

Latent Curve Analysis: A Manual for Research Data Analysts¹

The original version of this manual was written by
Fuzhong Li
Oregon Research Institute, Eugene, OR

It has been adapted for this presentation at the National Council on Family
Relations, 1999, by
Alan C. Acock²
Oregon State University

1 This manual represents collaborative work among its authors. This work was supported in part by a research grant for the secondary analysis of existing substance use data from the National Institute on Drug Abuse (Grant DA09548). We thank the Inter-university Consortium for Political and Social Research (ICPSR) for providing data from the National Youth Survey (NYS; Elliott, 1976) and Oregon Research Institute for providing data from the Young Adult Substance Use: Predictors and Consequences study, from which the examples for this manual were developed. The National Youth Survey was supported by Grant MH27552 from the National Institute of Mental Health and the Young Adult Substance Use: Predictors and Consequences study was supported by Grant DA09036 from the National Institute of Drug Abuse.

2 Contact Alan C. Acock, 322 Milam Hall, Department of Human Development and Family Sciences, Oregon State University, Corvallis OR 97331-5102. E-mail is alna.acock@orst.edu, phone 541.737.4992, and fax 541.737.1076. An electronic version of this manual is available at <http://osu.orst.edu/dept/hdfs/papers/paper.html>.

Preface

Methods of growth curve analysis have emerged as a versatile tool for studying longitudinal change. The methods are labeled using several names such as latent curve analysis, hierarchical linear modeling, multilevel modeling, and random coefficient models. In this manual, we focus on a structural equation modeling (SEM) approach described by a number of methodologists (i.e., McArdle & Epstein, 1987; Meredith & Tisak, 1990; Muthén, 1991; Willett & Sayer, 1994). Within this framework, we shall call it Latent Growth Curves analysis (LGC). LGC allows modeling trends of how groups and individuals change. Although the formal justifications are complex, fitting growth trajectory models can be carried out using standard computer software for SEM.

This manual presents a concrete explication of LGC that is intended to help data analysts who are studying change or rate of change. Despite an increase in the use of LGC, a practical manual of this type is not currently available to data analysts. Our goal is to provide a manual that provides the basic technical knowledge of programming LGC models using various SEM-based software. This manual is a work in progress. The complete manual will include conducting LGC on datasets that contain missing values and analyzing finite growth mixture models for unobserved heterogeneity in longitudinal change.

The manual is aimed at data analysts with a wide range of knowledge in statistical modeling. We assume that the user of this manual:

- Has read the "Introduction to Latent Growth Curves: A Gentle Introduction" (Acock and Li, 1999) or the equivalent, (<http://osu.orst.edu/dept/hdfs/papers/paper.html>)
- Understanding of analysis of variance, multiple regression, and factor analysis,
- Understanding the basic fundamentals of matrix algebra is helpful, but unnecessary, and
- Has access to one of the current SEM software programs presented in this manual and knows basic applications (e.g., being able to construct simple models such as a multiple regression model and a confirmatory factor analysis model).

The manual is divided into five sections.

- Section 1 presents an overview of LGC, with model specification for a two-factor, linear growth trajectory model. This is shorter, but more technical than the "Gentle Introduction" mentioned above.

- Section 2 introduces SEM software options that are used in this manual. These include Amos, EQS, LISREL/SIMPLIS, Mplus, Mx, and SAS's PROC CALIS. Also covered is the terminology and notation relevant to the use of LGC.
- Section 3 provides computer program codes (command files) for a two-factor (intercept, slope) LGC along with a schematic presentation of the model being tested. Selected portions of output are provided along with a brief, statistical interpretation of results.
- Section 4 covers using explanatory variables to predict the intercept and slope.
- Section 5 will be added to the complete manual. It covers fitting growth models with missing data. This includes LGC models with missing values caused by attrition and by design (i.e., multiple cohorts). This material is important because LGC rely on longitudinal data for which missing values is a constant problem.
- Section 6 will be added to the complete manual. It introduces more advanced topics representing a new growth curve modeling techniques—finite growth mixture modeling (Muthén, in press). Two examples are provided as applications of the new techniques: (a) growth mixture modeling and (b) piecewise growth mixture modeling. This material is important because it allows the researcher to empirically identify subgroups of people who have different LGC.

At the end of each section, references on the related methodological work and useful applications are provided. In working through the manual, users are encouraged to run the jobs (input files) of a particular software package themselves, both as an aid to the understanding of the material and to get practical experience in the use of the software in conducting LGC. The raw data sets used in this manual are available (<http://www.ori.org/~fuzhongl/data/home.htm>).

Although written primarily for research data analysts in social sciences, this manual can also be used as a class or supplementary manual for teaching growth curving modeling. Additionally, the manual may also be considered for use together with books that focus exclusively on latent growth curve modeling such as the one by Duncan, Duncan, Strycker, Li, and Alpert (1999).

SECTION 1: LATENT GROWTH CURVES ANALYSIS

Before preceding to the programming aspect of latent curve modeling, it is important to understand the basic feature of LGC. We provide a review on the elementary LGC for a two-factor (intercept and slope) linear growth trajectory model. The primary purpose is to introduce major parameters in LGC. Those experienced with LGC may skip this section.

Introduction

In many longitudinal studies, one of the important hypotheses involves whether the (latent) mean level of an attribute changes (e.g., increase or decrease) across time? For example, does the level of frustration increase over time for care providers to the adult handicapped? Does the average level of alienation change with prolonged drug abuse? The next question is what shape or functional form (e.g., straight line or quadratic) is indicated. Of particular additional interest is whether there are intra-individual differences in the developmental course (i.e., individual variability in initial level, rate of change, and final level). For example, some care providers may become increasingly frustrated, while others become less frustrated. Latent curve models provide a useful approach to address these research questions.

A Two-Factor Linear Growth Model

We begin with a brief introduction on the main ideas behind LGC using a two-factor linear growth model. Consider the example of a linear straight-line individual growth curve model, with p individuals ($p = 1, 2, \dots, N$) assessed at t times ($t = 0, 1, 2, 3$) on an observed outcome variable Y , we have:

$$Y_{pt} = \eta_{ip} + \pi_{sp} a_t + \varepsilon_{pt} \quad (1.1)$$

This linear model equation indicates that the observed outcome scores Y for each individual p at each time point t come from three unobserved sources of individual differences:

- η_{ip} represents the intercepts or initial levels i for each person, p . These are the same across all occasions t ,
- π_{sp} represents the slopes or growth rate for each person, p . These are multiplied by a set of coefficients a_t , that represent the time points such as 0,

- 1, 2, . . . T . Notice that 0 is used for the first measurement point,
- and ε_{pt} represents time-specific random errors.

Assuming a linear growth in Y variable across $T = 4$ evenly spaced times, the a_t coefficients can be fixed at 0, 1, 2, and 3 for the first, second, third, and fourth measurement times for the slope and 1, 1, 1, and 1 for the intercept:

$$\mathbf{a} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}' . \quad (1.2)$$

Setting $t_1 = 0$ on the factor loadings of η_s allows us to interpret η_i as the individual initial status at the start of the growth curve and η_s as the individual rate of change across time.

Equation 1.1 can be defined for each of the 4 time points:

EQUATION	TIME
$Y_{pt} = \eta_{ip} + \eta_{sp} \mathbf{a}_t + \varepsilon_{pt}$	General Model
$Y_{pt} = \eta_{ip} + \varepsilon_{pt}$	Time 1: $t_1 = 0$
$Y_{pt} = \eta_{ip} + \eta_{sp} \mathbf{1} + \varepsilon_{pt}$	Time 2: $t_2 = 1$
$Y_{pt} = \eta_{ip} + \eta_{sp} \mathbf{2} + \varepsilon_{pt}$	Time 3: $t_3 = 2$
$Y_{pt} = \eta_{ip} + \eta_{sp} \mathbf{3} + \varepsilon_{pt}$	Time 4: $t_4 = 3$

Both of the common factors, i.e., the intercept/initial level η_{ip} and slope π_{sp} have a mean, denoted by mean α_i (the group's average intercept at the initial level) and α_s (the group's average rate of change). They also have variances, denoted by σ_i^2 (often called D_i for individual variability at initial level) and σ_s^2 (often called D_s for individual variability in trajectories within the group), and their covariance, σ_{η_i, η_s} . What does all of this notation mean? The value of α_i tells us the initial level or intercept of the growth variable for the group. For the example, using frustration as our variable, α_i tells us the average level of frustration for the sample at the first time period, i.e., at the start. The value of α_s tells us the rate of change. If α_s is positive, this means the group is getting more frustrated over

time. If α_s is negative, this means the group is getting less frustrated over time. The variances tell us how much individuals in the group vary. A large value means the initial level, or rate of change, varies widely from person to person. Some individuals start at a high or a low level and some change much more than others do. Small variances mean that the group is homogeneous, starting at much the same point and changing at much the same rate. The covariance is included to allow for the intercept and slope to covary. This would happen if people who start at a low level of frustration increase more rapidly than people who start at a high level of frustration.³

We present this model in a path diagram form in Figure 1.1. Understanding Figure 1.1 is a big step toward understanding LGC modeling. The variables in rectangular boxes are the only variables you observe, frustration among caregivers. The y_1 is the level of frustration at time 1, when the caregivers start providing care; y_2 is the same variable measured at time 2, perhaps after 3 months of providing care. The time 3 variable y_3 is frustration measured at the next time (6-months) and y_4 is the frustration level measured at the fourth time (9-months). It is necessary to have the variable measured at least three times to estimate a

³ One can also view this linear model from a **multilevel perspective**. In this case, time is viewed as a Level 1 unit of analysis. That is, time becomes the context within which changes in that individual are observed. Level 1 is represented by equation 1.1. This tells us what is happening to the group by giving us their initial level and their rate of change. This may be all we want to know—do caregivers become more frustrated over time.

