Latent Growth Curve Analysis

John B. Willett & Kristen L. Bub
Harvard University Graduate School of Education

Learning and development are ubiquitous. When new skills are acquired, when attitudes and interests develop, people change. Measuring change demands a longitudinal perspective, with multiple waves of data collected on representative people at sensibly spaced intervals. Multi-wave data is usually analyzed by *individual growth modeling* using a *multilevel model for change* (Singer & Willett, 2003). Recently, innovative methodologists (McArdle, 1986; Meredith & Tisak, 1990; Muthen, 1991) have shown how the multilevel model for change can be mapped onto the general covariance structure model, such as that implemented in *LISREL*. This has lead to an alternative approach to analyzing change known as *latent growth modeling*. In this chapter, we describe and link these two analogous approaches.

Our presentation uses four waves of data on the reading scores of 1,740 Caucasian children from the *Early Childhood Longitudinal Study* (*ECLS-K*; NCES, 2002). Children's reading ability was measured in the Fall and Spring of kindergarten and first grade – we assume that test administrations were six months apart, with time measured from entry into kindergarten. Thus, in our analyses, predictor t -- representing time – has values 0, 0.5, 1.0 and 1.5 years. Finally, we know the child's gender (*FEMALE*: boy = 0, girl = 1), which we treat as a time-invariant predictor of change.[1]

I. Introducing Individual Growth Modeling

In Figure 1, we display empirical reading records for ten children selected from the larger dataset. In the top *left* panel is the growth record of child #15013, a boy, with observed reading score on the ordinate and time on the abscissa. Reading scores are represented by a "+" symbol and are connected by a smooth freehand curve summarizing the change trajectory. Clearly, this boy's reading ability improves

---

[1] The dataset is available at http://gseacademic.harvard.edu/~willetjo/.

during kindergarten and first grade. In the top *right* panel, we display similar smoothed change

trajectories for all ten children (dashed trajectories for boys, solid for girls, plotting symbols omitted to

reduce clutter). Notice the dramatic changes in children's observed reading scores over time, and how

disparate they are from child to child. The complexity of the collection, and because true reading ability

is obscured by measurement error, makes it hard to draw defensible conclusions about gender differences.

 However, perhaps the girls' trajectories do occupy higher elevations than those of the boys, on average.

[Insert Figure 1 here]

Another feature present in the reading trajectories in the top two panels of Figure 1 is the apparent

acceleration of the observed trajectories between Fall and Spring of first grade. Most children exhibit

moderate growth in reading over the first three waves, but their scores increase dramatically over the last

time period. The score of child #15013, for instance, rises modestly between waves 1 and 2 (20 to 28

points), modestly again between waves 2 and 3 (28 to 39 points), and then rapidly (to 66 points) by the

fourth wave. Because of this non-linearity, which was also evident in the entire sample, we transformed

children's reading scores before further analysis (Singer & Willett (2003, Chapter 6) comment on

choosing an appropriate transformation). We used a natural log transformation in order to "pull down"

the top end of the change trajectory disproportionally, thereby linearizing the accelerating raw trajectory.

In the lower panels of Figure 1, we redisplay the data in the newly transformed logarithmic

world. The *log-reading* trajectory of child #15013 is now approximately linear in time, with positive

slope. To dramatize this, we have superimposed a *linear* trend line on the transformed plot (by simply

regressing the log-reading scores on time using OLS regression analysis for that child). This trend line

has a positive slope, indicating that the log-reading score increases during kindergarten and first grade. In

the lower right panel, we display OLS-fitted linear trajectories for all ten children in the sub-sample and

reveal the heterogeneity in change that remains across children (albeit change in log-reading score). In

subsequent analyses, we model change in the log-reading scores as a linear function of time.

The individual change trajectory can be described by a "within-person" or "level-1" *individual*

growth model (Singer and Willett, 2003, Ch. 3). For instance, here we hypothesize that the log-reading

score, Y_{ij}, of child i on occasion j is a *linear* function of time, t:

$$Y_{ij} = \{\pi_{0i} + \pi_{1i}t_j\} + \varepsilon_{ij} \tag{1}$$

where $i = 1, 2, \ldots, 1740$ and $j = 1, 2, 3, 4$ (with, as noted earlier, $t_1 = 0$, $t_2 = 0.5$, $t_3 = 1.0$ and $t_4 = 1.5$

years, respectively). We have bracketed the *systematic* part of the model to separate the orderly temporal

dependence from the random errors, ε_{ij}, that accrue on each measurement occasion. Within the brackets,

you will find the *individual growth parameters*, π_{0i} and π_{1i}:

- π_{0i} is the *intercept* parameter, describing the child's true "initial" log-reading score on entry into

  kindergarten (because entry into kindergarten has been defined as the origin of time).

- π_{1i} is the *slope* ("rate of change") parameter, describing the child's true annual rate of change in

  log-reading score over time. If π_{1i} is positive, true log-reading score increases with time.

If the model is correctly specified, the individual growth parameters capture the *defining features* of the

log-reading trajectory for child i. Of course, in specifying such models, you need not choose a linear

specification -- many shapes of trajectory are available, and the particular one that you choose should

depend on your theory of change and on your inspection of the data (Singer & Willett, 2003, Ch. 6).

One assumption built deeply into individual growth modeling is that, while every child's change

trajectory has the *same functional form* (here, linear in time), different children may have *different values*

of the individual growth parameters. Children may differ in intercept (some have low log-reading ability

on entry into kindergarten, others are higher) and in slope (some children change more rapidly with time,

others less rapidly). Such heterogeneity is evident in the right hand lower panel of Figure 1.

