printed. Note that the order of the parameters in the vector corresponds to the order used by NAG during optimization and not the order of parameter specifications given by the user, or printed by Mx. Note also that using this option for problems with more than a few parameters can result in enormous NAGDUMP.OUT files.

4.4 Fitting Submodels: Saving Matrices and Files

One of the powerful features of Mx is its ability to start again where it left off. An example of this has already been described in Sections 4.3.8 and 4.3.10 above, where repeated attempts to optimize are made either from the current solution or from randomized starting values. Here we describe how repeated fits may be made from the solution, allowing for changes in the model. This can be done manually, writing out matrices to files, or automatically within the same run, using the `Multiple` command, or from a saved binary file. For large problems, use of binary files can save a lot of time.

4.4.1 Fitting Submodels Using Multiple Fit Option

Syntax:

```
Option Multiple
```

If the keyword `Multiple` is included on the `Options` line, the next Mx input file is assumed to have a special form. It will consist of a single group, ending with an `Options` line. The only commands that may be used under the `Multiple` option are `Specification`, `Pattern`, `Matrix`, `Value`, `Start`, `Equate`, `Fix`, `Free` and `Options`. Data may not be changed, matrices may not be added, and the matrix formulae specified in the `Covariance` or `Means` statement cannot be altered. At some time in the future, these restrictions may be lifted.

There is no group structure to the input stream following an `Multiple` command; all modifications to the model must refer to matrices with a number identifying the particular group in which the matrix is to be found. This changes the usual form of the `Specification, Matrix,` and `Pattern` commands to include a group specification, which must be placed directly after the keyword, *before* the letter indicating the matrix. Thus

```
Specification A
```

becomes

```
Specification 2 A
```

if `A` was specified in group 2. As an example of the use of this command, consider the simple factor model presented in Section 2.1.4. We could test the significance of the covariance of the two variables by fixing parameter #2 to zero. Obviously this could be achieved by modifying the entire input file and running it, but the following will fit both models, and fits the second starting at the solution provided by the first.

```
Simple MX example file
Data NGroups=1 NObservations=150 NInput_vars=2
CMatrix
1.2
.8 1.3
Matrices
A Full 2 1
D Diag 2 2
Model A*A' + D /
Specification A
1 0
Specification D
0 3
Options Multiple RSiduals
End
Specification 1 A
1 2
End
```

Considerable computer time can be saved using `Multiple`, since the solution of a model often has parameter estimates which are close to those estimated under nested models of the same type. In general, we recommend fitting models starting with the simplest, and working up to more complex models. When working from complex to simple, the `Drop` command (see next section) can be useful. If you have more than a single set of nested models to test, saving the general model in an Mx binary save file (see Section 4.4.3) can save considerable time and effort.

4.4.2 Dropping Parameters from Model

Syntax:

```
Drop {@value} <parnumlist>
```

where parnumlist is a list of parameter numbers as used in the `Specification` *command, and @value is an optional value to fix at*

Quite often, equality constraints between parameters lead to model specifications with the same parameter in many different matrices or several groups or both. Fixing all occurrences of the parameter with the `Fix` parameter can be time-consuming and error prone, so the `Drop` command may be used instead. By default, all parameters whose specification number matches a number in the list following the `Drop` command will be fixed to zero. For example:

```
Drop 5 8 7
Drop 11 to 20
```

would fix to zero all occurences of parameters 5, 8, 7 and 11 to 20. Note that matrix addresses cannot be used in this command. It is possible to supply an optional value to the drop command, so that for example

```
Drop @.5 2 3
```

would fix all occurences of parameters 2 and 3 to 0.5

4.4.3 Reading and Writing Binary Files

Syntax:

```
Save <filename>
Get <filename>
```

where filename is the name of the file to be saved or retrieved - the extension .mxs is recommended

When the multiple fit option is implemented (see Section 4.4.1) the entire input job specifications, data, and estimates may be stored in binary format for rapid retrieval and estimation in subsequent fitting of submodels. Note that `Save` *must* follow the entire job specification. Thus for example the following would save the results of fitting the model, together with the complete model specification, in the file `acesave.mxs`:

```
.
.    ! Mx commands for first job
.
Options Multiple
End
Save acesave.mxs ! First command in multiple fit number 1
Fix all
End
```

The `Fix all` command simply stops the optimizer from trying to improve on the current solution by fixing all the parameters. To use this binary save file in another command file, we could use the following:

```
Title - using a binary file
Get acesave.mxs
Fix A 1 1 1
End
```

By retrieving a binary file, Mx is automatically in the Multiple fit mode, so modifications can be made to the model and a further series of hypotheses tested. If `Get` is used in a separate job, a title line is required before this statement. Normally, parameter estimates after model fitting are stored, but if it is desired to save the user's starting values, it is possible to set the number of iterations to zero, use `Multiple`, and `Save` the starting values. When the binary file has been read with the `Get` command, it is necessary to remember to reset the `IT` parameter on the `Options` line.