What happens to an individual may be quite different from what happens, on average, to the group. In the second level, individuals are considered the Level 2 units of analysis. We normally have two questions, what explains why some individuals start at a low or high level and what explains why some individuals change more or less than others. With two equations specified, one for the initial status (intercept) parameter and one for the growth and change trajectory parameter, we can answer both questions. The equations are

Level 2 model:

$$\eta_{ip} = \alpha_{\eta_i} + D_{\eta_{ip}} \quad \text{Individual initial level}$$

$$\eta_{sp} = \alpha_{\eta_s} + D_{\eta_{sp}} \quad \text{Individual slope}$$

where D is expressed as the deviation score from its mean, i.e., $\eta - \alpha$. Combining Level 1 and Level 2 models resulting in:

$$y_{pt} = (\alpha_{\eta_i} + D_{\eta_{ip}}) + a_t(\alpha_{\eta_s} + D_{\eta_{sp}}) + \epsilon_{pt}$$

and rearrangement of terms gives

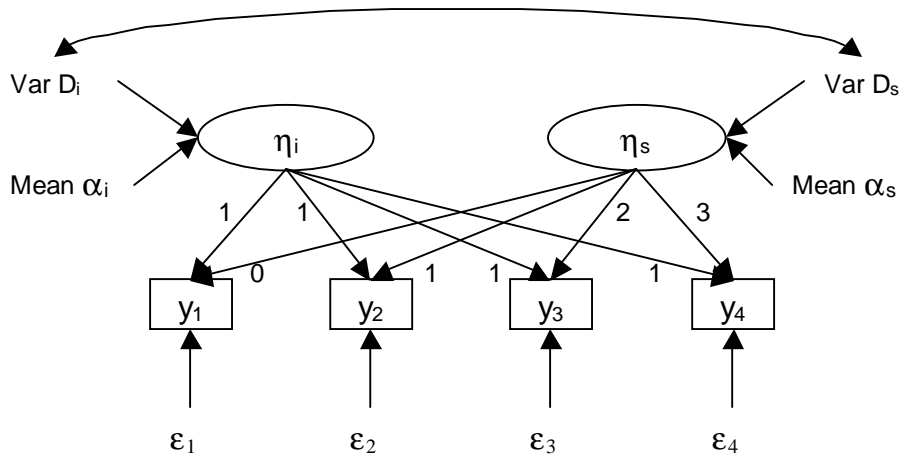
$$y_{pt} = (\alpha_{\eta_i} + a_t\alpha_{\eta_s}) + (D_{\eta_{ip}} + a_t D_{\eta_{sp}} + \epsilon_{pt})$$

The first term on the right side of the last equation represents the fixed component of growth (i.e., the group level characteristics of growth), the second term represents random component of growth (i.e., the individual level characteristics of growth).

Note that the model in the first two equations constitutes a two-level random coefficients model. It is also known as a fully unconditional trajectory model (Bryk & Raudenbush, 1992), as we are only examining the characteristics of growth/change (e.g., means, variances, covariance) but not attempting to predict growth/change using additional measures.

linear trend. Four measurement points allow us to estimate a quadratic trend. At the bottom of Figure 1.1, the ε_i represent measurement errors because it is reasonable to assume that our measurement of frustration is not perfectly reliable.

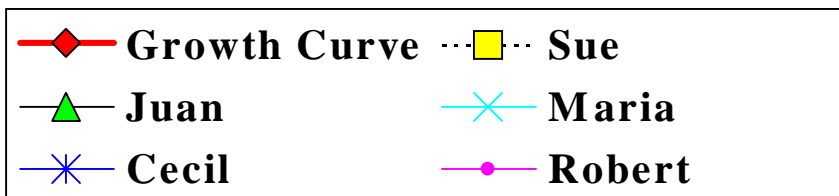
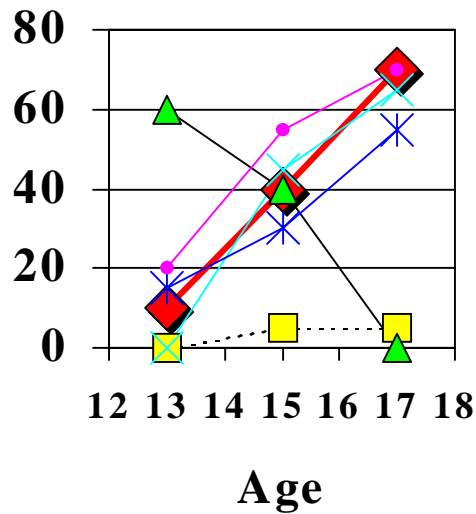
Figure 1.1--Conceptual Path Diagram



The η_i and η_s are enclosed in ellipses because they are latent variables (not directly observed like our four measures of frustration). The η_i represents the initial level of frustration (the intercept in equation 1.1). The η_s represents the linear trend, whether frustration increases or decreases with time (the slope in equation 1.1). The initial level and slope have a mean, $\text{Mean } \alpha_i$ for the intercept and $\text{Mean } \alpha_s$ for the slope. The mean tells us the average mean and slope for the group. At the individual level, each person may have a different intercept and slope.

Figure 1.2 illustrates the mean intercept and slope (the bold linear regression line) and the variance of several individual intercepts and slopes (the light linear regression lines). The bold line is the regression for the group and the five light lines are individual regression lines for five individuals.

Figure 1.2. Growth Curve and Individual Growth Curves



The LGC model in Equation 1.1 can be estimated through SEM-based software to provide maximum likelihood estimates of the growth model. This brings us to the next section, which summarizes SEM software to be used in this manual.

Concluding Comments

In Section 1, we have briefly reviewed the fundamentals of LGC, including its key parameters—latent means and variances. We did this in the context of a linear growth model. The remainder of the manual will be devoted to programming aspects of LGC using existing SEM software.

Relevant Work and Applications

Acock, A. C., and Li, F. 1999. Latent Growth Curve Analysis: A Gentle Introduction. Presented at the National Council on Family Relations, Irvine, CA.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). An introduction to latent variable growth

- curve modeling: Concepts, issues, and applications. Mahwah, NJ: Lawrence Erlbaum.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattell & J. Nesselroade (Eds.), Handbook of multivariate experimental psychology (2nd ed., pp. 561-614). New York: Plenum.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Development, *58*, 110-133.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. Psychometrika, *55*, 107-122.
- Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), Best methods for the analysis of change: Recent advances, unanswered questions, future directions (pp. 1-17). Washington, DC: American Psychological Association.
- Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, *92*, 726-748.
- Walker, Alexis J., Acock, A. C., Bowman, S. R., & Li, F. (1996). Amount of care given and caregiving satisfaction: A latent growth curve analysis. Journal of Gerontology: Psychological Sciences, *51B*:3. Pp 130-142.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychological Bulletin, *116*, 363-381.

SECTION 2: SOFTWARE

Each program has unique capabilities, but all of them can handle most applications of LGC. Programs are being updated at an increasingly fast pace and each generation emphasizes great ease of use by the analyst. Amos, EQS, LISREL can use path diagrams as model input/output. As such, conducting LGC can be done with mouse clicks instead of writing complex programming code (syntax). However, to avoid explaining each graphic interface and as a way of providing programming skills, this manual focuses on the traditional way of equation building methods using program-specific language. Readers may prefer to use the graphic interface of their chosen program, but understanding the implied programming code is useful.

Amos

Software. Amos (version 4) was written by James L. Arbuckle and distributed by SmallWaters Corporation, 1507 East 53rd Street, Suite 452, Chicago, IL 60615 (e-mail smallwaters@acm.org; url <http://www.smallwaters.com>).

The program does an excellent job with missing values using a full information approach, although other programs are implementing related procedures. Rigdon (1996) provides a software review of Amos. Duncan, Duncan, Stryker, Li, and Alpert (1999) provide detailed applications of Amos in the context of LGC.

Amos can be purchased as a stand alone program or as an add on to SPSS (url: www.spss.com). If purchased from SPSS, it can be started by clicking within the

SPSS menu. As with most other programs, the stand-alone versions can read SPSS, SAS, EXCEL, and many other types of datasets. There is a free student version. It is limited in how many variables can be analyzed.

Programming notation. Amos allows the user to specify a model and display parameter estimates under either Amos Basic or Amos Graphics modes. Problem specification is very flexible. For The LGC examples we use rely on the following notation:

- y , an observed (measured) variable,
- F , a latent variable such as the intercept or slope,
- E , an uniqueness associated with measurement of an observed variable,
- D , a residual associated with a latent variable such as the variance of the intercept or slope, and
- M is the mean of a mean of a latent variable such as the intercept or slope.

In Amos, relations between variables may be specified with Arrows formed from \rightarrow and \leftarrow signs, as in:

$y_1 \rightarrow y_2$ or $y_1 \leftrightarrow y_2$. Alternatively, relations may be specified with equations:

$$y_1 = y_2 + y_3 + (1)D_1,$$

where D_1 is the residual associated with y_1 . We could extend the equation by adding an extra pair of parentheses:

$$y_1 = (M) + y_2 + y_3 + (1)D_1.$$

EQS

Software. EQS (version 5.0) for Windows was written by Peter M. Bentler and Eric J. C. Wu and distributed by Multivariate Software, Inc., 4924 Balboa Blvd. #368, Encino, CA 9131 (e-mail eqs@netcom.com , URL <http://www.mvsoft.com>). Useful demonstrations of EQS applications includes work by Byrne (1994), Duncan et al. (1999), and Dunn, Everitt, and Pickles (1993). A student version is available at no charge from the WebPage. It is limited in the number of observed variables it can use.

