MiX99

Solving Large Mixed Model Equations

MTT

Reducing model dimensions

Timo Pitkänen, MTT

3.12.2014

Outline

- About rank reduction
- Example: Random regression test day model
- Demo!!

About rank reduction

- The aim of rank reduction is
  - Decrease number of unknown **parameters** in the model
  - Decrease model **complexity**

- When rank is reduced in a proper way the results between ”**full**” model and **reduced** model does not differ (at least significantly)
- Number of unknowns may still differ dramatically!

Example: Nordic test-day model

- **Test-day** observations from **Finland, Denmark** and **Sweden**
- Separate models for **HOL**, **JER** and **RDC**
- **HOL** and **RDC** : **27** traits (milk, protein, fat, 3 lactations, 3 countries)
- **JER 9** traits (Swedish and Danish data **combined**)
- HOL evaluation data has **143,4** million **records** from **7,1** million cows

Estimation of variance components for evaluation model (selected parts)

27 x 27 matrix

36 x 36 matrix

36 x 36 matrix

```
PEDIGREE    G am
RANDOM HTD H P

MILK(1)     = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
PROTEIN(1)  = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
FAT(1)      = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)

MILK(2)     = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
PROTEIN(2)  = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
FAT(2)      = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)

MILK(3)     = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
PROTEIN(3)  = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
FAT(3)      = H(t2 t3 t5 | HC2Y) P(t1..t5 | ani) G(t1..t5 | ani)
```

Herd specific lactation curve

Non-hereditary animal effect

Dimension of original co-variance matrices

Effect	Dimension
Herd-wise lactation curve (2 year-classes)	27
Non-hereditary **animal** effect	36
Genetic **animal** effect	36

- In the evaluation model the rank of these effects is reduced
- Details of rank reduction and the model :
  Across-Country Test-Day Model Evaluations for Nordic Holstein, Red Cattle, and Jersey. Lidauer & al. *J. Dairy. Sci.* (accepted, not published)

Genetic effect in the evaluation model after rank reduction

```
INTEGER  animal  HERD  DIM  COUNTRY  TRTGRP  HY  HTM  PROD_Y  PROD_YM AGEx5Y ...
REAL MILK PROT FAT ...
TABLEINDEX DIM
RANDOM HTM FIPE SDPE FIPE2
PEDIGREE ADDANI  am+p  1.0
TRAITGROUP TRTGRP
#first lact.
MILK(1) = ADDANI( t192 ... t206 | animal )@anml1
PROT(1) = ADDANI( t207 ... t221 | animal )@anml1
FAT(1)  = ADDANI( t222 ... t236 | animal )@anml1

#second lact.
MILK(2) = ADDANI( t237 ... t251 | animal )@anml1
PROT(2) = ADDANI( t252 ... t266 | animal )@anml1
FAT(2)  = ADDANI( t267 ... t281 | animal )@anml1

#third lact.
MILK(3) = ADDANI( t282 ... t296 | animal )@anml1
PROT(3) = ADDANI( t297 ... t311 | animal )@anml1
FAT(3)  = ADDANI( t312 ... t326 | animal )@anml1
```

