5 Structural equation models

5.1 Introduction

The single most important feature of the LISREL program is its facility to deal with a wide variety of models for the analysis of latent variables (LVs). In the social sciences, and increasingly in biomedical and public health research, LV models have become an indispensable statistical tool.

Because the whole framework of the LISREL model is based on relationships among LVs, it is worthwhile to briefly illustrate the concept of a latent variable. Latent variables are ubiquitous in some research domains, while in other contexts they are seldom used. In alcohol abuse studies, for example, they are a major focus of attention. It is the complexity of attitudes and traits underlying the alcoholism syndrome that is of greatest concern, rather than any specific behavior. As an example, questionnaire items are frequently collected that deal with the functioning of the subjects in a particular domain. Subsets of these items are often correlated. This implies that the subset reflects a common theme. For example, consider the following items from a hypothetical survey, used in a hypothetical model as shown in Figure 5.1:

- o Q1: How many alcoholic drinks do you generally consume on any occasion?
- o Q2: How many days in a typical week do you consume alcohol?
- o Q3: Do you frequently attend parties where alcohol is available?
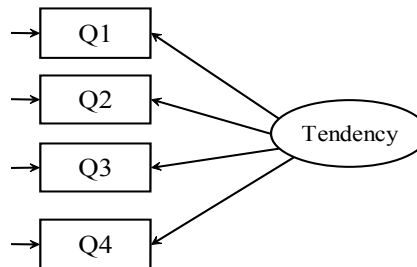- o Q4: Do you have alcoholic beverages with meals?

Figure 5.1   Path Diagram for Hypothetical SEM

In this example, a possible LV would be Tendency to Use Alcohol. This is a LV because Tendency to Use is a kind of unmeasurable propensity that is more than the combination of these items. The higher the individual Tendency LV score is, the more likely that the person will endorse questionnaire items regarding use and abuse of alcoholic beverages.

A LV is a statistical device used to summarize the information in a collection of correlated response variables. A LV describes the information of a set of items and reduces them to a single new measure. It is often assumed that the latent variable is superordinate to items on which it is based.

There are basically three major reasons for the utility of LV models. First, this kind of model can summarize information contained in many response variables by a few LVs. Consequently, the approach is parsimonious. Second, when properly specified a LV model can minimize the biasing effects of errors of measurement in estimating treatment effects. This means that the approach is often more accurate than is a traditional version of the same analysis. Third, LV models investigate effects between primary conceptual variables, rather than between any particular set of ordinary response variables. This means that a LV model is often viewed as more appropriate theoretically than is a simpler analysis with response variables only. A partial list of the sort of models that are subsumed under the framework of LISREL's general LV structure includes factor analysis, simultaneous equation models, standard growth curve processes, errors-in-variables models, virtually all forms of classical regression, univariate linear models and multivariate linear models, including the corresponding hypothesis tests on means and variances of classical experimental design. Literally hundreds of published articles appear each year that feature LV models, and an active program of statistical investigations on properties and extensions of LV models is carried out.

The hypothetical path diagram in Figure 5.2 shows seven x variables as indicators of three latent ξ variables. Note that x_3 is an indicator for both ξ_1 and ξ_2. There are two latent η variables, each with two y indicators. The model involves errors in equations (the ς s), and errors in variables (the ε s and δ s). A more detailed discussion of this model is given in Chapter 1 of the *LISREL 8: User's Reference Guide* (1996).
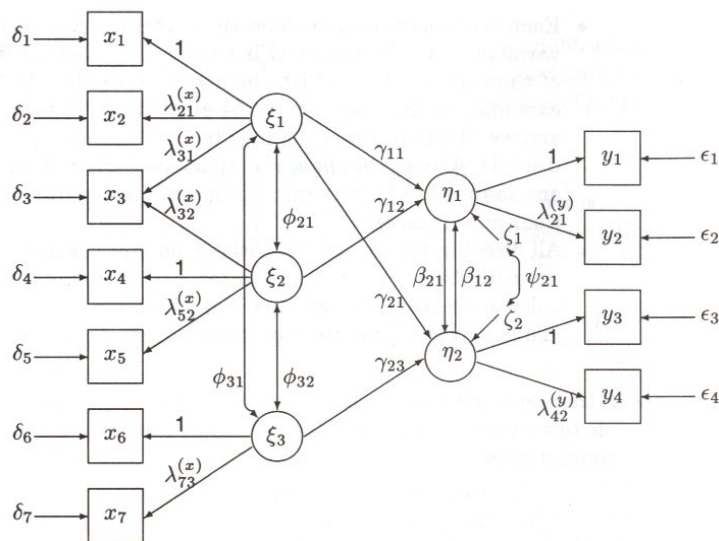

Figure 5.2   Path Diagram for Hypothetical SEM

The LISREL model for single samples (Jöreskog & Sörbom, 1996) is defined by two components, namely the structural equation model and the measurement model(s).

The structural equation model

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

(5.1)

where $\boldsymbol{\eta}$ is a $m \times 1$ vector of endogenous latent variables and where it is assumed that the $n \times 1$ vector $\boldsymbol{\xi}$ of exogenous latent variables has mean $\boldsymbol{\kappa}$ and covariance matrix $\boldsymbol{\Phi}$, and that the $m \times 1$ vector $\boldsymbol{\zeta}$ of error terms has zero mean and covariance matrix $\boldsymbol{\Psi}$, and $\text{cov}(\boldsymbol{\xi}, \boldsymbol{\zeta}') = \mathbf{0}$. If $|\mathbf{I} - \mathbf{B}| \neq 0$, and setting $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$, it follows that

$$\boldsymbol{\mu}_{\eta} = \mathbf{A}\left(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}\right)$$

(5.2)

and

$$Cov(\boldsymbol{\eta}) = \mathbf{A}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{A}'$$

.

(5.3)

Measurement models

The measurement models for the p endogenous observed variables, represented by the vector $\mathbf{y}$, and the q exogenous observed variables, contained in the vector $\mathbf{x}$, relate the observed (manifest) variables to the underlying factors (latent variables) and may be expressed as

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \;\; E(\boldsymbol{\varepsilon}) = \mathbf{0}, \;\; Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}_{\varepsilon}$$

$$\mathbf{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \;\; E(\boldsymbol{\delta}) = \mathbf{0}, \;\; Cov(\boldsymbol{\delta}) = \boldsymbol{\Theta}_{\delta}$$

(5.4)

respectively.

The mean vectors of the observed variables are

$$\boldsymbol{\mu}_y = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \mathbf{A}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\kappa}), \;\;\;\; \boldsymbol{\mu}_x = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\kappa}$$

(5.5)

In general, in a single population, $\boldsymbol{\tau}_y$, $\boldsymbol{\tau}_x$, $\boldsymbol{\alpha}$, and $\boldsymbol{\kappa}$ will not be identified without the imposition of further conditions. It further follows that

$$\boldsymbol{\Sigma}_y = \boldsymbol{\Lambda}_y \left[\mathbf{A}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{A}' \right] \boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon$$

(5.6)

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta$$

(5.7)

and

$$\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Lambda}_y \mathbf{A}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x'.$$

(5.8)

From (5.5) to (5.8), it follows that the covariance structure for the observed variables of the general LISREL model may be expressed as:

$$\boldsymbol{\Sigma} = Cov \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}$$

(5.9)

From (5.5), the mean structure of the observed variables of the general LISREL model follows as:

$$\boldsymbol{\mu} = E \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{bmatrix}.$$

(5.10)

LISREL fits the mean-and-covariance structure defined in (5.9) and (5.10) to the data on the observed variables of the LISREL model. In this regard, LISREL can handle simple random sample data as well as complex survey data.

Special cases of the general LISREL model are obtained by fixing and constraining the parameters which are the elements in the 13 parameter matrices $\left(\boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\tau}_x, \boldsymbol{\tau}_y, \boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x, \mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Theta}_\varepsilon, \boldsymbol{\Theta}_\delta, \boldsymbol{\Theta}_{\delta_\varepsilon} \right)$.

A large number of submodels is obtained by setting certain parameter matrices equal to the identity matrix or to zero. A few examples are:

o The measurement model for $\mathbf{x}$, $\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}$.

- A structural equation model where $\mathbf{y}$ and $\mathbf{x}$ are observed without error ($\boldsymbol{\Lambda}_y = \mathbf{I}, \boldsymbol{\Lambda}_x = \mathbf{I}$, $\boldsymbol{\Theta}_\varepsilon = \mathbf{0}, \boldsymbol{\Theta}_\delta = \mathbf{0}$)

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}$$

Kaplan (2000) pointed out that this model was a major innovation in econometric modeling. In the special case where $\mathbf{B} = \mathbf{0}$, one obtains the multivariate multiple regression model

$$\mathbf{y} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}$$

The general form of the LISREL model, due to its flexible specification in terms of fixed and free parameters and simple equality constraints, has proven to be so rich that it can handle a large variety of problems. Using the inequality constraints feature in LISREL, users constantly discover new models, such as nonlinear growth curves (see du Toit & Cudeck, 2001) and vector time series models with ARMA residuals (du Toit & Browne, 2001) that can be handled within the LISREL framework.

There are many articles on structural equation modeling. Hayduk (1996), for example, gives a long list of substantive areas where structural equation models are being used: addictions, criminology, education, family studies, health, marketing, psychology, and sociology to mention just a few. A very large number of technical and substantive articles using structural equation models have appeared in dozens of journals.

The next section describes how to draw a path diagram and create syntax using the graphical user interface of LISREL. An overview of the SIMPLIS syntax, which is used to specify LISREL models, is given in Section 5.3. Thereafter, two illustrative examples are discussed in Section 5.4. In Section 5.5, a simulation study and empirical comparisons are used to assess the results produced by LISREL in the case of complex survey data. An overview of the statistical theory implemented in LISREL for the analysis of complex survey data concludes this chapter.

5.2   Graphical User Interface

5.2.1   The new PTH window

The path diagram component of the graphical user interface (GUI) of the LISREL module consists of the options and dialog boxes of the **Setup** menu on the PTH window of LISREL 8.7. This GUI component allows you to interactively generate the syntax file by means of a path diagram, which is a graphical representation of a structural equation model. The **Setup** menu on the PTH window of LISREL 8.7 is reviewed in the next section while the four dialog boxes are reviewed separately in the subsequent sections. Thereafter, the use of the graphic pane of the PTH window is outlined.

The **Setup** menu on the PTH window provides access to a sequence of four dialog boxes that can be used to create a SIMPLIS or LISREL syntax file interactively by using a path diagram. A new PTH window is opened as follows. Open LISREL 8.7 and select the **New** option on the **File** menu to create the following window.
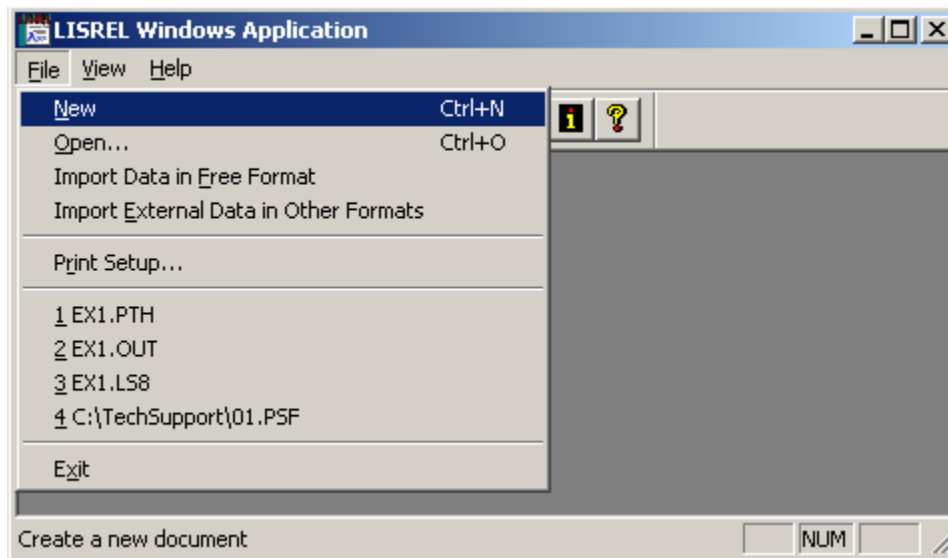
Click on the **New** option on the **File** menu to load the **New** dialog box and select the **Path Diagram** option from the **New** dialog box as shown below.
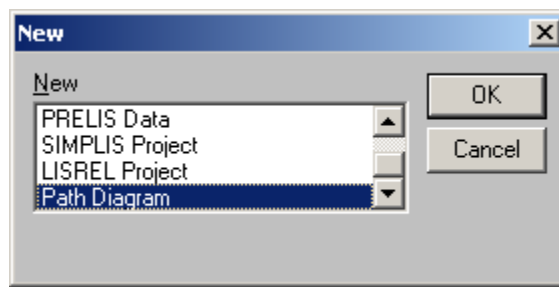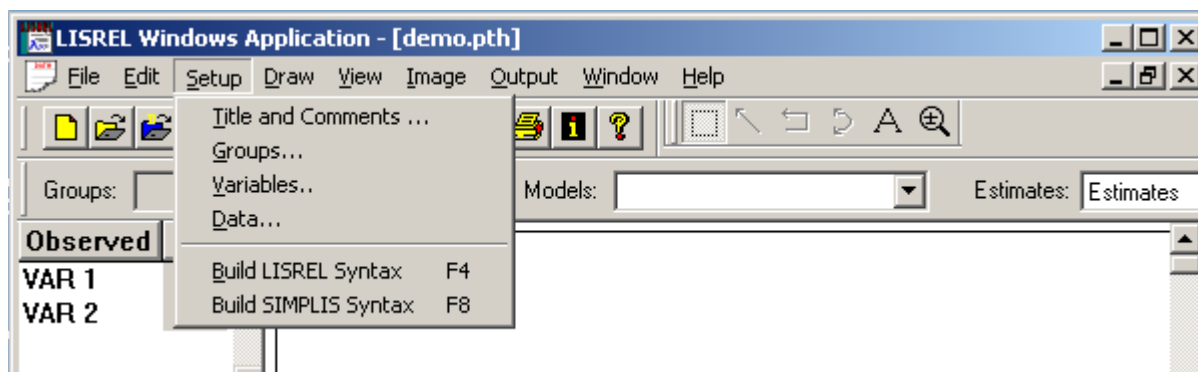
Click on the **OK** button to load the **Save As** dialog box and then enter, for example, the name **demo.pth** in the **File name** field to produce the following dialog box.
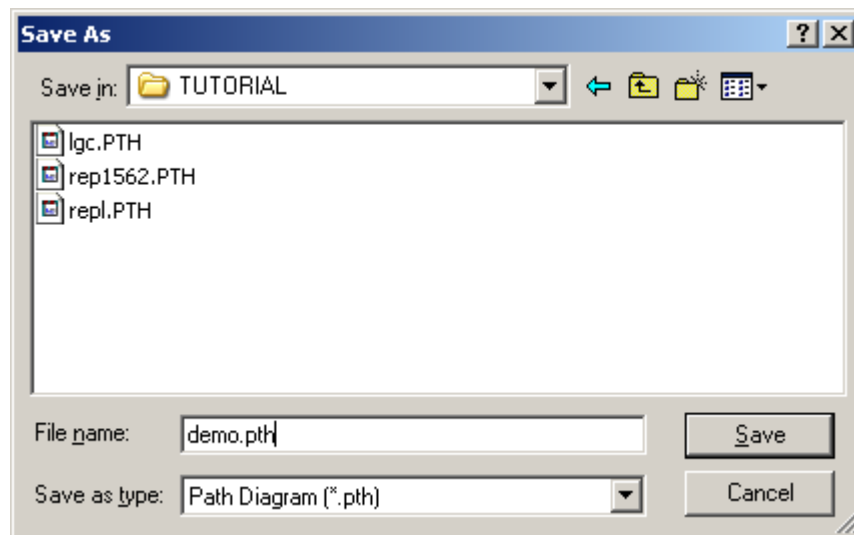

5.2.2 The Setup menu

Next, click on the **Save** button to open the PTH window for **demo.pth** and then click on the **Setup** menu to obtain the following window.

Typically, clicking on the **Title and Comments** option of the **Setup** menu will load the **Title and Comments** dialog box (see Section 5.2.2). However, you can click directly on the **Groups**, **Variables** or **Data** option to go to the **Groups** (see Section 5.2.3), the **Variables** (see Section 5.2.4), or the **Data** dialog box (see Section 5.2.5). Once you have completed the four sequential dialog boxes and drawn the path diagram, the SIMPLIS syntax file or the LISREL syntax file is generated by clicking on the **Build SIMPLIS Syntax** or the **Build LISREL Syntax** option respectively.

5.2.3   The Title and Comments dialog box

The **Title and Comments** dialog box allows you to specify a title and additional comments for the analysis. It is accessed by selecting the **Title and Comments** option on the **Setup** menu. This selection loads the following **Title and Comments** dialog box.
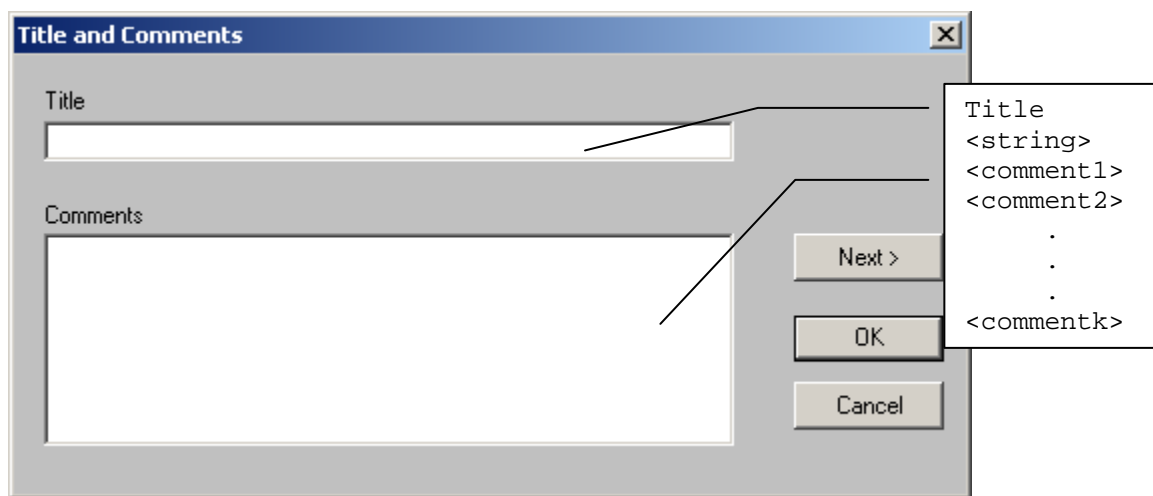

Note that the **Title and Comments** dialog box corresponds with the Title command as shown above.

Once you are done with the **Title and Comments** dialog box, click on the **Next** button to go to the **Group Names** dialog box.

5.2.4   The Group Names dialog box

The **Group Names** dialog box is usually accessed by clicking on the **Next** button of the **Title and Comments** dialog box. It is required for multiple group analysis and allows you to specify different group names.

Note that the **Group Names** dialog box corresponds with the Group command as indicated on the image below. For single group analysis, you can skip this dialog box by simply clicking on the **Next** button.
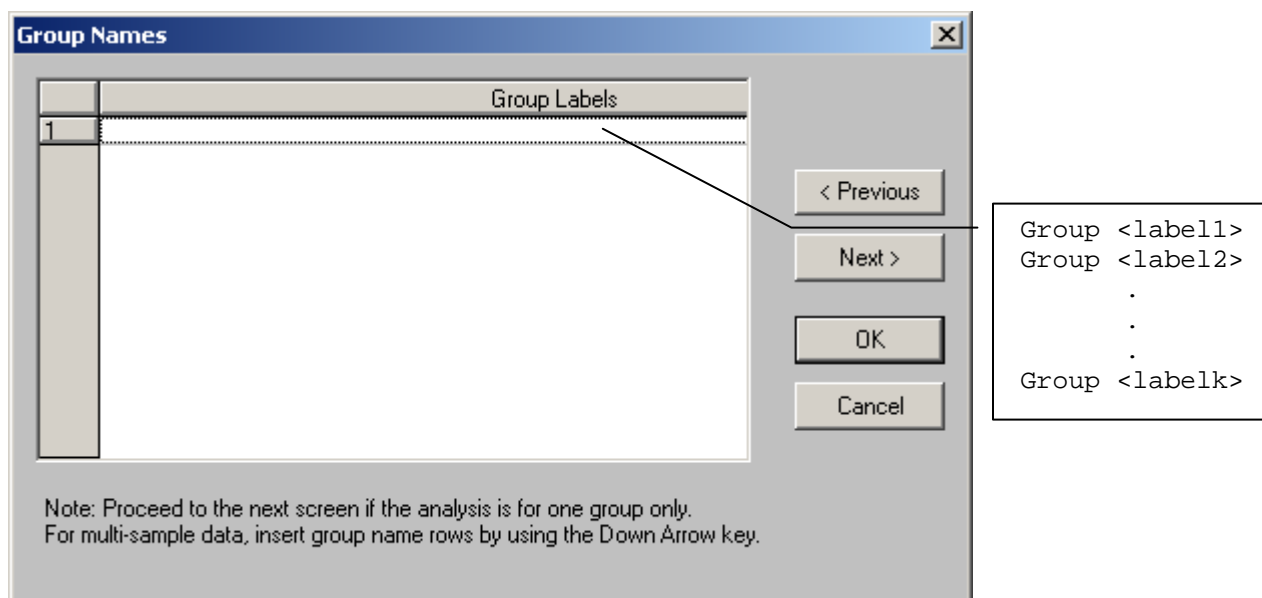
Once the **Group Names** dialog box has been completed, click the **Next** button to go to the **Labels** dialog box.
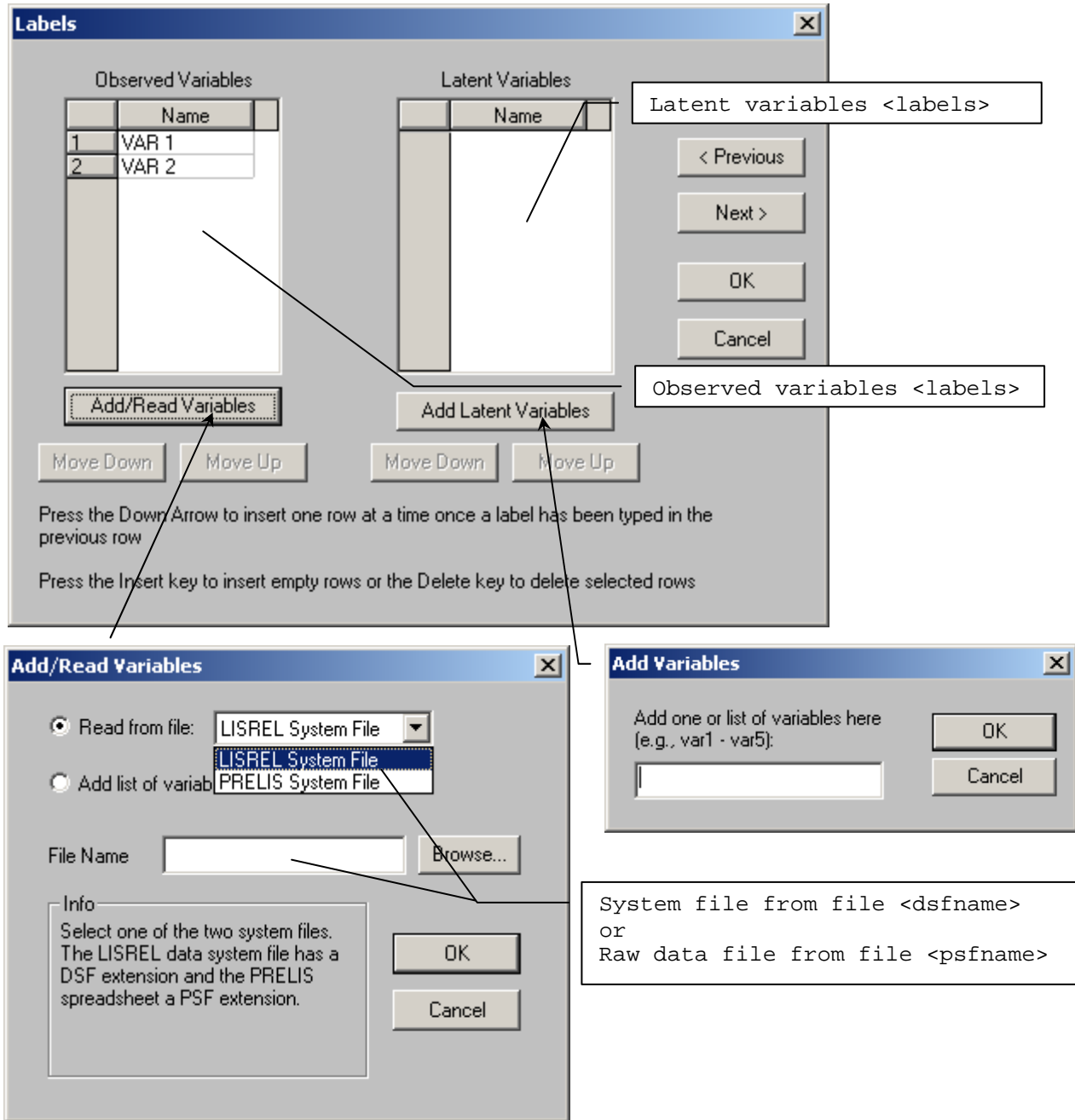
5.2.5   The Labels dialog box

The **Labels** dialog box allows you to specify the observed variables and latent variables of the model interactively. Access to this dialog box is obtained by clicking on the **Next** button of the **Group Names** dialog box. This selection loads the **Labels** dialog box as shown below.

Note that the **Labels** dialog box corresponds with the Observed variables and Latent variables commands as shown above. Note also that the **Add/Read Variables** dialog box, which is loaded by clicking on the **Add/Read Variables** button, corresponds with the System file from file and the Raw data file from file commands.

If a LISREL system file (DSF) or a PRELIS system file (PSF) is to be used, you can browse for the corresponding DSF or PSF by first selecting the **LISREL System File** option or the **PRELIS System File** option from the drop-down list box respectively and then clicking on the **Browse** button. Otherwise, you can add a list of variables by activating the **Add list of variables** radio button. When you are done with the **Add/Read Variables** dialog box, click the **OK** button to return to the **Labels** dialog box.

If the model includes any latent (unobservable) variables, you must specify labels for them by clicking on the **Add Latent Variables** button to load the **Add Variables** dialog box. Click the **OK** button after the label has been entered to return to the **Labels** dialog box. Once the labels for all the latent variables of the model have been specified, you click on the **Next** button to return to the **Data** dialog box.

Latent variables <labels>

Observed variables <labels>

System file from file <dsfname>
or
Raw data file from file <psfname>

5.2.6　The Data dialog box

Specify the data to be analyzed by using the **Data** dialog box. It is usually accessed by clicking on the **Next** button of the **Labels** dialog box. This action loads the following **Data** dialog box.
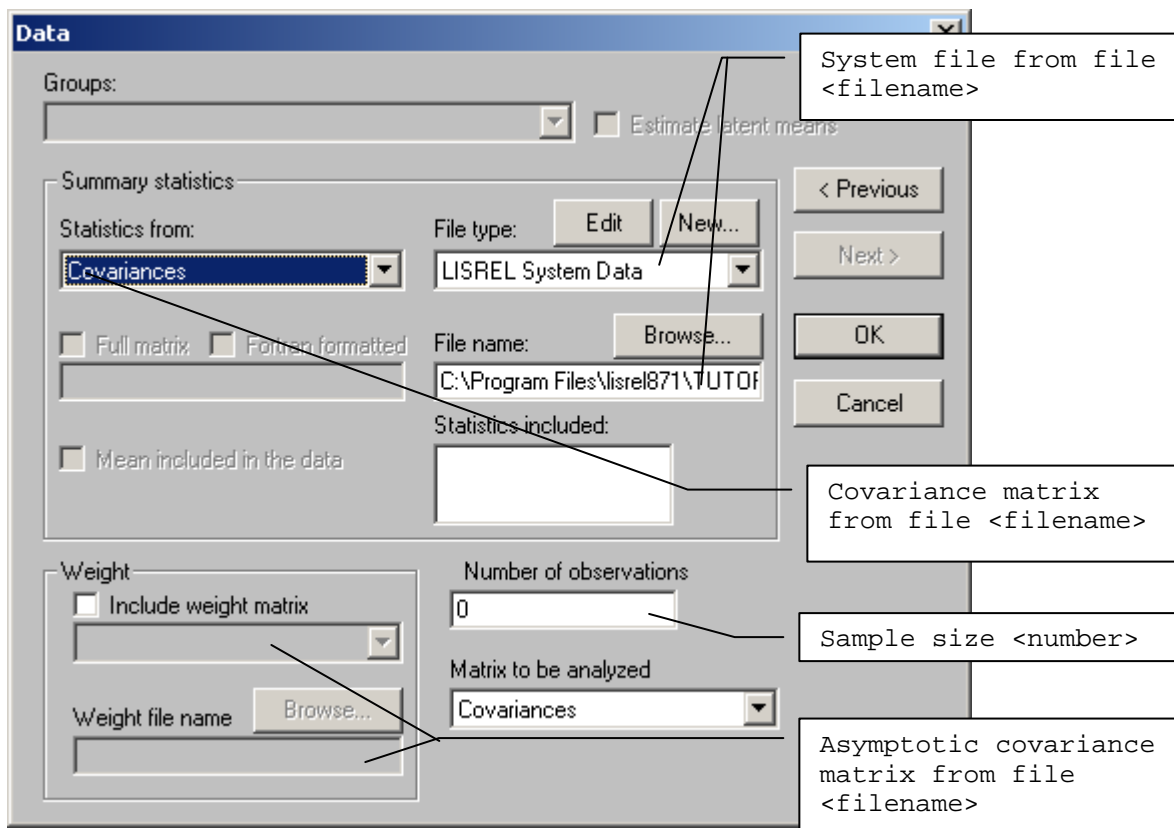
Note that the **Data** dialog box corresponds with the System file from file, Covariance matrix from file, Sample Size and Asymptotic covariance matrix from file commands as shown in the image above. If a DSF or a PSF is selected in the **Labels** dialog box and the covariance matrix is the matrix to be analyzed, the **Data** dialog box is redundant. In other words, in this case, you can click on the **OK** button to return to the **PTH** window without completing the **Data** dialog box.

In the case of a single-group analysis, the **Groups** drop-down list box is not accessible. In the case of a multiple group analysis, the **Groups** drop-down list box displays the labels of the different groups as specified in the **Group Names** dialog box. In this case, specify the data for each group by selecting the group name from the list box. If the latent variable means are to be compared across groups, click the **Estimate latent means** check box.

Select the desired data type from the **Statistics from** drop-down list box if the covariance matrix is not desired. Select the appropriate data file type from the **File type** drop-down list box if a DSF is not preferred and then use the **Browse** button to browse for the corresponding file. You can open the data file to be analyzed by clicking on the **Edit** button or specify the data by clicking on the **New** button. If the asymptotic covariance matrix or asymptotic variances of the sample moments is to be used in the analysis, you must check the **Include weight matrix** check box, select the type of weight matrix from the drop-down list box and browse for the appropriate file. If a DSF or a PSF is not to be processed, enter the number of observations in the **Number of observations** field. Select the desired moment matrix from the **Matrix to be analyzed** drop-down list box if a correlation matrix rather than a covariance matrix is to be analyzed.