Programing notation. Like Amos, EQS also uses straightforward command line. The EQS model specification consists primarily of five variable types:

- V , a measured variables,
- F , a latent variables or factors,
- E , a measured variable residuals or errors,

- D, a latent variable residuals disturbances (variance for LGC), and
- V999, a constant (the mean of the latent variables for the intercept and slope in our applications).

Users need only identify the V, F, E, and D variables, and their relations, within a regression framework. Note that LGC models are conceptualized in EQS as a series of regression equations involving dependent-observed (the Vs) and independent-latent (the Fs, V999, and Es) variables. For example, the following is the measurement equation for observed variable y1:

$$y1 = F1 + F2 + E1;$$

where F1 and F2 represent latent factors of intercept and slope, respectively, and E1 is the uniqueness for y1.

LISREL

Software. LISREL (version 8.30) was written by Karl G. Jöreskog and Dag Sörbom and distributed by Scientific Software International, Inc., 1525 East 53rd Street, Suite 530, Chicago, IL 60615 (e-mail techsupport@ssicentral.com; URL: <http://www.ssicentral.com/lisrel/mainlis.htm>). Useful demonstrations of LISREL application include work by Byrne (1989). A free student version can be downloaded from their WebPage. It is limited in the number of observed variables that can be used.

Programming notation. The current LISREL version (version 8.3) accepts two command languages in the input file: (a) LISREL and (b) SIMPLIS. The LISREL program is couched in matrix notation that is represented by Greek letters whereas the SIMPLIS uses plain language like Amos and EQS for SEM. To distinguish between these, we use the terms LISREL input and SIMPLIS input.

The specification of the LGC model involves only a portion of general LISREL models. In this manual, we shall use the LISREL model for the vector of endogenous variables Y in our programming. Accordingly, the following notation is used:

- y , designated as an observed variable
- η , designated as a latent variable (intercept and slope),
- ε , an uniqueness associated with measurement of an observed variable
- ψ , the variance of latent variable (intercept and slope), and
- α , a constant, the mean of the intercept and slope.

SIMPLIS, on the other hand, requires no complex equations, no Greek, and no matrix algebra. For SIMPLIS input, all that is required is to name all the

observed and latent variables and to formulate the model using a simple, regression-based command syntax that is similar to Amos or EQS. The model can be specified either as paths or as relationships (equations) in the input file or as a path diagram at run time. The SIMPLIS command language can be used with LISREL 8.3. In this manual, we shall use the following language notation for LGC:

- y , designated as an observed (measured) variable and
- Int and Slp designated as the latent factors of intercept and slope with latent means and variances.

Mplus

Software. Mplus (version 1.1) is a SEM program that replaces LISCOMP (Muthén, 1987). The program was written by Muthén and Muthén and distributed by Muthén & Muthén (Support@StatModel.com ; URL: <http://www.statmodel.com>). There is a demo version available that is limited in the number of observed variables you can use.

Programming notation. As in Amos and SIMPLIS, plain language can be used for Mplus programming. In this manual, we shall use the following language notation for LGC:

- y , designated as an observed (measured) variable,
- Int, is the latent variable of intercept,
- Slp is the latent variable of slope,
- E, an uniqueness associated with measurement of an observed variable,
- D, a residual associated with the prediction of latent variable (variance of the intercept or slope), and
- Const is designated as a constant, the mean of the intercept and slope.

Mx

Software. Mx (version x) was written by Neale M. C. (e-mail neale@gems.vcu.edu; URL <http://views.vcu.edu/mx>). The complete programs and documentation can be freely downloaded from the above WebPage. Like Amos, Mx can analyze a raw data structural equation model with missing data taken into account by using the maximum likelihood method. Hamagami (1997) provides a software review of Mx.

Programing notation. The Mx program allows input syntax that is similar to LISREL. This means that knowledge of LISREL will facilitate learning of the Mx program.

SAS@PROC CALIS

Software. SAS@PROC CALIS is a SAS System procedure that allows researchers to perform SEM analyses similar to those performed by many others. Information on SAS can be obtained through URL: <http://www.sas.com>. Hatcher (1994, 1996) provides an introduction to SAS PROC CALIS in SEM. At the time of this writing, the program has not been revised for several years. It is a part of the SAS System and there is no separate charge for it.

Programing notation. SAS PROC CALIS allows users to specify a model with either a regression-based command syntax that is similar to the one used in EQS or COSAN (Fraser, 1988) command syntax structure. Additionally, users can specify the RAM command and translate path diagrams through McArdle's reticular action model (McArdle & McDonald, 1984). In this manual, the following symbols are used as standard notation in PROC CALIS:

- y , is an observed (measured) variable,
- F , designated as a latent variable,
- E , a residual associated with measurement of an observed variable,
- D , a residual associated with the prediction of latent variable (variance of the intercept and slope), and
- Intcpt is a constant.

Concluding Comments

Section 2 provided an introduction to the existing major SEM-based software. As can be seen, the language syntax varies from program to program. In the Section 3, we will take a close look at how to build the growth curve model described in Section 1.

Relevant Work on Software Applications

- Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer-Verlag.
- Byrne, B. M. (1994). Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming. Thousand Oaks, CA: Sage Publications.
- Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah, NJ: Lawrence Erlbaum.
- Dunn, G., Everitt, B., & Pickles, A. (1993). Modeling covariances and latent variables using EQS. London: Champman & Hall.
- Fraser, C. (1988). COSAN user=s guide. Unpublished documentation. Centre for Behavioral Studies in Education. Armindale, Australia.
- Hamagami, F. (1997). A review of the Mx computer program for structural equation modeling. Structural Equation

- Modeling, 2, 157-175.
- Hatcher (1994). A step-by-step approach to using the SAS System for factor analysis and structural equation modeling. Cary, NC: SAS Institute.
- Hatcher (1996). Using SAS@ PROC CALIS for path analysis: An introduction. Structural Equation Modeling, 3, 176-192.
- McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. British Journal of Mathematical and Statistical Psychology, 37, 234-251.
- Rigdon, E. E. (1996). Amos 3.51. Structural Equation Modeling, 3, 87-91.

SECTION 3: Latent Curve Analysis: Amos Program Code for A Two-Factor Linear Model

This section provides program codes (command files) of the SEM software for the model depicted in Figure 1.1 in which the y variable is adolescent alcohol use at ages 12, 13, 14, and 15. The programming code is provided, followed by a path diagram showing the naming convention of the particular program, and selected output. A full information based approach using the maximum likelihood estimator is used to obtain parameter estimates. A brief interpretation of these parameter estimates is then given. We provide detailed information on Amos 4.0 and the material for the other programs appears in Appendix A.

Data selected for this two-factor model are from a longitudinal study (the Young Adult Substance Use: Predictors and Consequences (Hops, Andrew, Tildesley, & Duncan, 1998). Measures of alcohol use taken at ages 12 (in 1986) through 15 (in 1989) are used in this example. Missing data are excluded to simplify the presentation, resulting in a sample size of 118. The raw data set is available at <http://www.ori.org/~fuzhongl/data/home.htm>. The raw data is not necessary. All that is necessary is the covariances, means, and number of cases. This appears in Table 3.1 as an Excel file with the means, N of cases, and covariances. This Excel file can be read directly by Amos 4.0. The file is called manual.xls (its Excel name), and this matrix is entered like this on the page with a tab called sec3. This way of using Excel is useful. You can have multiple datasets in a single Excel file with the tabs used to label each of the datasets. This is useful for organizing a complex problem that involves using multiple datasets.

Table 3.1. An Excel File, manual.xls, containing data for two factor LGC

rowtype_	varname_	y1	y2	y3	y4
n		118	118	118	118
cov	y1	0.855			
cov	y2	0.473	0.82		
cov	y3	0.408	0.524	0.871	
cov	y4	0.3	0.446	0.496	0.761
mean		1.788	2.102	2.347	2.737

As we shall deal with outcome variables that are continuous in nature, the normal theory based maximum likelihood estimator is used for all illustrations. The estimator is often the default estimation method in most software programs and is available in all the software used in this manual.

Amos

The following is the Amos program code (input file) for a two-factor LGC model. Amos code for a linear LGC model with four observed indicators⁴

⁴ Version 4.0 introduced a new program code. Prior versions of Amos would use the following programming code:

```
$Input variables
```

```
y1
y2
y3
y4
```

```

Sub Main
  Dim Sem As New AmosEngine

  Sem.TableOutput
  Sem.Standardized

  Sem.BeginGroup "Manual.xls", "Sec3"
    Sem.Structure "y1 = (1)F1 + (0)F2 + (1)E1"
    Sem.Structure "y2 = (1)F1 + (1)F2 + (1)E2"
    Sem.Structure "y3 = (1)F1 + (2)F2 + (1)E3"
    Sem.Structure "y4 = (1)F1 + (3)F2 + (1)E4"
    Sem.Structure "F1=(M_Int) + (1)D1"
    Sem.Structure "F2=(M_Slp) + (1)D2"
    Sem.Structure "D1 <--> D2"
  Sem.FitModel
End Sub

```

This input file provides the Amos 4.0 AmosBasic code. This may look strange to people who have not worked with Visual Basic or similar programs, but it is easy to learn using the Amos 4.0 manual.

