Growth curves

A general form of the multivariate linear model may be formulated as follows. Rows of the data matrix $\mathbf{Y}$ of order $N \times p$ are independently distributed with the same covariance matrix $\mathbf{\Sigma}$ and with means of the form

$$E(\mathbf{Y}) = \mathbf{X\Xi P},$$

where $\mathbf{X}: N \times q$ and $\mathbf{P}: r \times p$ are fixed design matrices of ranks q and r respectively, and $\mathbf{\Xi}: q \times r$ is a matrix of parameters. If $r = p$ and $\mathbf{P}$ is nonsingular, this is a multivariate regression model with regression matrix $\mathbf{P}'\mathbf{\Xi}'$, with elements which are linear functions of the parameters in $\mathbf{\Xi}$.

This can be specified in LISREL as a submodel 3A in the form of

$$\begin{aligned}\boldsymbol{\eta} &= \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},\\ \mathbf{y} &= \boldsymbol{\tau}_y + \mathbf{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon},\\ x &= \boldsymbol{\tau}_x + \mathbf{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta},\end{aligned}$$

with Fixed-x. Put $\mathbf{\Lambda}_y = \mathbf{P}'$, $\mathbf{\Gamma}_y = \mathbf{\Xi}'$, $\mathbf{\Psi} = \mathbf{\Sigma}$ and leave $\boldsymbol{\tau}_y$, $\boldsymbol{\alpha}$, $\boldsymbol{\kappa}$ and $\mathbf{B}$ default.

This model is often used to estimate growth curves from panel data or longitudinal data in which the outcome variable is measured repeatedly on the same persons over several periods of time.

Consider a response variable *y* being measured on *N* individuals at *T* points in time $t_1, t_2, ..., t_T$. The raw data take the form of a data matrix $\mathbf{Y}$ of order $N \times T$, where y_{ij} is the observed measurement on individual *i* at time t_j. It is assumed that the rows of $\mathbf{Y}$ are independently distributed with the same covariance matrix $\mathbf{\Sigma}$. Also, the mean vectors of the rows are assumed to be the same, namely $\boldsymbol{\mu}' = [\mu_1, \mu_2, ..., \mu_T]$.

However, here we focus attention on the mean μ_t as a function of *t*. This gives a growth curve describing how the mean of *y* changes over time.

We consider polynomial growth curves of the form

$$\mu_t = \kappa_0 + \kappa_1 t + \kappa_2 t^2 + ... + \kappa_h t^h$$

or

$$\mu = \begin{bmatrix} 1 & t_1 & t_1^{\;2} & \dots & t_1^{\;h} \\ 1 & t_2 & t_2^{\;2} & \dots & t_2^{\;h} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_T & t_T^{\;2} & \dots & t_T^{\;h} \end{bmatrix} \begin{bmatrix} \kappa_0 \\ \kappa_1 \\ \vdots \\ \kappa_h \end{bmatrix}.$$

The degree of the polynomial h is assumed to be less than or equal to $T - 1$. When $h < T - 1$, the mean vector μ is constrained and there is not a one-to-one correspondence between $\mu_1, \mu_2, \dots, \mu_T$ and the polynomial coefficients $\kappa_0, \kappa_1, \dots, \kappa_h$. In this example we consider the estimation of these polynomial coefficients.

The above generalizes easily to the case of several groups of individuals with possibly different mean vectors. Suppose, for example, that there are two groups with n_1 and n_2 individuals in each group. Let the first n_1 rows of **Y** be the measurements on individuals in Group 1 and let the last n_2 rows of **Y** be the measurements on individuals in Group 2. The growth curves for the two groups may differ, so we assume that there are two distinct growth curves to be estimated, that is,

$$E\left(y_{it}^{(g)}\right) = \kappa_0^{(g)} + \kappa_1^{(g)} t + \dots + \kappa_h^{(g)} t^h, \quad g = 1, 2.$$

In multiple-group comparisons one may be interested in the following type of questions:

- Should the growth curves be represented by third-degree polynomials, or are quadratic or linear growth curves adequate?
- Should separate growth curves be used for different groups or do all groups have the same growth curve?

Often we conceive of the effect of treatment as represented by a parallel displacement of the whole growth curve for one group in relation to another. This cannot be taken for granted, however, but must be tested by means of data. In addition, growth curves can differ in terms of the degree of the polynomial but also in the shape for the same degree of polynomial. The covariance matrices $\mathbf{\Sigma}$ may be the same in different groups, or the correlation matrices may be the same and the standard deviations different, or the covariance matrices may be all different. The covariance matrices may also be structured in various ways.

