Reduced rank model

Model description

This example is from the old Finnish Ayrshire evaluation model for production. The example restricts to first lactation test day records for milk, protein and fat yields. Permanent environment and genetic lactation curves from the three traits have been combined, i.e., this is a reduced rank random regression test day model.

The model is

$$\mathbf{y} = \mathbf{Xb} + f_1(\mathbf{lc}, \mathbf{DIM}) + \mathbf{Zu} + f_2(\mathbf{pe}, \mathbf{DIM}) + f_3(\mathbf{a}, \mathbf{DIM}) + \mathbf{e},$$

where,

$\mathbf{y}$ is vector of milk, protein and fat yield observations,
$\mathbf{b}$ is vector of fixed herd times year interaction effects,
$\mathbf{X}$ is design matrix to link observations to appropriate fixed effects,
$f_1(\mathbf{l}, \mathbf{DIM})$ is fixed lactation curve regression function,
$\mathbf{u}$ is vector of random herd times test-day interaction effects,
$\mathbf{Z}$ is design matrix to link observations to appropriate random effects,
$f_2(\mathbf{pe}, \mathbf{DIM})$ is random permanent environment regression function,
$f_3(\mathbf{a}, \mathbf{DIM})$ is random additive genetic regression function, and
$\mathbf{e}$ is random residual vector.

The fixed regression function for lactation curve of trait j has form

$$f_1(\mathbf{lc}_j, DIM) = \sum_{r=0}^{3} \phi(DIM)_r \mathbf{lc}_{j,r} + \exp(-0.04 \cdot DIM) lc_{j,4},$$

where $\phi(DIM)_r$ contains Legendre polynomials. The random regression functions f_2 for permanent environment effects and f_3 for genetic animal effects has forms

$$f_2(\mathbf{pe}, DIM) = \sum_{r=0}^{5} \phi(DIM)_r \mathbf{pe}_r + \exp(-0.04 \cdot DIM)\mathbf{pe}_6 \text{ and}$$
$$f_3(\mathbf{a}, DIM) = \sum_{r=0}^{5} \phi(DIM)_r \mathbf{a}_r + \exp(-0.04 \cdot DIM)\mathbf{a}_6.$$

The coefficients differ by trait.

It is assumed that $\mathrm{var}(\mathbf{u}) = \mathbf{I} \otimes \mathbf{S}_0 = \mathbf{S}$, $\mathrm{var}(\mathbf{pe}) = \mathbf{I} \otimes \mathbf{P}_0 = \mathbf{P}$, $\mathrm{var}(\mathbf{a}) = \mathbf{A} \otimes \mathbf{G}_0 = \mathbf{G}$, $\mathrm{var}(\mathbf{e}) = \mathbf{I} \otimes \mathbf{R}_{0,\mathrm{i}} = \mathbf{R}$ and all random effects are non-correlated. Now, $\mathbf{S}_0$ is 3×3 -matrix of variance and covariance components for random herd x test-day effects, $\mathbf{P}_0$ and $\mathbf{G}_0$ are 7×7 -matrices of variance and covariance components for random permanent environment and genetic animal effects, respectively, and $\mathbf{R}_{0,\mathrm{i}}$ is 3×3 -matrix of variance and covariance components for the random residual class $i = 1, \ldots, 12$. $\mathbf{A}$ is the numerator relationship matrix.

Variance-covariance matrices are

$$\mathbf{S}_0 = \begin{bmatrix} 0.1615 & 0.1806 & 0.1146 \\ 0.1806 & 0.2493 & 0.1424 \\ 0.1146 & 0.1424 & 0.1925 \end{bmatrix}$$

$$\mathbf{P}_0 = \begin{bmatrix} 0.05567 & & & & & & \\ & 0.07250 & & & & & \\ & & 0.4089 & & & & \\ & & & 1.4890 & & & \\ & & & & 3.9420 & & \\ & & & & & 6.5690 & \\ & & & & & & 87.5704 \end{bmatrix}$$

$$\mathbf{G}_0 = \begin{bmatrix} 0.06547 & & & & & & \\ & 0.1136 & & & & & \\ & & 0.1438 & & & & \\ & & & 0.2139 & & & \\ & & & & 0.3464 & & \\ & & & & & 0.7556 & \\ & & & & & & 2.9177 \end{bmatrix}$$

$$\mathbf{R}_{0,1} = \begin{bmatrix} 0.09745 & -0.009369 & -0.01815 \\ -0.009369 & 0.1820 & -0.07951 \\ -0.01815 & -0.07951 & 1.5261 \end{bmatrix}$$