- The first line, beginning with "DIM . . ." establishes that Amos is being used by the program.
 - The Sem.TableOutput determines that a tabular output will be presented (TextOutput is the alternative).
 - The next line asks for standardized output; unstandardized output is automatic.
 - The line "Sem. Begin Group . . ." tells us the data is in an excel file, "manual.xls"
-

```

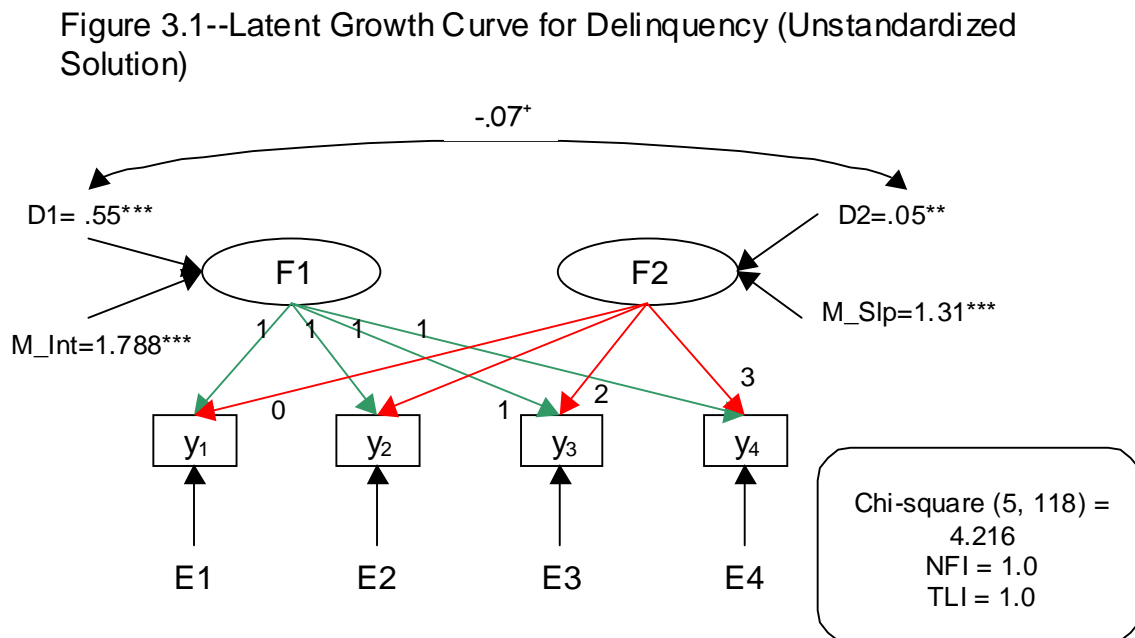
$Sample size = 118
$Covariances
0.855
0.473    0.820
0.408    0.524    0.871
0.300    0.446    0.496    0.760
$Means
1.788    2.102    2.347    2.737
$Standardized
$Structure
y1= (1)F1 + (0)F2 + (1)E1
y2= (1)F1 + (1)F2 + (1)E2
y3= (1)F1 + (2)F2 + (1)E3
y4= (1)F1 + (3)F2 + (1)E4
F1=(M_Int)+(1)D1
F2=(M_Slp)+(1)D2
D1 (Int)
D2 (Slp)
D1 <> D2 (Cor)

```

(see Table 3.1) and is in a worksheet called "Sec3" (you will see the tab with this name at the bottom of the table).

- Sem.Structure lines provide equations to represent the paths in the figure.
- The equation for y_1 consists of $(1)F1 + (0)F2 + (1)E1$. The values in parentheses are fixed for identification purposes. This is repeated for each y_t changing the coefficient for the slope, F2. The path from the error term, E1, to y_1 is fixed at 1. This is not shown in the Figure to simplify its appearance, but reflects the fact that y_1 contains all the error variance in E1.
- The equations for the intercept and slope, F1 and F2, follow. F1 depends on its mean and variance. The mean is an intercept and has no coefficient attached to it. The $1(D1)$ means that all of the individual variance goes into F1.
- The F2 equation is confusing using Amos notation. The mean slope (M_Slp) is an intercept in this equation and is shown as an intercept in the output. Don't be confused, this is the mean slope.
- The last line of code provides for the variance in the intercept (D1) and the variance in the slope (D2) to be correlated.

This code can be adapted to any LGC involving a linear growth. The results of running the code produces the growth model depicted in a path diagram form as shown in Figure 3.1.



Selected Amos output. The following provides selected output from Amos run.

Chi-square = 4.213
 Degrees of freedom = 5
 Probability level = 0.519

Regression Weights

			Estimate	S.E.	C.R.	P	Label
y1	<--	F1	1.00				
y1	<--	F2	0.00				
y2	<--	F1	1.00				
y2	<--	F2	1.00				
y3	<--	F1	1.00				
y3	<--	F2	2.00				
y4	<--	F1	1.00				
y4	<--	F2	3.00				

Intercepts

		Est	S.E.	C.R.	P	Label
	F1	1.78	0.08	21.82	0.00	M_Int
	F2	0.31	0.03	10.61	0.00	M_Slp

Covariances

		Est	S.E.	C.R.	P	Label
D1	<--> D2	-0.07	0.03	-1.93	0.05	

Correlations

		Est
D1	<--> D2	-0.41

Variances

	Est	S.E.	C.R.	P	Label
D1	0.55	0.11	5.05	0.00	
D2	0.05	0.02	2.68	0.01	
E1	0.35	0.08	4.14	0.00	
E2	0.30	0.05	5.66	0.00	
E3	0.36	0.06	6.09	0.00	
E4	0.21	0.07	2.84	0.00	

Interpretation. The overall model fit statistics is $\chi^2(5, N = 118) = 4.213, p = .519$. The RMSEA = .000, with a 90% confidence interval of .000 to .120. Amos provides various model fit indices such as commonly used comparative fit (CFI) index, Tucker Lewis Index (TLI), and the normed fit index (NFI). The value of

these indices for this model are high (CFI = 1.000, TLI = 1.0, and NFI = 1.000). In short, these statistics suggests an excellent fit of the model to the data.

Next, the focus is on the growth parameter estimates generated from the maximum likelihood estimator. The latent mean for the intercept is 1.78 ($p < .001$). This appears in the output under the "Intercepts" section for F1. The latent mean for the slope is .31 ($p < .001$). This also appears under the "Intercepts" section as the intercept for F2. Although this is defined as an intercept in the program code (Sem.Structure "F2=(M_slp) + 1 D2"), it really is the mean slope. The intercept indicates a statistically significant mean alcohol use at the initial level (i.e., at the age of 12) and the slope mean indicates a significant average increase, via a linear functional form, in the use of alcohol use during the four-year period. This alcohol use in adolescents is expected to increase by .31 each year, beginning with an average score of 1.78. The substantive significance of these values requires considering the scale on which alcohol use was measured.

We can get an expected value for alcohol use for each year. To obtain the expected mean, $E(M)$, for each measurement year, we have the following:

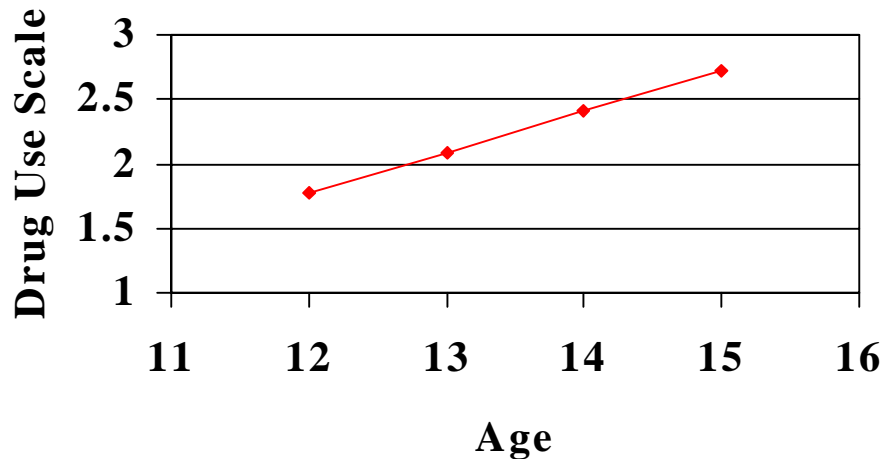
$$\begin{aligned}
 E(M_t) &= E(M_{int}) + E(M_{slp})a_t \\
 E(M_{1986}) &= 1.778 + (.313 \times 0) = 1.778 \\
 E(M_{1987}) &= 1.778 + (.313 \times 1) = 2.091 \\
 E(M_{1988}) &= 1.778 + (.313 \times 2) = 2.404 \\
 E(M_{1989}) &= 1.778 + (.313 \times 3) = 2.717
 \end{aligned}
 \tag{3.1}$$

These values can be obtained directly from Amos by adding one line to the program, `Sem.ImpliedMoments`. This gives the following output:

Implied Means

y4	y3	y2	y1
2.717	2.404	2.091	1.778

These predicted values are plotted (see Figure 3.2) and they show a linear increasing trend in adolescent alcohol use over the four-year measurement period.

Figure 3.2. Drug Usage LGC

The next question concerns the extent to which adolescents in the sample vary around their group average (mean) trajectories in alcohol use. This can be evaluated by looking at the variances (or equivalently, the standard deviations). Their corresponding variances ($D1 = .55$ and $D2 = .05$) are statistically significant, indicating significant individual variability in the initial level and rate of change (growth) in alcohol use across the four waves of measurement. The covariance ($D1 \leftrightarrow D2 = -.07$) or correlation ($-.04$) between the two latent factors is marginally significant at the two-tail level, $p < .10$,⁵ indicating the initial level and rate of change over time are slightly related. The correlation between the intercept and slope can be an important, but is often ignored. In this example, the negative correlation of $-.04$ means that adolescents who have higher initial alcohol use (intercept) have a slightly weaker rate of increase (slope). Each error term ($E1$ through $E4$) is also significant ($p < .05$). These reflect the unique portion of time-specific variance that is not accounted for by the model. The estimates are not shown in Figure 3.1 to simplify the presentation.

