

A Hierarchical Model with a Correlation between the Levels: An Example from Animal Breeding

Majbritt Felleki, Dalarna University, Sweden

The International Conference on Trends and Perspectives in
Linear Statistical Inference (LinStat2014)
24-28 August 2014, Linköping, Sweden



Sveriges
lantbruksuniversitet



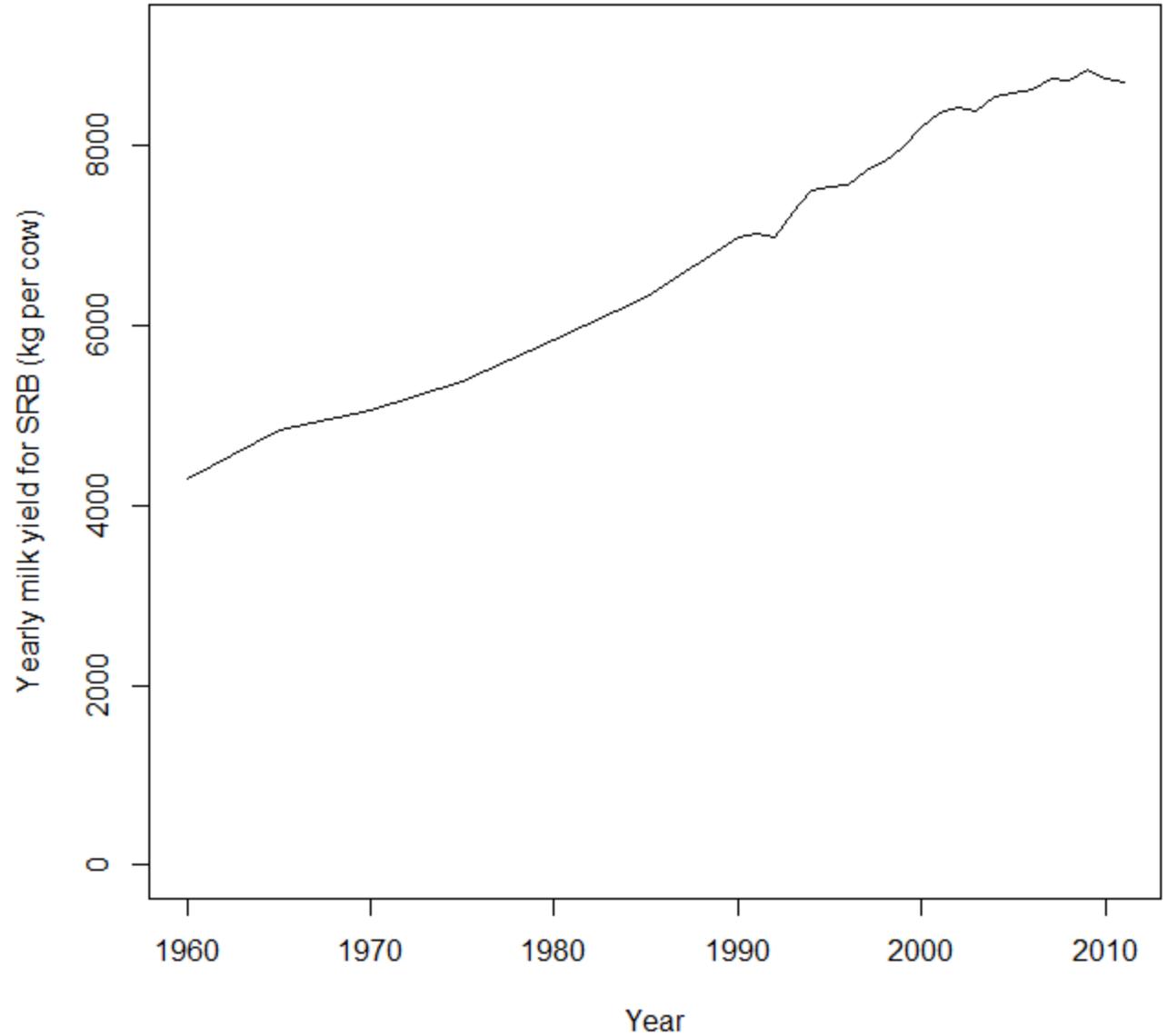
HÖGSKOLAN
DALARNA



Removed because of copyright

Removed because of copyright

Milk yield

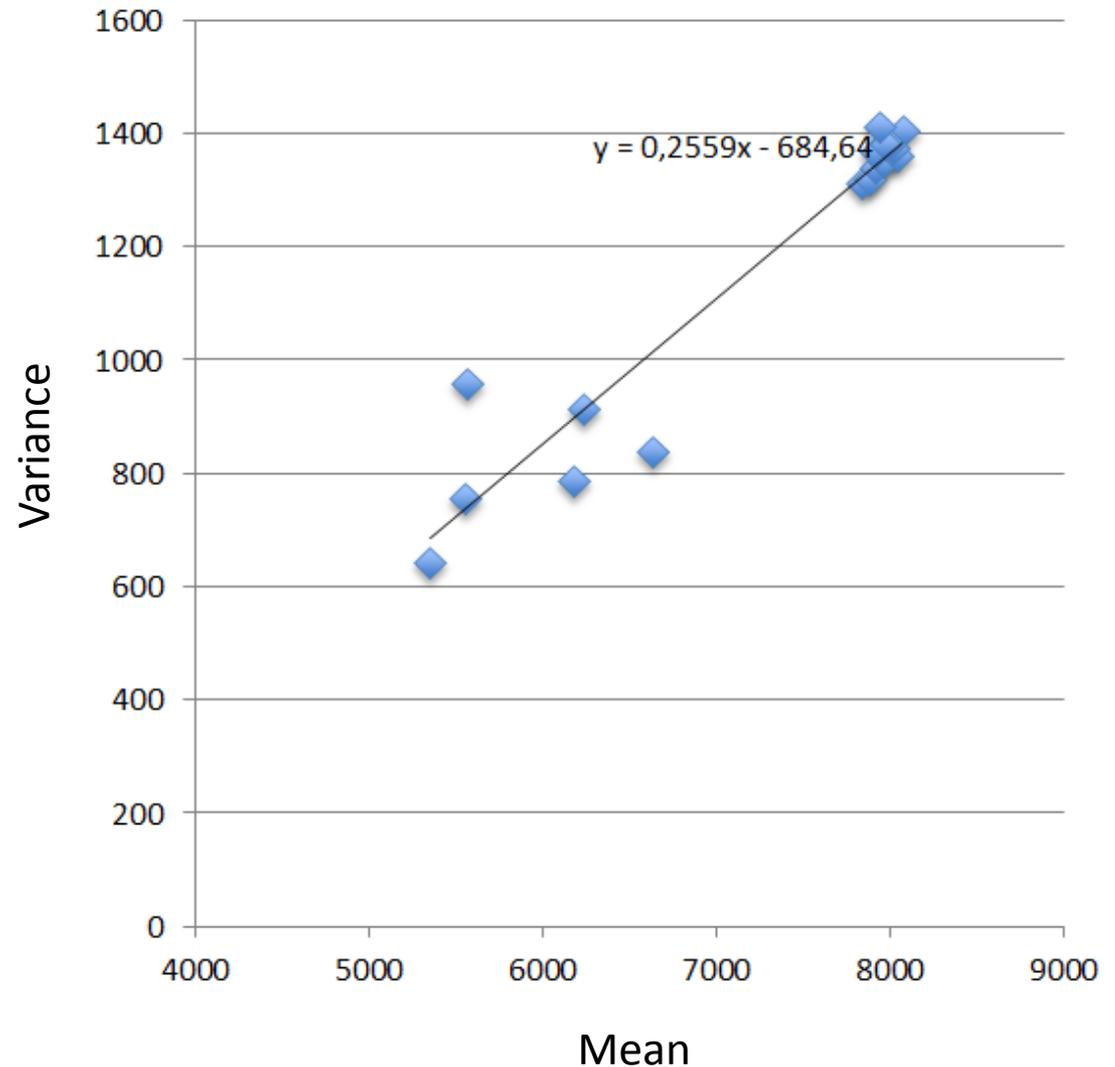


Data from
Swedish Dairy
Association
2011 and 2012
Cattle
