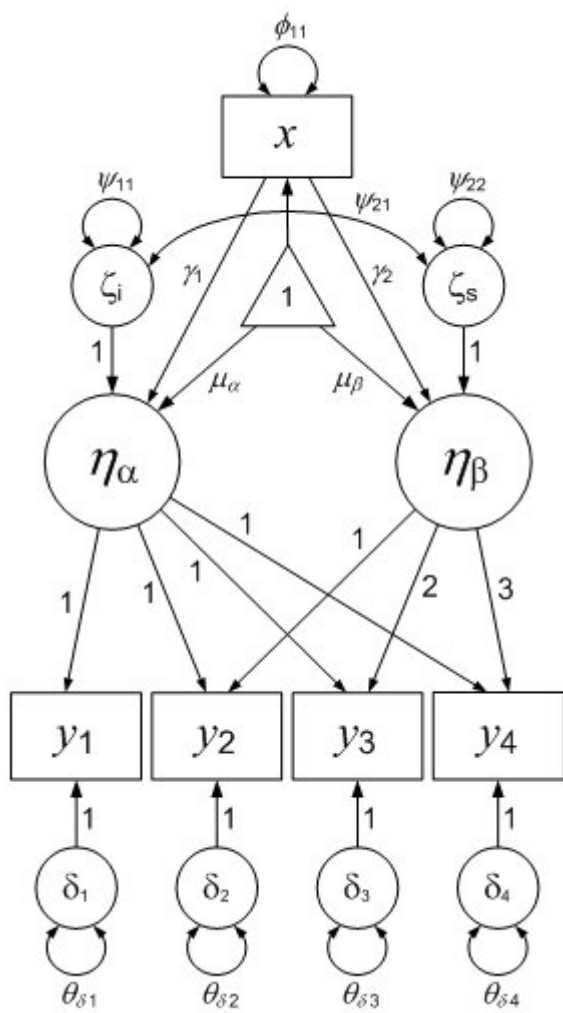


Simple Intercepts, Simple Slopes, and Regions of Significance in LCA 2-Way Interactions

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 two-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.



The unconditional LCA. Let y_{it} represent repeated measures of variable y for $i = 1, 2, \dots, N$ individuals at $t = 1, 2, \dots, T$ 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

$$\mathbf{y} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{1}$$

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

In most applications of LCA, the elements of $\mathbf{\Lambda}$ 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 Λ 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 Λ . The variance of the slope factor represents individual differences in the slope of the latent trajectory.

An expression for the latent curve factors is:

$$\boldsymbol{\eta} = \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\zeta} \quad (3)$$

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

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

A typical element of \mathbf{y} is:

$$y_{it} = (\mu_{\alpha} + \mu_{\beta} \lambda_{it}) + (\zeta_{\alpha_i} + \zeta_{\beta_i} \lambda_{it} + \varepsilon_{it}) \quad (5)$$

The conditional LCA. 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 . In such cases, the latent curve factors may be expressed as functions of the exogenous predictor x :

$$\boldsymbol{\eta} = \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\Gamma} \mathbf{x} + \boldsymbol{\zeta} \quad (6)$$

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

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

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

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} represents an interaction with time. To see why this is so, consider the scalar expressions for elements in $\boldsymbol{\eta}$ when there is only one exogenous predictor x :

$$\begin{aligned}\eta_{\alpha_i} &= \mu_{\alpha} + \gamma_1 x_i + \zeta_{\alpha_i} \\ \eta_{\beta_i} &= \mu_{\beta} + \gamma_2 x_i + \zeta_{\beta_i}\end{aligned}\tag{9}$$

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

$$y_{it} = \left(\mu_{\alpha} + \mu_{\beta} \lambda_t + \gamma_1 x_i + \gamma_2 \lambda_t x_i \right) + \left(\zeta_{\alpha_i} + \zeta_{\beta_i} \lambda_t + \varepsilon_{it} \right)\tag{10}$$

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_{β}) and the exogenous predictor x (i.e., γ_1), and the interaction of *time* and x (i.e., γ_2). Thus, the effect of *time* on y depends in part on the level of x . 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](#).

