Alternate Parameterization for Scalar and Non-scalar Sex-limitation Models in Mx

Sarah E. Medland[1,2]

[1] Queensland Institute of Medical Research, Australia
[2] Cognitive Psychophysiology Laboratory, University of Queensland, Australia

The purpose of this article is to present alternative parameterizations of scalar and non-scalar sex-limitation models in the Mx matrix algebra program (Neale et al., 2002). These models are designed for use with extended pedigrees and take advantage of the dynamic treatment of covariates within Mx. Example scripts are provided.

Mx (Neale et al., 2002) is a matrix algebra program specifically designed for use with twin and family data. One of the key features of the program is the flexibility of model specification, which allows researchers to test hypotheses specific to the trait or sample under investigation. The aim of the present note is to present an alternate parameterization of scalar and non-scalar sex limitation models that is especially suited to the extended twin-family design in which data is available for sibships of more than two siblings.

Sex-limitation, or sex differences in gene expression, may be scalar, non-scalar, or general in nature (Eaves et al., 1978; Neale & Cardon, 1992). Non-scalar sex-limitation models can be used to detect differences in the magnitude of genetic and environmental influences between males and females, and to determine whether the same set of genetic or environmental factors influence a trait in males and females. The scalar effects model tests the hypothesis that while the same genes are expressed in males and females the magnitude of these effects in males and females differs across the loci involved by some constant or scalar. The implementation of these models will be discussed in the following sections. Although the scalar model is a sub-model of the more general non-scalar model, the scalar model will be discussed first.

Scalar Sex-limitation of Total Variance

In analyses of data collected using the classical twin design, specification of total variance differences across sex (but equal heritability) is fairly elementary (Neale & Cardon, 1992). As shown below for a pair of dizygotic (DZ) twins, to allow for a scalar sex-limitation the variance of one sex (typically the male twins) is pre- and post-multiplied by a diagonal matrix (K) in which the scalar forms the diagonal elements.

$$K \times \begin{pmatrix} \sigma_A^2 + \sigma_C^2 + \sigma_E^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 \\ \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \sigma_A^2 + \sigma_C^2 + \sigma_E^2 \end{pmatrix} \times K' \quad \text{where} \quad K = \begin{matrix} \alpha & 0 \\ 0 & \alpha \end{matrix}$$

Where σ_A^2 is the variance due to additive genetic factors, σ_C^2 is variance due to common or shared environmental influences, σ_E^2 is the variance due to unique or non-shared environmental influences, and α is the scalar multiplier. For opposite or unlike sex twins where Twin 1 is male the variance multiplier for the second twin (matrix element 2 2) would be set to unity, that is $K = \begin{matrix} \alpha & 0 \\ 0 & 1 \end{matrix}$.

The extension of this parameterization to data collected using an extended twin-sibling design is relatively straightforward. For a sibship of size four (i.e., a pair of twins and two singleton siblings) the structure for the variance/covariance matrix is described in Table 1. The K matrix in this situation is an 8 by 8 diagonal matrix, containing the following elements: α, 1, α, 1, α, 1, α, 1.

An alternative approach is to compute the matrix of scalar corrections (the K matrix) for each sibship based on the sex-structure of the siblings. This approach takes advantage of the dynamic treatment of covariates within Mx (Neale et al., 2002). An example script demonstrating the implementation of this model is given in Appendix A. In this parameterization the sex of the individuals (coded as a binary variable 0/1) is included in the model as a definition variable, using the following command:

Definition_variable sex1 sex2 sex3 sex4;

The sex-covariates are assigned to a given matrix, in this case S (defined as a Full matrix of 1 by nsib, where nsib = the number of siblings), using the following command:

Table 1

The Structure for the Variance/Covariance Matrix for a Sibship of Size Four Using the Traditional Parameterization

	Twin 1 Male (M1)	Twin 1 Female (F1)	Twin 2 Male (M2)	Twin 1 Female (F2)	Sib 1 male (M3)	Sib 1 Female (F3)	Sib 2 Male (M4)	Sib 2 Female (F4)
M1	Var_{M1}							
F1	—	Var_{F1}						
M2	Cov_{M1-M2}	Cov_{F1-M2}	Var_{M2}					
F2	Cov_{M1-M2}	Cov_{F1-M2}	—	Var_{F2}				
M3	Cov_{M1-M3}	Cov_{F1-M3}	Cov_{M2-M3}	Cov_{F2-M3}	Var_{M3}			
F3	Cov_{M1-F3}	Cov_{F1-F3}	Cov_{M2-F3}	Cov_{F2-F3}	—	Var_{F3}		
M4	Cov_{M1-M4}	Cov_{F1-M4}	Cov_{M2-M4}	Cov_{F2-M4}	Cov_{M3-M4}	Cov_{F3-M4}	Var_{M4}	
F4	Cov_{M1-F4}	Cov_{F1-F4}	Cov_{M2-F4}	Cov_{F2-F4}	Cov_{M3-F4}	Cov_{F3-F4}	—	Var_{F4}

Specify S sex1 sex2 sex3 sex4.

