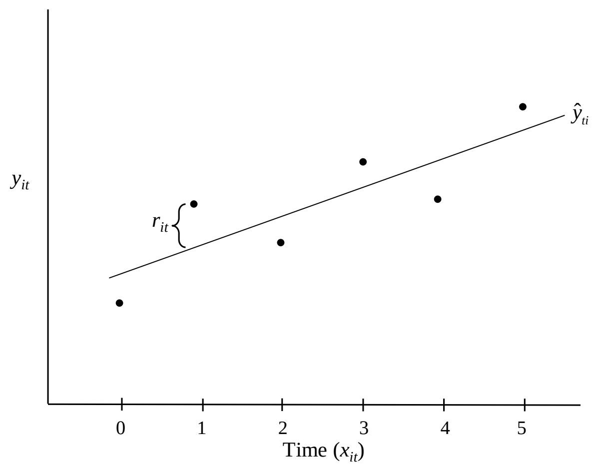
Latent Growth Curve Models

Meredith and Tisak (1984,1990) are generally credited with the inception of modern latent growth curve analysis by formalizing earlier work on exploratory factor analysis of growth (e.g., Baker, 1954; Rao, 1958; Tucker, 1958). They proposed latent variables with repeated measures as indicators, with and without special constraints on the loadings, in order to account for change over time. Latent growth curve models exploit the measurement model to estimate the variable of interest as some function of time.

Latent growth curve models provide information about the absolute level change at the individual and the sample average level. Conceptually, the basic building block is an individual regression where a score at each time point is regressed on time. The time variable can be represented by any age or date related variable, but is most commonly represented arbitrary codes beginning with 0 and ending with the final time point, T (e.g., $0, 1, 2, 3, \ldots T$).

The figure below illustrates a hypothetical slope for an individual case.



We could write this within-case regression equation as

$$y_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + r_{ti}$$

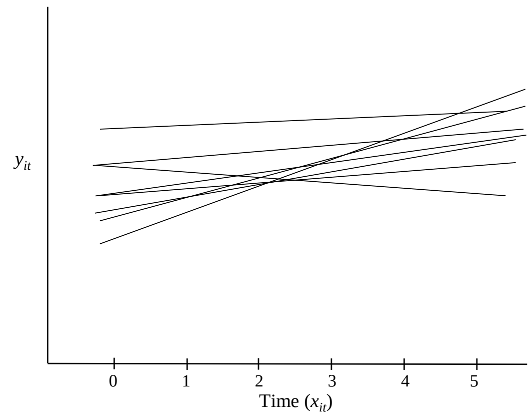
The intercept for this special equation represents the predicted value of *y* when the time variable equals 0, which would be the baseline score. The slope represents the change in *y* for each increment in time. You can also think about the slope as the average difference between time points.

If we were to imagine taking the intercept and slopes from a set of these individual regressions from all the cases in the data set, we could construct two other simple regression equations. Below are two such equations, one with the intercepts, β_0, serving as the outcome, and one with slopes, β_1, serving as the outcome. To start with a simple form, neither equation has any predictors.

$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$

Without any predictors, the intercept for a regression model is just the average. The intercept for the first equation, γ_{00}, represents the average of the intercepts, or the average baseline score. The intercept in the second equation, γ_{10}, represents the average of the slopes.

The variance of the residuals $Var(u_{0i})$ and $Var(u_{1i})$ represent variances in the baseline scores across individuals and variances in the slopes across individuals. These variances are often referred to as "random effects", whereas the average intercept and slope parameters are called "fixed effects". Because the model has both random and fixed effects, the term "mixed effects" is sometimes used. Growth curve models are generally portrayed as having a fundamental advantage over conventional repeated measures ANOVA, because individual differences in change or growth can be examined

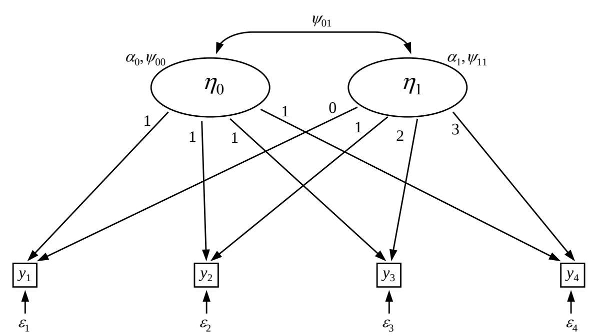


If we substitute the second set of equations into the first, we get a single "multilevel" equation.

