Genet. Res., Camb. (2003), **82**, *pp.* 207–222. *With 9 figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006426 Printed in the United Kingdom

Normal linear models with genetically structured residual variance heterogeneity: a case study

DANIEL SORENSEN[1]* AND RASMUS WAAGEPETERSEN[2]
[1] *Danish Institute of Agricultural Sciences, Department of Animal Breeding and Genetics, PB50, 8830 Tjele, Denmark*
[2] *Department of Mathematical Sciences, Aalborg University, 9220 Aalborg, Denmark*

(*Received 9 December 2002 and in revised form 30 June and 21 August 2003*)

Summary

Normal mixed models with different levels of heterogeneity in the residual variance are fitted to pig litter size data. Exploratory analysis and model assessment is based on examination of various posterior predictive distributions. Comparisons based on Bayes factors and related criteria favour models with a genetically structured residual variance heterogeneity. There is, moreover, strong evidence of a negative correlation between the additive genetic values affecting litter size and those affecting residual variance. The models are also compared according to the purposes for which they might be used, such as prediction of 'future' data, inference about response to selection and ranking candidates for selection. A brief discussion is given of some implications for selection of the genetically structured residual variance model.

1. Introduction

The normal mixed linear model commonly used in quantitative genetics postulates that the data and other random components are multivariate normally distributed, and that location parameters and data are linearly related. This basic structure is also a major building block when modelling takes place at the level of unobserved quantities, such as log frailties in log normal frailty models for the analysis of survival times (Korsgaard *et al.*, 1998) and liabilities in threshold models for the study of ordered categorical responses (Sorensen *et al.*, 1995). Typically, variance homogeneity is assumed but extensions that considered rather simple systematically structured departures from variance homogeneity were introduced in the 1990s (Foulley *et al.*, 1992; Gianola *et al.*, 1992; San Cristobal *et al.*, 1993; Foulley & Quaas, 1995). In particular, Foulley & Quaas (1995) propose models of heterogeneity for both residual and other components of variance. Recently, a significant extension of the model was suggested by San Cristobal-Gaudy *et al.* (1998), who introduced additive genetic effects influencing the log residual variances of the observations, thereby producing a genetically structured variance heterogeneity.

The model described by San Cristobal-Gaudy *et al.* (1998) is interesting from an evolutionary as well as from an animal breeding perspective. A model postulating that environmental sensitivity is partly under genetic control is relevant in studies of canalization (Waddington, 1957; Rendel, 1977), genetic assimilation (Waddington, 1953), reaction norms (Falconer & Mackay, 1996) and genotype by environment interaction. It can also provide an explanation for the increased levels of phenotypic variation often observed in experimental divergent lines selected for both higher and lower expressions of a trait (e.g. Clayton & Robertson, 1957). From an animal breeding point of view, there are at least two issues. First, if phenotypic variation is partly under genetic control, predictions of selection response based on the classical model may be incorrect, and one might wish to know under what conditions the possible error is important. Second, homogeneity of a final product often contributes to economic efficiency. It is therefore relevant to understand whether selection for a trait in a particular direction is likely to result in increased or decreased levels of phenotypic variation.

* Corresponding author. e-mail: sorensen@inet.uni2.dk

From an inferential point of view, the San Cristobal-Gaudy *et al.* (1998) model introduces considerable additional complexity. San Cristobal-Gaudy *et al.* (1998) use an EM algorithm for computing maximum likelihood estimates but several approximations are used in order to overcome computational difficulties. Furthermore, the distributions of maximum likelihood estimates and test statistics are hard to determine. Also, model checking based on residuals is complicated by the fact that usual standardized residuals are far from being independent standard normal.

In this paper, we present results of a case study in which normal mixed models with heterogeneous residual variances are fitted to pig litter size data originating from a selection experiment discussed in Sorensen *et al.* (2000). The paper has two main objectives. The first is to investigate the presence of additive genetic effects influencing the log residual variance and their possible correlation with the genetic effects influencing the expected litter sizes. Our second objective is to demonstrate that Bayesian methods provide an attractive alternative to traditional frequentist methods for complex models like the San Cristobal-Gaudy *et al.* (1998) model. Because conclusions concerning a genetically structured variance heterogeneity can be highly sensitive to the choice of model, we stress the importance of a thorough model assessment. In particular, we advocate the use of posterior predictive model assessment. We fit four models with different levels of complexity in the residual variance structure, with the simplest being the standard repeatability homogeneous variance additive genetic model. Fitting the simpler models first allows initial explorative analyses in which features of the data can be examined using various plots. The fitting of several models further enables the use of global measures of fit (Bayes factors and two related criteria) for assessing the possible superiority of the complicated models. We also compare the models in terms of performance for selection and prediction.

The paper is organized as follows. Section 2 provides a short description of the data, presents the four models under consideration and introduces statistics used for posterior predictive model assessment. In Section 3, we show the results of the analyses of the litter size data. Section 4 contains a discussion and conclusions.

The fitting of the highly parameterized Bayesian models using a Markov chain Monte Carlo (MCMC) algorithm (Robert & Casella, 1999; Sorensen & Gianola, 2002) requires some refinements in the MCMC algorithm in order to achieve efficient mixing. Details are given in the Appendix, which also includes a brief review of posterior predictive model assessment and of the three criteria of model comparison used in this study.

2. Methods

(i) *Data*

The data originate from a large scale selection experiment for total number of piglets born per litter (referred to as litter size hereinafter) carried out in the beginning of the 1990s and described in Sorensen *et al.* (2000). Briefly, selection of high intensity in a base population with 8988 litter size records was practiced only once, based on predicted additive genetic values obtained from a repeatability additive genetic model that included herd, season, type of insemination and parity as classification variables. Sows with up to nine parities from 82 registered breeding herds from the Danish pig breeding programme contributed records on litter size that were used to compute the additive genetic values. The selection experiment comprised one selected and one control line. Females in the selected and control lines produced two parities only. Animals from these lines were reared contemporaneously in a common research farm and were randomly allocated to pens. The complete data file consists of 10 060 litter size records from 4149 sows and the selected and the control lines include 1072 litter size records. The pedigree file includes 6437 individuals.

(ii) *Models fitted*

Four models of the form

$$\mathbf{y}|\mathbf{b}, \mathbf{p}, \mathbf{a}, (\sigma^2_{i,M})_{i=1,\ldots,n} \sim N(\mathbf{Xb}+\mathbf{Wp}+\mathbf{Za}, \\ \mathrm{diag}(\sigma^2_{i,M}, i=1,\ldots,n)), \quad M=1,\ldots,4 \qquad (1)$$

with increasing levels of complexity at the level of the log residual variance are fitted to the data vector $\mathbf{y}$ of length n. In Eqn 1, $\mathbf{b}$ is a vector containing the effects of four categorical covariates: parity (nine levels), season (four levels), herd (82 levels) and type of insemination (natural or artificial), $\mathbf{p}$ is a vector of permanent environmental effects with 4149 elements, $\mathbf{a}$ is a vector of additive genetic values with 6437 elements, and $\sigma^2_{i,M}$ is the residual variance for the ith observation under the Mth model. The matrices $\mathbf{X}$, $\mathbf{W}$ and $\mathbf{Z}$ are known incidence matrices.

Model 1 is the standard repeatability additive genetic model with homogeneous variance $\sigma^2_{i,1}=\exp(\tilde{b})$ for some parameter $\tilde{b}$ in $\mathbb{R}$. Model 2 allows different residual variances for different levels of the categorical covariates so that $\sigma^2_{i,2}$ is of the form

$$\sigma^2_{i,2} = \exp(\tilde{\mathbf{x}}_i'\tilde{\mathbf{b}}), \qquad (2)$$

where $\tilde{\mathbf{x}}_i'$ is the ith row in an incidence matrix $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{b}}$ is a parameter vector that includes effects of parity and type of insemination. In the case of Model 3, the residual variance is assumed to be partly under

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

genetic control. Thus,

$$\sigma^2_{i,3} = \sigma^2_{i,2} \exp(\mathbf{z}'_i \tilde{\mathbf{a}}) \tag{3}$$

where $\mathbf{z}'_i$ is the ith row of $\mathbf{Z}$ and $\tilde{\mathbf{a}}$ is a column vector with the 6437 additive genetic values affecting residual variation in litter size. Model 4 allows for an extra permanent environmental effect $\tilde{\mathbf{p}}$ so that

$$\sigma^2_{i,4} = \sigma^2_{i,3} \exp(\mathbf{w}'_i \tilde{\mathbf{p}}) \tag{4}$$

where $\mathbf{w}'_i$ is the ith row of $\mathbf{W}$.

(a) *Prior distributions*

The following prior distributions were assigned to the location parameters:

$$\mathbf{b} \sim N(0, \mathbf{I}_1 10^5), \quad \mathbf{a}|\sigma^2_a \sim N(0, \mathbf{A}\sigma^2_a), \quad \mathbf{p}|\sigma^2_p \sim N(0, \mathbf{I}_2\sigma^2_p). \tag{5}$$

In Eqn 5, $\mathbf{I}_1$ and $\mathbf{I}_2$ are identity matrices, and $\mathbf{A}$ is the known additive genetic relationship matrix of dimension 6437 × 6437. The scalars σ^2_a and σ^2_p are the additive genetic variances for litter size and the permanent environmental variance, respectively. For these, scaled inverted $vS\chi_v^{-2}$ prior distributions were chosen with $v = 4$ and $S = 0{\cdot}45$. This results in *a priori* means of σ^2_a and σ^2_p equal to 0·90 (the sensitivity of the posterior results to the choice of prior for σ^2_a and σ^2_p is studied in Section 3ii).

In case of Model 1 and Model 2 the parameters $\tilde{b}$ and $\tilde{\mathbf{b}}$ are *a priori* $N(0, 10^5)$ and $N(0, \mathbf{I}_3 10^5)$, respectively. For Model 3 the vectors $\mathbf{a}$ and $\tilde{\mathbf{a}}$ are assumed to have the following multivariate normal distribution:

$$\left.\begin{pmatrix} \mathbf{a} \\ \tilde{\mathbf{a}} \end{pmatrix}\right| \mathbf{G} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{G} \otimes \mathrm{A}\right) \tag{6}$$

where the 2 × 2 matrix $\mathbf{G}$ is

$$\mathbf{G} = \begin{bmatrix} \sigma^2_a & \rho\sigma_a\sigma_{\tilde{a}} \\ \rho\sigma_a\sigma_{\tilde{a}} & \sigma^2_{\tilde{a}} \end{bmatrix}. \tag{7}$$

In Eqn 7, the scalar $\sigma^2_{\tilde{a}}$ is the additive genetic variance for $\tilde{\mathbf{a}}$ and ρ is the coefficient of genetic correlation in the joint distribution of $\mathbf{a}$ and $\tilde{\mathbf{a}}$. For $\sigma^2_{\tilde{a}}$, the same scaled inverted χ^2 prior distribution is chosen as for σ^2_a, and ρ is *a priori* uniformly distributed in $(-1, 1)$. In the case of Model 4,

$$\tilde{\mathbf{p}}|\sigma^2_{\tilde{p}} \sim N(0, \mathbf{I}_2\sigma^2_{\tilde{p}}), \tag{8}$$

where $\sigma^2_{\tilde{p}}$ is the component of variance caused by permanent environmental effects affecting residual variation in litter size. The scaled inverted χ^2 prior distribution is also chosen for $\sigma^2_{\tilde{p}}$.