The matrix of sex-covariates is then multiplied (using a Kronecker multiplication) by a matrix of free parameters, in this case J (defined as Full matrix of 1 by nvar, where nvar = the number of variables in the analysis) which will become the scalar sex corrections (the diagonal of the K matrix). By adding a unit matrix (of dimensions 1 by nsib*nvar) to the Kronecker product (denoted by @ in Mx syntax) of S and J, we arrive at a matrix where the scalar correction equals 1 for the sex coded 0 and $(1 + \alpha)$ for the sex coded 1.

L = (S@J) + I;

This vector may then be placed in the leading diagonal of a square matrix, in this case K, using the vector to diagonal command:

K = \v2d (L);

Using this parameterization, the structure for the variance/covariance matrix for the situation described in Table 1 is given in Table 2. In addition to modeling a scalar sex limitation on the total variance it is also possible to model scalar sex limitation of the individual variance components. This parameterization (provided in Appendix B) allows a test of the equivalence of the individual scalars through equating the relevant parameters.

Non-scalar Sex-limitation

In non-scalar sex-limitation models the variance components of males and females are allowed to vary and an additional set of genes affecting only one sex may be postulated (Neale & Cardon, 1992). The variance covariance matrix for a pair of DZ opposite sex twins (where Twin 1 is female), is shown below:

$$\begin{pmatrix} \sigma^2_{Af} + \sigma^2_{Cf} + \sigma^2_{Ef} & \frac{1}{2}\left(\sigma_{Am} \times \sigma_{Af}\right) + \left(\sigma_{Cm} \times \sigma_{Cf}\right) \\ \frac{1}{2}\left(\sigma_{Af} \times \sigma_{Am}\right) + \left(\sigma_{Cf} \times \sigma_{Cm}\right) & \sigma^2_{Am} + \sigma^2_{Cm} + \sigma^2_{Em} + \sigma^2_{Gm} \end{pmatrix}$$

where the subscripts $_m$ and $_f$ denote male and female variance components. Thus, for example, σ^2_{Af} is the female variance due to additive genetic factors, σ^2_{Am} is

the male variance due to additive genetic factors, and σ^2_{Gm} refers to the variance due to a uniquely male set of genes. The non-scalar model is usually described as a general sex-limitation model when σ^2_{Gm} is estimated and a common sex-limitation model when σ^2_{Gm} is set to 0. The extension of this parameterization to the extended twin design is simple assuming the covariances are arranged as shown in Table 1.

An alternative approach utilizing the dynamic nature of Mx follows; an example script is given in Appendix C. Two matrices that specify the combination of male and female parameters will be used to define the variance/covariance matrix for each sibship. As was the case in the scalar sex limitation model above, in this parameterization the sex of the individuals is included in the model as definition variables (in this case it will be assumed that males are coded as 1, females as 0 and "missing" as .5). The matrix of sex-covariates (S: defined as Full nsib by 1) will be used to stipulate that the male path coefficients ($\sigma^2_{Am} + \sigma^2_{Cm} + \sigma^2_{Em}$, and σ^2_{Gm}) should be used for male participants.

To produce an equivalent matrix for females (P) the sex covariate matrix (S) is subtracted from matrix J, a unit matrix (defined as nsib by 1): P = J – S. The resulting matrix is a vector of sex coefficients where male is coded as 0, female as 1, and "missing" as .5, that will be used to specify female path coefficients. For example, in sibship j where the sex of the four siblings is male, female, female, male,

$$S_j = \begin{matrix} 1 \\ 0 \\ 0 \\ 1 \end{matrix} \quad \text{and,} \quad P_j = \begin{matrix} 0 \\ 1 \\ 1 \\ 0 \end{matrix}$$

These two matrices (S and P) are then multiplied (using Kronecker multiplication) by the male and female additive genetic and common environmental path coefficients (X, Y and U, V respectively). The addition of the two resultant matrices produces the correct combination of male and female parameters. A similar procedure is used to specify the variance due to the uniquely male set of genes (R):