$$y_{ti} = (\gamma_{00} + u_{0i}) + (\gamma_{10} + u_{1i})x_{ti} + r_{ti}$$
$$y_{ti} = \gamma_{00} + \gamma_{10}x_{ti} + u_{1i}x_{ti} + u_{0i} + r_{ti}$$

After the terms are rearranged, we can reconceptualize this equation a just a special regression model, with γ_{00} as the intercept, γ_{10} as the slope, and three types of residuals. One method of doing this analysis is to create separate records for each time point, sometimes called the "long data" format, which is the multilevel regression approach to growth curves.

For latent growth curves, the equivalent model can be estimated in SEM, but it requires exploiting the factor model in a special way to obtain intercept and slope means and variances.



The growth curve equations given above can be stated using our SEM notation.

$$y_{ti} = \lambda_{t0}\eta_{0i} + \lambda_{t1}\eta_{1i} + \varepsilon_{ti}$$

$$\eta_{0i} = \alpha_0 + \zeta_{0i}$$
$$\eta_{1i} = \alpha_1 + \zeta_{1i}$$

For the intercept interpretation to work out, we will always want to include the covariance between the intercept factor and the slope factor, $Cov(\eta_{0i}, \eta_{1i}) = \psi_{01}$. It can be useful to interpret the standardized value, the intercept-slope correlation.

The correlation between the intercept and slope factors can be complicated to interpret. Simply put, the correlation represents the association between an individual's score at baseline (assuming the common coding scheme for the slope loadings) and change over time. A positive correlation indicates higher baseline scores are associated with a higher slope value. A "higher slope value" is a less negative slope or a larger positive slope. Plotting the individual slopes from an analysis can be helpful. Here are a few hypothetical examples that might be helpful to examine.