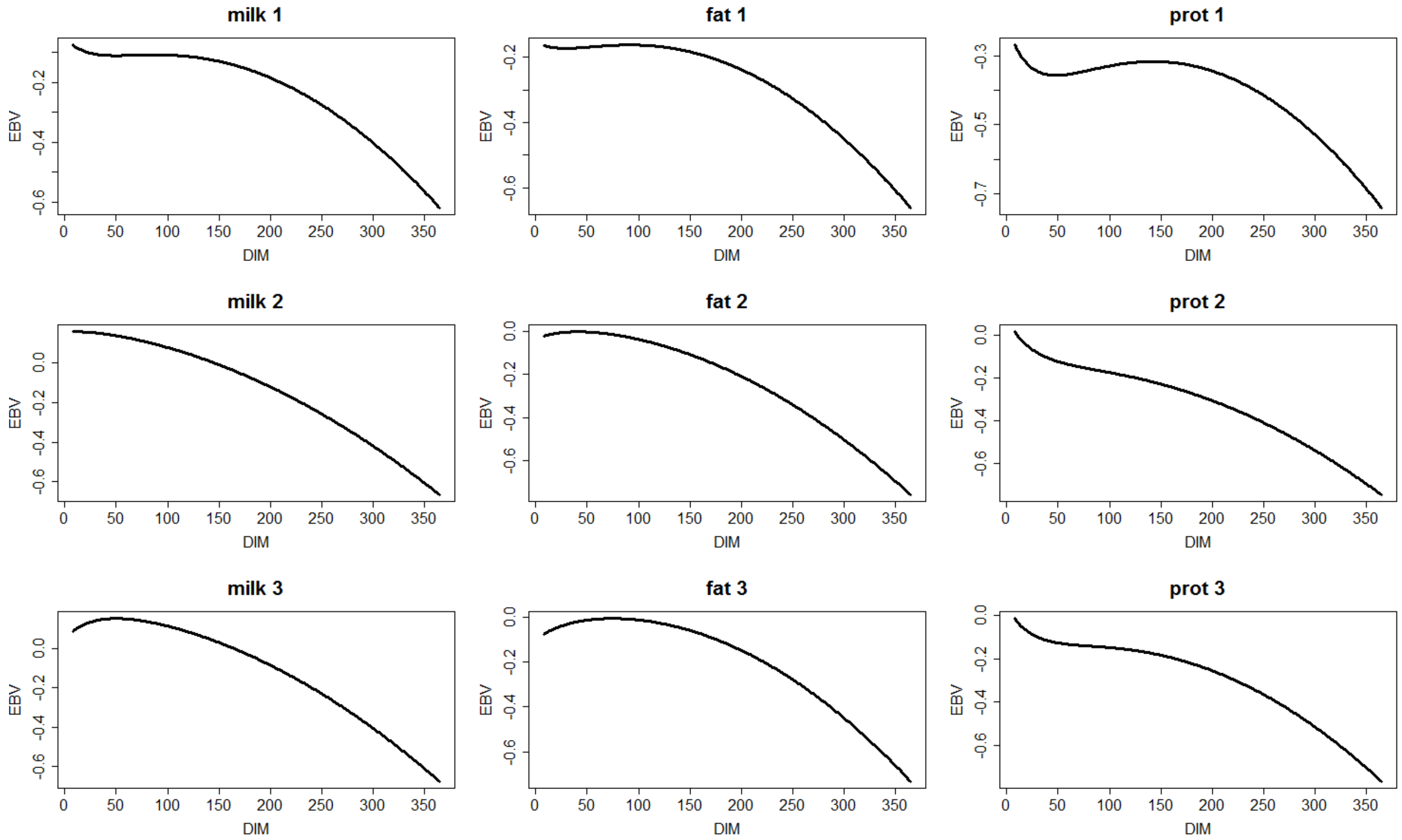
15 covariables, different covariables for each parity×trait

- Equations for genetic lactation curves are merged over traits
- Gen. level L for animal *i* at dim *d* lactation 1

$$\begin{bmatrix} L^{M}_{id} \\ L^{P}_{id} \\ L^{F}_{id} \end{bmatrix} = \begin{bmatrix} t_d^{192} & \cdots & t_d^{206} \\ t_d^{207} & \cdots & t_d^{221} \\ t_d^{222} & \cdots & t_d^{236} \end{bmatrix} \begin{bmatrix} u_{i_1} \\ \vdots \\ u_{i_{15}} \end{bmatrix}$$

- $\mathrm{VAR}(u) = \boldsymbol{I}_{15}$ (originally unstructured 36×36 matrix)

Genetic lactation curves for a random cow



Genetic effect in the evaluation model

- New co-variables are created by fitting ”**covariance functions**” to original covariance matrix of daily breeding values
- All co-variances between and within traits and lactations are modelled through co-variables
- Genetic correlation between countries is assumed to be 1
- Each country has **identical** set of co-variables
- Rank of G matrix is reduced from 36 to 15 (-58%)
- After rank reduction and covariance function fitting the G matrix is an identity matrix

Rank reduction of a non-hereditary effect

- New non-hereditary effect was created by combining variation explained by **herd specific (H)** and non-genetic **lactation curves (P)** and **error (E)**
- Lactation was divided into **12 intervals** and separate 3×3 **error co-variance** matrix was estimated for each interval
- Variation **between intervals** was moved to the **new** non-hereditary effect
- As result, only **one measurement error** co-variance matrix is used for each trait group