Statistics

Mean and variance

Milk yield

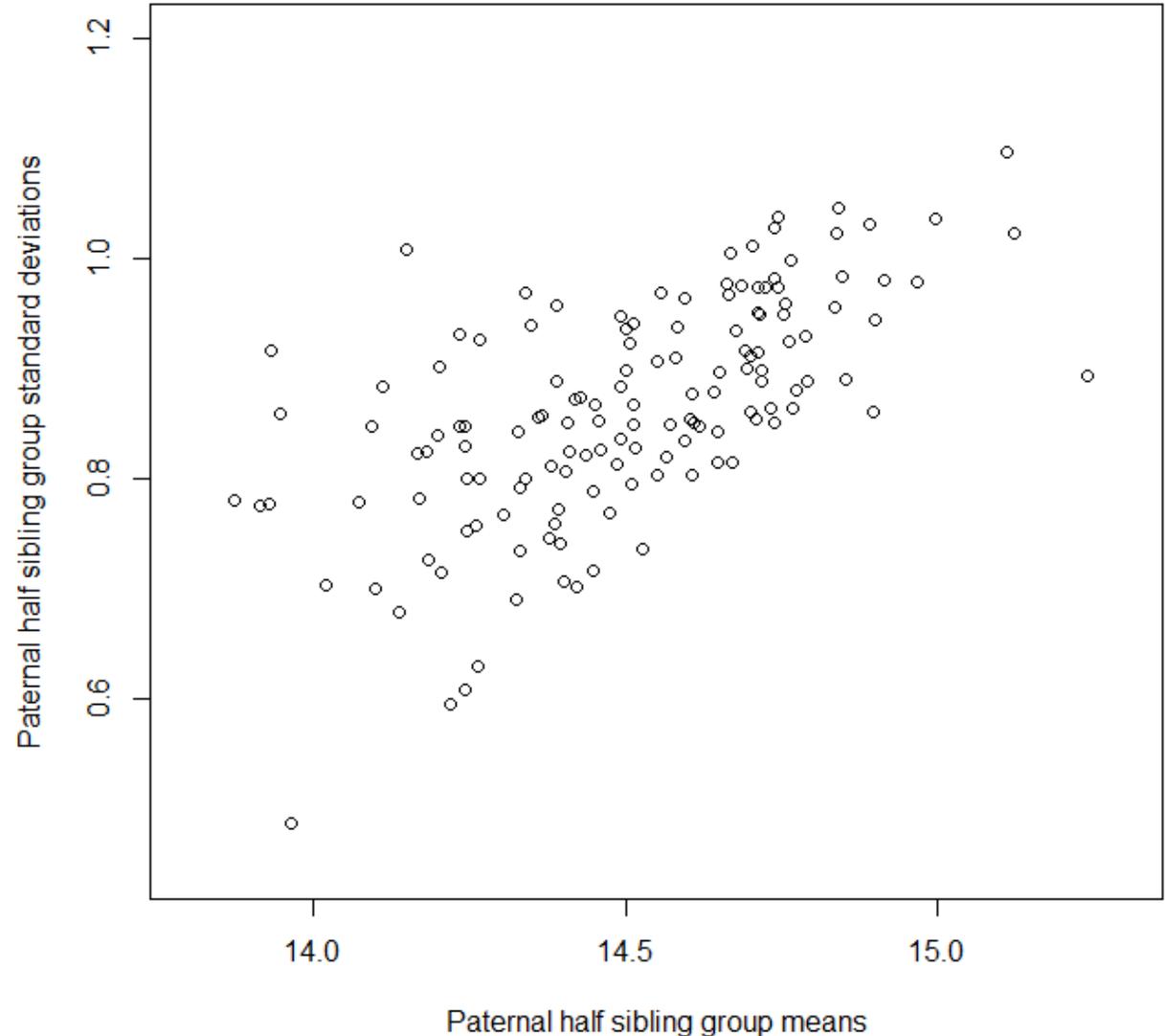
Figure by Erling Strandberg. Data from Strandberg & Danell (1989) and personal communication with Kjell Johansson, Växa Sverige.



Mean and variance

Teat count

Standard deviations of teat counts as a function of means for paternal half sib groups.
Data from Nordic Genetics.



Data

- Litter sizes, teat counts, milk yield, somatic cell count
- Fixed effects: herd, season, parity, date of birth, days in milk...
- Random effects: additive genetic and environmental

Hierarchical model

- "Oh, so the random effects have random effects!"

Hierarchical model

- Normal distributed responses and normal distributed random effects:

- $y = X\beta + Za + Wp + e$

- $e \sim N(0, \Phi)$, $\text{diag } \Phi = \varphi$,

$$\log \varphi = X_d \beta_d + Z a_d + W p_d$$

- $p \sim N(0, \sigma_p^2)$, $p_d \sim N(0, \sigma_{p_d}^2)$,

$$\begin{pmatrix} a \\ a_d \end{pmatrix} \sim N(0, G \otimes A), \quad G = \begin{pmatrix} \sigma_a^2 & \rho \sigma_a \sigma_{a_d} \\ \rho \sigma_a \sigma_{a_d} & \sigma_{a_d}^2 \end{pmatrix}$$