Growth curves can be estimated more efficiently and tests about the growth curves will be more powerful if the covariance structure, which arises naturally in repeated measurements, is taken into account. This covariance structure very often has an autoregressive nature. Therefore, we focus attention to the deviation $e_t = y_t - \mu_t$ of y_t from its mean value μ_t on the growth curve and consider various autoregressive models for this.

The first-order autoregressive model is

$$e_t = \beta_t e_{t-1} + z_t, \quad t = 2, 3, \dots, T,$$

where the residual z_t is uncorrelated with e_{t-1}. It is also assumed that $z_2, z_3, \dots, z_T$ are all uncorrelated.

It is readily verified that

$$Cov\left(y_t, y_{t-1}\right) = E\left(e_t, e_{t-1}\right) = \beta_t \sigma_{t-1}^2,$$

where

$$\sigma_{t-1}^2 = Var\left(y_{t-1}\right) = E\left(e_{t-1}^2\right),$$

and that

$$Cov(y_t, y_{t-k}) = \beta_t \beta_{t-1} \ldots \beta_{t-k+1} \sigma_{t-k}^2, \quad k = 1, 2, \ldots$$

Hence, the covariance matrix of **y** is (in the case T = 4)

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & & & \\ \beta_2\sigma_1^2 & \sigma_2^2 & & \\ \beta_3\beta_2\sigma_1^2 & \beta_3\sigma_2^2 & \sigma_3^2 & \\ \beta_4\beta_3\beta_2\sigma_1^2 & \beta_4\beta_3\sigma_2^2 & \beta_4\sigma_3^2 & \sigma_4^2 \end{bmatrix}.$$

It is seen that $\mathbf{\Sigma}$ is constrained; its 10 variances are functions of only seven parameters. Since the variances are free parameters, it is the six covariances that are functions of the three parameters β_2, β_3 and β_4.

This model is the perfect Markov simplex. If all β_i =1, we have the perfect Wiener simplex, see Jöreskog (1970a). Higher-order autoregressive models may also be considered. For example, a second-order model has

$$e_t = \beta_{t,t-1} e_{t-1} + \beta_{t,t-2} e_{t-2} + z_t, \quad t = 3, 4, \ldots, T.$$

The model can be estimated directly with LISREL using the following specification corresponding to the equations at the beginning of the document:

$$\begin{bmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\kappa} \end{bmatrix} + \begin{bmatrix} \mathbf{T}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{bmatrix} + \begin{bmatrix} \mathbf{z} \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{I} & \mathbf{P} \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{bmatrix} + \mathbf{0},$$

where $\mathbf{T}_\beta$ in the case of a second-order autoregressive model, is a square matrix whose subdiagonal non-zero elements are

$$\beta_{21}, \beta_{31}, \beta_{32}, \beta_{42}, \beta_{43}, \beta_{53}, \beta_{54}, \ldots, \beta_{T,T-2}, \beta_{T,T-1}.$$

The matrix **P** is the fixed matrix of order $T \times (h+1)$. When $\mathbf{T}_\beta$ is present in the above set of equations, the covariance matrix of **z** should be diagonal. The case of an unstructured $\mathbf{\Sigma}$ is obtained by setting $\mathbf{T}_\beta$ to zero and letting the covariance matrix of **z** to be free.

As an illustration of growth curve estimate, we use data on stature for boys and girls aged 3-7 (T = 5) (Tudenham & Snyder (1954)). The means, variances, and covariances of the stature measurements for the two groups are given in the table below.

We began by testing the hypothesis that the covariance matrix of the measured variables are equal for boys and girls. This was done in a previous example. The means were not involved and the likelihood ratio test statistics gave χ^2 = 21.59 with 15 degrees of freedom. The p – value is 0.12, so the hypothesis is not rejected.

```
         ETA 1      ETA 2      ETA 3      ETA 4      ETA 5      ETA 6
        --------   --------   --------   --------   --------   --------
 ETA 1    12.628
         (1.537)
           8.216

 ETA 2    12.569     14.506
         (1.590)    (1.766)
           7.907      8.216

 ETA 3    13.270     14.962     16.423
         (1.685)    (1.850)    (1.999)
           7.873      8.087      8.216

 ETA 4    14.133     15.760     17.378     19.797
         (1.825)    (1.992)    (2.155)    (2.410)
           7.743      7.913      8.063      8.216

 ETA 5    15.051     16.593     18.407     20.664     22.462
         (1.944)    (2.110)    (2.290)    (2.541)    (2.734)
           7.742      7.863      8.039      8.132      8.216
```