After all four dialog boxes are completed, click on the **OK** button to return to the PTH window.

5.2.7   The graphic pane of the PTH window

Once you are done with the four dialog boxes of the **Setup** menu of the **PTH** window, the graphic pane of the PTH window is used to create a path diagram of the model to be fitted to the data. An example of the graphic pane of the PTH window is shown below.
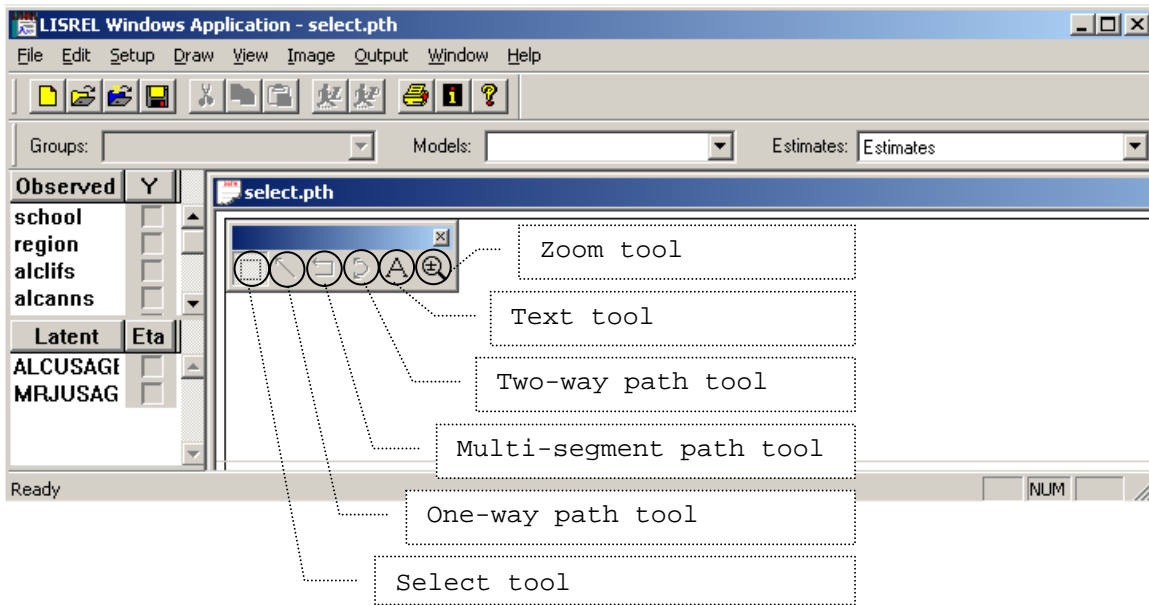
You may use the following sequential steps to create a path diagram of the structural equation model to be fitted to the data in the graphic pane of the window.

- o Use the **Select** tool to click, drag and drop the observed variables, one at a time, from the **Observed** list box into the graphic pane of the window.
- o If the model includes latent variables, use the **Select** tool to click, drag and drop the latent variables, one at a time, from the **Latent** list box into the graphic pane of the window.
- o Use the **One-way path** tool to specify the regression relationships between the observed and latent variables of the model.
- o If applicable, use the **Two-way path** tool to specify the covariance (correlation) relationships between the latent variables and the error variables of the model.
- o If certain parameters of the model are fixed to specific values, certain parameters are set equal to each other, a path needs to be removed from the model or certain graphic properties are desired, the right-click menu of a path is used. This menu is activated by first selecting the path by using the **Select** tool and then by right-clicking to view the following menu.
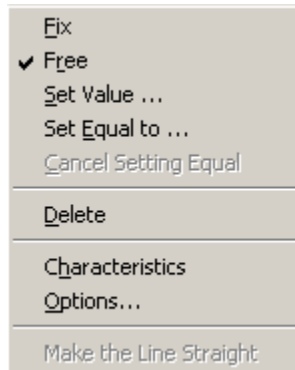
The options on the menu above are used as follows.

- o The **Fix** option is used to fix a parameter that was set free by default.
- o The **Free** option is used to free a parameter that was fixed by default.
- o The **Set Value…** option is used to specify the value for a fixed parameter.
- o The **Set Equal to…** option is used to set a parameter to be equal to another parameter(s).
- o The **Cancel Setting Equal** option is used to release an equality constraint that was specified.
- o The **Delete** option is used to specify the deletion of a path or a selected object.
- o The **Characteristics** option is used to obtain information about the parameter.
- o The **Options** option is used to modify the graphic properties of the path.
- o The **Make Line Straight** option is used to automatically straighten a one-way path.

Once the path diagram has been drawn, click on the **Build SIMPLIS Syntax** option or **Build LISREL Syntax** option on the **Setup** menu to generate the SIMPLIS syntax file or the LISREL syntax file respectively.

## 5.2.8 The Weight Cases and Survey Design dialog boxes

The **Weight Cases** and **Survey Design** dialog boxes can be accessed if a PSF file is the active window. This is accomplished by selecting the **Data** menu from the main menu bar.
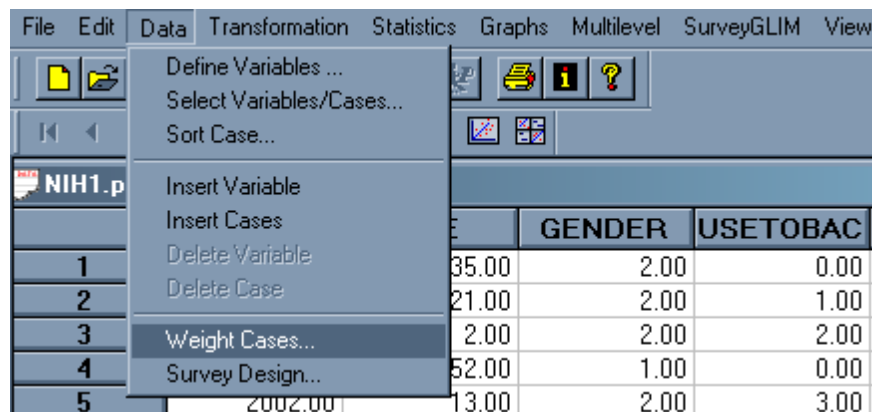
The **Weight Cases** option is used to calculate weighted sample statistics, for example means, covariances, and asymptotic covariances. It is assumed that these weights are normalized in the sense that the sum of sample weights equals the sample size.
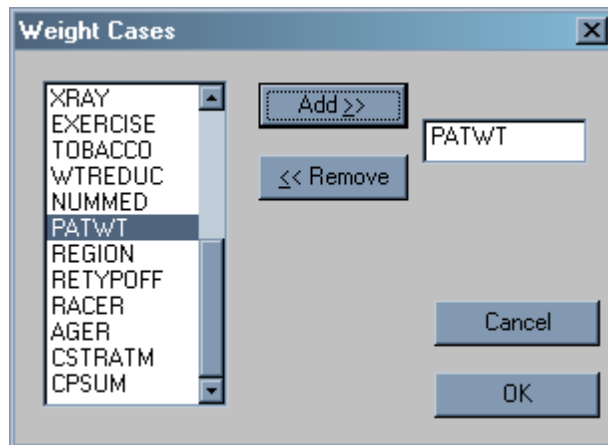
The **Survey Design** dialog box shown below is used to define the stratification and cluster variables and to select a design weight. This information is stored within the PSF file and is retrieved whenever a SEM, based on an continuous outcome variable, is fitted to the data contained in the PSF file.
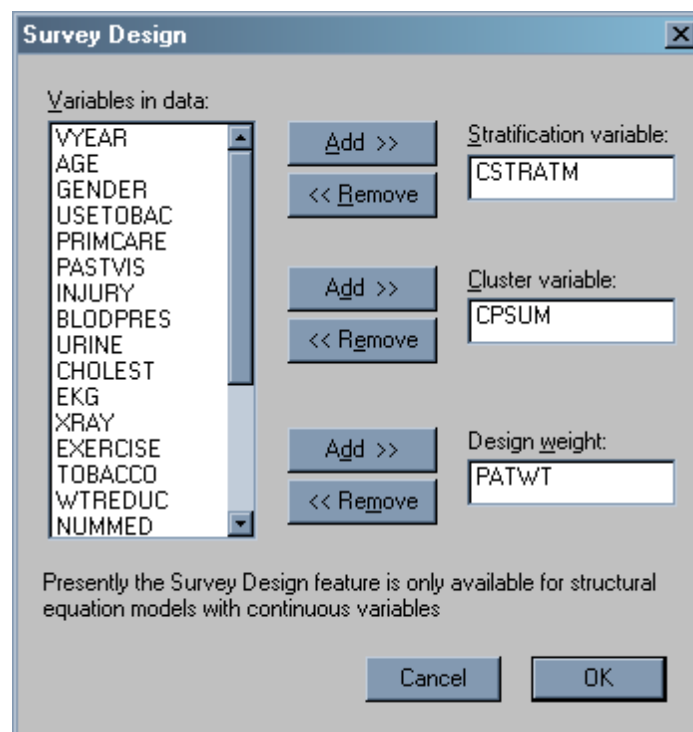
5.3   Syntax

5.3.1   The structure of the SIMPLIS syntax file

The SIMPLIS syntax file, which is generated by the graphical user interface of the LISREL module, can also be prepared manually by using the LISREL 8.7 text editor or any other text editor such as Notepad and WordPad. The general structure of the SIMPLIS syntax file depends on the data to be processed. If the raw data file to be processed is a PSF, the SIMPLIS syntax file has the following structure.

TITLE	Required
<string>	
RAW DATA FROM FILE <psfname>	Required
MISSING VALUE CODE <value>	Optional
STRATUM <label>	Optional
CLUSTER <label>	Optional
WEIGHT <label>	Optional
CASEWEIGHT <label>	Optional
$CLUSTER <label>	Optional
$PREDICT <labels>	Optional
LATENT VARIABLES	Optional
<labels>	
RELATIONSHIPS	Required
<relationships>	
SET <instruction>	Optional
LISREL OUTPUT <options>	Optional
PATH DIAGRAM	Optional
END OF PROBLEM	Required

where <string> denotes a character string, <label> denotes a case-sensitive variable name used in the raw data or moment matrix file, <labels> denotes a list of case-sensitive variable names used in the raw data or moment matrix file or for the latent variables of the model, <psfname> denotes the complete name (including the drive and folder names) of the PSF, <value> denotes any real number, <relationships> denotes a list of model expressions (see Section 5.3.24) and <instruction> denotes a parameter statement (see Section 5.3.26). <options> denotes a list of options for the analysis each of which either has the syntax:

<keyword> = <selection>

where \<keyword\> is one of AD, AL, BE, EP, GA, IT, KA, LX, LY, MA, ME, ND, NP, PH, PS, PV, RC, SI, SL, SV, TD, TE, TH, TM, TV, TX, TY or XO and \<selection\> denotes a number, a value or a name (see Section 5.3.13) or the syntax:

<option>

where \<option\> is one of ALL, AM, EF, FS, FT, MI, MR, NS, PC, PT, RO, RS, SC, SO, SS, WP, XA, XI or XM (see Section 5.3.14).

If the data to be analyzed are summarized in a DSF, the structure of the SIMPLIS syntax file is as follows.

TITLE	Optional
\<string\>	
SYSTEM FILE FROM FILE \<dsfname\>	Required
LATENT VARIABLES	Optional
\<labels\>	
RELATIONSHIPS	Required
\<relationships\>	
SET \<instruction\>	Optional
LISREL OUTPUT \<options\>	Optional
PATH DIAGRAM	Optional
END OF PROBLEM	Required

where \<dsfname\> denotes the complete name (including the drive and folder names) of the DSF (see Section 5.3.14).

The SIMPLIS syntax file has the following structure if the data file to be processed is in the form of a text file.

TITLE	Optional
\<string\>	
OBSERVED VARIABLES	Required if COV. or CORR.
\<labels\>	MATRIX is not selected
RAW DATA FROM FILE \<filename\>	Required
MISSING VALUE CODE \<value\>	Optional
STRATUM \<label\>	Optional
CLUSTER \<label\>	Optional
WEIGHT \<label\>	Optional
CASEWEIGHT \<label\>	Optional
$CLUSTER \<label\>	Optional
$PREDICT \<labels\>	Optional
COVARIANCE MATRIX FROM FILE \<filename\>	Required if RAW DATA or

	CORR. MATRIX is not selected
CORRELATION MATRIX FROM FILE <filename>	Required if RAW DATA or
	COV. MATRIX is not selected
ASYMPTOTIC COVARIANCE MATRIX FROM FILE	Optional
<acmfilename>	
MEANS FROM FILE <filename>	Optional
STANDARD DEVIATIONS FROM FILE <filename>	Optional
SAMPLE SIZE <number>	Required
LATENT VARIABLES	Optional
<labels>	
RELATIONSHIPS	Required
<relationships>	
SET <instruction>	Optional
LISREL OUTPUT <options>	Optional
PATH DIAGRAM	Optional
END OF PROBLEM	Required

where <filename> denotes the complete name (including the drive and folder names) of a text file, <acmfilename> denotes the complete name (including the drive and folder names) of the binary file containing the estimated asymptotic covariance matrix of the sample moments and <number> denotes a positive integer.

The three general structures of the SIMPLIS syntax file listed here assume a single-group structural equation model. In the case of a multiple group structural equation model, these structures apply to each GROUP command (see Section 5.3.10). The only exception is the END OF PROBLEM command, in the sense that only one should be specified as the final command of the SIMPLIS syntax file for the multiple group analysis.

The SYSTEM FILE FROM FILE command is a **required** command only if a DSF is used. If the data to be analyzed do not come from a DSF or PSF, then the OBSERVED VARIABLES paragraph, the SAMPLE SIZE command, and one of the RAW DATA FROM FILE, COVARIANCE MATRIX FROM FILE, or the CORRELATION MATRIX FROM FILE commands are **required**. The LATENT VARIABLES paragraph is **required** only if the model includes latent variables. The RELATIONSHIPS or PATHS paragraph is **required**. The remaining SIMPLIS commands are all **optional**.

One of the SYSTEM FILE FROM FILE or RAW DATA FROM FILE commands or the OBSERVED VARIABLES paragraph should be the first command following the TITLE paragraph. If the END OF PROBLEM command is included, it must be the final command. The other commands and paragraphs can be entered in any order.

In the following sections, the SIMPLIS commands and paragraphs are discussed separately in alphabetical order.

5.3.2 $CLUSTER command

The $CLUSTER command is used to specify the variable that contains the cluster information of nested data for which a multilevel structural equation modeling analysis is desired. It is an **optional** command. For example, in the case of a standard structural equation modeling analysis, the $CLUSTER command is omitted.

Syntax

$CLUSTER <label>

where <label> denotes the label of the cluster variable.

Example

Suppose that the primary sampling units of the complex survey are facility types and that the variable FACTYPE is used to indicate the facility type for each observation. Then, the corresponding $CLUSTER command is

$CLUSTER FACTYPE

5.3.3 $PREDICT command

The $PREDICT command is used to specify the explanatory variables for the fixed part of a multilevel structural equation model. It is an **optional** command. For example, in the case of a standard structural equation modeling analysis, the $PREDICT command is omitted.

Syntax

$PREDICT <labels>

where <labels> denotes the labels of the explanatory variables.

Example

Suppose that the age (AGE) and gender (GENDER) of each respondent are to be used as predictors for the fixed part of a multilevel structural equation model. For this example, the corresponding $PREDICT command is

$PREDICT = AGE GENDER;

5.3.4   ASYMPTOTIC COVARIANCE MATRIX FROM FILE command

The ASYMPTOTIC COVARIANCE MATRIX FROM FILE command is used to specify the name of the binary file that contains the estimated asymptotic covariance matrix of the sample moments. In the case of the Robust Maximum Likelihood (RML), the Weighted Least Squares (WLS) and the Diagonally Weighted Least Squares (DWLS) methods, it is a **required** command.

Syntax

> ASYMPTOTIC COVARIANCE MATRIX FROM FILE <acmfilename>

where <acmfilename> denotes the name of the binary file containing the estimated asymptotic covariance matrix of the sample moments. If the acmfilename contains blank spaces, it should be given in single quotes.

Example

Suppose that the name of the binary file with the estimated asymptotic covariance matrix of the sample variances and covariances is **NIH1.ACM** and that it is located in the folder **Projects\NIH1** on the E drive. In this case, the corresponding COVARIANCE MATRIX FROM FILE command is given by

> ASYMPTOTIC COVARIANCE MATRIX FROM FILE 'E:\Projects\NIH1\NIH1.ACM'

5.3.5   CASEWEIGHT command

The purpose of the CASEWEIGHT command is to allow the user to specify the variable containing the weights of the individual observations to be used to compute weighted means, sample variances and covariances (correlations) and asymptotic covariance matrices of the sample variances and covariances (correlations). It is assumed that these weights are normalized in the sense that they add up to the sample size. The CASEWEIGHT command is an optional command and corresponds with the selected variable on the **Weight Cases** dialog box (see Section 5.2.8).

Syntax

> CASEWEIGHT <label>

where <label> denotes the label of the variable containing the case weights.

Example

Suppose that the variable NEWWGT contains the weight for each observation. For this example, the CASEWEIGHT command is given by

> CASEWEIGHT NEWWGT

5.3.6   CLUSTER command

The CLUSTER command is used to specify the variable for the primary sampling units of the complex survey. It is an **optional** command. For example, in the case of a simple random sample, the CLUSTER command is omitted. The CLUSTER command corresponds with the CLUSTER variable section on the Survey Design dialog box (see Section 5.2.8).

Syntax

> CLUSTER <label>

where <label> denotes the label of the cluster variable.

Example

Suppose that the primary sampling units of the complex survey are types of facility and that the variable FACTYPE is used to indicate the facility type for each observation. Then, the corresponding CLUSTER command is

> CLUSTER FACTYPE

5.3.7   CORRELATION MATRIX paragraph

The correlation matrix to be processed can be specified as a part of the SIMPLIS syntax file by using the CORRELATION MATRIX paragraph. It is a **required** command only if the correlations to be processed are provided as part of the SIMPLIS syntax file. If the sample correlations are in the form of a text file, the CORRELATION MATRIX FROM FILE command rather than the CORRELATION MATRIX paragraph is used (see Section 5.3.6).

Syntax

> CORRELATION MATRIX
> <values>

where <values> denotes the sample correlations of the observed variables in free or fixed format. If the sample correlations are listed in a fixed format, a FORTRAN format statement should be included as the first line of the CORRELATION MATRIX paragraph.

Examples

> CORRELATION MATRIX
> 1.000
> 0.257 1.000
> 0.521 0.245 1.000
> 0.533 0.346 0.218 1.000

```
CORRELATION MATRIX
(4F6.3)
 1.000
 0.257  1.000
0.521  0.245  1.000
0.533  0.346  0.218 1.000
```

5.3.8   CORRELATION MATRIX FROM FILE command

If the correlation matrix to be processed is in the form of a text file, the CORRELATION MATRIX FROM FILE command is a **required** command and is used to specify the name of the text file that contains the sample correlations of the observed variables of the model. It is also possible to specify the correlation matrix as part of the SIMPLIS syntax file. In this case, the CORRELATION MATRIX paragraph instead of the CORRELATION MATRIX FROM FILE command is used (see Section 5.3.5).

Syntax

CORRELATION MATRIX FROM FILE <filename>

where <filename> denotes the name of the text file containing the sample correlation matrix. If the filename contains blank spaces, it should be given in single quotes.

Example

Suppose that the sample correlations are contained in the text file **SELECT.COR**, which is located in the folder **My Projects\SELECT** on the D drive. In this case, the corresponding CORRELATION MATRIX FROM FILE command is given by

CORRELATION MATRIX FROM FILE 'D:\My Projects\SELECT\SELECT.COR'

5.3.9   COVARIANCE MATRIX paragraph

The COVARIANCE MATRIX paragraph is used to provide the sample covariance matrix as a part of the SIMPLIS syntax file. If the covariance matrix to be analyzed is provided as part of the SIMPLIS syntax file, it is a **required** command. If the sample covariance matrix is in the form of a text file, the COVARIANCE MATRIX FROM FILE command rather than the COVARIANCE MATRIX paragraph is used (see Section 5.3.8).

Syntax

COVARIANCE MATRIX

<values>

where <values> denotes the sample variances and covariances of the observed variables in free format. If the sample variances and covariances are provided in a fixed format, a FORTRAN type X format statement should be included as the first line of the COVARIANCE MATRIX paragraph.

Examples

COVARIANCE MATRIX
25.001
33.257 57.251
26.385 32.674 61.323
39.533 38.552 44.227 72.052

COVARIANCE MATRIX
(4F6.3)
25.001
33.25757.251
26.38532.67461.323
39.53338.55244.22772.052

5.3.10 COVARIANCE MATRIX FROM FILE command

The COVARIANCE MATRIX FROM FILE command is used to specify the name of the text file that contains the sample covariance matrix of the observed variables of the model. It is a **required** command only if the covariance matrix to be analyzed is in the form of a text file.

Syntax

COVARIANCE MATRIX FROM FILE <filename>

where <filename> denotes the name of the text file containing the sample covariance matrix.

Example

Suppose that the name of the text file with the sample variances and covariances is **NIH1.COV** and that it is located in the folder **Projects\NIH1** on the E drive. In this case, the corresponding COVARIANCE MATRIX FROM FILE command is given by

COVARIANCE MATRIX FROM FILE 'E:\Projects\NIH1\NIH1.COV'

5.3.11 END OF PROBLEM command

The END OF PROBLEM command is usually the final command of a SIMPLIS syntax file and it indicates that no more commands or paragraphs are to be processed. It is an **optional** command.

Syntax

> END OF PROBLEM

5.3.12 GROUP command

The GROUP command is used to specify a model for each of the groups in a multiple-group structural equation model. A GROUP command is specified for each group to be included in the multiple group analysis. If no RELATIONSHIPS or PATHS paragraph and no SET command are specified for any group after the very first group, the structural equation model for the group is assumed to be identical (including equal parameters) to that of the previous group. In other words, if you want the parameters to be different from that of the previous group for a specific group, each parameter has to be specified explicitly in the RELATIONSHIPS or PATHS paragraph or SET commands for that specific group.

Syntax

> GROUP <string>

where <string> denotes the descriptive name of the group.

Examples

> GROUP Freshmen

> GROUP 1

5.3.13 LATENT VARIABLES paragraph

The LATENT VARIABLES paragraph is used to provide descriptive names to the latent variables of the model. It is a **required** command if the model includes latent variables. However, if the latent variable labels are in the form of a text file, the LATENT VARIABLES FROM FILE command instead of the LATENT VARIABLES paragraph is used (see Section 5.3.12).

Syntax

> LATENT VARIABLES
> <labels>

where <labels> denotes the descriptive names of the latent variables of the model. These names are provided in free or abbreviated format and only the first 8 characters of each name are utilized.

Examples

LATENT VARIABLES
JobSat OrgCom Perform

LATENT VARIABLES
FACTOR1 - FACTOR4

5.3.14   LATENT VARIABLES FROM FILE command

If the labels of the latent variables are in the form of a text file, the LATENT VARIABLES FROM FILE command is used to specify descriptive names for the latent variables of the model. In this specific case, it is a **required** command. The latent variable labels can also be specified as part of the SIMPLIS syntax file. In this regard, the LATENT VARIABLES paragraph rather than the LATENT VARIABLES FROM FILE command is used (see Section 5.3.11).

Syntax

LATENT VARIABLES FROM FILE <filename>

where <filename> denotes the name of the text file containing the descriptive names of the latent variables of the model.

Example

Suppose that the name of the text file containing the latent variable labels is **SELECT.LAB** and that it is located in the folder **Projects\SELECT** on the D drive. In this case, the corresponding LATENT VARIABLES FROM FILE command is given by

LATENT VARIABLES FROM FILE 'D:\Projects\SELECT\SELECT.LAB'

5.3.15   LISREL OUTPUT command

The LISREL OUTPUT command is used to request the results to be printed in terms of the LISREL model used in the analysis, to specify special analyses and to request additional results. It is an **optional** command. If the results in terms of the LISREL model are not desired, the OPTIONS command may be used to specify special analyses and to request additional results (see Section 5.3.19).

Syntax

> LISREL OUTPUT <options>

where <options> denotes a list of options for the analysis each of which either has the syntax:

> <keyword> = <selection>

where <keyword> is one of AD, AL, BE, DW, EP, GA, IT, KA, LX, LY, MA, ME, ND, NP, PH, PS, PV, RC, SI, SL, SV, TD, TE, TH, TM, TV, TX, TY or XO and <selection> denotes a number, a value or a name; or the syntax:

> <option>

where <option> is one of ALL, AM, EF, FS, FT, MI, MR, NS, PC, PT, RO, RS, SC, SO, SS, WP, XA, XI or XM. Keywords and options may be specified in any order.

Examples

> LISREL OUTPUT ND = 3 SC ME = DW
> LISREL OUTPUT BE = BETA.TXT GA = GAMMA.TXT PV = PV.TXT SV = SV.TXT ND = 6

All the keywords and options of the LISREL OUTPUT command are **optional**.

AD keyword

The purpose of the AD keyword is to specify the iteration number at which the admissibility of the solution will be checked and the iterations will stop if the check fails.

Syntax

> AD = <number>

where <number> denotes the iteration number or OFF if the check is to be turned off.

Default

> AD = 20

AL keyword

The AL keyword is used to specify the name of the text file to which the estimates of the endogenous latent variable means are to be written.

Syntax

AL = <filename>

where <filename> denotes the name of the text file to which the estimates are to be written.

ALL option

The purpose of the ALL option is to invoke the printing of all the results in the output file.

AM option

The automatic model modification procedure is invoked by specifying the AM option.

BE keyword

The purpose of the BE keyword is to specify the name of the text file for the estimate of the matrix of regression coefficients for the regression among the endogenous latent variables.

Syntax

BE = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

EF option

The EF option is used to invoke the printing of the estimated total and indirect effects in the output file.

EP keyword

The convergence criterion for the iterative algorithm, which is used to obtain parameter and standard error estimates, is specified by using the EP keyword.

Syntax

EP = <value>

where <value> denotes the convergence criterion.

Default

EP = 0.000001

FS option

The FS option is used to request a factor scores regression analysis.

FT option

The purpose of the FT option is to request an external text file containing measures of fit based on four different χ^2 test statistic values.

GA keyword

The GA keyword is used to specify the name of the text file for the estimated regression matrix of the regression of the endogenous latent variables on the exogenous latent variables.

Syntax

GA = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

IT keyword

The purpose of the IT keyword is to specify the maximum number of iterations for the iterative algorithm, which is used to compute parameter and standard error estimates.

Syntax

IT = <number>

where <number> denotes the maximum number of iterations.

Default

IT = <5q>

where <5q> is five times the number of free parameters of the model.

KA keyword

The KA keyword is used to specify the name of the text file for the estimates of the means of the exogenous latent variables.

Syntax

KA = <filename>

where <filename> denotes the name of the text file for the estimates.

LX keyword

The estimated matrix of factor loadings for the exogenous latent variables can be written to a text file by using the LX keyword.

Syntax

LX = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

LY keyword

The purpose of the LY keyword is to specify the name of the text file for the estimated matrix of factor loadings for the endogenous latent variables.

Syntax

LY = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

MA keyword

The MA keyword is used to specify the name of the text file for the moment matrix that was analyzed.

Syntax

MA = <filename>

where <filename> denotes the name of the text file for the moment matrix that was analyzed.

ME keyword

If the maximum likelihood method is not desired, other methods to fit the LISREL model to the data can be specified by using the ME keyword.

Syntax

ME = <method>

where <method> is one of the following:

IV   instrumental variables

TS   two-stage least squares

UL   unweighted least squares

GL   generalized least squares

ML   maximum likelihood

WL   weighted least squares

Default

ME = ML

MI option

The printing of the model modification indices in the output file is invoked by specifying the MI option.

MR option

The purpose of the MR option is to specify a MINRES exploratory factor analysis.

ND keyword

The purpose of the ND keyword is to specify the number of decimals for the results.

Syntax

ND = <number>

where <number> denotes the number of decimals desired.

Default

ND = 2

NP keyword

The NP keyword is used to specify the number of decimals for external text files to be produced.

Syntax

NP = <number>

where <number> denotes the number of decimals desired.

Default

NP = 3

NS option

The NS option is used to suppress the computation of internal starting values.

PC option

The PC option is used to invoke the printing of both the estimated asymptotic covariance and correlation matrices of the parameter estimators in the output file.

PH keyword

The PH keyword is used to specify the name of the text file for the estimated covariance (correlation) matrix of the exogenous latent variables.

Syntax

PH = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

PS keyword

The purpose of the PS keyword is to specify the name of the text file for the estimated covariance matrix of the error terms for the endogenous latent variables.

Syntax

PS = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

PT option

The PT option is used to invoke the printing of the technical details of the estimation method in the output file.

PV keyword

The PV keyword is used to specify the name of the text file for the estimates of all the free parameters of the LISREL model.

Syntax

>   PV = <filename>

where <filename> denotes the name of the text file to which the estimates are to be written.

RC keyword

The RC keyword is used to specify the ridge constant to be used if the matrix to be analyzed is not positive definite.

Syntax

>   RC = <value>

where <value> denotes the ridge constant.

Default

>   RC = 0.001

RO option

The purpose of the RO option is to invoke the use of the ridge constant for the moment matrix to be analyzed. The RO option will be invoked automatically if the matrix is not positive definite.

RS option

The RS option is used to invoke the printing of the residuals, standardized residuals, QQ-plot, and fitted covariance (or correlation, or moment) matrix in the output file.

SC option

The SC option is used to invoke the printing of the completely standardized solution in the output file.

SI keyword

The purpose of the SI keyword is to specify the name of the text file for the fitted moment matrix.

Syntax

> SI = <filename>

where <filename> denotes the name of the text file for the fitted moment matrix.

SL keyword

The SL keyword is used to specify the significance level of the model automated modification procedure expressed as a percentage when the automated modification procedure is desired.

Syntax

> SL = <number>

where <number> denotes the significance level expressed as a percentage.

Default

> SL = 1

SO option

The SO option is used to suppress the automated checking of the scale setting for each latent variable. It is needed for very special models where scales for latent variables are defined in a different way.

SS option

The SS option is used to invoke the printing of the standardized solution in the output file.

SV keyword

The purpose of the SV keyword is to specify the name of the text file for the standard error estimates of all the free parameters of the LISREL model.

Syntax

SV = <filename>

where <filename> denotes the name of the text file to which the standard error estimates are to be written.

TD keyword

The TD keyword is used to specify the name of the text file for the estimated covariance matrix of the measurement errors of the indicators of the exogenous latent variables.

Syntax

TD = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

TE keyword

The purpose of the TE keyword is to specify the name of the text file for the estimated covariance matrix of the measurement errors of the indicators of the endogenous latent variables.

Syntax

TE = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

TH keyword

The TH keyword is used to specify the name of the text file for the estimated covariance matrix between the measurement errors of the indicators of the endogenous latent variables and those of the exogenous latent variables.

Syntax

TH = <filename>

where <filename> denotes the name of the text file for the matrix of estimates.

TM keyword

The `TM` keyword can be used to specify the maximum number of CPU seconds allowed for the current analysis.

Syntax

TM = <number>

where <number> denotes the maximum number of CPU seconds allowed.

Default

TM = 172800

TV keyword

The purpose of the `TV` keyword is to specify the name of the text file for the t values of all the free parameters of the LISREL model.

Syntax

TV = <filename>

where <filename> denotes the name of the text file to which the t values are to be written.

TX keyword

The `TX` keyword is used to specify the name of the text file for the estimated vector of intercepts for the indicators of the exogenous latent variables.

