



Latent curve analysis with main and interaction effects

Contents

1.	Introduction	1
2.	The data	1
3.	Structural equation modeling	2
4.	Multilevel modeling	3
5.	The model	4
6.	Setting up the analysis	8
7.	Discussion of results	14

1 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.

2 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.

3 Structural equation modeling

A few of the variables in the data set `curran_NLSY_subset.LSF` saved in the **Complex Survey Sample examples** folder are shown below for the first 10 observations.

	genfem0	home_emo	home_cog	antiy1	antiy2	antiy3	antiy4	antiy5
1	1.00	0.80	-1.89	-9.00	2.00	-9.00	1.00	-9.00
2	1.00	-0.20	-1.89	2.00	-9.00	-9.00	7.00	-9.00
3	1.00	1.79	1.11	-9.00	1.00	-9.00	0.00	-9.00
4	1.00	-1.21	-3.89	2.00	-9.00	-9.00	3.00	-9.00
5	1.00	-2.21	-1.89	4.00	-9.00	-9.00	3.00	-9.00
6	0.00	-3.21	0.11	-9.00	-9.00	2.00	-9.00	1.00
7	0.00	1.79	0.11	1.00	-9.00	-9.00	1.00	-9.00
8	1.00	-2.21	-2.89	0.00	-9.00	-9.00	2.00	-9.00
9	0.00	-2.21	-3.89	5.00	-9.00	6.00	-9.00	3.00
10	1.00	-1.21	0.11	-9.00	-9.00	0.00	-9.00	2.00

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 LSF were selected from the survey data:

- 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.

- 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, based on a summation of 14 dichotomously scored items reported by the mother.
- o genxemo: A variable intended to represent the interaction between a child’s gender and level of emotional support: $\text{genxemo} = \text{genfemo} \times \text{home_emo}$.

4 Multilevel modeling

For the multilevel analysis, the LSF **curran_mlev.LSF** 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 LSF 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.

	MOM_ID	tim	Antiy	MOM_WT	genfemo	home_emo	home_cog	genxemo
1	1.00	1.00	2.00	493090.00	1.00	0.80	-1.89	0.80
2	1.00	3.00	1.00	493090.00	1.00	0.80	-1.89	0.80
3	1.00	5.00	0.00	493090.00	1.00	0.80	-1.89	0.80
4	1.00	7.00	2.00	493090.00	1.00	0.80	-1.89	0.80
5	2.00	0.00	2.00	598290.00	1.00	-0.20	-1.89	-0.20
6	2.00	3.00	7.00	598290.00	1.00	-0.20	-1.89	-0.20
7	3.00	1.00	1.00	159960.00	1.00	1.79	1.11	1.79
8	3.00	3.00	0.00	159960.00	1.00	1.79	1.11	1.79
9	3.00	5.00	2.00	159960.00	1.00	1.79	1.11	1.79
10	3.00	7.00	0.00	159960.00	1.00	1.79	1.11	1.79

The following variables were used in the multilevel analysis:

- 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 tim: This variable indicates the time of measurement and varies from 0 to 9.

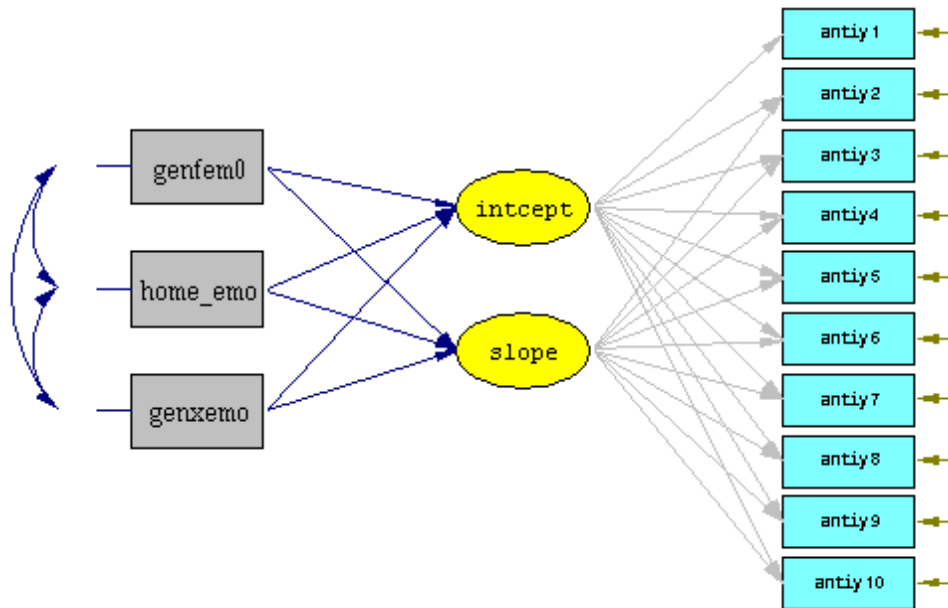
- Antiy: A measure of antisocial behavior in the child at a given measurement occasion. This continuous measure, representing the sum of six items assessing child antisocial behavior over the previous 3 months, was created with values ranging between 0 and 12, where a high value would indicate a higher level of antisocial behavior.
- Mom_Wt: The sampling weight for each mother.
- 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.
- home_cog: A measure of cognitive stimulation, based on a summation of 14 dichotomously scored items reported by the mother.
- genxemo: A variable intended to represent the interaction between a child’s gender and level of emotional support: $\text{genxemo} = \text{genfemo} \times \text{home_emo}$.