4.4.4 Writing Matrices to Files

Syntax:

```
MXa= <filename>
```

where `a` *is the single-letter name of the matrix to be written, or one of %E %M %P %V*

Mx will write matrices to files with this command. The first line has header information, including the Group number, the matrix name, type and dimensions. The matrix elements are then written in FORTRAN format (6D13.6). This file format is fully compatible with LISREL, so matrices output by Mx can be read in as starting values for LISREL and *vice versa* (see Section 3.3.1). If the matrix name is `%E`, the expected covariance matrix will be written to the file. If the matrix name is `%M`, the expected mean matrix will be written to a file. For groups with raw maximum likelihood fit functions, %P will write a series of columns of information about the likelihood of individual pedigrees, as follows:

1. the likelihood function for that vector of observations
2. the square root of the Mahalanobis distance, $(x-\mu)'\Sigma^{-1}(x-\mu)$
3. the number of the observation in the active (i.e. post selection) dataset. Note that with selection this may not correspond to the position of the vector in the data file
4. the length of the vector (i.e. the family size if it is a pedigree with one variable per family member).

Finally, it is also possible to save a VL file,with the `%V` keyword. While this is not normally useful (since it would recreate the original data file), following the select command it can be advantageous. Subsequent reading of the selected data can be faster than reading the whole dataset and performing selection again. A second advantage is where there are problems with the data so that observations are seen to be particularly unlikely. Mx will usually print out the observation, and the expected mean and standardized difference (see Section 5.4.2)

4.4.5 Creating RAMpath Graphics Files

Syntax:

```
Draw= <filename>
```

Structural equation models may be specified very simply in terms of three matrices. The first matrix, **S**, is symmetric and specifies all the two-headed arrows between all the variables (both latent and observed) in the diagram. The second matrix, **A**, is asymmetric and specifies all the single-headed arrows between all the variables in the model. Causal paths from variable i to variable j are specified in element $\mathbf{A}_{ji}$. For example, a path from variable 1 to variable 4 would be represented in element (4,1) of this matrix. The third matrix, **F**, is used to filter the observed variables from the latent variables for the purpose of model fitting. The development and application of this approach is succinctly described in the RAMpath manual (McArdle & Boker, 1990).

Iff[1] you specify a model with these three matrices, F, A and S, then RAMpath graphics files may be written and saved to a file with the `Draw` command. This file, largely consisting of a RAMpath input_model command, may be used as input for RAMpath to draw a diagram of your model to view or to produce publication-quality output on a postscript printer. An example of this approach is given in Section 5.2.3.

[1] Iff with two f's means 'if and only if'

Chapter 5

Example script files

There are very many different ways of setting up any particular model in Mx. As with programming, there is a compromise between compactness and comprehensibility that is set by the individual user. The most compact files are often the least comprehensible; the longest ones are the most prone to typographical errors, and can be very boring. Judicious use of comments can make for a short but comprehensible input file.

5.1 Using Calculation Groups

The examples in this section do not fit models; matrix formulae are simply evaluated and the results are printed.

5.1.1 General Matrix Algebra

Suppose we wish to find the inverse of the symmetric matrix:

$$\begin{array}{ccc} 1. & & \\ .23 & 2. & \\ .34 & .45 & 3. \end{array}$$

The following input file could be created:

```
Inverting Symmetric 3 x 3 example file
Data Calc NGroups=1
Matrices
A Symm 3 3
Compute A~ /
Matrix A
1.
```

```
.23 2
.34 .45 3.
Options RSiduals
End
```

The `RSiduals` command would print the expected matrix for this group, i.e. the result of evaluating the formula in the `Compute` command, i.e. the inverse of `A`. Note that the expected matrix of a calculation group is always considered to be *full* even if it is always symmetric for a particular example. Thus adding `Mx%e=inverted.mat` to the `Options` line above would write nine elements of the inverse of `A` matrix to the file `inverted.mat`

5.1.2 Assortative Mating 'D' Matrix

Multivariate phenotypic assortative mating (Van Eerdewegh, 1982; Vogler, 1985; Carey, 1986 Phillips et al, 1988; Fulker, 1988) leads to a predicted covariance matrix between husbands' and wives' phenotypes of the form:

$$\mathbf{\Sigma} = \left[\begin{array}{r|r} \mathbf{R_H} & \mathbf{R_H D' R_W} \\ \hline \mathbf{R_W D R_H} & \mathbf{R_W} \end{array}\right]$$

Thus the matrix $\mathbf{D}$ can be obtained from $\mathbf{R_W}^{-1}\mathbf{M}\mathbf{R_H}^{-1}$, where $\mathbf{M} = \mathbf{R_W D R_H}$, the off-diagonal block of correlations between phenotypes of husbands and phenotypes of wives. The following Mx input file will calculate matrix $\mathbf{D}$.

```
Calculation of full D matrix, 3 phenotypes husband & wife
Data Calculation NGroups=1
Matrices
A Symm 3 3
B Symm 3 3
C Full 3 3
Compute A~*C*B~ /
! Covariance of wives' variables
Matrix A
1
.4 .9
.3 .5 1.1
! Covariance of husbands' variables
Matrix B
1.2
.42 1.
.3 .47 .9
!Covariance between husband and wife variables.
Matrix C
.4 .1 .2
```

```
.05 .3 .12
.22 .11 .5
Options RSiduals
End
```

