MiX99

Solving Large Mixed Model Equations

Approximate reliabilities and MiX99

Ismo Strandén, MTT

Talk outline

- Approximation is 2 steps
- Some simple examples
- Random regression model example
- What is the difference

Approximation of reliabilities by ApaX

- Need elements from inverse of coefficient matrix of the mixed model equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix}^{-1} \quad \text{or} \quad \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix}^{-1}$$

- Approximations split the work into 2 steps:
  - Approximate amount of information due to records
  - Approximate amount of information due to pedigree information

- ➔ 2 steps in ApaX
  - 1st step is the same: record information
  - 2nd step can be chosen: pedigree information

The first step

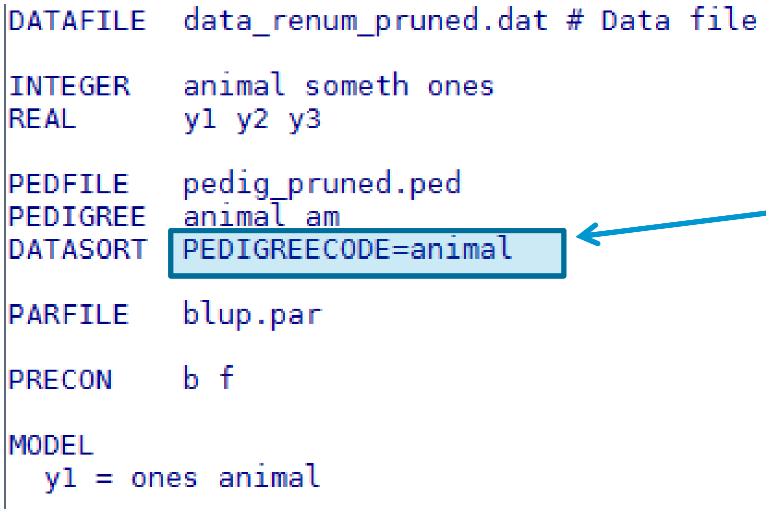
- Approximates information in estimation of contemporary groups

- Idea: all other effects have so much information that reliability is hardly affected
  - Effect having lowest estimation accuracy is accounted == contemporary group effect such as herd-year

- Memory is saved by requiring contemporary group to be a **within block effect**
  - all within blocks effects are used but in practice there **should be only one within block effect by trait**
    - If not, then these effects may cancel each other in the approximation → singularity warnings

Example:CLIM file and ApaX instructions

CLIM:



```
DATAFILE   data_renum_pruned.dat # Data file

INTEGER    animal someth ones
REAL       y1 y2 y3

PEDFILE    pedig_pruned.ped
PEDIGREE   animal am
DATASORT   PEDIGREECODE=animal

PARFILE    blup.par

PRECON     b f

MODEL
  y1 = ones animal
```

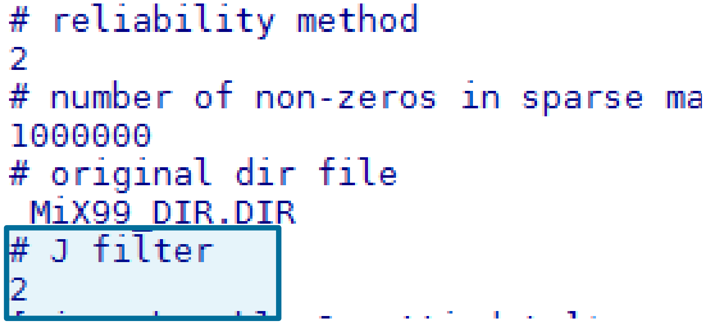
No withinblocks: no fixed effect accounting!

The only within block effect is ’animal’

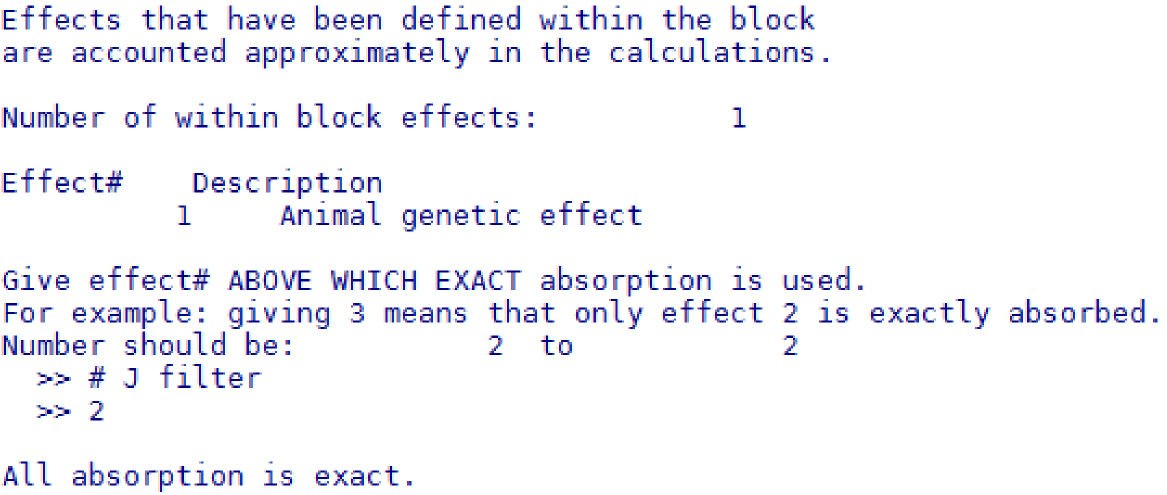
PEDIGREECODE needed: this tells Apax which is animal genetics;

NOTE: maternal effect models is case clearly indicates importance of iformation

ApaX:

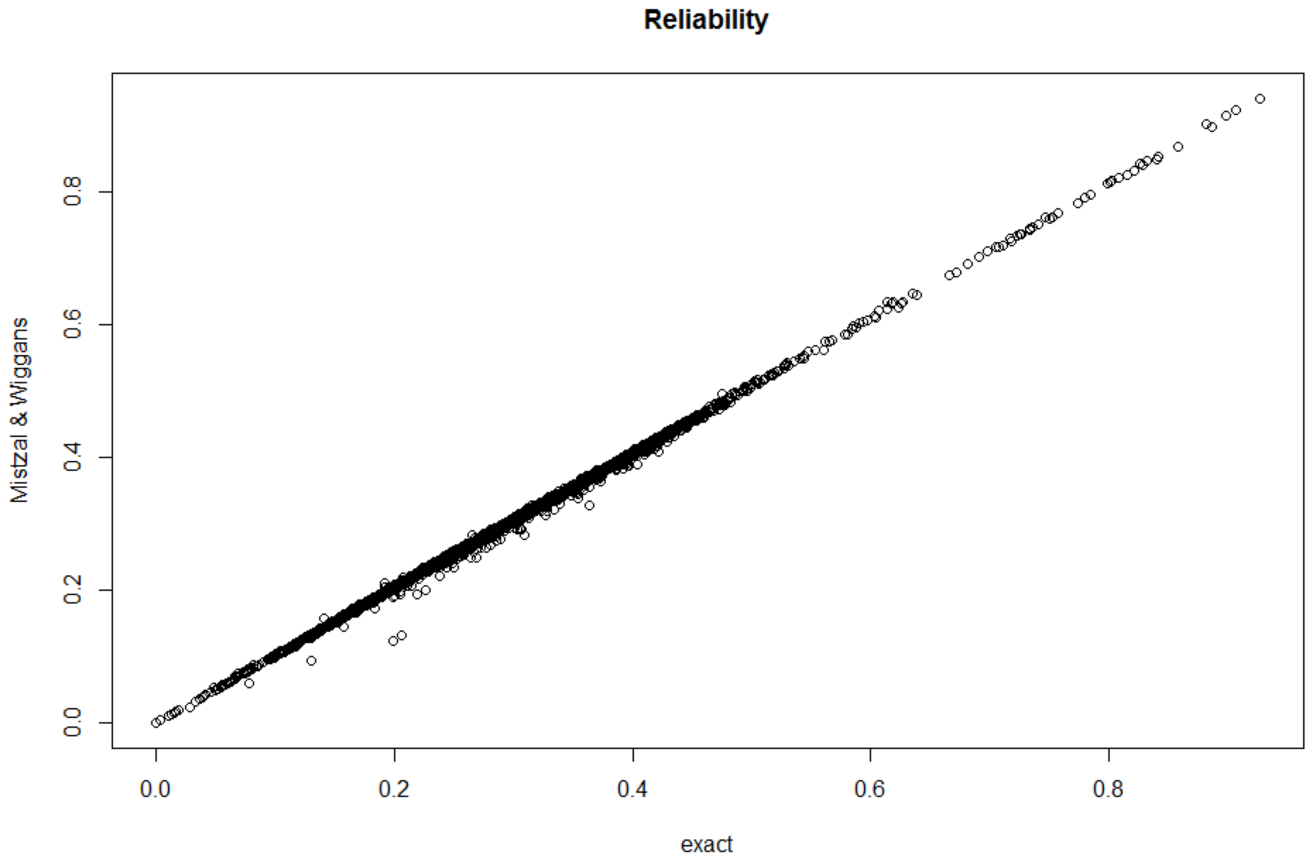


```
# reliability method
2
# number of non-zeros in sparse ma
1000000
# original dir file
 MiX99_DIR.DIR
# J filter
2
```

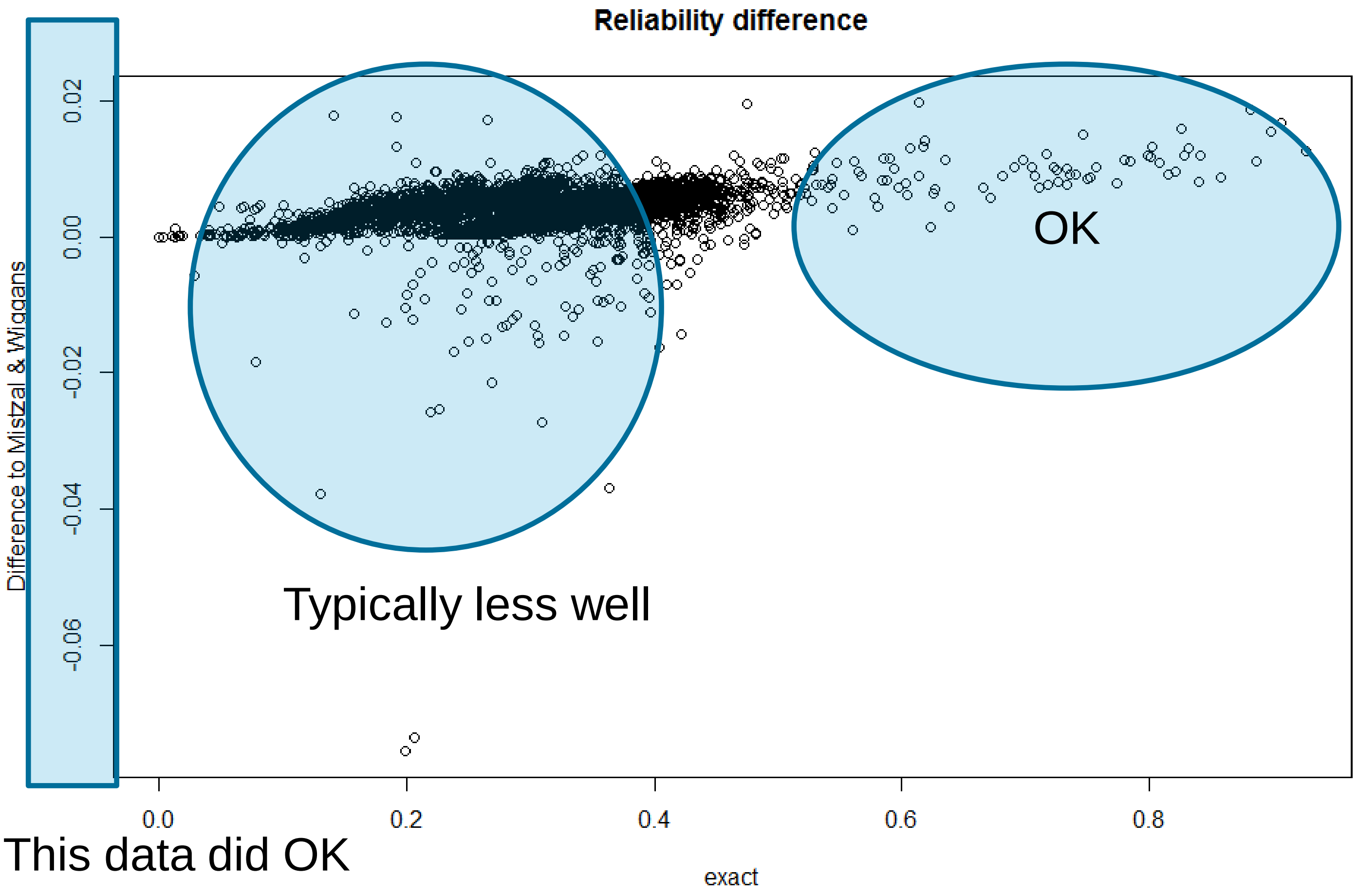


```
Effects that have been defined within the block
are accounted approximately in the calculations.

Number of within block effects:           1

Effect#    Description
          1     Animal genetic effect

Give effect# ABOVE WHICH EXACT absorption is used.
For example: giving 3 means that only effect 2 is exactly absorbed.
Number should be:           2  to            2
  >> # J filter
  >> 2

All absorption is exact.
```

