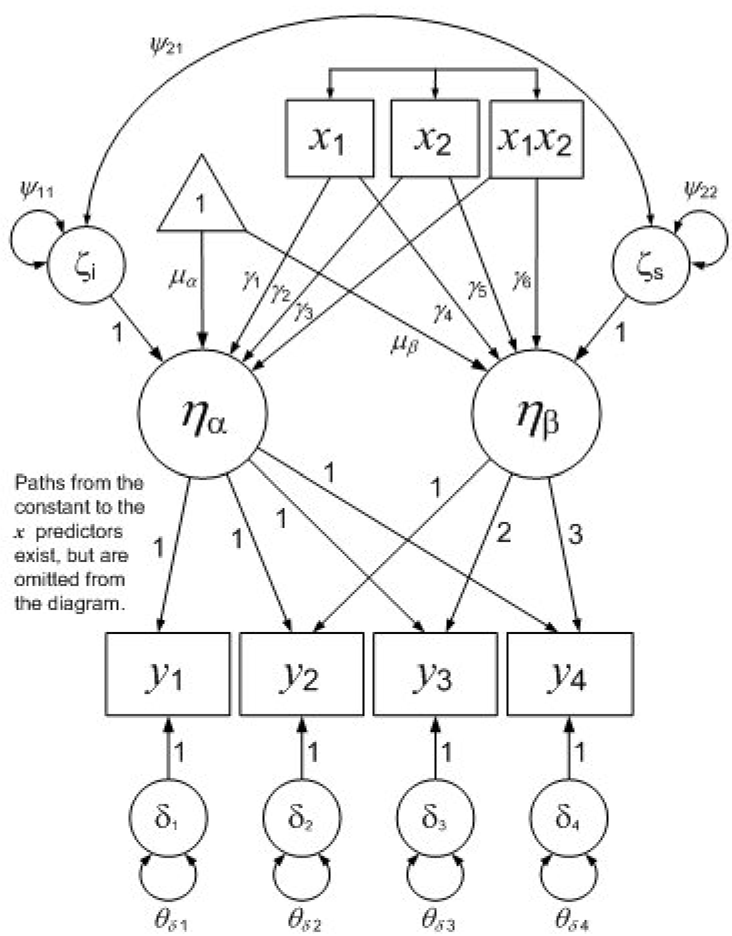
Simple Intercepts, Simple Slopes, and Regions of Significance in LCA 3-Way Interactions

Kristopher J. Preacher, **Patrick J. Curran**, and **Daniel J. Bauer**
University of North Carolina at Chapel Hill

This web page calculates simple intercepts, simple slopes, and the region of significance to facilitate the testing and probing of three-way interactions estimated in latent curve analysis (LCA) models. In LCA, repeated measures of a variable *y* are modeled as functions of latent factors representing *aspects of change* or *latent curves*, typically an intercept factor and one or more slope factors. We use the standard structural equation modeling (SEM) notation to define equations, and we assume that the user is knowledgeable both in the general SEM and in the testing, probing, and interpretation of interactions in multiple linear regression (e.g., Aiken & West, 1991). The following material is intended to facilitate the calculation of the methods presented in Curran, Bauer, and Willoughby (2004), and we recommend consulting this paper for further details, as well as our **companion web page** on 2-way interactions in LCA.



__The unconditional LCA__. Let y_{it} represent repeated measures of variable *y* for *i* = 1, 2, ..., *N* individuals at *t* = 1, 2, ..., *T* occasions (all of these techniques generalize to times varying over *i*, but for simplicity we assume that all individuals are measured at the same occasions; see Curran et al., 2004, p. 222 for details). In matrix notation, the general form of an LCA *measurement model* is

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad (1)$$

where $\mathbf{y}$ is a *T* x 1 vector of repeated measures for individual *i*, $\boldsymbol{\Lambda}$ is a *T* x *k* matrix of factor loadings (where *k* is the number of latent curve factors, here 2 to define a linear trajectory), $\boldsymbol{\eta}$ is a *k* x 1 vector of latent curve factors, and $\boldsymbol{\varepsilon}$ is a *T* x 1 vector of time-specific residuals.

In most applications of LCA, the elements of $\boldsymbol{\Lambda}$ are constrained to reflect linear growth, e.g.:

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \tag{2}$$

The first column of $\mathbf{\Lambda}$ contains loadings on the intercept factor. In LCA models, *time* is not explicitly included as a variable, but rather is incorporated in the model as elements of the second column of $\mathbf{\Lambda}$. The variance of the slope factor, in turn, represents individual differences in the slope of the latent trajectory. For more detail see Curran, et al. (2004).