The parameters $\mathbf{b}$, $\tilde{b}$ (or $\tilde{\mathbf{b}}$), σ^2_a, σ^2_p, ρ, $\sigma^2_{\tilde{a}}$ and $\sigma^2_{\tilde{p}}$ are assumed to be *a priori* independent, and given these parameters, also $\mathbf{p}$, $\tilde{\mathbf{p}}$ and $(\mathbf{a}^\top, \tilde{\mathbf{a}}^\top)$ are *a priori* independent.

(b) *Mean–variance relationships*

Models 3 and 4 allow for genetic dependence between the mean and the variance of the sampling distribution. Define the random variable

$$\mathbf{u}^\top = \tilde{\mathbf{a}}^\top - E(\tilde{\mathbf{a}}|\mathbf{a})^\top,$$

independent of $\mathbf{a}$, where $E(\tilde{\mathbf{a}}|\mathbf{a})^\top = (\sigma_{\tilde{a}}\rho/\sigma_a)\mathbf{a}^\top$. Then

$$\mathbf{u}|\sigma^2_{\tilde{a}}, \rho \sim N(\mathbf{0}, \sigma^2_{\tilde{a}}(1-\rho^2)\mathbf{A}),$$

and $(\mathbf{a}^\top, (\sigma_{\tilde{a}}\rho/\sigma_a)\mathbf{a}^\top + \mathbf{u}^\top)$ is distributed as $(\mathbf{a}^\top, \tilde{\mathbf{a}}^\top)$. If $\rho = 1$, we can write

$$\sigma^2_{i,3} = \sigma^2_{i,2} \exp((\sigma_{\tilde{a}}/\sigma_a)\mathbf{z}'_i\mathbf{a})$$

so that a deterministic relation between the mean and the sampling variation is obtained, given the model parameters. If $\rho = 0$,

$$\sigma^2_{i,3} = \sigma^2_{i,2} \exp(\mathbf{z}'_i\mathbf{u}),$$

which corresponds to a genetically structured variance homogeneity that is unrelated to the mean.

Notice finally that, if the components of $(\sigma_{\tilde{a}}\rho/\sigma_a)\mathbf{a} + \mathbf{u}$ are small, then

$$\exp(\mathbf{z}'_i((\sigma_{\tilde{a}}\rho/\sigma_a)\mathbf{a} + \mathbf{u})) \approx 1 + (\sigma_{\tilde{a}}\rho/\sigma_a)\mathbf{z}'_i\mathbf{a} + \mathbf{z}'_i\mathbf{u}$$

so that the mean–variance relationship is approximately linear.

(iii) *Posterior predictive assessment of the models for the litter size data*

The adequacy of a given statistical model may be assessed by comparing the observed value of some statistic with its sampling distribution under the model. This basic idea underlies posterior predictive model assessment (reviewed in Appendix i), in which a statistic possibly depending on unknown parameters is compared with its posterior predictive distribution. Below, we use standardized residuals to construct certain discrepancy statistics targeted to measure a specific putative structure in the data that the current model fails to address. In our analysis, we also consider various histograms and quantile plots based on posterior predictive realizations of the standardized residuals to check (for example) the normality assumption of our models. Posterior predictive assessment using the MCMC output is trivial and allows a graphical investigation of properties of the data which can be very revealing. It can also provide guidance regarding extensions of the model that are worth pursuing before embarking on the time consuming programming work needed to implement an extension of the model.

After fitting Model 1, possible variance heterogeneity associated with the four covariates parity, season, herd and insemination is studied using discrepancy statistics

$$T_{j,l}(\mathbf{y},\theta_1)=\frac{1}{m_{j,l}}\sum_{L_{ij}=l}\frac{(y_i-\mu_i)^2}{\sigma^2_{i,1}}-1,$$

$$j=\text{sea, ins, par, her};\ l=1,\ldots,n_j, \qquad (9)$$

where j is an index for the four covariates, l is an index for the n_j levels of the jth covariate and $L_{ij}=l$ if the ith record belongs to the lth level of the jth covariate. The vector θ_1 contains the parameters of Model 1, $m_{j,l}$ is the number of records with level l for the jth covariate, μ_i is the ith element in $\mathbf{Xb}+\mathbf{Wp}+\mathbf{Za}$, and $(y_i-\mu_i)^2/\sigma^2_{i,1}$ is the squared standardized residual associated with record i. Under Model 1, the terms under the summation sign are independent, single degree of freedom chi-square random variables; therefore the expected value of Eqn 9 is zero, so large or small values of $T_{j,l}$ indicate a possible variance heterogeneity associated with the jth covariate.

In order to study a possible association between residual variation and additive genetic values, the following discrepancy statistics are constructed. For a partitioning $-\infty=t_1<t_2<\cdots<t_{k-1}<t_k=\infty$ with corresponding intervals $I_j=[t_j,t_{j+1}[$ we consider statistics

$$T_j(\mathbf{y},\theta_2)=\frac{1}{m_{I_j}}\sum_{\mathbf{z}_i'\mathbf{a}\in I_j}\frac{(y_i-\mu_i)^2}{\sigma^2_{i,2}}-1,\quad j=1,2,\ldots,k, \qquad (10)$$

where m_{I_j} is the number of observations with $\mathbf{z}_i'\mathrm{a}\in I_j$ and θ_2 is the vector of parameters associated with Model 2. The statistic T_j is thus the average of squared standardized residuals whose corresponding genetic value falls into the jth interval minus one. A possible association between residual variation and additive genetic variation affecting litter size can be studied by comparing the joint posterior distributions of $(T_j(\mathbf{y},\theta_2),\ T_j(\mathbf{y}_{rep},\theta_2))$, $j=1,\ldots,k$. In Section 3i, we summarize the joint posterior predictive distributions using boxplots for posterior predictive realizations of $T_j(\mathbf{y},\theta_2)-T_j(\mathbf{y}_{rep},\theta_2)$ plotted against interval number for each of $k=10$ intervals, where $t_k=-1{\cdot}6+0{\cdot}4(k-2)$, $k=2,\ldots,9$. The length of the intervals were chosen to accommodate a similar number of observations in each (approximately 1000).

In Section 3i, the plot based on the discrepancy statistics T_j defined in Eqn 10 gives another form of insight concerning the relationship between residual variation and additive genetic values than that derived by the estimate of the correlation coefficient alone. However, in order to reveal a putative feature using a discrepancy statistic, the feature under study must induce a sufficient degree of structure in the data. For example, it is not obvious that posterior predictive model assessment is helpful for detection of variance heterogeneity due to the environmental effects $\tilde{\mathbf{p}}$, because the small number of records per sow implies that a possible pattern of variance heterogeneity is hard to separate from noise.

3. Results

The results reported in this section for each model are computed using MCMC samples obtained from running 1 000 000 iterations of the MCMC algorithm described in the appendix. In Section 3ii, we report confidence intervals for the Monte Carlo estimates of various posterior means in order to give an idea of the accuracy of the Monte Carlo computations.

(i) *Model building using posterior predictive model assessment*

After fitting Model 1, we perform an exploratory analysis using the discrepancy statistics $T_{j,l}$ to disclose possible variance heterogeneities associated with the categorical explanatory variables. The posterior predictive distribution of $(T_{\text{ins},1}(\mathbf{y},\theta_1)-T_{\text{ins},2}(\mathbf{y},\theta_1),\ T_{\text{ins},1}(\mathbf{y}_{rep},\theta_1)-T_{\text{ins},2}(\mathbf{y}_{rep},\theta_1))$ is displayed using the left scatter plot in Fig. 1. The plot indicates that a higher variance is associated with artificial insemination ($l=1$) than with natural insemination ($l=2$) because all points fall below the identity line. The right-hand plot in Fig. 1 shows pairs of boxplots based on posterior predictive samples of $T_{\text{par},l}(\mathbf{y},\theta_1)-T_{\text{par},l}(\mathbf{y}_{rep},\theta_1)$, $l=1,\ldots,9$. The plot suggests that residual variances are lower for parity one than for parities greater than one. Similar plots (not shown) do not indicate a pattern of variance heterogeneity associated with season and herd.

Based on the exploratory analysis for Model 1, we obtain Model 2 by letting the log residual variances depend on the insemination covariate and a parity covariate with six levels obtained by grouping all records for parity greater than or equal to six (there are rather few records with parity greater than six and our MCMC algorithm works best if the covariates for the residual variance do not have too different numbers of records for each level). As the next step in the model building process, we explore under Model 2 the possibility of a genetic association between residual variation and additive genetic variation for litter size. This involves a considerable extension of the model and a posterior predictive model assessment based on the discrepancy statistics T_j is helpful to decide whether such an effort is worth pursuing. A result of this explorative analysis is presented in Fig. 2, in

