Estimating social genetic effects

Mixed model

- Assumed "true" model: $P_i = A_{D,i} + E_{D,i} + \sum_{i \neq j} A_{S,j} + \sum_{i \neq j} E_{S,j}$

- Mixed animal model: $\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{e}$
  - $\mathbf{Z}_D\mathbf{a}_D$ = direct genetic effects of self
  - $\mathbf{Z}_S\mathbf{a}_S$ = social genetic effects of group members

Example

- Mortality due to cannibalism in chickens
- 4 chickens per cage
- **Z**-matrices for two cages

$$\mathbf{Z}_D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Z}_S = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Mixed model

Example with 4 individuals in a group

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{e}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [\mu] + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_{D,1} \\ A_{D,2} \\ A_{D,3} \\ A_{D,4} \end{bmatrix} + \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} A_{S,1} \\ A_{S,2} \\ A_{S,3} \\ A_{S,4} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

The residual summarizes both the direct and social non-genetic effects

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} E_{D,1} + E_{S,2} + E_{S,3} + E_{S,4} \\ E_{D,2} + E_{S,1} + E_{S,3} + E_{S,4} \\ E_{D,3} + E_{S,1} + E_{S,2} + E_{S,4} \\ E_{D,4} + E_{S,1} + E_{S,2} + E_{S,3} \end{bmatrix}$$

Mixed model: ASREML

- How to fit social effects into AsReml?
  - Use the “and()” statement in the model line
  - “and()” adds-up the Z-matrices

```
Data file groupsel.dat
self, m1, m2, m3, pheno
1 2 3 4 10
2 1 3 4 12
3 1 2 4 9
4 1 2 3 10
5 6 7 8 13
6 5 7 8 9
```

```
Pedigree file groupsel.ped
self sire dam
1 S1 D1
2 S2 D2
3 S3 D3
4 S4 D4
5 S5 D5
6 S6 D6
```

```
group selection
 self !P
 mate1 !P
 mate2 !P
 mate3 !P
 ptype 1
groupsel.ped !MAKE !SKIP 2
groupsel.dat !maxiter=20 !SKIP 2
ptype ~ mu !r self mate1 and(mate2) and(mate3)
0 0 1
self 2
2 0 US !GP !+3
.1
0.01 .1
self 0 AINV
```

Mixed model: residual variance structure

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} E_{D,1} + E_{S,2} + E_{S,3} + E_{S,4} \\ E_{D,2} + E_{S,1} + E_{S,3} + E_{S,4} \\ E_{D,3} + E_{S,1} + E_{S,2} + E_{S,4} \\ E_{D,4} + E_{S,1} + E_{S,2} + E_{S,3} \end{bmatrix}$$

Can we simply fit a single residual?

↓

What is the variance-covariance structure that emerges for the residual?

$$Var(e_i) = Var(E_{D,i} + E_{S,j} + E_{S,k} + E_{S,l})$$

$$= \sigma^2_{E_D} + 3\sigma^2_{E_S} = \sigma^2_{E_D} + (n-1)\sigma^2_{E_S}$$

$$Cov(e_i, e_j)_{within\_grp} = Cov(E_{D,i} + E_{S,j} + E_{S,k} + E_{S,l}; E_{D,j} + E_{S,i} + E_{S,k} + E_{S,l})$$

$$= Cov(E_{D,i}, E_{S,i}) + Cov(E_{S,j}, E_{D,j}) + Cov(E_{S,k}, E_{S,k}) + Cov(E_{S,k}, E_{S,k})$$

$$= 2\sigma_{E_{DS}} + 2\sigma^2_{E_S} = 2\sigma_{E_{DS}} + (n-2)\sigma^2_{E_S}$$

Two individuals have (n-2) group members in common → hence the (n-2)Var(E_S)

Cov(e_i,e_j) = 0 between groups

Mixed model: residual variance structure

- Within group, residuals are correlated
- There exist three biological VC
  - $Var(E_D)$
  - $Cov(E_D,E_S)$
  - $Var(E_S)$
- Statistically, we find only two VC
  - Var(e)
  - $Cov(e_i,e_j)_{within\_grp}$
- Hence, we cannot uniquely estimate all three biological VC
- $Cov(e_i,e_j)_{within\_grp} = 2Cov(E_D,E_S) + (n-2)Var(E_S)$
  - This can be either negative or positive
  - Probably positive in large groups
- Account for $Cov(e_i,e_j)_{within\_grp} \rightarrow$ allow for correlated residuals