Syntax

TX = <filename>

where <filename> denotes the name of the text file for the vector of estimates.

TY keyword

The `TY` keyword is used to specify the name of the text file for the estimated vector of intercepts for the indicators of the endogenous latent variables.

Syntax

TY = <filename>

where <filename> denotes the name of the text file for the vector of estimates.

WP option

The WP option is used to specify a column width of 132 for the output file. The default column width is 80 characters.

XA option

The purpose of XA option is to suppress the computation and printing of the additional χ^2 test statistic values. Only C1 (minimum fit function χ^2 value) will be computed. Standard error estimates are not affected. C1 is still an asymptotically correct chi-square for the GLS, ML, and WLS methods but not for ULS and DWLS methods. It is only intended for those who have very large models and cannot afford (or do not want) to let the computer run for an extended period.

XI option

The XI option is used to suppress the printing of the numerous measures of fit to the output file. If the XI option is specified, only the χ^2 value, degrees of freedom and corresponding p-value are printed.

XM option

The XM option is used to suppress the computation and printing of the modification indices. When a path diagram is requested, the indices will be computed, but will not be included in the output.

XO keyword

The purpose of the XO keyword is to specify the number of repetitions for which results should be written to the output file.

Syntax

    XO = <number>

where <number> denotes the number of repetitions for which results should be written to the output file.

Default

    XO = <nrep>

where <nrep> is the number of repetitions specified.

5.3.16 MEANS paragraph

The MEANS paragraph is used to provide the sample means of the observed variables of the model as part of the SIMPLIS syntax file. It is a **required** command only if a mean-and-covariance structure model is specified and the raw data file or DSF is not provided. Sample means can also be provided in the form of an external text file. In this case, the MEANS FROM FILE command rather than the MEANS paragraph is used (see Section 5.3.15).

Syntax

    MEANS
    <values>

where <values> denotes a list of sample means in free or fixed format. If a fixed format rather than a free format is used, a FORTRAN type format statement should be the first line of the MEANS paragraph.

Examples

    MEANS
    12.225 16.752 18.239 20.003 15.395

    MEANS
    (5F6.3)
    12.22516.75218.23920.00315.395

5.3.17 MEANS FROM FILE command

The MEANS FROM FILE command is used to specify the name of the text file that contains the sample means of the observed variables of the model. It is a **required** command only if a mean-and-covariance structure model is specified and the raw data file or DSF is not provided. Sample means can also be provided as part of the SIMPLIS syntax file. This is accomplished by using the MEANS paragraph instead of the MEANS FROM FILE command (see Section 5.3.14).

Syntax

    MEANS FROM FILE <filename>

where <filename> denotes the name of the text file containing the sample means.

Example

Suppose that the name of the text file with the sample means is **SELECT.MNS** and that it is located in the folder **Projects\SELECT** on the D drive. In this case, the corresponding MEANS FROM FILE command is given by

    MEANS FROM FILE 'D:\Projects\SELECT\SELECT.MNS'

5.3.18   MISSING VALUE CODE command

If the raw data to be processed include missing values, the MISSING VALUE CODE command is used to specify the global missing value. It is a **required** command if the Full Information Maximum Likelihood (FIML) method for data with missing values is to be used and the global missing value is not specified in the PSF.

Syntax

    MISSING VALUE CODE <value>

where <value> denotes the global missing value.

Example

Suppose that the missing values in a text data file are all listed as -100. The MISSING VALUE CODE command is then

    MISSING VALUE CODE -100

5.3.19   OBSERVED VARIABLES paragraph

The OBSERVED VARIABLES paragraph is used to provide descriptive names to the observed variables of the model. It is a **required** command, unless a PSF or a DSF is used. If the labels of the observed variables are in the form of a text file, the OBSERVED VARIABLES FROM FILE command instead of the OBSERVED VARIABLES paragraph is used (see Section 5.3.18).

Syntax

    OBSERVED VARIABLES
    <labels>

where <labels> denotes the descriptive names of the observed variables of the model. These names are provided in free or abbreviated format and only the first 8 characters of each name are utilized.

Examples

>   OBSERVED VARIABLES
>   Age Gender MSCORE SSCORE ESCORE

>   OBSERVED VARIABLES
>   JS1 – JS6 OC1 – OC10

5.3.20   OBSERVED VARIABLES FROM FILE command

If the labels of the observed variables are in the form of a text file, the OBSERVED VARIABLES FROM FILE command is used to specify descriptive names for the observed variables of the model. If a DSF or a PSF is not used, it is a **required** command. The labels of the observed variables can also be specified as part of the SIMPLIS syntax file. In this regard, the OBSERVED VARIABLES paragraph rather than the OBSERVED VARIABLES FROM FILE command is used (see Section 5.3.17).

Syntax

>   OBSERVED VARIABLES FROM FILE <filename>

where <filename> denotes the name of the text file containing the descriptive names of the observed variables of the model.

Example

Suppose that the name of the text file containing the latent variable labels is **ABUSE.LAB**, which is located in the folder **Projects\ABUSE** on the C drive. In this case, the corresponding OBSERVED VARIABLES FROM FILE command is given by

>   OBSERVED VARIABLES FROM FILE 'C:\Projects\ABUSE\ ABUSE.LAB'

5.3.21   OPTIONS command

The OPTIONS command is used to specify special analyses and to request additional results and it is an **optional** command. If the results in terms of the LISREL model are preferred, the LISREL OUTPUT command should be used (see Section 5.3.13).

Syntax

>   OPTIONS <options>

where <options> denotes a list of options for the analysis each of which either has the syntax:

<keyword> = <selection>

where <keyword> is one of AD, AL, BE, EP, GA, IT, KA, LX, LY, MA, ME, ND, NP, PH, PS, PV, RC, SI, SL, SV, TD, TE, TH, TM, TV, TX, TY or XO (see Section 5.3.13) and <selection> denotes a number, a value or a name; or the syntax:

<option>

where <option> is one of ALL, AM, DW, EF, FS, FT, MI, MR, NS, PC, PT, RO, RS, SC, SO, SS, WP, XA, XI or XM (see Section 5.3.13).

Examples

OPTIONS ND = 3 SC ME = DW AD = OFF

5.3.22   PATH DIAGRAM command

The PATH DIAGRAM command is used to generate a PTH file in which the results of the analysis are summarized in the form of a path diagram. It is an **optional** command.

Syntax

PATH DIAGRAM

5.3.23   PATHS paragraph

The PATHS paragraph may be used to specify the regression relationships of the structural equation model to be fitted to the data. Alternatively, the RELATIONSHIPS paragraph can be used to specify these relationships (see Section 5.3.24). It is a **required** command only if the RELATIONSHIPS paragraph is not used.

Syntax

PATHS
<paths>

where <paths> denotes a list of regression relationships each of which has the following syntax

<x> -> <y>

where <x> denotes a list of independent (exogenous) variable labels and <y> denotes a list of dependent (endogenous) variable labels. These lists can be in free format or in abbreviated form.

Example

```
PATHS
JS -> JS1 – JS7
OC -> OC1 OC3 OC7
OC -> JS
```

5.3.24   RAW DATA paragraph

The RAW DATA paragraph is used to provide the raw data to be analyzed as part of the SIMPLIS syntax file. It is a **required** command only if the raw data are provided as part of the SIMPLIS syntax file. If the raw data are in the form of a text file or a PSF, the RAW DATA FROM FILE command instead of the RAW DATA paragraph is used (see Section 5.3.23).

Syntax

```
RAW DATA
<values>
```

where <values> denotes the rows of the raw data matrix. These rows can be provided in free or fixed formats. However, in the case of a fixed format, the first line of the RAW DATA paragraph should be a FORTRAN format statement.

Examples

```
RAW DATA
1 7 5
2 1 7
5 5 5
2 2 4
3 3 1
5 6 6
7 7 7
1 1 1
1 2 1
6 6 7

RAW DATA
(2F6.3)
12.34514.417
16.24519.205
10.33411.276
15.11416.267
13.24715.589
```

5.3.25   RAW DATA FROM FILE command

The RAW DATA FROM FILE command is used to specify the name of the PSF or the text file containing the raw data. It is a **required** command only if a PSF or a text data file is to be processed. The raw data matrix can also be specified as part of the SIMPLIS syntax file. In this regard, the RAW DATA paragraph rather than the RAW DATA FROM FILE command is used (see Section 5.3.22).

Syntax

    RAW DATA FROM FILE <filename>

where <filename> denotes the name of the text data file or the PSF.

Example

Suppose that the name of the PSF containing the raw data is **SELECT.PSF** and that it is located in the folder **Projects\SELECT** on the E drive. In this case, the corresponding RAW DATA FROM FILE command is given by

    RAW DATA FROM FILE 'E:\Projects\SELECT\SELECT.PSF'

5.3.26   RELATIONSHIPS paragraph

The RELATIONSHIPS paragraph may be used to specify the regression relationships of the structural equation model. The PATH commands can also be used to specify these relationships (see Section 5.3.21). It is a **required** paragraph only if a PATH command is not used.

Syntax

    RELATIONSHIPS
    <relationships>

where <relationships> denotes a list of regression relationships each of which has the following syntax

    <y> = <x>

where <x> denotes a list of independent (exogenous) variable labels and <y> denotes a list of dependent (endogenous) variable labels. These lists can be in free format or in abbreviated form.

Example

```
PATHS
JS1 – JS7 = JS
OC1 OC3 OC7 = OC
JS = OC
```

5.3.27   SAMPLE SIZE command

The SAMPLE SIZE command is used to specify number of cases to be processed. It is a **required** command, unless a DSF or PSF is used.

Syntax

```
SAMPLE SIZE <number>
```

where <number> denotes the number of cases.

Example

```
SAMPLE SIZE 388
```

5.3.28   SET command

The SET command is used to specify the status and/or the value(s) of a parameter(s) of the model. It is an **optional** command.

Syntax

```
SET the <parameter> equal to <value>
SET the <parameter> Free
SET the <parameter1> and the <parameter2> Equal
```

where <parameter> is one of

```
Path <label1> -> <label2>
Variance of <label>
Covariance of <label1> and <label2>
Error Variance of <label>
Error Covariance of <label1> and <label2>
```

where <label>, <label1> and <label2> denote labels of observed or latent variables of the model and <value> denotes a real number.

Examples

> SET the Path Ses - >Alien67 and the Path Ses - >Alien71 Equal
>
> SET the Variance of Ses equal to 1.0
>
> SET the Error Covariance of ANOMIA67 and ANOMIA71 Free

5.3.29 STANDARD DEVIATIONS paragraph

The STANDARD DEVIATIONS paragraph is used to provide the sample standard deviations of the observed variables of the model as part of the SIMPLIS syntax file. It is a **required** command only if the covariance matrix is to be analyzed and a CORRELATION MATRIX FROM FILE command or CORRELATION MATRIX paragraph is used. Sample standard deviations can also be provided can also be provided in the form of an external text file. In this case, the STANDARD DEVIATIONS FROM FILE command rather than the STANDARD DEVIATIONS paragraph is used (see Section 5.3.28).

Syntax

> STANDARD DEVIATIONS
>
> <values>

where <values> denotes a list of sample standard deviations of the observed variables of the model in free or fixed format. If a fixed format rather than a free format is used, a FORTRAN format statement should be the first line of the STANDARD DEVIATIONS paragraph.

Examples

> STANDARD DEVIATIONS
>
> 13.61 14.76 14.13 14.90 10.90 3.749

> STANDARD DEVIATIONS
>
> (5F6.3)
>
> 12.22516.75218.23920.00315.395

5.3.30 STANDARD DEVIATIONS FROM FILE command

The STANDARD DEVIATIONS FROM FILE command is used to specify the name of the text file that contains the standard deviations of the observed variables of the model. It is a **required** command only if the covariance matrix is to be analyzed and a CORRELATION MATRIX FROM FILE command or a CORRELATION MATRIX paragraph is used. Sample standard deviations can also be provided as part of the SIMPLIS syntax file. This is accomplished by using the STANDARD DEVIATIONS paragraph instead of the STANDARD DEVIATIONS FROM FILE command (see Section 5.3.27).

Syntax

> STANDARD DEVIATIONS FROM FILE <filename>

where <filename> denotes the name of the text file containing the sample standard deviations.

Example

Suppose that the name of the text file with the sample standard deviations is **Depresion.std** and that it is located in the folder **Projects\DEPRESSION** on the E drive. In this case, the corresponding STANDARD DEVIATIONS FROM FILE command is given by

> STANDARD DEVIATIONS FROM FILE 'E:\Projects\DEPRESSION\Depresion.std '

5.3.31　STRATUM command

Complex surveys are typically obtained by stratifying the target population into subpopulations (strata). The STRATUM command allows the user to specify the stratification variable. Since other types of surveys are incorporated, the STRATUM command is an optional command. The STRATUM command corresponds with the STRATUM variable section on the Survey Design dialog box (see Section 5.2.8).

Syntax

> STRATUM <label>

where <label> denotes the label of the stratification variable.

Example

Suppose that the target population was stratified into census regions and that the variable CENREG is the variable used to indicate the census region for each observation. In this case, the STRATUM command is given by

> STRATUM CENREG

5.3.32　SYSTEM FILE FROM FILE command

The SYSTEM FILE FROM FILE command is used to specify the DSF to be processed. It is a **required** command only if a DSF is to be processed.

Syntax

> SYSTEM FILE FROM FILE <filename>

where <filename> denotes the name of the DSF.

Example

Suppose that the DSF file, **DEPRESSION.DSF**, which is located in the folder **Projects\DEPRESSION** on the F drive, is to be processed. In this case, the corresponding SYSTEM FILE FROM FILE command is given by

> SYSTEM FILE FROM FILE 'F:\Projects\DEPRESSION\DEPRESSION.DSF '

5.3.33 TITLE paragraph

The TITLE paragraph is used to specify a descriptive heading for the analysis. It is an **optional** command. If the TITLE paragraph is used, avoid using any words that correspond to other SIMPLIS commands or paragraphs in the string field.

Syntax

> TITLE
> <string>

where <string> denotes a character string.

Example

> TITLE
> A SIMPLIS syntax file for Example 6

5.3.34 WEIGHT command

Design weights are constructed for the ultimate sampling units of complex surveys. The purpose of the WEIGHT command is to allow the user to specify the design weight variable. Since surveys without design weights are permitted, the WEIGHT command is an optional command. The WEIGHT command corresponds with the DESIGN weight section on the Survey Design dialog box (see Section 5.2.8).

Syntax

> WEIGHT <label>

where <label> denotes the label of the design weight variable.

Example

Suppose that the variable USUWGT is used to capture the design weight for each observation. For this example, the WEIGHT command is given by

WEIGHT USUWGT
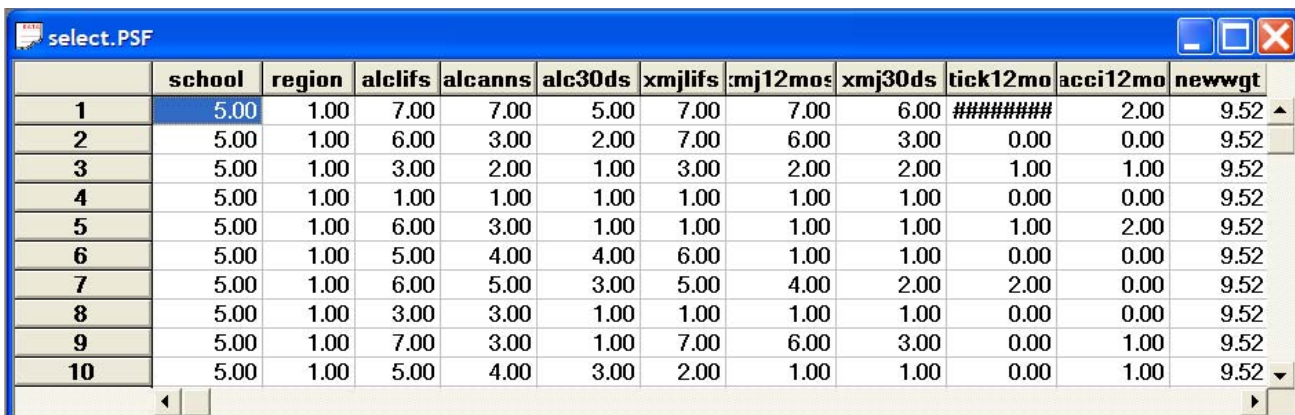
5.4   Examples

5.4.1   A structural equation model for the 2001 Monitoring the Future data

The data

The Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan has undertaken annual surveys designed to explore changes in important values, behaviors, and lifestyle orientations of contemporary American youth. The aims of these surveys are to provide a systematic, accurate description of the youth population of interest in a given year, and to explain relationships and trends observed over time. The Monitoring the Future surveys began in 1975. In the current example, data for 1608 respondents from the 2001 survey are used, and the focus is on relationships between the alcohol and marijuana use of respondents and traffic violations and/or accidents they were involved in.

Data for the first 10 participants on most of the variables used in this section are shown below in the form of a PSF named **select.psf** which can be found in the **TUTORIAL** folder.

| select.PSF	school	region	alclifs	alcanns	alc30ds	xmjlifs	mj12mos	xmj30ds	tick12mo	acci12mo	newwgt
1	5.00	1.00	7.00	7.00	5.00	7.00	7.00	6.00	########	2.00	9.52
2	5.00	1.00	6.00	3.00	2.00	7.00	6.00	3.00	0.00	0.00	9.52
3	5.00	1.00	3.00	2.00	1.00	3.00	2.00	2.00	1.00	1.00	9.52
4	5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	9.52
5	5.00	1.00	6.00	3.00	1.00	1.00	1.00	1.00	1.00	2.00	9.52
6	5.00	1.00	5.00	4.00	4.00	6.00	1.00	1.00	0.00	0.00	9.52
7	5.00	1.00	6.00	5.00	3.00	5.00	4.00	2.00	2.00	0.00	9.52
8	5.00	1.00	3.00	3.00	1.00	1.00	1.00	1.00	0.00	0.00	9.52
9	5.00	1.00	7.00	3.00	1.00	7.00	6.00	3.00	0.00	1.00	9.52
10	5.00	1.00	5.00	4.00	3.00	2.00	1.00	1.00	0.00	1.00	9.52

The following variables included in the PSF were selected from the survey data:

- school: This variable is used to indicate group membership of respondents within the 48 schools included in the survey.

- region: The 48 schools were drawn from 4 regions, and this variable indicates the region a school was drawn from.

- alclifs: The numerical response to the question "On how many occasions have you had alcoholic beverages to drink in your lifetime?"

- alc12mos: The numerical response to the question "On how many occasions have you had alcoholic beverages to drink in the past 12 months?"

- alc30ds: The numerical response to the question "On how many occasions have you had alcoholic beverages to drink in the past 30 days?"
- xmjlifs: The numerical response to the question "On how many occasions have you used marijuana in your lifetime?"
- xmj12mos: The numerical response to the question "On how many occasions have you used marijuana in the past 12 months?"
- xmj30ds: The numerical response to the question "On how many occasions have you used marijuana in the past 30 days?"
- tick12mo: The numerical response to the question "Within the last 12 months, how many times have you received a ticket (or been stopped and warned) for moving violations?"
- acci12mo: The numerical response to the question "Within the last 12 months, how many times you were involved in an accident while driving?"
- newwgt: The design weight of a student, computed as the inverse of the selection probability estimate of the region from which the student was selected. This selection probability estimate is merely the ratio of the sample frequency and the approximate population size of the region from which the student was selected.

The model

The five indicators or observed variables alclifs, alc30ds, xmjlifs, xmj12mos, and xmj30ds are modeled to measure the alcohol and marijuana usage. Alcohol and marijuana usage, represented by the latent variables ALCUSAGE and MRJUSAGE in our proposed model, are modeled as causes of the number of moving violations and accidents, as represented by the Eta variables ACCIDENT and TICKETS respectively. These two variables, in turn are measured without error by the two Y variables acci12mo and tick12mo. A path diagram of the model we intend fitting to the data is shown below.
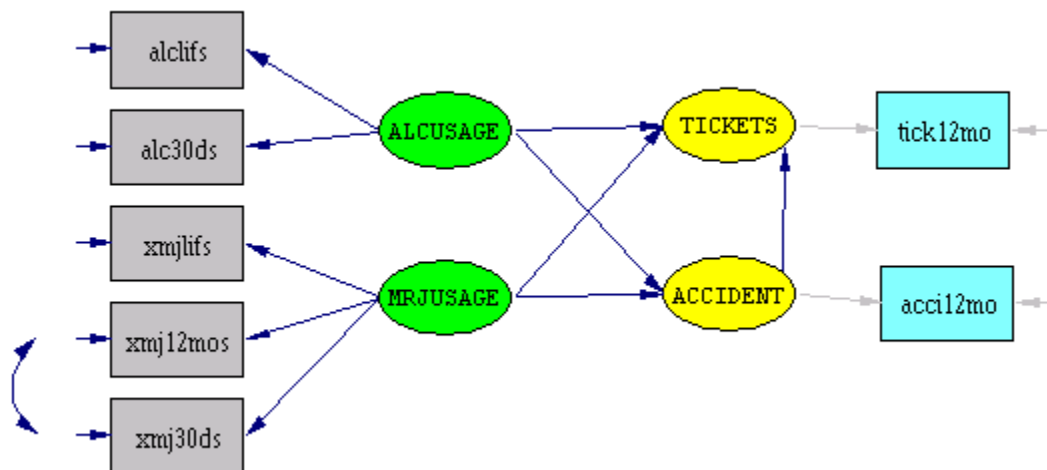
Mathematical Model

Measurement model

The measurement model for the latent variables ALCUSAGE, MRJUSAGE, ACCIDENT and TICKETS may be expressed as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{bmatrix} \begin{bmatrix} \mathbf{\eta} \\ \mathbf{\xi} \end{bmatrix} + \begin{bmatrix} \mathbf{\varepsilon} \\ \mathbf{\delta} \end{bmatrix}$$

where $\mathbf{y} = [\,\text{tick21mo} \;\; \text{acci12mo}\,]'$, $\mathbf{x} = [\,\text{alclifs} \;\; \text{alc30ds} \;\; \text{xmjlifs} \;\; \text{xmj12mos} \;\; \text{xmj30ds}\,]'$, $\mathbf{\eta} = [\,\text{TICKETS}$ ACCIDENT$\,]'$, $\mathbf{\xi} = [\,\text{ALCUSAGE MRJUSAGE}\,]'$, $\mathbf{\delta} = \begin{bmatrix} \delta_1 & \delta_2 & \delta_3 & \delta_4 & \delta_5 & \delta_6 \end{bmatrix}'$, $\mathbf{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 \end{bmatrix}'$,

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbf{\Lambda}_x = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \end{bmatrix}$$

where δ_1, δ_2, δ_3, δ_4, δ_5, δ_6, ε_1 and ε_2 denote measurement errors, and where λ_1, λ_2, λ_3, λ_4 and λ_5 denote unknown factor loadings.

Structural equation model

The structural equation model for the latent variables ALCUSAGE, MRJUSAGE, ACCIDENT and TICKETS is given by

$$\mathbf{\eta} = \mathbf{B}\mathbf{\eta} + \mathbf{\Gamma}\mathbf{\xi} + \mathbf{\zeta}$$

where $\mathbf{\zeta} = \begin{bmatrix} \zeta_1 & \zeta_2 \end{bmatrix}'$,

$$B = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{bmatrix}$$

where β, γ_1, γ_2, γ_3 and γ_4 denote unknown regression weights, and ζ_1 and ζ_2 denote error terms.

The survey design variables school and region will be used as stratification and cluster variable respectively, while the design weight as represented by the variable newwgt will also be included in the specification of the analysis, as illustrated next.

Preparing the data and setting up the analysis

The model is fitted to the data in **select.psf** by using the path diagram component (PTH window) of the LISREL GUI (See Section 5.2). After drawing the proposed model as a path diagram, SIMPLIS syntax is created and submitted. However, before we can fit the model, we need to specify the details of the complex survey design for the data in **select.psf**.

The first step is to open the PSF, which is accomplished as follows:

- o Use the **File**, **Open** option to activate the display of an **Open** dialog box.
- o Set the **Files of type** drop-down list box to **Prelis Data (*.psf)** and browse for the file **select.psf** in the **TUTORIAL** folder.
- o Select the file and click the **Open** button to open the PSF in a PSF window.

Preparing the data

Click on the **Survey Design** option on the **Data** menu to load the **Survey Design** dialog box. Select the variable region from the **Variables in data:** list box and click on the **Add** button of the **Stratification variable** section. Next, select the variable school and add this variable to the **Cluster variable** section in a similar fashion. Finally, add the weight variable by selecting the variable newwgt from the **Variables in data:** list box and add this variable in the **Design weight** section. The completed **Survey Design** dialog box is shown below. Click on the **OK** button to return to the PSF window, and click the **Save** option on the **File** menu.
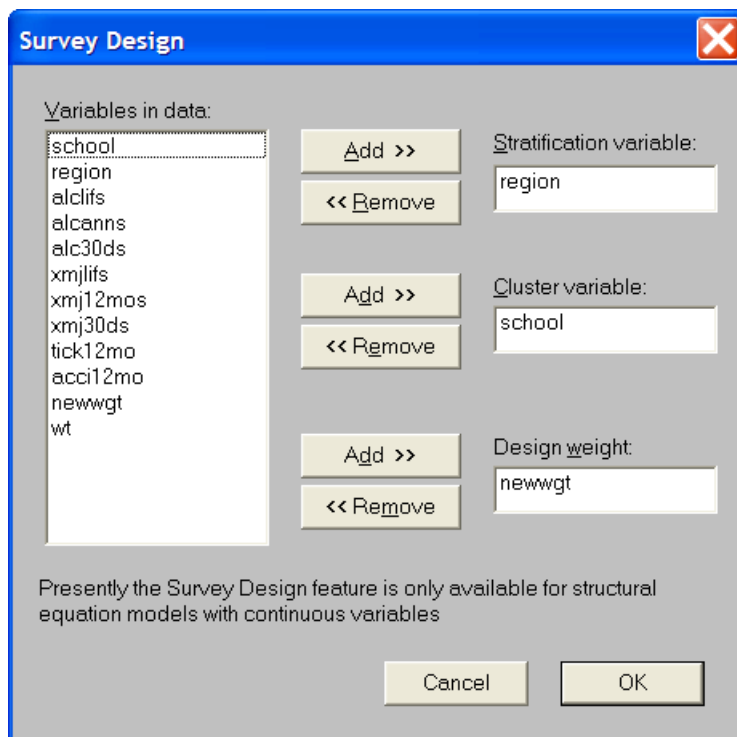
We now turn to creating a path diagram for the model to be fitted to these data. To open a new PTH window, select the **New** option on the **File** menu to load the **New** dialog box. Select the **Path Diagram** option from the list box on the **New** dialog box and provide a name for the path diagram, for example **select.pth**, in the **File name** string field of the **Save As** dialog box. Click on the **Save** button to open an empty PTH window.
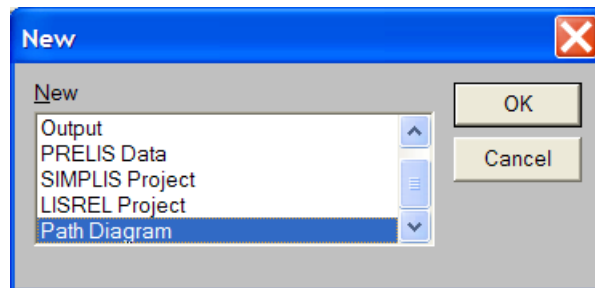

Select the **Title and Comments** option on the **Setup** menu to load the **Title and Comments** dialog box. Enter the title A model for traffic tickets and accidents in the **Title** string field, and click on the **Next** button to load the **Group Names** dialog box.