Hierarchical model

- Normal distributed responses and normal distributed random effects:

- $y = X\beta + Za + Wp + e$

- $e \sim N(0, \Phi)$, $\text{diag } \Phi = \varphi$,

$$\log \varphi = X_d \beta_d + Z a_d + W p_d$$

- $p \sim N(0, \sigma_p^2)$, $p_d \sim N(0, \sigma_{p_d}^2)$,

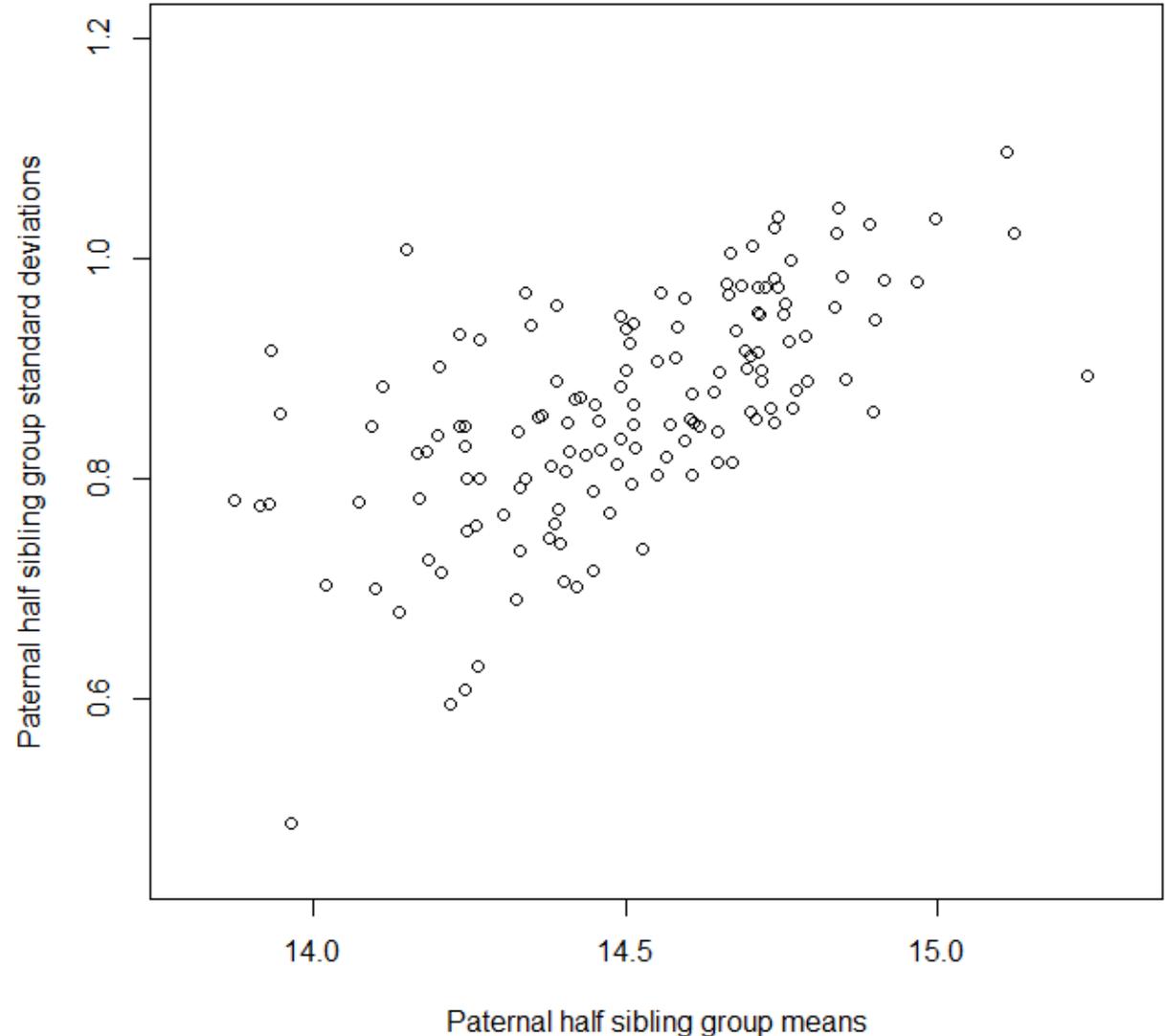
$$\begin{pmatrix} a \\ a_d \end{pmatrix} \sim N(0, G \otimes A), \quad G = \begin{pmatrix} \sigma_a^2 & \rho \sigma_a \sigma_{a_d} \\ \rho \sigma_a \sigma_{a_d} & \sigma_{a_d}^2 \end{pmatrix}$$