The relevant part of the output file looks like this:

```
  GROUP NUMBER:           1

 CALCULATION OF FULL D MATRIX, 3 PHENOTYPES HUSBAND & WIFE

  MATRIX A
 This is a SYMMETRIC matrix of order    3 by    3
    1.0000
    0.4000    0.9000
    0.3000    0.5000    1.1000

  MATRIX B
 This is a SYMMETRIC matrix of order    3 by    3
    1.2000
    0.4200    1.0000
    0.3000    0.4700    0.9000

  MATRIX C
 This is a FULL matrix of order    3 by    3
    0.4000    0.1000    0.2000
    0.0500    0.3000    0.1200
    0.2200    0.1100    0.5000

  EXPECTED MATRIX of this CALCULATION group
    0.4302   -0.2854    0.1497
   -0.3471    0.7770   -0.5237
    0.1361   -0.4908    0.7829
```

5.1.3 Pearson-Aitken Selection Formula

This example is a little more complex. In 1934, Aitken generalized Pearson's formulae for predicting the new mean and variance of variable when selection is performed on a correlated variable. Aitken's delightful paper shows how selection on a vector of variables, X_S, leads to changes in the covariance of correlated variables X_N that are not themselves selected. If the original (pre-selection) covariance matrix of X_S is $\mathbf{A}$, and the original covariance matrix of X_N is $\mathbf{C}$, and the covariance between

X_N and X_S is $\mathbf{B}$, then the original matrix may be written

$$\left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{B}' & \mathbf{C} \end{array}\right)$$

if selection transforms $\mathbf{A}$ to $\mathbf{D}$, the whole new matrix is given by:

$$\left(\begin{array}{r|r} \mathbf{D} & \mathbf{D}\mathbf{A}^{-1}\mathbf{B} \\ \hline \mathbf{B}'\mathbf{A}^{-1}\mathbf{D} & \mathbf{C}-\mathbf{B}'(\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1})\mathbf{B} \end{array}\right)$$

Likewise, if the original means are $(\mathbf{x_s} : \mathbf{x_n})'$ and selection modifies $\mathbf{x_s}$ to $\tilde{\mathbf{x_s}}$, the vector of means after selection is given by:

$$\mathbf{x_n} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{x_s} - \tilde{\mathbf{x_s}})$$

```
Pearson Aitken Selection formulae
! Idea is to give original means and covariances,
! and get new ones
!
! We have original matrix  A  B
!                          B' C
!
! Which we following transformation of A becomes
!                          D | DA~B
!                       B'A~D| C-B'(A~-A~DA~)B
!
Data Calc NGroups=2
Matrices
A Symm 2 2
B Full 2 2
C Symm 2 2
D Symm 2 2
Compute
     D | D*A~*B_
B'*A~*D| C-B'*(A~-A~*D*A~)*B /
!
! Note that the underscore above is the UNDER operator,
!  not A with a horizontal bar over it!
!
Matrix A
1.
.7 1.
Matrix B
.6 .42
.42 .6
```

```
Matrix C
1.
.352 1.
Matrix D
1.
.4 1.
Options RSiduals
End
!
! The means are done in group 2.  We have original means of
! the non-selected variables, n, which are transformed to
! n + A~Bs where s is the deviation of the means of the
! selected variables from their original values.
!
Second group to calculate new means
Data Calc
Matrices
A Symm 2 2 =A1
B Symm 2 2 =B1
S Full 2 1
N Full 2 1
Compute N + A~*B*S /
Matrix N
3
4
Matrix S
2 1
Options RSiduals
End
The new covariance matrix and mean vector are printed as the
expected matrices of the calculation groups.
```

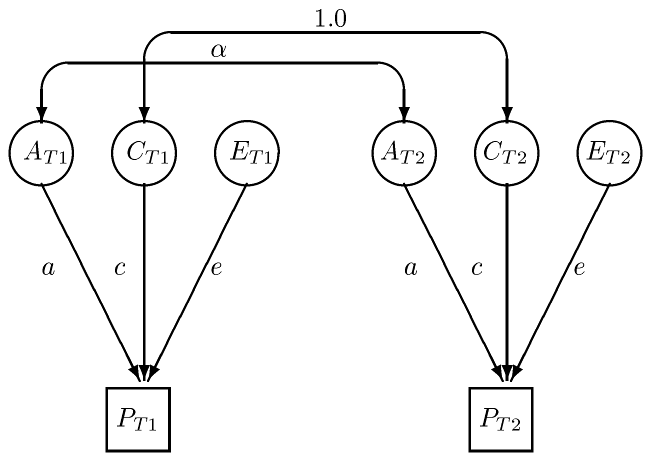
5.2 Model Fitting with Genetically Informative Data

The examples in this section demonstrate elementary use of Mx to fit models to data in the form of variance-covariance matrices.

5.2.1 ACE Genetic Model for Twin Data

If data are collected from identical or monozygotic (MZ) twins and from fraternal or dizygotic (DZ) twins, it is possible to estimate parameters of a model that includes