$$\mathbf{R}_{0,2} = \begin{bmatrix} 0.4145 & 0.07455 & 0.05787 \\ 0.07455 & 0.2191 & -0.1012 \\ 0.05787 & -0.1012 & 1.8572 \end{bmatrix}$$

$$\mathbf{R}_{0,3} = \begin{bmatrix} 1.4815 & 0.3577 & 0.2804 \\ 0.3577 & 0.9098 & 0.4838 \\ 0.2804 & 0.4838 & 2.0526 \end{bmatrix}$$

$$\mathbf{R}_{0,4} = \begin{bmatrix} 1.1855 & 0.2629 & 0.2052 \\ 0.2629 & 0.7765 & 0.4014 \\ 0.2052 & 0.4014 & 1.8991 \end{bmatrix}$$

$$\mathbf{R}_{0,5} = \begin{bmatrix} 1.0501 & 0.2454 & 0.1852 \\ 0.2454 & 0.6911 & 0.3587 \\ 0.1852 & 0.3587 & 1.6696 \end{bmatrix}$$

$$\mathbf{R}_{0,6} = \begin{bmatrix} 1.1883 & 0.2817 & 0.2062 \\ 0.2817 & 0.8215 & 0.4661 \\ 0.2062 & 0.4661 & 1.5676 \end{bmatrix}$$

$$\mathbf{R}_{0,7} = \begin{bmatrix} 1.1600 & 0.2789 & 0.1903 \\ 0.2789 & 0.8418 & 0.4735 \\ 0.1903 & 0.4735 & 1.2687 \end{bmatrix}$$

$$\mathbf{R}_{0,8} = \begin{bmatrix} 1.0881 & 0.2731 & 0.1830 \\ 0.2731 & 0.8270 & 0.4649 \\ 0.1830 & 0.4649 & 1.1950 \end{bmatrix}$$

$$\mathbf{R}_{0,9} = \begin{bmatrix} 1.1613 & 0.2807 & 0.1896 \\ 0.2807 & 0.8570 & 0.5101 \\ 0.1896 & 0.5101 & 1.0130 \end{bmatrix}$$

$$\mathbf{R}_{0,10} = \begin{bmatrix} 1.4100 & 0.3756 & 0.2597 \\ 0.3756 & 1.0649 & 0.6634 \\ 0.2597 & 0.6634 & 1.1334 \end{bmatrix}$$

$$\mathbf{R}_{0,11} = \begin{bmatrix} 1.6402 & 0.3637 & 0.2662 \\ 0.3637 & 1.0388 & 0.6832 \\ 0.2662 & 0.6832 & 1.1880 \end{bmatrix}$$

$$\mathbf{R}_{0,12} = \begin{bmatrix} 1.4886 & 0.2991 & 0.2474 \\ 0.2991 & 0.8203 & 0.5801 \\ 0.2474 & 0.5801 & 1.1190 \end{bmatrix}$$

Input files

Datafile:

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.

RRANK.dat

```
 48  29  1  2901  7  82  40100001 10 3000210  8  1  5  36  3  15.8194  -9.0000  -9.0000
 48  29  1  2901  7  83  40100001 11 3000211  8  1  5  64  4  14.9325  12.3286  11.4682
 48  29  1  2901  7  84  40100001 12 3000212  8  1  5  91  5  14.0044  -9.0000  -9.0000
 48  29  1  2901  8  85  40100001 13 3000201  8  1  5 128  7  15.0562  12.2857  11.0045
 48  29  1  2901  8  86  40100001 14 3000202  8  1  5 161  7  16.9331  -9.0000  -9.0000
 48  29  1  2901  8  87  40100001 15 3000203  8  1  5 183  8  15.0562  11.8429  10.0773
 48  29  1  2901  8  88  40100001 16 3000204  8  1  5 221  9  14.6231  -9.0000  -9.0000
 48  29  1  2901  8  89  40100001 17 3000205  8  1  5 248  9  14.5612  11.1571   9.1773
 48  29  1  2901  8  90  40100001 18 3000206  8  3  5 281 10  12.7462  -9.0000   9.0000
 48  29  1  2901  8  91  40100001 19 3000207  8  3  5 308 11  13.2413  -9.0000   9.0000
 48  29  1  2901  8  92  40100001 20 3000208  8  4  5 338 12   7.5694   6.5571   6.2455
 48  32  1  3201  7  81  40100001  9 3000209  6  1  5   8  1   9.5081   8.7571   5.5091
  .   .  .    .  .   .         .  .       .  .  .  .   .  .        .        .        .
  .   .  .    .  .   .         .  .       .  .  .  .   .  .        .        .        .
```

Column 1: Herd (integer)
Column 2: Animal (integer)
Column 3: Lactation (integer)
Column 4: Animal x Lactation (integer)
Column 5: Heard x Year (integer)
Column 6: Heard x Test-day (integer)
Column 7: Hslope (integer)
Column 8: Year x Month (integer)
Column 9: Month x 4 year class (integer)
Column 10: Age class (integer)
Column 11: Days carried calf (integer)
Column 12: Days dry (integer)
Column 13: Days in milk (integer)
Column 14: Residual number (integer)
Column 15: Milk yield (real)
Column 16: Protein yield (real)
Column 17: Fat 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.