Removed because of copyright

Mean and variance

Teat count

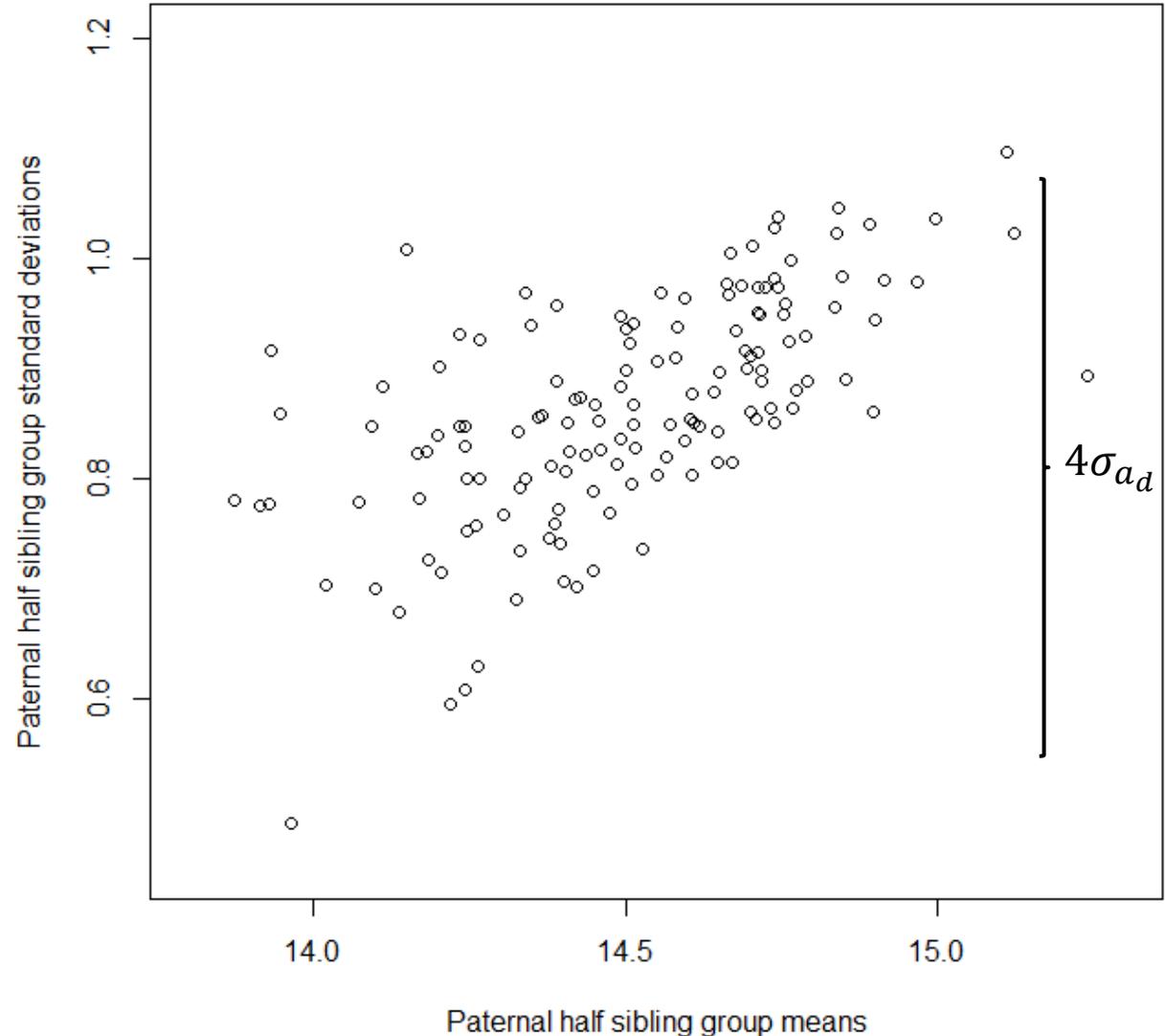
Standard deviations of teat counts as a function of means for paternal half sib groups.
Data from Nordic Genetics.



