AARHUS
UNIVERSITY

CENTER FOR QUANTITATIVE
GENETICS AND GENOMICS QGG

Modelling of grouped recorded data for different population structures in pig breeding

Mahmoud Shirali[1], Bjarne Nielsen[2], Tage Ostersen[2], Ole Christensen[1], Per Madsen[1], Guosheng Su[1], Just Jensen[1]

[1] Center for Quantitative Genetics & Genomics, Aarhus University, Denmark
[2] SEGES, Pig Research Centre, Denmark

Dan Avl

Miljø- og
Fødevareministeriet
gudp

LANDBRUG FØDEVARER SEGES

Objective: To investigate the effect of population structure in the data on grouped analysis:

- *Model group records with varying group sizes*
- *Considering multiple random effects*
- *Asses genetic parameter and breeding values*

1) *Develop methodology based on analysing simulated data*
2) *Applying the method to real data*

Modelling pooled data with varying group size and different random effects

- **Individual record:**

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_{\mathbf{l}}\mathbf{l} + \mathbf{Z}_{\mathbf{c}}\mathbf{c} + \mathbf{Z}_{\mathbf{a}}\mathbf{a} + \mathbf{e}$$

- **Group record:**

$$\mathbf{T}\mathbf{y} = \mathbf{T}(\mathbf{X}\mathbf{b} + \mathbf{Z}_{\mathbf{l}}\mathbf{l} + \mathbf{Z}_{\mathbf{c}}\mathbf{c} + \mathbf{Z}_{\mathbf{a}}\mathbf{a} + \mathbf{e})$$

$$\sigma_{a*}^2 = \mathbf{z}_{\mathrm{a}}^{*}\mathbf{A}\mathbf{z}_{\mathrm{a}}^{*'}\sigma_a^2 = \sum_{j=1}^{n_g}\mathbf{A}_{\mathrm{jj}}\sigma_a^2 + n_g(n_g - 1)\mathbf{A}_{\mathrm{jj'}}\sigma_a^2$$

$$\sigma_{l*}^2 = \mathbf{z}_{\mathbf{l}}^{*}\mathbf{z}_{\mathbf{l}}^{*'}\sigma_l^2 = \sum_{k=1}^{N_L} n_{lk}^2\sigma_l^2$$

$$\sigma_{c*}^2 = \mathbf{z}_c^{*}\mathbf{z}_c^{*'}\sigma_c^2 = n_g^2\sigma_c^2$$

$$\sigma_{e*}^2 = n_g\sigma_e^2$$

Su et al. 2018

Real data

- **Low within pen relationship (Case 1)**
  - 23824 animals for analysis in 709 sections in 15 years
  - Majority 1 littermate per pen.
  - 47868 animals in pedigree
  - Group size : **8 to 14** pigs per pen

- **Large within pen relationship (Case 2)**
  - 100285 pigs with phenotype
  - 112959 pigs with pedigree
  - Group size : **6 to 15** pigs per pen

Simulation of data

- Pedigree & data structure from low and large within pen relationship data are used
- Breeding values

$$a_i = \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + m_i$$

- Mendelian sampling term has expected value of 0 and variance of

$$\mathrm{var}(m) = \tfrac{1}{2}\left[1 - \frac{(F_s + F_d)}{2}\right]\sigma_A^2$$

- No selection
- Taking into account inbreeding
- 10 replicates

	σ_a^2	σ_c^2	σ_l^2	σ_e^2	h^2
simulated	300	70	40	490	0.33

Simulation: genetic parameter estimation

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{Sl} + \mathbf{e}$$

$$\mathbf{y}^* = \mathbf{X}^*\mathbf{b} + \mathbf{Z}^*\mathbf{a} + \mathbf{W}^*\mathbf{c} + \mathbf{S}^*\mathbf{l} + \mathbf{e}^*$$

Scenario	**Analysis**	σ_a^2	σ_c^2	σ_l^2	σ_e^2
	Simulated	300	70	40	490
Case 1	Individual	298 (18)	71 (5)	44 (11)	490 (13)
	Pooled	241 (60)	70 (26)	76 (91)	495 (254)
Case 2	Individual	298 (15)	71 (2)	40 (3)	492 (6)
	Pooled	302 (43)	72 (7)	31 (16)	491 (75)

Simulation: accuracy of breeding value estimation

Scenario	Individual	Pooled
Case 1	0.67 (0.02)	0.34 (0.03)
Case 2	0.70 (0.01)	0.48 (0.02)