Figure 5.1: The parameter α is fixed at 1 for MZ twins and at .5 for DZ twins. All latent variables have a variance of 1.0



the effects of additive genes (A), shared or family environment (C) and random or unique environment (E). This model is shown in Figure 5.1 as a path diagram. A LISREL input file for the model was provided by Heath *et al* (1989), which could be adapted very easily using `Equate` statements in Mx. However, the MX script illustrates use of the `Specification` command to modify it slightly. Heath *et al* allow for phenotypic interaction, but this is left for a later example (Section 5.2.5)

```
A C E model fitted to the Heath (1989) data
!     on alcohol consumption.
! Group 1.... Monozygotic twin pairs
    ! Anything after a '!' is a comment
Data NGroups=2 NInput_variables=2 NObservations=171
Labels Alc_t1 Alc_t2
CMatrix
1.28
0.766 1.194
Matrices
A Full 2 6
B Symmetric 6 6
```

```
Covariance_model A*B*A' /
Specification A
1 2 3 0 0 0
0 0 0 1 2 3
Labels Col A
At1 Ct1 Et1 At2 Ct2 Et2
Labels Row A
Pt1 Pt2
Matrix B
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 0 0 0 1
Labels Row B
At1 Ct1 Et1 At2 Ct2 Et2
Labels Col B
At1 Ct1 Et1 At2 Ct2 Et2
Options RSiduals
End

Group 2.... Dizygotic twin pairs
Data NInput_variables=2 NObservations=101
CMatrix
1.077
0.463 0.962
Labels Alc_t1 Alc_t2
Matrices
A Full 2 6 =A1
B Symm 6 6
Covariance_model A*B*A' /
Matrix B
1
0 1
0 0 1
.5 0 0 1
0 1 0 0 1
0 0 0 0 0 1
Labels Row B
At1 Ct1 Et1 At2 Ct2 Et2
Labels Col B
At1 Ct1 Et1 At2 Ct2 Et2
```

```
Start .6 All
Options RSiduals
End
```

5.2.2 Power Calculation for the Classical Twin Study

Our next example uses the same model as in the preceding Section, but in this case we calculate the power of the classical twin study to detect the effects of common environmental variation. The particular case we wish to examine is where the true population variance comprises 20% additive genetic, 30% shared environment, and 50% random environment variance, but we fit a model *without* shared environment variation to simulated MZ and DZ covariance matrices. This example is discussed in some detail in Neale & Cardon (1992); here we reproduce their results obtained with rather more effort using LISREL and some statistical tables.

There are two stages to the power calculation. First, fixed values of the parameters a, c and e are given, and a two-group Mx script simply calculates the expected covariances under the model, and saves them to two files, mzsim.cov and dzsim.cov. The next problem (preferably in the same input file, though this isn't essential) fits a model of additive genetic and random environmental variance only. The complete input file looks like this:

```
A C E model for power calculation   Part 1: simulate the data
Data Calc NGroups=2       !Note that it is just a calculation group
Matrices                  !Define matrices for the model
A Full 1 3
B Symmetric 6 6
I Identity 2 2
Compute (I@A)*B*(I@A)' /  !Formula for expected MZ cov
Labels Col A
At1 Ct1 Et1
Matrix A
.5477 .4472 .7071     ! To simulate:
!                       30% additive genetic   (.5477^2=.3)
!                       20% common environment (.4472^2=.2)
!                       50% random environment (.7071^2=.5)
!
!  Note that I@A will generate a matrix of the form
!  .6  0 .8  0  0  0
!   0  0  0 .6  0 .8
!  because I is an identity matrix and @ is the kronecker product.
!  as required for the model.  We reduce the chance of errors
!  when modifying the script by adopting this formulation here.
!
Matrix B
```

```
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 0 0 0 1
! Some labels to make the output a bit prettier
Labels Row B
At1 Ct1 Et1 At2 Ct2 Et2
Labels Col B
At1 Ct1 Et1 At2 Ct2 Et2
! This is the crucial bit where we save the file
Options mx%e=mzsim.cov
Options NO_output
End

Group 2.... Dizygotic twin pairs
Data Calc
Matrices
A Full 1 3 =A1
B Symm 6 6
I Iden 2 2
Compute (I@A)*B*(I@A)' /
Matrix B
1
0 1
0 0 1
.5 0 0 1
0 1 0 0 1
0 0 0 0 0 1
! And next we save the DZ file
Options mx%e=dzsim.cov
Options NO_output  ! No output for this group as nothing new, really
End

! The simulation is complete.  Now the next, stacked, problem
! will fit the wrong model to the simulated data

A C E model for power calculation
Data NGroups=2 NInput_variables=2 NObservations=1000
Labels Twin_1 Twin_2
CMatrix Full File=mzsim.cov
Matrices
```

```
A Full 1 3
B Symmetric 6 6
I Identity 2 2
Covariance_model (I@A)*B*(I@A)' /
Specification A
1 0 3            ! False model with no genetic effects
Labels Col A
At Ct Et
Labels Row A
Pt1 Pt2
Matrix B
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 0 0 0 1
Labels Row B
At1 Ct1 Et1 At2 Ct2 Et2
Labels Col B
At1 Ct1 Et1 At2 Ct2 Et2
Options RSiduals
End

Group 2.... Dizygotic twin pairs
Data NInput_variables=2 NObservations=1000
CMatrix Full File=dzsim.cov
Labels Twin_1 Twin_2
Matrices
A Full 1 3 =A1
B Symm 6 6
I Iden 2 2
Covariance_model (I@A)*B*(I@A)' /
Matrix B
1
0 1
0 0 1
.5 0 0 1
0 1 0 0 1
0 0 0 0 0 1
Labels row B
At1 Ct1 Et1 At2 Ct2 Et2
Labels col B
```

```
At1 Ct1 Et1 At2 Ct2 Et2
Start .7 ALL     ! Note different starting values from simulation
Options RSiduals Power= .05,1  ! For .05 level of sig. and 1 d.f.
End
```