Mean and variance

Teat count

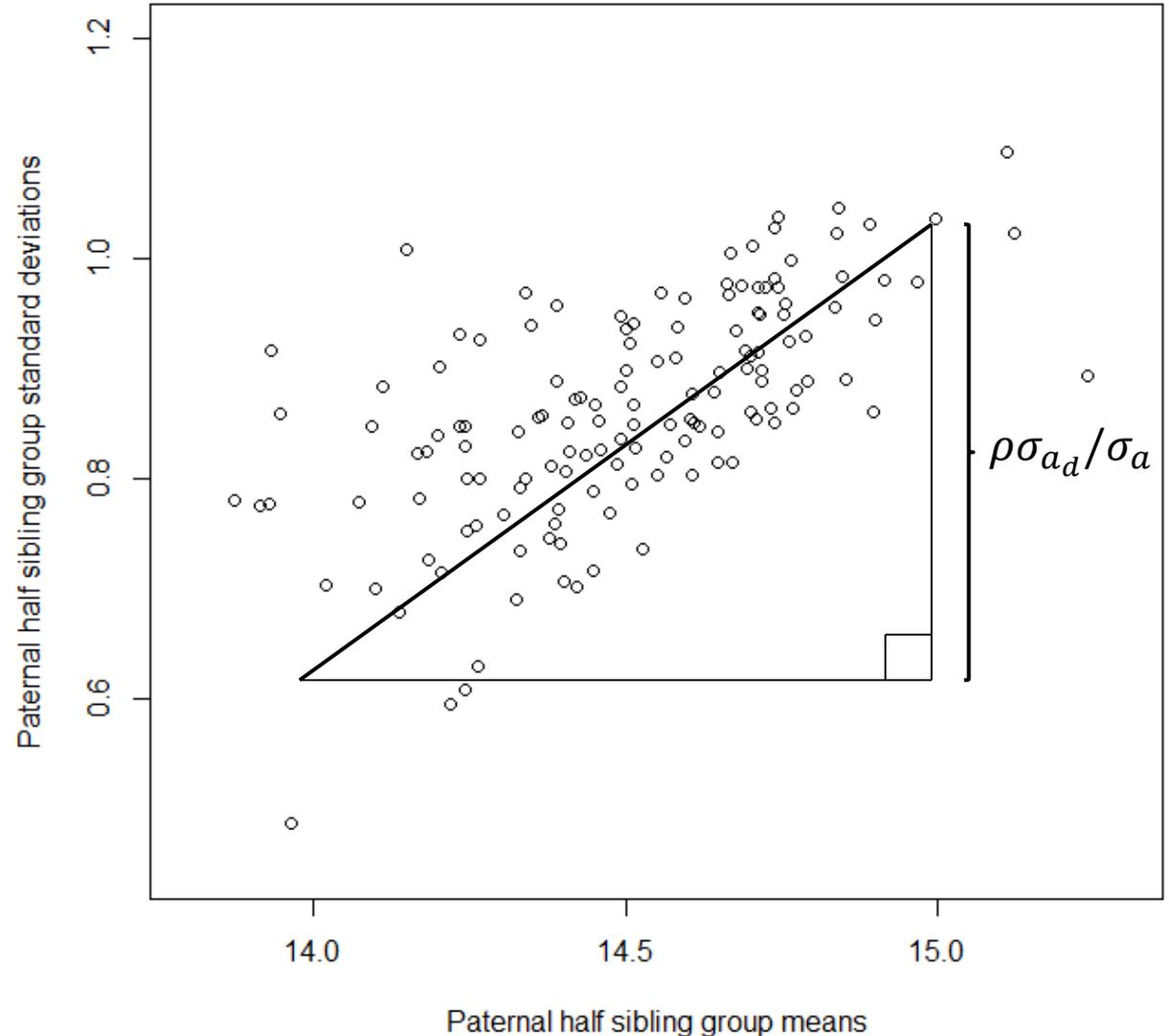
Standard deviations of teat counts as a function of means for paternal half sib groups.
Data from Nordic Genetics.



Mean and variance

Teat count

Standard deviations of teat counts as a function of means for paternal half sib groups.
Data from Nordic Genetics.