An expression for the latent curve factors is:

$$\boldsymbol{\eta} = \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\zeta} \tag{3}$$

where $\boldsymbol{\mu}_{\boldsymbol{\eta}}$ is a *k* x 1 vector of latent curve factor means and $\boldsymbol{\zeta}$ is a *k* x 1 vector of residuals. Scalar expressions for elements in $\boldsymbol{\eta}$ with no exogenous predictors are:

$$\begin{aligned} \eta_{\alpha_i} &= \mu_\alpha + \zeta_{\alpha_i} \\ \eta_{\beta_i} &= \mu_\beta + \zeta_{\beta_i} \end{aligned} \tag{4}$$

A typical element of **y** is:

$$y_{it} = \left(\mu_\alpha + \mu_\beta \lambda_t\right) + \left(\zeta_{\alpha_i} + \zeta_{\beta_i} \lambda_t + \varepsilon_{it}\right) \tag{5}$$

<u>*The conditional LCA*</u>. One of the primary advantages of the LCA framework is that the factors representing intercept and slope can serve as endogenous (dependent) variables in other model equations. The figure above represents just such a *conditional* LCA model, in which the intercept and slope representing the latent trajectory of the repeated measures of *y* are modeled as dependent variables regressed on x_1, x_2, and the product of x_1 and x_2 to represent the interactive effect of two exogenous predictors on the latent curve factors. In such cases, the latent curve factors may be expressed as functions of the exogenous predictors x_1, x_2, and x_1x_2:

$$\boldsymbol{\eta} = \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \tag{6}$$

where $\boldsymbol{\Gamma}$ is a *k* x *p* matrix of regression parameters linking the *k* latent curve factors to the *p* exogenous predictors and **x** is a *p* x 1 vector of exogenous predictors x_1, x_2, and x_1x_2. Substituting Equation 6 into Equation 1 yields a reduced form equation for **y**:

$$\mathbf{y} = \mathbf{\Lambda}\left(\boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}\right) + \boldsymbol{\varepsilon} \tag{7}$$

$$\mathbf{y} = \left(\boldsymbol{\Lambda}\boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x} \right) + \left(\boldsymbol{\Lambda}\boldsymbol{\zeta} + \boldsymbol{\varepsilon} \right) \tag{8}$$

The first parenthetical term in Equation 8 is referred to as the *fixed component* and the second parenthetical term as the *random component*.

The prediction of $\boldsymbol{\eta}$ with time-invariant predictors $\mathbf{x}$ <u>represents a three-way interaction with *time*</u>. To see why this is so, consider the scalar expressions for elements in $\boldsymbol{\eta}$ when the exogenous predictors in $\mathbf{x}$ include x_1, x_2, and x_1x_2:

$$\begin{aligned} \eta_{\alpha_i} &= \mu_\alpha + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{1i} x_{2i} + \zeta_{\alpha_i} \\ \eta_{\beta_i} &= \mu_\beta + \gamma_4 x_{1i} + \gamma_5 x_{2i} + \gamma_6 x_{1i} x_{2i} + \zeta_{\beta_i} \end{aligned} \tag{9}$$

a typical element of $\mathbf{y}$ is:

$$\begin{aligned} y_{it} = \big(& \mu_\alpha + \mu_\beta \lambda_t + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{1i} x_{2i} \\ & + \gamma_4 \lambda_t x_{1i} + \gamma_5 \lambda_t x_{2i} + \gamma_6 \lambda_t x_{1i} x_{2i} \big) \\ & + \left(\zeta_{\alpha_i} + \zeta_{\beta_i} \lambda_t + \varepsilon_{it} \right) \end{aligned} \tag{10}$$

The fixed component of Equation 10 can be seen to contain an intercept term (i.e., μ_α), conditional main effects for *time* (i.e., μ_α), x_1 (i.e., γ_1), and x_2 (i.e., γ_2), conditional two-way interaction effects for x_1x_2 (i.e., γ_3), *time* and x_1 (i.e., γ_4), and *time* and x_2 (i.e., γ_5), and the three-way interaction of *time*, x_1, and x_2 (i.e., γ_6). Thus, the effect of *time* on y depends in part on the levels of x_1 and x_2. Given this, we can draw upon classical techniques for testing and plotting conditional effects in multiple regression. See our supporting material for probing interactions in standard regression **here**.