Residual variance structure

Two groups of 4 individuals each

$$Var(\mathbf{e}) = \mathbf{R}\sigma_e^2$$

with $R_{ii} = 1$

$R_{ij} = \rho$ when i and j are group members

$R_{ij} = 0$ when i and j are in different groups

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho & \rho & 0 & 0 & 0 & 0 \\ \rho & 1 & \rho & \rho & 0 & 0 & 0 & 0 \\ \rho & \rho & 1 & \rho & 0 & 0 & 0 & 0 \\ \rho & \rho & \rho & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \rho & \rho & \rho \\ 0 & 0 & 0 & 0 & \rho & 1 & \rho & \rho \\ 0 & 0 & 0 & 0 & \rho & \rho & 1 & \rho \\ 0 & 0 & 0 & 0 & \rho & \rho & \rho & 1 \end{bmatrix}$$

and

$$\sigma_e^2 = \sigma_{E_D}^2 + (n-1)\sigma_{E_S}^2$$

$$\rho = [2\sigma_{E_{DS}} + (n-2)\sigma_{E_S}^2] \Big/ [\sigma_{E_D}^2 + (n-1)\sigma_{E_S}^2]$$

Residual variance structure in ASREML

- Use the CORU statement in the R-structure definition
  - The starting value refers to ρ

```
Data file groupsel.dat
group,nr,self,m1,m2,m3,phero
1 1 2 3 4 10
1 2 2 1 3 4 12
1 3 3 1 2 4 9
1 4 4 1 2 3 10
2 1 5 6 7 8 13
2 2 6 5 7 8 9
2 3 7 5 6 8 12
2 4 8 5 6 7 14
```

Include group in the data file, and a consecutive nr within the group

```
group selection
 grp !I 2500
 nr !I 4
 self !P
 mate1 !P
 mate2 !P
 mate3 !P
 ptype 1
groupsel.ped !MAKE !SKIP 2
groupsel.dat !maxiter=20 !SKIP 2
ptype ~ mu !r self mate1 and(mate2) and(mate3)
1 2 1
2500 grp ID
4 nr CORU 0.1
self 2
2 0 US !GP !+3
.1
0.01 .1
self 0 AINV
```

Drawback: correlated residuals are computationally demanding and may converge slow

Residual variance structure for large n

- ρ = [2Cov(E_D,E_S) +(n-2)Var(E_S)] / Var(e)
  - This is likely to be positive for large n
  - "Group members are similar" → you can fit a random group effect instead
  - This yields a simpler but equivalent model as long as ρ > 0.

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{A}_S + \mathbf{Z}_g\mathbf{g} + \mathbf{e}^*$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{A}_S + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} e_1^* \\ e_2^* \\ e_3^* \\ e_4^* \\ e_5^* \\ e_6^* \\ e_7^* \\ e_8^* \end{bmatrix}$$

$$Var(\mathbf{e}^*) = \mathbf{I}\sigma^2_{e^*}$$

$$Cov(y_i, y_j \mid A) = 2\sigma_{E_{DS}} + (n-2)\sigma^2_{E_S}$$

$$\Rightarrow \sigma^2_g = 2\sigma_{E_{DS}} + (n-2)\sigma^2_{E_S}$$

$$\sigma^2_g + \sigma^2_{e^*} = \sigma^2_e$$

$$\Rightarrow \sigma^2_{e^*} = \sigma^2_e - \sigma^2_g = \sigma^2_{E_D} - 2\sigma_{E_{DS}} + \sigma^2_{E_S}$$

Note: this redefines the residual and its variance → comparison of studies

Problem: This is valid only when ρ > 0
How do you know beforehand that ρ > 0?

Ignoring non-genetic social effects

What happens if you ignore E_S?