The relevant part of the output is at the end, where we see that for the specified sample sizes of 1000 pairs each of MZ and DZ twins, the χ^2 goodness-of-fit is 11.35, as found by Neale & Cardon:

```
 Chi-squared fit of model =    11.349
  Degrees of freedom =           4
  Probability =      0.023
 Akaike's Information Criterion =     3.349

 Power of this test, at the 0.0500 significance level
 with   1. df is 0.920572
 Based on your combined observed sample size of        2000.
 The following sample sizes would be required to reject
 the hypothesis:

  Power         Total N
  .25             292.
  .50             677.
  .75            1223.
  .80            1383.
  .90            1852.
  .95            2290.
  .99            3238.
```

Because we requested that this statistic be treated as a χ^2 for 1 degree of freedom for the purposes of calculating power at the .05 level of significance (`POwer=.05, 1`), the output gives the power of the test, followed by the sample sizes that would be required to obtain certain commonly used levels of power.

5.2.3 RAM Specification of Model for Twin Data: Graphic Output

Here is a two group example, a phenotypic interaction PACE model (see Section 5.2.5 for more details), specified using the three matrix approach of McArdle & Boker (1990). Details of this method, and the syntax of the `Draw` command can be found in Section 4.4.5.

The title for the diagram is taken from the title of the group in the Mx input file, and the labels for the variable are taken from labels of the columns of the **S** matrix. The draw commands in this file produce two files, mx.ram and dz.ram. I don't like the way RAMpath draws interaction between the phenotypes, but there is a certain irrefutable logic in having causal arrows always going out the bottom of a variable and in the top. You can easily edit the RAMpath file to get rid of the @ signs if you like. Note that specifying models à la RAMpath is didactically very clear but computationally very inefficient, since the inverse of the maximally dimensioned **A** matrix is required. Using calc groups can reduce this burden if CPU time is at a premium but diagrams are desired.

```
Phenotypic interaction PACE model, Heath 1989 data
! Demonstration of RAM specification and output
Data NGroups=2 NInput_vars=2 NObservations=171
Cmatrix Symm
1.280
0.766 1.194
Matrices
s Symm 8 8
i Iden 8 8
a Full 8 8
f ZIden 2 8
Covariances F*(I-A)~*S*(I-A)~'*F'/
Matrix s
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
Label row a
a1 c1 e1 a2 c2 e2 p1 p2
Label col a
a1 c1 e1 a2 c2 e2 p1 p2
Label row s
a1 c1 e1 a2 c2 e2 p1 p2
Label col s
a1 c1 e1 a2 c2 e2 p1 p2
Specify a
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
```

```
1 2 3 0 0 0 0 4  ! This is where the parameters are
0 0 0 1 2 3 4 0
Start .6 a 7 1 a 7 3
Options RSiduals Draw=mz.ram
End

Group 2.... DZ twins
Data NInput_vars=2 NObservations=101
Cmatrix Symm
1.077
0.463 0.962
Matrices
s SY 8 8
i Iden 8 8
a Full 8 8 = a1
f ZIden 2 8
Covariances F*(I-A)~*S*(I-A)~'*F'/
Matrix s
1
0 1
0 0 1
.5 0 0 1
0 1 0 0 1
0 0 0 0 0 1
0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
Label row s
a1 c1 e1 a2 c2 e2 p1 p2
Label col s
a1 c1 e1 a2 c2 e2 p1 p2
Bound -.99 .99 3
Bound 0 5 1 2
Options RSiduals Draw=dz.ram
End
```

5.2.4 Cholesky Decomposition for Multivariate Twin Data

Any positive definite matrix may be transformed to a product of a lower triangular matrix and its transpose, i.e.

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}'$$

This *Cholesky* decomposition is unique for a given positive definite matrix $\mathbf{\Sigma}$. It has a simple graphic representation as a path diagram (Figure 5.2) where the first latent factor loads on all variables, the second on all variables except the first, the

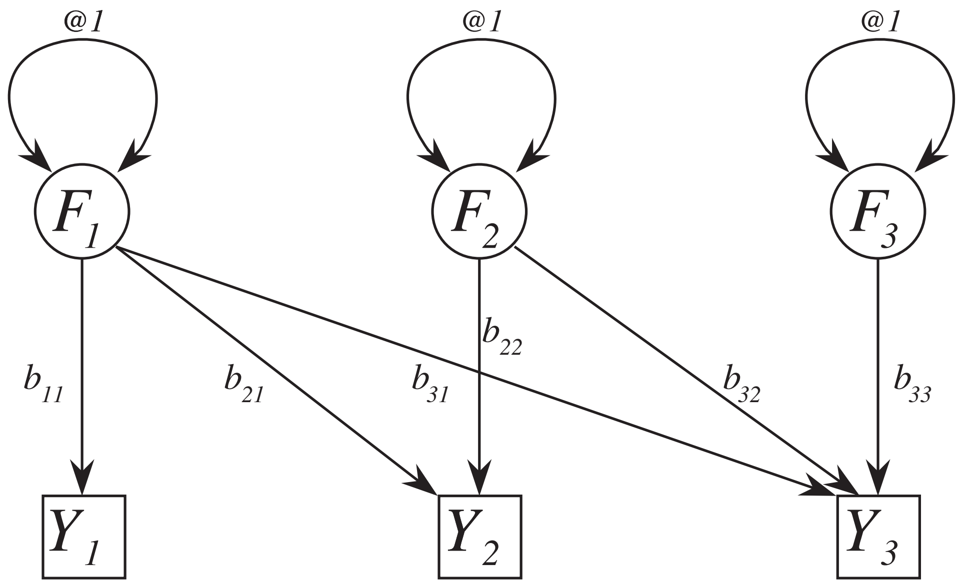


Figure 5.2: Cholesky or triangular factor model for three variables Y1 Y2 and Y3

third on all variables except the first two, and so on.