Misztal&Wiggans (mthd 2) vs. exact



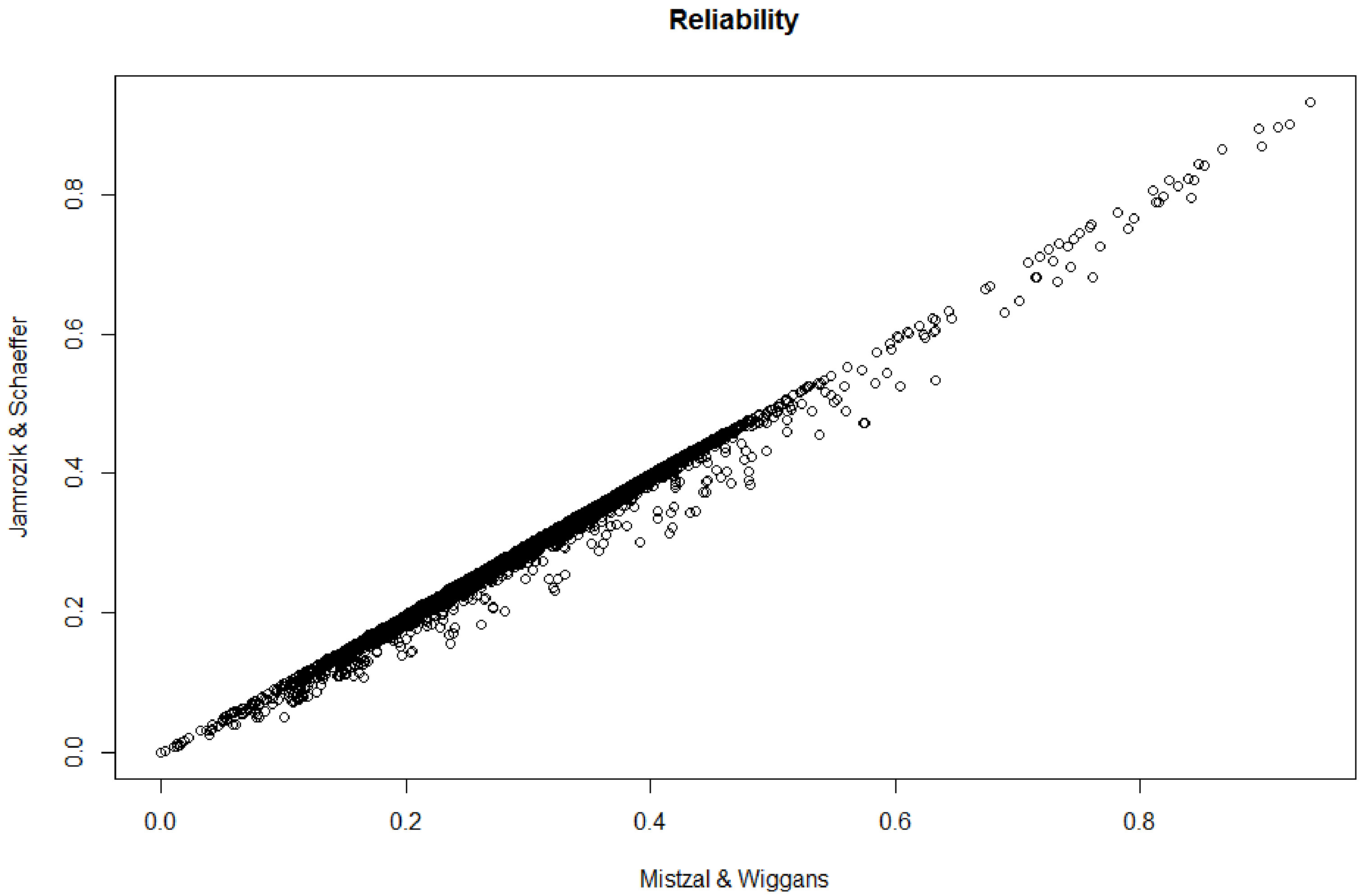
Difference: r^2(M&W) – r^2(exact)



