

Using MxScript.exe to create Mx Scripts for Structural Equation Models with Measured Genotypes

MxScript is a Pascal program for DOS/Windows to facilitate the specification of structural equation models with measured genotypes (see Van den Oord EJCG and Snieder H. Including quantitative trait loci....) in Mx (Neale, Boker, Xie, Maes, 1999). To start the program type/select: MxScript [input file] <ENTER/RUN>. The [input file] is optional and requires a DOS name (maximum of 8 characters, a dot, and then an extension of at most 3 characters). If there is no input file the program will ask questions about the data structure and parameter matrices you want to use. Depending on previous answers, certain questions may be skipped. For instance, when the measurement part is not used no further questions will be asked about the matrices in this part of the model and if there is only one subject per family no questions will be asked about family variables. If an input file is specified the program will read the answers directly from that file. An example is shown in Appendix 1. The answers can be changed to specify another model but the sequence of the questions should remain the same. A lower case y means “yes”. Blanks (or any other characters) are read as “no” or 0 when a number is expected. The same routing is used as in the case when no input file is specified. For instance, when the measurement part is not used, all answers regarding the matrices in this part of the model will be ignored.

MxScript generates a script that consists of four groups when there are no family variables and five groups when there are family variables. The groups plus their function and further changes that are required are:

Group number	Function	Free parameters	Starting values
1	General part subject variables	yes	yes
2	Genetic effects	yes	
3	Stratification effects subject variables	Yes	
4	Family variables	Yes	yes
4 or 5	Fit model to data		

Thus, no changes are required in the final group. Changes in the other groups merely consist of freeing the parameters/matrix elements (groups 1-5). In groups 1 and 4 starting values need to be assigned to free matrix elements, and values to some fixed matrix elements to scale the latent variables. To illustrate this we will modify the script shown Appendix 2 that was made with MxScript to run the model from figure 1b while controlling for stratification. A full explanation of all commands in the script can be found in the Mx manual and we will confine ourselves to a short description. The parts that were changed are printed bold.

Group 1: General part subject variables

Calculation ng=4

Matrices

```
a Full 2 1 Free ! Means subject variables; Alpha
B Full 2 2 Fixed ! Causal effects subject variables on each other; Beta
P Symm 2 2 Free ! Residual (co)variances subject variables; Psi
C Symm 2 2 Free ! Covariances between family members; C
```

End Matrices

Free B 2 1

Matrix a ! specify (starting) values for (free) parameters

0.5

0.5

Matrix B ! (starting) values: Main diagonal B should ALWAYS be zero

0 0.0

0.3 0

Matrix P ! (starting) values

```

1.5
0.1 1.5
Matrix C          ! starting values
0.2
0.1 0.1
Output

```

The first group pertains to the general matrices that are not subscripted in table 1 of the article and represent the overall model in all genotype groups and population strata. The first two lines indicate the title and group type. The part between Matrices-End Matrices shows the matrices. These matrices are always declared fixed meaning that no parameters are estimated and that all elements are zero. Exceptions are matrix C containing the covariances between subjects from the same family, and vector \mathbf{v} that contains the means of the family variables. There are two ways to free/estimate matrix elements. First, all elements in a matrix can be estimated by declaring the matrix **Free** in the Matrices-End Matrices section. When only some matrix elements need to be estimated this can be done in the part after <End Matrices> using the statement (p. 73 Mx Manual): **Free** <matrix name> <row number> <column number>. The default starting values for free parameters and the default values for fixed parameters are zero. This can be changed the replacing the zeroes by other numbers in the matrices that follow the statement Matrix <matrix name>.

The second group specifies the genetic effect on the matrices.

```

Group 2: Genetic effects
Calculation
Matrices
A Full 2 2 Fixed ! Means; Alpha = 2 horizontally stacked 2 x 1 subvectors
End Matrices
Free A 1 1 A 1 2
Output

```

The complete matrices Phi and Theta should NEVER be declared Free. These are stacked symmetric matrices and by freeing the full matrix the submatrices would not be symmetric anymore. Note that here the Free statement lists two elements that are free, and that there are no starting values because the default value of zero is acceptable for the genetic parameters that represent deviations from the general model.

The third groups pertains to the stratification effects:

```

Group 3: Stratification effects on subject variables
Calculation
Matrices
A Full 2 14 Fixed ! Means; Alpha = 20 horizontally stacked 2 x 1 subvectors
End Matrices
Free A 1 1 to A 1 14
Output

```

Again Phi, Theta, and C are stacked symmetric matrices that should never be declared fully Free. Here the Free statement for matrix A lists a range of parameters.

No changes have to be made to the fourth group.

```

Group 4: Fit model to data
Data NInput=22
Rectangular File=example1.dat
Labels
F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14
Y1_1 Y2_1 G1_1 G2_1
Y1_2 Y2_2 G1_2 G2_2

```

....