Similar to the mean prediction, predictions on adolescent variation around the average means over time can be calculated. The expected variance reflecting individual difference can be written as:

⁵ Inspecting the output, Amos reports the significance as $p = .05$. Since the critical ratio is slightly less than the 1.96 needed for $p = .05$, we assume Amos is rounding. This is misleading since p is greater than $.05$, so I used the $p < .10$ probability in the text.

$$\begin{aligned}
E(V_t) &= E[(Y_{pt} - M_t) \times (Y_{pt} - M_t)'] \\
&= E[(\eta_{ip} + a_t \eta_{sp} + \varepsilon_{pt})(\eta_{ip} + a_t \eta_{sp} + \varepsilon_{pt})'] \\
&= V_{\eta_{ip}} + a_t^2 V_{\eta_{sp}} + 2a_t \text{Cov}(\eta_{ip}, \eta_{sp}) + V(\varepsilon_{pt})
\end{aligned}$$

where,

$E(V_t)$ is the estimated variance at time t

Y_{pt} is an individual's score at time t

(3.2)

M_t is the mean at time t

a_t is the coefficient (0, 1, 2, 3)

$V_{\eta_{ip}}$ is the variance of the intercept

$V_{\eta_{sp}}$ is the variance of the slope

$\text{Cov}(\eta_{ip}, \eta_{sp})$ is the covariance of the intercept and slope

$V(\varepsilon_{pt})$ is the error variance, E_t , for y at time t

Note that the distribution of the uniqueness variances can be heteroscedastic over time. Inserting the variance values from the Amos output we obtain a different expected variances for each year:

$$E(V_{1986}) = (.55) + (0)^2 \times .05 + (2 \times 0 \times (-.07)) + .35 = .90$$

$$E(V_{1987}) = (.55) + (1)^2 \times .05 + (2 \times 1 \times (-.07)) + .30 = .76$$

$$E(V_{1988}) = (.55) + (2)^2 \times .05 + (2 \times 2 \times (-.07)) + .36 = .83$$

$$E(V_{1989}) = (.55) + (3)^2 \times .05 + (2 \times 3 \times (-.07)) + .21 = .79$$

This shows little evidence of heteroscedasticity. The variance in the adolescents' usage decreases after the first year. However, the variance is stable for the next three years. In some applications, heteroscedasticity of the variance can be an interesting finding. For example, it would be interesting if adolescents were homogeneous at the start of the process but became increasingly heterogeneous over time. This was not the case for our data.

Comparison of SEM Program Estimations

The following provides a summary of parameter estimates in LGC across the various SEM software. As can be seen, the estimates are very close. Standard errors are given in parentheses.

Table 3.2. Comparison of Estimates from Six Programs

	<i>Amos</i>	<i>EQS</i>	<i>LISREL</i>	<i>Mplus</i>	<i>Mx</i>	<i>PROC CALIS</i>
<i>Mean intercept</i>	1.778 (.082)	1.778 (.082)	1.778 (.082)	1.778 (.081)	1.778 (.082)	1.778 (.082)
<i>Mean slope</i>	.313 (.030)	.313 (.030)	.313 (.030)	.313 (.029)	.313 (.030)	.313 (.030)
<i>Intercept (variance)</i>	.551 (.109)	.551 (.109)	.551 (.109)	.547 (.108)	.551 (.109)	.551 (.109)
<i>slope (variance)</i>	.047 (.018)	.047 (.018)	.047 (.018)	.047 (.017)	.047 (.018)	.047 (.018)
<i>Intercept, slope (covariance)</i>	-.067(.035)	-.067 (.035)	-.067 (.035)	-.066 (.034)	-.067 (.035)	-.067 (.035)
<i>Error variance (t1)</i>	.354 (.084)	.354 (.086)	.354 (.086)	.351 (.085)	.354 (.086)	.354 (.086)
<i>Error variance (t2)</i>	.303 (.054)	.303 (.054)	.303 (.054)	.300 (.053)	.303 (.054)	.303 (.054)
<i>Error variance (t3)</i>	.366 (.060)	.366 (.060)	.366 (.060)	.363 (.059)	.366 (.060)	.366 (.060)
<i>Error variance (t4)</i>	.211 (.074)	.210 (.074)	.210 (.074)	.209 (.073)	.210 (.074)	.210 (.074)

Concluding Comments

This Section 3 presented a LGC of a two-factor linear curve model. The use of other software is shown in Appendix A. The linearity in change/growth is defined by setting the basis vector to known values (0 1 2 3). This is one of many possible fixed vectors we could use to examine for a specific curve. Other options include

- using linear polynomial coefficients (i.e., -3 -1 1 3), Using this scheme makes the intercept parameter represent the average of the four measurement points rather than the first time point as most people do.
- specifying a quadratic trend with an extra factor using loadings (i.e., 0 1 4 9). This would result in two slope factors, one linear (0 1 2 3) and the other quadratic (0 1 4 9). If this model did better than the linear model, this would be evidence of a nonlinear growth curve. To estimate the quadratic effect we need at least four time points. The model including the quadratic effect would be compared to the linear model using the difference in chi-square and the difference in the degrees of freedom.
- estimating the basis vector to determined the shape of the curve (0 1 * *), where * indicates freely estimate loadings. The model is identified by fixing the first two parameters. This would allow the growth curve to go up or down at both the 3rd and 4th year.

On the programming aspects, users who are familiar with the matrix operation, both LISREL and Mx are excellent choices. Users who are comfortable with traditional multiple regression equation by equation modeling method may find programs such as EQS, Amos, SAS@PROC CALIS easier to use.

Relevant Work and Useful Applications

- MaCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*, 215-253.
- McArdle, J. J. (1998). Modeling longitudinal data by latent growth curve methods. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 359-406). Mahwah, NJ: Lawrence Erlbaum.
- McArdle, J. J., & Hamagami, E. (1991). Modeling incomplete longitudinal data using latent growth structural equation models. In L. Collins & J. L. Horn (Eds.), *Best methods for analysis of change* (pp. 276-304). Washington, DC: American Psychological Association.
- Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research*, *29*, 263-288.
- Walker, Alexis J., Acock, A. C., Bowman, S. R., & Li, F. (1996). Amount of care given and caregiving satisfaction: A latent growth curve analysis. *Journal of Gerontology: Psychological Sciences*, *51B*:3. Pp 130-142.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363-381.
-

SECTION 4: Latent Curve Analysis with Covariates

The estimation of a latent growth curve is a widely used procedure and many studies stop at this point. You may want to go a step farther and try to explain the

growth process. There are two questions. First, what variables explain the initial level of the process (intercept). Do adolescents who have a low initial level of delinquency have a low level of family conflict? Parents with more education? A strong social support network? The research question is what explains variance in the initial level of the process. This question could be answered by cross sectional data without doing latent growth curve analysis. We could use a sample of 12 year olds and use the protective factors as predictors of their level of delinquency.

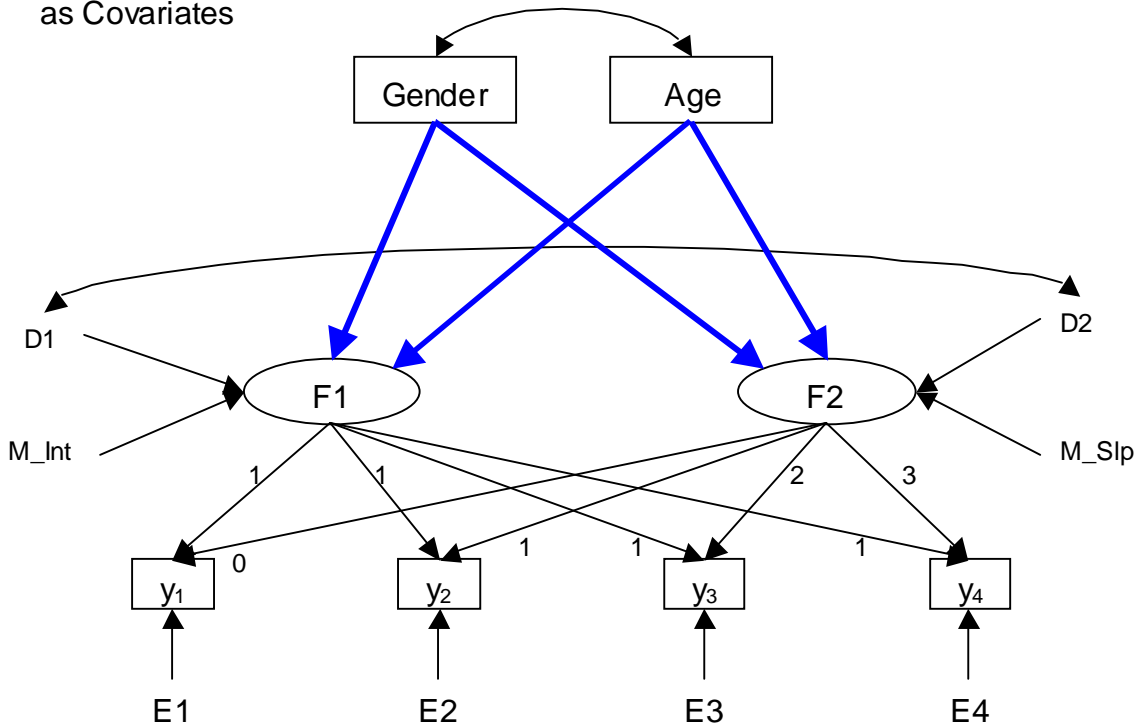
The second question is what variables explain the growth process. Why do some adolescents rapidly develop a high level of delinquency? Why do others have no increase or even a decline in their delinquency? Some of the same variables that explain the initial level of the process, may also explain the growth. Perhaps, adolescents with strong social support systems have less increase in delinquency. Perhaps, having a highly educated mother mitigates the growth in delinquency. Similarly, a low level of family conflict may be important.