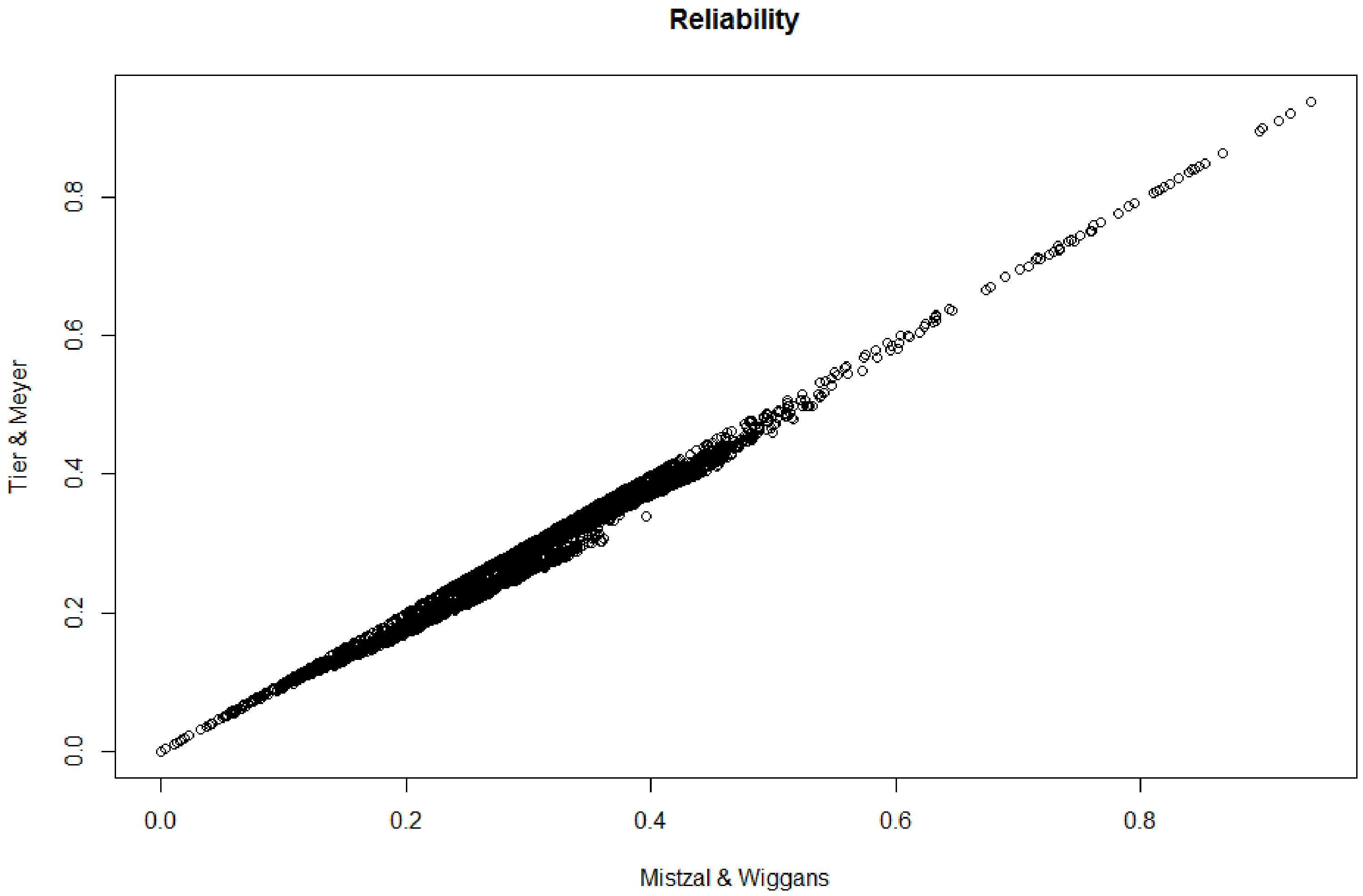
Typically less well

This data did OK

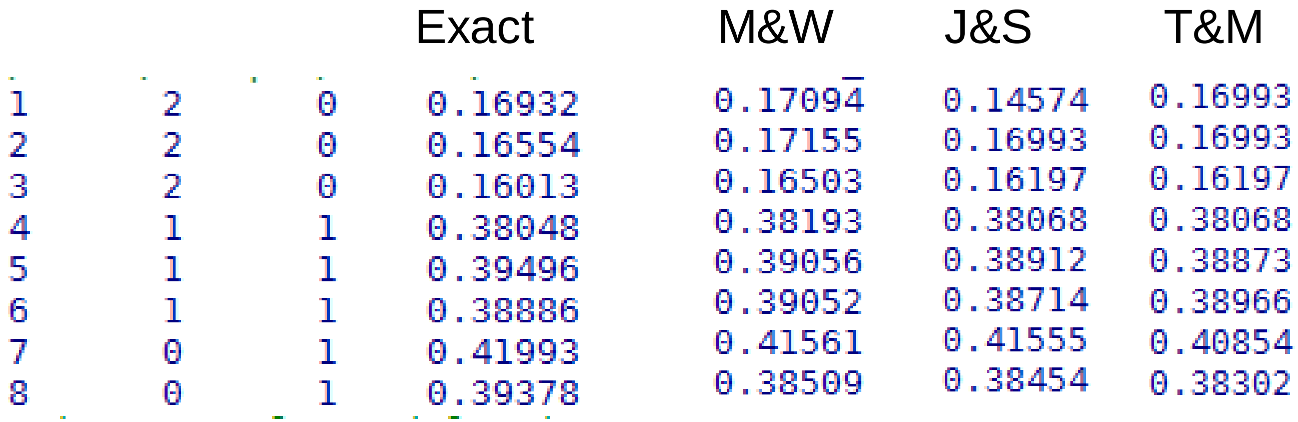
Misztal&Wiggans vs. Jamrozik&Schaeffer



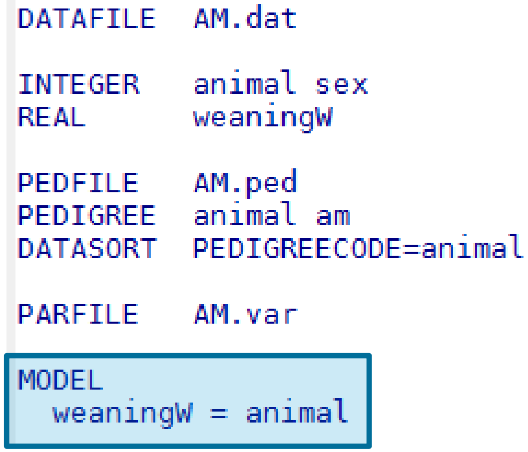
Misztal&Wiggans vs. Tier&Meyer



Simple animal model example

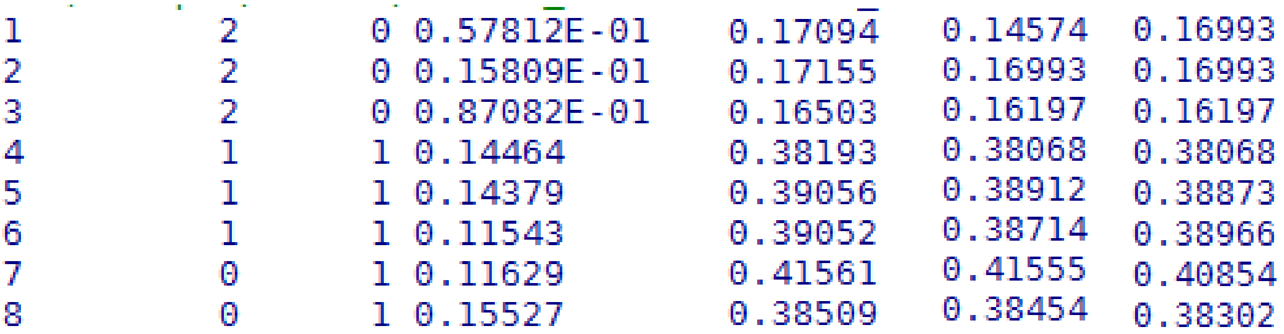


```
                              Exact       M&W       J&S       T&M
1       2       0       0.16932     0.17094   0.14574   0.16993
2       2       0       0.16554     0.17155   0.16993   0.16993
3       2       0       0.16013     0.16503   0.16197   0.16197
4       1       1       0.38048     0.38193   0.38068   0.38068
5       1       1       0.39496     0.39056   0.38912   0.38873
6       1       1       0.38886     0.39052   0.38714   0.38966
7       0       1       0.41993     0.41561   0.41555   0.40854
8       0       1       0.39378     0.38509   0.38454   0.38302
```

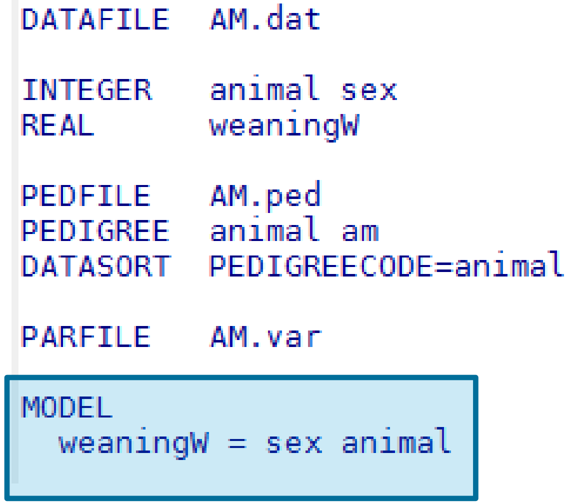


```
DATAFILE   AM.dat

INTEGER    animal sex
REAL       weaningW

PEDFILE    AM.ped
PEDIGREE   animal am
DATASORT   PEDIGREECODE=animal

PARFILE    AM.var

MODEL
  weaningW = animal
```

When only additive genetic effect, approximations work.



```
1          2        0 0.57812E-01    0.17094   0.14574   0.16993
2          2        0 0.15809E-01    0.17155   0.16993   0.16993
3          2        0 0.87082E-01    0.16503   0.16197   0.16197
4          1        1 0.14464        0.38193   0.38068   0.38068
5          1        1 0.14379        0.39056   0.38912   0.38873
6          1        1 0.11543        0.39052   0.38714   0.38966
7          0        1 0.11629        0.41561   0.41555   0.40854
8          0        1 0.15527        0.38509   0.38454   0.38302
```

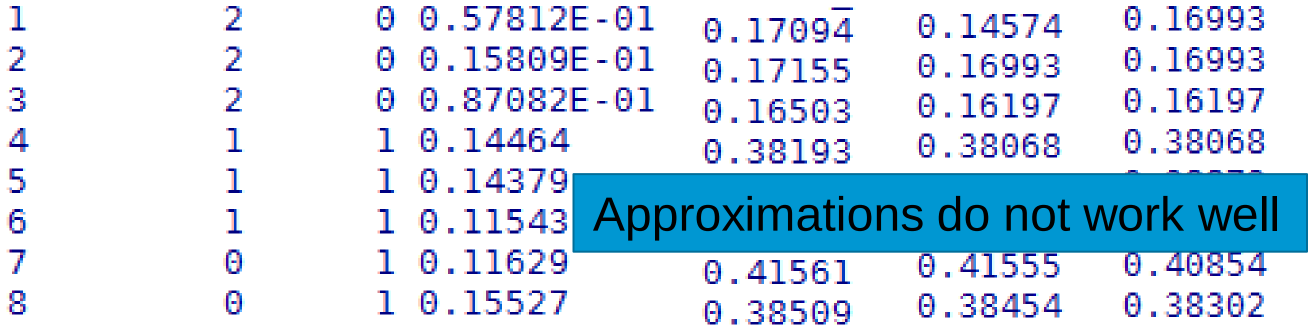


```
DATAFILE   AM.dat

INTEGER    animal sex
REAL       weaningW

PEDFILE    AM.ped
PEDIGREE   animal am
DATASORT   PEDIGREECODE=animal

PARFILE    AM.var

MODEL
  weaningW = sex animal
```

The added fixed effect not accounted in the approximations.
Need to define it as a within block effect.