Click on the **Next** button to load the **Labels** dialog box. Click on the **Add/Read Variables** button to load the **Add/Read Variables** dialog box, and select the **PRELIS System File** option in the **Read from file:** drop-down list box. Click on the **Browse** button to load the **Browse** dialog box and select the file **select.psf** in the **TUTORIAL** folder. Click on the **OK** button to return to the **Labels** dialog box.
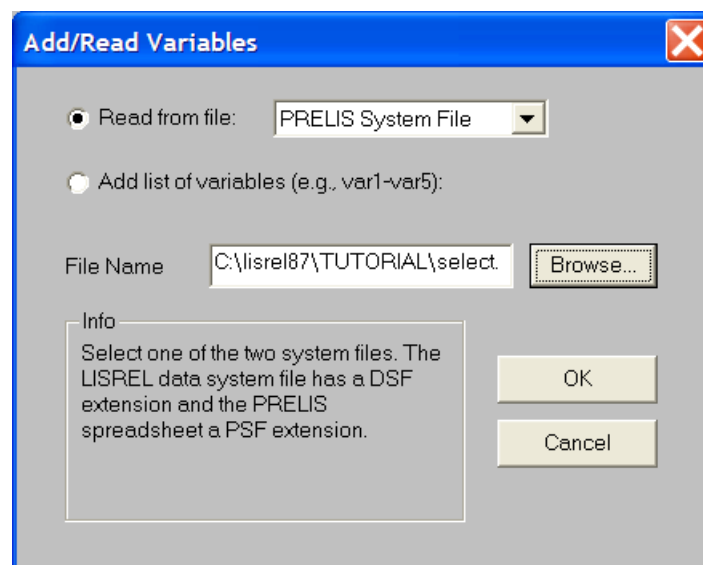
Click on the **Add Latent Variables** button to load the **Add Variables** dialog box. Enter the label ALCUSAGE in the string field and click **OK**. Enter the labels MRJUSAGE, ACCIDENT, and TICKETS in the same way.
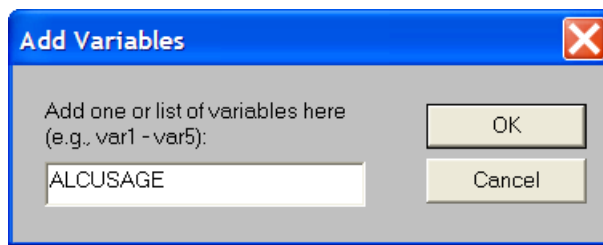
The completed **Labels** dialog box is shown below. Click on the **OK** button to return to the PTH window for **select.pth**.
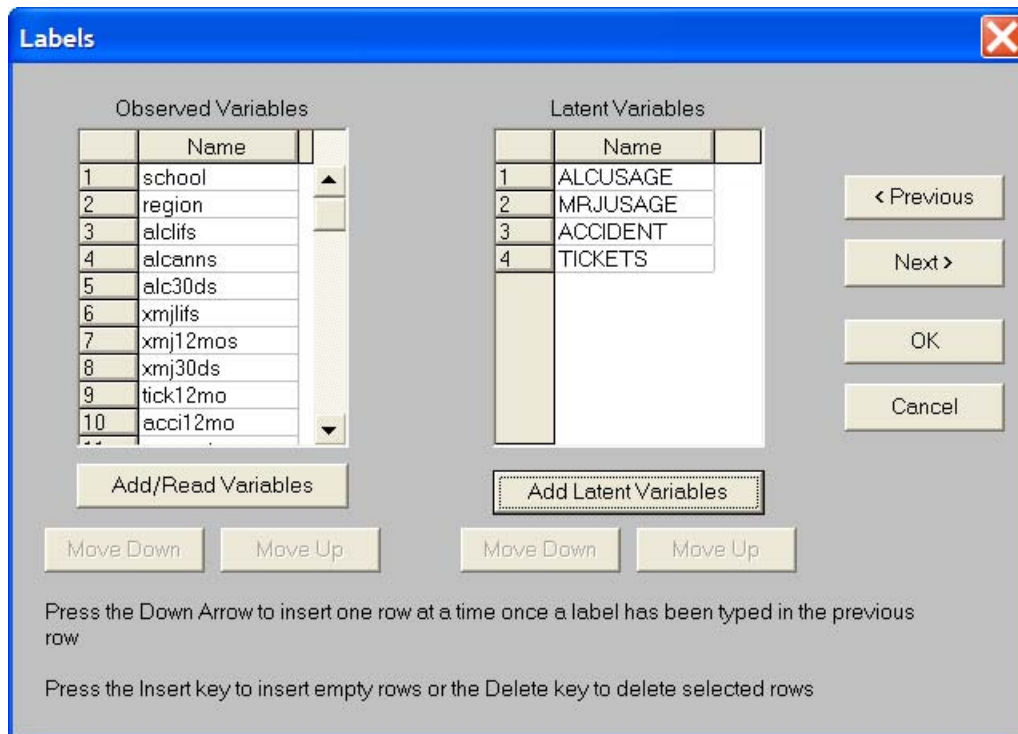
Setting up the analysis

At this point, an empty PTH window is displayed, with variable names listed to the left. Check the **Y** check boxes of acci12mo and tick12mo respectively. Check the **Eta** check boxes of ACCIDENT and TICKETS respectively to obtain the window shown below.
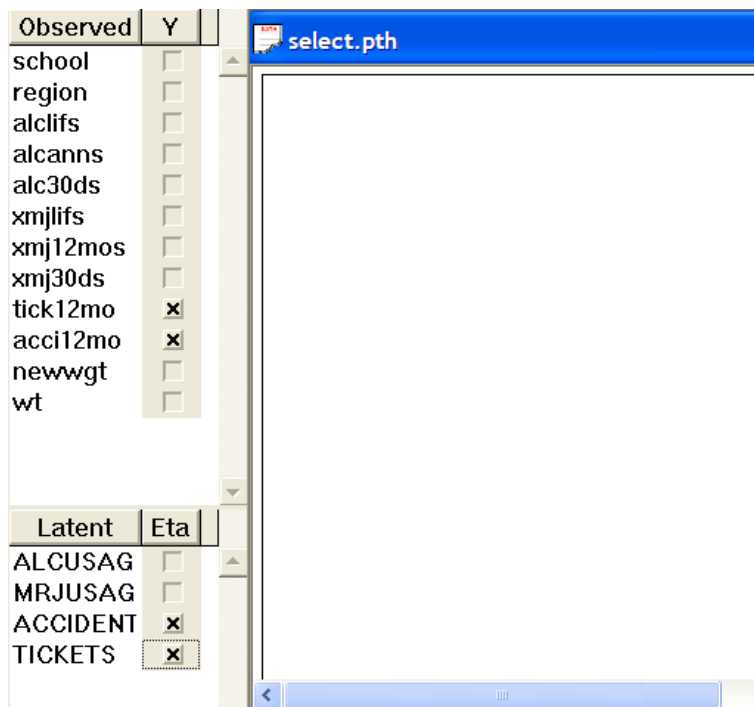
Next, click, drag and drop the labels of the Y variables one at a time into the PTH window. Position these variables to the right of the PTH window. Click, drag and drop the labels of the latent variables ACCIDENT and TICKETS one at a time into the PTH window to obtain the window shown below. Note that labels of variables dragged to the PTH window are shown against a colored background.
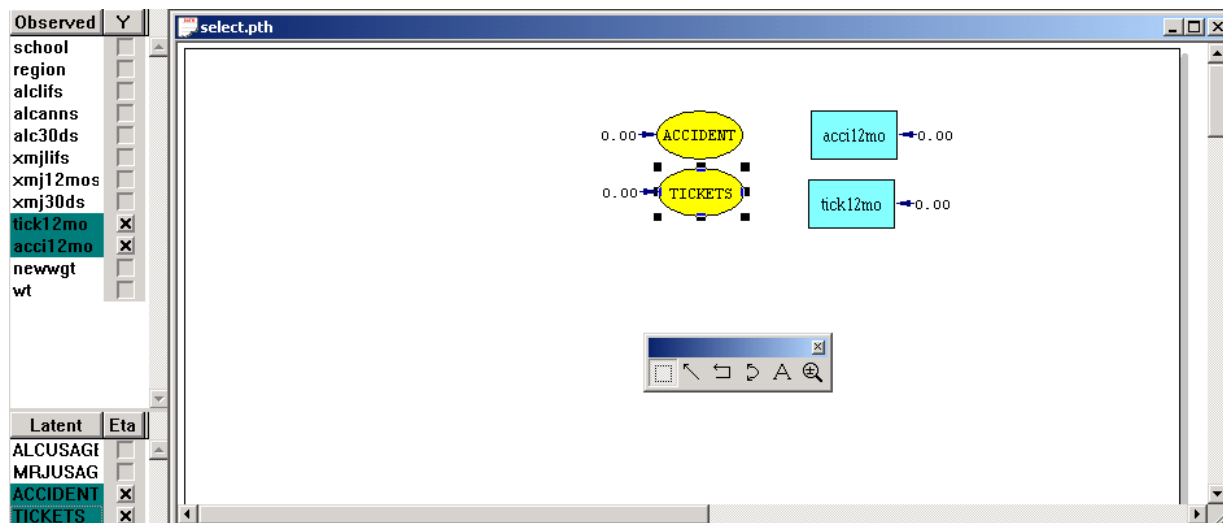
We now add the rest of the observed variables (alclifs, alc30ds, xmjlifs, xmj12mos, and xmj30ds) one at a time into the PTH window, positioning them to the left of the PTH window. The last variables to be added are the latent variables ALCUSAGE and MRJUSAGE.
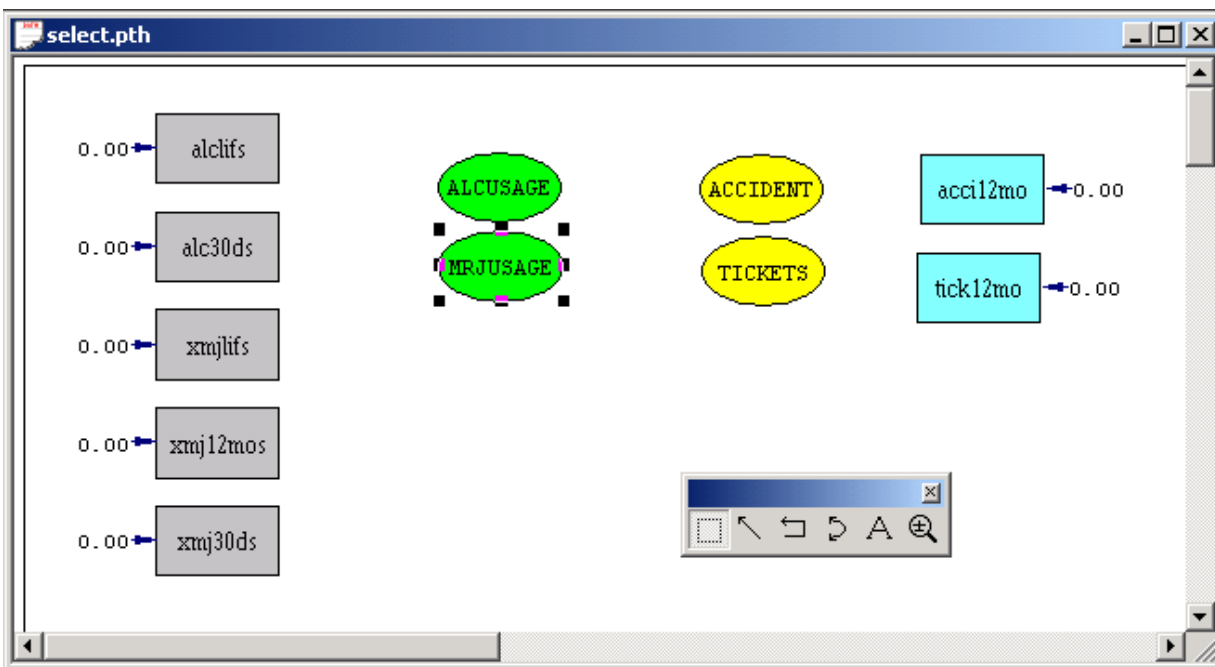
The next step is to add the paths between the variables dragged in the PTH window. Select the arrow icon on the **Drawing** toolbar, and click and drag indicator paths from the latent variable ALCUSAGE to alclifs and alc30ds respectively. To do so, start by clicking inside the ellipse representing ALCUSAGE and do not release the mouse button before the cursor is inside the rectangle representing alclifs or alc30ds. Click and drag similar indicator paths from the latent variable MRJUSAGE to xmjlifs, xmj12mos and xmj30ds respectively.

Structural paths from the latent variable ALCUSAGE to both ACCIDENT and TICKETS, and from the latent variable MRJUSAGE to ACCIDENT and TICKETS are added in the same way. Also add indicator paths from the latent variable ACCIDENT to acci12mo, from the latent variable TICKETS to tick12mo, and from TICKETS to ACCIDENT. The model should now look like the image below.
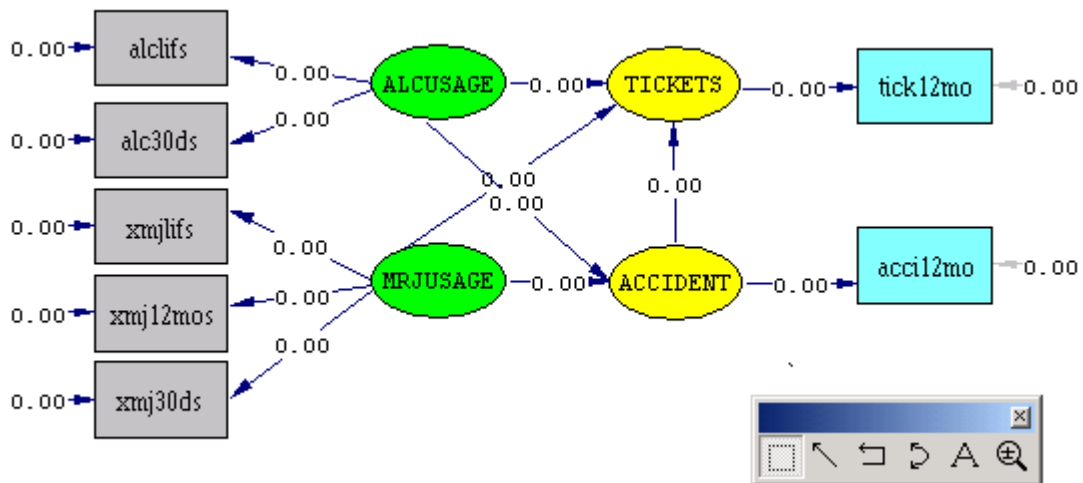
The two indicator paths ACCIDENT to acci12mo, and TICKETS to tick12mo have to be fixed to a value of 1.0. To do so, deselect the arrow icon on the **Drawing** toolbar by clicking on the selection icon to its left. Next, right click on the path between ACCIDENT to acci12mo and select the **Set Value** option from the pop-up menu that appears. Set the value to 1.0 and click **OK** to return to the PTH window.
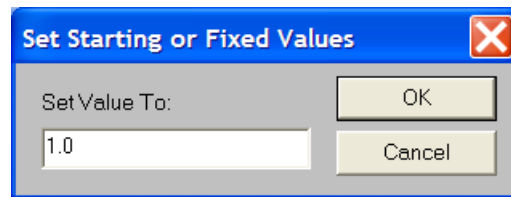

Right click on this path again, and select the **Fix** option from the pop-up menu. Note that the color representing the path has changed in the PTH window. Set the path between TICKETS to tick12mo to 1.0 in the same way.
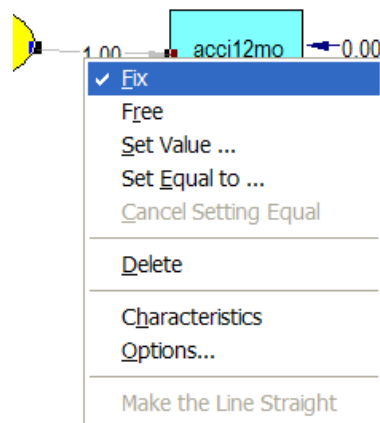

Finally, set the error variances of acci12mo and tick12mo to zero by right-clicking the error arrows and selecting the **Fix** option from the pop-up menu.
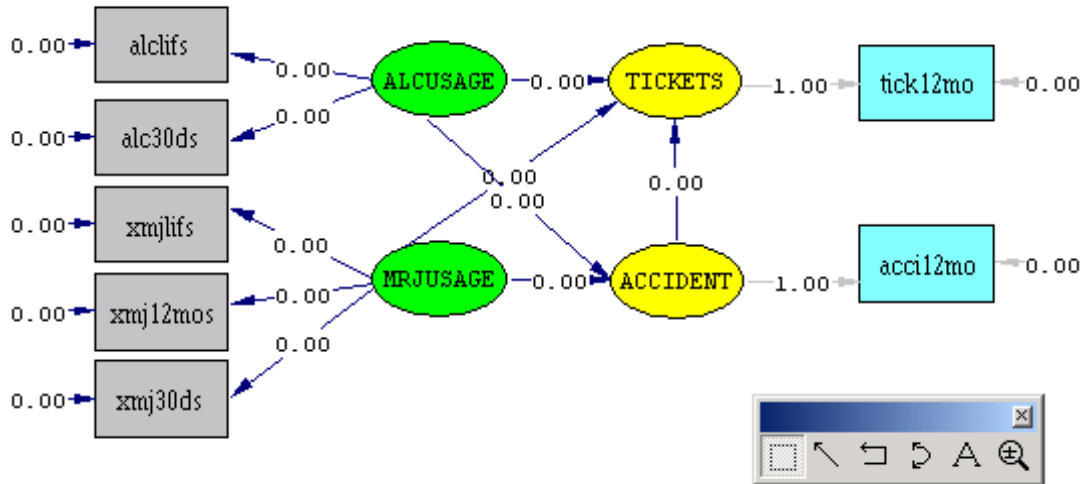
The last paths to be added to the path diagram are the covariance between the measurement errors of xmj12mos and xmj30ds. Select the double arrow icon on the **Drawing** toolbar, and click and drag a path between the error arrows of xmj12mos and xmj30ds. Be sure to position the cursor over each arrow before activating and releasing the mouse button.
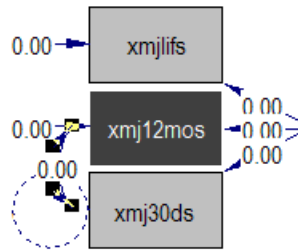
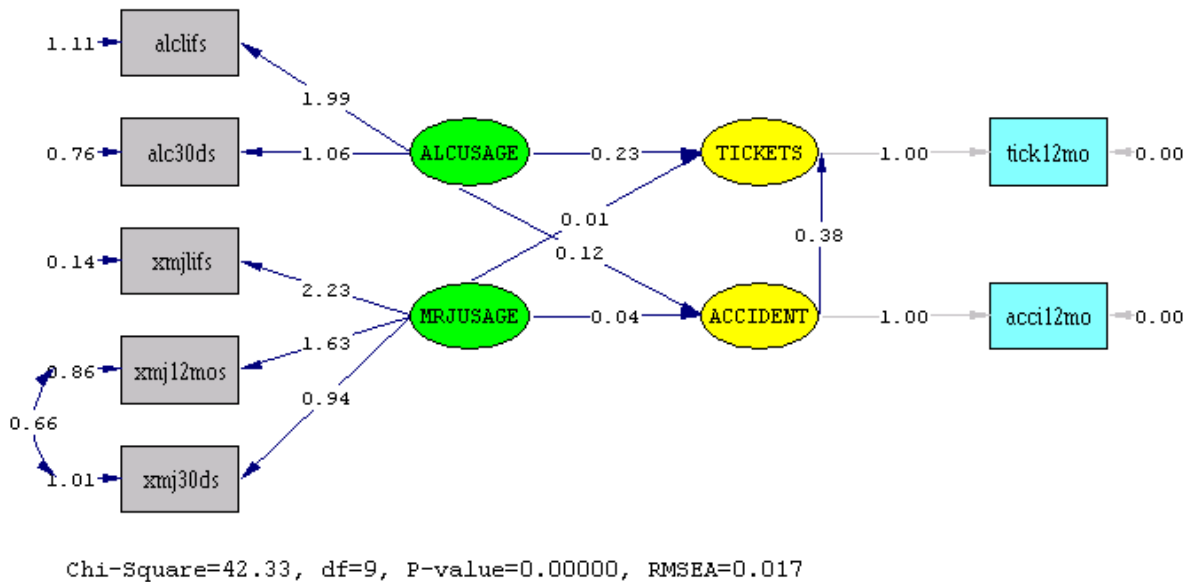The path diagram should look like the following image.
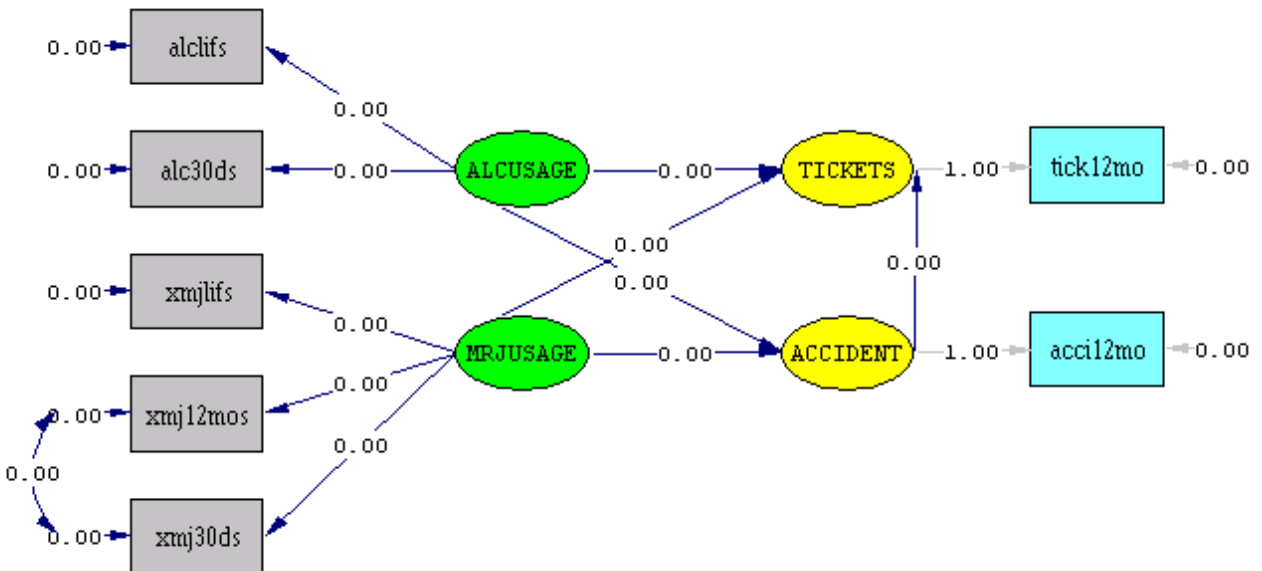
Select the **Build SIMPLIS Syntax** option on the **Setup** menu. The generated syntax is automatically displayed in a SPJ window, as shown below.

Click on the **Run LISREL** icon on the main toolbar to produce the following PTH window.

Discussion of results

Portions of the output file **select.out** are shown below.

From the results, it is evident that the five factor loadings are statistically significant if a 1% level of significance is used. In addition, the error covariance for xmj12mos and xmj30ds is also significant at a 1% level of significance. In other words, the results do not indicate any misspecifications in the measurement model of the latent variables ALCUSAGE and MRJUSAGE.

```
          LAMBDA-X

              ALCUSAGE    MRJUSAGE
              --------    --------
   alclifs      1.99        - -
               (0.05)
               39.21
   alc30ds      1.06        - -
               (0.04)
               28.72
   xmjlifs      - -         2.23
                           (0.05)
                           47.85
   xmj12mos     - -         1.63
                           (0.05)
                           32.32
   xmj30ds      - -         0.94
                           (0.07)
                           14.31

          THETA-DELTA
              alclifs    alc30ds    xmjlifs    xmj12mos    xmj30ds
              --------   --------   --------   --------    --------
   alclifs      1.11
               (0.16)
               6.72
   alc30ds      - -       0.76
                         (0.06)
                         12.37
   xmjlifs      - -       - -        0.14
                                    (0.08)
                                    1.76
   xmj12mos     - -       - -        - -        0.86
                                               (0.08)
                                               10.75
   xmj30ds      - -       - -        - -        0.66       1.01
                                               (0.04)     (0.11)
                                               17.07       9.45
```

Since $\hat{\beta} = 0.38$ $(t = 11.03)$, it follows that a student's number of accidents exerts a significant positive influence $(p < 0.01)$ on his/her number of traffic tickets. Thus an increase in the number of accidents corresponds to an increase in the number of traffic tickets. Similarly, it follows that the alcohol usage of a student is a significant antecedent $(p < 0.01)$ of both the number of accidents

and the number of traffic tickets of the student. On the other hand, the marijuana usage is not a statistical significant antecedent of both the student's number of accidents and traffic tickets.

The R^2 value for the number of accidents follows as $1 - 0.44 = 0.56$. In other words, the alcohol usage and marijuana usage of a student explains approximately 56% of the variation in the student's number of accidents. Similarly, it follows that they explain approximately 27% of the variation in the number of traffic accidents of the student.

```
            BETA

                 ACCIDENT    TICKETS
                 --------    --------
    ACCIDENT        - -        - -

    TICKETS         0.38       - -
                   (0.03)
                   11.03

            GAMMA

                 ALCUSAGE    MRJUSAGE
                 --------    --------
    ACCIDENT        0.12       0.04
                   (0.03)     (0.02)
                    4.70       1.93

    TICKETS         0.23       0.01
                   (0.05)     (0.04)
                    4.90       0.17

        PSI
        Note: This matrix is diagonal.

                 ACCIDENT    TICKETS
                 --------    --------
                    0.44       0.73
                   (0.03)     (0.05)
                   16.10      14.42
```

```
        Global Goodness of Fit Statistics, Missing Data Case

    Chi-Square Scale Factor =        0.52319E-01
    Number of Strata    =                 4
    Number of Clusters =                 28


      -2ln(L) for the saturated model =        276552.786
      -2ln(L) for the fitted model    =        277361.917

                  Degrees of Freedom = 9
        Full Information ML Chi-Square   = 42.33 (P = 0.00)
    Root Mean Square Error of Approximation (RMSEA) = 0.017
  90 Percent Confidence Interval for RMSEA = (0.012 ; 0.022)
      P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
```

From the results above, it is evident that the χ^2 test statistic value for the null hypothesis of a perfect fit is significant if a 1% level of significance is used. There is sufficient evidence that the theoretical model does not fit the data perfectly. However, the RMSEA point estimate of 0.017 indicates that the model does provide a close fit to the data (Browne & Cudeck, 1993).

5.4.2   Implementation of sampling weights in a linear growth curve model

The data

A linear growth curve model with two dummy coded covariates (Lang1 and Lang2) is fitted to a simulated dataset contained in the PSF **surveysem.psf**, contained in the **MISSINGEX** folder. It is assumed that the data are stratified according to 48 counties.

Within each county three schools are selected as primary sampling units (PSUs). In school number 1 four students are selected; in schools 2 and 3, three students are selected from each school. Students were selected on the basis of their initial achievement in an aptitude test (Score1) and measurements were repeated over six time intervals for five students from each school and over four time intervals for the remaining five.

The table below (Weight3) shows the weight calculations based on standardized initial scores.

Interval	Lower	Upper	% Expected	% Selected	Weight3
1	-Inf	-1.00	15.87	10.00	1.587
2	-1.00	-0.70	8.33	10.00	0.833
3	-0.70	-0.20	17.88	10.00	1.788
4	-0.20	0.00	7.93	10.00	0.793
5	0.00	0.30	11.79	10.00	1.179
6	0.30	1.00	22.34	10.00	2.234
7	1.00	1.30	6.19	10.00	0.619
8	1.30	1.80	6.09	10.00	0.609
9	1.80	2.30	2.52	10.00	0.252
10	2.30	Inf	1.07	10.00	0.107

Ten students were selected from each school as follows:

- Four from racial group 1 with Weight2 = 7.0/4.0
- Three from racial group 2 with Weight2 = 2.0/3.0
- Three from racial group 3 with Weight2 = 1.0/3.0

Final weights are obtained as follows:

Final_wt = Weight3*Weight2*10.0.

Multiplication of the weights by a factor of 10.0 was done to illustrate that a constant scaling of the weights does not affect parameter estimates, standard error estimates or the chi-square goodness of fit statistic value.

The data were simulated according to the following model.

$$Score_{ij} = a_{0i} + a_{1i}t_j + e_{ij}, \quad j = 1, 2...6$$

In the model, $t_j = (j-1)$ and i denotes student number i. In simulating the data set, it was assumed that

$$a_{0i} = \alpha_0 + \gamma_1 Lang1 + \gamma_2 Lang2 + u_{0i}$$
$$a_{1i} = \alpha_1 + u_{1i}$$

where

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}$$
$$\psi_{11} = \text{var}(u_{0i}) = 2$$
$$\psi_{21} = \text{cov}(u_{0i}, u_{1i}) = 0.6$$
$$\psi_{22} = \text{var}(u_{1i}) = 0.4$$
$$e_{ij} \sim N(0, I\sigma^2), \quad \sigma^2 = 1$$

and

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1.0 \end{pmatrix}$$

The first 10 records of the data set are shown below.

|	County	School	Score1	Score2	Score3	Score4	Score5	Score6	Lang1	Lang2	Final_wt
1	1.000	1.000	-1.836	-0.893	-1.210	-3.238	-1.161	-1.300	0.000	0.000	27.765
2	1.000	1.000	-0.066	2.988	0.918	3.626	-9.000	-9.000	0.000	0.000	27.765
3	1.000	1.000	0.739	3.351	0.693	-0.808	0.226	1.713	0.000	0.000	27.765
4	1.000	1.000	1.911	3.171	3.353	6.250	-9.000	-9.000	0.000	0.000	27.765
5	1.000	2.000	0.653	0.427	-2.252	-2.683	-3.232	-6.432	1.000	0.000	10.577
6	1.000	2.000	0.715	1.523	2.675	6.776	-9.000	-9.000	1.000	0.000	10.577
7	1.000	2.000	-0.311	0.462	3.083	0.968	3.284	4.457	1.000	0.000	10.577
8	1.000	3.000	1.392	0.081	0.923	0.048	-9.000	-9.000	0.000	1.000	5.289
9	1.000	3.000	0.828	1.414	4.930	5.560	4.106	6.317	0.000	1.000	5.289
10	1.000	3.000	-1.075	-0.624	-1.772	0.359	-9.000	-9.000	0.000	1.000	5.289

surveysem.PSF

Note that even-numbered cases have missing values on Score5 and Score6. This is indicated with a missing value code of -9.00.

The model

The conceptual path diagram for the model is shown below. The paths from the latent variables intcept and time to the dependent variables Score1 to Score6 are shown in gray to indicate that the corresponding coefficients are fixed values.
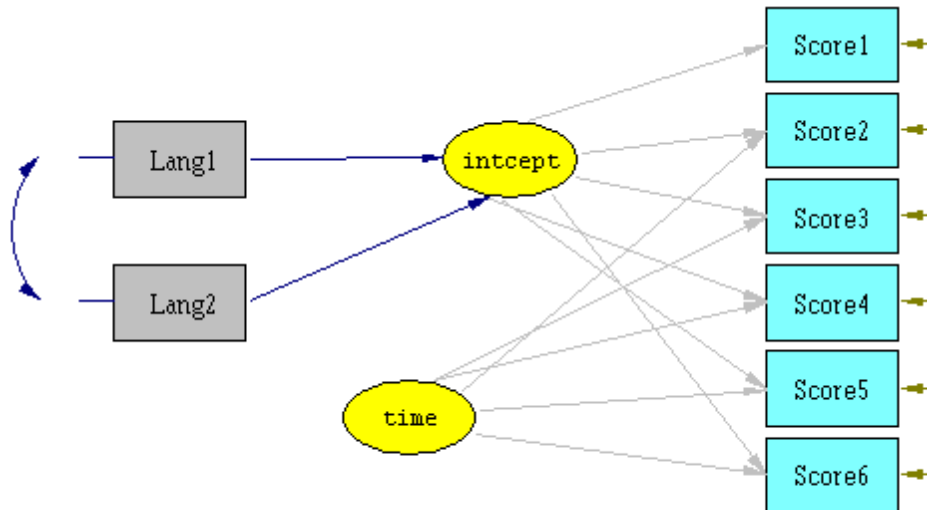
The conceptual path diagram for the structural part of the model is given below and indicates that allowance is made for the latent variables to be correlated.
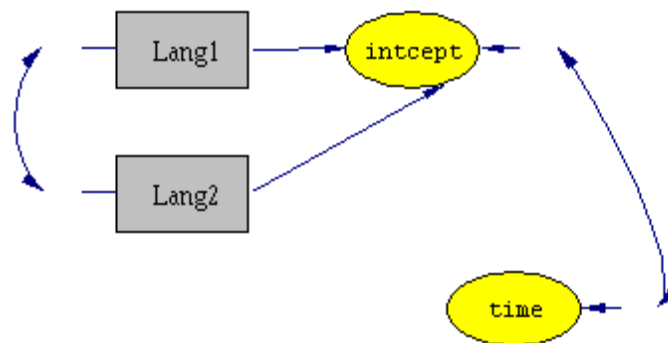
Mathematical Model

Measurement model

The measurement model for the latent variables may be expressed as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{bmatrix} \begin{bmatrix} \mathbf{\eta} \\ \mathbf{\xi} \end{bmatrix} + \begin{bmatrix} \mathbf{\varepsilon} \\ \mathbf{\delta} \end{bmatrix}$$

where $\mathbf{y} = [\text{Score1} \quad \text{Score2} \quad .... \quad \text{Score6}]'$, $\mathbf{x} = [\text{Lang1} \quad \text{Lang2}]'$, $\boldsymbol{\eta} = [\text{intcept} \quad \text{time}]'$, $\boldsymbol{\xi} = [\text{Lang1}$
$\text{Lang2}]'$, $\boldsymbol{\delta} = [\delta_1 \quad \delta_2]'$, $\boldsymbol{\varepsilon} = [\varepsilon_1 \quad ... \quad \varepsilon_6]'$, and $Cov(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}$. Also

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix},$$

$$Cov(\boldsymbol{\xi}) = \begin{bmatrix} \phi_{11} & \phi_{21} \\ \phi_{21} & \phi_{22} \end{bmatrix},$$

$$E(\boldsymbol{\xi}) = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix},$$

$$\boldsymbol{\Lambda}_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and $Cov(\boldsymbol{\delta}) = \mathbf{0}$.