RRANK.ped

```
29      4    -2    48
32     10    -2    48
34     10    29    48
35     11    29    48
38      3    -2    49
40      3    -2    49
41      4    38    49
44     11    38    49
45     11    44    49
48      5    44    49
56      3    -2    68
57      4    -2    68
58      5    -2    68
61     11    56    68
64     10    57    68
 .      .     .     .
 .      .     .     .
```

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

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.

RRANK.var

```
1   1   1    0.161500000
1   1   2    0.180600000
1   1   3    0.114600000
1   2   2    0.249300000
1   2   3    0.142400000
1   3   3    0.192500000
2   1   1    0.0556749484195297
2   2   1    0
2   2   2    0.0725025611506946
2   3   1    0
2   3   2    0
2   3   3    0.408905016097884
2   4   1    0
2   4   2    0
2   4   3    0
2   4   4    1.4890080533471
2   5   1    0
2   5   2    0
2   5   3    0
2   5   4    0
2   5   5    3.94201313167673
2   6   1    0
2   6   2    0
2   6   3    0
.   .   .    .
.   .   .    .
```

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

Variance component file for multiple residuals:

The separate residual variance component file has variances and covariances for all the residual variance classes in the model. The file has a line for each (co)variance.

RRANK.res

```
1  1  1   0.974470060838579E-01
1  2  1  -0.936891579815379E-02
1  2  2   0.181990477999392
1  3  1  -0.181469885530098E-01
1  3  2  -0.795123860949361E-01
1  3  3   1.52616610883689
2  1  1   0.414463121208139
2  2  1   0.745476018420651E-01
2  2  2   0.219142096613889
2  3  1   0.578676886038661E-01
2  3  2  -0.101246380110280
.  .  .    .
.  .  .    .
```

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

Covariable table file:

The covariable table file has covariables for regression effects in the model. First column contains the index that connects an observation in the data file to the corresponding set of covariables in the table. The rows have to be sorted in ascending order by the covariable index. Within the smallest and largest index line, index lines must not be missing.

RRANK.cov

```
 8 1.0 -1.22474 1.58114 -1.87083 0.72615 0.81823 0.51296 -0.35104 . .
 9 1.0 -1.21788 1.55464 -1.80838 0.69768 0.80690 0.49815 -0.34053 . .
10 1.0 -1.21102 1.52829 -1.74681 0.67032 0.79594 0.48392 -0.33033 . .
11 1.0 -1.20416 1.50209 -1.68611 0.64404 0.78532 0.47027 -0.32042 . .
12 1.0 -1.19730 1.47603 -1.62628 0.61878 0.77503 0.45716 -0.31078 . .
 .   .        .       .        .       .       .       .       . . .
 .   .        .       .        .       .       .       .       . . .
```

Column 1: Days in milk (integer)
Column 2: Intercept term of 3th order Legendre polynomial for fixed lactation curve (real)
Column 3: Slope term of 3th order Legendre polynomial for fixed lactation curve (real)
Column 4: Quadratic term of 3th order Legendre polynomial for fixed lactation curve (real)

Column 5: Cubic term of 3th order Legendre polynomial for fixed lactation curve (real)
Column 6: Wilmink term ($e^{-0.04}$) for fixed lactation curve (real)
Column 2: Intercept term of 5th order Legendre polynomial for permanent environment random regression effect of milk yield (real)
Column 3: Slope term of 5th order Legendre polynomial for permanent environment random regression effect of milk yield (real)
Column 4: Quadratic term of 5th order Legendre polynomial for permanent environment random regression effect of milk yield (real)
.
.

Command file:

CLIM instructions for a reduced rank model analysis. Everything beyond '#' sign is considered as a comment.