<u>*y_t on λ_t regressions at x_1 and x_2*</u>. The regression of y on *time* for specific values of x_1 and x_2 we term *y_t on λ_t regressions at x_1 and x_2*. Taking the expectation of Equation 10 and rearranging clarifies the role of x_1 and x_2 when they moderate the magnitude of the regression of y on *time*:

$$\begin{aligned} \hat{\mu}_{y_t|x_1x_2} = & \left(\hat{\mu}_\alpha + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \hat{\gamma}_3 x_1 x_2 \right) \\ & + \left(\hat{\mu}_\beta + \hat{\gamma}_4 x_1 + \hat{\gamma}_5 x_2 + \hat{\gamma}_6 x_1 x_2 \right) \lambda_t \end{aligned} \tag{11}$$

Note that Equation 11 has the form of a simple regression of y on λ_t where the first parenthetical term is the intercept of the simple regression and the second parenthetical term is the slope of the simple regression. We will refer to the first parenthetical term as the *simple intercept* and the second term as the *simple slope*. It can be seen that the simple intercept and simple slope are *compound coefficients* that result from the linear combination of other parameters. To further explicate this, we can re-express Equation 11 in terms of sample estimates such that

$$\hat{\mu}_{y_t|x_1x_2} = \hat{\omega}_0 + \hat{\omega}_1\lambda_t \quad (12)$$

where

$$\begin{aligned}\hat{\omega}_0 &= \hat{\mu}_\alpha + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \hat{\gamma}_3 x_1 x_2 \\ \hat{\omega}_1 &= \hat{\mu}_\beta + \hat{\gamma}_4 x_1 + \hat{\gamma}_5 x_2 + \hat{\gamma}_6 x_1 x_2\end{aligned} \quad (13)$$

These general expressions for the simple intercept ($\hat{\omega}_0$) and simple slope ($\hat{\omega}_1$) define the conditional regression of *y* on λ_t as a function of x_1 and x_2. Because these are sample estimates, we must compute standard errors to conduct inferential tests of these effects. The computation of these standard errors is one of the key purposes of our calculator.

The preceding material addresses the strategy of probing the three-way interaction of *time*, x_1, and x_2 such that conditional trajectories are examined for chosen values of x_1 and x_2. Alternatively, the effect of x_1 on *y* can be seen to depend on *time* and x_2, and the effect of x_2 on *y* can be seen to depend on *time* and x_1. Although tests of these effects can be highly informative, our primary interest is likely to be in conditional trajectories calculated at specific levels of x_1 and x_2. Consequently, we do not explore these alternative expressions here.

Summary. We are primarily interested in the estimation of the simple intercept ($\hat{\omega}_0$) and the simple slope ($\hat{\omega}_1$) of the conditional regression of outcome *y* on *time* as a function of the moderators x_1 and x_2. We have developed a calculator for y_t on λ_t regressions at x_1 and x_2 (see Curran, Bauer, & Willoughby, 2004 for details). We now turn to a brief description of the values that can be calculated using our table below.

The Region of Significance. The first available output is the region of significance of the simple slope describing the relation between the outcome *y* and *time* as a function of moderators x_1 and x_2. We do not provide the region of significance for the simple intercept given that this is rarely of interest in practice. The region of significance defines the specific values of the moderator at which the regression of *y* on *time* transitions from non-significance to significance. Although this region can be easily obtained when testing a two-way interaction, these are much more complex to compute for a three-way interaction (see Curran, Bauer, & Willoughby, 2004 for futher details). As is proposed in Curran et al. (2004, p. 227), the table allows for the calculation of the region of significance of the regression of *y* on *time* across values of x_1 at a particular value of x_2. This is a melding of the simple slopes and region approach. There are lower and upper bounds to the region. In many cases, the regression of *y* on *time* is significant at values of the moderator that are *less than* the lower bound and *greater than* the upper bound, and the regression is non-significant at values of the moderator falling *within* the region. However, there are some cases in which the opposite holds (e.g., the significant slopes fall within the region). Consequently, the output will explicitly denote how the region should be defined in terms of the significance and non-significance of the simple slopes. There are also instances in which the region cannot be mathematically obtained, and an error is displayed if this occurs for a given application. However, this region is calculated for a *specific conditional value of* x_2. The region can be re-calculated at several different conditional values of x_2 (e.g., ±1SD) to gain a better understanding of the structure of the three-way interaction. By default, the region is calculated at α= .05, but this may be changed by the user. Finally, the point estimates and standard errors of both the simple intercepts and the simple slopes are automatically calculated precisely at the lower and upper bounds of the region.