We have coded the trajectories in the right-hand panels of Figure 1 by child gender. Displays like

these help to reveal systematic differences in change trajectory from child to child, and help you to assess

whether inter-individual variation in change is related to individual characteristics, like gender. Such

"level-2" questions -- about the effect of *predictors* of change -- translate into questions about "between-

person" relationships among the individual growth parameters and predictors like gender. Inspecting the right hand lower panel of Figure 1, for instance, you can ask whether boys and girls differ in their initial scores (do the intercepts differ by gender?) or in the rates at which their scores change (do the slopes differ by gender?).

Analytically, we can handle this notion in a second "between-person" or "level-2" statistical model to represent *inter-individual differences in change*. In the level-2 model, we express how we believe the individual growth parameters (standing in place of the full trajectory) depend on predictors of change. For example, we could investigate the impact of child gender on the log-reading trajectory by positing the following pair of simultaneous level-2 statistical models:

$$\pi_{0i} = \gamma_{00} + \gamma_{01}FEMALE_i + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}FEMALE_i + \zeta_{1i}$$

(2)

where the level-2 residuals, ζ_{0i} and ζ_{1i}, represent those portions of the individual growth parameters that are "unexplained" by the selected predictor of change, *FEMALE*. In this model, the γ coefficients are known as the "fixed effects" and summarize the population relationship between the individual growth parameters and the predictor. They can be interpreted like regular regression coefficients. For instance, if the initial log-reading ability of girls is higher than boys (i.e., if girls have larger values of π_{0i}, on average) then γ_{01} will be positive (since *FEMALE* = 1, for girls). If boys have lower annual rates of change (i.e., if boys have smaller values of π_{1i}, on average), then γ_{11} will be negative. Together, the level-1 and level-2 models in (1) and (2) make up the *multilevel model for change* (Singer & Willett, 2003, Ch. 3).

Researchers investigating change must fit the multilevel model for change to their longitudinal data. Many methods are available for doing this (see Singer & Willett, 2003, Chs. 2 and 3), the simplest of which is exploratory, as in Figure 1. To conduct data-analyses *efficiently*, the level-1 and level-2 models are usually fitted *simultaneously* using procedures now widely available in major statistical packages. The models can also be fitted using covariance structure analysis, as we now describe.

II.   Latent Growth Modeling

Here, we introduce latent growth modeling by showing how the multi-level model for change can be mapped onto the general covariance structure model. Once the mapping is complete, all parameters of the multi-level model for change can be estimated by fitting the companion covariance structure model using standard covariance structure analysis (CSA) software, such as AMOS, LISREL, EQS, MPLUS, etc.

To conduct latent growth analyses, we lay out our data in *multivariate* format, in which there is a single row in the dataset for each person, with multiple (*multi-*) variables (*-variate*) containing the time-varying information, arrayed horizontally. With four waves of data, multivariate format requires four columns to record each child's growth record, each column associated with a measurement occasion. Any time-invariant predictor of change, like child gender, also has its own column in the dataset. Multivariate formatting is not typical in longitudinal data analysis (which usually requires a "person-period" or "univariate" format), but is required here because of the nature of covariance structure analysis. As its name implies, CSA is an analysis of *covariance structure* in which, as an initial step, a sample covariance matrix (and mean vector) is estimated to summarize the associations among (and levels of) selected variables, including the multiple measures of the outcome across the several measurement occasions. The data-analyst then specifies statistical models appropriate for the research hypotheses, and the mathematical implications of these hypotheses for the structure of the underlying population covariance matrix and mean vector are evaluated against their sample estimates. Because latent growth analysis compares sample and predicted covariance matrices (and mean vectors), the data must be formatted to support the estimation of covariance matrices (and mean vectors) – in other words, in a *multivariate* format.

Note, finally, that there is no unique column in the multivariate dataset to record time. In our multivariate format dataset, values in the outcome variable's first column were measured at the *start* of

kindergarten, values in the second column were measured at the beginning of spring in kindergarten, etc.

The time values – each corresponding to a particular measurement occasion and to a specific column of

outcome values in the dataset – are noted by the analyst and programmed directly into the CSA model. It

is therefore more convenient to use latent growth modeling to analyze change when panel data are *time-*

structured – when everyone has been measured on an identical set of occasions and possesses complete

data. Nevertheless, you can use latent growth modeling to analyze panel datasets with limited violations

of time-structuring, by regrouping the full sample into sub-groups who share identical time-structured

profiles and then analyzing these subgroups simultaneously with CSA *multi-group* analysis.

II.1   Mapping the Level-1 Model for Individual Change onto the CSA Y-Measurement Model

In (1), we specified that the child's log-reading score, Y_{ij}, depended linearly on time, measured

from kindergarten entry. Here, for clarity, we retain symbols t_1 through t_4 to represent the measurement

timing but you should remember that each of these time symbols has a known value (0, 0.5, 1.0 and 1.5

years, respectively) that is used when the model is fitted. By substituting into the individual growth

model, we can create equations for the value of the outcome on each occasion for child *i*:

$$
\begin{aligned}
Y_{i1} &= \pi_{0i} + \pi_{1i}t_1 + \varepsilon_{i1} \\
Y_{i2} &= \pi_{0i} + \pi_{1i}t_2 + \varepsilon_{i2} \\
Y_{i3} &= \pi_{0i} + \pi_{1i}t_3 + \varepsilon_{i3} \\
Y_{i4} &= \pi_{0i} + \pi_{1i}t_4 + \varepsilon_{i4}
\end{aligned}
\tag{3}
$$

that can easily be rewritten in simple matrix form, as follows:

$$
\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \end{bmatrix} \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}
\tag{4}
$$