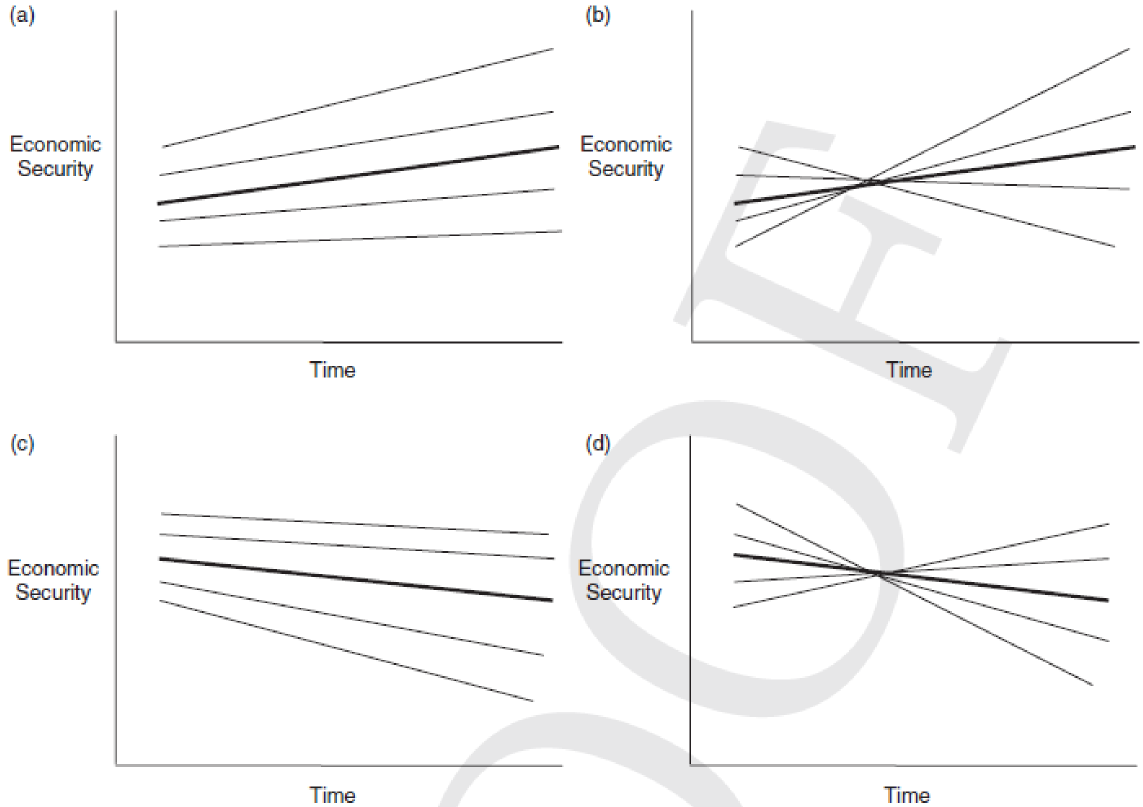


Figure 7.4 Four possible hypothetical relationships between intercept and slope for a growth model of perceived economic security. (a) average increase with positive correlation between intercept and slope; (b) average increase with negative correlation between intercept and slope; (c) average decrease with positive correlation between intercept and slope; (d) average decrease with negative correlation between intercept and slope.

Time Coding

By far, $0, 1, 2, 3, \ldots T-1$ is the most commonly used coding scheme for the slope factor loadings, although other coding schemes are certainly possible. In some instances, a researcher may be interested in an intercept that can be interpreted as the middle time point (e.g., λ_{t1} = -2, -1, 0, 1, 2) or the last time point (e.g., λ_{t1} = -4, -3, -2, -1, 0). Other time variables are certainly acceptable. Age or grade can be used, for example. Be aware that the interpretation of the intercept is for the predicted value of y when the time score equals 0. So, using age as the time variable gives the predicted score at birth. This may make sense in some cases but in many cases it will not make sense. Unequal spacing of time is another issue to be aware of. If there is a two year gap between the third and four wave of an otherwise annual survey, the codes should correctly correspond (e.g., 0, 1, 2, 4, 5).

Intraclass correlation coefficient

The intraclass correlationcoefficient, which gauges the proportion of between case and within case variance (between-person vs. within-person, if you like), is often used to describe how much scores are clustered within cases. It is not provided by any SEM programs, but is easily computed.

$$\rho = \frac{\psi_{00}}{Var\left(y_{ti}\right)} = \frac{\psi_{00}}{\psi_{00} + \theta_{(tt)}}$$

The variance estimate of the intercept factor (representing baseline score variability if common time scores are used) is given by ψ_{00}, and the measurement error variance is given by $\theta_{(tt)}$. In this case, the measurement error variances is constrained to be equal across time points (homogeneity of error variance) to obtain a single value, but an average of the freely estimated values might be a reasonable substitute.

Reliability

A closely related calculation is the reliability of the of the parameter estimate. The general concept is higher reliability reflects greater precision of the estimate, and parameter estimates have higher reliability, there will be greater statistical power. This can be calculated for the intercept or the slope. There are a few different approaches to assessing reliability, and I give only one of the possible here (Raudenbush & Bryk, 1986; 2002).

$$\rho_{RB} = \frac{\psi_{11}}{\psi_{11} + \theta_{(tt)} / T}$$

In the equation, ψ_{11} is the variance estimate of the slope factor and T is the number of time points. I discuss a couple of other possible proposed computations in the text. Notice that higher reliability occurs when the more variability of slopes between cases is large relative to the variability within cases and when there are more time points.

Model fit

Poor fit of a latent growth curve does not reflect the degree of change over time and it does not even necessarily reflect the validity of the linear form. The lack of fit is a function of the average deviation of observed values from the linear slope as illustrated in the individual growth figure above. Variance of the measurement residuals in this context is due to several factors (Bollen, 2007; Wu, West, & Taylor, 2009), including random measurement error in the observed variable, occasion-specific systematic variance, occasion-specific nonsystematic variance, and the correctness of the functional form (i.e., linear in the present model).

References and Further Reading

Baker, G. A. (1954). Factor analysis of relative growth. *Growth, 18*, 137-143.
Bollen, K. A. (2007). On the origins of latent curve models. In R. Cudeck & R. MacCallum (Eds), *Factor analysis at 100* (pp. 79-98). Mahwah, NJ: Lawrence Erlbaum Associates.
Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New York: John Wiley & Sons.
Meredith, W., & Tisak, J. (1984). *Statistical considerations in Tuckerizing curves with emphasis on growth curves and cohort sequential analysis*. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.
Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.
Newsom, J.T. (2015). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction*. New York: Routledge.
Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrika, 51*, 83-90.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, NJ: Sage
Raudenbush, S., & Bryk, A.S. (1986). "A hierarchical model for studying school effects." *Sociology of education*. 1-17.
Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23,* 19-23.
Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods*, *14*, 183.