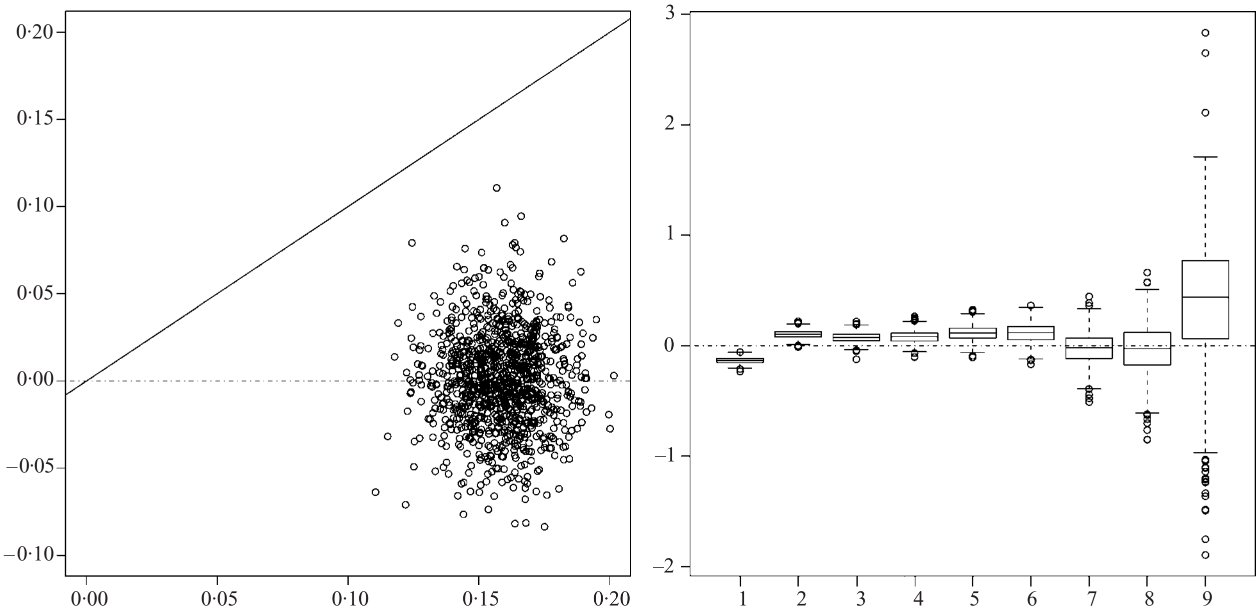

Fig. 1. (Left) Plot of simulated posterior predictive realizations of ($T_{\text{ins},1}$ ($\mathbf{y}$, θ_1) − $T_{\text{ins},2}$ ($\mathbf{y}$, θ_1), $T_{\text{ins},1}$ ($\mathbf{y}_{rep}$, θ_1) − $T_{\text{ins},2}$ ($\mathbf{y}_{rep}$, θ_1)) (solid line is the identity). (Right) Boxplots for posterior predictive realizations of $T_{\text{par},l}(\mathbf{y}, \theta_1) - T_{\text{par},l}(\mathbf{y}_{rep}, \theta_1)$, $l = 1,\ldots, 9$.

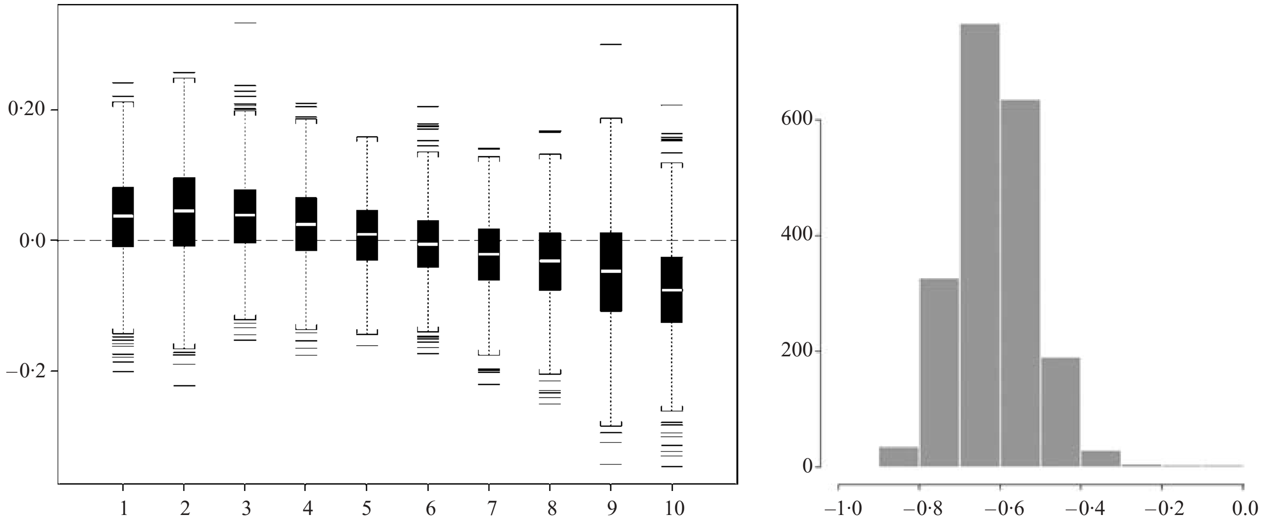

Fig. 2. (Left) Boxplots for posterior predictive realizations of $T_j(\mathbf{y}, \theta_2) - T_j(\mathbf{y}_{rep}, \theta_2)$ (see Section 2iii) plotted against interval number $j = 1,\ldots, 10$. (Right) Estimated marginal posterior distribution for ρ under Model 4 (see Section 3ii).

which the boxplots for posterior predictive realizations of $T_j(\mathbf{y}, \theta_2) - T_j(\mathbf{y}_{rep}, \theta_2)$ show a negative association with additive genetic values for the trait, indicating that sows of high genetic merit are likely to show less environmental variability.

Figure 2 (left) motivates extending Model 2 to account for an association between additive genetic values and residual variation. Figure 2 (left) also highlights the fact that, with the exception of the first three intervals, the genetic association between residual variation and additive genetic variation is fairly linear throughout the whole range of additive genetic values.

Finally, Fig. 3 shows quantile plots for posterior realizations of the standardized residuals $(y_i - \mu_i)/\sigma_{i,4}$, $i = 1,\ldots, n$, under Model 4, in which each of the sets of residuals are computed using approximately independent posterior realizations of the model parameters. Figure 3 shows that the marginal distribution of each of the posterior realizations of the residuals are fairly close to the standard normal distribution.

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

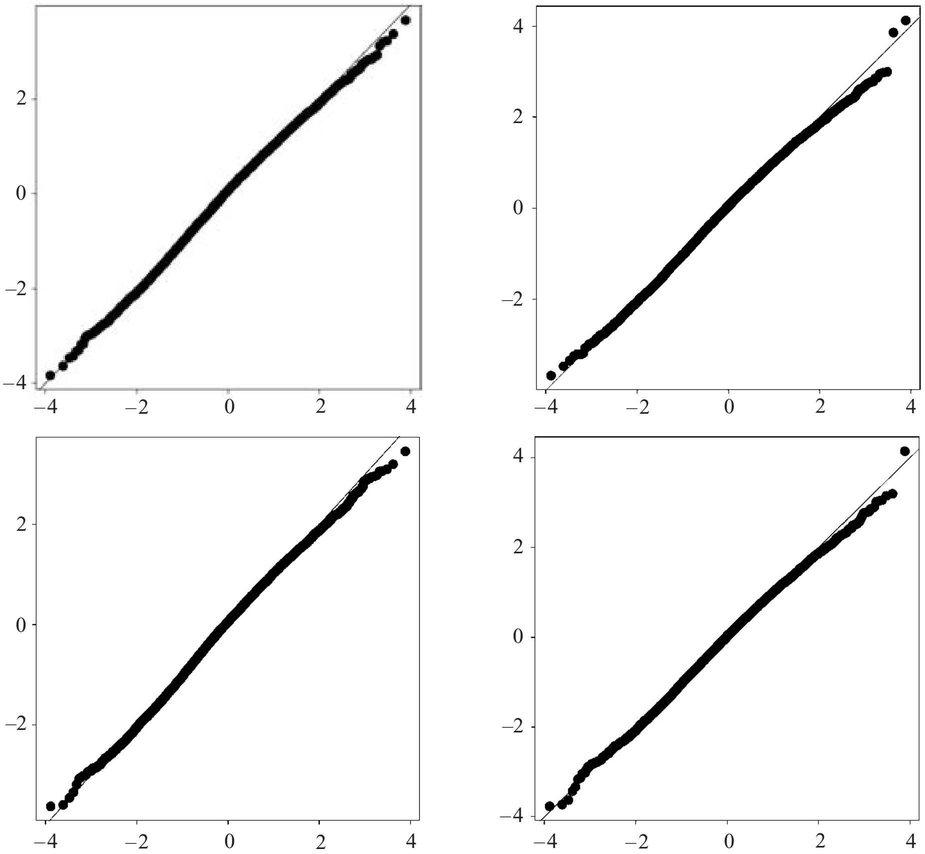

Fig. 3. Quantiles of four posterior realizations of $(y_i-\mu_i)/\sigma_{i,4}$, $i=1,\ldots,n$ against quantiles of the standard normal distribution. The solid line is the identity.

(ii) *Posterior distributions*

Table 1 shows Monte Carlo estimates of posterior means and 95% posterior intervals for chosen parameters based on Models 1, 2, 3 and 4. For Model 1, $\tilde{b}_0$ is the log residual variance and, for Models 2, 3 and 4, $\tilde{b}_0$ is the log residual variance for a record with parity one and natural insemination. The differences $\tilde{\delta}_j=\tilde{b}_{j,2}-\tilde{b}_{j,1}$, $j=$ins, par, are the effects on the log residual variance of moving from level one to two for insemination and parity, respectively. The pattern of variance heterogeneity between artificial insemination and natural insemination and parity two and parity one records as measured by $\tilde{\delta}_{\text{ins}}$ and $\tilde{\delta}_{\text{par}}$ is similar across all four models. The posterior intervals for these differences are bounded away from zero so there is strong evidence for variance heterogeneity associated with insemination and parity. The posterior mean of heritability based on Model 1 is 0·16. The posterior means of heritability based on Model 2 are 0·17, 0·20, 0·13 and 0·15 for levels one one, one two, two one and two two of parity and insemination, respectively.

The Monte Carlo estimates of the posterior mean of the additive genetic variance σ^2_a are of similar magnitude in the case of Models 1 and 2 and are a little larger for Models 3 and 4. Estimates of the permanent environmental variance σ^2_p are similar for Models 1, 3 and 4.

Estimated marginal posterior distributions σ^2_a, σ^2_p, $\sigma^2_{\tilde{a}}$, and $\sigma^2_{\tilde{p}}$ based on Model 4 are given in Fig. 4. The dark lines superimposed in each of the four figures is the density of the prior scaled inverted distribution with parameters $v=4$ and $S=0{\cdot}45$ and a prior mode equal to 0·30. The estimated marginal distribution under Model 4 for the correlation coefficient ρ is shown in the right plot in Fig. 2. The posterior intervals in Table 1 and the estimated marginal distributions for $\sigma^2_{\tilde{a}}$ and ρ which are bounded away from zero provide strong evidence for the presence of additive genetic values which affect the residual variance and are negatively correlated with the additive genetic values affecting litter size. The posterior distribution of $\sigma^2_{\tilde{p}}$ under Model 4 also provides evidence for the existence of permanent environmental effects influencing residual variation.