While this matrix equation is unlike the representation in (1), *it says exactly the same thing* -- that

observed values of the outcome, Y, are related to the times (t_1, t_2, t_3, and t_4), to the individual growth

parameters (π_{0i} and π_{1i}), and to the measurement errors (ε_{i1}, ε_{i2}, ε_{i3}, and ε_{i4}). The only difference between

(4) and (1) is that all values of the outcome and of time, and all parameters and time-specific residuals, are

arrayed neatly as vectors and matrices. (Don't be diverted by the strange vector of zeros introduced

immediately to the right of the equals sign – it makes no difference to the *meaning* of the equation, but it

will help our subsequent mapping of the multilevel model for change onto the general CSA model).

In fact, the new growth model representation in (4) maps straightforwardly onto the CSA *Y-Measurement Model*, which, in standard *LISREL* notation, is:

$$\mathbf{Y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \, \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{5}$$

where $\mathbf{Y}$ is a vector of observed scores, $\boldsymbol{\tau}_y$ is a vector intended to contain the population means of $\mathbf{Y}$,

$\boldsymbol{\Lambda}_y$ is a matrix of factor loadings, $\boldsymbol{\eta}$ is a vector of latent (endogenous) constructs, and $\boldsymbol{\varepsilon}$ is a vector of

residuals.[2] Notice that the new matrix representation of the individual growth model in (4) matches the

CSA Y-Measurement Model in (5) providing that the observed and latent score vectors are set to:

$$\mathbf{Y} = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} \tag{6}$$

and providing that parameter vector $\boldsymbol{\tau}_y$ and loading matrix $\boldsymbol{\Lambda}_y$ are specified as containing the following

constants and known times:

$$\boldsymbol{\tau}_y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \end{bmatrix} \tag{7}$$

---

[2]   Readers unfamiliar with the general *CSA* model should consult Bollen (1989).

Check this by substituting from (6) and (7) into (5) and multiplying out -- you will conclude that, with

this specification of score vectors and parameter matrices, the CSA Y-Measurement Model can act as, or

contain, the individual growth trajectory from the multi-level model for change.

Notice that (1), (3), (4), (5) and (6) all permit the measurement errors to participate in the

individual growth process. They state that level-1 residual ε_{i1} disturbs the true status of the ith child on the

first measurement occasion, ε_{i2} on the second occasion, ε_{i3} on the third, and so on. However, so far, we

have made no claims about the underlying distribution from which the errors are drawn. Are the errors

normally distributed, homoscedastic and independent over time within-person? Are they heteroscedastic

or auto-correlated? Now that the individual change trajectory is embedded in the Y-Measurement Model,

we can easily account for level-1 error covariance structure because, under the usual CSA assumption of a

multivariate normal distribution for the errors, we can specify the CSA parameter matrix Θ_ε to contain

hypotheses about the covariance matrix of $\boldsymbol{\varepsilon}$. In an analysis of change, we usually compare nested

models with alternative error structures to identify which error structure is optimal. Here, we assume that

level-1 errors are distributed normally, independently, and heteroscedastically over time within-person:[3]

$$\boldsymbol{\Theta}_\varepsilon = \begin{bmatrix} \sigma_{\varepsilon 1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon 2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon 3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon 4}^2 \end{bmatrix} \tag{8}$$

Ultimately, we estimate all level-1 variance components on the diagonal of Θ_ε and reveal the action of

measurement error on reading score on each occasion.

The key point is that judicious specification of CSA score and parameter matrices forces the

level-1 individual change trajectory into the Y-Measurement Model in a companion covariance structure

analysis. Notice that, unlike more typical covariance structure (confirmatory factor) analyses, the $\boldsymbol{\Lambda_y}$

---

[3]   Supplementary analyses suggested that this was reasonable.

matrix in (7) is *entirely specified as a set of known constants and times* rather than as *unknown latent factor loadings* to be estimated. Using the Y-Measurement Model to represent individual change in this way "forces" the individual-level growth parameters, π_{0i} and π_{1i}, into the endogenous construct vector $\boldsymbol{\eta}$, creating what is known as the *latent growth vector*, $\boldsymbol{\eta}$. This notion – that the CSA $\boldsymbol{\eta}$-vector can be *forced* to contain the individual growth parameters – is critical in latent growth modeling, because it suggests that level-2 inter-individual variation in change can be modeled in the CSA Structural Model, as we now show.

II.2   Mapping the Level-2 Model for Inter-Individual Differences in Change onto the CSA Structural Model

In an analysis of change, at level-2, we ask whether inter-individual heterogeneity in change can be predicted by other variables, such as features of the individual's background and treatment. For instance, in our data-example, we can ask whether between-person heterogeneity in the log-reading trajectories depends on the child's gender. Within the growth-modeling framework, this means that we must check whether the individual growth parameters – the true intercept and slope standing in place of the log-reading trajectories – are related to gender. Our analysis therefore asks: Does initial log-reading ability differ for boys and girls? Does the annual rate at which log-reading ability changes depend upon gender? In latent growth modeling, level-2 questions like these, which concern the distribution of the individual growth parameters *across individuals* and their relationship to predictors, are addressed by specifying a CSA *Structural Model*. Why? Because it is in the CSA structural model that the vector of unknown endogenous constructs $\boldsymbol{\eta}$ -- which now contains the all-important individual growth parameters, π_{0i} and π_{1i} -- is hypothesized to vary across people.