Hierarchical model

- Normal distributed responses and normal distributed random effects:

- $y = X\beta + Za + Wp + e$

- $e \sim N(0, \Phi)$, $\text{diag } \Phi = \varphi$,

$$\log \varphi = X_d \beta_d + Z a_d + W p_d$$

- $p \sim N(0, \sigma_p^2)$, $p_d \sim N(0, \sigma_{p_d}^2)$,

$$\begin{pmatrix} a \\ a_d \end{pmatrix} \sim N(0, G \otimes A), \quad G = \begin{pmatrix} \sigma_a^2 & \rho \sigma_a \sigma_{a_d} \\ \rho \sigma_a \sigma_{a_d} & \sigma_{a_d}^2 \end{pmatrix}$$

HGLM
(LMM, GLMM)
and
DHGLM

Monographs on Statistics and Applied Probability 106

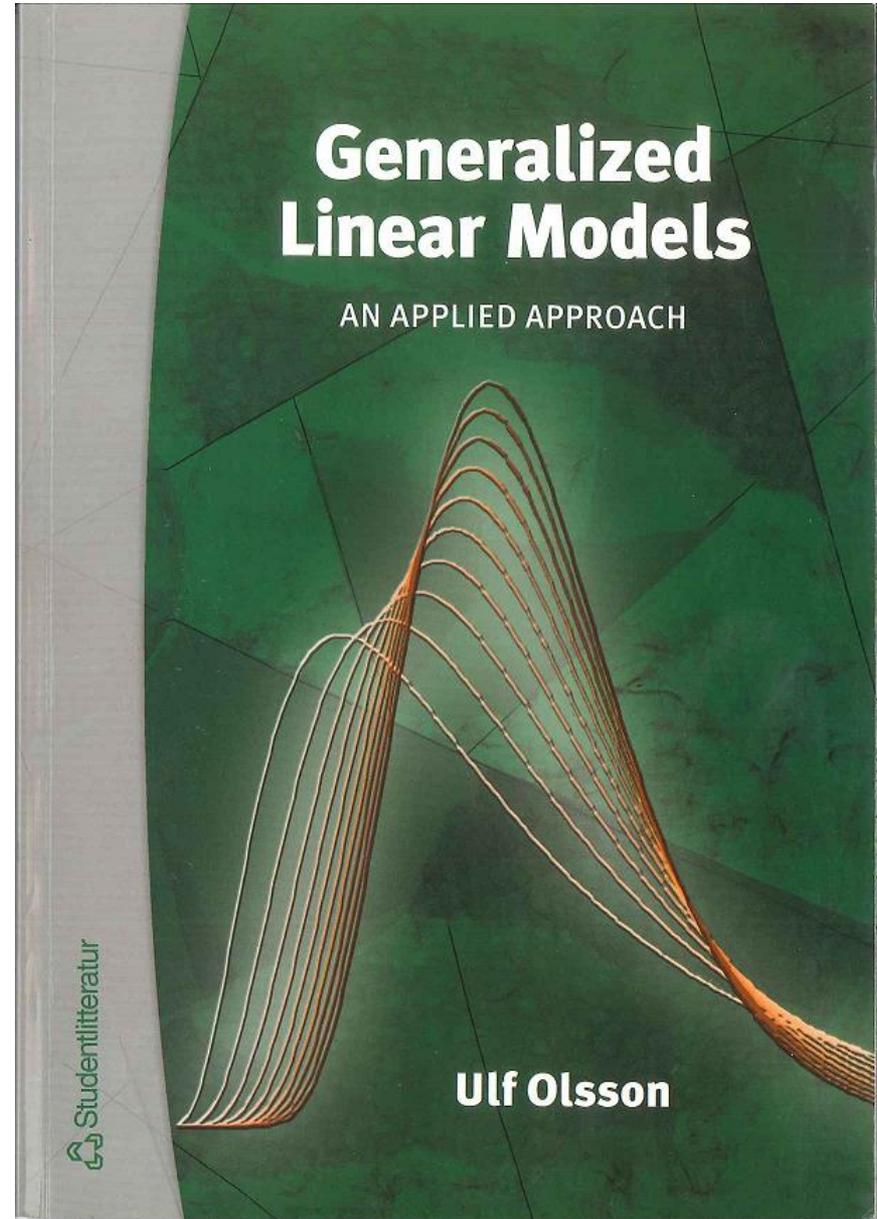
**Generalized Linear
Models with
Random Effects**
Unified Analysis via H-likelihood