5 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).

The LISREL model consists of a measurement and structural part.

Measurement model

The measurement model may be expressed as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \Lambda_y & \mathbf{0} \\ \mathbf{0} & \Lambda_x \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{bmatrix}$$

where

$$\mathbf{y} = [\text{antiy1} \quad \text{antiy2} \quad \dots \quad \text{antiy10}]'$$

$$\mathbf{x} = [\text{genfemo} \quad \text{hom_emo} \quad \text{genxemo}]'$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_{10}]'$$

$$\boldsymbol{\delta} = [\delta_1 \quad \delta_2 \quad \delta_3]'$$

$$\boldsymbol{\xi} = [\text{genfemo} \quad \text{hom_emo} \quad \text{genxemo}]'$$

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

and

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

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

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

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

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

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

and

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

The unknown model parameters are therefore $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \alpha_1, \alpha_2, \phi_{11}, \phi_{12}, \dots, \phi_{33}, \psi_{11}, \psi_{12}, \psi_{22}$, and $\text{var}(\varepsilon_1)$.

Multilevel model

A general two-level model for a response variable y depending on a set of r predictors x_1, x_2, \dots, 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 $i = 1, 2, \dots, N$ denotes the level-2 units, and $j = 1, 2, \dots, n_i$ 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 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$, and $\mathbf{e}_{i1}, \mathbf{e}_{i2}, \dots, \mathbf{e}_{in_i}$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(1)}$.

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.

$$\begin{aligned}
antiy_{ij} = & \beta_0 + \beta_1 * genfemo_{ij} + \beta_2 * home_emo_{ij} + \beta_3 * genxemo_{ij} + \\
& \beta_4 * tim_{ij} + \beta_5 * (genfemo_{ij})(tim_{ij}) + \beta_6 * (home_emo_{ij})(tim_{ij}) + \\
& \beta_7 * (genxemo_{ij})(tim_{ij}) + u_{i0} + u_{i1} * tim_{ij} + e_{ij}
\end{aligned}$$

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 $\beta_1, \beta_2, \dots, \beta_7$ 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.

6 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.