Structural equation model

The structural equation model for the latent variables intcept and time is given by ($\mathbf{B} = \mathbf{0}$)

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

where $\boldsymbol{\zeta} = [\zeta_1 \quad \zeta_2]'$, with

$$Cov(\boldsymbol{\zeta}) = \begin{bmatrix} \psi_{11} & \psi_{21} \\ \psi_{21} & \psi_{22} \end{bmatrix},$$

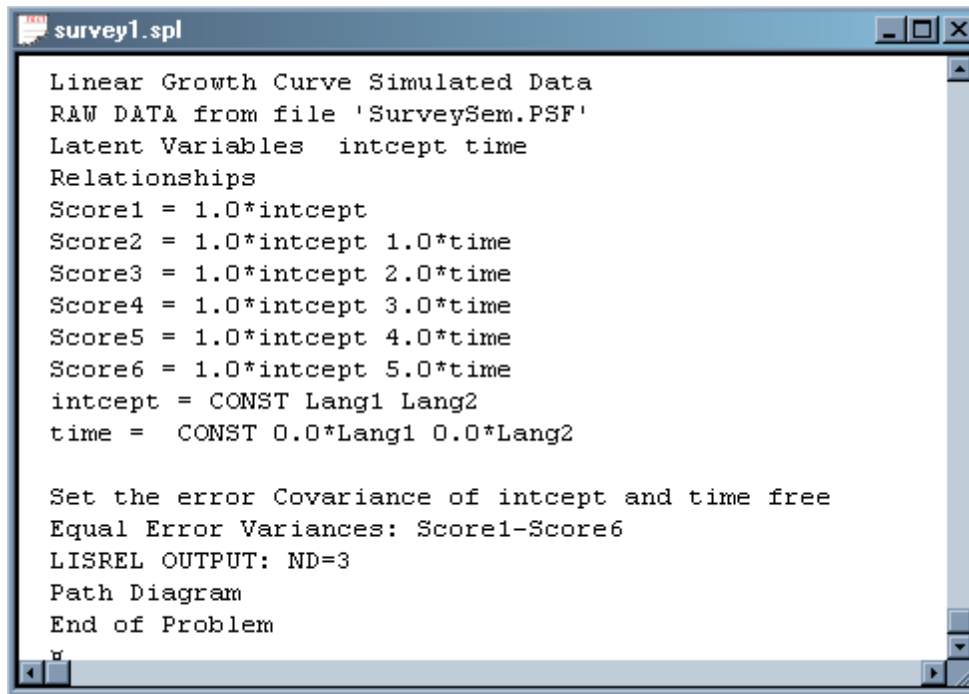$$E(\boldsymbol{\eta}) = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix},$$

and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ 0 & 0 \end{bmatrix}$$

where γ_1 and γ_2 denote unknown regression weights. The thirteen unknown model parameters are γ_1, γ_2, ψ_{11}, ψ_{12}, ψ_{22}, α_1, α_2, σ_ε^2, ϕ_{11}, ϕ_{21}, ϕ_{22}, κ_1, and κ_2.

The unrestricted model has $36 + 8 = 44$ parameters, since there are 8 observed variables and the number of non-duplicated elements of a covariance matrix of order 8 is 36. Therefore, the number of degrees of freedom equals $44 - 13 = 31$.

SIMPLIS syntax for the model is shown next. Note that 5.0*time, for example, indicates that the coefficient for the time→ Score6 path is fixed at a value of 5.0.

```
survey1.spl                                         _□×

  Linear Growth Curve Simulated Data
  RAW DATA from file 'SurveySem.PSF'
  Latent Variables  intcept time
  Relationships
  Score1 = 1.0*intcept
  Score2 = 1.0*intcept 1.0*time
  Score3 = 1.0*intcept 2.0*time
  Score4 = 1.0*intcept 3.0*time
  Score5 = 1.0*intcept 4.0*time
  Score6 = 1.0*intcept 5.0*time
  intcept = CONST Lang1 Lang2
  time =   CONST 0.0*Lang1 0.0*Lang2

  Set the error Covariance of intcept and time free
  Equal Error Variances: Score1-Score6
  LISREL OUTPUT: ND=3
  Path Diagram
  End of Problem
```

An experienced LISREL user may prefer to type the SIMPLIS commands to fit a specific model. However, we provide an outline of the steps required to build the syntax by drawing a path diagram. With more experience, users will find many shortcuts in developing the syntax. For example, it may be easier to type in the 1.0*, 2.0*, etc. values after syntax has been generated than to fix and set each path to a specific value.

Example: Implementation of sampling weights in a linear growth curve model

Describing the data

From the main menu bar, select the **Data**, **Survey Design** option and add the variable Final_wt in the **Design weight***:* box. Click **OK** when done.
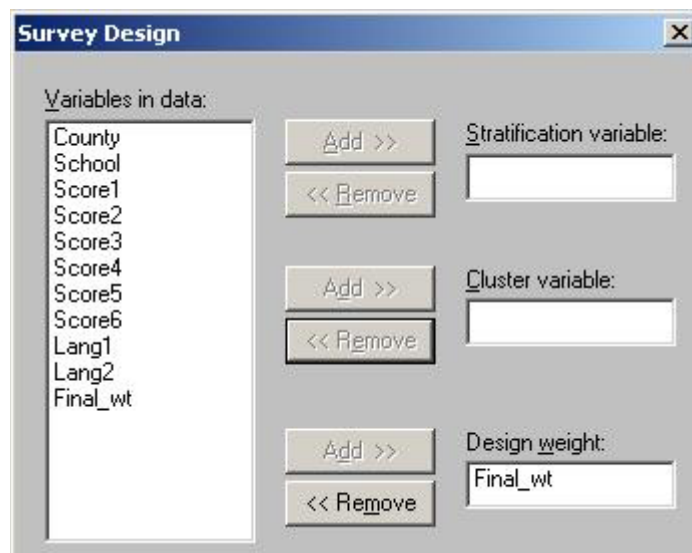
Next use the **Data**, **Define Variables** option to select the variables Score5 and Score6, then click the **Missing Values** button to invoke the **Missing Values** dialog box. Enter -9.0 as shown below. Click **OK**, then use the **File**, **Save** option to ensure that these changes are contained in the PSF file.
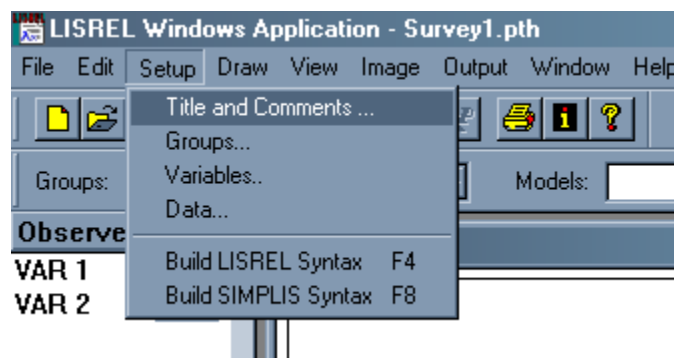
Setting up the analysis

To generate the SIMPLIS commands interactively, we proceed as follows. Using the **File**, **New** option, select the **Path Diagram** option from the **New** dialog box.
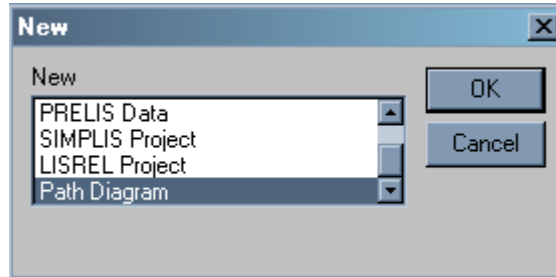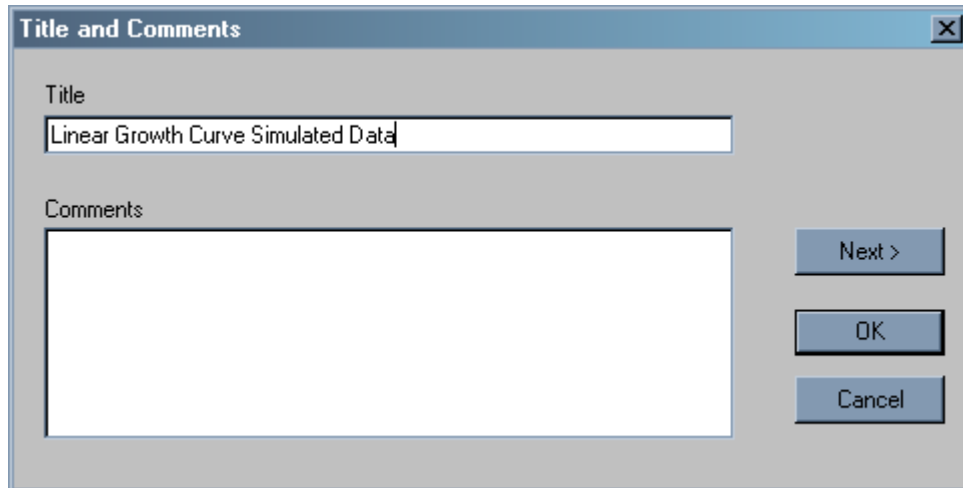

Save the new path diagram in the **MISSINGEX** folder as **Survey1**.**pth**.
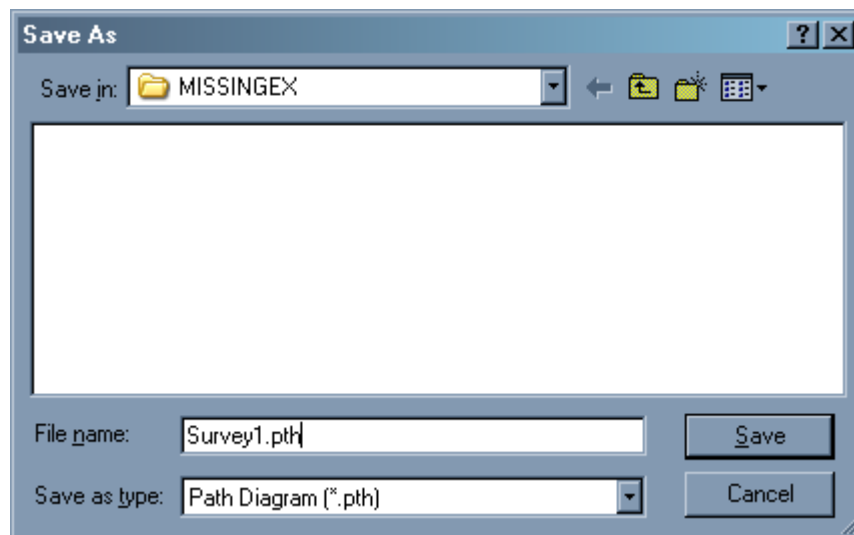

Click **Save** when done. From the **Setup** menu, select **Title and Comments**.

This action loads the **Title and Comments** dialog box shown below. Enter a title and any (optional) comments as shown below.

Click **Next** to proceed to the **Groups** dialog and, since this is a single-group example, click **Next** again to activate the **Labels** dialog shown below. To add a list of observed variables, click the **Add/Read Variables** button below the **Observed Variables** list box.
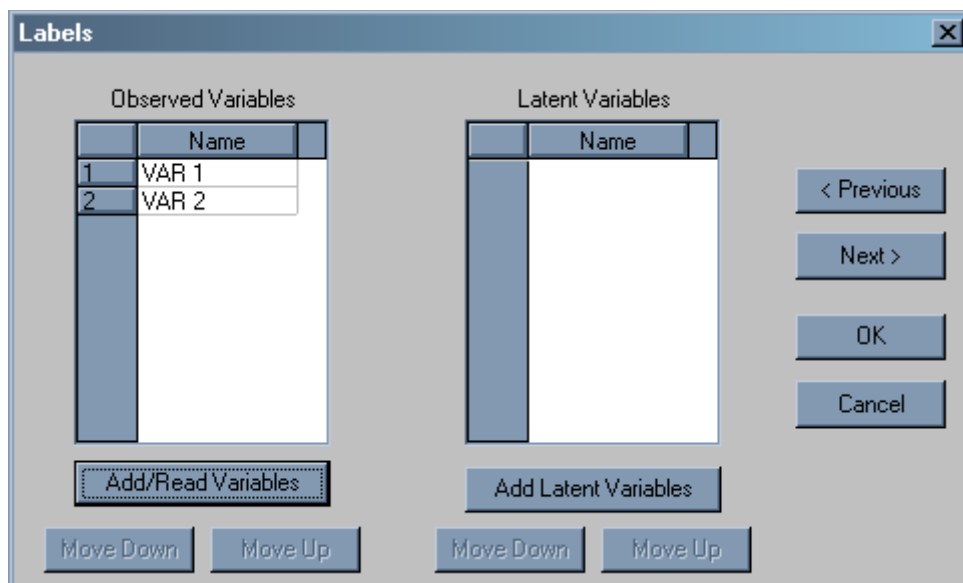
From the **Add/Read Variables** dialog box, click the **Read from file:** radio button and select **PRELIS System File**. Next, use the **Browse** button to locate and select **surveysem.psf**.
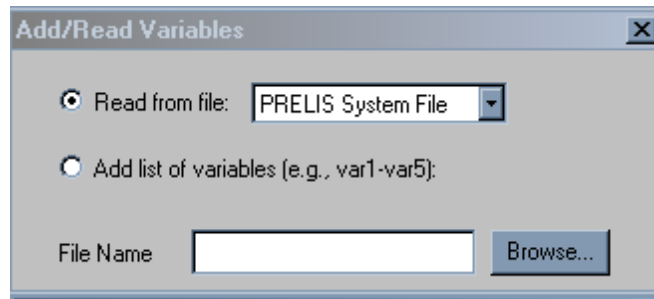
Click the **Open** button once the desired PSF is selected. The observed variable names will be displayed in the **Labels** dialog box. To add a list of latent variables, click the **Add Latent Variables** button and type the names of the latent variables one at a time.
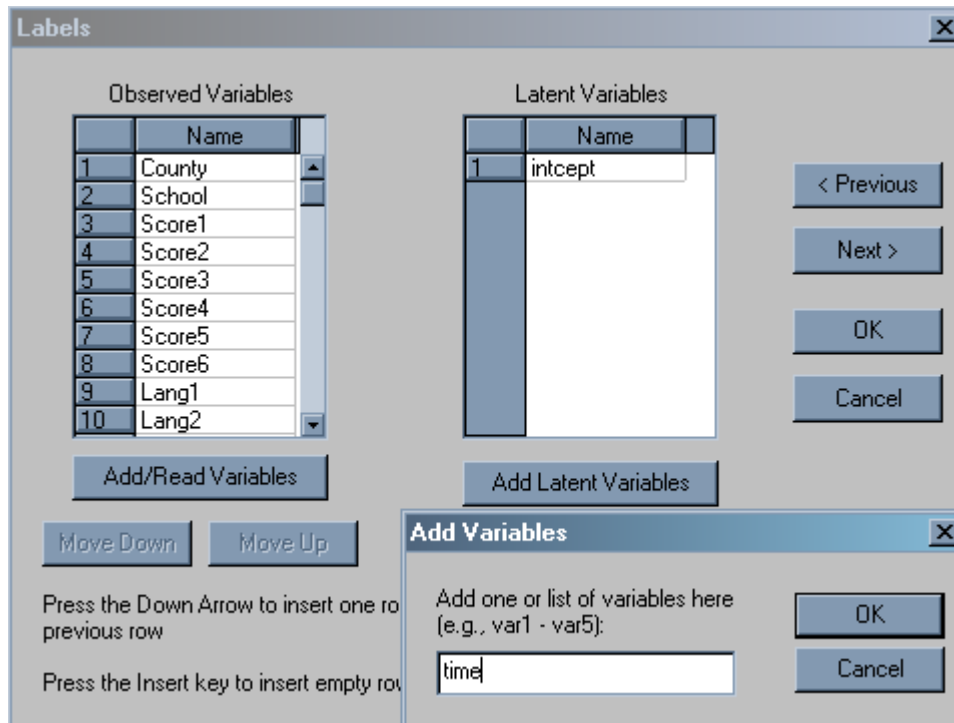
Click **OK** when the latent variables intcept and time are entered. The left hand side of the path diagram window should display the observed and latent variables. If not, select the **View**, **Toolbars** option and from the drop-down list **Select Variables**. Click on the check boxes on the right hand side of the variable names to define Score1 to Score6 as Y (dependent) variables and intcept and time as Eta (endogenous latent) variables.

Observed	Y	
County	☐	
School	☐	
Score1	☒	
Score2	☒	
Score3	☒	
Score4	☒	
Score5	☒	
Score6	☒	
Lang1	☐	
Lang2	☐	
Final_wt	☐	
CONST	☐	
Latent	**Eta**	
intcept	☒	
time	☒	

We start drawing the path diagram by dragging the names Score1 to Score6 to the path diagram window. Next, drag intcept and time to the middle of the path diagram window. A variable is dragged to the path diagram window by left-clicking on the variable name and then moving it with the left mouse button held down. Finally, drag Lang1, Lang2 and CONST to the left of intcept and time. Click the one-sided arrow on the drawing bar and connect arrows from intcept to the variables Score1, …, Score6.
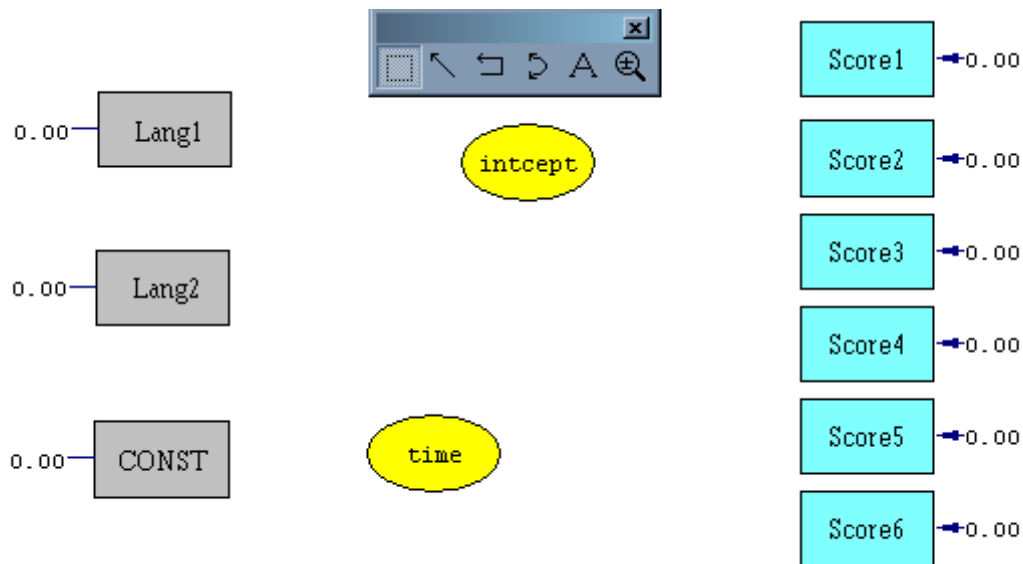
With the left mouse button down, start in the ellipse and drag the arrow to within a rectangle representing one of the Score variables before releasing the mouse button. Unselect the arrow by clicking on the square on the drawing bar. Once this is done, move the mouse pointer to each arrow

and right-click. From the pop-up menu, select **Fix**. Repeat this for each path from intcept to a Score value.
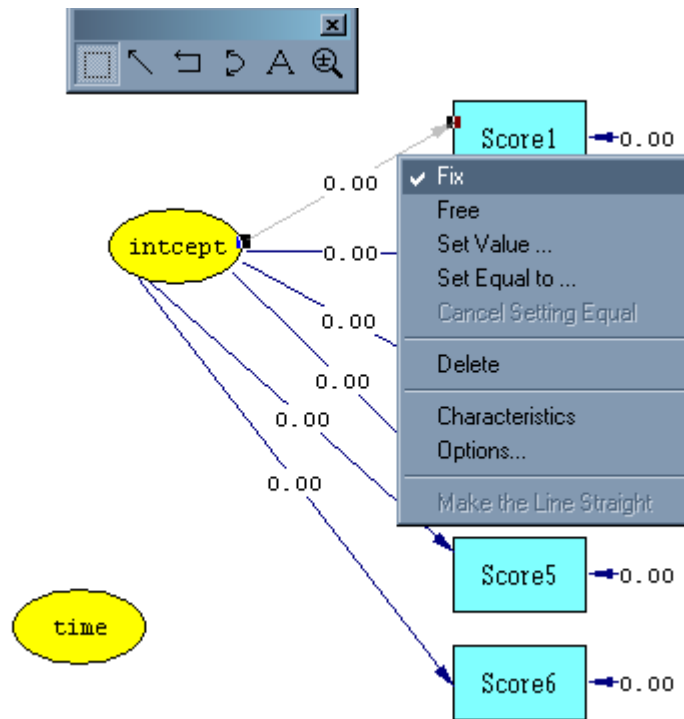
Once all the paths are fixed, start with the path from intcept to Score1. Right-click on the arrow, and select the **Set Value** option from the pop-up menu.
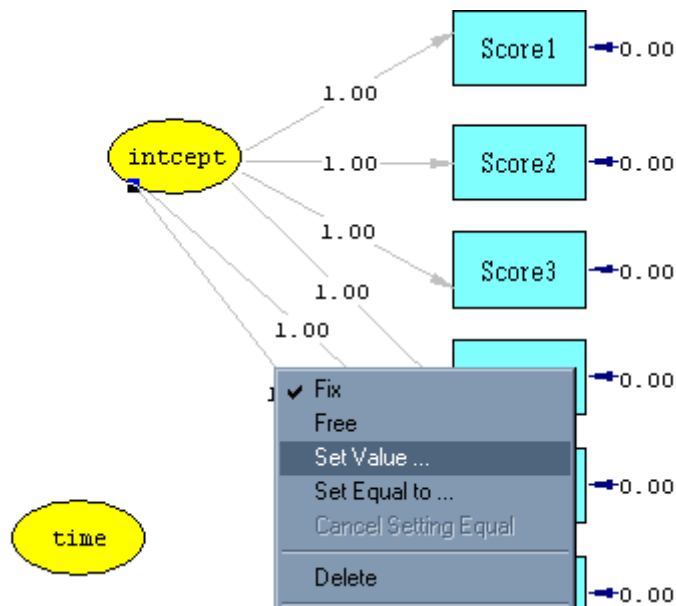
Change the value of each path to 1 as shown below.
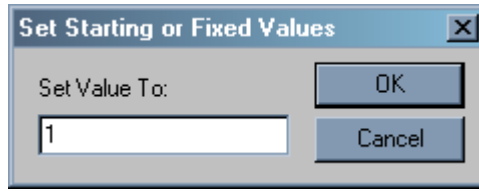
Repeat the above procedure by drawing paths from time to Score2, Score3, …, Score6. Fix these paths and set the path coefficients to 1, 2, 3, 4, and 5 respectively, as shown below.
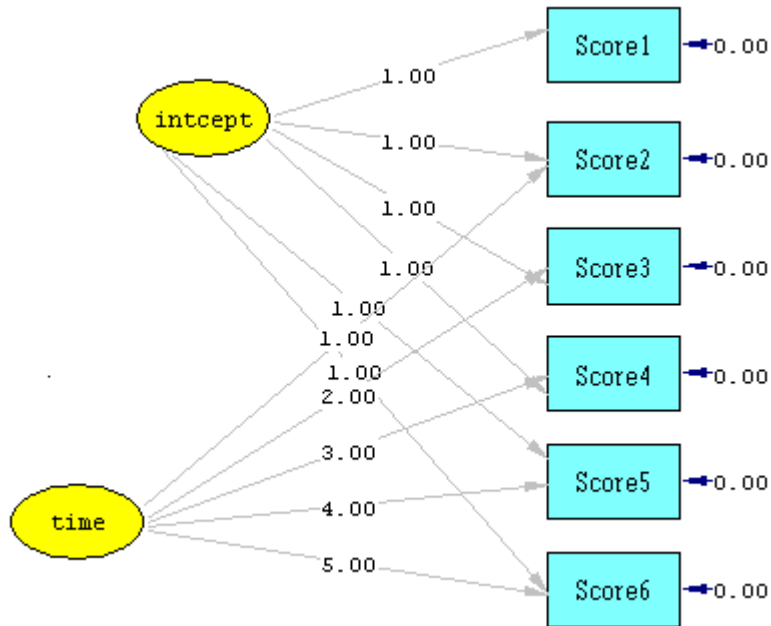
The path diagram is completed by drawing arrows from Lang1, Lang2 and the SIMPLIS variable CONST to intcept and time as shown. Once this is done, select the two-sided arrow (error covariance or factor correlation) to add a covariance path between Lang1 and Lang2.
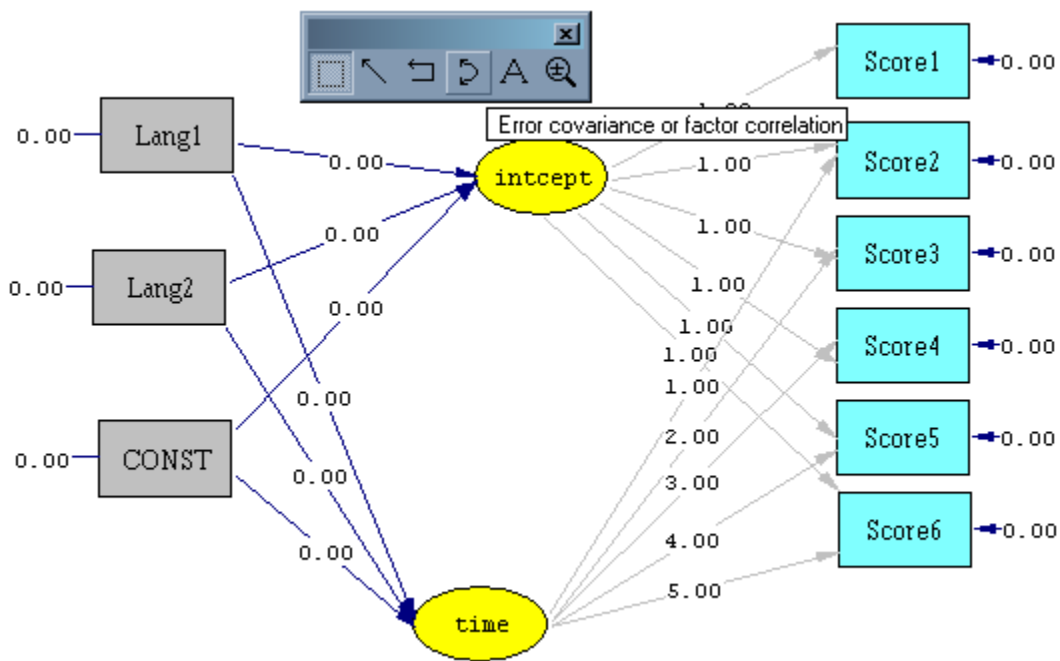
To draw this path, left-click on the horizontal link between 0.00 and the Lang1 rectangle. Drag the path to the line connecting 0.00 and Lang2 before releasing the mouse button.
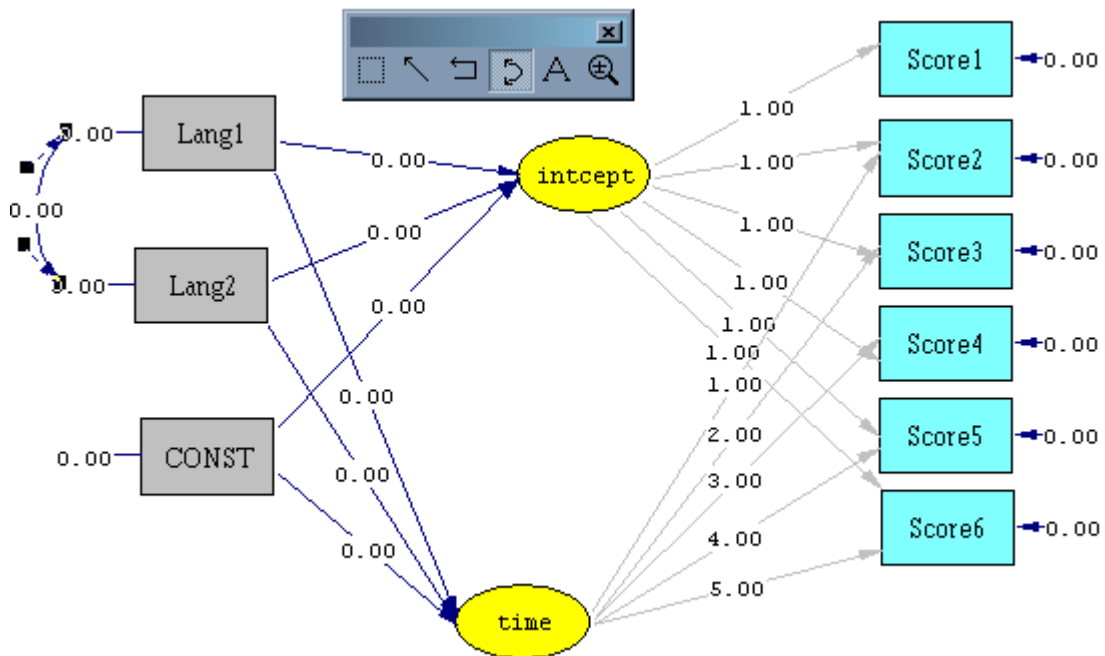
To build the corresponding SIMPLIS syntax, select **Setup**, **Build SIMPLIS Syntax** from the main menu.
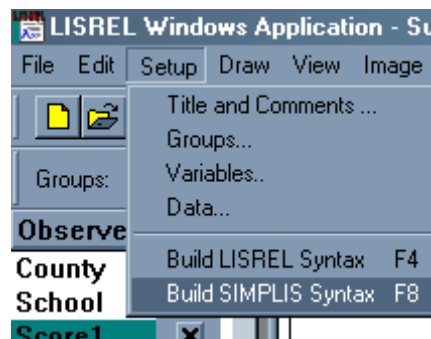
The syntax shown below is generated.


```
RAW DATA from file 'C:\Program Files\lisrel870\MISSINGEX\SurveySem.PSF'
Latent Variables  intcept time
Relationships
Score1 = 1.00*intcept
Score2 = 1.00*intcept 1.00*time
Score3 = 1.00*intcept 2.00*time
Score4 = 1.00*intcept 3.00*time
Score5 = 1.00*intcept 4.00*time
Score6 = 1.00*intcept 5.00*time
intcept = Lang1 Lang2 CONST
time = Lang1 Lang2 CONST
Path Diagram
End of Problem
```

In our model, we assume that the coefficient of the latent variable time is not influenced by Lang1 or Lang2 and we change the syntax by adding 0.00* in front of Lang1 and Lang2 as shown below.


```
Linear Growth Curve Simulated Data
RAW DATA from file 'C:\Program Files\lisrel870\MISSINGEX\SurveySem.PSF'
Latent Variables  intcept time
Relationships
Score1 = 1.00*intcept
Score2 = 1.00*intcept 1.00*time
Score3 = 1.00*intcept 2.00*time
Score4 = 1.00*intcept 3.00*time
Score5 = 1.00*intcept 4.00*time
Score6 = 1.00*intcept 5.00*time
intcept = Lang1 Lang2 CONST
time = 0.00*Lang1 0.00*Lang2 CONST
! Commands below added |
Set the error Covariance of intcept and time free
Equal Error Variances: Score1-Score6
Path Diagram
```

Since it is assumed that the error variances associated with the dependent variables are homogeneous, we manually add the command

> Equal Error Variances: Score1 – Score6

In addition, it is assumed that intcept and time are correlated. The equivalent SIMPLIS command is also typed in, as shown in the syntax file.

Once these modifications to the command file are completed, click the **Run LISREL** icon to obtain a path diagram.

Discussion of results – implementation of sampling weights in a linear growth curve model

The path diagram shows the parameter estimates and χ^2 goodness of fit statistic (χ^2 = 72.42, df = 31) under the assumption that stratification and selection of clusters do not effect standard errors or model fit.