Youngjo Lee
John A. Nelder
Yudi Pawitan

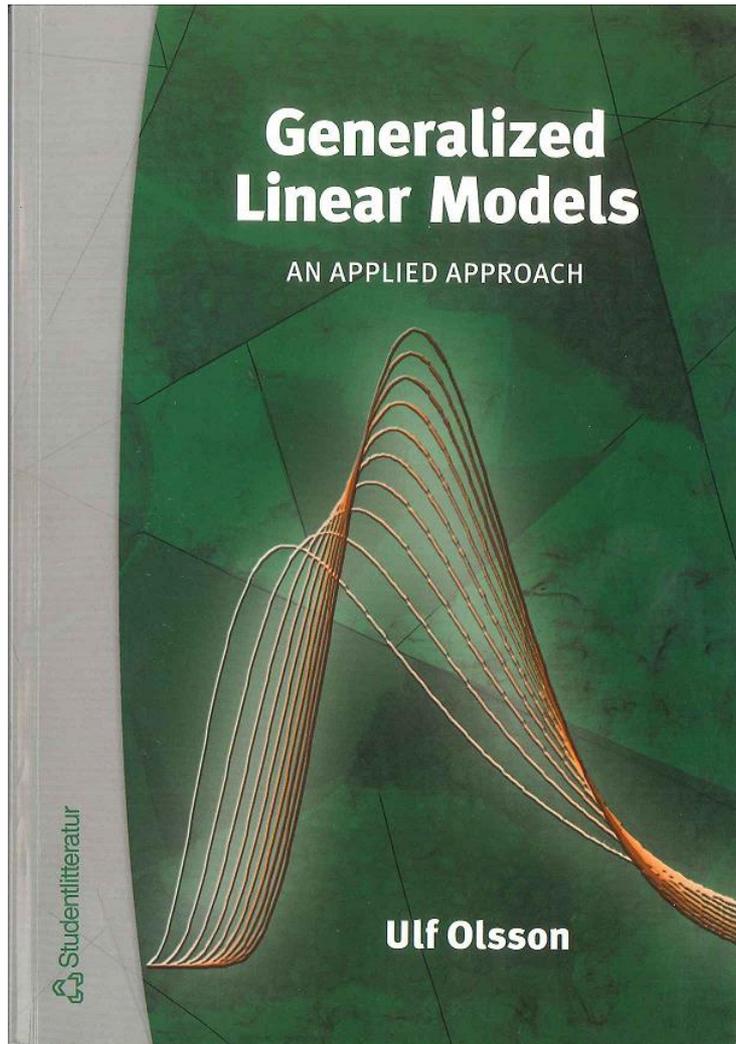
 **Chapman & Hall/CRC**
Taylor & Francis Group

GLM



GLM

VS (D)HGLM



Estimation

- $h = \log f(y, a, a_d, p, p_d)$
- β and a are estimated from h
- β_d and a_d are estimated from the adjusted profile likelihood

$$p_{\beta,a}(h) = \left[h - \log \det \left\{ -\frac{\partial^2 h}{\partial(\beta,a)^2} / (2\pi) \right\} \right]_{(\beta,a)=(\hat{\beta},\hat{a})}$$

- The variance components are estimated from

$$p_{\beta,a,\beta_d,a_d}(h) = \left[h - \log \det \left\{ -\frac{\partial^2 h}{\partial(\beta,a,\beta_d,a_d)^2} / (2\pi) \right\} \right]_{(\beta,a,\beta_d,a_d)=(\hat{\beta},\hat{a},\hat{\beta}_d,\hat{a}_d)}$$

Estimation

- Mean part

$$y \sim N(X\beta + Za + Wp, \Phi\sigma^2)$$

- Residual variance part

$$\frac{\hat{e}^2}{1-h} \sim \Gamma\left(\frac{1-h}{2}, \frac{1-h}{2\varphi}\right),$$

$$\log \varphi + \frac{1}{\varphi} \left(\frac{\hat{e}^2}{1-h} - \varphi \right)$$

$$\sim N\left(X_d\beta_d + Za_d + Wp_d, \frac{2}{1-h}\sigma_d^2\right)$$

Results - data

Removed because of copyright

Results - data

Removed because of copyright

Results - simulations

Removed because of copyright

Results - simulations

Removed because of copyright

Discussion and future research

- The algorithm, when running, is fast
- Huge data can be analyzed
- The code required is almost user friendly
- The results are unbiased in case of (a not yet specified number of) repeated observations

Discussion and future research

- Dependent on scale
- Approximation of DHGLM (because of A)
- Get rid of the approximation
- Different additive genetic effects for mean and residual variance

References

- Rönnegård, L., Felleki, M., Fikse, W. F., Mulder, H. A. & Strandberg, E. (2010). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution* 42(1), 8.
- Felleki, M., Lee, D., Lee, Y., Gilmour, A. R. & Rönnegård, L. (2012). Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genetics Research* 94(06), 307–317.
- Rönnegård, L., Felleki, M., Fikse, W. F., Mulder, H. A. & Strandberg, E. (2013). Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *Journal of Dairy Science* 96(4), 2627–2636.