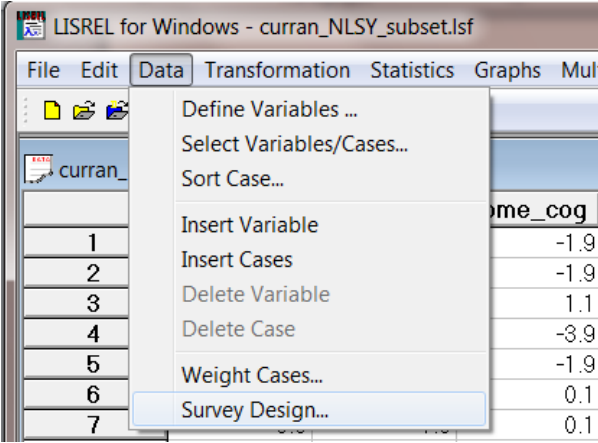
```

curran1.spl
Linear Growth Curve NLSY Data Curran et al (2004)
RAW DATA from file Curran_NLSY_subset.LSF
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

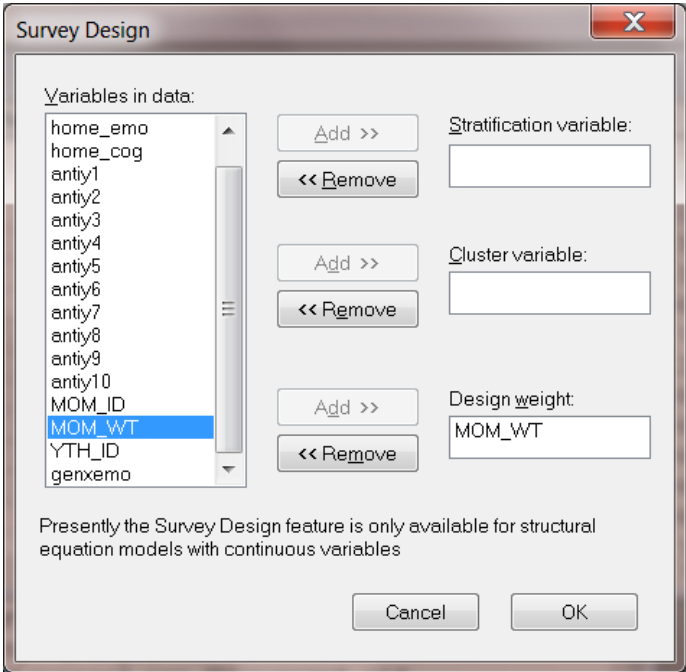
```

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. 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



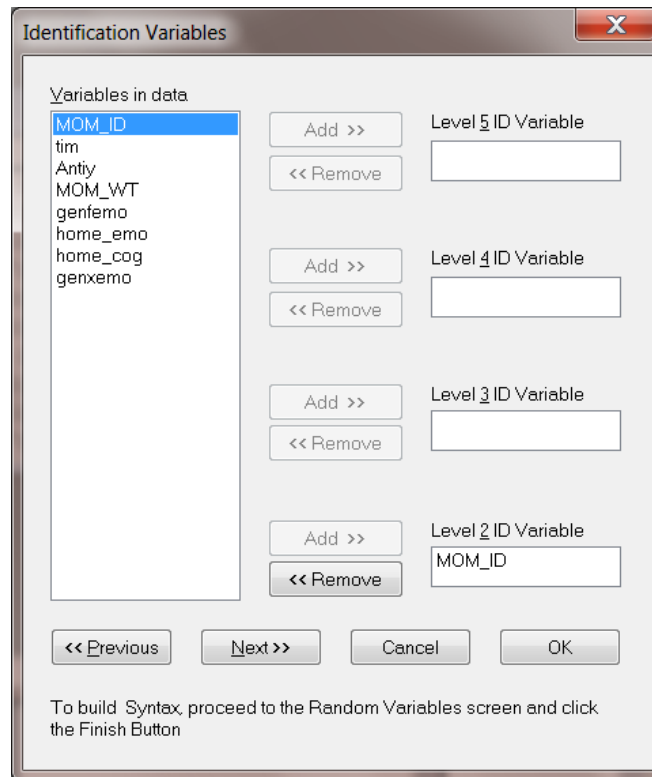
and enter the variable MOM_WT in the **Design weight** field as shown below.



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 discussed in the *Multilevel Examples Guide*. Briefly, the level-2 ID is identified as Mom_ID (see **Identification Variables** dialog box below), 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 **Select Response and Fixed Variables** dialog box below. The weight variable Mom_Wt used in the second of the multilevel analyses discussed here is entered in the **Weight Variables** dialog box.



Weight Variables X

Variables in data

- MOM_ID
- tim
- Antiy
- MOM_WT**
- genfemo
- home_emo
- home_cog
- genxemo

Add >> << Remove

Add >> << Remove

Add >> << Remove

Add >> << Remove

Add >> << Remove

Level 5 Weight:

Level 4 Weight:

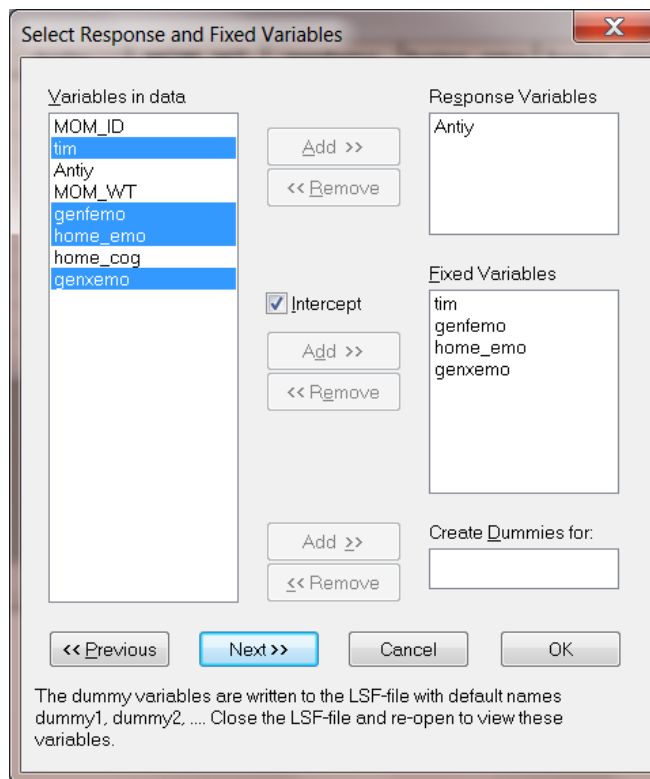
Level 3 Weight:

Level 2 Weight:

Level 1 Weight: MOM_WT

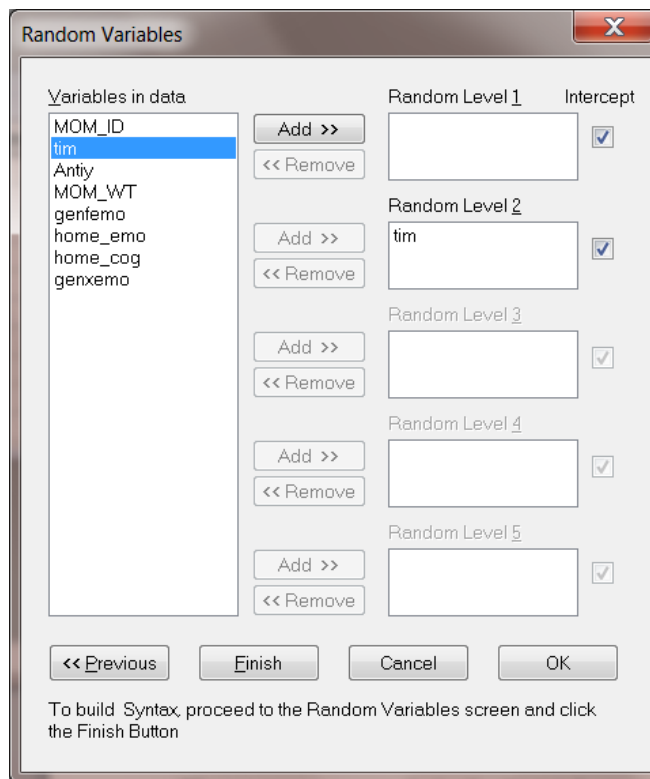
<< Previous Next >> Cancel OK

To build Syntax, proceed to the Random Variables screen and click the Finish Button



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 by the GUI.

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:

```

curran_mlev1.PRL
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=15 OUTPUT=STANDARD ;
TITLE=Curran 2004 paper using weights;
SY='C:\LISREL9\Examples\MISSINGEX\curran_mlev.LSF';
ID2=MOM_ID;
WEIGHT1=MOM_WT;
RESPONSE=Antiy;
FIXED=intcept tim genfemo home_emo genxemo;
RANDOM1=intcept;
RANDOM2=intcept tim;