Recall that the CSA Structural Model stipulates that endogenous construct vector $\boldsymbol{\eta}$ is potentially related to both itself and to exogenous constructs $\boldsymbol{\xi}$ by the following model:

$$\eta = \alpha + \Gamma\xi + B\eta + \zeta \tag{9}$$

where α is a vector of intercept parameters, Γ is a matrix of regression coefficients that relate exogenous predictors ξ to outcomes η, B is a matrix of regression coefficients that permit elements of the endogenous construct vector η to predict each other, and ζ is a vector of residuals. In a covariance structure analysis, we fit this model to data, simultaneously with the earlier measurement model, and estimate parameters α, Γ and B. The rationale behind latent growth modeling argues that, by structuring (9) appropriately, we can force parameter matrices α, Γ and B to contain the fixed effects central to the multilevel modeling of change.

So, what to do? Inspection of the model for systematic inter-individual differences in change in (2) suggests that the level-2 component of the multilevel model for change can be reformatted in matrix form, as follows:

$$\begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} \gamma_{01} \\ \gamma_{11} \end{bmatrix} [FEMALE_i] + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix} + \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \tag{10}$$

which is identical to the CSA Structural Model in (9), providing that we force the elements of the CSA B parameter matrix to be zero throughout:

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{11}$$

and that we permit the α vector and the Γ matrix to be free to contain the fixed effects parameters from the multilevel model for change:

$$\alpha = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_{01} \\ \gamma_{11} \end{bmatrix} \tag{12}$$

and providing we can force the potential predictor of change – child gender -- into the CSA exogenous construct vector, ξ. In this new level-2 specification of the structural model, the latent intercept vector, α,

contains the level-2 fixed-effects parameters γ_{00} and γ_{10}, defined earlier as the population intercept and

slope of the log-reading trajectory for boys (when FEMALE = 0). The Γ matrix contains the level-2

fixed-effects parameters γ_{01} and γ_{11}, representing increments to the population average intercept and slope

for girls, respectively. By fitting this CSA model to data, we can estimate all four fixed effects.

When a time-invariant predictor like *FEMALE* is present in the structural model, the elements of

the latent residual vector ζ in (10) represent those portions of true intercept and true slope that are

unrelated to the predictor of change -- the "adjusted" values of true intercept and slope, with the linear

effect of child gender partialled out. In a covariance structure analysis of the multilevel model for

change, we assume that latent residual vector ζ is distributed normally with zero mean vector and

covariance matrix $\boldsymbol{\Psi}$,

$$\boldsymbol{\Psi} = Cov\{\boldsymbol{\zeta}\} = \begin{bmatrix} \sigma_{\zeta_0}^2 & \sigma_{\zeta_{01}} \\ \sigma_{\zeta_{10}} & \sigma_{\zeta_1}^2 \end{bmatrix} \tag{13}$$

which contains the residual (partial) variances and covariance of true intercept and slope, controlling for

the predictor of change, *FEMALE*. We estimate these level-2 variance components in any analysis of

change.

But there is one missing link that needs resolving before we can proceed. How is the

hypothesized predictor of change, *FEMALE*, loaded into the exogenous construct vector, ξ? This is easily

achieved via the so-far-unused CSA *X-Measurement Model*. And, in the current analysis, the process is

disarmingly simple because there is only a single infallible predictor of change, child gender. So, in this

case, while it may seem a little weird, the specification of the X-Measurement Model derives from a

tautology:

$$FEMALE_i = (0) + (1)(FEMALE_i) + (0) \tag{14}$$

Which, while not affecting predictor *FEMALE*, facilitates comparison with the CSA X-Measurement

Model:

$$\mathbf{X} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \tag{15}$$

By comparing (14) and (15), you can see that the gender predictor can be incorporated into the analysis

by specifying an X-Measurement Model in which:

- Exogenous score vector **X** contains one element, the gender predictor, *FEMALE*, itself.

- The **X**-measurement error vector, δ, contains a single element whose value is fixed at zero,

  embodying the assumption that gender is measured infallibly (with "zero" error).

- The $\boldsymbol{\tau}_x$ mean vector contains a single element whose value is fixed at zero. This forces the mean

  of *FEMALE* (which would reside in $\boldsymbol{\tau}_x$ if the latter were not fixed to zero) into the CSA latent

  mean vector, κ, which contains the mean of the exogenous construct, ξ, in the general CSA model.

- The matrix of exogenous latent factor loadings $\boldsymbol{\Lambda_x}$ contains a single element whose value is fixed

  at 1. This forces the metrics of the exogenous construct and its indicator to be identical.

