Comparison of statistical models to analyze the genetic effect on within-litter variance in pig

*Dörte Wittenburg**, *Volker Guiard*†, *Friedrich Teuscher*, *Norbert Reinsch*

Research Institute for the Biology of Farm Animals, Genetics & Biometry,
Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

1 Introduction

The uniformity of special traits is economically important in animal production. A high variability of piglet birth weight within litter impacts the survival rate negatively (e.g. Högberg & Rydhmer, 2000). Therefore, present research in livestock genetics examines genetic effects on the so called within-litter variance (e.g. Roehe & Kalm, 2000; Damgaard *et al.*, 2003).

The present study investigates genetic effects on the within-litter variance in consideration of a sex effect on piglet birth weight. The objectives are to decrease the within-litter variance and piglet mortality and to keep litter size high. The within-litter variance is defined as the sow's trait. The trait is fitted by animal models to estimate heritabilities. Transformations of the within-litter variance are described by linear mixed models (LMMs). We use restricted maximum likelihood (REML; Searle *et al.*, 1992) to estimate the variance components. Furthermore, the untransformed trait is fitted by a generalized linear mixed model (GLMM) with log-link function. The presented models are compared by residual diagnostic.

2 Material and methods

The German breeding organization BHZP provided a dataset consisting of 103 266 birth weights of 9 439 litters of German Landrace (LR; 5 379 litters) and German Edelschwein (ES; 4 060 litters). The data were recorded from January 2002 to August 2006. Births of $N = 3\,914$ sows (LR 2 245; ES 1 669) in the first to eighth parity were observed. It is assumed, that birth weights are independently and normally distributed within litter. A preliminary analysis of piglet birth weight resulted in a significant sex effect on the phenotype. After generating the sample variances per litter and sex we applied an adjusted F-test (Kenward & Roger, 1997) to exclude a sex effect on the variability of birth weight within litter. Thus, the *pooled* sample variance of male and female birth weights within litter is defined as the

---

*`wittenburg@fbn-dummerstorf.de`

sow's trait. This trait is proportional to a χ^2-distributed random variable with $(n_{i,\text{litter}} - 2)$ degrees of freedom, where $n_{i,\text{litter}}$ denotes the litter size of sow $i \in \{1, \dots, N\}$ per litter. The genetic effect is examined separately for each line.

The distribution of the pooled sample standard deviation S is well approximated by a normal distribution, see Figure 1(a). Therefore, the within-litter standard deviation may be described by a linear mixed model, i.e.

$$S = X\beta + Zu + e\,. \tag{M1}$$

Model (M1) consists of random effects u, fixed effects β (farm-year-season, parity, interaction line of boar × line of sow) and the random deviation $e \sim N(0, W^{-1}\sigma_e^2)$. The matrix W is a diagonal matrix of weights. The inverse of the conditional variance of the standard deviation given the random effects is defined as weight per sow and litter. The vector u includes the genetic effect of sow $u_a \sim N(0, A\sigma_a^2)$ with the numerator relationship matrix A, effect of father of litter $u_p \sim N(0, I\sigma_p^2)$ and the permanent environment $u_{pe} \sim N(0, I\sigma_{pe}^2)$, which affects all litters of one sow. Moreover, it is $V = Var(S) = ZGZ' + W^{-1}\sigma_e^2$ and $Var(u) = G$.

We apply the software ASReml 2.0 (Gilmour *et al.*, 2006) to estimate the variance components. The weights $2\left(n_{i,\text{litter}} - 2\right)/\widehat{s}^2_{i,\text{litter}}$ are included, where $\widehat{s}^2_{i,\text{litter}}$ denotes the predicted sample variance per litter. The weights are iteratively adjusted to improve the estimation of variance components and the prediction of the trait.

Moreover, the distribution of the logarithmized pooled sample variance is well approximated by a normal distribution. The animal model for $\ln S^2$ consists of the same components as mentioned above. In this case constant weights $\left(n_{i,\text{litter}} - 2\right)/2$ are used for the estimation of variance components.