The estimates of the posterior characteristics are subject to Monte Carlo error. Estimates of the Monte Carlo error yield the following confidence intervals for the estimated posterior means under Model 2 (1·35; 1·39) (σ^2_a) (0·70; 0·73) (σ^2_p), (1·87; 1·87) ($\tilde{b}_0$), (−0·16; −0·15) ($\tilde{\delta}_{\text{ins}}$) and (0·33; 0·34) ($\tilde{\delta}_{\text{par}}$). For

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

Table 1. *Monte Carlo estimates of posterior means (first row for each model) and 95% posterior intervals (second row for each model) of chosen parameters of Models 1, 2, 3 and 4. See Sections 2 and 3ii for explanation of symbols*

Model	σ_a^2	σ_p^2	$\tilde{b}_0$	$\tilde{\delta}_{\text{ins}}$	$\tilde{\delta}_{\text{par}}$	$\sigma_{\tilde{a}}^2$	$\sigma_{\tilde{p}}^2$	ρ
1	1·40	0·60	2·00	–	–	–	–	–
	1·02; 1·81	0·29; 0·90	1·96; 2·04	–	–	–	–	–
2	1·37	0·71	1·87	−0·15	0·34	–	–	–
	0·97; 1·81	0·39; 1·06	1·80; 1·95	−0·22; −0·08	0·25; 0·42	–	–	–
3	1·58	0·60	1·78	−0·16	0·34	0·11	–	−0·57
	1·13; 2·00	0·31; 0·96	1·65; 1·90	−0·24; −0·09	0·25; 0·43	0·08; 0·15	–	−0·72; −0·41
4	1·62	0·60	1·77	−0·17	0·35	0·09	0·06	−0·62
	1·20; 2·05	0·30; 0·92	1·65; 1·89	−0·25; −0·09	0·26; 0·44	0·06; 0·13	0·05; 0·09	−0·80; −0·43

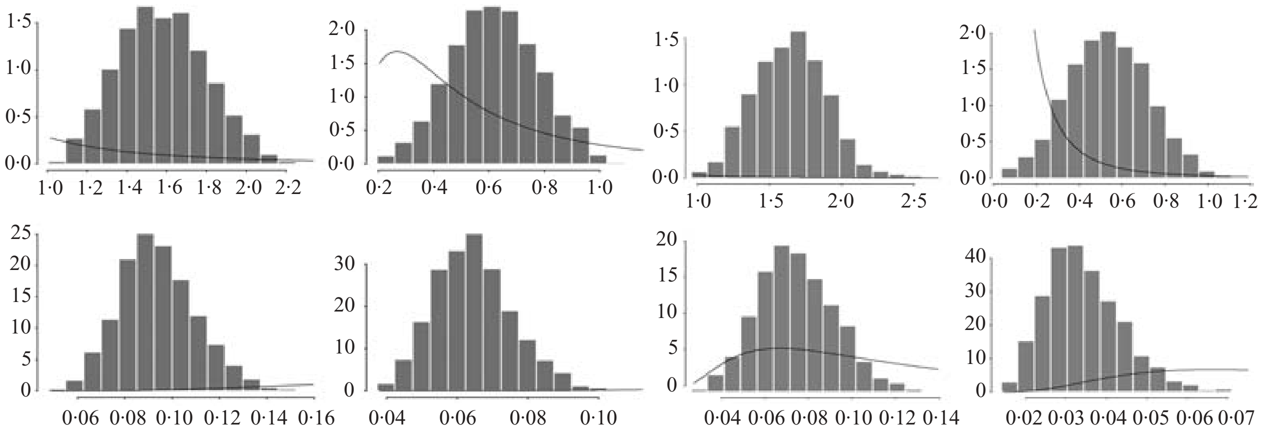

Fig. 4. (Left) Monte Carlo estimates of σ_a^2 and σ_p^2 (top) and $\sigma_{\tilde{a}}^2$ and $\sigma_{\tilde{p}}^2$ (bottom). The thick lines represent the prior scaled inverted χ^2 densities with parameters $\nu=4$ and $S=0·45$. (Right) As left but with alternative choice of prior ($\nu=4$ and $S=0·10$).

Model 4, we obtain (1·55; 1·68) (σ_a^2) (0·56; 0·64) (σ_p^2), (1·77; 1·78) ($\tilde{b}_0$), (−0·17; −0·16) ($\tilde{\delta}_{\text{ins}}$), (0·34; 0·35) ($\tilde{\delta}_{\text{par}}$), (0·09; 0·10) ($\sigma_{\tilde{a}}^2$), (0·06; 0·07) ($\sigma_{\tilde{p}}^2$) and (−0·64; −0·61) ($\rho$). The conclusions above regarding the posterior means are not changed by consideration of the Monte Carlo error.

In order to study the influence of the prior distribution on the inferences, a model identical to Model 4 was fitted, except that the scale parameter S of the scaled inverted χ^2 prior distributions was set equal to 0·1 instead of 0·45. This results in a prior mode equal to 0·067. The posterior means and 95% posterior intervals for σ_a^2, σ_p^2, $\sigma_{\tilde{a}}^2$, $\sigma_{\tilde{p}}^2$ and ρ with $S=0·1$ are 1·64 (1·67; 1·90), 0·54 (0·17; 0·91), 0·07 (0·04; 0·11), 0·03 (0·02; 0·06), and −0·60 (−0·78; −0·41). Except for $\sigma_{\tilde{p}}^2$ only relatively small changes compared to Table 1 are observed; see also the right plot in Fig. 4.

(a) *Further model assessment*

One may argue that the strong evidence of a negative correlation between genetic values for litter size and residual variance is an artefact caused by failure of the assumption of normality for litter size, which cannot exceed some minimum and maximum values. One could therefore anticipate a right truncated distribution of litter sizes and consequently a small variance for high producing sows. Moreover, it is not unlikely that sows producing extremely low litter sizes have been temporarily exposed to conditions, such as disease, that determined low productivity at a given parity. Once the condition is removed, the sow reverts to some normal level. This mechanism, not accounted for by any of the four models, would generate large variation among low producing individuals.

Figure 5 shows histograms of litter sizes for high producing sows (left) and for low producing sows (right). The left histogram is moderately skewed but is, however, based on the raw records with no correction for the effects of explanatory variables and genetic values. More incisively, we can consider the distribution of residuals for records with high values of additive genetic values for litter size. Figure 6 shows histograms based on posterior realizations of the standardized residuals $(y_i-\mu_i)/\sigma_{i,4}$ for which the associated posterior realizations of additive genetic

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

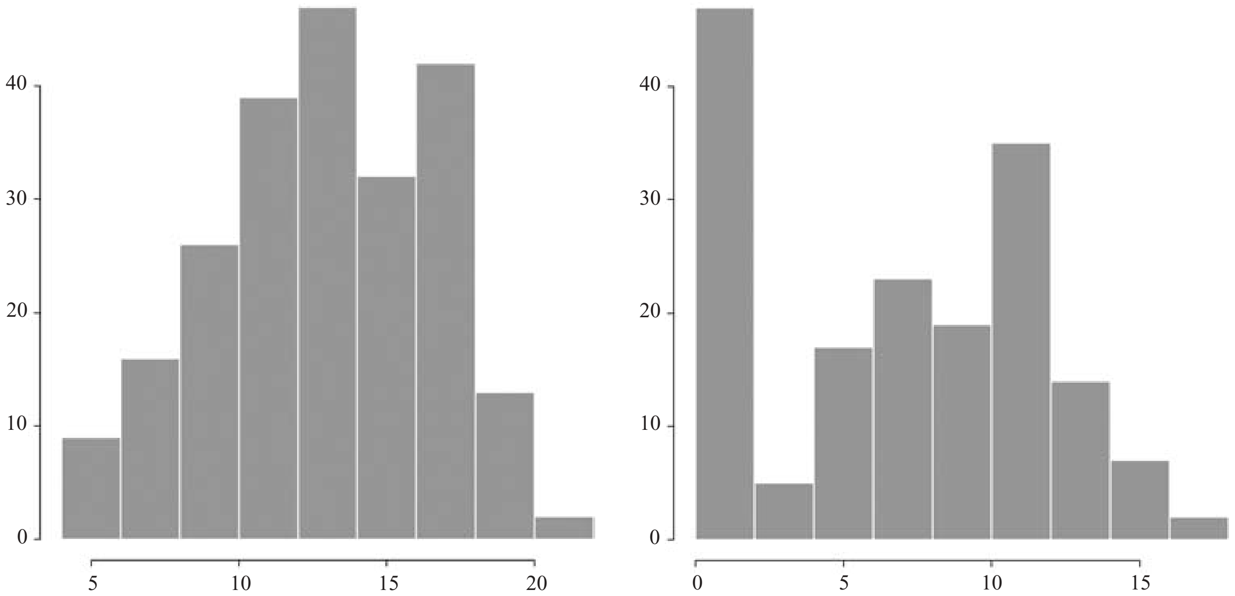

Fig. 5. Histograms of litter sizes for sows with more than one litter, and that produced at least one litter of size greater than or equal to 18 (left), or at least one litter of size one or two (right).

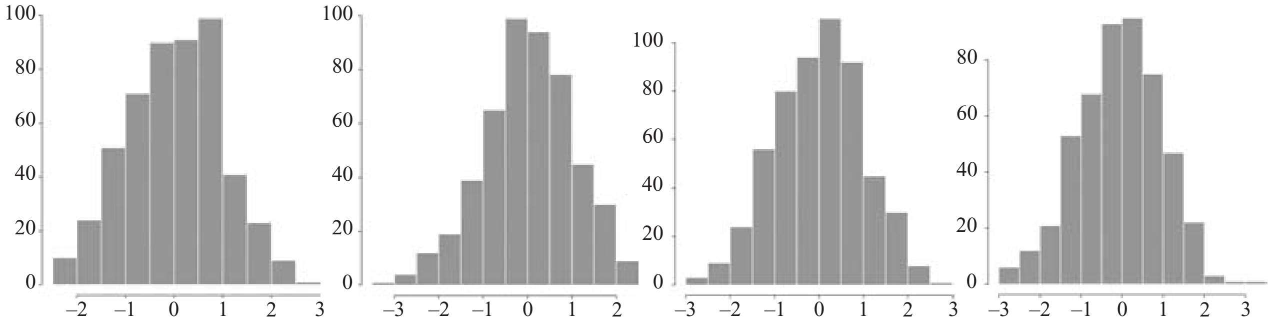

Fig. 6. Histograms based on posterior realizations of standardized residuals $(y_i - \mu_i)/\sigma_{i,4}$ with associated genetic value $\mathbf{z}_i'\mathbf{a}$ among the 200 largest.

values $\mathbf{z}_i'\mathbf{a}$ were among the 200 largest (this produces samples of residuals with an average size of around 500). A truncation effect is not apparent in these histograms.

Among low producing sows, there is a relatively high frequency of litter sizes of 1 or 2, causing asymmetry (Fig. 5, right). In order to study the influence of these low records on the inferences, Model 4 was fitted to a reduced data set that did not include the 64 litter size records equal to 1 or 2. The posterior means and 95% posterior intervals for σ^2_a, $\sigma^2_{\tilde{a}}$ and ρ based on the reduced data set are 1·41 (1·02; 1·86), 0·08 (0·05; 0·10), and −0·49 (−0·71; −0·26). Naturally, the posterior means of the variances are smaller when inferences are based on the truncated data set. However the decline is fairly small. Also the posterior distribution of the correlation coefficient is only mildly affected by exclusion of the low records.

