Adjustment for heterogeneous variance

Model description

This example uses the data given by Schaeffer & Dekkers: Random regressions in animal models for test-day production in dairy cattle (Proc. 5th WCGALP, 18:443-446, 1994).

The model with the breeding values (mean model) is:

$$\mathbf{y}_i\lambda_\mathrm{i} = f_1(\mathbf{p}_i, \mathbf{DIM}) + \mathbf{Xb}_i + f_2(\mathbf{a}_i, \mathbf{DIM}) + \mathbf{e}_i,$$

where,

$\mathbf{y}_i$ is vector of milk yield observations of stratum i,
λ_i is an adjustment factor of stratum i,
$f_1(\mathbf{p}_i, \mathbf{DIM})$ is fixed general regression function,
$\mathbf{b}_i$ is vector of fixed month and herd x year interaction effects of stratum i,
$\mathbf{X}$ is design matrix to link observations to appropriate fixed effects,
$f_2(\mathbf{a}_i, \mathbf{DIM})$ is random additive genetic regression function, and
$\mathbf{e}_i$ is random residual vector of stratum i.

The fixed regression function has form

$$f_1(\mathbf{p}_i, DIM) = DIM \cdot p_{i,1} + \ln(305/DIM) \cdot p_{i,2}$$

and the random regression function f for animal j in stratum i has form

$$f_2(\mathbf{a}_{i,j}, DIM) = a_{i,j,1} + DIM \cdot a_{i,j,2} + \ln(305/DIM) \cdot a_{i,j,3}.$$

It is assumed that $\mathrm{var}(\mathbf{a}) = \mathbf{A}\otimes\mathbf{G}_0 = \mathbf{G}$, $\mathrm{var}(\mathbf{e}) = \mathbf{I}\sigma_\mathrm{e}^2 = \mathbf{R}$ and $\mathrm{cov}(\mathbf{a}, \mathbf{e}) = \mathrm{cov}(\mathbf{e}, \mathbf{a}) = \mathbf{0}$, where $\mathbf{G}_0$ is 3×3 -matrix of variance and covariance components for random animal effects and $\mathbf{A}$ is the numerator relationship matrix.

The variance components are

$$\mathbf{G}_0 = \begin{bmatrix} 44.791 & -0.133 & 0.351 \\ -0.133 & 0.073 & -0.010 \\ 0.351 & -0.010 & 1.068 \end{bmatrix}$$

and $\sigma_\mathrm{e}^2 = 100$.

The model, which describes the heterogeneity of variances (variance model) is:

$$\mathbf{s}_i = \boldsymbol{\beta}_{1i} + \boldsymbol{\beta}_{2i} + \boldsymbol{\varepsilon}_i,$$

where,

$\mathbf{s}_i$ is vector of resembled observations for variance model in stratum i,
$\boldsymbol{\beta}_{1i}$ is vector of fixed month effects in stratum i,
$\boldsymbol{\beta}_{2i}$ is vector of random herd x year interaction effects in stratum i, and
$\boldsymbol{\varepsilon}_i$ is random residual vector in stratum i.

It is assumed that $\mathrm{var}(\boldsymbol{\beta}_\mathbf{2}) = \mathbf{I}\otimes R_i(\rho)$, $\mathrm{var}(\boldsymbol{\varepsilon}) = \mathbf{I}\sigma_\varepsilon^2 = \mathbf{R}$ and $\mathrm{cov}(\boldsymbol{\beta}, \boldsymbol{\varepsilon}) = \mathrm{cov}(\boldsymbol{\varepsilon}, \boldsymbol{\beta}) = \mathbf{0}$, where $R_i(\rho)$ indicates first order autoregressive process.

The variance components are

$$R_i(\rho) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where $\sigma^2 = 0.0785$, $\rho = 0.7$, and $\sigma_\varepsilon^2 = 0.5909$.

Input files

Mean model:

Data file:

The data file contains information on the data to be analyzed together with class and regression variables for the model. The data file can be in free format (columns are separated by at least one space) or in binary format.