Explaining change is a major strength of latent growth analysis. Explaining change is also a theoretically pivotal goal of most theoretical orientations. We are at least as interested in how things change as we are in where they are located at any particular point.

The dependent variable for both sets of questions is NOT what we are use to having. Instead of having delinquency be the dependent variable, we have regression coefficients being the dependent variables. The regression intercept is the dependent variable for the first set of questions. The regression slope is the dependent variable for the second set of questions. The independent variables are predicting the intercept and slope rather than the variable, *per se*. Once we understand this distinction, the programming and interpretation are not especially difficult.

Empirical Example. Our empirical example involves adolescent smoking behavior. We are interested in estimating the latent growth curve for adolescent smoking behavior. Then, we are interested in explaining it. We propose that females are less likely to smoke initially and have a lower growth rate than males. We also propose that the older the adolescent is at the start of the study the higher their initial smoking, but that age at the start of the study has no effect on the growth curve. Our research hypotheses are fairly simple in this example, but the procedures we use can be applied to any research questions you have that fit this model. Our model appears in Figure 4.1.

Figure 4.1--Latent Growth Curve for Cigarette Use with Gender and Age as Covariates



The new features of this model are highlighted by the bold blue lines. Gender and age directly effects both the intercept and slope.

Our data are 275 adolescents who were measured on cigarette usage at four consecutive years. The adolescents varied from 13 to 16 the first year of assessment. Gender is coded 1 for male, 0 for female; cigarette use is coded 1 for never used, 2 for used prior to past 6 months, 3 for current use less than 4 times a month, 4 for current use between 4 and 29 times a month, and 5 is current use of 30 or more times a month. Table 4.1 presents our data in an Excel spreadsheet. Notice that this is the same file we used for the first example, manual.xls, but this data is in the tab labeled sheet2:

Figure 4.2. Excel Spreadsheet

rowtype	varname_	cigs1	cigs2	cigs3	cigs4	gender	age
n		275	275	275	275	275	275
cov	cigs1	1.776					
cov	cigs2	1.624	1.98				
cov	cigs3	1.451	1.604	1.885			
cov	cigs4	1.265	1.46	1.556	1.899		
cov	gender	-0.0373	-0.00897	-0.00982	-0.03554	0.245	
cov	age	0.47	0.452	0.404	0.308	-0.0347	1.068
mean		2.2582	2.4655	2.5273	2.6945	0.4255	0.14236

The Amos code follows:

```

Sub Main
Dim Sem As New AmosEngine

Sem.TableOutput
Sem.Standardized

Sem.BeginGroup "Manual.xls", "Sheet2"
Sem.Structure "cigs1 = (1)F1 + (0)F2 + (1)E1"
Sem.Structure "cigs2 = (1)F1 + (1)F2 + (1)E2"
Sem.Structure "cigs3 = (1)F1 + (2)F2 + (1)E3"
Sem.Structure "cigs4 = (1)F1 + (3)F2 + (1)E4"
Sem.Structure "gender = (M_Gender) + (1)E5"
Sem.Structure "age = (M_Age) + (1)E6"
Sem.Structure "F1=(M_Int) + gender + age + (1)D1"
Sem.Structure "F2=(M_Slp) + gender + age + (1)D2"
Sem.Structure "D1 <--> D2"
Sem.Structure "E5 <--> E6"

```

Sem.FitModel

End Sub

This code is almost the same as the code Amos code used in Section 3. The covariance matrix appears in Sheet2 of the Excel file. We defined gender as "gender = (M_gender) + (1)E5" and age as "age=(M_age) + (1)E6." We need the means here because we are analyzing a moment matrix. Thus, E5 is simply the variance in gender and E6 is the variance in age. Although we have not reason for gender and age to be correlated, we have provided for this possibility using "E5 <-> E6." Finally, we added the gender and age variables as predictors of the intercept and slope ["F1=(M_Int) + gender + age + (1)D1" and "F2=(M_Slp) + gender + age + (1)D2"].

Amos Output

Computation of degrees of freedom

```

Number of distinct sample moments = 27
Number of distinct parameters to be estimated = 18
Degrees of freedom = 27 - 18 = 9

```

Minimum was achieved

```

Chi-square = 13.807
Degrees of freedom = 9
Probability level = 0.129

```

Regression Weights:	Estimate	S.E.	C.R.
F1 <----- age	0.446	0.074	6.045
F1 <----- gender	-0.064	0.154	-0.418
F2 <----- age	-0.046	0.020	-2.267
F2 <----- gender	0.008	0.042	0.183
cigs1 <----- F1	1.000		
cigs1 <----- F2	0.000		
cigs2 <----- F1	1.000		
cigs2 <----- F2	1.000		
cigs3 <----- F1	1.000		
cigs3 <----- F2	2.000		
cigs4 <----- F1	1.000		
cigs4 <----- F2	3.000		

Intercepts

	Estimate	S.E.	C.R.	P	Label
gender	0.43	0.03	14.26	0.00	M_Gender

age	0.14	0.06	2.28	0.02 M_Age
F1	2.23	0.10	22.04	0.00 M_Int
F2	0.15	0.03	5.23	0.00 M_Slp

Covariances

		Estimate	S.E.	C.R.	P
E5	<--> E6	-0.03	0.03	-1.12	0.26
D1	<--> D2	-0.11	0.03	-3.35	0.00

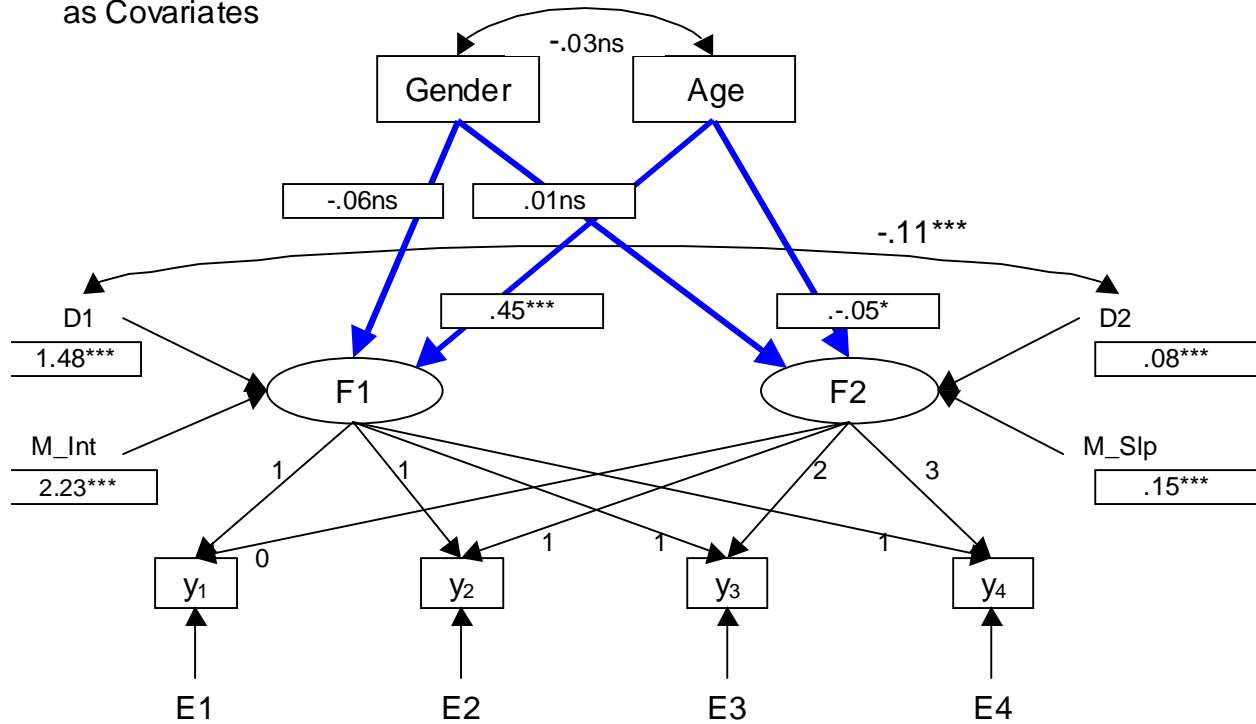
Correlations

		Estimate
E5	<--> E6	-0.07
D1	<--> D2	-0.30

Variances

	Estimate	S.E.	C.R.	P
E5	0.24	0.02	11.70	0.00
E6	1.06	0.09	11.70	0.00
D1	1.48	0.14	10.66	0.00
D2	0.08	0.01	6.32	0.00
E1	0.11	0.05	2.41	0.02
E2	0.32	0.04	8.92	0.00
E3	0.30	0.04	8.36	0.00
E4	0.29	0.06	4.98	0.00

Figure 4.3--Latent Growth Curve for Cigarette Use with Gender and Age as Covariates



The interpretation of this model is straightforward. The model provides a very good fit to the data with a chi-square (9, N=275) = 13.807; $p = .129$ and all goodness of fit measures at an acceptable level.⁶ The initial level of cigarette use is 2.23 and it increases at a rate of .15 per year. Gender is not important. Boys neither start at a higher rate of cigarette use, nor increase at a higher rate than girls. Adolescent cigarette use is not a gender issue. Age is important for both the intercept and the slope. The older the adolescent when the study started, they higher their cigarette usage (.45; $p < .001$). The older the adolescent was at the start of the study, however, the less their cigarette usage increases over time. This suggests that if we looked at cigarette usage over the age range of these adolescents, starting when they were 12 and continuing until they were 20, there would be a nonlinear relationship.