To evaluate the presented models we analyzed the studentized standardized residuals $r_{i,\text{litter}}$. One needs the difference between observed and predicted value, i.e. $\epsilon_{i,\text{litter}} = s_{i,\text{litter}} - \widehat{s}_{i,\text{litter}}$, and the corresponding entry of the residual covariance matrix $Var(\epsilon) = K\left(V - Q\right)K'$ with the matrices $Q = X\left(X'V^{-1}X\right)^{-}X'$ and $K = I - ZGZ'V^{-1}$ to calculate the studentized residuals $r_{i,\text{litter}}$ of model (M1). Substituting the unknown parameters by their estimates yields

$$r_{i,\text{litter}} = \frac{\epsilon_{i,\text{litter}}}{\sqrt{\widehat{Var}(\epsilon_{i,\text{litter}})}}\,. \tag{R}$$

The studentized residuals of the logarithmized sample variance are calculated similarly.

Considering the χ^2-distribution of the pooled sample variance S^2 the following generalized linear mixed model is applicable on the pooled sample variance

$$S^2 = \exp\{X\beta + Zu\}\,\varepsilon\,. \tag{M2}$$

The vectors u and β coincides with those used in model (M1). The conditional expectation $\mu = \mathbb{E}(S^2|u)$ is transformed into the linear predictor η with use of the link function g. We use the log-link function, hence $\eta = g(\mu) = X\beta + Zu$. The random deviations $\varepsilon_{i,\text{litter}}$ $(i = 1, \dots, N)$ are independently and gamma distributed with expectation one and variance $2\phi/(n_{i,\text{litter}} - 2)$. The dispersion parameter ϕ describes additional variance, which is not covered by the model. The estimation of variance components was done in ASReml 2.0 by an adapted REML method for GLMM (Wolfinger & O'Connell, 1993) with regard to the gamma distributed error ε.

The studentized residuals $r_{i,\text{litter}}$ of model (M2) are evaluated on the scale caused by the link function g. The deviations are $\epsilon_{i,\text{litter}} = \Delta^{-1}_{i,\text{litter}}\left(s^2_{i,\text{litter}} - \widehat{s}^2_{i,\text{litter}}\right)$, where $\Delta^{-1} = [\partial\eta/\partial\mu]$ denotes the matrix of first derivatives at the estimated parameters. Using the log-link it holds $\Delta_{i,\text{litter}} = \widehat{s}^2_{i,\text{litter}}$. The residual variance is $Var(\epsilon) = K\left(\widetilde{V} - Q\right)K'$ with the matrices K and Q as mentioned above and $\widetilde{V} = ZGZ' + \phi W^{-1}$. The diagonal matrix W consists of the weights $(n_{i,\text{litter}} - 2)/2$. Thus, the studentized residuals are calculated similarly to (R).

The estimation of heritability was done in the LMMs by

$$h^2_{\text{litter}} = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_p + \sigma^2_{pe} + \overline{w}^{-1}_{\text{litter}}\sigma^2_e}\,. \tag{H}$$

Substituting σ^2_e by ϕ in (H) the estimator h^2_{litter} is also used to estimate heritability in the GLMM. To achieve an appropriate estimator for heritability we construct a special weight $\overline{w}_{\text{litter}}$ based on average values per litter. Hence the average litter size and – in case of model (M1) – the average predicted standard deviation per litter are included in the estimation.

3 Results and discussion

The estimation of heritabilities and their standard errors (see Table 1) yielded to similar results within line. The heritability was about 8 % in line LR and 10 % was estimated in line ES. These estimates were transformed into breeding successes δ_{litter}. The breeding success was determined relatively to the estimated standard deviation within litter and amounts to at least 1 %. The annual proportion selected was 56.25 %. Because the proportion selected is fairly high, the 1 % can be interpreted as a lower bound of theoretically possible breeding success.