Chi-Square=72.42, df=31, P-value=0.00004, RMSEA=0.016
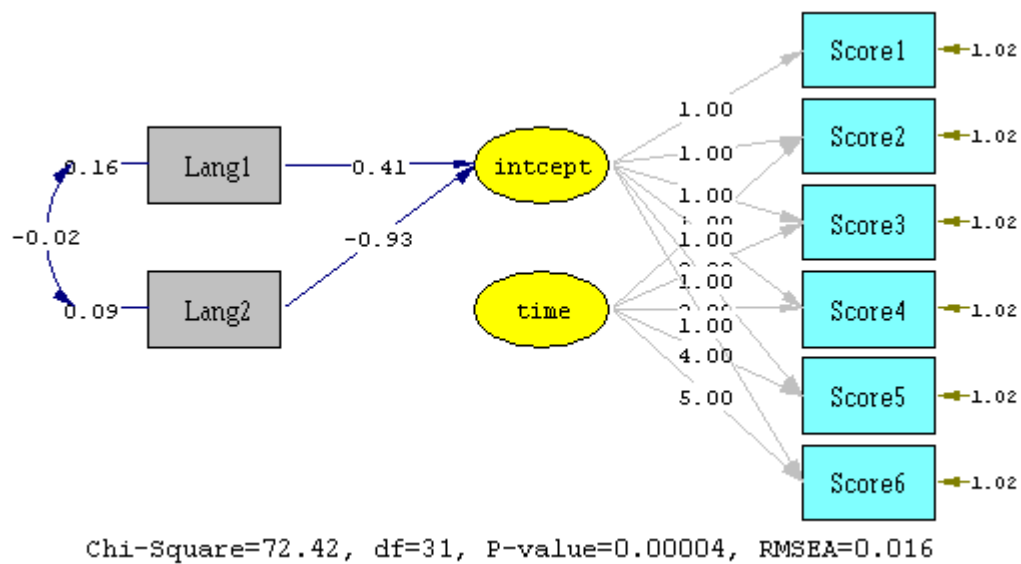
In the display below, the structural part of the model is shown. The numeric values are the *t*-values (*t*: parameter estimate/std. error) for each path. For example, the *t*-value for $\hat{\gamma}_1 = 0.41$ equals 2.28.

| Structural Model | Estimates: T-values |

survey1.PTH
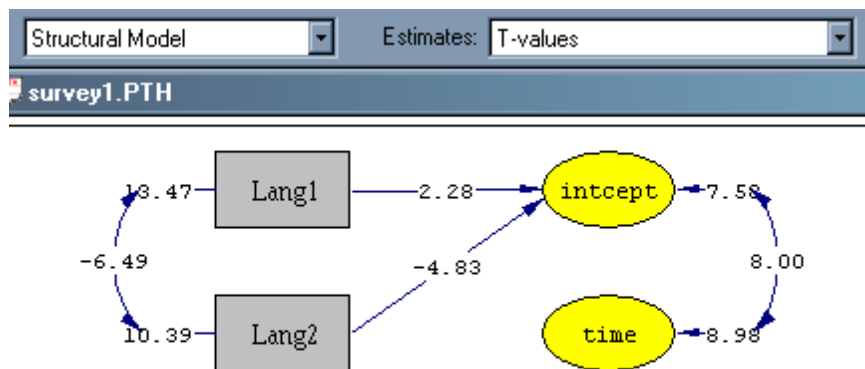
Example: Incorporating stratum and cluster variables in the model

Next, select the **Survey Design** option from the **Data** menu and add County and School respectively as stratification and cluster variables. Click **OK** and then select the **File**, **Save** option.

Once this is done, make **survey1.spl** the active window before clicking the **Run Lisrel** icon.

Discussion of results – Incorporating stratum and cluster variables in the model

From the path diagram below we note that the χ^2 statistic has decreased from 72.42 to 31.34 and that the t-value corresponding to $\hat{\gamma}_1$ decreased from 2.28 to 1.71. In the latter case, the t-value indicates a non-significant coefficient for the Lang1, intcept path.
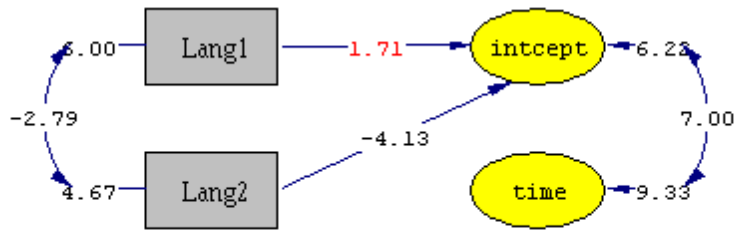
Chi-Square=31.34, df=31, P-value=0.44911, RMSEA=0.001

Finally, selected parts of the LISREL output are shown below.


```
survey1.OUT                                    _ □ ×

   Linear Growth Curve Simulated Data

   LISREL Estimates (Maximum Likelihood)

           GAMMA
                    Lang1          Lang2
                   --------       --------
    intcept          0.414         -0.934
                    (0.777)        (0.728)
                     0.532         -1.282

           PHI
                    Lang1          Lang2
                   --------       --------
    Lang1            0.160
                    (0.086)
                     1.864
    Lang2           -0.020          0.090
                    (0.023)        (0.062)
                    -0.868          1.451
```

The 95% confidence intervals for the estimated parameters show that the population parameter values are all contained in the respective intervals. For example, a 95% confidence interval for α_1 gives $0.491 \pm 2(0.05) = (0.391; 0.591)$ with corresponding population value of 0.5.

```
survey1.OUT                                                    _ □ ×

          PSI
                intcept          time
               --------        --------
   intcept        2.038
                 (1.055)
                  1.931
     time         0.666           0.382
                 (0.307)         (0.132)
                  2.174           2.899

          THETA-EPS
   Score1        Score2        Score3        Score4        Score5        Score6
  --------      --------      --------      --------      --------      --------
   1.023         1.023         1.023         1.023         1.023         1.023
  (0.153)       (0.153)       (0.153)       (0.153)       (0.153)       (0.153)
   6.677         6.677         6.677         6.677         6.677         6.677

          ALPHA
                intcept          time
               --------        --------
                  0.986           0.491
                 (0.536)         (0.159)
                  1.842           3.088
```

The χ^2 goodness-of-fit statistic equals 31.34, with a corresponding *p*-value of 0.45, indicating that
the model fits the data quite well.

```
survey1.OUT                                                    _ □ ×

          Global Goodness of Fit Statistics, Missing Data Case

       Chi-Square Scale Factor =         0.29463E-01
       Number of Strata     =                 48
       Number of Clusters =                  144


          -2ln(L) for the saturated model =        92753.788
          -2ln(L) for the fitted model     =        93817.557

                    Degrees of Freedom = 31
        Full Information ML Chi-Square   = 31.341 (P = 0.449)
     Root Mean Square Error of Approximation (RMSEA) = 0.00149
     90 Percent Confidence Interval for RMSEA = (0.0 ; 0.0107)
          P-Value for Test of Close Fit (RMSEA < 0.05) = 1.000
```

5.5 Evaluation

5.5.1   Simulation study based on a linear growth curve model

Introduction

In Section 4.5.2, a simulation study based on a linear growth model for continuous outcomes was discussed (see Asparouhov, 2004). This study is repeated by fitting a structural equation model to the data which is mathematically equivalent to a multilevel model.

The data

An unbalanced design, consisting of 500 univariate observations that are clustered within 100 level-two units, was used. Half of the level two units have four observations and the other half have six observations. The times of the observations are equally spaced starting at 0 and ending with 3 for the clusters with 4 observations and ending with 5 for the clusters with 6 observations. The linear growth model has random intercept and slope coefficients. For a complete description of the data, see Section 4.5.2.

The model

A path diagram of the model is shown below.
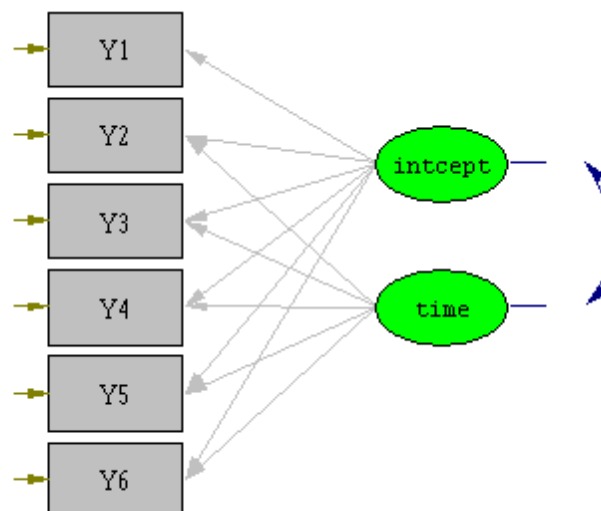
The LISREL model is

$$\mathbf{x} = \mathbf{\Lambda}_x \mathbf{\xi} + \mathbf{\delta}$$

where
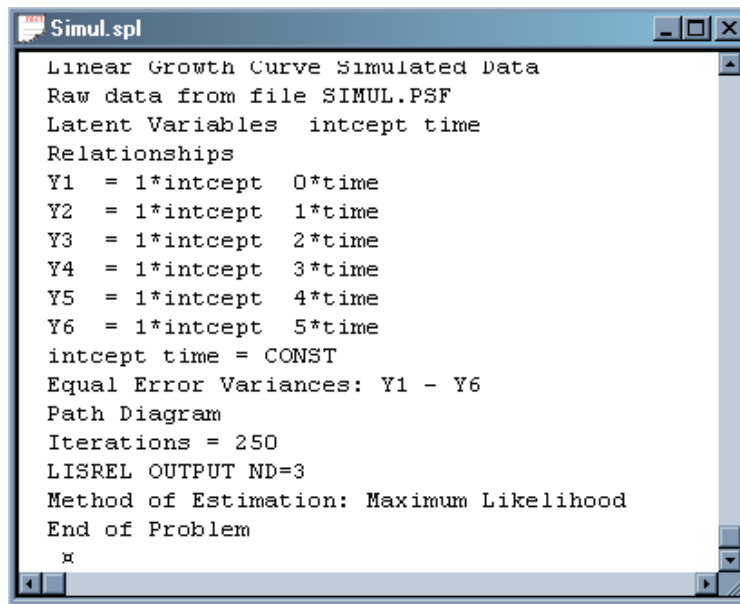
$$\mathbf{\Lambda}_x = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix},$$

$$E(\mathbf{\xi}) = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix}, \quad Cov(\mathbf{\xi}) = \begin{bmatrix} \phi_{11} & \phi_{21} \\ \phi_{21} & \phi_{22} \end{bmatrix},$$

and $Cov(\mathbf{\delta}) = \sigma^2 \mathbf{I}$.

Setting up the analysis

The SIMPLIS syntax for fitting this growth model is as follows.

```
Simul.spl                                          _ □ ×
Linear Growth Curve Simulated Data
Raw data from file SIMUL.PSF
Latent Variables  intcept time
Relationships
Y1   = 1*intcept   0*time
Y2   = 1*intcept   1*time
Y3   = 1*intcept   2*time
Y4   = 1*intcept   3*time
Y5   = 1*intcept   4*time
Y6   = 1*intcept   5*time
intcept time = CONST
Equal Error Variances: Y1 - Y6
Path Diagram
Iterations = 250
LISREL OUTPUT ND=3
Method of Estimation: Maximum Likelihood
End of Problem
  ¤
```

The command intcept time = CONST indicates the estimation of the population intercept and slope coefficients. The command Equal Error Variances: Y1 – Y6 specifies a homogeneous error variance term on level 1 of the model.

Discussion of results

Results of the simulation study are summarized in Tables 5.1 and 5.2.

Table 5.1: Bias and Coverage in LISREL and Mplus

Parameter	True Value	Bias		Coverage	
		LISREL	Mplus	LISREL	Mplus
β_0	0.5	0.017	0.017	0.908	0.908
β_1	0.1	0.002	0.002	0.938	0.942
ϕ_{11}	1	-0.024	-0.024	0.845	0.848
ϕ_{22}	0.2	-0.006	-0.006	0.892	0.902
ϕ_{21}	0.3	-0.006	-0.006	0.936	0.940
σ^2	1	-0.008	-0.008	0.908	0.910

The bias and coverage produced by LISREL and Mplus are virtually identical.

As part of the simulation study, the unadjusted and adjusted χ^2 goodness-of-fit statistics for each of the 500 simulations were computed. The adjusted χ^2 was obtained by the multiplication of the unadjusted χ^2 with the χ^2 scale factor, as described in Section 5.6. The mean values and rejection rates of these statistics for the 500 samples are given in Table 5.2. The rejection rate (expressed as a percentage) denotes the number of times the χ^2 statistic exceeded $\chi^2_{21;0.05}$. The degrees of freedom, 21, is obtained as the number of non-duplicated elements of the covariance matrix plus the number of means minus the number of parameters estimated. Therefore $df = 21 + 6 - 6$.

Table 5.2: Mean values and rejection rates of χ^2 goodness-of-fit statistics

Unadjusted χ^2	Adjusted χ^2
32.98	22.49
32 %	9 %

The expected mean for a χ^2_{21} random variate is 21. The mean for the adjusted χ^2, which is higher than the expected mean, explains the rejection rate being higher than 5 %. This result implies that more research on the correction factor of the χ^2 test statistic under complex sampling is indicated.

5.5.2   Latent curve analysis with main and interaction effects

Introduction

Curran, Bauer & Willoughby (2004) considered the testing of main effects and interactions in latent curve analysis. Their goal was to illustrate that classic techniques, as applied in multiple regression, can be generalized to the case of latent curve analyses. As part of the paper, an example was used to illustrate the testing of a categorical by continuous interaction in an unbalanced latent curve model with missing data over time. In this section, the same model is fitted, with and without sampling weights, in order to evaluate the impact of ignoring weights on an analysis: an analysis option not available to the authors of the paper at the time of publication.

The data

The example in the paper was based on data from the National Longitudinal Study of Youth. Specific details on the selection of the sample can be found in Curran (1997). The sample consists of information on 405 children at 4 occasions. At the start of the study, children in the sample were between 6 and 8 years old. Information is not available for all children on all occasions: while 405 were interviewed initially, the second, third, and fourth occasions provided information on 374, 297, and 294 children respectively. Only 221 children were interviewed on all four occasions. As such, the participant attrition over time, combined with the variability in age at the start of the study and the fact that measurement occasions were approximately two years apart, makes this an example of an unbalanced design with missing data.

In this section, the same data are used. Two analyses will be performed: a multilevel and a SEM analysis, the latter to verify the validity of the comparison of our results with that of Curran, Bauer & Willoughby (2004). In addition, models will be fitted with and without sampling weights.

The data were used in different formats for the structural equation and multilevel models. A short description of each data format is given below.

Structural equation modeling

A few of the variables in the data set **curran_NLSY_subset.psf** are shown below for the first 10 observations.

The emotional support at home and the level of antisocial behavior exhibited by these children were of special interest. The authors focused on three questions of interest: the form of the mean developmental trajectory of antisocial behavior over time, the possibility of meaningful individual variability in trajectories around these mean values, and the possible effect of interaction between the gender of a child and the level of emotional support on antisocial behavior.

The following variables included in the PSF were selected from the survey data:

- o MOM_ID: This variable represents the identification number of the mother and serves as grouping variable for all measurements for a specific child. There are 405 mothers included in this subset of the NLSY data.
- o MOM_WT: The sampling weight for each mother.
- o antiy1 – antiy10: A measure of antisocial behavior in the child. For each of these variables, a continuous measure representing the sum of six items assessing child antisocial behavior over the previous 3 months was created with values ranging between 0 and 12, where a high value would indicate a higher level of antisocial behavior.
- o genfemo: The gender of the child, coded "0" for a female, and "1" for a male.
- o home_emo: A measure of emotional support of the child in the home. This continuous measure, ranged from 0 to 13 with higher values reflecting higher levels of support, was measured at the first measurement occasion. It is centered around the mean level of emotional support.
- o home_cog: A measure of cognitive stimulation, computed as a summation of 14 dichotomously scored items reported by the mother, which ranges between 0 and 14.
- o genxemo: A variable intended to represent the interaction between a child's gender and level of emotional support: genxemo = genfemo$\times$ home_emo.

Multilevel modeling

For the multilevel analysis, the PSF **curran_mlev.psf** was used as basis of the analysis. Data on all the variables used in this model are shown below for the first 10 respondents. Note that, in contrast

to the PSF used for the SEM model, antisocial behavior is now represented by a single variable containing the stacked measurements over the 4 measurement occasions. For the first child, for example, the values 2, 1, 0, and 2 respectively were observed at the measurement occasions, where the latter is indicated by the variable tim.

curran_mlev.PSF						
	Mom_ID	**Mom_Wt**	**Antiy**	**tim**	**genfemo**	**home_emo**	**genxemo**
1	288.000	493088.000	2.000	1.000	1.000	0.795	0.795
2	288.000	493088.000	1.000	3.000	1.000	0.795	0.795
3	288.000	493088.000	0.000	5.000	1.000	0.795	0.795
4	288.000	493088.000	2.000	7.000	1.000	0.795	0.795
5	297.000	598291.000	2.000	0.000	1.000	-0.205	-0.205
6	297.000	598291.000	7.000	3.000	1.000	-0.205	-0.205
7	343.000	159958.000	1.000	1.000	1.000	1.795	1.795
8	343.000	159958.000	0.000	3.000	1.000	1.795	1.795
9	343.000	159958.000	2.000	5.000	1.000	1.795	1.795
10	343.000	159958.000	0.000	7.000	1.000	1.795	1.795

The following variables were used in the multilevel analysis:

- Mom_ID: This variable represents the identification number of the mother and serves as grouping variable for all measurements for a specific child. There are 405 mothers included in this subset of the NLSY data.
- Mom_Wt: The sampling weight for each mother.
- antiy: A measure of antisocial behavior in the child at a given measurement occasion. This continuous measure, representing the sum of six items assessing child antisocial behavior over the previous 3 months, was created with values ranging between 0 and 12, where a high value would indicate a higher level of antisocial behavior.
- tim: This variable indicates the time of measurement, and varies from 0 to 9.
- genfemo: The gender of the child, coded "0" for a female, and "1" for a male.
- home_emo: A measure of emotional support of the child in the home. This continuous measure ranged from 0 to 13, with higher values reflecting higher levels of support, was measured at the first measurement occasion. It is centered around the mean level of emotional support.
- genxemo: A variable intended to represent the interaction between a child's gender and level of emotional support: genxemo = genfemo× home_emo.

The model

Curran, Bauer & Willoughby (2004) shows how a structural equation model-based latent curve analysis and a hierarchical linear model for these data can be formulated to produce equivalent results. They point out, however, that there are subtle but important differences in both model estimation and interpretation due to the way in which time is incorporated into the model. These

differences are of particular importance in the case of conditional growth models, where one or more exogenous variables predict the random growth curve parameters. Main effects of the random trajectories imply that exogenous variables interact with time in the prediction of repeated measures for both cases. While both predictors and time are used as exogenous variables in the hierarchical linear model, the interaction between time and any predictor is explicitly modeled as a cross-level interaction. The latent curve analysis does not use time as a variable as such. Instead, it is incorporated into the model via the factor loading matrix. In this section, the two models and data sets constructed for use in the analyses will further illustrate this difference. To accommodate the differences in models fitted and data sets used, the structural equation model and the multilevel, or hierarchical linear, model, will be discussed separately in the rest of this section.

Structural equation model

We first consider the structural equation model. The model shown below corresponds to Figure 2 in Curran, Bauer & Willoughby (2004), and represents cohort-sequential conditional linear latent curve model with 10 time points, regressed on the main effect of gender, the main effect of emotion, and the interaction between gender and emotion. The variables intcept and slope represent the latent intercept and latent slope of the trajectory respectively.
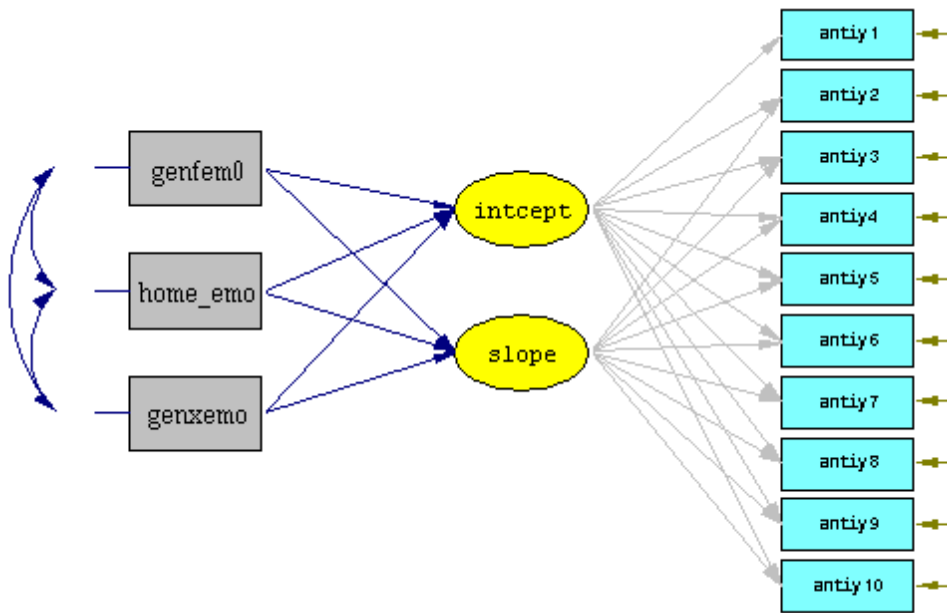
In the Y part of the model, we include the dependent variables antiy1 to antiy10. It is assumed that antiy1 to antiy10 are indicators of the endogenous (ETA) latent variables intcept and slope.

The covariates genfemo, home_emo, and genxemo are also assumed to have relationships with both the intercept and the slope of the trajectory and form the X part of the model. Finally, we allow the intcept and slope variables to be correlated. This path cannot be seen on the basic path diagram,

which is the type of diagram used here (to view this path, select the **Structural Model** option from the **Models:** drop-down list in the PTH window).

Mathematical model

The LISREL model consists of a measurement and structural part.

Measurement model

The measurement model may be expressed as

$$
\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{bmatrix} \begin{bmatrix} \mathbf{\eta} \\ \mathbf{\xi} \end{bmatrix} + \begin{bmatrix} \mathbf{\varepsilon} \\ \mathbf{\delta} \end{bmatrix}
$$

where $\mathbf{y} = \begin{bmatrix} antiy1 & antiy2 & \ldots & antiy10 \end{bmatrix}'$, $\mathbf{x} = \begin{bmatrix} genfemo & hom\_emo & genxemo \end{bmatrix}'$ $\mathbf{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \ldots & \varepsilon_{10} \end{bmatrix}'$, $\delta = \begin{bmatrix} \delta_1 & \delta_2 & \delta_3 \end{bmatrix}'$, $\mathbf{\xi} = \begin{bmatrix} genfemo & hom\_emo & genxemo \end{bmatrix}'$,

$$
\mathbf{\eta} = \begin{bmatrix} \text{intcept} \\ \text{slope} \end{bmatrix}
$$

and

$$
\mathbf{\Lambda}'_y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{bmatrix}.
$$

$Cov(\mathbf{\varepsilon})$ is a diagonal matrix with diagonal elements $var(\varepsilon_1)$, $var(\varepsilon_2)$, ..., $var(\varepsilon_{10})$ where we constrained these elements to be equal, while $\mathbf{\Lambda}_x$ is a 3×3 identity matrix and

$$
Cov(\mathbf{\xi}) = \mathbf{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} \\ \phi_{21} & \phi_{22} & \phi_{32} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}.
$$

Structural equation model

The structural equation model for the latent variables intcept and slope is given by ($\mathbf{B} = \mathbf{0}$)

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

,

where $\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 & \zeta_2 \end{bmatrix}'$, with

$$Cov(\boldsymbol{\zeta}) = \begin{bmatrix} \psi_{11} & \psi_{21} \\ \psi_{21} & \psi_{22} \end{bmatrix},$$

$$E(\boldsymbol{\zeta}) = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix},$$

and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_4 & \gamma_5 & \gamma_6 \end{bmatrix}.$$

The unknown model parameters are therefore γ_1, γ_2, γ_3, γ_4, γ_5, γ_6, α_1, α_2, ϕ_{11}, ϕ_{12}, ..., ϕ_{33}, ψ_{11}, ψ_{12}, ψ_{22}, and $var(\varepsilon_1)$.

Multilevel model

A general two-level model for a response variable y depending on a set of r predictors $x_1, x_2, \ldots, x_r$ can be written in the form

$$y_{ij} = \mathbf{x}'_{(f)ij}\boldsymbol{\beta} + \mathbf{x}'_{(2)i}\mathbf{u}_i + \mathbf{x}'_{(1)ij}\mathbf{e}_{ij}$$

where $i = 1, 2, \ldots, N$ denotes the level-2 units, and $j = 1, 2, \ldots, n_i$ the level-1 units. Thus y_{ij} represents the response of individual j, nested within level-2 unit i. The model shown here consists of a fixed and a random part. The fixed part of the model is represented by the vector product $\mathbf{x}'_{(f)ij}\boldsymbol{\beta}$, where $\mathbf{x}'_{(f)ij}$ is a typical row of the design matrix of the fixed part of the model with, as elements, a subset of the r predictors. The vector $\boldsymbol{\beta}$ contains the fixed, but unknown parameters to be estimated. The vector products $\mathbf{x}'_{(2)i}\mathbf{u}_i$ and $\mathbf{x}'_{(1)ij}\mathbf{e}_{ij}$ denote the random part of the model at levels 2 and 1 respectively. For example, $\mathbf{x}'_{(2)i}$ represents a typical row of the design

matrix of the random part at level 2, and $\mathbf{u}_i$ the vector of random level-2 coefficients to be estimated. The product $\mathbf{x}'_{(1)ij}\mathbf{e}_{ij}$ serves the same purpose at level 1. It is assumed that $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_N$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Phi}_{(2)}$, and $\mathbf{e}_{i1},\mathbf{e}_{i2},\ldots,\mathbf{e}_{in_i}$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Phi}_{(1)}$.

Within this hierarchical framework, the model fitted to the data uses the participant's gender, level of emotional support at home, and the interaction between these variables to predict the variability in intercept and slope over time of antisocial behavior trajectories.

$$antiy_{ij} = \beta_0 + \beta_1 * genfemo_{ij} + \beta_2 * home\_emo_{ij} + \beta_3 * genxemo_{ij} +$$
$$\beta_4 * tim_{ij} + \beta_5 * (genfemo_{ij})(tim_{ij}) + \beta_6 * (home\_emo_{ij})(tim_{ij}) +$$
$$\beta_7 * (genxemo_{ij})(tim_{ij}) + u_{i0} + +u_{i1} * tim_{ij} + e_{ij}$$

where β_0 denotes the average expected level of antisocial behavior for a female child at the first measurement occasion with a score of 0 on the measure of emotional support at home. The coefficients $\beta_1, \beta_2, \ldots, \beta_7$ are the estimated coefficients associated with the fixed part of the model which contains the predictor variables genfemo, home_emo, and the interaction term genxemo. The random part of the model is represented by, u_{i0}, $u_{i0}*tim_{ij}$, and e_{ij}, which denote the variation in average level of antisocial behavior between children, in slope over measurements occasions, and between measurements taken at different occasions, where the occasions form the lowest level of the hierarchy.

Setting up the analysis

Structural equation model

The SIMPLIS syntax for the model is shown below. Note that 5.0*slope, for example, indicates that the coefficient for the slope→ antiy6 path is fixed at a value of 5.0.

```
curran1.spl                                              _ □ ×

Linear Growth Curve NLSY Data Curran et al (2004)      ▲
RAW DATA from file 'Curran_NLSY_subset.PSF'
Latent Variables  intcept slope
Relationships
antiy1 = 1.0*intcept
antiy2 = 1.0*intcept 1.0*slope
antiy3 = 1.0*intcept 2.0*slope
antiy4 = 1.0*intcept 3.0*slope
antiy5 = 1.0*intcept 4.0*slope
antiy6 = 1.0*intcept 5.0*slope
antiy7 = 1.0*intcept 6.0*slope
antiy8 = 1.0*intcept 7.0*slope
antiy9 = 1.0*intcept 8.0*slope
antiy10= 1.0*intcept 9.0*slope
intcept = CONST genfem0  home_emo  genxemo
slope =   CONST genfem0  home_emo  genxemo
Set the error Covariance of intcept and slope free
Equal Error Variances: antiy1-antiy10
LISREL OUTPUT: ND=3 IT=0 PV=curran1.par SV=curran1.std
Path Diagram
End of Problem

   ¤                                                    ▼
◄                                                     ►
```

An experienced LISREL user may prefer to type the SIMPLIS commands to fit a specific model. Alternatively, the syntax can be created by drawing a path diagram. This is done in the same way as shown in Section 5.4.2. To add a weight variable, as is the case in the second of the structural equation models considered here, select the **Data, Survey Design** option from the main menu bar

```
LISREL Windows Application - curran_NLSY_subset.PSF

File  Edit  Data  Transformation  Statistics  Graphs  Multilevel  SurveyG

        Define Variables ...
        Select Variables/Cases...
        Sort Case...

        Insert Variable
        Insert Cases                    e_emo     home_cog
curra   Delete Variable
        Delete Case
    1                                    0.795      -1.894
    2                                   -0.205      -1.894
    3   Weight Cases...                  1.795       1.106
    4   Survey Design...                -1.205      -3.894
    5              1.000                -2.205      -1.894
    6              0.000                -3.205       0.106
```
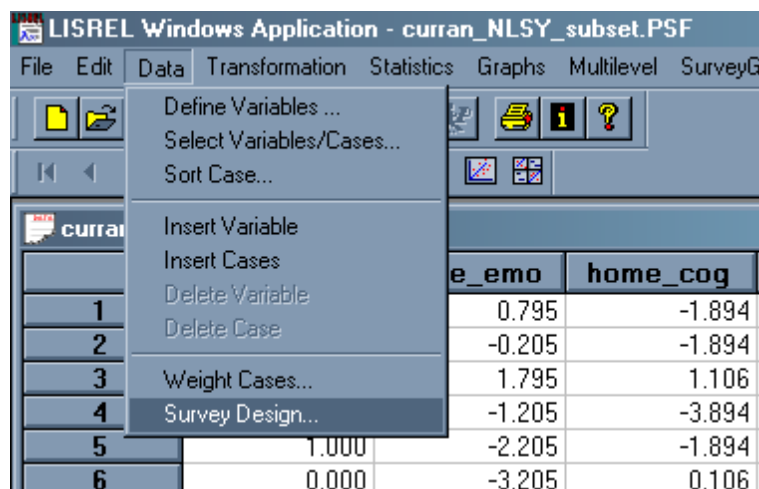
and enter the variable MOM_WT in the **Design weight** field as shown below.
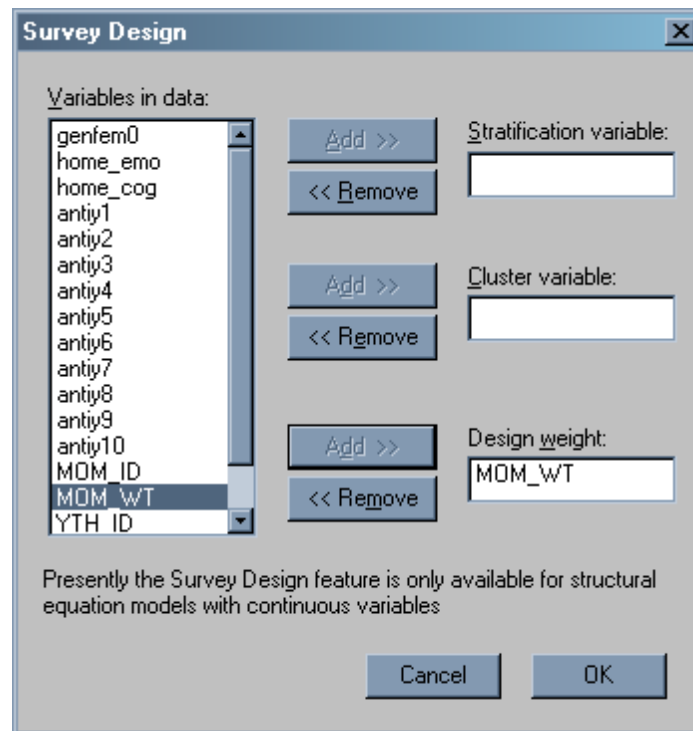
To run the model, click the **Run LISREL** icon button on the main menu bar.