Mx expects 1 data record for each family that should have the sequence:

- family or x variables
- dummy variables for family types
- y variables for subject 1
- genetic dummy variables subject 1
- y variables for subject 2
- genetic dummy variables subject 2
-
- y variables for final subject in the family
- genetic dummy variables for final subject in the family

In the example the data are in the file example1.dat. Because no family variables were used in this example, they are not expected in the data file. Missing value for y variables have to be indicated with a dot (.). Mx doesn't allow missing genetic and family type dummy variables. In families where the number of subjects is smaller than the number of subjects in the largest families, "missing" subjects can be included by coding the dummy variables zero and the y variables with a dot.

To run the script type/select: Mx [script name] <ENTER/RUN>. Mx will write the output to a file with extension [.mxo]. It may be preferable to read the parameter estimates from groups 1-3 (or groups 1-4 if there are family variables). In these groups the Mx matrix name is the first letter of the Greek matrix name reported in Table 1 of the article. In the data group the matrices are given a letter in alphabetic order. This makes it more difficult to establish which parameter estimates are reported.

If you encounter bugs please send me an email (E.vandenOord@fss.uu.nl) and I'll fix it.

Appendix 1: Example input file.

```
Answers(use lower case y for yes)    Questions
n      Are there multiple indicators for one of the subject (=y) variables?
2      Number of (latent) subject variables
      Total number of indicators for latent subject variables
2      Number of subjects in largest family
n      Are there variables (=x) that are identical for all subjects in the family?
      Are there multiple indicators for one of the family variables?
      Number of (latent) family variables
      Total number of indicators for latent family variables
2      Number of dummy variables for genotypes
y      Do you want to specify stratification effects?
14     Number of dummy variables for family types
y      Do the subject variables have causal effects on each other? BETA
      Do you want the mean of the family variable(s) to have an effect on the subject variables? TAU
y      GENETIC: means of subject variables? ALPHA
      GENETIC: causal relations between subject variables? BETA
      GENETIC: causal effect of family on subject variables? GAMMA
      GENETIC: residual (co)variances of subject variables? PSI
      GENETIC: intercepts of subject indicator variables? NU
      GENETIC: factor loadings subject indicator variables? LAMBDA
      GENETIC: measurement errors subject indicator variables? THETA
      STRATIFICATION: means of subject variables? ALPHA
      STRATIFICATION: causal relations between subject variables? BETA
      STRATIFICATION: causal effect of family on subject variables? GAMMA
      STRATIFICATION: residual (co)variances of subject variables? PSI
      STRATIFICATION: intercepts of subject indicator variables? NU
      STRATIFICATION: factor loadings subject indicator variables? LAMBDA
      STRATIFICATION: measurement errors subject indicator variables? THETA
Name of file with your data? (DOS name)
Example1.dat
Name of the file for the Mx script? (DOS name)
Example1.mx
```

Appendix 2: Example Mx script.

```
Group 1: General part subject variables
Calculation ng=4
Matrices
  a Full  2  1 Fixed  ! Means subject variables; Alpha
  B Full  2  2 Fixed  ! Causal effects subject variables on each other; Beta
  P Symm  2  2 Fixed  ! Residual (co)variances subject variables; Psi
  C Symm  2  2 Free   ! Covariances between family members; C
End Matrices
Matrix a                ! specify (starting) values for (free) parameters
  0.0
  0.0
Matrix B                ! (starting) values: Main diagonal B should ALWAYS
remain zero
  0  0.0
  0.0  0
Matrix P                ! (starting) values
  0.0
  0.0  0.0
Matrix C                ! starting values
  0.0
  0.0  0.0
Output

Group 2: Genetic effects
Calculation
Matrices
  A Full  2  2 Fixed  ! Means; Alpha = 2 horizontally stacked 2 x 1
subvectors
End Matrices
Output

Group 3: Stratification effects on subject variables
Calculation
Matrices
  A Full  2  14 Fixed ! Means; Alpha = 14 horizontally stacked 2 x 1
subvectors
End Matrices
Output

Group 4: Fit model to data
Data NInput=22
Rectangular File=example1.dat
Labels
  F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14
  Y1_1 Y2_1 G1_1 G2_1
  Y1_2 Y2_2 G1_2 G2_2
Definition
  F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14
  G1_1 G2_1
  G1_2 G2_2 /
Matrices
  a Full  2  1 =a1
  b Full  2  2 =A2
  c Full  2  20 =A3
  d Full  2  2 =B1
  e Symm  2  2 =P1
  f Symm  2  2 =C1
  g Iden  2  2   ! identity matrix subject variables
  h Full  14  1   ! family type dummy variables
```

```

i Full 2 1      ! genetic dummy variables subject 1
j Full 2 1      ! genetic dummy variables subject 2
End Matrices
Means
(g-(d))~*(a+b*i+c*h)  _      ! mean subject 1
(g-(d))~*(a+b*j+c*h) /      ! mean subject 2
Covariances
(g-(d))~*(e)*((g-(d))~)' | (f)  _ ! variance subject 1
(f) | (g-(d))~*(e)*((g-(d))~)' / ! variance subject 2
Specify h
F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14
Specify i
G1_1 G2_1
Specify j
G1_2 G2_2
Options _
End

```