Table: Berkeley Guidance study

Observed and fitted means					
Age	3	4	5	6	7
Girls					
Observed	95.45	102.99	110.26	117.25	123.41
Fitted	95.46	102.95	110.27	117.18	123.39
Boys					
Observed	96.71	104.27	111.13	117.47	124.01
Fitted	96.70	104.32	111.11	117.54	124.03
Observed (above) and fitted (below) covariance matrices					
Girls					
3	12.110 12.628				
4	12.454 12.569	15.132 14.506			
5	13.491 13.270	16.074 14.962	18.148 16.423		
6	14.061 14.133	16.424 15.760	18.567 17.378	20.612 19.797	
7	14.822 15.051	17.133 16.593	19.587 18.407	21.534 20.664	23.426 22.462
Boys					
3	13.177 12.628				
4	12.693 12.570	13.838 14.504			
5	13.055 13.270	13.784 14.963	14.592 16.423		
6	14.211 14.134	15.049 15.757	16.118 17.379	18.923 19.793	
7	15.294 15.051	16.018 16.592	17.156 18.408	19.738 20.663	21.437 22.461

We now test the hypothesis of equal mean vectors for boys and girls without assuming equal covariance matrices. For this we use Submodel 1 in the form of

$$x = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \xi + \boldsymbol{\delta}.$$

Take $\boldsymbol{\Lambda}_x = \mathbf{I}$ and $\boldsymbol{\Theta}_\delta = \mathbf{0}$. Then the mean vector is $\boldsymbol{\tau}_x$ and the covariance matrix is $\boldsymbol{\Phi}$. The command file (**EX104B.LIS** in the **LISREL Examples** folder) for this is as follows:

```
Example 10.4b: Testing Equality of Mean Vectors. Input for Girls
DA NG=2 NI=5 NO=70
CM FI=GIRLS.COV
ME FI=GIRLS.MEA
MO NX=5 NK=5 TX=FR LX=ID TD=ZE
OU
Example 10.4b: Testing Equality of Mean Vectors. Input for Boys
DA NO=66
CM FI=BOYS.COV
ME FI=BOYS.MEA
MO TX=IN
OU
```

The test gives χ^2 =25.87 with 5 degrees of freedom, so the hypothesis is rejected. Boys and girls have different mean vectors.

We proceed by estimating a growth curve for boys and girls under the assumption that the covariance matrices are equal. The model is the one defined by

$$\begin{bmatrix}\mathbf{e}\\ \boldsymbol{\eta}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \boldsymbol{\kappa}\end{bmatrix} + \begin{bmatrix}\mathbf{T}_\beta & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{e}\\ \boldsymbol{\eta}\end{bmatrix} + \begin{bmatrix}\mathbf{z}\\ \mathbf{0}\end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix}\mathbf{I} & \mathbf{P}\end{bmatrix}\begin{bmatrix}\mathbf{e}\\ \boldsymbol{\eta}\end{bmatrix} + \mathbf{0},$$

with $\mathbf{T}_\beta = \mathbf{0}$ and $Cov(\mathbf{z})$ equal in the two groups but otherwise unconstrained. We assume that the growth curves are cubic, i.e. h = 3. Measuring time as age – 5, the matrix **P** (one can of course use orthogonal polynomials instead):

$$P = \begin{bmatrix} 1 & -2 & 4 & -8\\ 1 & -1 & 1 & -1\\ 1 & 0 & 0 & 0\\ 1 & 1 & 1 & 1\\ 1 & 2 & 4 & 8\end{bmatrix}.$$

In LISREL we take p = 5, m = 9, $\boldsymbol{\Lambda}_y = \begin{bmatrix}\mathbf{I} & \mathbf{P}\end{bmatrix}$, $\boldsymbol{\Theta}_\varepsilon = \mathbf{0}$, α is a 9×1 vector, where the last four elements are the coefficients of the growth curve polynomial, $\mathbf{B}(9\times 9) = \mathbf{0}$, and $\boldsymbol{\Psi}$ a 9×9 symmetric matrix where the last four rows are fixed zeros.