In multivariate genetic analysis, a Cholesky decomposition of separate genetic and environmental covariance matrices is possible. Thus the additive genetic, shared environment and random environment factors in the simple ACE model (Figure 5.1) has a multivariate counterpart where the phenotypes P_1 and P_2 are replaced by vectors of observed phenotypes, and the latent variables A C and E are replaced by matrices of factors. The path coefficients a c and e are replaced by triangular matrices of factor loadings according to the Cholesky decomposition. There are a number of ways of approaching such a model in Mx. The standard way, to set it up as one would a LISREL model, would prove very awkward to alter to allow for increased or decreased numbers of variables. Here we use an input file that separates the genetic, shared and random environmental factors into different calculation groups that generate genetic, shared and random environment covariance matrices. It is then a simple matter to form the expected covariance matrices for twin data as partitioned matrices containing linear combinations of these matrices. It is then simplicity itself to fix a parameter to zero, as only one character has to be changed from a 1 to a zero.

The example includes data from individuals without cotwins, as well as MZ and DZ twin pairs. Estimates from a model such as this depend on the size of the observed variances, and can be difficult to interpret. Estimates of the proportion of variance and covariance due to each source can be obtained using the eldiv operator

(%) (groups 7-9).

```
!
! Trivariate Cholesky 'Independent Pathways' model.
!
! Additive Genetic factors are group 1 matrix H
! Random Environ. factors are group 2 matrix E
! Common Environ. factors are group 3 matrix C
!
! Data are Extraversion, Neuroticism and CESD Depression
!
This group JUST calculates, no fitting is done.
! Genetic factors
Data Calculation NGroups=9
Matrices
H Lower 3 3 Free
Compute H*H'  /
End

Title 2nd Group also for calculation
! Specific Envt.
Data Calculation
Matrices
E Lower 3 3 Free
Compute E*E'  /
End

Title 3rd Group also for calculation
! Shared Envt.
Data Calculation
Matrices
C Lower 3 3 Free
Compute C*C'  /
End

!
! Now get to the actual data, and use the results of
!  groups 1, 2 and 3 (%E1 %E2 & %E3)
!
Unmatched twins
Data NInput_vars=3 NObservations=449
Cmatrix Symm
 .102216
-.021648 .0827805
```

```
-.005792 .0177622 .0138577
Matrices
G Symm 3 3 = %E1
E Symm 3 3 = %E2
C Symm 3 3 = %E3
Covariances (G + E + C) /
End

MZ twins with cotwins  Extraversion Neuroticism and Depression
Data NInput_vars=6 NObservations=456
Cmatrix
.0887349
-.018651  0.08036
-.001031 .0185154 .0139989
.0366848 -.008858 -.002873 0.097803
.0013777 .0291787 .0110424 -.014907 .0818947
-0.00122 .0103675 .0050655 -.005258 .0172896 0.013687
Matrices
G Symm 3 3 = %E1
E Symm 3 3 = %E2
C Symm 3 3 = %E3
Covariances
             (G + E + C| G + C _
                  G + C| G + E + C )
/
Options RSiduals
End

!
!
DZ twins with cotwins  Extraversion Neuroticism and Depression
Data NInput_vars=6 NObservations=357
CMatrix
0.100808
-0.02249 .0785764
-.003347 .0173292 .0150047
-0.00471 .0025537 -.002045 0.102181
.0032861 .0081708 .0078504 -.023836 .0808707
-.001514 .0053548 .0036505 -.005488 .0164667 .0143083
Matrices
G Symm 3 3 = %E1
E Symm 3 3 = %E2
C Symm 3 3 = %E3
```

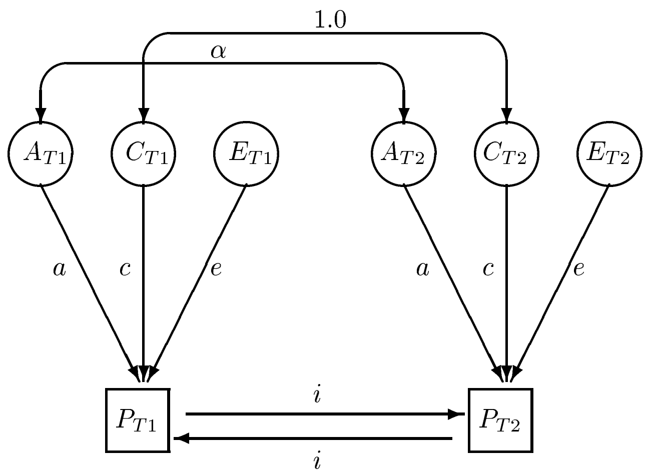
```
h Diag 1 1
Covariances
!
! Covariance matrix bit.  Note H as diag matrix of .5's
! that models DZ genet cov.  By using the Kron operator,
! every element of G is multiplied by .5, in accordance with
! genetic theory that DZ twins share only half their genes.
!
      (G + E + C| H @ G + C _
       H @ G + C| G + E + C )
/
Matrix Half
.5
Start .5 All
Boundary -1 1 All
Options RSiduals IT=200
End

Calculation of standardized solution - Additive Genetic
Data Calc
Matrices
A Symm 3 3 =%E1
P Symm 3 3 =%E4
Compute A%P /
Options RSiduals
End

Calculation of standardized solution - Shared Environment
Data Calc
Matrices
C Symm 3 3 =%E3
P Symm 3 3 =%E4
Compute C%P /
Options RSiduals
End

Calculation of standardized solution - Random Environment
Data Calc
Matrices
E Symm 3 3 =%E2
P Symm 3 3 =%E4
Compute E%P /
Options RSiduals
```