Thus, by specifying a CSA X-Measurement Model in which the score vectors are:

$$\mathbf{X} = \left[FEMALE_i \right], \boldsymbol{\delta} = \left[0 \right] \tag{16}$$

and the parameter matrices are fixed at:

$$\boldsymbol{\tau}_x = \left[0 \right], \boldsymbol{\Lambda}_x = \left[1 \right] \tag{17}$$

we can make the CSA exogenous construct ξ represent child gender. And, since we know that exogenous

construct ξ is a predictor in the CSA Structural Model, we have succeeded in inserting the predictor of

change, child gender, into the model for inter-individual differences in change. As a final consequence of

(14) through (17), the population mean of the predictor of change appears as the sole element of the CSA

exogenous construct mean vector, κ:

$$\boldsymbol{\kappa} = Mean\{\boldsymbol{\xi}\} = \left[\mu_{FEMALE} \right] \tag{18}$$

and the population variance of the predictor of change appears as the sole element of CSA exogenous

construct covariance matrix $\mathbf{\Phi}$:

$$\mathbf{\Phi} = Cov\{\xi\} = \left[\sigma^2_{FEMALE}\right] \tag{19}$$

Both of these parameter matrices are estimated when the model is fitted to data. And, although we do not

demonstrate it here, the X-Measurement Model in (14) through (19) can be reconfigured to accommodate

multiple time-invariant predictors of change, and even several indicators of each predictor construct if

available. This is achieved by expanding the exogenous indicator and construct score vectors to include

sufficient elements to contain the new indicators and constructs and the parameter matrix $\mathbf{\Lambda_x}$ is expanded

to include suitable loadings (Willett and Singer (2003; Ch. 8) give an example with multiple predictors).

So, the CSA version of the multilevel model for change – now called the *latent growth model* – is

complete. It consists of the CSA *X-Measurement*, *Y-Measurement*, and *Structural Models*, defined in (14)

through (19), (4) through (8), and (9) through (13), respectively and is displayed as a path model in

Figure 2. In the figure, by fixing the loadings associated with the outcome measurements to their constant

and temporal values, we emphasize how the endogenous constructs were forced to become the individual

growth parameters, which are then available for prediction by the hypothesized predictor of change. We

fitted the latent growth model in (4) through (14) to our reading data on the full sample of 1740 children

using LISREL (see Appendix). Table 1 presents full maximum-likelihood (FML) estimates of all

relevant parameters from latent regression-weight matrix $\mathbf{\Gamma}$ and parameter matrices $\mathbf{\Phi}$, $\mathbf{\alpha}$ and $\mathbf{\Psi}$.

[Insert Figure 2 and Table 1 here]

The estimated level-2 fixed effects are in the first four rows of Table 1. The first and second rows

contain estimates of parameters γ_{00} and γ_{10}, representing true initial log-reading ability ($\hat{\gamma}_{00} = 3.170$,

p<.001) and true annual rate of change in log-reading ability ($\hat{\gamma}_{10} = 0.583$, p<.001) for boys (for whom

FEMALE = 0). Anti-logging tells us that, on average, boys: (a) begin kindergarten with an average

reading ability of 23.8 ($=e^{3.1700}$), and (b) increase their reading ability by 79% ($=100(e^{0.5828} - 1)$) per year.

The third and fourth rows contain the estimated latent regression coefficients γ_{01} (0.073, p<.001) and γ_{11} (-0.010, p>.10), which capture differences in change trajectories between boys and girls. Girls have a higher initial level of 3.243 (= 3.170 + 0.073) of log-reading ability, whose anti-log is 25.6 and a statistically significant couple of points higher than the boys. However, we cannot reject the null hypothesis associated with γ_{11} (-0.010, p>.10) so, although the estimated annual rate of increase in log-reading ability for girls is 0.572 (= 0.5828 - 0.0096), a little smaller than boys, this difference is not statistically significant. Nonetheless, anti-logging, we find that girls' reading ability increases by about 78% ($=100(e^{0.5828}-1)$) per year, on average. We display fitted log-reading and reading trajectories for prototypical boys and girls in Figure 3 -- once de-transformed, the trajectories are curvilinear and display the acceleration we noted earlier in the raw data.

[Insert Figure 3 here]

Next, examine the random effects. The fifth through eighth rows of Table 1 contain estimated level-1 error variances, one per occasion, describing the measurement fallibility in log-reading score over time. Their estimated values are 0.022, 0.023, 0.021, and 0.008, respectively, showing considerable homoscedasticity over the first three occasions but measurement error variance decreases markedly in the spring of first grade. The tenth through twelfth rows of Table 1 contain the estimated level-2 variance components, representing estimated partial (residual) variances and partial covariance of true initial status and rate of change, after controlling for child gender. We reject the null hypothesis associated with each variance component, and conclude that there is predictable true variation remaining in both initial status and rate of change.

Conclusion: Extending the Latent Growth Modeling Framework

In this chapter, we have shown how a *latent growth modeling* approach to analyzing change is created by mapping the multilevel model for change onto the general CSA model. The basic latent

growth modeling approach that we have described can be extended in many important ways:

- **You can include any number of waves of longitudinal data**, by simply increasing the number of rows in the relevant score vectors. Including more waves generally leads to greater precision for the estimation of the individual growth trajectories and greater reliability for measuring change.

- **You need not space the occasions of measurement equally**, although it is most convenient if everyone in the sample is measured on the *same set* of irregularly spaced occasions. However, if they are not, then latent growth modeling can still be conducted by first dividing the sample into sub-groups of individuals with identical temporal profiles and using multi-group analysis to fit the multi-level model for change simultaneously in all sub-groups.

- **You can specify curvilinear individual change**. Latent growth modeling can accommodate polynomial individual change of any order (provided sufficient waves of data are available), or any other curvilinear change trajectory in which individual status is linear in the growth parameters.

- **You can model the covariance structure of the level-1 measurement errors explicitly**. You need not accept the independence and homoscedasticity assumptions of classical analysis unchecked. Here, we permitted level-1 measurement errors to be heteroscedastic, but other, more general, error covariance structures can be hypothesized and tested.