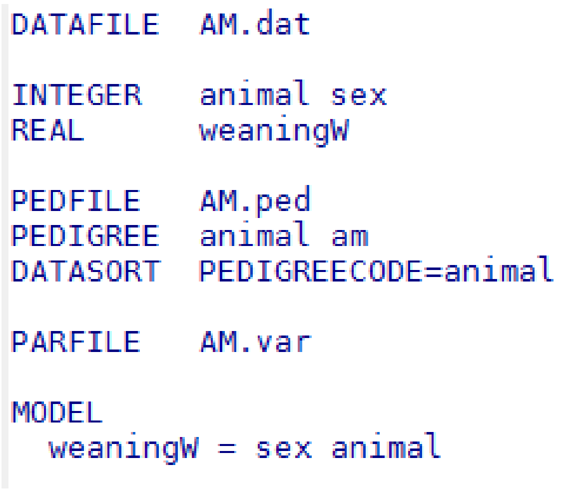
Simple animal model example

Exact M&W J&S T&M



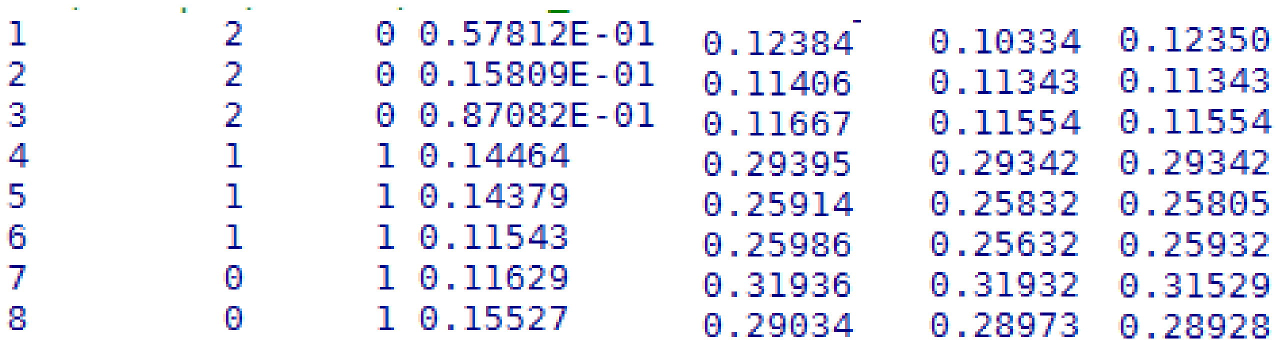
```
1     2     0 0.57812E-01   0.17094   0.14574   0.16993
2     2     0 0.15809E-01   0.17155   0.16993   0.16993
3     2     0 0.87082E-01   0.16503   0.16197   0.16197
4     1     1 0.14464       0.38193   0.38068   0.38068
5     1     1 0.14379
6     1     1 0.11543
7     0     1 0.11629       0.41561   0.41555   0.40854
8     0     1 0.15527       0.38509   0.38454   0.38302
```

Approximations do not work well

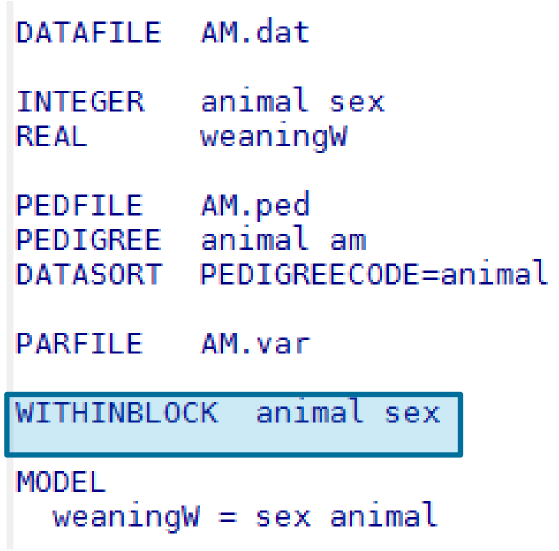


```
DATAFILE   AM.dat

INTEGER    animal sex
REAL       weaningW

PEDFILE    AM.ped
PEDIGREE   animal am
DATASORT   PEDIGREECODE=animal

PARFILE    AM.var

MODEL
  weaningW = sex animal
```

No fixed effect accounted in reliability approximation



```
1     2     0 0.57812E-01   0.12384   0.10334   0.12350
2     2     0 0.15809E-01   0.11406   0.11343   0.11343
3     2     0 0.87082E-01   0.11667   0.11554   0.11554
4     1     1 0.14464       0.29395   0.29342   0.29342
5     1     1 0.14379       0.25914   0.25832   0.25805
6     1     1 0.11543       0.25986   0.25632   0.25932
7     0     1 0.11629       0.31936   0.31932   0.31529
8     0     1 0.15527       0.29034   0.28973   0.28928
```

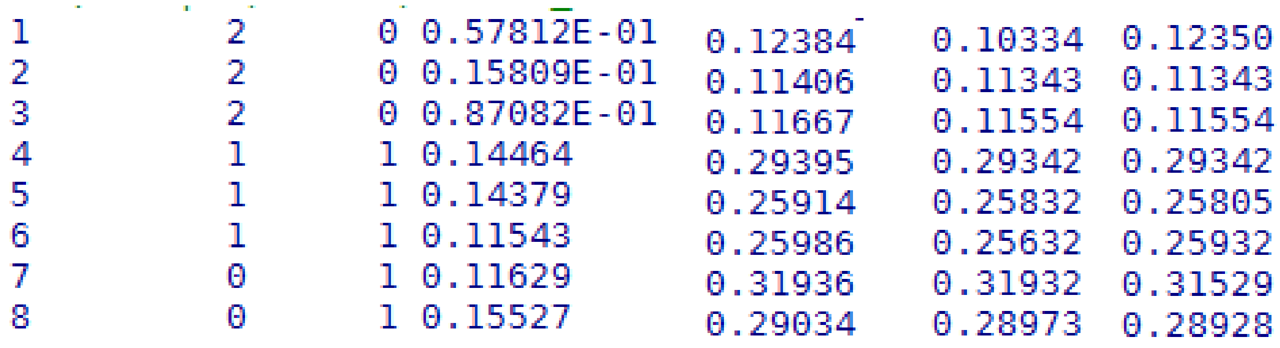


```
DATAFILE   AM.dat

INTEGER    animal sex
REAL       weaningW

PEDFILE    AM.ped
PEDIGREE   animal am
DATASORT   PEDIGREECODE=animal

PARFILE    AM.var

WITHINBLOCK  animal sex

MODEL
  weaningW = sex animal
```

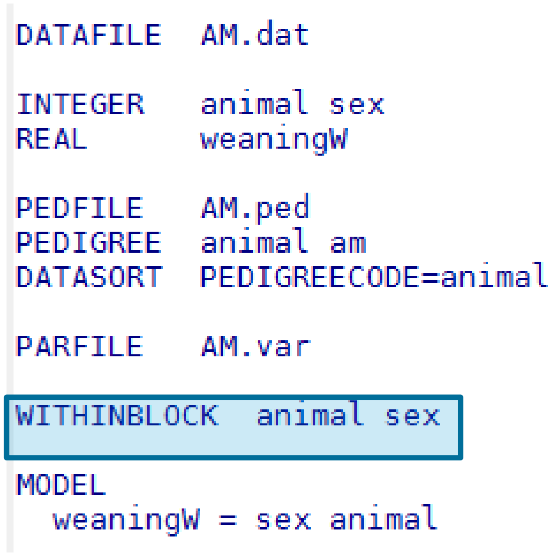
When fixed effect is account,
approximation do better but still different from exact.