Figure 5.3: Path for additive genetic (A), shared environment (C) and specific environment (E) effects on phenotypes (P) of pairs of twins ($T1$ and $T2$). Path i models direct phenotypic effects of a twin on his or her cotwin. The parameter α is fixed at 1 for MZ twins and at .5 for DZ twins.




```
End
```

5.2.5 PACE Model: Reciprocal Interaction between Twins

Figure 5.2.5 shows a path diagram similar to the ACE model for twin data shown in Figure 5.1; There is now a path (i) from the phenotype of a twin to that of his or her cotwin. This is reciprocal interaction between dependent variables. It can easily be shown (see Jöreskog and Sörbom, 1989; also Section 1.2) that a matrix representation of this process is to use the formulation $(\mathbf{I}-\mathbf{B})^{-1}$, where $\mathbf{B}$ is a full matrix whose element b_{jk} represents the causal effects of variable k on variable j. In this case, the parameter i has been bounded to lie between -1 and 1, though this is not necessary.

```
Phenotypic interaction model, fit to Heath 1989 data.
! Female MZ Twins
Data NGroups=2 NInput_vars=2 NObservations=171
CMatrix
```

```
1.28
0.766 1.194
Labels Alc_t1 Alc_t2
Matrices
G Full 2 6
P Symm 6 6
B Full 2 2
I Identity 2 2
Covariance_model (I-B)~*(G*P*G')*(I-B)~'/
Labels row G
B4int_t1 B4int_t2
Labels col G
At1 Ct1 Et1 At2 Ct2 Et2
Specification G
1 2 3 0 0 0
0 0 0 1 2 3
Labels row P
At1 Ct1 Et1 At2 Ct2 Et2
Labels col P
At1 Ct1 Et1 At2 Ct2 Et2
Matrix P
1
0 1
0 0 1
1 0 0 1
0 1 0 0 1
0 0 0 0 0 1
Labels col B
B4int_t1 B4int_t2
Labels row B
Alc_t1 Alc_t2
Specification B
0 4
4 0
Options RSiduals
End

Phenotypic interaction model, Heath alcohol data
! Female DZ twins
Data NInput_vars=2 NObservations=101
Cmatrix Symm
1.077
0.463 0.962
```

```
Matrices
G Full 2 6 =G1
P Symm 6 6
B Full 1 2 =B1
I Identity 2 2
Covariance_model (I-B)~*(G*P*G')*(I-B)~'/
Labels row P
At1 Ct1 Et1 At2 Ct2 Et2
Labels col P
At1 Ct1 Et1 At2 Ct2 Et2
Matrix P
1
0 1
0 0 1
.5 0 0 1
0 1 0 0 1
0 0 0 0 0 1
Start .6 g 1 1 - g 2 6
Boundaries -.99 .99 4
Boundaries 0 5 1 2 3
Options RSiduals
End
```

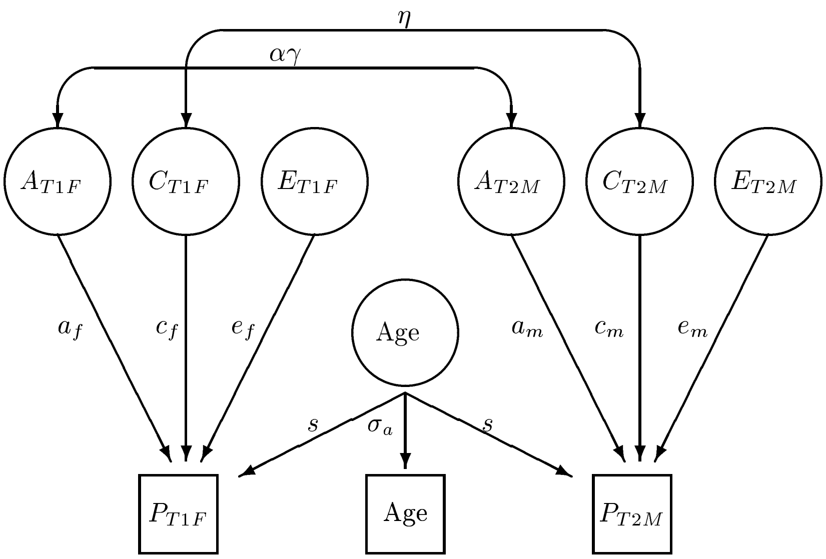
5.2.6 Scalar and Non-scalar Sex Limitation; Age Regression

The model and data in this section were taken from Neale & Martin (1989) who fitted a model of scalar sex-limitation to data from 5 groups of twins who reported subjective impressions of drunkenness following a challenge dose of alcohol. The model here involves additive genetic, shared and random environment components (A C and E; see example in Section 5.2, and Figure 5.1) but allows these to differ in their effects on the phenotypes of males and females. In addition, the regression of the phenotypes on Age is modelled, so that there are parameters for the variance of age (σ_a^2) and for the regressions (s). A path diagram of the model is shown in Figure 5.4.