The possibility of an artifact was investigated further by simulating data based on Model 2 with parameters given by the posterior means in the second row in Table 1 and on Model 3 with $\sigma^2_{\tilde{a}} = 0{\cdot}11$, ρ equal to −0·75, 0 or 0·75, and the other parameter values as for Model 2. After discretizing the data to the nearest integer, data points larger than 21 were set equal to 21, and those smaller than 1 were set equal to 1. Plots similar to the left plot in Fig. 2 but obtained for the simulated datasets are shown in Fig. 7. The leftmost plot shows, correctly, the lack of association between the residuals defined by Eqn 10 and additive genetic values, in data simulated using Model 2. So also does the rightmost plot, based on data generated with Model 3, $\rho = 0$. The second and third plots, based on Model 3 with $\rho = -0{\cdot}75$ and $\rho = 0{\cdot}75$, respectively, show the expected negative and positive associations.

These results provide evidence against the conjecture that the inferred correlation is an artefact. The model's postulate of the presence of additive genetic values affecting residual variation, correlated with additive genetic values influencing litter size, must be allowed to stand until further investigation.

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

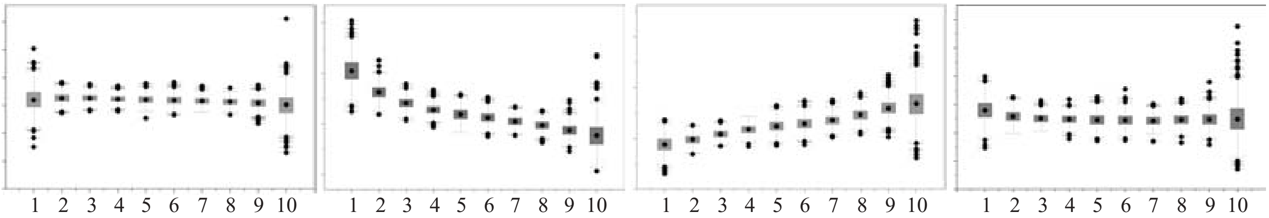

Fig. 7. Boxplots for posterior predictive realizations of $T_j(y, \theta_2)$ (see Section 2iii) plotted against interval number $j=1,\ldots,10$. From left to right: data simulated under Model 2; Model 3, $\rho=-0{\cdot}75$; Model 3, $\rho=0{\cdot}75$; Model 3, $\rho=0{\cdot}0$.

(iii) *Model comparison based on Bayes factors, posterior Bayes factors and deviance information criterion*

In this section, we compare the models using Bayes factors, posterior Bayes factors and the deviance criterion (DIC). These global criteria trade off model fit with model complexity and are reviewed in Appendix ii. The first and second rows in Table 2 show logarithms of the Bayes factors and of the posterior Bayes factors relative to Model 1. The third row shows differences of DICs from the DIC of Model 1. The following conclusions can be drawn: (i) The three criteria provide the same ranking of the models; (ii) Model 4 is by far the most favoured in all cases. For any reasonable set of prior probabilities assigned to the four models, the posterior probability for Model 4 is practically equal to one (even assigning a prior probability to Model 4 equal to 10^{-30} yields a posterior probability equal to 0·999); (iii) the biggest difference is observed between Model 2 and Model 3, and this is consistent for the three criteria. Thus, the presence of additive genetic values affecting residual variation is given high credibility by all three methods.

As mentioned in Appendix iib, the Monte Carlo estimator of the Bayes factor is known to be numerically unstable (Newton & Raftery, 1994). This is disclosed in Fig. 8 (left), which shows Monte Carlo estimates of the likelihood prior mean under Model 4 computed from samples that increase in size in steps of one, from 1 to 10 000 (the samples were obtained by subsampling each hundredth state of the MCMC output). Notice that, despite the downward jumps of the estimates, the ranking of the models quickly stabilizes to a consistent pattern. Also, the estimate of the logarithm of the likelihood posterior mean (Fig. 8, right) is influenced by occasional very large values in the posterior sample of the likelihood.

(iv) *Model comparison based on the models' predictive ability*

So far, the models have been used to understand a specific aspect of nature, namely, factors affecting the residual variance. The fact that Model 4 is assigned high credibility using posterior predictive model checking, or investigating the relevant posterior distributions, or via comparisons based on Bayes factors and related quantities, does not necessarily imply that it works much better than the simple models for prediction of 'future data', or unobservables such as additive genetic values.

Table 2. *Natural logarithms of Bayes factors* (*first row*), *posterior Bayes factors* (*second row*) *and DIC. All figures are expressed as differences from Model 1*

	Model 2	Model 3	Model 4
$\ln B_{i1}$	130	307	407
$\ln B^{p}_{i1}$	149	478	561
$\mathrm{DIC}_i-\mathrm{DIC}_1$	−154	−395	−450

(a) *Prediction of 'future data'*

The ability of the models to predict 'future data' is studied as follows. The 589 records from the second generation in the selection and control line are excluded from the full data set and 95% posterior predictive intervals for these observations are computed using the remaining (10 060 − 589) records. The proportion of excluded records falling into their predictive intervals are 0·95, 0·931, 0·941 and 0·941 for Models 1, 2, 3 and 4, respectively. Thus the models rank similarly in terms of the coverage properties of their predictive intervals.

(b) *Prediction of additive genetic values*

In quantitative genetic experiments, one may use models for prediction of additive genetic values either to infer response to selection or to select parents of the next generation. Define response to selection as the difference in average additive genetic value of individuals in the selected line and of unselected individuals and consider inferring selection response by means of the four models. Because response to selection is an unobserved random variable, it is not possible to define a discrepancy statistic that acts as a benchmark for testing the models. However, we carry out an informal test, by defining 'observed selection

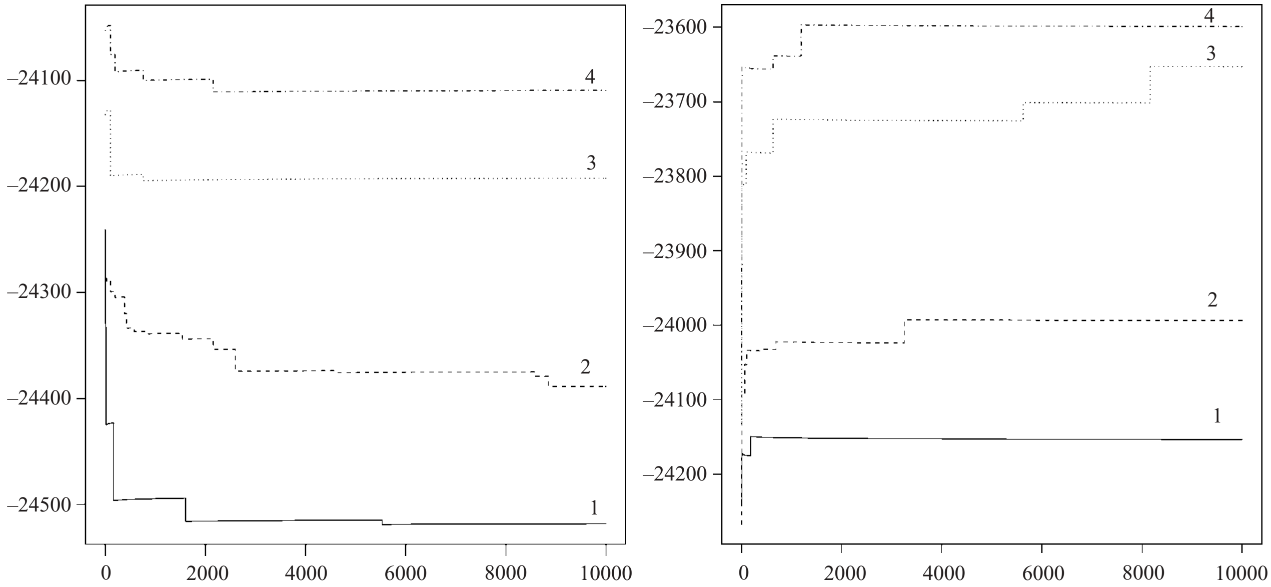

Fig. 8. Estimates of the logarithms of the likelihood prior mean (left) and the likelihood posterior mean (right) plotted against increasing sample size.

response' as the difference between the raw averages of the records in the selected and in the control line. This 'observed selection response' is equal to 0·40 piglets per litter. The means of the posterior distribution of response to selection and the 95% posterior intervals for Models 1, 2, 3 and 4 are 0·47 (0·26; 0·69), 0·43 (0·22; 0·65), 0·37 (0·17; 0·58) and 0·37 (0·15; 0·57), respectively. The models provide a similar picture of the response to selection, and the observed selection response falls comfortably within the 95% posterior intervals for all the models.

One way of studying the consequences of using the different models on selection decisions is to look at the number of individuals that are selected in common by the four models. Table 3 shows the number of individuals that overlap when the top 50 animals are selected on the basis of the posterior mean $E(\mathbf{a}|\mathbf{y}, M_i)$ of the additive genetic values computed with each of the four models. An interesting pattern emerges from the figures in the table. The largest degree of overlap is observed in comparisons involving the two models that do not include $\tilde{\mathbf{a}}$ (Models 1 and 2) and the two models that include $\tilde{\mathbf{a}}$ (Models 3 and 4). All the other comparisons where one of the models includes $\tilde{\mathbf{a}}$ and the other does not, show smaller amounts of overlap. Thus, inclusion of a genetic term as a factor influencing variance heterogeneity seems to have a bearing on selection decisions.

One might also wish to quantify the consequences of using the 'wrong' mode on selection response. Assume that Model 4 is the 'correct' model among those studied, and that selection of parents on the basis of $E(\mathbf{a}|\mathbf{y}, M_i)$ takes place with Models 1, 2 or 3. For model M_i, $i=1$, 2, 3, the selection response is estimated by the average of the $E(a_j|\mathbf{y}, M_4)$, where j is among the indices of the 50 animals selected using mode M_i. The overlap of 34, 38 and 47 individuals among the highest scoring 50 (Table 3) translates into a decrease in selection response relative to the selection response obtained with Model 4 of 5% if Model 1 is used, of 4% if Model 2 used, and of 0% if Model 3 is used.