Simple animal model example

Exact M&W J&S T&M

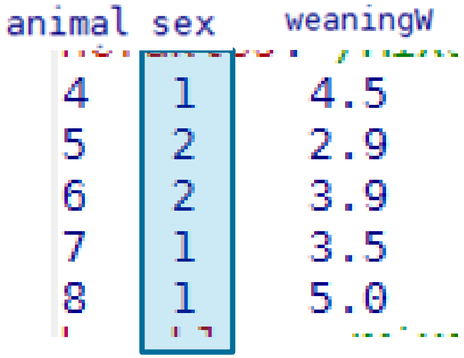


```
1      2     0 0.57812E-01  0.12384  0.10334  0.12350
2      2     0 0.15809E-01  0.11406  0.11343  0.11343
3      2     0 0.87082E-01  0.11667  0.11554  0.11554
4      1     1 0.14464      0.29395  0.29342  0.29342
5      1     1 0.14379      0.25914  0.25832  0.25805
6      1     1 0.11543      0.25986  0.25632  0.25932
7      0     1 0.11629      0.31936  0.31932  0.31529
8      0     1 0.15527      0.29034  0.28973  0.28928
```



```
DATAFILE    AM.dat

INTEGER     animal sex
REAL        weaningW

PEDFILE     AM.ped
PEDIGREE    animal am
DATASORT    PEDIGREECODE=animal

PARFILE     AM.var

WITHINBLOCK   animal sex

MODEL
  weaningW = sex animal
```

Data:



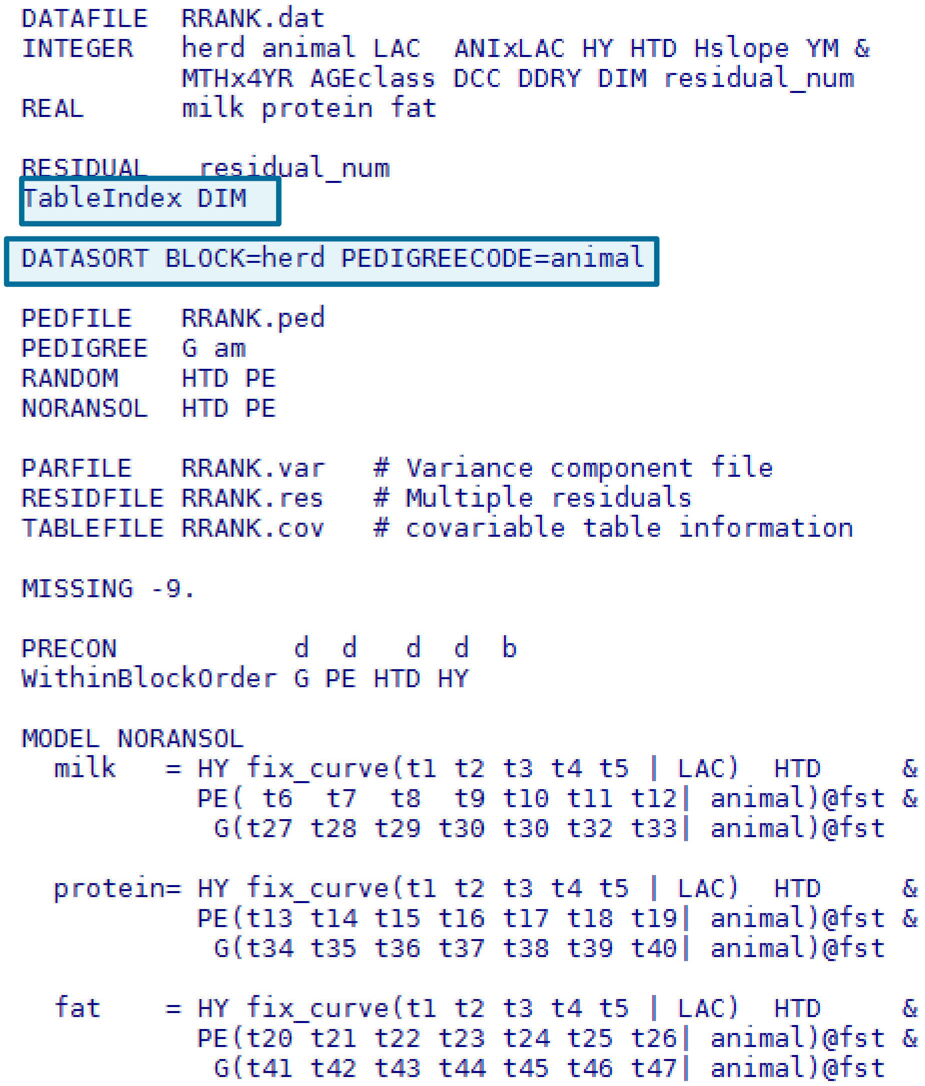
```
animal sex   weaningW
4      1     4.5
5      2     2.9
6      2     3.9
7      1     3.5
8      1     5.0
```

Fixed effect has not many observations by level → not penalized enough → approximations give too large reliabilities.