- **You can model change in several domains simultaneously**, including both exogenous and endogenous domains. You simply extend the empirical growth record and the measurement models to include rows for each wave of data available, in each domain.

- **You can model *intervening effects***, whereby an exogenous predictor may act directly on endogenous change and also indirectly via the influence of intervening factors, each of which may be time-invariant or time varying.

In the end, you must choose your analytic strategy to suit the problems you face. Studies of change can be designed in enormous variety and the multilevel model for change can be specified to account for all manner of trajectories and error structures. But, it is always wise to have more than one way to deal with data -- latent growth modeling often offers a flexible alternative to more traditional approaches.

Bibliography

Bollen, K. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

McArdle, J. J. (1986). Dynamic But Structural Equation Modeling Of Repeated Measures Data. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook Of Multivariate Experimental Psychology*, Volume 2. New York: Plenum, 561-614.

Meredith, W., & Tisak, J. (1990). Latent Curve Analysis. *Psychometrika,* 55, 107-122.

Muthen, B. O. (1991). Analysis Of Longitudinal Data Using Latent Variable Models With Varying Parameters. In L. M. Collins & J. L Horn (Eds.), *Best Methods For The Analysis Of Change: Recent Advances, Unanswered Questions, Future Directions*. Washington, DC: American Psychological Assocation, 1-17.

NCES (2002). *Early Childhood Longitudinal Study*. Washington, D.C.: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford University Press.

Appendix: Specimen LISREL Program

```
/*Specify the number of variables (indicators) to be read from the external data-file of raw data*/
        data ni=6

/*Identify the location of the external data-file*/
        raw fi = C:\Data\ECLS.dat

/*Label the input variables and select those to be analyzed*/
        label
        id Y1 Y2 Y3 Y4 FEMALE
        select
        2 3 4 5 6 /

/*Specify the hypothesized covariance structure model*/
        model      ny=4 ne=2 ty=ze ly=fu,fi te=di,fi                        c
                   nx=1 nk=1 lx=fu,fi tx=fr td=ze ph=sy,fr                  c
                   al=fr ga=fu,fr be=ze ps=sy,fr

/*Label the individual growth parameters as endogenous constructs (eta's)*/
        le
        pi0 pi1

/*Label the predictor of change as an exogenous construct (ksi) */
        lk
        FEMALE

/*Enter the required "1's" and measurement times into the Lambda-Y matrix*/
        va 1 ly(1,1) ly(2,1) ly(3,1) ly(4,1)
        va 0.0 ly(1,2)
        va 0.5 ly(2,2)
        va 1.0 ly(3,2)
        va 1.5 ly(4,2)

/*Enter the required scaling factor "1" into the Lambda-X matrix*/
        va 1.0 lx(1,1)

/*Free up the level-1 residual variances to be estimated*/
        fr te(1,1) te(2,2) te(3,3) te(4,4)

/*Request data-analytic output to 5 decimal places*/
        ou nd=5
```

Table 1. Trajectory of change in the logarithm of children's reading score over four measurement occasions during kindergarten and first grade, by child gender. Parameter estimates, approximate *p*-values, and goodness-of-fit statistics from the multi-level model for change, obtained with latent growth modeling (n=1740).

Effect	Parameter	Estimate
Fixed Effects:		
	γ_{00}	3.1700***
	γ_{10}	0.5828***
	γ_{01}	0.0732***
	γ_{11}	-0.0096
Variance Components:		
	$\sigma^2_{\varepsilon_1}$	0.0219***
	$\sigma^2_{\varepsilon_2}$	0.0228***
	$\sigma^2_{\varepsilon_3}$	0.0208***
	$\sigma^2_{\varepsilon_4}$	0.0077***
	$\sigma^2_{\zeta_0}$	0.0896***
	$\sigma^2_{\zeta_1}$	0.0140***
	$\sigma_{\zeta_0 \zeta_1}$	-0.0223***
Goodness of Fit:		
	χ^2	1414.25***
	df	7

~ = p<.10, * = p<.05, ** = p<.01, *** = p<.00

Figure Captions

<u>Figure 1</u>. Observed raw and transformed trajectories of reading score over kindergarten and first grade

for ten children (boys = dashed; girls = solid). *Top panel*: (a) raw reading score versus time for child

#15013, with observed data points (+'s) connected by a smoothed freehand trajectory, (b) smoothed

freehand trajectories for all 10 children. *Bottom panel*: (a) log-reading score versus time for child

#15013, with an OLS-estimated linear change trajectory, (b) OLS-estimated linear trajectories for all

children.

<u>Figure 2</u>. Path diagram for the hypothesized latent growth in reading score. Rectangles represent

observed indicators, circles represent latent constructs, and arrows and their associated parameters