Figures 1(c) and 1(d) contrast the studentized residuals with the predicted traits of model (M1) and (M2), respectively, exemplary for line ES. The residuals in Figure 1(c) are evenly distributed and their fluctuation margin is small. The estimates of skewness and kurtosis are 0.430 and 0.356, respectively. In contrast, the residuals in Figure 1(d) show a light trend, which was not completely balanced by the log-link function. The estimates of skewness and kurtosis are 1.212 and 2.218, respectively. An alternative link function might improve this residual appearance. But having a gamma distributed error ε the usual link functions such as inverse and identity are not appropriate to reduce the skewness.

Figure 1(b) outlines the role of weighting in line ES. Analyzing model (M1) without weights, that means $e \sim N(0, I\sigma_e^2)$, the absolute residual values of sows within 5 % of the lowest litter sizes turned out to be larger than in a weighted analysis and vice versa. The predicted trait was often slightly misestimated. Note, that the weights were defined as the inverse conditional variances of S_i $(i = 1, \dots, N)$ and thus, by weighting the trait we achieve an asymptotic residual variance of one. However, the weighted analysis of the sample standard deviation only showed negligible differences in respect to the distribution of the studentized residuals in comparison to an unweighted analysis.

The evaluation of estimated breeding values u_a of model (M1) and (M2) yielded to a rank correlation of 0.95 (line LR) and 0.98 (line ES). The rank correlation between weighted and unweighted analysis of model (M1) was 0.94 (line LR) and 0.96 (line ES). Thus, in terms of breeding values no essential differences between the presented models are recognizable.

4 Conclusions

This study showed, that a significant sex effect on the piglet's phenotype but not on its variability can be appropriately handled by the pooled within-litter variance of male and female birth weights. The LMMs for S and $\ln S^2$ as well as the GLMM for S^2 are suitable to analyze the genetic effect on the within-litter variance. The residual diagnostic, that means the analysis of the residual distribution and calculating the skewness and kurtosis of the studentized residuals, favors model (M1) to fit the within-litter variance. The estimated heritabilities are quite small, nevertheless there is potential to decrease the variability of birth weight within litter through selection.

	Trait	σ_a^2 (se)	h^2_{litter} (se)	δ_{litter} (%)
Line LR	S	778.704 (219.653)	0.073 (0.020)	1.02
	$\ln S^2$	0.065 (0.016)	0.084 (0.020)	1.29
	S^2	0.041 (0.010)	0.085 (0.020)	1.03
Line ES	S	836.727 (165.927)	0.108 (0.021)	1.37
	$\ln S^2$	0.055 (0.010)	0.088 (0.015)	1.21
	S^2	0.050 (0.010)	0.107 (0.021)	1.27

Table 1: Estimation of additive-genetic variance σ_a^2 and heritability h^2_{litter}; breeding success δ_{litter} is given relatively to the estimated standard deviation within litter

References

Damgaard, L. H., Rydhmer, L., Løvendahl, P. & Grandinson, K. (2003), Genetic parameters for within-litter variation in piglet birth weight and change in within-litter variation during suckling. *Journal of Animal Science* **81**, 604–610.

Gilmour, A. R., Cullis, B. R., Harding, S. A. & Thompson, R. (2006), *ASReml Update: What's new in Release 2.00.* VSN International Ltd, Hemel Hempstead, UK.