Random regression models

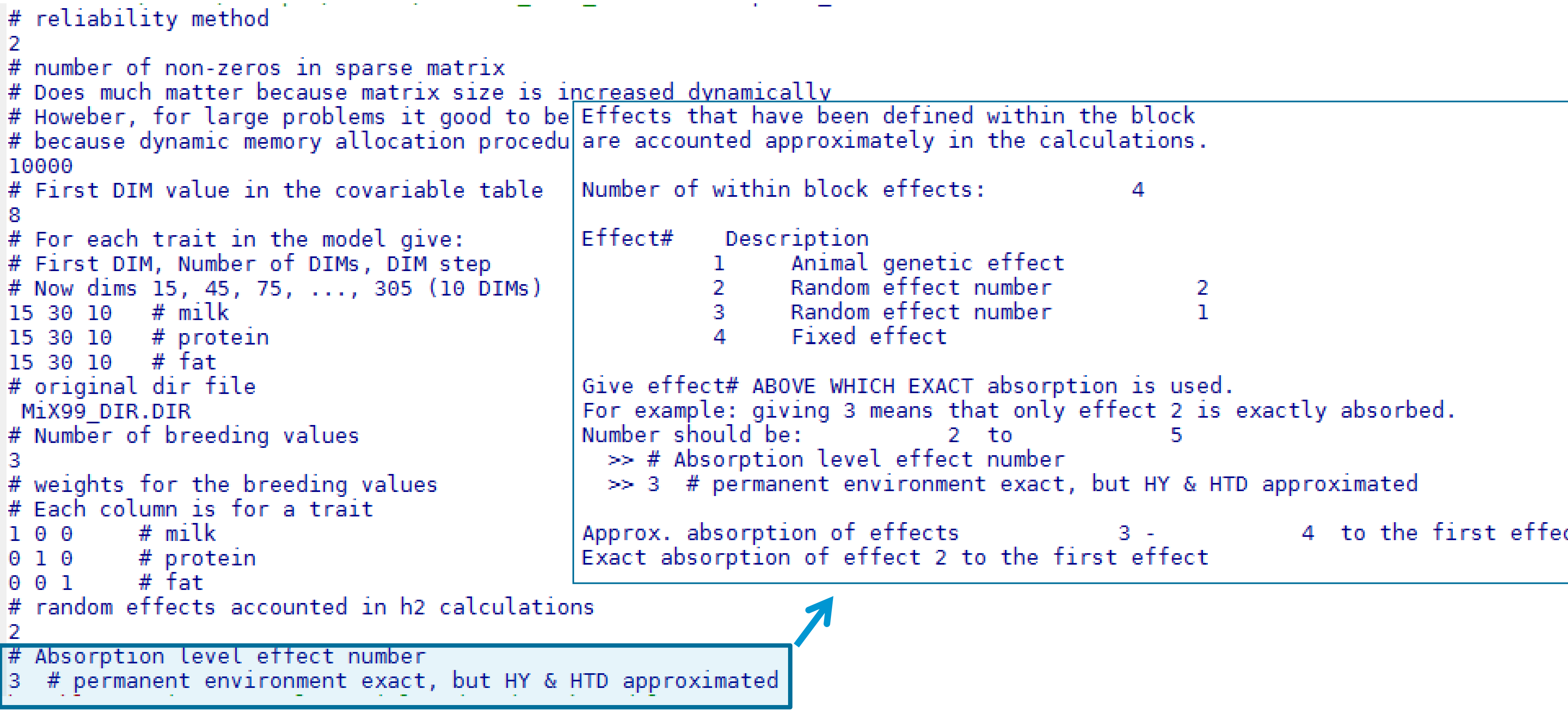
- Interest is in functions not values of random regression
- Functions can be used only by table file
  - Need values of function at many instances (eg. by DIM)
  - Table file tells values of function at different values
- ApaX needs to be told which values to use
  - First instance
  - Number of instances
  - Step

CLIM file



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

RESIDUAL   residual_num
TableIndex DIM

DATASORT BLOCK=herd PEDIGREECODE=animal

PEDFILE   RRANK.ped
PEDIGREE  G am
RANDOM    HTD PE
NORANSOL  HTD PE

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

MISSING -9.

PRECON           d  d   d  d  b
WithinBlockOrder G PE HTD HY

MODEL NORANSOL
  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 t30 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
```

ApaX instruction file



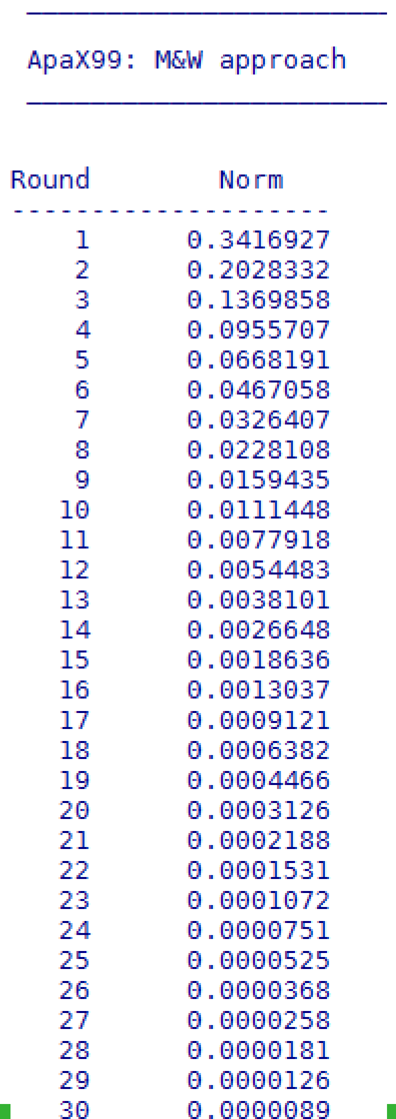
```
# reliability method
2
# number of non-zeros in sparse matrix
# Does much matter because matrix size is increased dynamically
# Howeber, for large problems it good to be
# because dynamic memory allocation procedu
10000
# First DIM value in the covariable table
8
# For each trait in the model give:
# First DIM, Number of DIMs, DIM step
# Now dims 15, 45, 75, ..., 305 (10 DIMs)
15 30 10    # milk
15 30 10    # protein
15 30 10    # fat
# original dir file
 MiX99_DIR.DIR
# Number of breeding values
3
# weights for the breeding values
# Each column is for a trait
1 0 0     # milk
0 1 0     # protein
0 0 1     # fat
# random effects accounted in h2 calculations
2
# Absorption level effect number
3  # permanent environment exact, but HY & HTD approximated
```

```
Effects that have been defined within the block
are accounted approximately in the calculations.

Number of within block effects:            4

Effect#    Description
          1     Animal genetic effect
          2     Random effect number           2
          3     Random effect number           1
          4     Fixed effect

Give effect# ABOVE WHICH EXACT absorption is used.
For example: giving 3 means that only effect 2 is exactly absorbed.
Number should be:           2  to            5
  >> # Absorption level effect number
  >> 3  # permanent environment exact, but HY & HTD approximated

Approx. absorption of effects           3 -           4  to the first effec
Exact absorption of effect 2 to the first effect
```

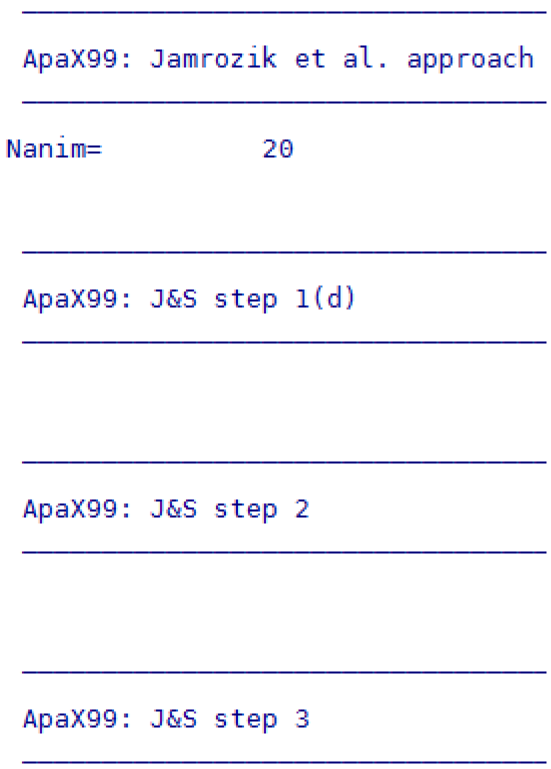
Controls level of approximation in the first step.

Now: HTD and HY were approximately accounted.

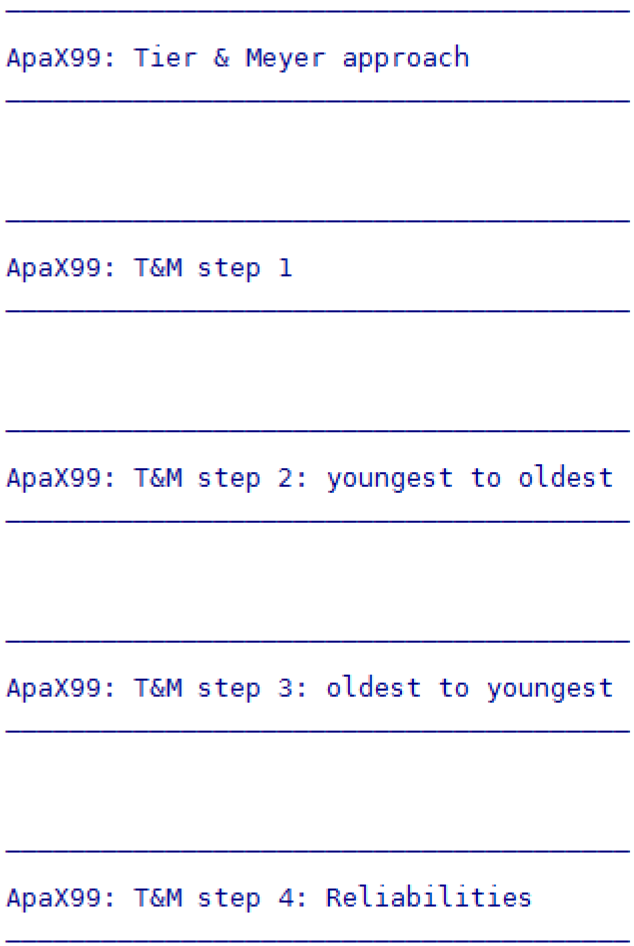
Different steps by method