⁶ This model was estimated using a covariance matrix as the input. The covariance matrix was computed using SPSS and the covariance matrix and means were put in an Excel file using three decimal places. When this model is estimated directly on the raw data, Amos is able to keep many more significant figures and produces more accurate results. To give you an idea of the difference this makes, the chi-square when Amos is applied to the raw data is 11.78. This does not change any conclusions and the parameter estimates are very similar both ways, but it is probably better to always analyze raw data, when this is practical.

APPENDIX A: Two Factor Latent Growth Curves Using other Software

EQS

The following is the EQS program code for the two-factor linear curve model.

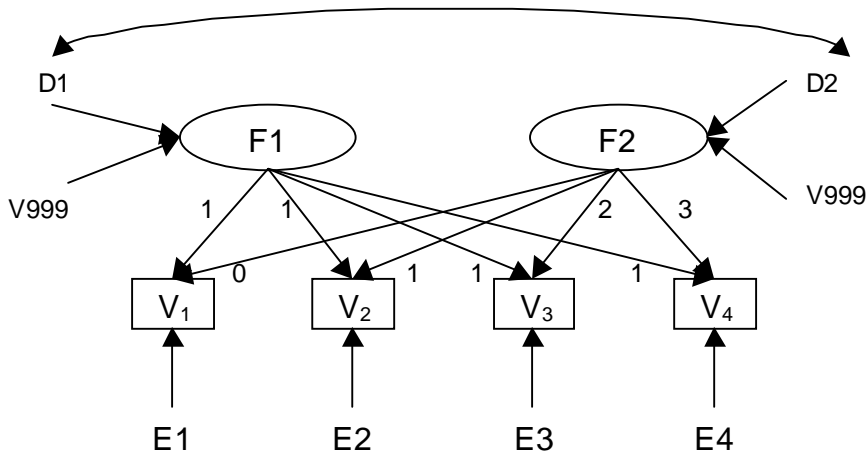
```

/TITLE
EQS code for a linear LGC model with four observed indicators
/SPECIFICATIONS
Cases=118; Variable=4; Method=ml; Matrix=cov; Analysis=moment;
/Labels
V1=y1; V2=y2; V3=y3; V4=y4;
F1=Int; F2=Slp;
/EQUATIONS
V1=F1+0F2+E1;
V2=F1+1F2+E2;
V3=F1+2F2+E3;
V4=F1+3F2+E4;
F1=*V999+D1;
F2=*V999+D2;
/VARIANCES
E1=*; E2=*; E3=*; E4=*;
D1=*; D2=*;
/COVARIANCE
D1,D2=*;
/Matrix
0.855
0.473    0.820
0.408    0.524    0.871
0.300    0.446    0.496    0.760
/Mean
1.788    2.102    2.347    2.737
/PRINT
Fit=all;
/END

```

The observed Y variables are labeled as Vs and the latent factors of the intercept and slope are labeled as Int and Slp. For LGC, a moment matrix is analyzed as indicated by "Analysis=moment". Notice that the means as well as the covariances are entered as the data. The model construction begins with EQUATIONS. The "*"s" indicates the parameter is freely estimated in the model. The line Fit=all prints out various fit statistics including chi-square statistic. The corresponding LGC model in a path diagram is shown in Figure A.1.

Figure A.1--Latent Growth Curve for Delinquency (EQS Notation)



Selected EQS output. The selected output for this EQS run is provided below.

```

CHI-SQUARE =          4.204 BASED ON          5 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS          0.52037
INT   =F1   =          1.778*V999  + 1.000 D1
          .082
          21.731
SLP   =F2   =          .313*V999  + 1.000 D2
          .030
          10.564
    
```

VARIANCES OF INDEPENDENT VARIABLES

E		D	
C		---	
E1-Y1	.354	D1-INT	.551
	.086		.109
	4.136	D2-SLP	5.052
E2-Y2	.303		.047
	.054		.018
	5.659		2.681
E3-Y3	.366		
	.060		
	6.090		
E4-Y4	.210		
	.074		
	2.839		

COVARIANCES AMONG INDEPENDENT VARIABLES

E		D	
C		---	
D2	- SLP		-.067

```

          D1 - INT          .035
                        -1.935
CORRELATIONS AMONG INDEPENDENT VARIABLES
-----

```

```

          E          D          ---
          C
          D2 - SLP          -.414
          D1 - INT

```

Interpretation. The overall model fit statistics can be found in the top portion of the selected output. This is followed by the parameter estimates. Since the basic interpretation is the same as in the Amos run, that can be used as a guide for interpretation of the output.

LISREL

LISREL code for a linear LGC model with four observed indicators is

```

LISREL code for a linear LGC model with four observed indicators
Da Ni=4 No=118 Ma=cm
Cm
0.855
0.473    0.820
0.408    0.524    0.871
0.300    0.446    0.496    0.760
Mean
1.788    2.102    2.347    2.737
La
y1 y2 y3 y4
Model Ny=4 Ty=ze Ne=2 Te=sy,fi al=fr be=ze ps=sy,fr
Le
Int Slp
va 1 ly(1,1) ly(2,1) ly(3,1) ly(4,1)
va 0 ly(1,2)
va 1 ly(2,2)
va 2 ly(3,2)
va 3 ly(4,2)
fr te(1,1) te(2,2) te(3,3) te(4,4)
Ou nd=3

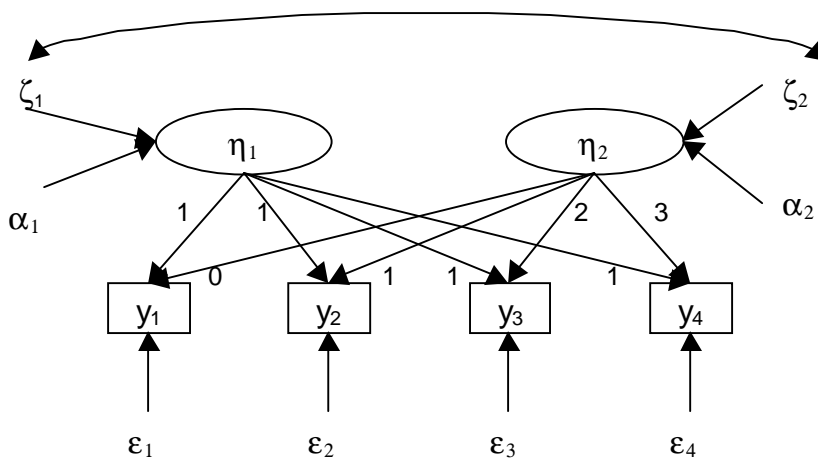
```

The model construction starts with the line "Model." LISREL uses matrices to define parameter specification. In this model, we have four observed variables as indicated by $N_y = 4$. T_y (τ_y) is a vector of observed intercepts and is specified as zero for identification purposes (because we are interested in latent means). Two endogenous latent factors (η_1 and η_2) are listed ($N_e = 2$) and they are labeled as Int and Slp after the "Le" line. The Ly elements are fixed at the values in the "va"

lines. Because the model does not involve any exogenous predictors the Beta matrix (BE) is fixed (i.e., a null matrix). The ψ matrix involving the latent factor variances (ψ_{11} and ψ_{22}) and covariance (ψ_{21}) is specified as symmetric and free (ps=sy,fr). The latent means for η_1 and η_2 are parameterized in the α vector. The uniqueness variances (ϵ_{τ}) are specified as symmetric and fixed (Te = sy,fi) but the diagonal elements are free as indicated by the line fr te(1,1) te(2,2) te(3,3) te(4,4).

In the output line (Ou), we have nd=3 requesting three decimal. In a path diagram form, this model is show in Figure A.2.

Figure A.2--Latent Growth Curve for Delinquency (LISREL Notation)



For those familiar with LISREL notation, we present the equations in the form of

- as a series of regression equations,
- in matrix form (η 's and ϵ 's as vectors), and
- in expanded matrix form, for the model shown in Figure A.2 and the LISREL code.

The regression equations are:

$$\begin{aligned}
 y_1 &= \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \epsilon_1 \\
 y_2 &= \lambda_{21}\eta_1 + \lambda_{22}\eta_2 + \epsilon_2 \\
 y_3 &= \lambda_{31}\eta_1 + \lambda_{32}\eta_2 + \epsilon_3 \\
 y_4 &= \lambda_{41}\eta_1 + \lambda_{42}\eta_2 + \epsilon_4
 \end{aligned}
 \tag{5.1}$$

where $\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41} = 1$ for η_1 and $\lambda_{12}, \lambda_{22}, \lambda_{32}$, and λ_{42} are set to 0, 1, 2, 3, respectively, for η_2 . In matrix algebra we have:

$$Y = \Lambda\eta + \varepsilon \tag{5.2}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \tag{5.3}$$

Selected LISREL output.

The selected output for the LISREL analysis is listed below.

CHI-SQUARE WITH 5 DEGREES OF FREEDOM = 4.204 (P = 0.520)

PSI

	Int	Slp
	---	----
Int	0.551 (0.109) 5.052	
Slp	-0.067 (0.035) -1.935	0.047 (0.018) 2.678

THETA-EPS

	y1	y2	y3	y4
	-----	-----	-----	-----
	0.354 (0.086) 4.137	0.303 (0.054) 5.658	0.366 (0.060) 6.089	0.211 (0.074) 2.843

ALPHA

	Int	Slp
	-----	-----
	1.778 (0.082) 21.728	0.313 (0.030) 10.565

SIMPLIS

The following is the SIMPLIS program code. Because the SIMPLIS produces identical estimates as LISREL. Its output is not presented.