y_t on λ_t regressions at x_1 . The regression of y on *time* for specific values of x we term y_t on λ_t regressions at x_1 . Taking the expectation of Equation 10 and rearranging clarifies the role of x when x moderates the magnitude of the regression of y on *time*:

$$\hat{\mu}_{y_t|x} = \left(\hat{\mu}_{\alpha} + \hat{\gamma}_1 x \right) + \left(\hat{\mu}_{\beta} + \hat{\gamma}_2 x \right) \lambda_t\tag{11}$$

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

$$\hat{\mu}_{y_t|x} = \hat{\omega}_0 + \hat{\omega}_1 \lambda_t\tag{12}$$

where

$$\begin{aligned}\hat{\omega}_0 &= \hat{\mu}_{\alpha} + \hat{\gamma}_1 x \\ \hat{\omega}_1 &= \hat{\mu}_{\beta} + \hat{\gamma}_2 x\end{aligned}\tag{13}$$

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 . 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 calculators.

y_t on x_1 regressions at λ_t . Conversely, the effect of x on y can be seen to depend on *time*. This regression of y on x for specific values of *time* we term y_t on x_1 regressions at λ_t . Rearranging Equation 11 clarifies the role of *time* when *time* moderates the magnitude of the regression of y on x :

$$\hat{\mu}_{y_t|\lambda_t} = (\hat{\mu}_\alpha + \hat{\mu}_\beta \lambda_t) + (\hat{\gamma}_1 + \hat{\gamma}_2 \lambda_t) x \quad (14)$$

Note that Equation 14 has the form of a simple regression of y on x where the first parenthetical term is a simple intercept and the second parenthetical term is a simple slope. As with y_t on λ_t regressions at x_1 , y_t on x_1 regressions at λ_t may be expressed in terms of compound coefficients:

$$\hat{\mu}_{y_t|\lambda_t} = \hat{\omega}_0 + \hat{\omega}_1 x \quad (15)$$

where

$$\begin{aligned} \hat{\omega}_0 &= \hat{\mu}_\alpha + \hat{\mu}_\beta \lambda_t \\ \hat{\omega}_1 &= \hat{\gamma}_1 + \hat{\gamma}_2 \lambda_t \end{aligned} \quad (16)$$

The sample estimates of the simple intercept ($\hat{\omega}_0$) and simple slope ($\hat{\omega}_1$) define the conditional regression of y on x as a function of λ_t . Again, $\hat{\omega}_0$ and $\hat{\omega}_1$ are general expressions for simple intercepts and simple slopes for the regression of y on x conditional on λ_t and, despite similarity in notation, are not to be confused with the simple intercept and simple slope of the regression of y on λ_t conditional on x .

Summary. We are primarily interested in two cases: (1) 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 moderator x or (2) the estimation of the simple intercept and the simple slope of the conditional regression of outcome y on x as a function of *time*. When comparing the calculation of the simple intercepts and slopes across these two cases, it is clear that they share a common computational form, and this is why we have used the same notation to define the simple intercept and slope for each case. However, to simplify the use of our tables in practice, we have developed calculators separately for y_t on λ_t regressions at x_1 and y_t on x_1 regressions at λ_t , although the underlying analytics are all identical (see Curran, Bauer, & Willoughby, 2004 for details). We now turn to a brief description of the values that can be calculated using our tables below.

The Region of Significance. For y_t on λ_t regressions at x_1 , 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 a moderator x . 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 x at which the regression of y on *time* transitions from non-significance to significance. There are lower and upper bounds to the region. In many cases, the regression of y on *time* is significant at values of x 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 (i.e., the significant slopes fall within the region). Consequently, the output will explicitly note how the region should be interpreted 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. By default, the region is calculated at $\alpha = .05$, 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.

The region of significance is also available for y_t on x_1 regressions at λ_t , in which case the region defines the specific values of *time* at which the slope of the regression of y on x transitions from non-significance to significance.

Simple Intercepts and Simple Slopes. The second available output is the calculation of point estimates and standard errors for up to three simple intercepts and simple slopes of the regression of y on *time* at specific levels of x (or the regression of y on x at specific levels of *time*). In the tables we refer to these specific values as *conditional values*. There are a variety of potential conditional values of the moderator that may be chosen for the computation of the simple intercepts and slopes. If x is dichotomous (e.g., 0 or 1 to denote *gender*), we could select the first and second conditional values to be equal to 0 and 1 to compute the regression of y on *time* for males and for females (leaving the third conditional value blank). If the moderator is continuous, we might select values of x that are one standard deviation above the mean, equal to the mean, and one standard deviation below the mean. For y_t on x_1 regressions at λ_t it probably makes the most sense to choose values of λ_t actually used in the model, although this is not strictly required. Whatever the conditional values chosen, these specific values are entered in the section labeled "Conditional Values," and this will provide the corresponding simple intercepts and simple slopes of the regression of y on *time* at those specific values of x (or the regression of y on x at those specific values of *time*). The calculation of simple intercepts and slopes at specific values is optional; the user may leave any or all of the conditional value fields blank.