Each record, i.e., line in a free format file, has been divided to two parts: 1) Integer number columns and 2) real number columns. Columns of real numbers are always after the integer number columns.

mean_model.dat

```
1  1  1  11  1   73.0   1.4298500   26.0
2  1  1  11  1  123.0   0.9081270   23.0
3  1  1  11  1  178.0   0.5385280   21.0
4  2  1  12  2  184.0   0.5053760    1.0
5  4  1  12  2  165.0   0.6143660   14.0
1  2  2  21  1   34.0   2.1939499   29.0
2  2  2  21  1   84.0   1.2894900   18.0
2  3  2  21  1   58.0   1.6598700   25.0
3  3  2  21  1  113.0   0.9929240   19.0
4  3  2  22  2  158.0   0.6577170   15.0
4  3  2  22  2  218.0   0.3358170   11.0
1  3  3  31  1    8.0   3.6408701   37.0
.  .  .   .  .     .      .             .
.  .  .   .  .     .      .             .
```

Column 1: Month (integer)
Column 2: Animal ID (integer)
Column 3: Herd (integer)
Column 4: Herd x Year (integer)
Column 5: Year (integer)
Column 6: Covariable for regression effect, days in milk (real)
Column 7: Covariable for regression effect, ln(305/days in milk) (real)
Column 8: Milk yield (real)

Pedigree file:

All the pedigree information must be given in the pedigree file. Each animal in the pedigree must have a record with four integers of which the forth integer is optional.

mean_model.ped

```
 1    9   7   1
 2   10   8   1
 3    9   2   2
 4   10   8   3
 5   11   7   3
 6   11   1   4
 7    0   0   8
 8    0   0   8
 9    0   0   8
10    0   0   8
11    0   0   8
```

Column 1: Animal ID
Column 2: Sire ID
Column 3: Dam ID
Column 4: Herd code

Variance component file:

The variance components file has variances and covariances for all the random effects in the model. The variance component file has a line for each (co)variance. Order of lines in the file is irrelevant.

mean_model.var

```
  1   1   1     44.791
  1   2   1     -0.133
  1   3   1      0.351
  1   2   2      0.073
  1   2   3     -0.010
  1   3   3      1.068
  2   1   1    100.000
```

Column 1: Random effect number (integer)
Column 2: Row index (integer)
Column 3: Column index (integer)
Column 4: (Co)variance value (real)

Command file:

Everything beyond '#' sign is considered as a comment.

MEAN_MODEL.DIR

```
TITLE:
Heterogenous variance adjustment by Meuwissen approach: a demonstration
INT-VAR:
        Month Animal Herd HerdYear Year
REAL-VAR:
        Covar_1  Covar_2  Milk
TRAITS:
        1 1
TRGRP:
        1 -
SORT_R: Block sorting variable, Animal sorting variable
              3                              2
FIXRAN: Number of different fixed and random effect
                              2          3
MODEL:  Subgrp. Trait Weight Month HerdYear Animal, Gamma1, Gamma2
         1        3     -     1      4       2        2       2
BLKORD: Order of effects within herd
         -      2       1        1       1
RANDOM: Animal, Gamma1, Gamma2
               1        1       1
PEDIGR: Number: Animal, Gamma1, Gamma2
               3        1        1       1
REGRES: Number Beta1, Beta2, Month HerdYear  Animal, Gamma1,Gamma2
          7      1      2      cl     cl      cl       1       2
COMBINE:
         n
RELMET:
         am
INPFIL:
         mean_model.dat
VAR:
         5 3 f
MISSVA:
         0.0
SCALE:
         n
PEDFIL:
         mean_model.ped
PARFIL:
         mean_model.var
TMPDIR:
         tmpMiX
SOLFOR: Animal effects
             y
SOLUNF:
             n
PRECON: WpW, XpX
        animal herd-year :   XpX
           d      d           b
NPROC:  Number of processors used by the sover program
             2
COMBLK  Number of blocks in common area when using parallel processing
             1
```