Real data analysis: low within pen

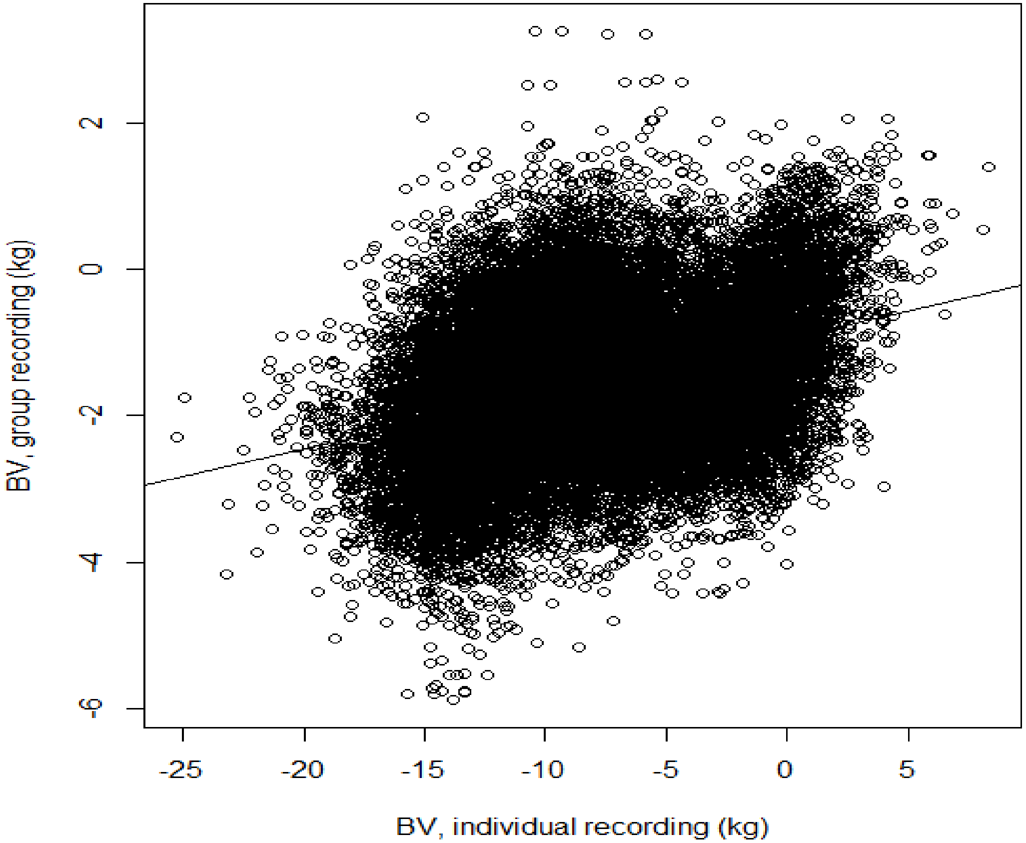
Modelling feed efficiency

$$y_{ij} = YHM_i + b_1 SBW_j + b_2 ADG_j + a_j + c_j + e_{ij}$$

$$y^*_{ij} = YHM^*_i + b^*_1 n^*_j + b^*_2 SBW^*_j + b^*_3 ADG^*_j + a^*_k + e^*_{ijk}$$

Trait	σ^2_c	σ^2_a	σ^2_e	**Total**	**h²**
Individual RFI	6.6	16.7	73.2	96.5	0.17
Sum group RFI	-	18.8	93.0	111.8	0.17

Accuracy: low within relationship



$$Cor(BV_{individual}, BV_{group}) = \mathbf{0.36}$$

Real data analysis: large within pen relationship

Modelling growth

$$y_{ijk} = YHM_i + SEX_j + a_k + c_k + l_k + e_{ijk}$$

$$y^*_{ijkl} = YHM^*_i + SEX^*_j + b^*_1 n^*_k + a^*_l + c^*_l + l^*_l + e^*_{ijkl}$$

Trait	**Analysis**	σ^2_a	σ^2_c	σ^2_l	σ^2_e
ADG	Individual	2691 (113)	694 (18)	424 (19)	4219 (51)
	Pooled	2909 (370)	434 (88)	484 (74)	5661 (864)

The accuracy of breeding value estimation = ***0.65***

Conclusions

- ***Developed model for pooled data considers:***
  - *multiple fixed and random effects*
  - *varying group sizes*

- ***Developed model can be used for genetic parameter estimation***
  - *With larger standard error*

- ***A close genetic relationship within a pen is favourable for breeding value estimation***

- ***Genetic parameter estimation with group records on feed intake and growth is feasible***