Table 2

The Structure for the Variance/Covariance Matrix for a Sibship of Size Four Using the Alternate Parameterization

	Twin 1 (T1)	Twin 2 (T2)	Sibling 1 (S1)	Sibling 2 (S2)
T1	Var_{T1}			
T2	$Cov_{T1\text{-}T2}$	Var_{T2}		
S1	$Cov_{T1\text{-}S1}$	$Cov_{T2\text{-}S1}$	Var_{S1}	
S2	$Cov_{T1\text{-}S2}$	$Cov_{T2\text{-}S2}$	$Cov_{S1\text{-}S2}$	Var_{S2}

(S@ X+P@U) — for the additive genetic path coefficients.
(S@Y+P@V) — for the common environmental path coefficients.
(S@R) — for the additional set of uniquely male genes.

These matrices are then multiplied by their transpose and the resultant additive genetic parameters are multiplied by a matrix (in this case *H*) which contains the additive genetic correlations between siblings (i.e., 1 between monozygotic (MZ) twins and .5 between DZ twins):

((S@X+P@U)*(S@X+P@U)').H + ((S@R)*(S@R)').H +((S@Y+P@V)* (S@Y+P@V)')

The result is the sum of the additive genetic and common environmental structures corrected for the sex of the siblings to which the unique environmental structure is added.

As the specific environmental influences are not correlated between siblings a different approach is used for their inclusion in the variance/covariance matrix. To obtain the male-specific environment structure the male path coefficient (*Z*) is multiplied by its transpose and the result is then multiplied (using Kronecker multiplication) by a matrix (*O*) which contains the elements of *S* on its diagonal (O = \v2d(S')). A similar process is followed for females: the path coefficient is multiplied by its transpose and the result is multiplied by matrix *T* (T = \v2d(P')). The resulting diagonal matrices are added together to produce the specific environmental variance structure:

O@(Z*Z')+T@(W*W').

Thus the complete variance covariance matrix is given by:

Covariances ((S@X+(P@U))*(S@X+(P@U))').H + ((S@R)*(S@R)').H +((S@Y+(P@V))* (S@Y+(P@V))') + O@(Z*Z')+T@(W*W').

Using this parameterization, the structure for the variance/covariance matrix for the situation described in Table 1 is given in Table 2.

Conclusion

The computational efficiency of alternate parameterization is expected to be similar to that of a traditional parameterization in which definition variables have been used. However, the computational efficiency of traditional parameterization without definition variables is expected to be greater than that of the alternative parameterization presented here. In small data sets, the traditional formulation may result in low numbers of siblings in each cell. This may be particularly problematic in the analysis of categorical data where such an approach may result in empty cells in the contingency table. For example, in a variable with three categories of decreasing prevalence, (i.e., mild, moderate, and severe) the number of families required to ensure that there are sufficient male and female siblings who fall into the severe category within each of the zygosity groups may prohibit the inclusion of sibling data in this analysis. By tailoring the structure of the variance-covariance matrix to each family the alternative parameterization allows the analysis of male and female data together, which will help to minimize this problem.

References

Eaves, L. J., Last, K. A., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human behaviour. *Heredity, 41*(3), 249–320.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2002). *Mx: Statistical Modelling* (6th ed.). VCU Box 900126, Richmond, VA 23298, http://www.vcu.edu/mx/: Department of Psychiatry, VCU.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Appendix A

! SCALAR SEX LIMITATION MODEL FOR 2 TWINS AND 2 SIBLINGS

```
#define nvar 1     !number of variables — this is a univariate example so 1
#define nsib 4     !size of largest sibship — in this case 4
#define tvar 4     !nvar * nsib
```