Title: SIMPLIS code for a linear LGC model with four observed indicators
 Observed Variables: y1 y2 y3 y4
 Covariance Matrix:
 0.855

```

0.473    0.820
0.408    0.524    0.871
0.300    0.446    0.496    0.760
Means:
1.788    2.102    2.347    2.737
Latent Variables: Int Slp
Sample size = 118
Relationships:
    y1 =1*Int
    y2 =1*Int
    y3 =1*Int
    y4 =1*Int
    y1=0*Slp
    y2=1*Slp
    y3=2*Slp
    y4=3*Slp
Int = Const
Slp = Const
Number of decimals = 3
LISREL output
End of Problem

```

Mplus

Mplus reads the mean vector and covariance matrix through an external file. As shown below, the file to be read in is named as *Acovm.dat* in this example. The means and covariance matrix (in the file of *covm.dat*) for Mplus looks like the following:

```

1.788    2.102    2.347    2.737
0.855
0.473    0.820
0.408    0.524    0.871
0.300    0.446    0.496    0.760

```

The first line contains the vector of means (y_1 through y_4) followed by the covariance matrix. Users can copy the matrix and the mean vector into a notepad along with the program code provided.

```

Title: Mplus code for a linear LGC model with four observed indicators
Data: File is covm.dat;
      Type is covariance means;
      Nobservations = 118;
Variable: Names are y1 y2 y3 y4;
Analysis: Type = meanstructure;
Model:   Int by y1@1 y2@1 y3@1 y4@1;

```

```

Slp by y1@0 y2@1 y3@2 y4@3;
      [y1-y4@0];
      [Int Slp];
Output: samp standardized;

```

The types of data is specified as a "covariance means" because both the covariance matrix and the means must be included. The observed Y variables are given in the "Variable" line. The "Type" in the "Analysis" line indicates the model contains a mean structure. The model construction starts with the command line Model. The latent variable of intercept (Int) is defined by the four observed indicators with fixed loadings (the symbol @ indicates the parameter is fixed at the value indicated). The slope factor is defined by the known values specifying a linear growth function (0, 1, 2, and 3). The intercepts of the observed variables at the four time points are fixed to zero by the "[y1-y4@0];" line, although the growth factor means (for the intercept and slope) are estimated. Variables in square brackets refer to the intercepts or means of these variables. The output line requests descriptive statistics (samp) and standardized estimates (standardized).

Selected Mplus output. The selected output of the Mplus program is provided below.

```

Chi-Square Test of Model Fit
  Value                4.252
Degrees of Freedom      5
  P-Value                .514
Means
INT                    1.778 .081    22.429
SLP                    .313  .029    10.413
Variances
INT                    .547   .108    5.366
SLP                    .047   .017    2.674
SLP WITH
INT                    -.066   .034   -2.109
Residual Variances
Y1                     .315   .085    2.857
Y2                     .300   .053    6.094
Y3                     .363   .059    5.705
Y4                     .209   .073    4.166

```

Mx

The following is the Mx program. This is a powerful program and it is free, but it is more complicated to use than the other programs.

```

Title: Mx code for a linear LGC model with four observed indicators
Data NGroups=1 Ninput_vars=4 Nobs=118
Labels y1 y2 y3 y4
CMatrix Symm
0.855
0.473    0.820
0.408    0.524    0.871
0.300    0.446    0.496    0.760
Means
1.788    2.102    2.347    2.737
matrices
A full 4 2 fix
U Symm 4 4
M full 2 1 free
D Symm 2 2 free
R Stan 2 2 free
Mean A*M /
Compute A*D*R*D'*A' + U /
Matrix A    !Fixing A matrix with constant values
1. 0
1. 1.
1. 2.
1. 3.
Labels Row M
Inter_M Slope_M
Specify M !specifying latent mean vector
1 2
Matrix M !starting values
1.29 .06
Labels Row D
Int_Var Slo_Var
Labels Col D
Int_Var Slo_Var
Specify D !specifying latent variance matrix
3
0 4
Matrix D !starting values
1.
0 1.
Specify U !Specifying the error term matrix
7
0 8
0 0 9
0 0 0 10
Matrix U !Providing starting value for error terms
1.
0 1.
0 0 1.
0 0 0 1.
Bound .001 10 D 1 1 D 2 2
Labels Col A Int Slp
Option Se
End Group;

```

Selected Mx output. The following presents the portion of the Mx output from the input command above.

```

Chi-squared fit of model >>>>>>      4.208
Degrees of freedom >>>>>>>>>>>>      5
Probability >>>>>>>>>>>>>>>>>>>>  0.520
RMSEA >>>>>>>>>>>>>>>>>>>>>>>>  0.000
MATRIX M
This is a FULL matrix of order      2 by      1
      1
  INT_M      1.7780
  SLP_M      0.3132
MATRIX M (Rough) STANDARD ERRORS
This is a FULL matrix of order      2 by      1
      1
  INT_M      0.0820
  SLP_M      0.0298
MATRIX R
This is a STANDARDISED matrix of order      2 by      2
      1      2
  1      1.0000
  2     -0.4140      1.0000
MATRIX R (Rough) STANDARD ERRORS
This is a STANDARDISED matrix of order      2 by      2
      1      2
  1      1.0000
  2      0.1354      1.0000
MATRIX U
This is a SYMMETRIC matrix of order      4 by      4
      1      2      3      4
  1      0.3543
  2      0.0000      0.3030
  3      0.0000      0.0000      0.3658
  4      0.0000      0.0000      0.0000      0.2105
MATRIX U (Rough) STANDARD ERRORS
This is a SYMMETRIC matrix of order      4 by      4
      1      2      3      4
  1      0.0872
  2      0.0000      0.0544
  3      0.0000      0.0000      0.0612
  4      0.0000      0.0000      0.0000      0.0756

```

SAS@PROC CALIS

The following is the SAS program code.

```

Data LGC (type=cov);
  Input _type_ $ _name_ $ y1 y2 y3 y4;
  Cards;
N      .   118      118      118      118
Mean   .   1.788    2.102    2.347    2.737

```

```

Cov  y1  .117  .  .  .
Cov  y2  .087  .158  .  .
Cov  y3  .070  .093  .220  .
Cov  y4  .061  .086  .138  .258
;
Proc Calis ucov aug data = LGC(type=cov);
Lineqs
y1 = F1 + 0F2 + E1;
y2 = F1 + 1F2 + E2;
y3 = F1 + 2F2 + E3;
y4 = F1 + 3F2 + E4;
F1 = intcpt intercep + D1,
F2 = intcpt intercep + D2;
Std
E1-E4 = Error1 Error2 Error3 Error4,
D1-D2 = Int Slp;
Cov
D1-D2 = Cor;
Run;

```

Selected PROC CALIS output

The following provides portion of the selected output from the SAS PROC CALIS run.

```

Chi-square = 4.2233          df = 5          Prob>chi**2 = 0.5177
F1          =      1.7780*INTERCEP + 1.0000 D1
  Std Err      0.0815 INTCPT
  t Value      21.8224
F2          =      0.3132*INTERCEP + 1.0000 D2
  Std Err      0.0295 SLOPE
  t Value      10.6154

```

Variiances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
D1	INT	0.551232	0.109094	5.053
D2	SLP	0.047368	0.017692	2.677
D2 D1	COR	-0.066797	0.034566	-1.932
D2 D1	COR	-0.413378		
E1	ERRO1	0.354560	0.085627	4.141
E2	ERRO2	0.302478	0.053481	5.656
E3	ERRO3	0.366211	0.060115	6.092
E4	ERRO4	0.209971	0.074019	2.837

APPENDIX B: Two Factor Latent Growth Curves With Additional Predictors Using other Software

EQS

The following is the EQS program code for the two-factor linear curve model.

```

/TITLE
Testing a model with predictors. Gender and Age as predictors of LGC for
Cigarette Use
/Specifications
cas=275; var=6; me=ml; ana=mom; ma=raw;
da='c:/mydocu~1/li/predict.dat';
/labels
v1=cig1; v2=cig2; v3=cig3; v4=cig4;
v5=gender; v6=age;
/equations
v1 = F1 + 0F2 + e1;
v2 = F1 + 1F2 + e2;
v3 = F1 + 2F2 + e3;
v4 = F1 + 3F2 + e4;
v5 = *v999 + e5;
v6 = *v999 + e6;
f1 = *v999 + *v5 + *v6 + D1;
f2 = *v999 + *v5 + *v6 + D2;
/var
e1 to e6 = *; D1 to D2 = *;
/cov
D1,D2 = *; e5,e6 = *;
/print
fit=all
/end
/END

```

The raw data appears in the file "predict.dat." Here are the first few cases.

3.000	4.000	2.000	2.000	1.000	15.000
2.000	3.000	3.000	4.000	.000	13.000
3.000	3.000	4.000	5.000	1.000	13.000
5.000	4.000	2.000	2.000	1.000	15.000
2.000	2.000	2.000	2.000	.000	15.000
1.000	1.000	1.000	1.000	.000	16.000
1.000	1.000	1.000	1.000	.000	13.000
1.000	1.000	1.000	1.000	.000	16.000
5.000	4.000	3.000	4.000	.000	15.000
2.000	2.000	2.000	2.000	.000	15.000
3.000	2.000	2.000	2.000	.000	13.000