In the fifth group, two commented out lines (beginning with !) indicate how the model could be modified to test the fit of a model that allowed for different genes in the two sexes. Two features are of particular note in this example. First is the use of the group-specific matrix constraint (e.g. `M Full3 7 =M2`) which makes for a brief input file. Second is the use of boundary constraints to prevent the estimation of parameters of opposite sign in the two sexes. Considerable effort is required to implement this simple constraint in LISREL (see Neale & Martin, 1989).

```
Age correction Sex limitation model MZ FEMALES
```

Figure 5.4: Path model for sex-limited additive genetic (A), shared environment (C) and specific environment (E) effects on phenotypes (P) of pairs of twins ($T1$ and $T2$) The parameter α is fixed at 1 for MZ twins and at .5 for DZ twins. Either γ or η may be estimated with data from twins, but not both.



```
Data NGroups=5 NInput_vars=7 NObservations=43
CMatrix File=DRUNK.COV
Matrices
F Full 3 7
P Symm 7 7
Covariances F*P*F' /
Specification F
1 2 3 0 0 0 4
0 0 0 3 2 1 4
0 0 0 0 0 0 9
Matrix P
1
0 1
0 0 1
0 0 0 1
0 1 0 0 1
1 0 0 0 0 1
0 0 0 0 0 0 1
Options RSiduals
End

Age correction Sex limitation model MZ MALES
Data NInput_vars=7 NObservations=42
CMatrix File=DRUNK.COV
Matrices
P Symm 7 7 =P1
M Full 3 7
Covariances M*P*M' /
Specification M
5 6 7 0 0 0 4
0 0 0 7 6 5 4
0 0 0 0 0 0 9
Options RSiduals
End

Age correction Sex limitation model DZ FEMALES
Data NInput_vars=7 NObservations=44
CMatrix File=DRUNK.COV
Matrices
F Full 3 7 =F1
P Symm 7 7
Covariances F*P*F' /
Matrix P
```

```
1
0 1
0 0 1
0 0 0 1
0 1 0 0 1
.5 0 0 0 0 1
0 0 0 0 0 0 1
Options RSiduals
End

Age correction Sex limitation model DZ MALES
Data NInput_vars=7 NObservations=38
CMatrix File=DRUNK.COV
Matrices
M Full 3 7 =M2
P Symm 7 7 =P3
Covariances M*P*M' /
Options RSiduals
End

Age correction Sex limitation model DZ OPP SEX
Data NInput_vars=7 NObservations=39
CMatrix File=DRUNK.COV
Matrices
P Symm 7 7
O Full 3 7
Covariances O*P*O' /
Specification O
1 2 3 0 0 0 4
0 0 0 7 6 5 4
0 0 0 0 0 0 9
Matrix P
1
0 1
0 0 1
0 0 0 1
0 1 0 0 1
.5 0 0 0 0 1
0 0 0 0 0 0 1
!Free P 6 1
!Boundary 0 0.5 P 6 1
Boundary 0 5 1 2 3 5 6 7
Start .6 All
```

```
Options RSiduals
End
```

5.2.7 Multivariate Assortative Mating: Modelling D

In this section we return to the transformation of the data described in Section 5.1.2. The question is now, how do we fit a model with parameters in the **D** matrix, so that we can explore the significance of marital assortment both within and between variables. For example, is there selection of wealthy men by attractive women, or is it just that attractiveness and wealth are correlated, and partners choose each other on the grounds of wealth alone? Neale & McArdle (1990) published a transformation of the matrix equation in Section 5.1.2 which allowed the fitting of LISREL and COSAN models to marital data. While LISREL is widely available, the implementation of the model is not straightforward, involving phantom latent variables (Rindskopf, 1984). The model is very easy to implement in Mx, as is shown below. We have already shown how estimates of parameters in the **D** matrix may be obtained directly (Section 5.1.2); here we show how to test specific hypotheses about direct and indirect homogamy. Phillips et al. (1988) reports data on general intelligence, educational level, extraversion, anxiety, tough-mindedness and independence in both husbands and wives. The first diagonal element of **D** therefore represents direct homogamy for intelligence; by fixing this parameter to zero we test the statistical significance of the process.

```
Assortative mating: Phillips data, test that d 1 1 is zero
Data NGroups=1 NInput_vars=12 NObservations=334
CMatrix File=asmat.cov
Matrices
h Symm 6 6 fr
w Symm 6 6 fr
d Full 6 6 fr
M2       h | h*d'*w_
     w*d*h | w        /
Start 1. h 1 1 h 2 2 h 3 3 h 4 4 h 5 5 h 6 6
Start 1. w 1 1 w 2 2 w 3 3 w 4 4 w 5 5 w 6 6
Fix  d 1 1
Options RSiduals
End
```

5.3 Fitting Models with Non-linear Constraints

5.3.1 Principal Components

Any symmetric matrix **C** may be decomposed to a product $\mathbf{ABA}'$ where **B** is a real diagonal matrix and **A** is a real orthogonal matrix, i.e. $\mathbf{AA}' = \mathbf{I}$. The elements