Solver option file:

Everything beyond '#' sign is considered as a comment.

stpA

```
# RAM: RAM demand: H=high, M=medium, L=low
   H
 # STOP: Maximum_number_of_iterations, Stopping_criteria
   20           1.0e-7     D        f
 # RESID: Calculate residuals? (Y/N)
   N
 # VALID: N=no, P=prediction, S=sum of effects, Y=YD, D=DYD, I=IDD
   N
 # HETVAR: adjust for heterogeneous variance (N, S, C)
   S
   # HETADJ: Beta1, Beta2, Month HerdYear  Animal, Gamma1, Gamma2
          #   0      0      0      0        1        1       1
              1      1      1      1        1        1       1
 # TYPSOL: type of solution files? (N,Y,A)
   N
```

Creation of data for variance model:

Command file:

Everything beyond '#' sign is considered as a comment.

HVDATA.DIR

```
#  Name of the data file
   mean_model.dat
#  Number of integer columns; type of file (f/u)
   5  f
#  Column with the trait group code
   0
#  Column with the block variable for the auto regress. proc. (HERD)
   3
#  Heterogeneity Model:
#  Fixed across block strata; random within block strata
#  (Month)                    (Year within Herd)
    1                              5
#  Create pedigree file for REML-analysis? (y/n)
   n
#  Standardisation of multiplicative adjustment factors? (y/n)
   n
   #      Name of file with the class codes of the base
   #      ID.Base.Classes
#  Approximation of rank across traits (y/n)
   y
#  Random HY effect: Combining of second and third lactation
#                    in the variance model
   1
#  Directory for the temporary files
   ./tmpHV
```

Variance model:

Data file:

Data file HV.data is made by mix99hv-program.

Autocorrelation structure:

Normal pedigree file is replaced by information about the autocorrelation structure. The file HV.pedi is made by mix99hv-program.

Variance component file:

The variance components file has variances and covariances for all the random effects in the model. The variance component file has a line for each (co)variance. Order of lines in the file is irrelevant.

variance_model.var

```
  1   1   1     0.078485
  2   1   1     0.590860
```

Column 1: Random effect number (integer)
Column 2: Row index (integer)
Column 3: Column index (integer)
Column 4: (Co)variance value (real)

Command file:

Everything beyond '#' sign is considered as a comment.

VARIANCE_MODEL.DIR

```
# title
  Variance model
# INT-VAR
  BLOCK HERDxYR TRGRP MONTH
# REAL-VAR
  HET_OBS  WEIGHT
# traits
   1
# trait-subgroups,  input column
   1                    3
# input column of block code and animal code
   1   2
# number of fixed- and random effect columns in the model lines
   1   1
# MODEL:
# subgrp  t wt |   Month, Herd
    1     1  2       4     2
# BLKORD
              -     1
# RANDOM:
# Herd
  1
# pedigree for animal effects: Num:  Herd
                                 1     1
# REGRESS:
# num:   Month,  Herd
  2       cl      cl
# COMBINE:
  n
# RELMET:
  AR
# ROH
  0.7
# input file
  ../tmpHV/HV.data
# int-col. real-col. form
    4         2       u
# code for missing real values
  -8192.0
# scaling (y/n)
   n
# pedigree file
  ../tmpHV/HV.pedi
# parameter file
  variance_model.var
# directory for the temporary files
  ../tmpMiX/B1
# solution files for random effects: Herd
                                      y
# binary solution file
  n
# block preconditioner WpW, XpX  (=D_M)
  d   d
# number of processors
  2
# last pedigree blocks in common
  0
```

Solver option file:

Everything beyond '#' sign is considered as a comment.