RRANK.clm

```
TITLE 3 Ayrshire herds, first lactation, milk protein fat, rank 7

DATAFILE    RRANK.dat
INTEGER     herd animal LAC ANIxLAC HY HTD Hslope YM &
            MTHx4YR AGEclass DCC DDRY DIM residual_num
REAL        milk protein fat

DATASORT    BLOCK=herd PEDIGREECODE=animal

PEDFILE     RRANK.ped         # Pedigree file
PEDIGREE    G am+p 0.3333333 # am=animal model, random genetic groups
RANDOM      HTD PE
NORANSOL    HTD PE

PARFILE     RRANK.var         # Variance component file
RESIDFILE   RRANK.res         # Multiple residuals
TABLEFILE   RRANK.cov         # Covariable table information

RESIDUAL    residual_num
TABLEINDEX DIM

MISSING     -9.
PRECON      d  d  d  d  b
WITHINBLOCKORDER G PE HTD HY

MODEL
  milk    = HY fix_curve(t1 t2 t3 t4 t5 | LAC)  HTD &
            PE( t6  t7  t8  t9 t10 t11 t12| animal)@fst &
            G(t27 t28 t29 t30 t31 t32 t33| animal)@fst

  protein = HY fix_curve(t1 t2 t3 t4 t5 | LAC)  HTD &
            PE(t13 t14 t15 t16 t17 t18 t19| animal)@fst &
            G(t34 t35 t36 t37 t38 t39 t40| animal)@fst

  fat     = HY fix_curve(t1 t2 t3 t4 t5 | LAC)  HTD &
            PE(t20 t21 t22 t23 t24 t25 t26| animal)@fst &
            G(t41 t42 t43 t44 t45 t46 t47| animal)@fst
```

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.

MiX99-solver program has been run by command `mix99s -s`, meaning that in all parts of the program default values are used.

"Solfix"-file contains solutions for fixed effects.

Fact.	Trt	Level	N-Obs	Solution	Factor	Trait
1	1	1	122	5.2513	T1(LAC)	MILK
1	2	1	59	4.7220	T1(LAC)	PROTEIN
1	3	1	59	4.9463	T1(LAC)	FAT
2	1	1	122	-4.2418	T2(LAC)	MILK
2	2	1	59	-1.7821	T2(LAC)	PROTEIN
2	3	1	59	-2.1270	T2(LAC)	FAT
3	1	1	122	-0.66598E-01	T3(LAC)	MILK
3	2	1	59	-0.86012	T3(LAC)	PROTEIN
3	3	1	59	-0.19078	T3(LAC)	FAT
4	1	1	122	-0.67025	T4(LAC)	MILK
4	2	1	59	0.43670E-01	T4(LAC)	PROTEIN
4	3	1	59	-0.12096	T4(LAC)	FAT
5	1	1	122	-7.3449	T5(LAC)	MILK
5	2	1	59	1.8808	T5(LAC)	PROTEIN
5	3	1	59	-2.3557	T5(LAC)	FAT

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.

7	19	5.2425	3.5825	2.0408
8	53	6.9468	4.1820	3.0713
9	22	5.9377	3.4939	2.2117
10	28	5.6559	3.5134	1.8584

Column 1: Level code
Column 2: Number of observations
Column 3: Solution for trait 1 milk and factor HY
Column 4: Solution for trait 2 protein and factor HY
Column 5: Solution for trait 3 fat and factor HY

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

29	2	11	-0.53313E-01	-0.39332E-01	-0.32179E-01	-0.16458E-01	.	.	.
32	0	9	0.36233E-01	-0.40278E-01	0.20700E-01	0.55185E-01	.	.	.
34	0	8	0.86730E-02	0.23274E-01	-0.66625E-01	0.42099E-01	.	.	.
35	0	11	0.19903E-04	0.49056E-01	-0.19022	0.12564	.	.	.
38	2	12	0.66930E-01	0.90357E-01	-0.68414E-01	0.14589	.	.	.
40	2	8	0.40998E-01	0.39146E-01	0.78203E-02	-0.49407E-01	.	.	.
41	0	9	-0.15436E-01	-0.49803E-01	0.13587	-0.12185	.	.	.
44	2	3	0.40230E-01	-0.51780E-02	-0.59744E-02	0.52057E-01	.	.	.
45	0	8	0.34592E-02	-0.71975E-01	0.59179E-01	-0.47872E-01	.	.	.
48	0	4	-0.14091E-01	-0.49553E-01	0.30043E-01	0.35543E-01	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

Column 1: Animal ID
Column 2: Number of descendants
Column 3: Number of observations
Column 4: Solution for trait 1 milk and factor T27 (animal)
Column 5: Solution for trait 1 milk and factor T28 (animal)
Column 6: Solution for trait 1 milk and factor T29 (animal)
Column 7: Solution for trait 1 milk and factor T30 (animal)
Column 8: Solution for trait 1 milk and factor T31 (animal)
Column 9: Solution for trait 1 milk and factor T32 (animal)
Column10: Solution for trait 1 milk and factor T33 (animal)