G1: Data from the MZ Twins and Their Siblings

```
Data NGroups = 2 NI = 10
Rectangular File = mz.dat
labels family zygosity sex1 trait1 sex2 trait2 sex3 trait3 sex4 trait4
select trait1 trait2 trait3 trait4 sex1 sex2 sex3 sex4 ;
Definition_variables sex1 sex2 sex3 sex4 ;
Begin Matrices ;
X Lower nvar nvar Free     ! Genetic structure
Z Lower nvar nvar Free     ! Specific environmental structure
Y Lower nvar nvar Free     ! Common environmental structure
M Full nsib nvar Free      ! Means
S Full 1 nsib              ! Sex definition variables
B Full nvar 1              ! Sex correction for means model
I Unit 1 tvar              ! 1 to add to sex correction
J Full 1 nvar Free         ! Scalar sex correction
H Full 1 1                 ! .5
End matrices ;
Specify S sex1 sex2 sex3 sex4 ;
Matrix H .5 ;
Begin Algebra;
L = (S@J) + I ;
K = \v2d(L) ;
A = X*X' ;
E = Z*Z' ;
C = Y*Y' ;
T = A+C+E;
End Algebra ;
Means M + (B*S)' ;
Covariances K*(  T        | A+C      | H@A+C      | H@A+C_
                 A+C      | T        | H@A+C      | H@A+C_
                 H@A+C    | H@A+C    | T          | H@A+C_
                 H@A+C    | H@A+C    | H@A+C      | T          )*K' ;
Start 13 M 1 1 — M nsib nvar
Start 0.1 J 1 1 — J 1 nvar
Start 4 Y 1 1
Start 6 X 1 1
Start 4 Z 1 1
op rs
End
```

G2: Data from the DZ Twins and Their Siblings

```
Data NI = 10
Rectangular File = dz.dat
labels family zygosity sex1 trait1 sex2 trait2 sex3 trait3 sex4 trait4
select trait1 trait2 trait3 trait4 sex1 sex2 sex3 sex4 ;
Definition_variables sex1 sex2 sex3 sex4 ;
Begin Matrices = Group 1
End matrices ;
Specify S sex1 sex2 sex3 sex4 ;
Begin Algebra;
L = (S@J) + I ;
K = \v2d(L) ;
End Algebra ;
```

Means M + (B*S)' ;
Covariances K*(T | H@A+C | H@A+C | H@A+C_
                H@A+C | T | H@A+C | H@A+C_
                H@A+C | H@A+C | T | H@A+C_
                H@A+C | H@A+C | H@A+C | T   )*K' ;
End

Appendix B

! SCALAR SEX LIMITATION MODEL FOR 2 TWINS AND 2 SIBLINGS

#define nvar 1   !number of variables — this is a univariate example so 1
#define nsib 4   !size of largest sibship — in this case 4
#define tvar 4   !nvar * nsib

G1: Data from the MZ Twins and Their Siblings

```
Data NGroups = 2 NI = 10
Rectangular File = mz.dat
labels family zygosity sex1 trait1 sex2 trait2 sex3 trait3 sex4 trait4
select trait1 trait2 trait3 trait4 sex1 sex2 sex3 sex4 ;
Definition_variables sex1 sex2 sex3 sex4 ;
Begin Matrices ;
X Lower nvar nvar Free         ! Genetic structure
Z Lower nvar nvar Free         ! Specific environmental structure
Y Lower nvar nvar Free         ! Common environmental structure
M Full nsib nvar Free          ! Means
S Full 1 nsib                  ! Sex definition variables
B Full nvar 1                  ! Sex correction for means model
I Unit 1 tvar                  ! 1 to add to sex correction
J Full 1 nvar Free             ! Scalar sex correction for A
K Full 1 nvar Free             ! Scalar sex correction for C
L Full 1 nvar Free             ! Scalar sex correction for E
H Stand nsib nsib              ! To specify the additive genetic correlation
U Unit nsib nsib
T Diag nsib nsib
End matrices ;
Specify S sex1 sex2 sex3 sex4 ;
Value .5 H 2 1 — H nsib 3 ;
Value 1 H 2 1 ;
Value 1 T 1 1 — T nsib nsib ;
Begin Algebra;
O = \v2d((S@J) + I) ;
P = \v2d((S@K) + I) ;
Q = \v2d((S@L) + I) ;
A = X*X' ;
E = Z*Z' ;
C = Y*Y' ;
End Algebra ;
Means M + (B*S)' ;
Covariances (O*(H@A)*O')+(P*(U@C)*P')+(Q*(T@E)*Q') ;
Start 13 M 1 1 — M nsib nvar
Start 0.1 J 1 1 — J 1 nvar
Start 0.1 K 1 1 — K 1 nvar
Start 0.1 L 1 1 — L 1 nvar
Start 4 Y 1 1
Start 6 X 1 1
Start 4 Z 1 1
op rs
End
```