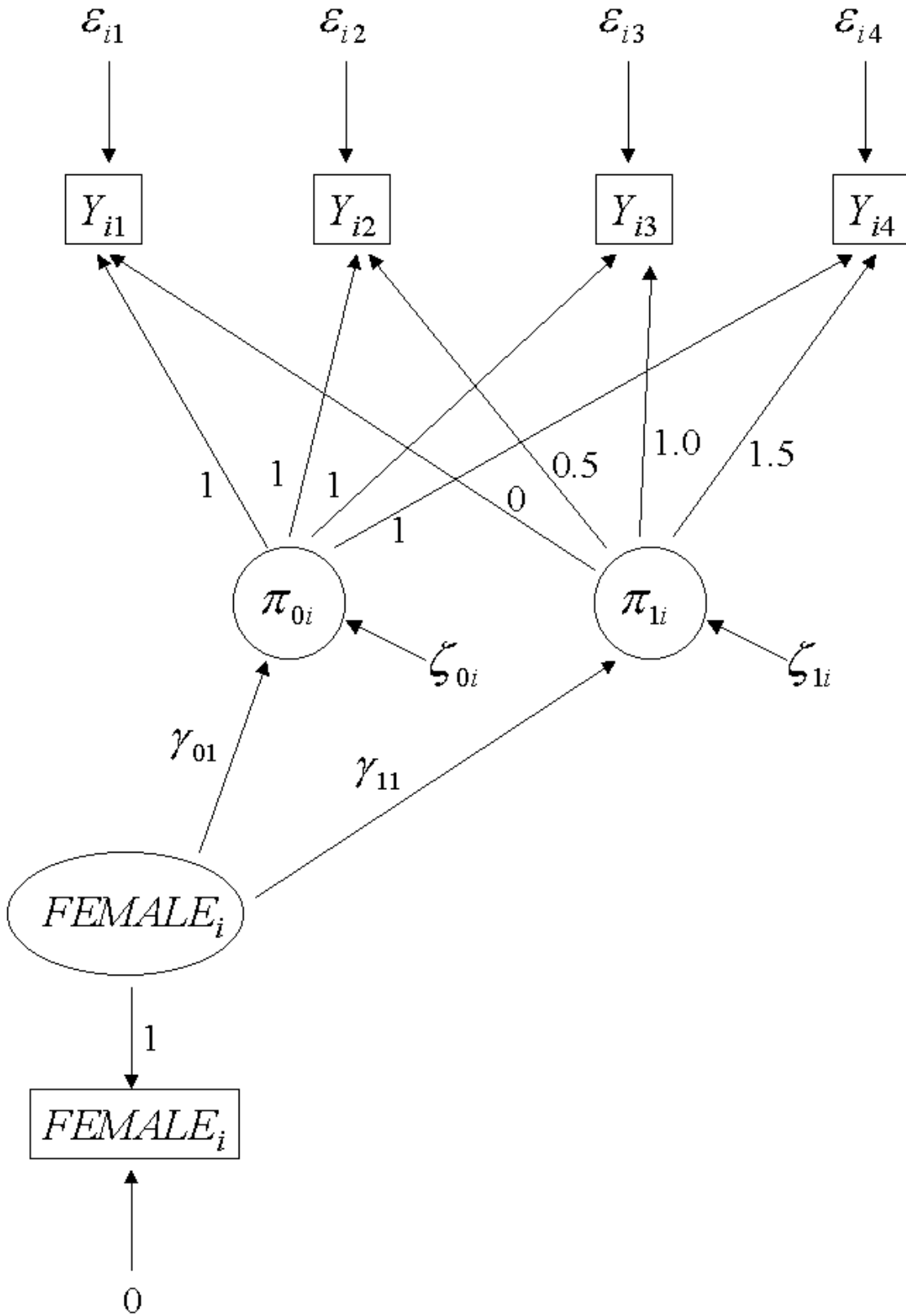indicate hypothesized relationships.

<u>Figure 3</u>. Fitted log-reading and reading trajectories over kindergarten and first grade for prototypical
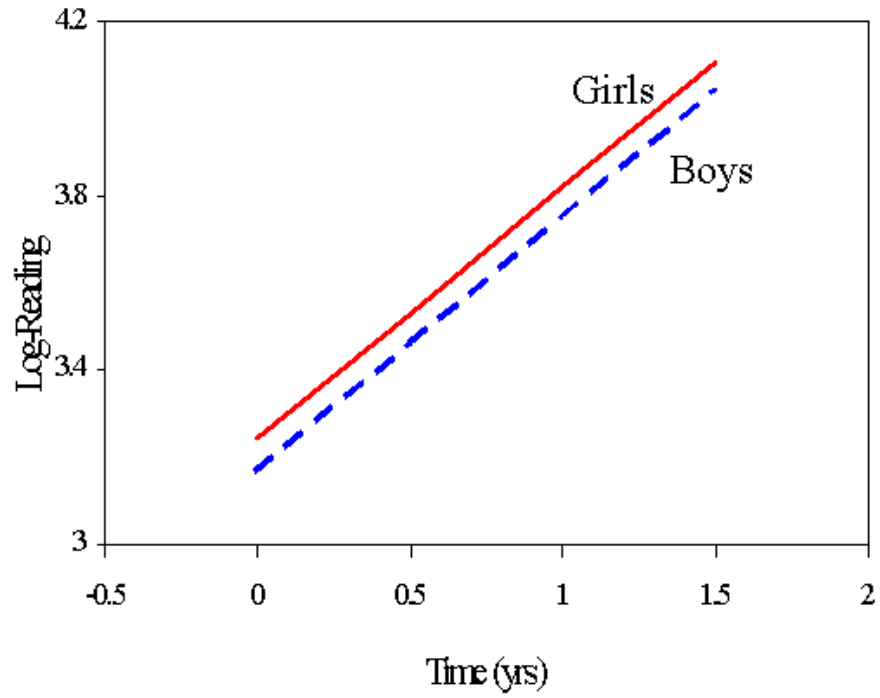
Caucasian children, by gender.