<u>*Simple Intercepts and Simple Slopes*</u>. The second available output is the calculation of point estimates and standard errors for up to two simple intercepts and simple slopes of the regression of *y* on *time* at specific levels of the moderators. In the table we refer to these specific values as *conditional values*. We can choose from a variety of potential conditional values of x_1 and x_2 for the computation of the simple intercepts and slopes. If x_1 or x_2 is dichotomous, we could select conditional values of 0 and 1 to compute the regression of *y* on *time* within group 0 and group 1. If x_1 or x_2 is continuous, we might select conditional values that are one standard deviation above the mean of x_1 or x_2 and one standard deviation below the mean of x_1 or x_2. Whatever the conditional values chosen, these specific values are entered in the sections labeled "Conditional Values of x_1" and "Conditional Values of x_2," and this will provide the corresponding simple slopes of *y* on *time* at those values of x_1 and x_2. The calculation of simple intercepts and slopes at specific values of the moderator is optional; the user may leave any or all of the conditional value fields blank.

<u>*Using the Calculator*</u>. Simple intercepts, simple slopes, and the region of significance can be obtained by following these five steps:

1. We strongly suggest writing out by hand the equation provided at the top of the table (this equation is essentially the same as Equation 11). This will significantly aid in keeping track of the necessary values to enter into the tables.
2. Enter the sample values for the path coefficients that correspond to the simple intercept and simple slope of interest. For interpretational purposes, it is essential that any extra continuous covariates included in the model be centered prior to analysis and that a useful reference group be chosen for categorical covariates. This will ensure that any plots, if requested, will be accurate.
3. Enter the asymptotic variances of the required path coefficients under "Coefficient Variances"; note that these are the squared standard errors. Also enter the necessary asymptotic covariances under "Coefficient Covariances." All of these values can be obtained from the asymptotic covariance matrix of the regression parameters available in most standard SEM packages. More information on obtaining the ACOV matrix can be found **here**.
4. The region of significance and the simple intercept and simple slope calculated at the boundaries of this region are provided by default. At a minimum, the user must provide the sample regression parameters, and asymptotic variances and covariances. One available option is the selection of the probability value upon which to calculate the region. The default value is α= .05, but this can be changed to any appropriate value (e.g., .10 or .025).
5. If the calculation of additional *simple intercepts* and *simple slopes* is desired for specific conditional values of x_1 and x_2, enter the conditional values of x_1 and x_2 at which to estimate these values. If x_1 or x_2 is dichotomous and was originally coded 0 and 1 to denote group membership, enter 0 and 1 for the first and second conditional values, and leave the third cell blank. If x_1 or x_2 is continuous, any two conditional values may be selected as described above (results for more than two conditional values may be obtained by re-entering additional conditional values and recalculating). If these conditional value fields are all left blank, no simple intercepts or simple slopes will be provided.

Once all of the necessary information is entered into the table, click "Calculate." The status box will identify any errors that might have been encountered. If no errors are found, the results will be presented in the output window. Although the results in the output window cannot be saved, the contents can be copied and pasted into any word processor for printing.

<u>*R code for creating plots*</u>. Below the output window are two additional windows. If conditional values of λ_t (Points to Plot) and x_1, as well as at least one conditional value of x_2, are entered, clicking on

"Calculate" will also generate R code for producing a plot of the interaction between *time* and x_1 at the lowest value of x_2 (R is a statistical computing language). This R code can be submitted to a remote Rweb server by clicking on "Submit above to Rweb." A new window will open containing a plot of the interaction effect. The user may make any desired changes to the generated code before submitting, but changes are not necessary to obtain a basic plot. Indeed, this window can be used as an all-purpose interface for R.

Assuming enough information is entered into the interactive table, the second output window below the table will include R syntax for generating confidence bands. The user is expected to supply lower and upper values for the moderator x_1 (-10 and +10 by default). As above, this R code can be submitted to a remote Rweb server by clicking on "Submit above to Rweb." A new window will open containing a plot of confidence bands.

References

Curran, P. J., Bauer, D. J, & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods*, *9*(2), 220-237.