```
_______________________

ApaX99: M&W approach
_______________________


Round          Norm
--------------------
    1       0.3416927
    2       0.2028332
    3       0.1369858
    4       0.0955707
    5       0.0668191
    6       0.0467058
    7       0.0326407
    8       0.0228108
    9       0.0159435
   10       0.0111448
   11       0.0077918
   12       0.0054483
   13       0.0038101
   14       0.0026648
   15       0.0018636
   16       0.0013037
   17       0.0009121
   18       0.0006382
   19       0.0004466
   20       0.0003126
   21       0.0002188
   22       0.0001531
   23       0.0001072
   24       0.0000751
   25       0.0000525
   26       0.0000368
   27       0.0000258
   28       0.0000181
   29       0.0000126
   30       0.0000089
```

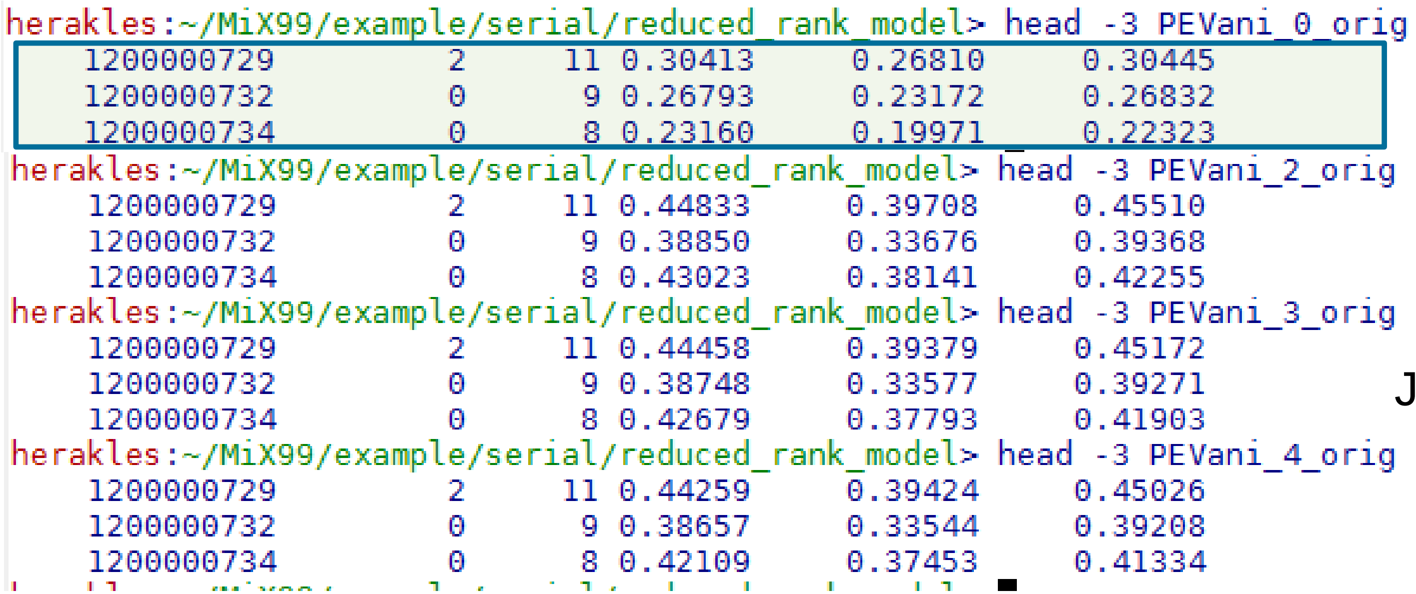


```
_________________________________

ApaX99: Jamrozik et al. approach
_________________________________

Nanim=          20


_________________________________

ApaX99: J&S step 1(d)
_________________________________



_________________________________

ApaX99: J&S step 2
_________________________________



_________________________________

ApaX99: J&S step 3
_________________________________
```



```
_______________________________________

ApaX99: Tier & Meyer approach
_______________________________________



_______________________________________

ApaX99: T&M step 1
_______________________________________



_______________________________________

ApaX99: T&M step 2: youngest to oldest
_______________________________________



_______________________________________

ApaX99: T&M step 3: oldest to youngest
_______________________________________



_______________________________________

ApaX99: T&M step 4: Reliabilities
_______________________________________
```

Solutions



```
herakles:~/MiX99/example/serial/reduced_rank_model> head -3 PEVani_0_orig
    1200000729          2       11 0.30413      0.26810      0.30445
    1200000732          0        9 0.26793      0.23172      0.26832
    1200000734          0        8 0.23160      0.19971      0.22323
```

Exact

```
herakles:~/MiX99/example/serial/reduced_rank_model> head -3 PEVani_2_orig
    1200000729          2       11 0.44833      0.39708      0.45510
    1200000732          0        9 0.38850      0.33676      0.39368
    1200000734          0        8 0.43023      0.38141      0.42255
```

Misztal & Wiggans

```
herakles:~/MiX99/example/serial/reduced_rank_model> head -3 PEVani_3_orig
    1200000729          2       11 0.44458      0.39379      0.45172
    1200000732          0        9 0.38748      0.33577      0.39271
    1200000734          0        8 0.42679      0.37793      0.41903
```

Jamrozik&Schaeffer

```
herakles:~/MiX99/example/serial/reduced_rank_model> head -3 PEVani_4_orig
    1200000729          2       11 0.44259      0.39424      0.45026
    1200000732          0        9 0.38657      0.33544      0.39208
    1200000734          0        8 0.42109      0.37453      0.41334
```

Tier&Meyer

Approximations have their price: fast but reliabilities often larger than exact.

ApaX conclusions

- ApaX combines information from 2 sources:
  - Records
  - Pedigree

- Approximations in both sources of information

- Approxmations do not produce exact solutions

- Approximations tend to work well when data is dense:
  - Number of observations per fixed effect level large
  - Number of animals with observations large

- Approximations usually better when reliability is large