```

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.PRL
OPTIONS OLS=YES CONVERGE=0.001000 MAXITER=15 OUTPUT=STANDARD ;
TITLE=Curran 2004 paper using weights;
SY='C:\LISREL9 Examples\MISSINGEX\curran_mlev.LSF';
ID2=MOM_ID;
WEIGHT1=MOM_WT;
RESPONSE=Antiy;
FIXED=intcept tim genfemo home_emo genxemo tim*genfemo tim*home_emo tim*genxemo;
RANDOM1=intcept;
RANDOM2=intcept tim;

```

7 Discussion of results

The results of both the SEM and multilevel models are summarized in Table 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 6.

Table 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 × genfemo (γ_4)	0.013	0.013	0.013	0.035	0.035	0.035
tim × home_emo (γ_5)	0.012	0.012	0.012	0.010	0.010	0.010
tim × genfemo × 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
<i>Var(intcept)</i> (ϕ_{11})	*	0.669	0.669	*	0.203	0.204
<i>Var(time slope)</i> (ϕ_{22})	*	0.019	0.019	*	0.009	0.009
<i>Cov(intcept, tim)</i> (ϕ_{21})	*	0.076	0.076	*	0.035	0.035
σ^2 (var(ε))	*	1.758	1.758	*	0.097	0.098

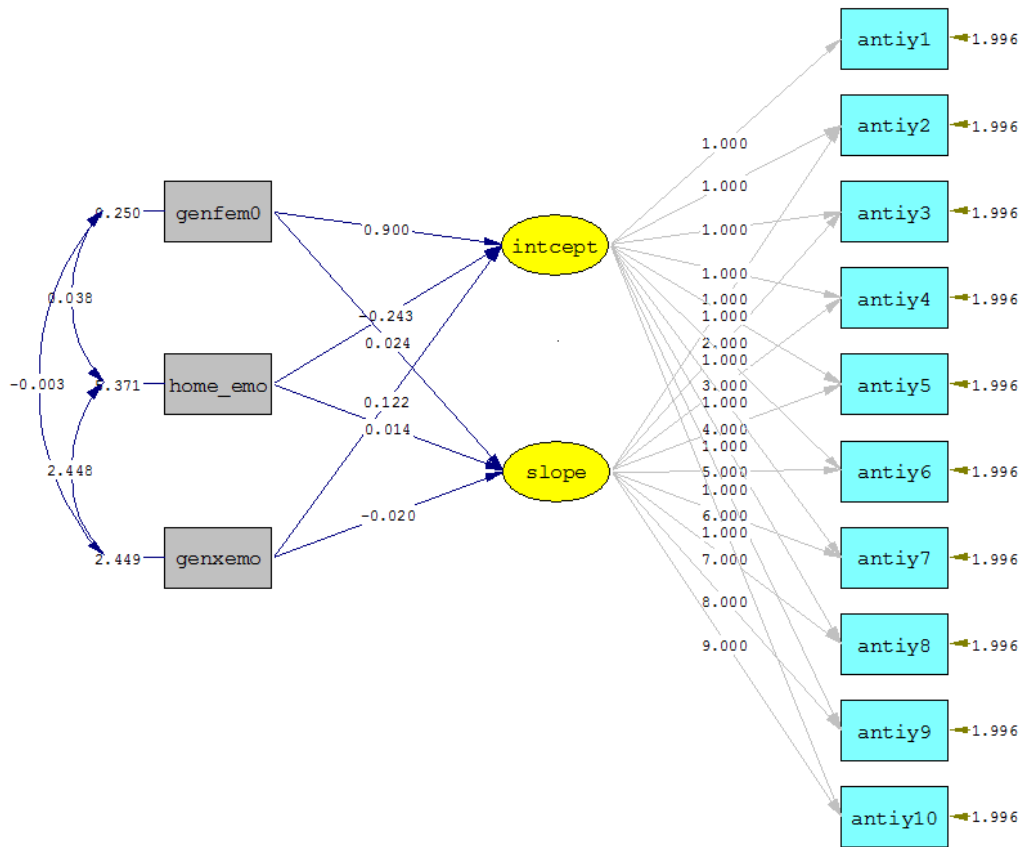
* Not reported in the Curran *et. al.* paper

Table 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
<i>Var(intcept)</i> (ϕ_{11})	0.487	0.415	0.317	0.330
<i>Var(time slope)</i> (ϕ_{22})	0.230	0.022	0.014	0.015
<i>Cov(intcept, tim)</i> (ϕ_{21})	0.091	0.103	0.060	0.063
σ^2 ($\text{var}(\varepsilon)$)	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.

From the path diagram, $\chi^2 = 249.52$, 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. A matrix that is not positive definite can have a large impact on the estimated chi-square value. 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. The χ^2 -statistic in this case is equal to 155.493.



Chi-Square=249.52, df=83, P-value=0.00000, RMSEA=0.070