- Simply fit $\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{e}$ with $\text{Var}(\mathbf{e}) = \mathbf{I}\,\text{Var}(e)$
- This assumes that $\rho = 2\text{Cov}(E_D,E_S) + (n-2)\text{Var}(E_S) = 0$
- Either: social effects are assumed fully heritable, $\text{Var}(E_S) = 0$
- or $(n-2)\text{Var}(E_S) = -2\text{Cov}(E_D,E_S)$
- These are very strong a priori assumptions
- This is not an issue of statistical significance or not, always allow for E_S

Consequences of ignoring E_S

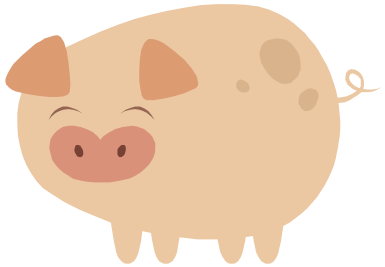
- $\text{Var}(E_S)$ ends up in $\text{Var}(A_S) \rightarrow$
  - Severe overestimation of (social) genetic variance
- Bijma et al., 2007b
  - Estimated $\rho = 0.09$ ($P<0.001$)
  - Using $\text{Var}(\mathbf{e}) = \mathbf{I}\,\text{Var}(e) \rightarrow$ Var(TBV) overestimated by a factor of 2.6!

Message

Estimated genetic parameters for social effects are extremely sensitive to what other components you fit in the model

Model selection is a key issue

Ignoring non-genetic social effects



- Ignoring social effects may bias estimation of *classical* heritability <u>when group members are related</u>
- Feed intake pigs, n = 8, Bergsma et al 2008
- Average relatedness within group, r = 0.18
- Classical model $y = Xb + Za + e$
  - Estimated $h^2 = 0.41$
- Accounting for group effect $y = Xb + Za + Z_g g + e$
  - Estimated $h^2 = 0.18$
  - Physically pens were identical
- Due to social effects and large n, group members are similar ($\rho > 0$)
  - Similar group members $\rightarrow$ similar relatives $\rightarrow h^{2}\uparrow$

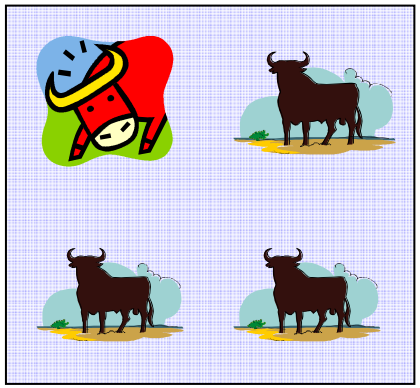
Statistical models for socially affected traits

- Which fixed and random effect to include?
  - The social effect is “phenotypic”
    - It may have fixed, random, and genetic components
  - Not all biological components may be estimable (e.g. $Var(E_S)$, $Cov(E_D,E_S)$ and $Var(E_D)$)
    - Derive the resulting variance structure within and between groups
  - Statistical significance of correction factors is not the primary issue
    - We know that $h^2 \neq 100\%$, account for E, also when p>0.1

Statistical models for socially affected traits

- The social effect is “phenotypic”
  - Include fixed effects for the group member
    - $Y = \{X_D b_D + Z_D a_D + e_D\} + \{X_S b_S + Z_S a_S + e_S\}$
    - Sex, age or breed <u>of the group member</u>
      - These are usually easy to estimate
  - Include random effects for the group member
    - $Y = \{X_D b_D + Z_D a_D + .. + e_D\} + \{X_S b_S + Z_S a_S + .. + e_S\}$
    - Litter of the group member (non-genetic maternal social effect)
    - Permanent effects (repeated records)
    - Mother of the group member (genetic maternal social effect)
      - These are not always easy to estimate
      - Test sensitivity of your social VC for other model components
  - Derive the theoretically expected (residual) variance structure
    - And allow for it in your statistical model

Example: mixed breeds in beef cattle



- Allow for a social fixed effect of the breed of the group members
  - Angus or Hereford (A,H)

$$\mathbf{y} = \mathbf{X}_D\mathbf{b}_D + \mathbf{X}_S\mathbf{b}_S + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{e}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} H_D \\ A_D \end{bmatrix} + \begin{bmatrix} 0 & 3 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} H_S \\ A_S \end{bmatrix} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{e}$$

Breed of self Breed of group members

Use the “and()” statement in Asreml

Estimability of social genetic VC

- Little research has been done
- So far
  - Relatedness within and between groups is critical
  - FS-groups is impossible, irrespective of pedigree
    - You cannot distinguish direct from social effects
  - HS-groups without FS in the data is also impossible
  - Be carefull with structuring families across groups (e.g. Wolf PNAS paper → see Bijma et al., 2007b)
  - Random groups (with respect to relatedness) works well
    - But is probably not optimal
  - Combining two families per group is an option
  - This may be useful when tagging is difficult (marine species?)
  - You cannot fit a **fixed** group effect
    - Problematic???
    - Avoid confounding of physically good pens with certain families
- Data requirements
  - ~4 times more than for direct effect only (random groups)
  - More if groups are larger