Multilevel model

Specifying the multilevel model is straightforward, and proceeds as shown in Chapter 4. Briefly, the level-2 ID is identified as Mom_ID, the outcome is antiy, and the fixed part of the model consists of the variables genfemo, home_emo, genxemo and tim as shown in the two dialog boxes below. The weight variable Mom_Wt used in the second of the multilevel analyses discussed here is entered on the **Identification Variables** dialog box.
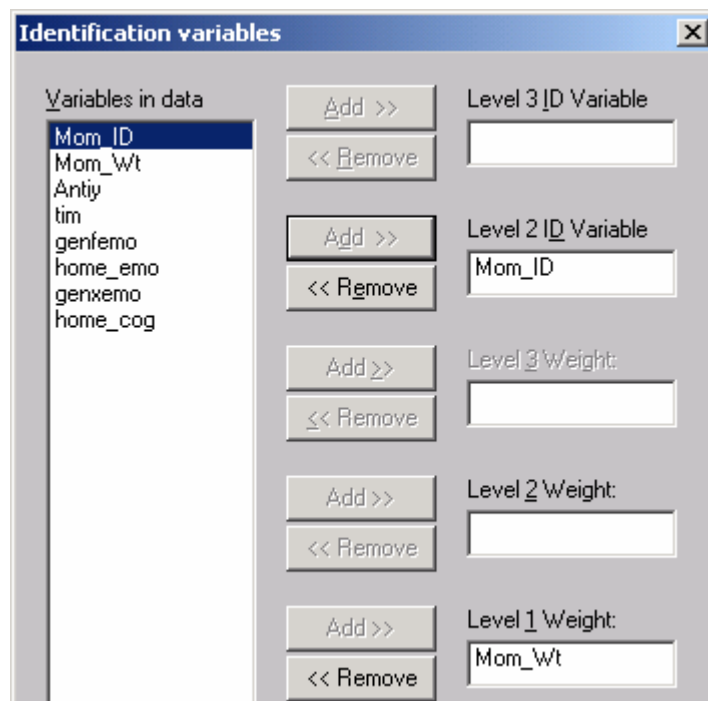
Note that, in the **Select Response and Fixed Variables** dialog box shown, the required interaction between tim and the three variables genfemo, home_emo, genxemo is not included – this will be added manually to the syntax file created via the dialog boxes.
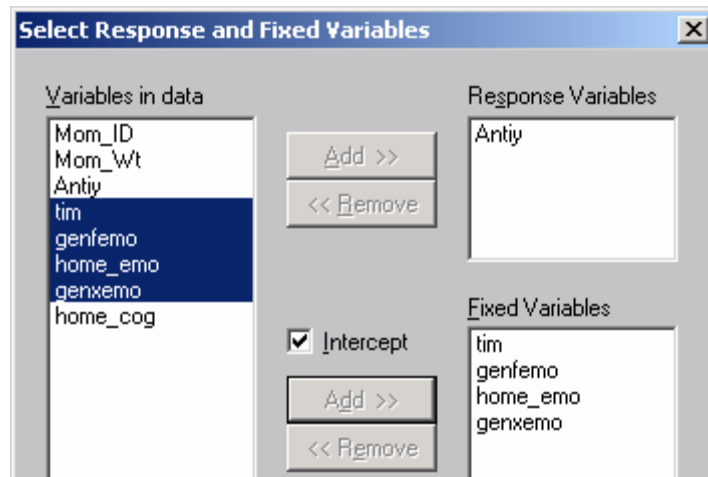
To estimate both a random intercept and a random slope, add the variable tim to the **Random Level 2** field on the **Random Variables** dialog box as shown below.
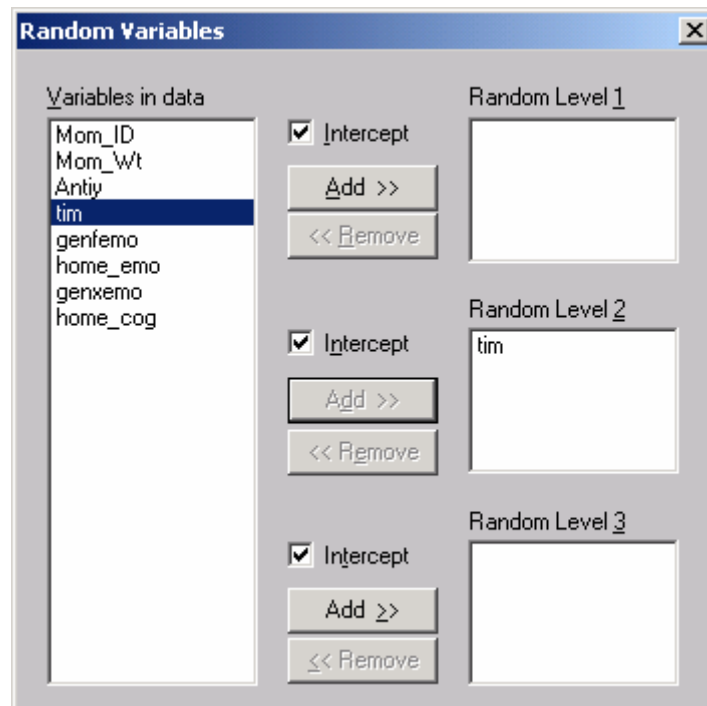
The syntax generated via the **Finish** button on the **Random Variables** dialog box is shown below:

Finally, type the additional interaction terms tim*genfemo tim*home_emo tim*genxemo into the syntax file to obtain the final syntax as shown below. The analysis is started by clicking the **Run Prelis** icon button on the main menu bar.

```
curran_mlev1.PR2                                                    _ |□| X|

    OPTIONS OLS=YES CONVERGE=0.000100 MAXITER=100 OUTPUT=STANDARD ;
    TITLE=Curran 2004 Paper plus Weights;
    SY='C:\Program Files\lisrel870\MISSINGEX\curran_mlev.PSF';
    ID2=Mom_ID;
    RESPONSE=Antiy;
    FIXED=intcept tim genfemo home_emo genxemo tim*genfemo tim*home_emo tim*genxemo;
    RANDOM1=intcept;
    RANDOM2=intcept tim;
    ¤
```

Discussion of results

The results of both the SEM and multilevel models are summarized in Table 5.3. The table also contains the results from Curran, Bauer & Willoughby (2004). These analyses did not take the sampling weight Mom_Wt into account. The results of the weighted analysis, in which this variable was incorporated into the estimation procedure, are reported in Table 5.4.

Table 5.3: Unweighted analyses: comparison of results

Coefficient	Estimates			Standard errors		
	CBW paper	Multilevel	SEM	CBW paper	Multilevel	SEM
genfemo (γ_1)	0.829	0.829	0.829	0.161	0.161	0.161
home_emo (γ_2)	-0.194	-0.194	-0.194	0.048	0.048	0.048
genxemo (γ_3)	0.044	0.044	0.044	0.070	0.070	0.070
INTCPT (α_0)	1.217	1.217	1.217	0.114	0.115	0.114
tim$\times$ genfemo (γ_4)	0.013	0.013	0.013	0.035	0.035	0.035
tim$\times$ home_emo (γ_5)	0.012	0.012	0.012	0.010	0.010	0.010
tim$\times$ genfemo $\times$ home_emo (γ_6)	-0.029	-0.029	-0.029	0.015	0.015	0.015
tim (α_1)	0.066	0.066	0.066	0.024	0.025	0.025
$Var(intcept)$ (ϕ_{11})	*	0.669	0.669	*	0.203	0.204
$Var(time\ slope)$ (ϕ_{22})	*	0.019	0.019	*	0.009	0.009
$Cov(intcept, tim)$ (ϕ_{21})	*	0.076	0.076	*	0.035	0.035
σ^2 (var(ε))	*	1.758	1.758	*	0.097	0.098

*   Not reported in the Curran *et. al.* paper

Table 5.4: Weighted analyses: comparison of results

Coefficient	Estimates		Standard errors	
	Multilevel	**SEM**	**Multilevel**	**SEM**
genfemo (γ_1)	0.901	0.900	0.201	0.205
home_emo (γ_2)	-0.244	-0.243	0.021	0.058
genxemo (γ_3)	0.122	0.122	0.112	0.115
INTCPT (α_0)	1.203	1.203	0.120	0.115
$tim \times genfemo$ (γ_4)	0.024	0.024	0.047	0.048
$tim \times home\_emo$ (γ_5)	0.014	0.014	0.015	0.015
$tim \times genfemo$ $\times home\_emo$ (γ_6)	-0.020	-0.020	0.026	0.038
tim (α_1)	0.073	0.073	0.030	0.030
$Var(intcept)$ (ϕ_{11})	0.487	0.415	0.317	0.330
$Var(time\ slope)$ (ϕ_{22})	0.230	0.022	0.014	0.015
$Cov(intcept,tim)$ (ϕ_{21})	0.091	0.103	0.060	0.063
σ^2 ($var(\varepsilon)$)	1.966	1.996	0.234	0.217

The goodness-of-fit of the models fitted can also be compared. For the weighted structural equations model, the following path diagram was obtained:
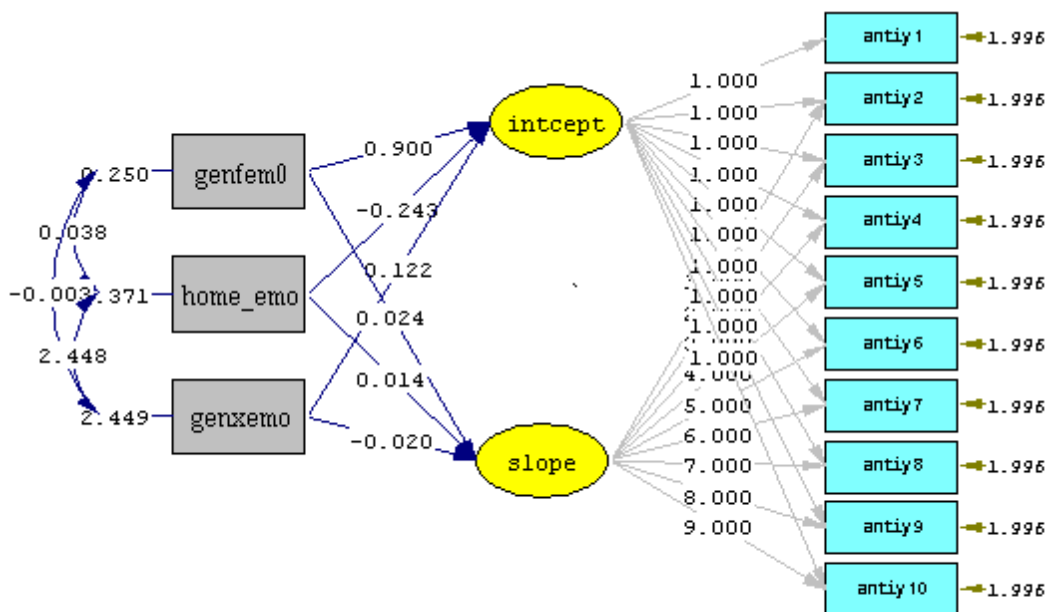
Chi-Square=99.48, df=83, P-value=0.10488, RMSEA=0.022

From the path diagram, $\chi^2 = 99.48$, with 83 degrees of freedom. The corresponding χ^2-statistic value for the unweighted model is 107.2978, with 83 degrees of freedom. For the analyses that included the weight variable, there are differences in parameter estimates if we compare the multilevel model results with those of the structural equation model. These differences are most evident in the covariance matrix of the latent variables. LISREL produced a warning message that this matrix is not positive definite. On further examination, it was found that the variable antiy10 contained only 8 non-missing observations. Refitting of the models, using the first 9 variables antiy1 − antiy9, showed that the multilevel and SEM results are identical in this case.

5.5.3 Replicate weights

Introduction

Survey data sets often include a column W_0 of design weights and additional columns $W_1, W_2, \ldots, W_{R-1}$ of replicate weights. Typically, a researcher may repeatedly fit the same model to the data by working through the sequence of weight variables $W_0, W_1, \ldots, W_{R-1}$. Means of the R sets of parameter estimates and their standard errors may subsequently be computed to obtain more accurate parameter and standard error estimates. In this section, we illustrate how to use LISREL in the case of replicate weights.

The data
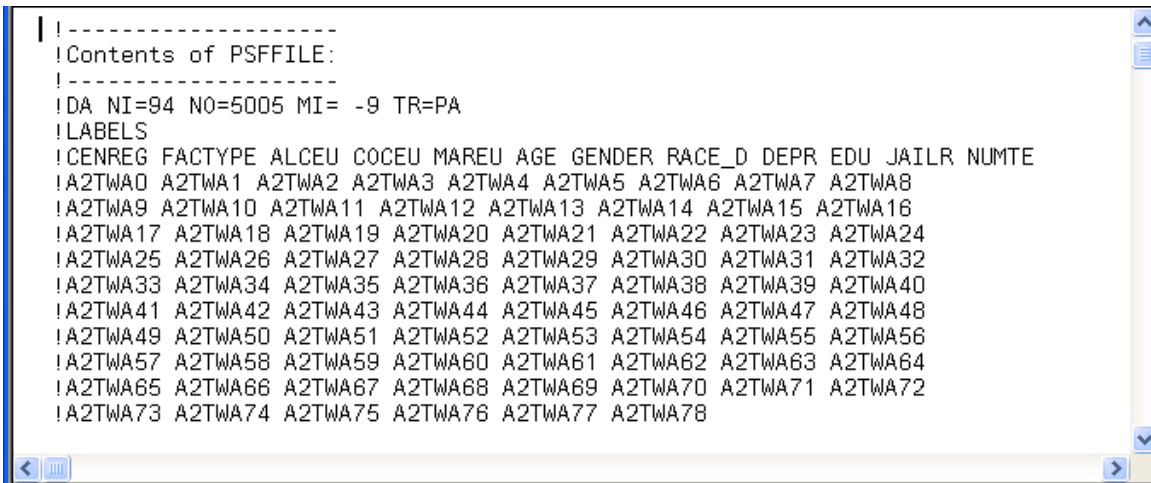
The data set used here is described in Section 3.4.1. A few of the variables are shown below for the first 10 observations in the data set.

	CENREG	FACTYPE	ALCEU	COCEU	MAREU	AGE	GENDER
1	3.0	2.0	1.0	-9.0	1.0	23.0	1.0
2	4.0	2.0	1.0	-9.0	1.0	26.0	2.0
3	2.0	3.0	1.0	1.0	1.0	34.0	1.0
4	3.0	4.0	1.0	-9.0	1.0	13.0	1.0
5	2.0	4.0	1.0	-9.0	-9.0	28.0	1.0
6	4.0	4.0	-9.0	-9.0	-9.0	28.0	1.0
7	4.0	4.0	1.0	-9.0	1.0	44.0	1.0
8	2.0	4.0	1.0	0.0	1.0	33.0	1.0
9	1.0	4.0	1.0	1.0	1.0	27.0	1.0
10	1.0	3.0	1.0	1.0	1.0	45.0	1.0

The contents of the PSF are obtained by selecting the **Statistics**, **Data Screening** option. A portion of the output is shown below.

The following variables included in the PSF were selected from the survey data:

- CENREG: This variable indicates the census region and has four categories, these being "Northeast", "Midwest", "South", and "West" respectively.
- FACTYPE: The facility treatment type has four categories, too, representing facilities with "residential treatment", "outpatient methadone treatment", "outpatient non-methadone treatment", and "more than one type of treatment" respectively.
- ALCEU: An indicator variable with value "1" if the respondent has ever used alcohol, and "0" otherwise.

```
!--------------------
!Contents of PSFFILE:
!--------------------
!DA NI=94 NO=5005 MI= -9 TR=PA
!LABELS
!CENREG FACTYPE ALCEU COCEU MAREU AGE GENDER RACE_D DEPR EDU JAILR NUMTE
!A2TWA0 A2TWA1 A2TWA2 A2TWA3 A2TWA4 A2TWA5 A2TWA6 A2TWA7 A2TWA8
!A2TWA9 A2TWA10 A2TWA11 A2TWA12 A2TWA13 A2TWA14 A2TWA15 A2TWA16
!A2TWA17 A2TWA18 A2TWA19 A2TWA20 A2TWA21 A2TWA22 A2TWA23 A2TWA24
!A2TWA25 A2TWA26 A2TWA27 A2TWA28 A2TWA29 A2TWA30 A2TWA31 A2TWA32
!A2TWA33 A2TWA34 A2TWA35 A2TWA36 A2TWA37 A2TWA38 A2TWA39 A2TWA40
!A2TWA41 A2TWA42 A2TWA43 A2TWA44 A2TWA45 A2TWA46 A2TWA47 A2TWA48
!A2TWA49 A2TWA50 A2TWA51 A2TWA52 A2TWA53 A2TWA54 A2TWA55 A2TWA56
!A2TWA57 A2TWA58 A2TWA59 A2TWA60 A2TWA61 A2TWA62 A2TWA63 A2TWA64
!A2TWA65 A2TWA66 A2TWA67 A2TWA68 A2TWA69 A2TWA70 A2TWA71 A2TWA72
!A2TWA73 A2TWA74 A2TWA75 A2TWA76 A2TWA77 A2TWA78
```

- COCEU: An indicator variable with value "1" if the respondent has ever used cocaine, and "0" otherwise.
- MAREU: An indicator variable with value "1" if the respondent has ever used marijuana, and "0" otherwise.
- AGE: This variable denotes age at admission to a facility.
- GENDER: The respondent's gender is denoted by this indicator variable that assumes a value of "1" for female respondents.
- RACE_D: The original variable RACE recoded so that '1" denotes white and "0" other ethnic groups.
- DEPR: This indicator variable is coded "1" if the respondent is depressed, and "0" otherwise.
- EDU: A categorical variable representing the respondent's level of education at admission. It has 5 categories, these being (from 1 to 5) "less than 8 years", "8 – 11 years or less than High School graduate", "High School graduate / GED", "some college", and "college graduate / postgraduate".
- JAILR: This indicator variable indicates whether the respondent had a prison or jail record prior to admission.

- o NUMTE: A count variable, indicating the total number of treatment episodes prior to admission.
- o A2TWA0 – A2TWA78: These variables are abstract final full sample weights. A more complete description follows below.

Variance estimation frequently relies on one of two techniques: Taylor series linearization or replication weights. Replicate weights are based on the same ideas as the jackknife, and has recently come into use in US government surveys, where replicate weights are provided instead of information on PSUs. In such cases, replicate weights may be used to disguise and/or prevent identification of individuals within PSUs to preserve privacy.

Handling of missing data

Missing values for the variables ALCEU, COCEU, …, NUMTE in the file **replicwts**.**psf** are coded -9.0. For these variables, 20.8 % of the possible values are not observed. Use of listwise deletion would result in retaining only 97 of the selected 5005 cases. Therefore, we use the full information maximum likelihood (FIML) procedure as described in Section 5.7.

To define -9.0 as the global missing value code, select the **Data**, **Define Variables** option from the main menu. Select the variable ALCEU (or any other variable in the list) and click the **Missing Values** button to activate the **Missing Values for** dialog box. Type -9.0 in the **Global missing value** text box and select **pairwise** as the deletion method. Click **OK** when done.
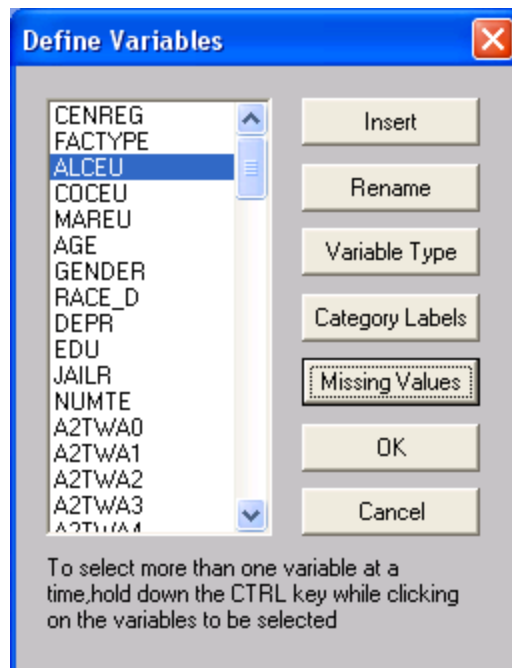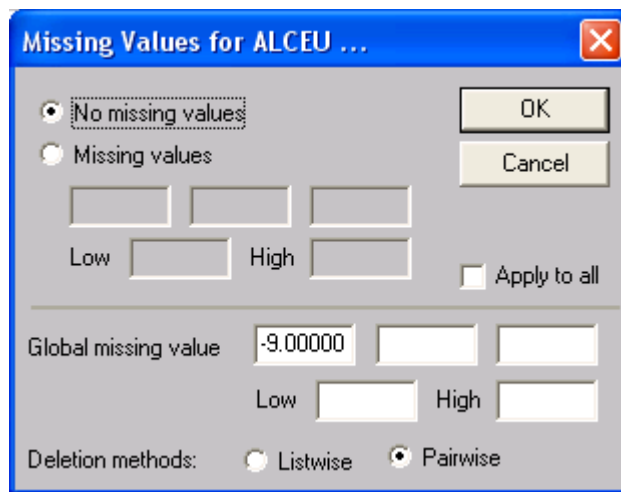
Handling of zero weights

Descriptive statistics of the weight variables A2TWA0 to A2TWA78 revealed that these variables contained zero as possible values. If a zero value is encountered in any row of the data set, it was replaced by the average value of all the non-zero weights in the corresponding row.
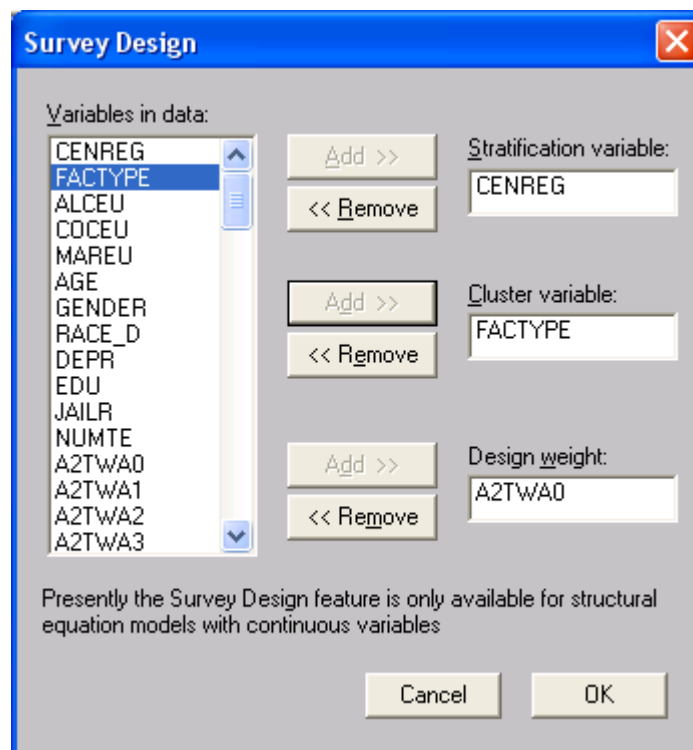
We found that multiple imputation failed in estimating weights if 0 was regarded as a missing value code. The reason for this is that all the weight variables are highly correlated and therefore the covariance matrix of the weight variables is essentially singular.

To define CENREG and FACTYPE as stratification and cluster variables, select the **Data**, **Survey Design** option from the main menu and add the variables as shown in the dialog box below. For our first analysis, select A2TWA0 as the **Design weight** variable.

The model

The path diagram representation of the model fitted to the data is shown below. It is assumed that each of the X-variables AGE, GENDER, RACE_D, DEPR, EDU and JAILR is a perfect indicator of the corresponding latent variable, these being age, gender, race_d, depr, educ and jailr. This implies that the error variances of the X-variables are zero, and that the path coefficients age → AGE, gender → GENDER, race_d → RACE_D, depr → DEPR, educ → EDU and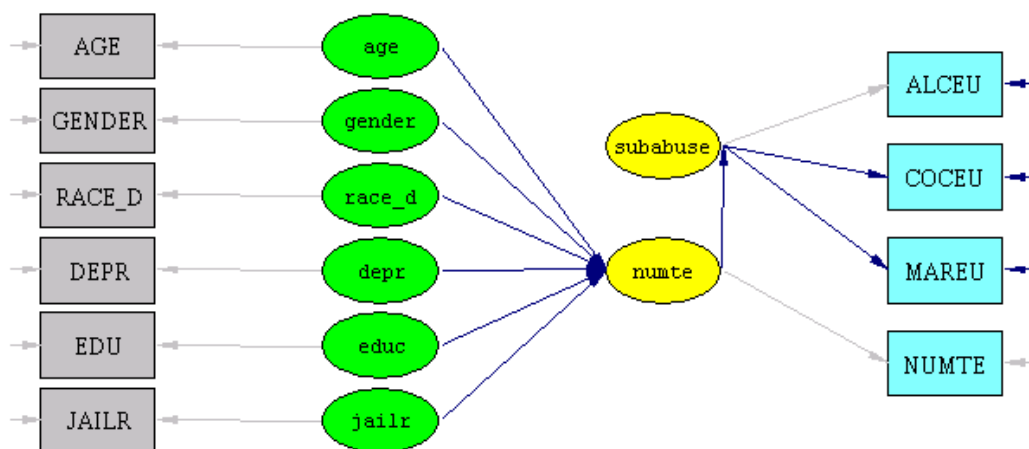 jailr → JAILR are all equal to one. In the LISREL terminology, age, gender, race_d, depr, edu and jailr are exogenous (KSI) latent variables. In principle, several indicators of depression and education, if available, could be incorporated into this model.

In the Y part of the model, we include the dependent variables ALCEU, COCEU, MAREU and NUMTE. It is assumed that ALCEU, COCEU, and MAREU are indicators of the endogenous latent (ETA) variable subabuse while NUMTE is a perfect indicator of the ETA variable numte. The path subabuse → ALCEU is set equal to 1 to fix the scale of the endogenous latent variable subabuse.

Finally, we assume that numte can be predicted by age, …, jailr, and in turn, subabuse is predicted by numte. In this context, the latent variable numte is a so-called mediating variable.

Mathematical model

The LISREL model consists of a measurement and structural part.

Measurement model

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_x \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{bmatrix}$$

where $\mathbf{x} = (AGE, GENDER, RACE\_D, DEPR, EDU, JAILR)'$, $\boldsymbol{\Lambda}_x$ is a 6×6 identity matrix and $\boldsymbol{\xi} = (age, gender, race\_d, depr, edu, jailr)'$. Also, $Cov(\boldsymbol{\delta}) = \mathbf{0}$ and $Cov(\boldsymbol{\xi}) = \boldsymbol{\Phi}$.

Furthermore,

$$\boldsymbol{\eta} = \begin{bmatrix} subabuse \\ numte \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} ALCEU \\ COCEU \\ MAREU \\ NUMTE \end{bmatrix},$$

and

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \end{bmatrix}.$$

Finally, $Cov(\boldsymbol{\varepsilon})$ is a diagonal matrix with diagonal elements $var(ALCEU)$, $var(COCEU)$, $var(MAREU)$ and 0.

Structural equation model

The structural model can be written as

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\varepsilon},$$

where

$$\mathbf{B} = \begin{bmatrix} 0 & \beta_{12} \\ 0 & 0 \end{bmatrix},$$

and

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \end{bmatrix},$$

that is,

$$subabuse = \beta_{12} \times numte + t_1$$
$$numte = \gamma_{21} \times age + \gamma_{22} \times gender + \gamma_{23} \times race\_d + \gamma_{24} \times depr + \gamma_{25} \times educ + \gamma_{26} \times jailr.$$

Also

$$Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$$
$$= \begin{bmatrix} \psi_{11} & \psi_{21} \\ \psi_{21} & \psi_{22} \end{bmatrix}.$$

The unknown model parameters are therefore $\lambda_{y,21}$, $\lambda_{y,31}$, β_{12}, γ_{21}, γ_{22}, γ_{23}, γ_{24}, γ_{25}, γ_{26}, ϕ_{11}, ϕ_{21}, ϕ_{22}, ..., ϕ_{66}, ψ_{11}, ψ_{21}, ψ_{22}, $var(\varepsilon_1)$, $var(\varepsilon_2)$, and $var(\varepsilon_3)$. In subsequent output, these 36 parameters will be denoted by the symbols LY21, LY31, BETA12, GAMMA21, GAMMA22, ..., GAMMA26, PHI11, PHI21, ..., PHI66, PSI11, PSI21, PSI22, TE11, TE22 and TE33.

Example: Setting up the analysis using SIMPLIS syntax

It is relatively easy to specify the model described above with SIMPLIS syntax. We start by indicating that the raw data is to read from the file **replicwts.psf**. Note that this is followed by a REWIND command which allows LISREL to repeatedly read the raw data from the same file.

The program commands AGE = 1.0*age, ..., JAILR=1.0*jailr specify that each of the X-variables are assumed to be exactly equal to the corresponding latent variable. Note that the part 1.0*latent variable indicates that the path coefficient is fixed at the value of 1.0. In contrast, a command such as COCEU = (0.5)*subabuse indicates that 0.5 is a starting value (preliminary estimate) of $\lambda_{y,21}$. Since we assume that the X-variables measure the corresponding KSI latent variables without error, we set the error variances of NUMTE to JAILR to 0.

```
replic1.spl

Raw Data from file 'replicwts.psf' REWIND

Latent Variables subabuse numte age gender race_d depr educ jailr
Relationships:

ALCEU= 1.0*subabuse
COCEU = subabuse
MAREU = subabuse

NUMTE = 1.0*numte
AGE=1.0*age
GENDER = 1.0*gender
RACE_D = 1.0*race_d
DEPR = 1.0*depr
EDU = 1.0*educ
JAILR = 1.0*jailr

Set the Error Variance of NUMTE to 0.0
Set the Error Variance of AGE to 0.0
Set the Error Variance of GENDER to 0.0
Set the Error Variance of RACE_D to 0.0
Set the Error Variance of DEPR to 0.0
Set the Error Variance of EDU to 0.0
Set the Error Variance of JAILR to 0.0
```

Next, we assume that subabuse is predicted by numte, and, in turn, numte is predicted by age, gender, race_d, depr, educ and jailr. This part of the syntax is shown below.

```
replic1.spl

!--------------------------------!
! Use numte as mediating variable !
!--------------------------------!

subabuse = numte
numte = age gender race_d depr educ jailr
Set the Error Covariance of numte and subabuse Free

LISREL OUTPUT: RP=1 ND=4 AD=OFF SV=replic1.std PV=replic1.par
Path Diagram
End of Problem

  ※
```

To allow for the estimation of the covariance of the errors between the ETA latent variables numte and subabuse, we SET the error covariance between these variables free.