Table 3. *Number of individuals that are selected in common by each pair of models*

Model comparison	1 *vs* 2	1 *vs* 3	1 *vs* 4	2 *vs* 3	2 *vs* 4	3 *vs* 4
Overlap	43	35	34	38	38	47

(v) *Model checking using 'conventional' residuals*

Our model assessment is based on the posterior predictive realizations of standardized residuals and derived discrepancy statistics. A conceptually simpler approach is to consider one single set of standardized residuals obtained by replacing μ_i and $\sigma^2_{i,M}$ by point estimates. However, in the complex models considered with a huge number of random effects, such residuals are far from being standard normal and are essentially useless for model assessment. Figure 9 (left) shows a quantile plot of standardized residuals obtained by replacing μ_i and $\sigma^2_{i,4}$ with their posterior means under Model 4. The standardized residuals are far from being standard normal and an overfitting effect is apparent because the empirical variance of

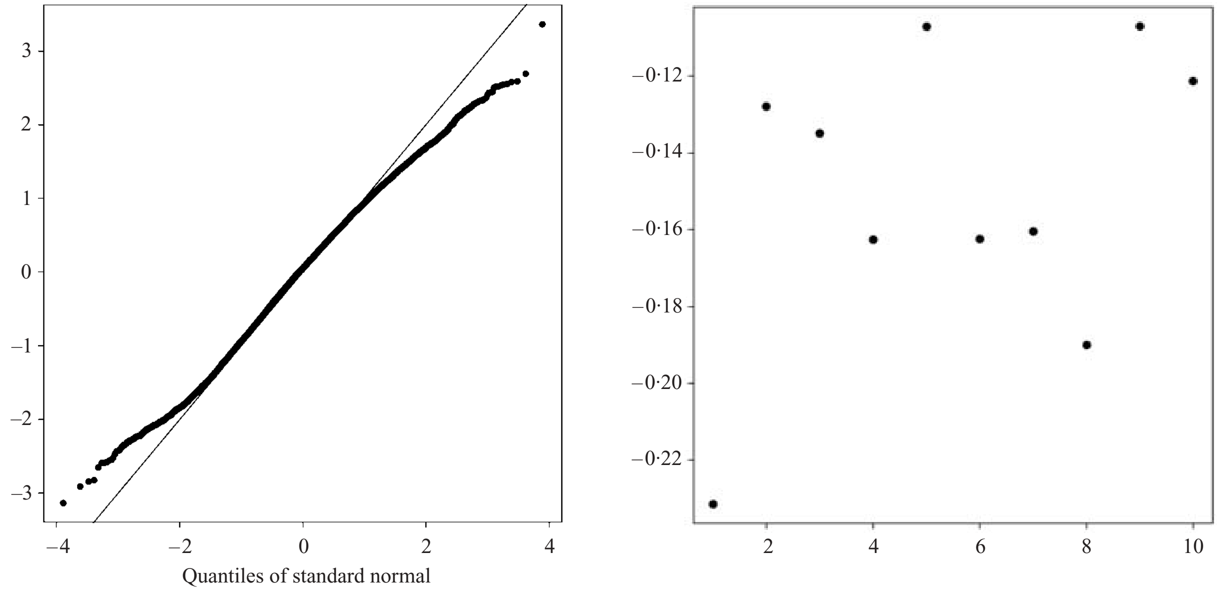

Fig. 9. (Left) Quantiles of standardized residuals obtained using point estimates of μ_i and $\sigma^2_{i,4}$ against quantiles of the standard normal distribution. (Right) The statistics $\bar{T}_j$ plotted against interval number $j=1,\ldots,10$.

the standardized residuals is only 0·82. The right-hand plot in Fig. 9 shows a plot of statistics $\bar{T}_j$ similar to the T_j in Section 2iii but with the unknown quantities replaced by posterior expectations, i.e.

$$\bar{T}_j = \frac{1}{m_{I_j}} \sum_{\mathbf{z}_i' E(\mathbf{a}|y)\in I_j} \frac{(y_i - E(\mu_i|\mathbf{y}, M_2))^2}{\exp(E(\log\sigma^2_{i,2}|\mathbf{y}, M_2))} - 1,$$
$$j=1,\ldots, 10,$$

where the intervals I_j are those used for the discrepancy statistics T_j and m_{I_j} here denotes the number of observations with $\mathbf{z}_i' E(\mathbf{a}|y) \in I_j$. No pattern of genetically structured variance heterogeneity is visible. Apparently, the subtle structures in the data are hidden by the process of posterior averaging.

4. Discussion

In this work, four models with increasingly complex residual variance structures are fitted to pig litter size data in order to investigate sources of variance heterogeneity and in particular the possible presence of additive genetic effects influencing the log residual variance. The models are compared using global criteria that trade off model fit with model complexity. The three such criteria used in this study generate the same ranking of the models and all give very strong evidence for a genetically structured variance heterogeneity. All criteria favour in particular Model 4, which also includes variance heterogeneity owing to permanent environmental effects. In agreement with this, the posterior distributions for $\sigma^2_{\tilde{a}}$, $\sigma^2_{\tilde{p}}$ are bounded away from zero under Model 4. The posterior distribution of ρ under Model 4 provides further evidence of a strong negative correlation between the additive genetic values influencing litter size and the additive genetic values affecting residual variation. The models are also compared according to the purposes for which they might be used. It is shown that models that rank very differently according to the global measures of fit are hardly distinguishable in terms of their ability to predict ‘future data’ or to infer response to selection. Yet a different result emerges when the models are used for selecting parents for breeding for larger litter size. Therefore, depending on the context, a simple model might be adequate even though it fails to address features of the data accounted for by the more complex models. This thought was put forward by Rubin (1984); we agree.

In the remaining part of the discussion, we provide introductory remarks about implications for selection of the above findings and alternative modelling approaches.

(i) *Implications for selection*

Models with genetically structured variance heterogeneity might contribute to an understanding of the process affecting mean–variability relations in natural and domestic populations. Previous work was based on a simple model (Lerner, 1954; Lewontin, 1964; Zhivotovsky & Feldman, 1992), assuming that environmental sensitivity decreases with the number of heterozygous loci. An extension postulates that pleiotropic effects at a finite number of loci act additively on the mean and variance (Gavrilets & Hastings, 1994; Hill, 2002). San Cristobal-Gaudy *et al.*

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

(1998) consider instead an infinitesimal model with correlation between the additive genetic values affecting the mean and those affecting the log residual variance and predict response to selection for canalization. Here, drawing from their work, we give a brief overview of changes of mean and variance caused by selection on an index designed to increase the mean of the trait and to reduce its variance.

Consider a simplified version

$$y|a,\tilde{a} \sim N(a,\ \exp(b+\tilde{a})) \text{ and } (a,\tilde{a})|\sigma_a^2, \sigma_{\tilde{a}}^2, \rho \sim N\left((0,0), \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{\tilde{a}} \\ \rho\sigma_a\sigma_{\tilde{a}} & \sigma_{\tilde{a}}^2 \end{bmatrix}\right) \tag{11}$$

of the genetically structured heterogeneous variance model. The phenotypic variance (variance of the marginal distribution of y) is $\sigma^2 = \sigma_a^2 + \exp(b+\sigma_{\tilde{a}}^2/2)$. Consider selecting on the index

$$I(y) = \bar{y} + kS_y^2, \tag{12}$$

where $\bar{y}$ is the average of the n records of an individual, $S_y^2 = \sum(y-\bar{y})^2/(n-1)$, the sample variance of the records of the individual, and k is a relative weight in units of inverse phenotypic standard deviations, assumed known. One might be interested in increasing the phenotype and decreasing its variance, in which case k would be negative. This index is arrived at empirically and no claims are made about its optimality properties. Under the model defined by Eqn 11, selection by truncation does not lead to mathematically tractable expressions. Instead, following Gavrilets & Hastings (1994), it will be assumed that directional selection can be described by the linear fitness function

$$w(y) = 1 + sI(y), \tag{13}$$

where s is a small quantity that defines the strength and direction of selection. Eqn 13 holds for weak selection, because s can be made arbitrarily small so that $w(y)$ is positive with probability essentially equal to one. The expectation of Eqn 13 over the distribution of y is

$$E[w(y)] = \bar{w} = 1 + s\left[k\exp\left(b + \frac{\sigma_{\tilde{a}}^2}{2}\right)\right]. \tag{14}$$

The fitness of genotype $(a, \tilde{a})$ is defined as

$$\begin{aligned} w(a,\tilde{a}) &= E[(1+sI(y)|a,\tilde{a})] \\ &= 1 + s\{a + k\exp[b+\tilde{a}]\}. \end{aligned} \tag{15}$$

Eqn 15 depends on the mean and variance of the conditional distribution of y given genotype $(a, \tilde{a})$, because Eqn 12 has a term that depends on the variance of the phenotypic records of the individual. If selection operates on Eqn 12, the mean value of a, or response to selection, is

$$\begin{aligned} R_a &= \iint a\frac{w(a,\tilde{a})}{\bar{w}} p(a,\tilde{a})da d\tilde{a} \\ &= \frac{1}{\bar{w}} \iint a\{1 + s[a + k\exp(b+\tilde{a})]\}p(a,\tilde{a})da d\tilde{a} \\ &= \frac{s}{\bar{w}}\sigma_a^2 + \frac{s}{\bar{w}}k\rho\sigma_a\sigma_{\tilde{a}}\exp(b+\sigma_{\tilde{a}}^2/2) \\ &= \frac{s}{\bar{w}}h^2\sigma^2 + \frac{s}{\bar{w}}k\rho\sigma_a\sigma_{\tilde{a}}(1-h^2)\sigma^2, \end{aligned} \tag{16}$$

where $h^2 = \sigma_a^2/\sigma^2$. The first term is the direct contribution from change in additive genetic value a. It can be positive or negative, depending on the direction of selection, which defines the sign of s. The second term is due to the correlation between a and $\tilde{a}$, and its sign depends on that of $sk\rho$. With small $\sigma_{\tilde{a}}^2/2$, the influence of these terms depends on the relative sizes of $\rho\sigma_a\sigma_{\tilde{a}}$ and σ_a^2. For the litter size data and using the posterior means in Table 1 from Model 3, $\rho\sigma_a\sigma_{\tilde{a}} = -0{\cdot}23$ and $\sigma_a^2 = 1{\cdot}55$, so the effect of the second term is fairly small. Similarly, the mean of $\tilde{a}$ with selection based on Eqn 12 is

$$\begin{aligned} R_{\tilde{a}} &= \iint \tilde{a}\frac{w(a,\tilde{a})}{\bar{w}} p(a,\tilde{a})da d\tilde{a} \\ &= \frac{s}{\bar{w}}\rho\sigma_a\sigma_{\tilde{a}} + \frac{s}{\bar{w}}k\sigma_{\tilde{a}}^2(1-h^2)\sigma^2. \end{aligned} \tag{17}$$

The mean value of the residual variance before selection is given by $\exp(b+\sigma_{\tilde{a}}^2/2)$. In the selected group, the expected value of the residual variance is

$$\iint \exp(b+\tilde{a})\frac{w(a,\tilde{a})}{\bar{w}} p(a,\tilde{a})da d\tilde{a}.$$