Non-heredity effect in the evaluation model

```
#DNK1
MILK(2) = SDPE(  t90 ..  t98 | animal )@1
PROT(2) = SDPE(  t99 .. t107 | animal )@1
FAT(2)  = SDPE( t108 .. t116 | animal )@1
#DNK 2
MILK(5) = SDPE( t117 .. t125 | animal )@2
PROT(5) = SDPE( t126 .. t134 | animal )@2
FAT(5)  = SDPE( t135 .. t143 | animal )@2
#DNK 3
MILK(8) = SDPE( t144 .. t152 | animal )@3
PROT(8) = SDPE( t153 .. t161 | animal )@3
FAT(8)  = SDPE( t162 .. t170 | animal )@3
```

Each trait is modelled with 9 parameters

- Equations for non-hereditary lactation curves are merged within lactations
- Non-gen. level P for animal *i* at dim *d*

$$\begin{bmatrix} P_{id}^{M1} \\ P_{id}^{P1} \\ P_{id}^{F1} \end{bmatrix} = \begin{bmatrix} t_d^{90} & \cdots & t_d^{98} \\ t_d^{99} & \cdots & t_d^{107} \\ t_d^{108} & \cdots & t_d^{116} \end{bmatrix} \begin{bmatrix} u_{i_1}^1 \\ \vdots \\ u_{i_9}^1 \end{bmatrix}$$

$$\begin{bmatrix} P_{id}^{M2} \\ P_{id}^{P2} \\ P_{id}^{F2} \end{bmatrix} = \begin{bmatrix} t_d^{117} & \cdots & t_d^{125} \\ t_d^{126} & \cdots & t_d^{134} \\ t_d^{135} & \cdots & t_d^{143} \end{bmatrix} \begin{bmatrix} u_{i_1}^2 \\ \vdots \\ u_{i_9}^2 \end{bmatrix}$$

$$\begin{bmatrix} P_{id}^{M3} \\ P_{id}^{P3} \\ P_{id}^{F3} \end{bmatrix} = \begin{bmatrix} t_d^{144} & \cdots & t_d^{152} \\ t_d^{153} & \cdots & t_d^{161} \\ t_d^{162} & \cdots & t_d^{170} \end{bmatrix} \begin{bmatrix} u_{i_1}^3 \\ \vdots \\ u_{i_9}^3 \end{bmatrix}$$

$$\mathrm{Var} \begin{bmatrix} u_{i_1}^1 \\ \vdots \\ u_{i_9}^3 \end{bmatrix} = \text{Full } 27 \times 27 \text{ matrix}$$

Dimension reduction summary

Effect	Original dimension	Reduced
Herd-wise lactation curve	27	Explained by non-hereditary effect
Non-heredity **animal** effect	36	27
Genetic **animal** effect	36	15

- Rank of genetic effect is reduced by 58%
- Rank of non-hereditary effect is reduced by 25%
- Non-hereditary effect explains also herd-wise lactation curve and part of error variation
- The total number of equations needed is reduced by ~244 million in test-day model

Structure of MME and memory requirements

```
 S T R U C T U R E   O F   M M E                    REDUCED   FULL(approx)
 ===================================================================
   Number of General Regression Equations ...            45
   Number of Across-Block Fixed Equations ...        49 501
   Number of Within-Block Fixed Equations ...     4 624 344
   Number of Across-Block Random Equations ..             0
   Number of Within-Block Random Equations .. 243 090 066   304 683 150
   Number of Animal Equations ................132 570 270   318 168 648
   Number of Phantom Parent Group Equations ..       5 745        13 788
   Total Number of Equations .................380 436 594   624 508 701 -40%

   Number of Animals in A-Inverse ............  8 838 018
   Number of Phantom Parent Groups ............        383


 Memory Requirements (Mb) for mix99s:        Reduced   Full (approx)
---------------------------------------------------------------------
          Data and Model Information .....       10       ?
          I/O-Buffer .....................       18       ?
          Vectors ........................   11 612     19 353
                                             ------     -------
          Total Allocated Memory .........   11 640     19 381++
```

MiX99 DEMO: Test-day model

- Test-day observations from 100 herds
- Milk, protein and fat observations
- Lactations 1 to 3
- Two milking systems: robot and conventional