Raw Reading Scores:
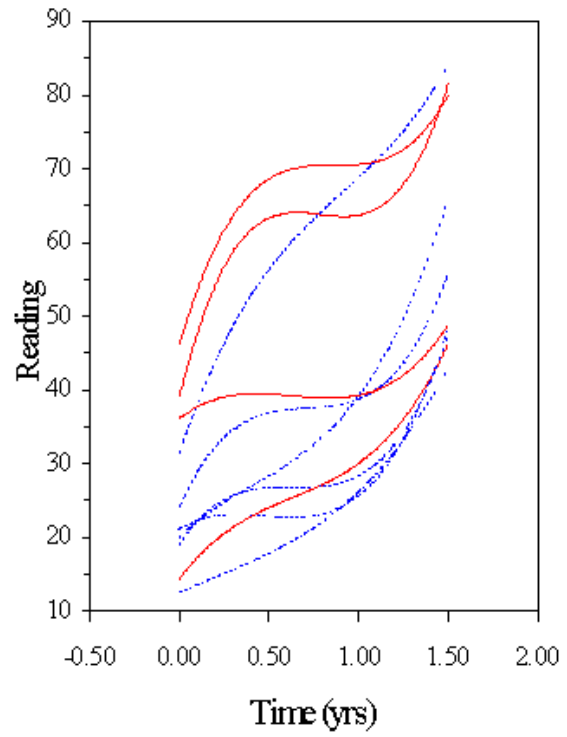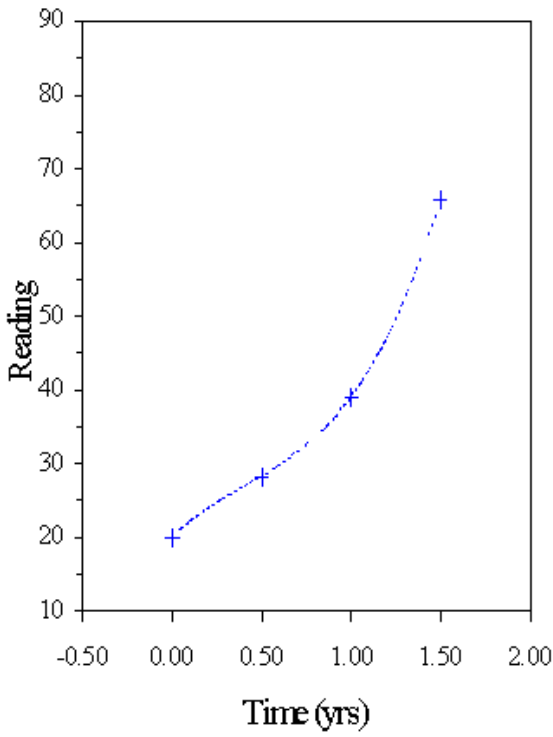
Log-Transformed Reading Scores:
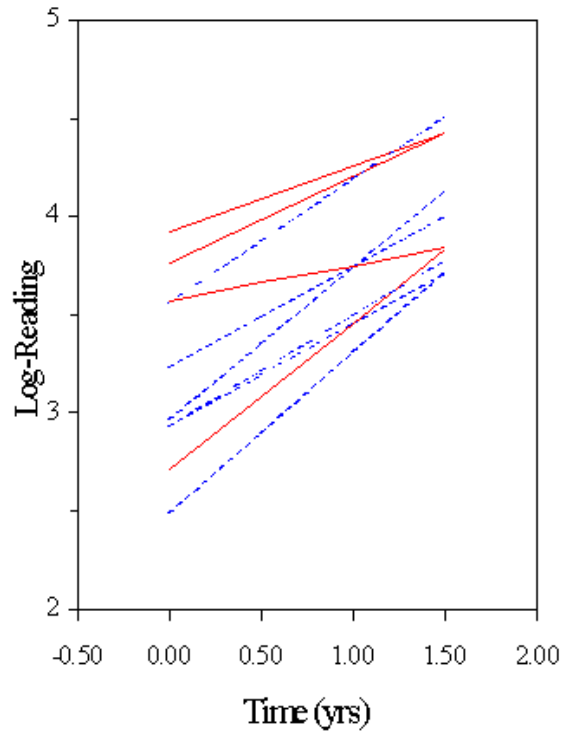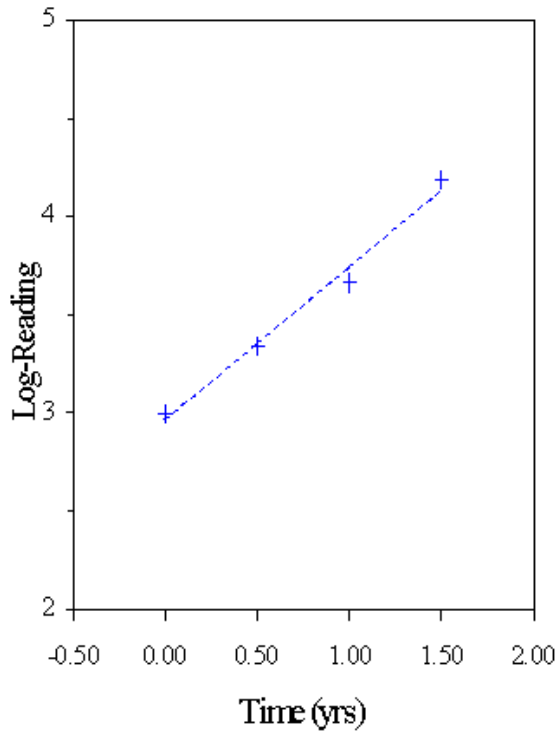
Fitted Log-Reading Scores

Fitted Reading Scores