This integral can be obtained in closed form but a simpler expression results from a first order Taylor series expansion about $R_{\tilde{a}}$. This yields

$$E[\exp(b+\tilde{a})] = \exp(b+R_{\tilde{a}}),$$

where the expectation is taken with respect to the distribution of $(a, \tilde{a})$ in the selected group. The change in the mean residual variance is then, approximately,

$$\exp(b)[\exp(\sigma_{\tilde{a}}^2/2) - \exp(R_{\tilde{a}})]. \tag{18}$$

To gain a rough idea of the magnitude of the change in the mean residual variance, we computed the average $\tilde{a}$ among the 50 females with highest a. This yields $R_{\tilde{a}} = -0{\cdot}30$, which, again using the posterior mean of $\sigma_{\tilde{a}}^2$ from Model 3 in Table 1, results in a relative decline of

$$\frac{\exp(0{\cdot}05) - \exp(-0{\cdot}30)}{\exp(0{\cdot}05)} = 0{\cdot}29,$$

which is of the same order of magnitude as the decline of the additive genetic variance owing to the

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

Bulmer effect (Bulmer, 1980) among selected parents, assuming the infinitesimal model with homogeneous variance. Hill (2002) studied the effect of truncation selection on phenotype on changes of mean and phenotypic variance. Hill (2002) worked with a finite number of loci and therefore the effect of selection is due to changes in gene frequencies. The expressions for changes in mean and variance also include two terms as in Eqns 16 and 17, one arising from the effect of genes affecting the mean and the other from the effect of genes affecting the variance. Hill (2002) wonders whether terms such as $\sigma^2_{\tilde{a}}$ and $\rho\sigma_a\sigma_{\tilde{a}}$ are of sufficient magnitude to matter in prediction equations. Here, we provide evidence of their relevance. It is important to study further the dynamics of the genetically structured heterogeneous variance model under selection and to obtain a better understanding of the different factors intervening in the changes of genetic parameters.

(ii) *Extensions and alternative models*

It has been suggested to us to extend Models 2, 3 and 4 for the residual variances by adding independent terms e_i to the log residual variances $\log\sigma^2_{i,M}$, whereby a different variance for each record is obtained. If, for example, $S^2_i=\sigma^2_{i,M}\exp(e_i)$ is taken to be $\sigma^2_{i,M}\nu\chi_\nu^{-2}$, a scaled χ_ν^{-2}, and we assume

$$y_i|\mu_i, S^2_i \sim N(\mu_i, S^2_i)$$

then, integrating over the distribution $S^2_i|\sigma^2_{i,M}, \nu$, by the mixture property of the t_ν distribution (see, for example, Sorensen & Gianola, 2002, pages 28, 595), we obtain

$$y_i|\mu_i, \sigma^2_{i,M} \sim t_\nu(\mu_i, \sigma^2_{i,M}).$$

So the inclusion of e_i in the log residual variance essentially corresponds to choosing a more heavytailed sampling distribution. In the light of Fig. 3, this does not seem to be relevant for the litter size data.

Only normal sampling models are considered in this work. Perez-Enciso *et al.* (1993), using approximate methods, compared the quality of fit of Poisson and normal models, and did not obtain clear differences. Owing to underdispersion, our data are in fact in conflict with a Poisson sampling distribution. However, a biologically interesting alternative is to assume that the ith sow has a 'potential' n_i for producing litters of a certain size and that the litter size y_{ij} for the jth parity is Binomial(n_i, p_{ij}). The variable n_i could be assigned a Poisson prior distribution and $\log[p_{ij}/(1-p_{ij})]$ could be modelled via a mixed linear structure. Given the p_{ij}, the y_{ij} would then be marginally Poisson but the correlation caused by the common n_i would lead to a smaller within sow variation among the y_{ij} than if the y_{ij} had been independent. Inference about the binomial parameter n_i has been discussed for example in Draper & Guttman (1971) and Raftery (1988).

Appendix

(i) *Posterior predictive model assessment*

Given data $\mathbf{y}$, a basic idea for assessing the fit of a model M with sampling density $p(\cdot|\theta_M, M)$ depending on a known parameter θ_M is to compare the observed value of some univariate discrepancy statistic $T(\mathbf{y}, \theta_M)$ with its sampling distribution under $p(\cdot|\theta_M, M)$. If the observed value is atypical in the sense of being located in the extreme tails of its sampling distribution, we tend to reject the model. Equivalently, we might consider whether zero is an atypical value in the distribution of the difference $T(\mathbf{y}, \theta_M)-T(\mathbf{y}_{rep}, \theta_M)$ where $\mathbf{y}_{rep}$ is replicate data generated from the model $p(\cdot|\theta_M, M)$. When θ_M is unknown, a common plug-in approach is to replace θ_M by an estimate $\hat{\theta}_M$ and then proceed as if θ_M was known and equal to $\hat{\theta}_M$.

A Bayesian inference concerning θ_M is based on the posterior density

$$p(\theta_M|\mathbf{y}, M) \propto p(\theta_M|M)p(\mathbf{y}|\theta_M, M),$$

where $p(\theta_M|M)$ is the prior. Each θ_M generated from the posterior density is a candidate for the unknown value of the parameter so, instead of considering just one fixed value θ_M in the model assessment, we should consider a range of θ_M values generated from the posterior density. This is the idea of posterior predictive model assessment (Rubin, 1984; Gelman *et al.*, 1996), in which one considers the joint posterior predictive distribution of $T(\mathbf{y}, \theta_M)$ and $T(\mathbf{y}_{rep}, \theta_M)$. That is θ_M is generated from the posterior of θ_M and, given θ_M, $\mathbf{y}_{rep}$ is generated from $p(\cdot|\theta_M, M)$. As above, one might for example check whether zero is an extreme value in the posterior predictive distribution of $T(\mathbf{y}, \theta_M)-T(\mathbf{y}_{rep}, \theta_M)$. The marginal distribution of $\mathbf{y}_{rep}$ is given by the so-called posterior predictive density

$$\begin{aligned} p(\cdot|\mathbf{y}, M) &= \int p(\cdot|\theta_M, M)p(\theta_M|\mathbf{y}, M)d\theta_M \\ &= E_{\theta_M|\mathbf{y}, M}[p(\cdot|\theta_M, M)], \end{aligned} \tag{19}$$

i.e. the posterior expectation of $p(\cdot|\theta_M, M)$. Realizations of $\mathbf{y}_{rep}$ differ form the observed data $\mathbf{y}$ by the inherent sampling variation of the distribution $[\mathbf{y}_{rep}|\theta_M, M]$ and the posterior uncertainty of the parameters θ_M under model M, but differences may also occur because of failure of M to fit data $\mathbf{y}$.

In practice, we obtain MCMC draws $\theta_M^{(k)}$, $k=1,\ldots, K$, from the posterior distribution of θ_M and

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

subsequently generate replicate data $\mathbf{y}_{rep}^{(k)}$ given $\theta_M^{(k)}$. The posterior predictive distribution of $T(\mathbf{y}, \theta_M) - T(\mathbf{y}_{rep}, \theta_M)$ can then be studied using a boxplot of the differences $T(\mathbf{y}, \theta_M^{(k)}) - T(\mathbf{y}_{rep}^{(k)}, \theta_M^k)$. A less compact alternative is to consider a scatterplot of the pairs $(T(\mathbf{y}, \theta_M^{(k)}), T(\mathbf{y}_{rep}^{(k)}, \theta_M^k))$. The observed value of the discrepancy statistic is then atypical compared with its posterior predictive distribution if the points in the scatterplot are far from the identity line.

(ii) *Global measures of fit*

At the end of an exploratory exercise, several models might be available. Often, the best fitting model has a relatively large number of parameters and it is relevant to study whether the complexity of the model is supported by the available data. As a global measure of fit of a model, one might consider the prior mean of the likelihood, the posterior mean of the likelihood evaluated at the observed data or the posterior mean of the log likelihood. Criteria for model comparison based on these quantities are, respectively, the Bayes factor, the posterior Bayes factor (Aitkin, 1991) and the deviance information criterion (DIC) (Spiegelhalter *et al*., 2002).

(a) *Bayes factors*

The marginal or prior predictive density of the data given model M_i is given by

$$p(\mathbf{y}|M_i) = \int p(\mathbf{y}|\theta_i, M_i) p(\theta_i|M_i) d\theta_i = E[p(\mathbf{y}|\theta_i, M_i)]. \tag{20}$$

This density can be interpreted as the probability of obtaining the observed data under model M_i, before the data became available or as the prior mean of the likelihood. The Bayes factor for two models is the ratio between the prior means of the likelihood under each of the models

$$\begin{aligned} B_{ij} &= \frac{p(\mathbf{y}|M_i)}{p(\mathbf{y}|M_j)} = \frac{\Pr(M_i|\mathbf{y})/\Pr(M_j|\mathbf{y})}{\Pr(M_i)/\Pr(M_j)} \\ &= \frac{\text{posterior odds ratio}}{\text{prior odds ratio}} \end{aligned} \tag{21}$$

and provides a measure of whether the data have increased the odds on M_i relative to M_j. Useful reviews can be found in O'Hagan (1994) and Kass & Raftery (1995). In contrast to the two methods described below, Bayes factors have the advantage of building on a set of logical foundations that provide coherence. However, results of model comparison using the Bayes factor may be very influenced by the prior distribution. If the prior distribution accurately represents the information about θ available to the scientist prior to the experiment then, for a Bayesian, this influence should not be a matter of concern. However, the Bayes factor can give misleading inferences when vague proper prior distributions are used. In particular, an improper prior distribution for $[\theta_i|M_i]$ leads to impropriety of Eqn 20 and to pathologies of B_{ij} in Eqn 21. Partly because of these reasons, several other criteria of model comparison have been suggested in the literature and two of these, outlined below, are used in this work.

With the advent of MCMC, many methods for computing Bayes factors have been proposed in the literature. A recent comparative review is in Han & Carlin (2001). Here, we use the Monte Carlo consistent estimator proposed in Newton & Raftery (1994), which is easy to compute but not very stable numerically (the stability of the estimates is studied in Section 3iii). Obtaining more stable Monte Carlo estimates is in general not straightforward.