The command file (**EX104C.LIS**) is as follows:

```
Example 10.4c: Estimating Third Degree Growth Curve for Girls
Assuming Sigma(Girls) = Sigma(Boys)
DA NG=2 NI=5 NO=70
CM FI=GIRLS.COV
ME FI=GIRLS.MEA
MO NY=5 NE=9 AL=FR PS=SY,FI TE=ZE
MA LY
1 0 0 0 0 1 -2 4 -8
0 1 0 0 0 1 -1 1 -1
0 0 1 0 0 1  0 0  0
0 0 0 1 0 1  1 1  1
0 0 0 0 1 1  2 4  8
FI AL 1 - AL 5
FR PS 1 1 - PS 5 5
MA PS
12.110
12.454     15.132
13.491     16.074     18.148
14.061     16.424     18.567     20.612
14.822     17.133     19.587     21.534     23.426 /
OU NS RS SE TV AD=OFF
Example 10.4c: Estimating Third Degree Growth Curve for Boys
Assuming Sigma(Girls) = Sigma(Boys)
DA NO=66
CM FI=BOYS.COV
ME FI=BOYS.MEA
MO AL=PS LY=PS PS=IN
OU
```

Since NE is larger than NY, starting values must be provided and NS must be entered on the OU command. It is sufficient to provide starting values for the first five rows of $\mathbf{\Psi}$ for the first group. These starting values are taken to be the covariance matrix for girls. Note that the slash is necessary because $\mathbf{\Psi}$ is actually of order 9×9. $\mathbf{\Lambda}_y$ is the same fixed matrix in both groups. $\mathbf{\Psi}$ is specified to be invariant. The joint covariance matrix of boys and girls is in the upper left 5×5 submatrix of $\mathbf{\Psi}$. All other elements of $\mathbf{\Psi}$ are fixed zeros. The polynomial coefficients are the last 4 elements of $\boldsymbol{\alpha}$. The first 5 elements of $\boldsymbol{\alpha}$ are fixed zeros. AD is set to OFF because $\mathbf{\Lambda}_y$ does not have full column rank and $\mathbf{\Psi}$ is not positive definite.

This model gives an overall χ^2 of 23.29 with 17 degrees of freedom. The *p*-value is 0.14. The polynomial growth curves are estimated as

Girls:
$$\mu_t = 110.278 + 7.165t - 0.214t^2 - 0.046t^3$$
$$(0.486) \quad (0.119) \quad (0.030) \quad (0.025)$$

Boys:
$$\mu_t = 111.112 + 6.537t - 0.186t^2 - 0.074t^3$$
$$(0.500) \quad (0.122) \quad (0.03`) \quad (0.025)$$

The quantities in parentheses are the standard errors of the polynomial coefficients. In the table given earlier in this example, the observed and fitted means and the observed and fitted variances and covariances are given. It is seen that the differences between the observed and fitted quantities are generally small so the overall fit of the model can be regarded as good.

Do boys and girls have the same growth curves? We can test this hypothesis by re-estimating the model under the constraint that the polynomial coefficients are the same and calculating the difference in χ^2-values. This gives χ^2 = 24.46 with 4 degrees of freedom, so the hypothesis is rejected. Inspection of the polynomial coefficients in relation to their standard errors suggest that each coefficient is different for boys and girls. Also, the coefficient κ_3 for girls is not significant. So a quadratic curve would be sufficient fir girls. However, in the range of t from -2 to 2 (t = age – 5), the two curves are in fact close to each other.

Further analysis can be done by including $\mathbf{T}_\beta$ in

$$\begin{bmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\kappa} \end{bmatrix} + \begin{bmatrix} \mathbf{T}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{bmatrix} + \begin{bmatrix} \mathbf{z} \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{I} & \mathbf{P} \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{bmatrix} + \mathbf{0},$$

and letting $\boldsymbol{\Psi}$ be diagonal. A test of the Wiener simplex structure gives an overall χ^2 = 59.63 with 27 degrees of freedom. Although the Wiener simplex is fairly consistent with the observed covariance matrices, the fit is not sufficiently good. A test of the Markov simplex gives χ^2 = 46.93 with 23 degrees of freedom. This model does not fit the data either. For this data it seems best to retain the model with a joint but unstructured covariance matrix.