Sarah E. Medland

G2: Data from the DZ Twins and Their Siblings

```
Data NI = 10
Rectangular File = dz.dat
labels family zygosity sex1 trait1 sex2 trait2 sex3 trait3 sex4 trait4
select trait1 trait2 trait3 trait4 sex1 sex2 sex3 sex4 ;
Definition_variables sex1 sex2 sex3 sex4 ;
Begin Matrices = Group 1
S Full 1 nsib          ! Sex definition variables
H Stand nsib nsib     ! .5
End matrices ;
Specify S sex1 sex2 sex3 sex4 ;
Value .5 H 2 1 - H nsib 3 ;
Value 1 T 1 1 - T nsib nsib ;
Begin Algebra;
O = \v2d((S@J) + I) ;
P = \v2d((S@K) + I) ;
Q = \v2d((S@L) + I) ;
End Algebra ;
Means M + (B*S)' ;
Covariances (O*(H@A)*O')+(P*(U@C)*P')+(Q*(T@E)*Q') ;
Options multiple
End
EQ J 1 1 1 K 1 1 1 L 1 1 1
! Reduces the model to a total variance scalar sex limitation model
End
```

Appendix C

! NON-SCALAR SEX LIMITATION MODEL FOR 2 TWINS AND 2 SIBLINGS

```
#define nvar 1    !number of variables — this is a univariate example so 1
#define nsib 4    !size of largest sibship — in this case 4
```

G1: Data from the MZ Twins and Their Siblings

```
Data NGroups = 2 NI = 10
Rectangular File = mz.dat
labels family zygosity sex1 trait1 sex2 trait2 sex3 trait3 sex4 trait4
select trait1 trait2 trait3 trait4 sex1 sex2 sex3 sex4 ;
Definition_variables sex1 sex2 sex3 sex4 ;
Begin Matrices ;
X Lower nvar nvar Free   ! Male genetic structure
Y Lower nvar nvar Free   ! Male common environmental structure
Z Lower nvar nvar Free   ! Male specific environmental structure
R Lower nvar nvar Free   ! An additional uniquely male set of genes
U Lower nvar nvar Free   ! Female genetic structure
V Lower nvar nvar Free   ! Female common environmental structure
W Lower nvar nvar Free   ! Female specific environmental structure
M Full nsib nvar Free    ! Means
S Full nsib 1            ! Sex definition variables
B Full nvar 1            ! Sex correction for means model
J Unit nsib 1            ! Matrix of 1s to create sex correction
L Unit nvar nvar         ! To generalise to the multivariate case
F Stand 4 4
! F is used to specify the additive genetic correlations between siblings
End matrices ;
Specify S sex1 sex2 sex3 sex4 ;
Value 1 F 2 1
Value .5 F 3 1 — F nsib 3
Begin Algebra;
H = F@L;
A = (U*U')+(V*V')+(W*W') ;      ! FEMALE VARIANCE
```

```
C = (X*X')+(Y*Y')+(Z*Z')+(R*R');          ! MALE VARIANCE
P = J-S;
O = \v2d(S');
T = \v2d(P');
End Algebra ;
Means M + (B*S')' ;
Covariances       ((S@X+(P@U))* (S@X+(P@U))').H + ((S@R)*(S@R)').H +
((S@Y+(P@V))* (S@Y+(P@V))') + O@(Z*Z')+T@(W*W') ;
Start 130 M 1 1 — M nsib nvar
Start 4 Y 1 1 V 1 1 R 1 1
Start 6 X 1 1 U 1 1
Start 4 Z 1 1 W 1 1
op rs
End
```

G2: Data from the DZ Twins and Their Siblings

```
Data NGroups = 2 NI = 10
Rectangular File = dz.dat
labels family zygosity sex1 trait1 sex2 trait2 sex3 trait3 sex4 trait4
select trait1 trait2 trait3 trait4 sex1 sex2 sex3 sex4 ;
Definition_variables sex1 sex2 sex3 sex4 ;
Matrices = Group 1;
S Full nsib 1       ! Sex definition variables
F Stand 4 4
! F is used to specify the additive genetic correlations between siblings
End matrices ;
Specify S sex1 sex2 sex3 sex4 ;
Value .5 F 2 1 — F nsib 3
Begin Algebra;
H = F@L;
P = J-S;
O = \v2d(S');
T = \v2d(P');
End Algebra ;
Means M + (B*S')' ;
Covariances       ((S@X+(P@U))* (S@X+(P@U))').H + ((S@R)*(S@R)').H +
((S@Y+(P@V))* (S@Y+(P@V))') + O@(Z*Z')+T@(W*W') ;
Options Multiple
End
Drop R 1 1 1
! Dropping Gm — testing the general vs the common sex limitation model
End
```