Points to Plot. Given the calculation of one or more simple slopes, it is common to plot these relations graphically to improve interpretability of effects. The final available output is the calculation of a lower and upper value associated with each of the simple slopes to aid in the graphing of these using any standard software package (e.g., Excel, SPSS, etc.). These are provided to simply aid in the graphing of effects; no inferential tests apply here. For the regression of y on x at specific levels of *time*, the user enters any two values of x in order to plot the regression line between y and x at specific values of λ_t . Although any pair of values can be used, we recommend using the lower and upper *specified* values of λ_t . However, many other specific values can be chosen that may be more appropriate for a particular research application. For the regression of y on *time* at specific levels of x , the user enters any two values of λ_t in order to plot the regression line between y and λ_t at specific values of x . Although any pair of values can be used, we recommend using either the lower and upper *observed* values of x , the lower and upper *possible* values of x , or one *sd* below and above the mean of x . However, again, many other specific values can be chosen.

Using the Calculators. Simple intercepts, simple slopes, and the region of significance can be obtained by following these seven steps:

1. Select whether you wish to investigate y_t on λ_t regressions at x_1 or y_t on x_1 regressions at λ_t as described above, and select the relevant table for calculations. Again, the underlying computations are identical; we present two separate tables to ease the use of these methods in practice.
2. We strongly suggest writing out by hand the equation provided at the top of each table for the given application at hand (these equations are essentially the same as Equations 11 and 14). This will significantly aid in keeping track of the necessary values to enter into the tables. We also suggest referring to the path diagram above.
3. Enter the sample values for the parameter estimates 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.

4. 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 parameter estimates available in most standard SEM packages. More information on obtaining the ACOV matrix can be found [here](#)
5. 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 parameter estimates and the asymptotic variances and covariances. One available option is the selection of the probability value upon which to calculate the region. The default value is $\alpha = .05$, but this can be changed to any appropriate value (e.g., .10 or .025).
6. If the calculation of additional simple intercepts and simple slopes is desired for specific conditional values of *time* or *x* other than the values defined as part of the region of significance, enter the conditional values at which to estimate these values. If y_t on λ_t regressions at x_1 are required and *x* is dichotomous (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* is continuous, up to three conditional values may be selected as described above (results for more than three conditional values may be obtained by re-entering additional conditional values and recalculating). Conversely, if y_t on x_1 regressions at λ_t are desired, up to three conditional values of λ_t may be selected as described above. If these conditional value fields are all left blank, no simple intercepts or simple slopes will be provided.
7. If the *points to plot* are desired, simply enter a lower and upper value of λ_t (for the first table) or *x* (for the second table) in the appropriate box. Any values can be used. If these fields are left blank, no points to plot will be provided.

Once all of the necessary information is entered into the table, simply 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. The results in the output window can be pasted into any word processor for printing.

R Code for Creating Simple Slopes Plot. Below the output window are two additional windows. If conditional values of *x* and λ_t are entered, clicking on "Calculate" will also generate R code for producing a plot of the interaction effect (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 either *x* or λ_t (-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.

R Code for Creating Confidence Bands / Regions of Significance Plot. Assuming enough information is entered into the interactive table, the second output window below the table will include R syntax for generating confidence bands, continuously plotted confidence intervals for simple slopes corresponding to all conditional values of the moderator. The x-axis of the resulting plot will represent conditional values of the moderator (*x*), and the y-axis represents values of the simple slope of *y* regressed on *time*.

If the moderator is dichotomous, only two values along the x-axis (corresponding to the codes used for grouping) would be interpretable. Therefore, in cases where the focal predictor is continuous and the moderator is dichotomous, we suggest using the lower table instead, treating *time* as the moderator for the confidence bands / regions of significance plot. Regardless of what variable is treated as the moderator, the user is expected to supply lower and upper values for the moderator (-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.

All material on these pages not otherwise credited is ©2003 by Kristopher J. Preacher.