stpV

```
# RAM: RAM demand: H=high, M=medium, L=low
  H
# STOP: Maximum_number_of_iterations, Stopping_criteria
  50           1.0e-7        D          f
# RESID: Calculate residuals? (Y/N)
  N
# VALID: N=no, P=prediction, S=sum of effects, Y=YD, D=DYD, I=IDD
  N
# HETVAR: adjust for heterogeneous variance (N,S,C)
  N
# TYPSOL: type of solution files? (N,Y,A,H)
  H
```

Cycle between mean model and variance model:

Solver option file:

Everything beyond '#' sign is considered as a comment.

stpC

```
# RAM: RAM demand: H=high, M=medium, L=low
  H
# STOP: Maximum_number_of_iterations, Stopping_criteria (a/r/d)
  2500           1.0e-7                      r       f
# RESID: Calculate residuals? (Y/N)
   N
# VALID: N=no, P=prediction, S=sum of effects, Y=YD, D=DYD, I=IDD
   N   # calculate of individual daughter deviations
# HETVAR: adjust for heterogeneous variance (N, S, C)
   C
   # HETADJ: Beta1, Beta2, Month HerdYear  Animal, Gamma1, Gamma2
        #     0      0      0      0        1        1       1
              1      1      1      1        1        1       1
   # STOPC: Max cycles; stopping criteria for lamda values (CD)
      150                    1.0e-8
# TYPSOL: type of solution files? (N,Y,A)
   Y
```

Solution files

Structure of the formatted solution files depends on the model. Therefore, explanation of the content of those files is given in the printout of the particular run of the program.

Solver option files for heterogeneous variance adjustment analysis have been given in 'Input files for Example 10'.

"Solreg"-file contains solutions for general regressions.

Trt	Reg-No	Solution	Trait	Covariable
1	1	-0.55740E-01	Milk	Covar_1
1	2	3.5341	Milk	Covar_2

Column 1: Trait number
Column 2: Regression number within trait
Column 3: Solution
Column 4: Name of trait
Column 5: Name of covarible

"Solfix"-file contains solutions for across-block fixed effects.

Fact.	Trt	Level	N-Obs	Solution	Factor	Trait
1	1	1	3	19.415	Month	Milk
1	1	2	4	8.1896	Month	Milk
1	1	3	5	13.074	Month	Milk
1	1	4	6	9.1899	Month	Milk
1	1	5	5	9.7699	Month	Milk

Column 1: Factor number
Column 2: Trait number
Column 3: Level code
Column 4: Number of observations
Column 5: Solution
Column 6: Name of factor
Column 7: Name of trait

"Solf01"-file contains solutions for first within-block fixed effect.

11	3	13.337
12	2	7.9732
21	4	10.129
22	2	9.9101
31	4	17.008
32	4	8.0923
41	1	8.0985
42	3	12.008

Column 1: Level code
Column 2: Number of observations
Column 3: Solution for trait 1 Milk and factor HerdYear

"Solani"-file contains solutions for genetic animal effect.

1	1	3	-0.19640	0.35310E-01	-0.17384E-01
2	1	4	0.17984	-0.41071E-01	0.19009E-01
3	0	6	-0.13680	-0.27646E-04	-0.63727E-02
4	0	5	0.39596	-0.20431E-02	0.23862E-01
5	0	3	-0.15555	-0.52320E-02	-0.98592E-02
6	0	2	-0.18066	0.14085E-01	-0.14089E-01
7	2	0	-0.10667	0.10471E-01	-0.86369E-02
8	2	0	0.19194	-0.14371E-01	0.14290E-01
9	2	0	-0.18489	0.25290E-01	-0.14471E-01
10	2	0	0.19194	-0.14371E-01	0.14290E-01
11	2	0	-0.92337E-01	-0.70193E-02	-0.54687E-02

Column 1: Animal ID
Column 2: Number of descendants
Column 3: Number of observations
Column 4: Solution for trait 1 Milk and factor Animal
Column 5: Solution for trait 1 Milk and factor Covar_1(Animal)
Column 6: Solution for trait 1 Milk and factor Covar_2(Animal)