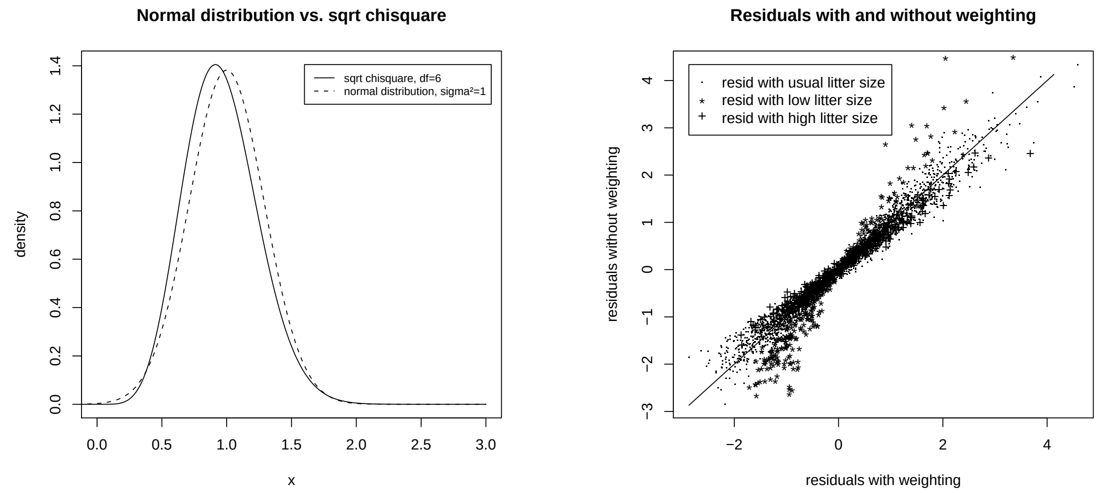
Högberg, A. & Rydhmer, L. (2000), A genetic study of piglet growth and survival. *Acta Agriculturæ Scandinavica, Section A, Animal Science* **50**, 300–303.

Kenward, M. G. & Roger, J. H. (1997), Small sample inference for fixed effects from Restricted Maximum Likelihood. *Biometrics* **53**, 983–997.

Roehe, R. & Kalm, E. (2000), Estimation of genetic and environmental risk factors associated with pre-weaning mortality in piglets using generalized linear mixed models. *Animal Science* **70**, 227–240.

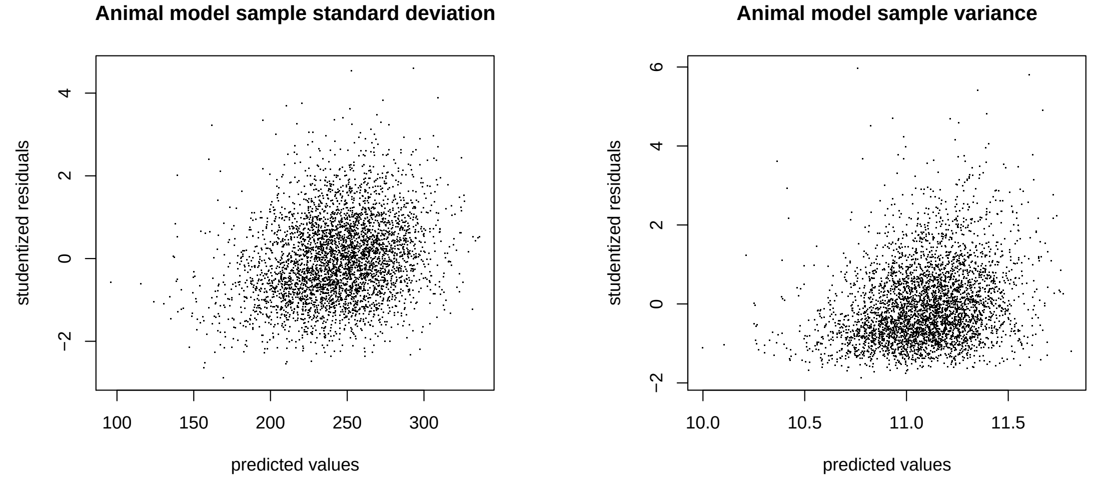
Searle, S. R., Casella, G. & McCulloch, C. E. (1992), *Variance Components.* Wiley, New York.

Wolfinger, R. & O'Connell, M. (1993), Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computations and Simulations* **48**, 233–243.



(a) Density of sample standard deviation S vs. density of a normal distribution $N(\sigma, \nu)$ with $\mathbb{E}(S) \approx \sigma$, $\nu = \sigma^2/(2 \cdot \mathrm{df})$

(b) Studentized residuals with and without weighting the within-litter standard deviation



(c) Studentized residuals of the linear mixed model (M1)

(d) Studentized residuals of the generalized linear mixed model (M2)

Figure 1: Analysis of variability of piglet birth weight within litter (German Edelschwein)