(b) *Posterior Bayes factors*

The posterior Bayes factor (Aitkin, 1991) for comparison of two models M_i and M_j is given by

$$B_{ij}^p = \frac{\int p(\mathbf{y}|\theta_i, M_i) p(\theta_i|\mathbf{y}, M_i) d\theta_i}{\int p(\mathbf{y}|\theta_j, M_j) p(\theta_j|\mathbf{y}, M_j) d\theta_j} = \frac{p(\mathbf{y}|\mathbf{y}, M_i)}{p(\mathbf{y}|\mathbf{y}, M_j)}, \tag{22}$$

i.e. the ratio of posterior predictive densities (19) under model M_i and M_j, respectively, evaluated at $\mathbf{y}$. Aitkin (1991) studies the frequentist properties of (22) and shows that for the case of nested models, it reduces to a general class of penalised likelihood ratio tests which includes, among others, Akaike's information criterion (Akaike, 1973). Aitkin (1991) proposes to use and interpret B_{ij}^p in the same way as the Bayes factor B_{ij}. Computation of Eqn 22 from the MCMC output is immediate by simply averaging a posterior sample of the likelihood.

(c) *Deviance information criterion*

Instead of using the posterior expectation of the likelihood as for the posterior Bayes factor, the DIC (Spiegelhalter *et al*., 2002) uses the posterior expectation of the log likelihood as a measure of model fit. For a particular model M, the DIC is defined as

$$DIC = 2\bar{D} - D(\bar{\theta}_M),$$

where

$$\begin{aligned} \bar{D} &= -2 \int \log p(\mathbf{y}|\theta_M) p(\theta_M|\mathbf{y}, M) d\theta_M \\ &= E_{\theta_{M|\mathbf{y}}, M}[D(\theta_M)], \end{aligned} \tag{23}$$

is the posterior expectation of the so-called deviance $D(\theta_M) = -2 \log p(\mathbf{y}|\theta_M)$. The second term in the right

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

hand side of (23) is the deviance evaluated at the posterior mean of the parameter vector θ_M. The DIC is obtained by adding $\bar{D}$ which measures model fit and $\bar{D}-D(\bar{\theta}_M)$ which according to Spiegelhalter *et al.* (2002) is a measure of model complexity. Models with a smaller DIC should be favoured because this indicates a better fit and a lower degree of model complexity. In common with B^p_{ij}, DIC is very easily calculated using the MCMC output.

(iii) *MCMC algorithm*

For a standard normal mixed model with homogeneous error variance and conjugate priors, a common choice of MCMC algorithm is a Gibbs sampler. In the extended model with heterogeneous variances we cannot use Gibbs updates for $\tilde{\mathbf{b}}$, $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{p}}$. Instead, we use so-called Langevin–Hastings updates (Rossky *et al.*, 1978; Besag, 1994; Roberts & Tweedie, 1997; Christensen *et al.*, 2001) combined with a reparameterization for $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{p}}$. Briefly, for a target density π and a state s, the Langevin–Hastings proposal distribution is $N(s+\frac{h}{2}\nabla(s), h\mathbf{I})$, where $\nabla(s)=\partial/\partial s \log \pi(s)$ is the gradient of the log-target density and h is a user specified proposal variance. Especially for high dimensional target distributions, the use of the gradient in the proposal distribution can lead to much better convergence properties than, for example, when the simple random walk Metropolis proposal distribution $N(s, h\mathbf{I})$ is used. For ease of programming, we also use Langevin–Hastings (and a reparameterization) for $\mathbf{a}$ and $\mathbf{p}$. The reparameterization used for $(\mathbf{a}^\top, \tilde{\mathbf{a}}^\top)$ is $(\mathbf{a}^\top, \tilde{\mathbf{a}}^\top)=L_G \otimes TD^{1/2}(\gamma^\top, \tilde{\gamma}^\top)$ where $(\gamma^\top, \tilde{\gamma}^\top)$ has a standard multivariate normal distribution, L_G is the lower-triangular Cholesky factor of $\mathbf{G}$, and T and D correspond to the factorization $\mathbf{A}=TDT^\top$ of $\mathbf{A}$ (Henderson, 1976). For $\mathbf{p}$ and $\tilde{\mathbf{p}}$, we use $\mathbf{p}=\sigma_p\delta$ and $\tilde{\mathbf{p}}=\sigma_{\tilde{p}}\tilde{\delta}$, where also $(\delta^\top, \tilde{\delta}^\top)$ is a standard normal vector. With Langevin–Hastings updates, it is easier to obtain a well-mixing MCMC algorithm for the posterior distribution of $(\gamma^\top, \tilde{\gamma}^\top)$ and $(\delta^\top, \tilde{\delta}^\top)$ than for the original variables $(\mathbf{a}^\top, \tilde{\mathbf{a}}^\top)$, $\mathbf{p}$ and $\tilde{\mathbf{p}}$. Posterior samples of $(\mathbf{a}^\top, \tilde{\mathbf{a}}^\top)$, $\mathbf{p}$ and $\tilde{\mathbf{p}}$ are simply obtained by transforming the posterior samples of $(\gamma^\top, \tilde{\gamma}^\top)$ and $(\delta^\top, \tilde{\delta}^\top)$.

The algorithm used for posterior sampling is a fixed scan hybrid Monte Carlo algorithm (also known as Metropolis-within-Gibbs) where $\mathbf{b}$, $\tilde{\mathbf{b}}$, $(\gamma^\top, \tilde{\gamma}^\top)$, $(\delta^\top, \tilde{\delta}^\top)$, $(\sigma^2_p, \sigma^2_{\tilde{p}})$, and $(\sigma^2_a, \sigma^2_{\tilde{a}}, \rho)$ are updated in turn using Gibbs for $\mathbf{b}$, Langevin–Hastings for $\tilde{\mathbf{b}}$, $(\gamma^\top, \tilde{\gamma}^\top)$ and $(\delta^\top, \tilde{\delta}^\top)$, random walk updates (on the log scale) for the variance parameters, and a random walk update for ρ.

The Langevin–Hastings updates are straightforward to program but one disadvantage is that suitable values of the proposal variances must be chosen from pilot runs of the algorithm. It is further our experience that a Langevin–Hastings update is not suitable for a multivariate full conditional distribution with very different marginal variances.

We thank H. Stern, S. L. Lauritzen and M. San Cristobal for useful discussions on posterior predictive model checking, on approaches to modelling litter size data and on genetic properties of the genetically structured residual variance model.

References

Aitkin, M. (1991). Posterior Bayes factors (with discussion). *Journal of the Royal Statistical Society Series B* **53**, 111–142.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (eds B. N. Petrov & F. Csaki), pp. 267–281. Budapest: Akademiai Kiado.

Besag, J. (1994). Contribution to the discussion paper by Grenander and Miller. *Journal of the Royal Statistical Society B* **56**, 591–592.

Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics*. Oxford University Press.

Christensen, O. F., Møller, J. & Waagepetersen, R. P. (2001). Geometric ergodicity of Metropolis–Hastings algorithms for conditional simulation in generalised linear mixed models. *Methodology and Computing in Applied Probability* **3**, 309–327.

Clayton, G. A. & Robertson, A. (1957). An experimental check on quantitative genetical theory. II. The long-term effects of selection. *Journal of Genetics* **55**, 152–170.

Draper, N. & Guttman, I. (1971). Bayesian estimation of the binomial parameter. *Technometrics* **13**, 667–673.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Longman.

Foulley, J. L. & Quaas, R. L. (1995). Heterogeneous variances in Gaussian linear mixed models. *Genetics, Selection, Evolution* **27**, 211–228.

Foulley, J. L., San Cristobal, M., Gianola, D. & Im, S. (1992). Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics and Data Analysis* **13**, 291–305.

Gavrilets, S. & Hastings, A. (1994). A quantitative genetic model for selection on developmental noise. *Evolution* **48**, 1478–1486.

Gelman, A., Meng, X. L. & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6**, 733–807.

Gianola, D., Foulley, J. L., Fernando, R. L., Henderson, C. R. & Weigel, K. A. (1992). Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *Journal of Dairy Science* **75**, 2805–2823.

Han, C. & Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes Factors: a comparative review. *Journal of the American Statistical Association* **96**, 1122–1132.

Henderson, C. R. (1976). A simple method for the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**, 69–83.

Hill, W. G. (2002). Direct effects of selection on phenotypic variability of quantitative traits. In *7th World Congress on Genetics Applied to Livestock Production*, August 19–23, Montpellier, France, Communication No. 19-02.

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

Korsgaard, I. R., Madsen, P. & Jensen, J. (1998). Bayesian inference in the semiparametric log normal model using Gibbs sampling. *Genetics, Selection, Evolution* **30**, 241–256.

Lerner, I. M. (1954). *Genetic Homeostasis*. Edinburgh: Oliver and Boyd.

Lewontin, R. C. (1964). The interaction of selection and linkage. II. Optimal model. *Genetics* **50**, 757–782.

Newton, M. A. & Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society Series B* **56**, 1–48.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics: Bayesian Inference*, vol. 2B. Edward Arnold.

Perez-Enciso, M., Tempelman, R. J. & Gianola, D. (1993). A comparison between linear and Poisson mixed models for litter size in Iberian pigs. *Livestock Production Science* **35**, 303–316.

Raftery, A. E. (1988). Inference for the binomial N parameter: a hierarchical Bayes approach. *Biometrika* **75**, 223–228.

Rendel, J. M. (1977). Canalisation in quantitative genetics. In *Proceedings of the International Conference on Quantitative Genetics* (eds E. Pollak, O. Kempthorne & T. B. Bailey), pp. 23–28. Iowa State University Press.

Robert, C. P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag.

Roberts, G. O. & Tweedie, R. L. (1997). Exponential convergence of Langevin diffusions and their approximations. *Bernoulli* **2**, 314–363.

Rossky, P. J., Doll, J. D. & Friedman, H. L. (1978). Brownian dynamics as smart Monte Carlo simulation. *Journal of Chemical Physics* **69**, 4628–4633.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* **12**, 1151–1172.

San Cristobal, M., Foulley, J. L. & Manfredi, E. (1993). Inference about multiplicative heteroscedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genetics, Selection, Evolution* **25**, 3–30.

San Cristobal-Gaudy, M., Elsen, J. M., Bodin, L. & Chevalet, C. (1998). Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genetics, Selection, Evolution* **30**, 423–451.

Sorensen, D. & Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer-Verlag.

Sorensen, D., Andersen, S., Gianola, D. & Korsgaard, I. R. (1995). Bayesian inference in threshold models using Gibbs sampling. *Genetics, Selection, Evolution* **27**, 229–249.

Sorensen, D., Vernersen, A. & Andersen, S. (2000). Bayesian analysis of response to selection: a case study using litter size in Danish Yorkshire pigs. *Genetics* **156**, 283–295.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B* **64**, 583–639.

Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution* **7**, 118–126.

Waddington, C. H. (1957). *The Strategy of the Genes*. London: Allen and Unwin.

Zhivotovsky, L. A. & Feldman, M. W. (1992). On the difference between mean and optimum of quantitative characters under selection. *Evolution* **46**, 1574–1578.

https://doi.org/10.1017/S0016672303006426 Published online by Cambridge University Press