Finally, we use the LISREL OUTPUT: command to specify the number of repetitions (RP = 1), number of decimals (ND = 4), admissibility check off (AD = OFF), and save the standard error estimates (SV = replic1.std) and parameter estimates (PV = replic1.par) to text files **replic1.std** and **replic1.par** respectively.

Discussion of results – SIMPLIS syntax

The estimated path coefficients for weight = A2TWA0 are shown in the path diagram given below. Although not presented here, all coefficients are statistically significant. The χ^2-statistic for goodness of fit is 24.13, degrees of freedom is 19, and the *p*-value is 0.19103.

Chi-Square=24.13, df=19, P-value=0.19103, RMSEA=0.000
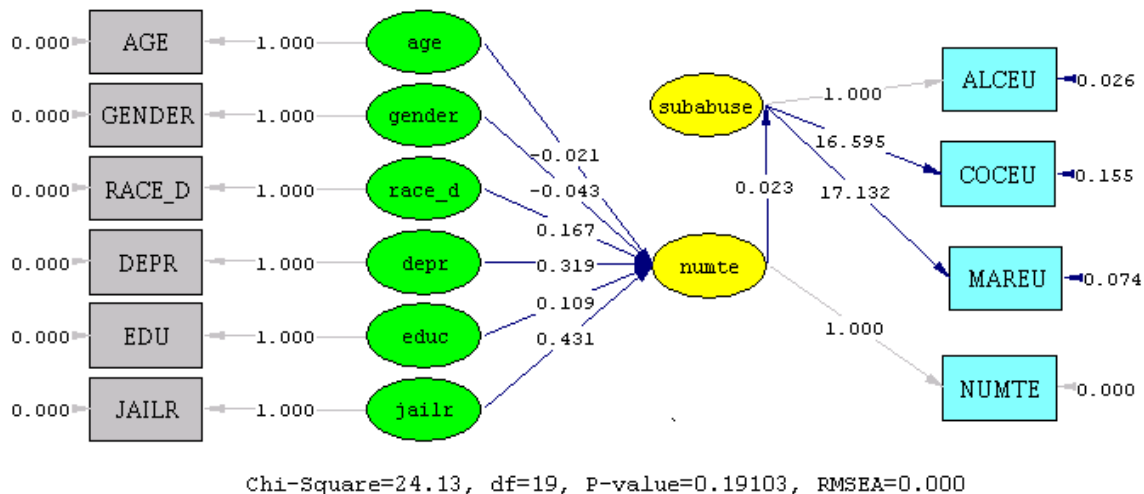
Contents of the files **replic1.par** and **replic1.std** are given below. The estimates are preceded by three numbers N1, N2, and N3, where N1 = repetition number, while N2 and N3 are zero if convergence has been attained. The parameter estimate and standard error values are given in scientific notation. For example,

$$0.165949D + 02 = 0.165942 \times 10^2 = 16.5942$$
$$0.234136D + 0 = 0.23413 \times 10^0 = 0.23413$$

In general, $D + k$ implies that the decimal point should be moved k positions to the right, whereas $D - k$ implies the insertion of k zero values just after the decimal point. For example, $0.868283D - 03 = 0.000868283$.

```
replic1.par
       1  0  0
   0.165949D+02  0.171319D+02  0.227140D-01-0.212969D-01-0.429678D-01  0.167480D+00
   0.318534D+00  0.108624D+00  0.430552D+00  0.119557D+03-0.120186D+00  0.189763D+00
   0.414549D+00  0.599595D-02  0.189501D+00  0.296351D-01  0.326566D-01-0.252343D-01
   0.248619D+00  0.237541D+01-0.299360D-02-0.331195D-01-0.570921D-02  0.875714D+00
  -0.204854D+00-0.476042D-01  0.165119D-01  0.549779D-02-0.484843D-01  0.247219D+00
   0.725569D-02-0.318796D+00  0.143960D+02  0.264990D-01  0.155231D+00  0.736832D-01
```

```
replic1.std
       1  0  0
   0.234136D+00  0.271913D+00  0.868283D-03  0.554108D-03  0.421755D-02  0.103334D-01
   0.106231D-01  0.370566D-02  0.154622D-01  0.430864D+00  0.944686D-02  0.761436D-03
   0.170966D-01  0.434721D-03  0.135845D-02  0.152616D-01  0.412085D-03  0.687255D-03
   0.148745D-03  0.208347D-01  0.616706D-03  0.131988D-02  0.973212D-03  0.283429D-02
   0.112422D-01  0.529228D-03  0.741812D-03  0.538325D-03  0.125730D-02  0.204008D-03
   0.276414D-03  0.100000D+01  0.560179D+00  0.305170D-03  0.107011D-02  0.922522D-03
```

Example: Use of replicate weights
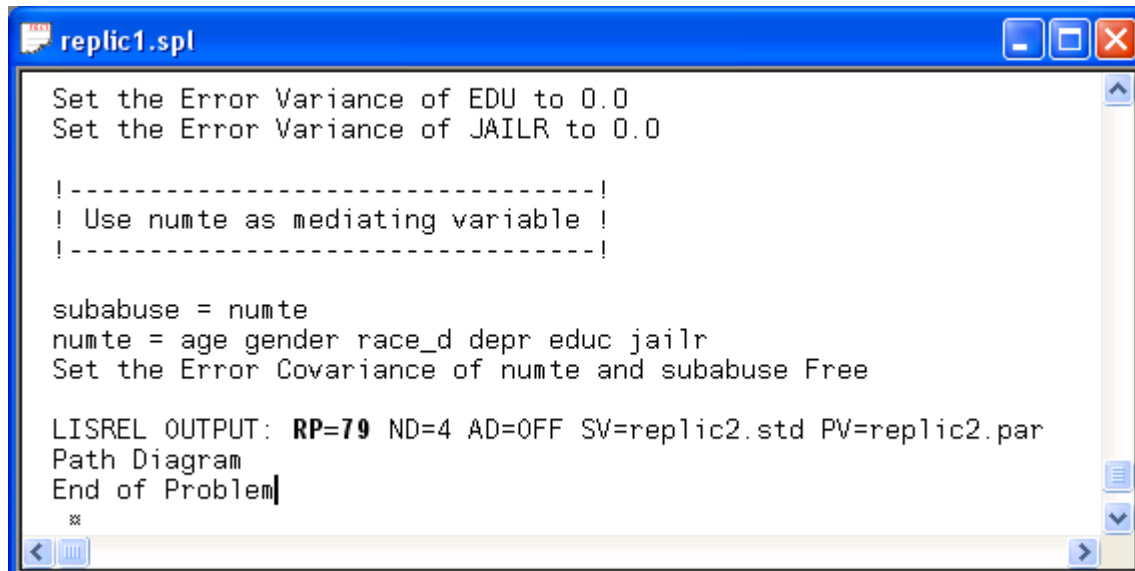
We now demonstrate the use of replicate weights in LISREL, and start by making the PSF window the active window in order to invoke the PSF menu bar. From the **Data** menu, select **Survey Design** and remove the stratum variable CENREG and the cluster variable FACTYPE by clicking on the corresponding **Remove** buttons. Select the **File**, **Save** option to save this change.

Next, edit the SIMPLIS syntax file by changing RP = 1 to RP = 79 (shown in bold typeface below) and use the filenames **replic2.std** and **replic2.par** for saving the standard errors and parameter estimates respectively.

It is important to note that LISREL assumes that there are a total of 79 weight variables in the data set, starting with the **Design weight** variable name (in the present case A2TWA0) selected in the **Survey Design** dialog box. Additionally, it is assumed that the weight columns follow one another.

```
replic1.spl

Set the Error Variance of EDU to 0.0
Set the Error Variance of JAILR to 0.0

!-------------------------------!
! Use numte as mediating variable !
!-------------------------------!

subabuse = numte
numte = age gender race_d depr educ jailr
Set the Error Covariance of numte and subabuse Free

LISREL OUTPUT: RP=79 ND=4 AD=OFF SV=replic2.std PV=replic2.par
Path Diagram
End of Problem
```

Discussion of results – using replicate weights

The results below show the parameter estimates and parameter standard error estimates for the model fitted with design weight A2TWA0, including stratification (CENREG) and clustering (FACTYPE) variables, and the corresponding average values using replicate weights without CENREG and FACTYPE.

In general, parameter estimates for the two estimation methods are quite close. Standard error estimates, on the other hand, tend to be larger for the replicate method. This is clearly a topic that requires further research.

Parameter	Estimates		Standard Errors	
	(A2TWAO)	(Replicates)	(A2TWAO)	(Replicates)
LY21	16.59493	16.55000	3.49175	4.58610
LY31	17.13186	16.48700	4.05513	4.60610
BETA12	0.02271	0.02073	0.01295	0.00534
GAMMA21	-0.02130	-0.02207	0.00826	0.00662
GAMMA22	-0.04297	-0.03559	0.06290	0.07759
GAMMA23	0.16748	0.22387	0.15410	0.09480
GAMMA24	0.31853	0.43322	0.15843	0.13714
GAMMA25	0.10862	0.13610	0.05526	0.04669
GAMMA26	0.43055	0.57974	0.23059	0.16802
PSI11	0.00726	0.00633	0.00412	0.00187
PSI21	-0.31880	-0.28876	0.00000	0.01568
PSI22	14.39595	14.27700	8.35414	4.16910
TE11	0.02650	0.02649	0.00455	0.00346
TE22	0.15523	0.15176	0.01596	0.00997
TE33	0.07368	0.07615	0.01376	0.00883

Parameter	Estimates		Standard Errors	
	(A2TWAO)	(Replicates)	(A2TWAO)	(Replicates)
PHI11	119.55744	119.55000	6.42561	5.20570
PHI21	-0.12019	-0.12074	0.14088	0.11214
PHI22	0.18976	0.18974	0.01136	0.00508
PHI31	0.41455	0.41367	0.25497	0.11399
PHI32	0.00600	0.00596	0.00648	0.00482
PHI33	0.18950	0.18948	0.02026	0.00522
PHI41	0.02964	0.02911	0.22760	0.19762
PHI42	0.03266	0.03263	0.00615	0.00737
PHI43	-0.02523	-0.02497	0.01025	0.00695
PHI44	0.24862	0.24873	0.00222	0.00112
PHI51	2.37541	2.37230	0.31071	0.32812
PHI52	-0.00299	-0.00302	0.00920	0.01064
PHI53	-0.03312	-0.03312	0.01968	0.01036
PHI54	-0.00571	-0.00588	0.01451	0.01641
PHI55	0.87571	0.87582	0.04227	0.02947
PHI61	-0.20485	-0.20241	0.16766	0.15295
PHI62	-0.04760	-0.04758	0.00789	0.00615
PHI63	0.01651	0.01673	0.01106	0.00663
PHI64	0.00550	0.00588	0.00803	0.00945
PHI65	-0.04848	-0.04834	0.01875	0.01407
PHI66	0.24722	0.24768	0.00304	0.00149

5.6   Theory

5.6.1   Introduction

We assume that the population from which the sample data are obtained can be stratified into H strata. Within each stratum h, n_h clusters or primary sampling units (PSUs) are drawn and within the h-th stratum and k-th cluster, n_{h_k} ultimate sampling units (USUs) are drawn with design weights w_{hkl}, where l denotes the l-th USU within the k-th cluster, which in turn is nested within stratum h.

In the subsequent sections we discuss parameter estimation and the Taylor linearization method employed in LISREL to produce robust standard error estimates under single stage sampling.

5.6.2   Parameter estimation

Assume that $\mathbf{y}_{hkl}$ is distributed as a $p \times 1$ multivariate normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In structural equation modeling, it is hypothesized that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\gamma})$ and $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a vector of q unknown parameters to be estimated.

Example 1: CFA model with one factor

Consider the model $\boldsymbol{\Sigma}(\mathbf{y}) = \boldsymbol{\lambda}\boldsymbol{\varphi}\boldsymbol{\lambda}' + \mathbf{D}_\theta$. For this model, we have $\boldsymbol{\gamma} = \left(\lambda_1, \lambda_2, ..., \lambda_p, \phi, \theta_1, \theta_2, ..., \theta_p\right)'$. To avoid any indeterminacy in estimating the unknown parameters, one would typically set the variance (ϕ) of the latent factor equal to 1, or alternatively fix one of the factor loadings, for example λ_1, to 1. In the latter case, $\boldsymbol{\gamma} = (\lambda_2, ..., \lambda_p, \psi, \theta_1, \theta_2, ..., \theta_p)'$.

Complete data

In the case of complete data, parameter estimation is relatively straightforward and can be summarized by the following two steps.

Step 1

Calculate the natural logarithm of the likelihood function, $\ln L$, where

$$\ln L = \sum_{h=1}^{H} \sum_{k=1}^{n_h} \sum_{l=1}^{n_{hk}} w_{hkl} \ln f\left(\mathbf{y}_{hkl} \mid \boldsymbol{\gamma}\right)$$

$$(5.11)$$

where

$$f\left(\mathbf{y}_{hkl} \mid \boldsymbol{\gamma}\right) = \left(2\pi\right)^{-p/2} \left|\boldsymbol{\Sigma}\right|^{-1/2} \exp\left(-\frac{1}{2}tr\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\gamma}\right)\mathbf{G}\right\}\right)$$

(5.12)

and

$$\mathbf{G} = \left(\mathbf{y}_{hkl} - \boldsymbol{\mu}(\boldsymbol{\gamma})\right)\left(\mathbf{y}_{hkl} - \boldsymbol{\mu}(\boldsymbol{\gamma})\right)'$$

(5.13)

Step 2

Obtain an estimate $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ by solving the set of simultaneous equations

$$\left.\frac{\partial \ln L}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \mathbf{0}$$

(5.14)

In general, no closed-form solution to the set of equations in (5.14) exists, and therefore parameter estimates are obtained iteratively using the Fisher scoring algorithm:

$$\hat{\boldsymbol{\gamma}}^{(t+1)} = \hat{\boldsymbol{\gamma}}^{(t)} + \mathbf{I}_n^{-1}\left(\hat{\boldsymbol{\gamma}}^{(t)}\right)\mathbf{g}\left(\hat{\boldsymbol{\gamma}}^{(t)}\right)$$

(5.15)

where $\hat{\boldsymbol{\gamma}}^{(t)}$ denotes the parameter values at iteration t, $t = 1, 2, \cdots$, and $\mathbf{g}(\cdot)$ denotes the gradient vector and where $\mathbf{I}_n(\cdot)$ denotes the Fisher information matrix for the elements of $\boldsymbol{\gamma}$.

In other words,

$$\mathbf{g}(\boldsymbol{\gamma}) = \frac{\partial \ln L}{\partial \boldsymbol{\gamma}}$$

(5.16)

and

$$\mathbf{I}_n(\boldsymbol{\gamma}) = -E\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}\right]$$

(5.17)

Iterations are continued until $\left|\hat{\gamma}_i^{(t+1)} - \hat{\gamma}_i^{(t)}\right| \prec \varepsilon \quad \forall i = 1, 2, \cdots, q$ where ε is a small scalar value, *e.g.* 10^{-6}.

Incomplete data

Since the case of no missing values may be considered as a special case of a general framework that handles missing values, specific expressions for the gradient vector and information matrix will be given in this section.

Suppose that the data set contains n cases, where $n = \sum_{h=1}^{H} \sum_{k=1}^{n_h} \sum_{l=1}^{n_{hk}} 1$, then one could alternatively write

(5.11) as $\ln L = \sum_{\alpha=1}^{n} w_\alpha \ln f\left(\mathbf{y}_\alpha | \boldsymbol{\gamma}\right)$, where α denotes the subscript hkl.

Suppose that the data can be split into two subsets of sizes n_1 and n_2, such that subset 1 contains no missing values and subset 2 contains no values for the p-th element of $\mathbf{y}_\alpha$, $\alpha \in \left[\text{subset 2}\right]$. Let w_{1j} and $\mathbf{y}_{1j}$, $j = 1, 2, \cdots, n_1$, denote the weights and observations in subset 1, while w_{2j} and $\mathbf{y}_{2j}$, $j = 1, 2, \cdots, n_2$, denote the weights and observations in subset 2. In this case, we can write $\ln L$ as

$$\sum_{i=1}^{2} \sum_{j=1}^{n_i} w_{ij} \ln f\left(\mathbf{y}_{ij} | \boldsymbol{\gamma}\right)$$

(5.18)

Suppose that, in general, there are NPAT patterns of missingness, $i = 1, 2, \ldots,$ NPAT and that the number of observations within a pattern equals n_i, $j = 1, 2, \ldots, n_i$. Let p_i indicate the number of variables with non-missing data.

Example 2

Pattern	y1	y2	y3	y4	p_i	w_{ij}
1	19.3	36.4	18.7	.	3	1.8
1	17.6	34.3	19.1	.	3	1.7
1	16.4	32.8	17.9	.	3	1.4
2	20.1	40.2	19.6	57.8	4	0.8
2	19.4	39.3	20.1	58.9	4	0.7
3	.	36.5	.	62.4	2	1.1
4	.	.	.	60.2	1	1.3

For the data shown above, NPAT $= 4$, $n = 7$, $n_1 = 3$, $n_2 = 2$, $n_3 = 1$ and $n_4 = 1$.

In general,

$$\ln L = -\frac{1}{2}\sum_{i=1}^{NPAT}\sum_{j=1}^{n_i} w_{ij}\left\{p_i\ln 2\pi + \ln|\mathbf{\Sigma}_i| + tr\mathbf{\Sigma}_i^{-1}\left(\mathbf{y}_{ij}-\mathbf{\mu}_i\right)\left(\mathbf{y}_{ij}-\mathbf{\mu}_i\right)'\right\}$$

(5.19)

Let

$$w_{i\cdot} = \sum_{j=1}^{n_i} w_{ij}$$

(5.20)

and

$$\bar{\mathbf{y}}_{w_i} = \frac{1}{w_{i\cdot}}\sum_{j=1}^{n_i} w_{ij}\mathbf{y}_{ij}$$

(5.21)

Example 3

For the data set given in Example 1, $w_{1\cdot} = 1.8 + 1.7 + 1.4 = 4.9$, and

$$\bar{\mathbf{y}}_{w_1} = \frac{1}{4.9}\left\{1.8\begin{bmatrix}19.3\\36.4\\18.7\end{bmatrix}+1.7\begin{bmatrix}17.6\\34.3\\19.1\end{bmatrix}+1.4\begin{bmatrix}16.4\\32.8\\17.9\end{bmatrix}\right\}=\begin{bmatrix}17.88\\34.64\\18.61\end{bmatrix}$$

Ignoring the weights, $\bar{\mathbf{y}}_1' = (17.77\ 34.50\ 18.57)$.

Use of (5.20) and (5.21) leads to the following expression for (5.19):

$$\ln L = -\frac{1}{2}\sum_{i=1}^{NPAT}\left\{w_{i\cdot}\left(p_i\ln 2\pi + \ln|\mathbf{\Sigma}_i|\right)+tr\mathbf{\Sigma}_i^{-1}\sum_{j=1}^{n_i}w_{ij}\left(\mathbf{y}_{ij}\mathbf{y}_{ij}'-\mathbf{y}_{ij}\mathbf{\mu}_i'-\mathbf{\mu}_i\mathbf{y}_{ij}'+\mathbf{\mu}_i\mathbf{\mu}_i'\right)\right\}$$

$$= -\frac{1}{2}\sum_{i=1}^{NPAT}\left\{w_{i\cdot}\left(p_i\ln 2\pi + \ln|\mathbf{\Sigma}_i|\right)+tr\mathbf{\Sigma}_i^{-1}\left\{\sum_{j=1}^{n_i}w_{ij}\mathbf{y}_{ij}\mathbf{y}_{ij}'-w_{i\cdot}\bar{\mathbf{y}}_{w_i}\mathbf{\mu}_i'-\mathbf{\mu}_i w_{i\cdot}\bar{\mathbf{y}}_{w_i}'-w_{i\cdot}\bar{\mathbf{y}}_{w_i}\bar{\mathbf{y}}_{w_i}'+w_{i\cdot}\bar{\mathbf{y}}_{w_i}\bar{\mathbf{y}}_{w_i}'+w_{i\cdot}\mathbf{\mu}_i\mathbf{\mu}_i'\right\}\right\}$$

Therefore

$$\ln L = -\frac{1}{2}\sum_{i=1}^{NPAT}\left\{w_{i\cdot}\left(p_i\ln 2\pi + \ln|\mathbf{\Sigma}_i|\right)+tr\mathbf{\Sigma}_i^{-1}\left(\mathbf{G}_{w_i}+\mathbf{S}_{w_i}\right)\right\}$$

(5.22)

where

$$\mathbf{G}_{w_i} = \left(\bar{\mathbf{y}}_{w_i} - \boldsymbol{\mu}_i\right)\left(\bar{\mathbf{y}}_{w_i} - \boldsymbol{\mu}_i\right)'$$

(5.23)

and

$$\mathbf{S}_{w_i} = \frac{1}{w_{i\cdot}}\sum_{j=1}^{n_i} w_{ij}\left(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{w_i}\right)\left(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{w_i}\right)'$$

(5.24)

or equivalently,

$$w_{i\cdot}\mathbf{S}_{w_i} = \sum_{j=1}^{n_i} w_{ij}\mathbf{y}_{ij}\mathbf{y}_{ij}' - w_{i\cdot}\bar{\mathbf{y}}_{w_i}\bar{\mathbf{y}}_{w_i}'$$

Gradient vector

$$\mathbf{g}(\gamma_r) = \sum_{i=1}^{NPAT}\sum_{j=1}^{n_i} w_{ij}\left\{\left(\mathbf{y}_{ij} - \boldsymbol{\mu}_i\right)'\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\gamma_r} + \frac{1}{2}tr\boldsymbol{\Sigma}_i^{-1}\left[\left(\mathbf{y}_{ij} - \boldsymbol{\mu}_i\right)\left(\mathbf{y}_{ij} - \boldsymbol{\mu}_i\right)' - \boldsymbol{\Sigma}_i\right]\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_r}\right\}$$

(5.25)

Use of (5.20), (5.21), (5.22) and (5.23) gives

$$\begin{aligned}
\mathbf{g}(\gamma_r) &= \sum_{i=1}^{NPAT}\left\{\sum_{j=1}^{n_i}\left(w_{ij}\mathbf{y}_{ij} - w_{ij}\boldsymbol{\mu}_i\right)'\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\gamma_r} + \frac{1}{2}tr\boldsymbol{\Sigma}_i^{-1}\left(w_{i\cdot}\mathbf{S}_{w_i} + w_{i\cdot}\mathbf{G}_{w_i} - w_{i\cdot}\boldsymbol{\Sigma}_i\right)\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_r}\right\} \\
&= \sum_{i=1}^{NPAT} w_{i\cdot}\left\{\left(\bar{\mathbf{y}}_{w_i} - \boldsymbol{\mu}_i\right)'\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\gamma_r} + \frac{1}{2}tr\boldsymbol{\Sigma}_i^{-1}\left(\mathbf{S}_{w_i} + \mathbf{G}_{w_i} - \boldsymbol{\Sigma}_i\right)\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_r}\right\}
\end{aligned}$$

(5.26)

Information matrix

$$\begin{aligned}
\mathbf{I}_n(\gamma_{rs}) &= -E\left(\frac{\partial^2\ln L}{\partial\gamma_r\partial\gamma_s}\right) \\
&= \sum_{i=1}^{NPAT}\sum_{j=1}^{n_i} w_{ij}\left\{\frac{\partial\boldsymbol{\mu}_i'}{\partial\gamma_r}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\gamma_s} + \frac{1}{2}tr\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_r}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_s}\right\} \\
&= \sum_{i=1}^{NPAT} w_{i\cdot}\left\{\frac{\partial\boldsymbol{\mu}_i'}{\partial\gamma_r}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\gamma_s} + \frac{1}{2}tr\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_r}\boldsymbol{\Sigma}_i^{-1}\frac{\partial\boldsymbol{\Sigma}_i}{\partial\gamma_s}\right\}
\end{aligned}$$

(5.27)

Let

$$\mathbf{r}_{ij} = \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{y}_{ij} - \boldsymbol{\mu}_{ij} \right)$$

(5.28)

and

$$\mathbf{P}_{ij} = \mathbf{r}_{ij} \mathbf{r}_{ij}' - \boldsymbol{\Sigma}_{i.}$$

(5.29)

From (5.26) it follows that

$$[\mathbf{g}(\boldsymbol{\gamma})]_r = \sum_{i=1}^{NPAT} \sum_{j=1}^{n_i} [\mathbf{g}_{ij}]_r$$

(5.30)

where

$$[\mathbf{g}_{ij}]_r = w_{ij} \left\{ \mathbf{r}_{ij}' \frac{\partial \boldsymbol{\mu}_i}{\partial \gamma_r} + \frac{1}{2} tr \left\{ \mathbf{P}_{ij} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \gamma_r} \right\} \right\}, \quad i = 1, 2, ..., NPAT; \ j = 1, 2, ..., n_i$$

(5.31)

Approximate covariance matrix of estimators

An approximate covariance matrix of the parameter estimators is derived as follows. From (5.14), (5.16) and (5.30) it follows that $\hat{\boldsymbol{\gamma}}$ is the solution to the set of equations

$$\hat{\mathbf{w}}(\hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{NPAT} \sum_{j=1}^{n_i} \mathbf{g}_{ij}(\hat{\boldsymbol{\gamma}}) = \mathbf{0}$$

(5.32)

Note that each $\mathbf{g}_{ij}$ is associated with a specific case *hkl*, where *h* denotes strata, *k* clusters and *l* ultimate sample units. Using a first-order Taylor expansion of $\hat{\mathbf{w}}(\hat{\boldsymbol{\gamma}})$ at $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$, it follows that

$$\mathbf{0} = \hat{\mathbf{w}}(\hat{\boldsymbol{\gamma}}) \approx \hat{\mathbf{w}}(\boldsymbol{\gamma}) + \frac{\partial \hat{\mathbf{w}}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$$

(5.33)

Taking variances on both sides, it further follows that

$$Cov\left(\hat{\mathbf{w}}(\hat{\boldsymbol{\gamma}})\right) \approx \frac{\partial \hat{\mathbf{w}}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'} Cov(\hat{\boldsymbol{\gamma}}) \left(\frac{\partial \hat{\mathbf{w}}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'}\right)'.$$

(5.34)

Thus, provided that (*cf.* (5.32)) $\dfrac{\partial^2 \ln L}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \dfrac{\partial}{\partial \boldsymbol{\gamma}} \left[\dfrac{\partial \mathbf{g}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'} \right]$ is a non-singular matrix,

$$Cov(\hat{\boldsymbol{\gamma}}) \approx \left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right]^{-1} Cov\left(\hat{\mathbf{w}}(\hat{\boldsymbol{\gamma}})\right) \left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right]^{-1},$$

where $E\left[\dfrac{\partial^2 \ln L}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right] = -\mathbf{I}_n(\boldsymbol{\gamma})$.

Therefore, an approximate expression for the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}$ is given by

$$Cov(\hat{\boldsymbol{\gamma}}) \approx \mathbf{I}_n^{-1}(\boldsymbol{\gamma}) \mathbf{G} \mathbf{I}_n^{-1}(\boldsymbol{\gamma})$$

(5.35)

where $\mathbf{G} = Cov\left(\hat{\mathbf{w}}(\hat{\boldsymbol{\gamma}})\right)$.

Using results derived by Binder (1983) and Fuller (1975), it follows that

$$\mathbf{G} = \frac{n-1}{n-q} \sum_{h=1}^{H} \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} \left(\mathbf{t}_{h_{i.}} - \bar{\mathbf{t}}_{h_{..}}\right)\left(\mathbf{t}_{h_{i.}} - \bar{\mathbf{t}}_{h_{..}}\right)'$$

(5.36)

where

- $n_h = \sum\limits_{j=1}^{n_{hi}} m_{h_{ij}}$, with $m_{h_{ij}}$ the number of cases with identical response patterns within stratum h, cluster i, and USU j. If $f_{h_{ij}} = 1$ for all h, then $m_{h_{ij}} = 1$ for all h, i and j.

- $f_h = \dfrac{n_h}{N_h}$, the sampling rate for stratum h.

- $\mathbf{t}_{h_{ij}} = \mathbf{g}_{h_{ij}}(\hat{\boldsymbol{\gamma}})$, where $\mathbf{g}_{h_{ij}}(\hat{\boldsymbol{\gamma}})$ (*cf.* (5.31)) is the h_{ij}-th contribution to the gradient vector $\mathbf{g}(\boldsymbol{\gamma})$

as defined by (5.30).

- $\mathbf{t}_{h_i.} = \sum\limits_{j=1}^{m_{h_{ij}}} \mathbf{t}_{h_{ij}}$

- $\bar{\mathbf{t}}_{h_{..}} = \dfrac{1}{n_h} \sum\limits_{i=1}^{n_h} \mathbf{t}_{h_i.}$

In practice, we assume a zero contribution to **G** for strata that contain a single PSU (cluster). Additionally, if the data do not contain a stratification variable, the PSUs are assumed to be the strata, and the observations (ultimate sampling units) within each PSU, clusters. Likewise, if there is no variable to define clusters, the observations within each stratum are treated as being the primary sampling units.

Adjustment to the chi-square goodness of fit statistics

Simulation studies indicated that the χ^2_{LR}-statistic based on the log-likelihood (*cf.* (5.22)) in general yields too high a rejection rate. Let (*cf.* (5.35))

$$ d = tr\left(\mathbf{I}_n\left(\hat{\boldsymbol{\gamma}}\right) Cov\left(\hat{\boldsymbol{\gamma}}\right)\right), $$

(5.37)

where $\mathbf{I}_n\left(\hat{\boldsymbol{\gamma}}\right)$ denotes the information matrix defined by (5.27).

A correction to the χ^2-statistic for testing model fit is given by

$$ \chi^2_{robust} = c \times \chi^2_{LR}, $$

where

$$ c = \frac{q}{d}, $$

and where q denotes the number of parameters to be estimated.

5.7 References

Asparouhov, T. (2004). Weighting for Unequal Probability of Selection in Latent Variable Modeling. Mplus Web Notes: No. 7.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys, *International Statistical Review*, **51**, 279-292.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models*. Sage Publications.

Curran, P.J. (1997). Comparing three modern approaches to longitudinal data analysis: An examination of a single developmental sample. Symposium conducted at the meeting of the Society for Research on child Development, Washington, DC.

Curran, P.J., Bauer, D.J., & Willoughby, M.T. (2004). Testing Main Effects and Interactions in Latent Curve Analysis. *Psychological Methods*, **9(2)**, 220-237.

Du Toit, S.H.C. & Browne, M.W. (2001). The covariance structure of a vector ARMA time series. In: R. Cudeck, S.H.C. du Toit, & D. Sörbom (Eds.), *Structural Equation Modeling: Present and Future*, Lincolnwood, IL: Scientific Software International, Inc.

Du Toit, S.H.C. & Cudeck, R. (2001). The analysis of nonlinear random coefficient regression models with LISREL using constraints. In: R. Cudeck, S.H.C. du Toit, & D. Sörbom (Eds.), *Structural Equation Modeling: Present and Future*, Lincolnwood, IL: Scientific Software International, Inc.

Fuller, W.A. (1975). Regression Analysis for Sample Survey. *Sankhya*, Series **C**, **37**, 117-132.

Hayduk, L. (1996). *LISREL issues, debates and strategies*, Baltimore: John Hopkins University Press.

Jöreskog, K.G. & Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling.* Chicago: Scientific Software International.

Kaplan, D. (2000). *Structural Equation Modeling: Foundations and Extensions* (*Advanced Quantitative Techniques in the Social Sciences*). Sage.