Dealing with variable group size

- Case 1
  - Underlying parameters do not depend on group size (no "true" G x group-size interaction)
  - Genetic VC are constant [in particular: $V(A_S) \neq f(n)$]
  - Issue is impact of *n* on non-genetic section of model

Correlated residual model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{e}$$

$$Var(\mathbf{e}) = \mathbf{R}\sigma_e^2,\ R_{ii} = 1,\ R_{ij,within} = \rho$$

$$\rho = [2\sigma_{E_{DS}} + (n-2)\sigma_{E_S}^2]/\sigma_e^2$$

$$\sigma_e^2 = \sigma_{E_D}^2 + (n-1)\sigma_{E_S}^2$$

ρ and Var(e) depend on group size

↓

Fit heterogeneous Var(e) and ρ

Should be possible in AsReml

Dealing with variable group size

Random group effect model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \mathbf{Z}_g\mathbf{g} + \mathbf{e}^*$$

$$Var(\mathbf{e}^*) = \mathbf{I}\sigma^2_{e^*},\ \sigma^2_{e^*} = \sigma^2_{E_D} - 2\sigma_{E_{DS}} + \sigma^2_{E_S}$$

$$Var(g) = [2\sigma_{E_{DS}} + (n-2)\sigma^2_{E_S}]$$

Var(e) does not depend on group size

but Var(g) does

↓

Fit heterogeneous Var(g)

Consequences of ignoring heterogeneity of variance have not been investigated

The genetic term is “automatically” heterogeneous because the number of group members in Z_S varies → If non-genetic heterogeneity of variance exists, you may expect it to end up in (and thus inflate) the genetic terms

Dealing with variable group size

- Case 2
  - Variance of social effects depend on group size, $Var(A_S) = f(n)$

Consider: $\sigma^2_{TBV} = \sigma^2_{A_D} + 2(n-1)\sigma_{A_{DS}} + (n-1)^2\sigma^2_{A_S}$

- In large groups, heritable variance is very large → this may not make sense
- The social effects per individual must become smaller → $Var(A_S)$ must go down with n
- This is not true GxE-interaction, just scaling or "dilution"
- i.e. $Corr(A_{S,i,n=4}, A_{S,i,n=5}) = 1$, but $Var(A_{s,i,n=4}) > Var(A_{s,i,n=5})$
- We found such results for growth in pigs

Dealing with variable group size

- Accounting for decreasing Var(A_S) with n
  - Diluting social effects depending on n –1

$$y_i = fixed + A_{D,i} + c_{n-1} \sum_{n-1} A_{S,j} + e_i$$

- c is an unknown constant depending on (n – 1) (Arango et al., 2005, JAS)
- c = 1 $\rightarrow$ Var(A_S) is independent of n $\rightarrow$ Var(TBV) increases with n
- c = 1/(n –1) $\rightarrow$ the sum of social effects is constant
  $\rightarrow$ Var(TBV) is independent of n
- find best c iteratively using AsReml
  - E.g. c = $(n–1)^x$, vary x from 0 to –1
  - This is like random regression, slope = age * bv_{slope}
    -but “age” is unknown $\rightarrow$ iterate to ReML value

An alternative model useful for BVE (Abe Huisman)

- Interesting when:
  - If your BVE-software does not allow for social effects
  - Your data file becomes too large when you add all group members
- Idea
  - Direct effects are expressed in self
  - Social effects are expressed in group members
  - → use conventional bivariate analysis with two traits:
    1. Own performance
    2. Mean performance of group members
  - This fits in ordinary BVE software
- Issues
  - This model does not properly fit when group members are related
  - → not at all robust for VCE, seems less important for BVE
  - This has not been extensively tested!

Summary on VCE

- Social variance components can be estimated
- They are sensitive to BIAS
  - Confounding with non-genetic social effects
  - Fit fixed effects also for social component
  - Sensitivity analysis is important
  - Think of the biological interpretation of your model
- More research is needed on
  - Optimum designs for analysis (relatedness)
  - Varying group size
  - …….